



Vigilada Mineducación

USO DE KERNELS EN SERIES DE TIEMPO PARA LA DETECCIÓN DE
PRÁCTICAS MANIPULATIVAS EN MERCADOS FINANCIEROS

JOSE DANIEL HERRERA OCHOA

Tesis de grado

Director

OLGA LUCIA QUINTERO MONTOYA

UNIVERSIDAD EAFIT

ESCUELA DE INGENIERÍAS

MAESTRÍA EN CIENCIAS DE LOS DATOS Y LA ANALÍTICA

MEDELLÍN

2023

A todas aquellas que durante este proceso me apoyaron con su paciencia, palabras de aliento y comprensión. Especialmente en aquellos momentos de mayor debilidad.

AGRADECIMIENTOS

A todo el equipo docente de la Universidad que facilitó y permitió mi proceso formativo, especialmente a Olga Lucía Quintero Montoya por su paciencia y vocación, las cuales fueron claves en el desarrollo de la presente propuesta.

CONTENIDO

| | |
|--|----|
| INTRODUCCIÓN | 11 |
| PLANTEAMIENTO DEL PROBLEMA..... | 12 |
| JUSTIFICACIÓN..... | 13 |
| OBJETIVOS..... | 15 |
| GENERAL | 15 |
| ESPECÍFICOS | 15 |
| MARCO TEÓRICO O MARCO CONCEPTUAL..... | 16 |
| DISEÑO METODOLÓGICO | 23 |
| Plan de trabajo..... | 25 |
| Equipo y roles | 25 |
| Presupuesto..... | 25 |
| Productos esperados..... | 25 |
| Plan de gestión de datos | 26 |
| Aspectos Éticos | 26 |
| DESARROLLO | 27 |
| FÉNOMENOS DE MANIPULACIÓN EN MERCADOS FINANCIEROS | 27 |
| <i>Pump and dump</i> | 28 |
| <i>Wash trade</i> | 28 |
| <i>Poop and scoop</i> | 28 |
| <i>Spoofing</i> | 29 |
| <i>Layering</i> | 29 |
| HERRAMIENTAS PARA EL ANÁLISIS DE SEÑALES | 29 |
| Espectro de potencia | 32 |
| Espectrograma | 36 |
| Escalograma..... | 39 |
| MANIPULACIONES FINANCIERAS USANDO KERNELS DE ANÁLISIS DE SEÑALES | 44 |
| Espectro de potencia | 53 |
| Espectrograma | 55 |
| Espectrograma | 61 |

| | |
|--|-----|
| VENTANEO DE LAS SEÑALES | 72 |
| AJUSTE Y SELECCIÓN DE PARÁMETROS PARA LOS KERNEL | 83 |
| RECONOCIMIENTO DE PRÁCTICAS MANIPULATIVAS USANDO MÁQUINAS DE APRENDIZAJE | 93 |
| Acercamientos con Aprendizaje de Máquina..... | 104 |
| Selección de Características | 109 |
| Exploración de modelos de clasificación | 119 |
| Entrenamiento de modelos..... | 131 |
| Evaluación de modelos..... | 134 |
| Consideraciones Finales..... | 147 |
| CONCLUSIONES | 150 |
| REFERENCIAS | 154 |
| ANEXOS..... | 165 |
| ANEXO A – FUENTES Y <i>SCRIPTS</i> | 165 |
| ANEXO B – SELECCIÓN DE CLÚSTERES | 166 |
| ANEXO C – SELECCIÓN INICIAL DE CARACTERÍSTICAS..... | 167 |
| ANEXO D – SELECCIÓN final DE CARACTERÍSTICAS | 168 |

LISTA DE FIGURAS

| | |
|---|----|
| Figura 1. Framework para el aprendizaje de máquina mediante la aplicación de kernel..... | 24 |
| Figura 2. Representación en el dominio temporal de una señal compleja..... | 31 |
| Figura 3. Representación de una discontinuidad en una señal..... | 33 |
| Figura 4. Resolución Espectral..... | 34 |
| Figura 5. Representación gráfica del espectro de potencia..... | 35 |
| Figura 6. Representación gráfica de un espectrograma..... | 36 |
| Figura 7. Estimación de los segmentos de un espectrograma..... | 38 |
| Figura 8. Segmentos resultantes luego de solapar muestras..... | 39 |
| Figura 9. Representación gráfica del proceso de cálculo de un escalograma..... | 41 |
| Figura 10. Representación gráfica de la CWT mediante un escalograma..... | 42 |
| Figura 11. Espectro de Potencia de una acción con manipulación tipo pump and dump..... | 48 |
| Figura 12. Espectrograma de una acción con manipulación tipo pump and dump con ventana de tiempo más corta..... | 49 |
| Figura 13. Escalograma de una acción con manipulación tipo pump and dump con ventana temporal a un año..... | 50 |
| Figura 14. Representación de la ventana de Kaiser con diversos valores de Beta..... | 55 |
| Figura 15. Espectrograma basado en STFT de una acción con manipulación tipo wash trade con ventana de 3 días..... | 57 |
| Figura 16. Espectrograma basado en FFT de una acción con manipulación tipo wash trade con ventana de 3 días..... | 58 |
| Figura 17. Diversos porcentajes de solapamiento para las ventanas..... | 60 |
| Figura 18. Comparación del espectrograma generado mediante CWT y la DWT..... | 63 |
| Figura 19. Representación de la capacidad de resolución y el soporte frecuencial de las wavelets analíticas..... | 65 |
| Figura 20. Wavelets Analíticas Recomendadas..... | 66 |

| | |
|--|-----|
| Figura 21. Detalle de como la Wavelet Morse mantiene un buen balance entre resolución frecuencial y temporal..... | 68 |
| Figura 22. Comportamiento de la wavelet Morse con diversos parámetros | 70 |
| Figura 23. Espectro de potencia de una acción con manipulación tipo pump and dump en diversas ventanas de volatilidad | 83 |
| Figura 24. Espectrograma basado en STFT y FFT de una acción con manipulación tipo pump and dump con volatilidad de 30 días | 85 |
| Figura 25. Escalograma basado en CWT de una acción con manipulación tipo wash trade con diversos valores de Beta | 89 |
| Figura 26. Escalograma basado en CWT de una acción con manipulación tipo wash trade con diversos valores Gamma | 90 |
| Figura 27. Representación del filtrado de eventos reflectivos en una señal | 95 |
| Figura 28. Representación de las diversas vistas de un escalograma | 97 |
| Figura 29. Representación de clústeres mediante k-medoids | 106 |
| Figura 30. Grafo de k-distancias para la estimación clústeres en DBSCAN..... | 107 |
| Figura 31. Clústeres por agrupamiento espectral mediante la matriz de Laplace | 108 |
| Figura 32. Proporción de la varianza por característica | 115 |
| Figura 33. Análisis PCA sobre las características frecuenciales..... | 118 |
| Figura 34. Ejemplo de árbol de clasificación..... | 121 |
| Figura 35. Ejemplo de árbol de clasificación..... | 127 |
| Figura 36. Ejemplo de k-fold cross-validation | 133 |
| Figura 37. Matriz de confusión y estadísticos de desempeño | 135 |
| Figura 38. Ejemplo de curva ROC para modelo KNN Fino ($k = 1$)..... | 137 |
| Figura 39. Ejemplo del proceso de optimización de Bayes..... | 142 |
| Figura 40. Gráfico de error de clasificación | 143 |
| Figura 41. Selección de predictores mediante algoritmo ReliefF | 145 |
| Figura 42. Curva ROC para modelo KNN Optimizado..... | 146 |
| Figura 43. Representación de los clasificadores más relevantes para el modelo KNN Optimizado..... | 147 |

LISTA DE TABLAS

| | |
|--|-----|
| Tabla 1. Análisis de diversos modelos de clasificación..... | 120 |
| Tabla 2. Desempeño de los modelos de clasificación | 138 |
| Tabla 3. Selección de las características en modelo KNN mediante el algoritmo “ReliefF” | 144 |

RESUMEN

De forma intuitiva se podría pensar que cualquier desviación en los datos de las negociaciones de los activos financieros podría ser fácilmente detectada debido a la base estadística en la que se fundamentan las ciencias de las finanzas. Sin embargo, los mercados en los que se negocian los activos financieros operan bajo el principio de oferta y demanda, como también bajo el principio de oportunidad. Elementos que los hacen muy susceptible a la manipulación de precios.

Por tal motivo, cada vez es más relevante considerar técnicas que permitan identificar elementos en una serie de datos financiera que faciliten la búsqueda de patrones o de artefactos que den cuenta si una acción ha sido manipulada o no.

Se propone entonces el uso de kernels para la descomposición y el filtrado de señales en series de tiempo, ya que mediante el uso de esta técnica se pueden obtener elementos de la serie como el poder y la frecuencia, que posteriormente pueden facilitar la caracterización de una acción que ha sido objeto de operaciones fraudulentas o manipulativas, y mediante diversas técnicas de aprendizaje de máquina lograr una detección más oportuna a partir de dicha caracterización; particularmente en entornos de negociación dinámicos y en constante evolución.

Para tal fin, se contrastará el desempeño de los kernels frente a técnicas tradicionales, eligiendo los más adecuados. De igual manera, se evaluarán diversas técnicas de aprendizaje de máquina y se elegirá aquella que mejor aprenda y represente los patrones o artefactos en las operaciones fraudulentas. Procurando de esta manera la elevación de los estándares de negociación en los mercados financieros, como también profundizando en las aplicaciones que la descomposición y el filtrado de señales con kernels puede llegar a tener, no solo como herramienta de visualización de datos, sino también como insumos para técnicas de aprendizaje de máquina.

Palabras clave: análisis de señales; series de tiempo; descomposición de señales; kernels

ABSTRACT

Intuitively, one might think that any deviation in trading data could be easily detected due to the statistical basis on which finance sciences are based. However, the markets in which financial assets are traded operate under the principle of supply and demand, as well as the principle of opportunity. Elements that make them very susceptible to price manipulation.

For this reason, it is increasingly relevant to consider techniques that allow the identification of elements in financial time series that can deliver information that show whether a stock has been subject of manipulative practices or not.

The use of kernels for signals decomposition and filtering in financial time series is then proposed. By using this technique elements of the time series such as power and frequency can be obtained, which can later facilitate the characterization of a stock that has been subject of fraudulent or manipulative trading. Then considering diverse machine learning techniques, achieve a timelier detection based on said characterization, particularly in dynamic and constantly evolving trading environments.

For this purpose, the performance of the kernels will be contrasted against traditional techniques, choosing the most appropriate ones. In the same way, various machine learning techniques will be evaluated and the one that best learns and represents the patterns or artifacts in fraudulent operations will be chosen. Trying in this way to raise trading standards in financial markets, as well as delving into the applications that the decomposition and filtering of signals with kernels can have, not only as a data visualization tool, but also as inputs. for machine learning techniques.

Keywords: signal analysis; time series; signal decomposition; kernels

INTRODUCCIÓN

Las prácticas manipulativas en los mercados financieros tienen como propósito principal alterar de forma intencional y fraudulenta el precio, el volumen o la tendencia de un activo financiero para obtener un beneficio particular. (AMV, 2022)

Dichas prácticas afectan la transparencia, la eficiencia y la estabilidad de los mercados, así como la confianza de los inversionistas y el bienestar social. Se hace necesario el desarrollo de estrategias que permitan identificar y caracterizar las prácticas manipulativas que se presentan en los mercados bursátiles, con el fin de emitir las sanciones respectivas y prevenir la desestabilización de los mercados. (AMV, 2022)

El objetivo general de esta investigación es analizar los diferentes tipos y categorías de manipulación de mercados, así como algunos ejemplos de las prácticas más comunes, con el fin de proponer un modelo para detectar y describir estas conductas en los mercados financieros. Se busca dar respuesta a lo siguiente:

- ¿Qué es son las manipulaciones de mercado y cómo se clasifican?
- ¿Qué herramientas se emplean para el procesamiento de señales financieras?
- ¿Qué métodos existen para describir y detectar las manipulaciones de mercado?
- ¿Qué tipo de modelo se puede aplicar para detectar y hacer escalable la detección de manipulaciones en activos financieros?

Se considera una metodología de investigación mixta, combinando técnicas cualitativas y cuantitativas. Se parte de una revisión bibliográfica, en la que se define el concepto, las tipologías y los efectos de las manipulaciones en los mercados, así como también los métodos existentes para su detección y medición.

Se emplean datos históricos de las transacciones realizadas en sistemas de negociación estadounidense, ya que existe información de referencia suficiente para validar modelos que faciliten la detección de prácticas manipulativas.

A priori se favorecen las técnicas derivadas en aprendizaje de máquina, ya que su implementación ha demostrado la escalabilidad en diversos problemas de detección y clasificación de señales como será descrito más adelante.

En un inicio se presenta el marco teórico sobre la manipulación de mercado, sus tipos, categorías y ejemplos. Luego se describen las herramientas empleadas para detectar y medir la manipulación de mercado. A continuación, se presentan los resultados obtenidos mediante las diversas herramientas aplicadas a los datos. Posteriormente, se discuten las implicaciones y las limitaciones del estudio. Finalmente, en el quinto capítulo se exponen las conclusiones y las recomendaciones derivadas de la investigación.

PLANTEAMIENTO DEL PROBLEMA

Diversas técnicas de aprendizaje de máquina han demostrado la mejora del rendimiento en gran variedad de tareas comunes por un gran margen, siendo el campo de las finanzas una de las áreas del conocimiento más favorecidas con dichas aplicaciones (Golmohammadi et al., 2012). Lo anterior, se debe a que las finanzas al igual que la gran mayoría de técnicas de aprendizaje de máquina tienen una fuerte fundamentación estadística (Ngai et al., 2011).

Se podría pensar que cualquier alteración en los precios de los activos financieros debería ser fácilmente detectable, pues la negociación de este tipo de activos y su posterior valoración tienen como premisa las relaciones estadísticas entre los datos que describen dichos activos.

Sin embargo, en casos de fraude o manipulaciones de mercado, los defraudadores procuran al máximo disfrazar sus operaciones de legalidad para que éstas pasen desapercibidos frente a los mecanismos de detección implementados por los entes de control (Lin, 2016).

Detectar prácticas manipulativas es fundamental, pues es premisa de los mercados garantizar ambientes de negociación justos para todos y que se garantice la protección de los intereses de todos los inversionistas (Khanna & Sonti, 2014).

Sin embargo, pese a los aportes que el aprendizaje de máquina ha ofrecido en el campo de las finanzas, entre los mecanismos de detección actuales aún predominan las implementaciones basadas en reglas duras, y el uso de técnicas de aprendizaje de máquina aún no es generalizado. Lo que deja una gran brecha, pues constantemente se presentan nuevos esquemas de defraudación, particularmente gracias al auge de las negociaciones de alta frecuencia.

Adicionalmente, es importante destacar que la cantidad de operaciones que pueden ser realizadas en una fracción de tiempo es tan abrumadora, que emplear reglas duras para la detección de conductas manipulativas resulta inviable, ya que los recursos y la tecnología existentes son insuficientes para lograr un monitoreo efectivo (Aksenov et al., 2020). No hay manera de realizar comparaciones, ni de validar esquemas previamente identificados debido a los altos volúmenes de negociación.

Por otra parte, debe considerarse, que cada tipo de manipulación de mercado tiene sus propias características, y resulta obvio que puede ser complejo desarrollar un modelo que permita detectar todo tipo de manipulaciones (Aksenov et al., 2020).

Por lo anterior, se vuelve relevante considerar técnicas que permitan identificar y aprender elementos que den cuenta de patrones o artefactos inusuales en operaciones cuyo fin sea defraudar o manipular los mercados financieros, especialmente no solo en eventos puntuales, sino de manera sostenida en el tiempo.

JUSTIFICACIÓN

Contar con técnicas más eficaces para la detección de prácticas manipulativas de mercados, siempre será relevante, pues los reguladores tienen como premisa el aseguramiento de los más altos estándares de negociación, con el fin de facilitar y garantizar las mejores condiciones para cada uno de los actores que intervienen. (Lin, 2016)

Como fue mencionado previamente, diversas técnicas de aprendizaje de máquina han demostrado la mejora del rendimiento en gran variedad de tareas comunes por un gran margen.

Sin embargo, para lograr una detección eficaz de prácticas manipulativas en los mercados empleando técnicas de aprendizaje de máquina hay varios desafíos que enfrentar. Primero, los sets de datos son muy grandes y heterogéneos. Segundo, los patrones anormales no vienen etiquetados, por lo que la cantidad de información disponible para entrenamiento es limitada, (considerando el hecho de que existen diversas estrategias en constante evolución). Tercero, contar con sets de datos etiquetados hace que los modelos de detección estén sesgados hacia los sets de datos de entrenamiento. Por último, concentrarse en un esquema de manipulación específico y el uso de ciertos parámetros particulares, hace que la detección esté sesgada hacia un esquema manipulativo en particular. Por tanto, el aprendizaje no supervisado parece ser el enfoque más pertinente para superar tales desafíos y lograr una detección más eficaz. (Abbas et al., 2018; Rizvi et al., 2020b)

No obstante, en los últimos años, la descomposición de señales en series de tiempo mediante kernels ha demostrado ser una técnica eficiente para la detección de señales anómalas particularmente en aquellas de alta frecuencia. Debido a que estas técnicas están diseñadas para extraer componentes y características de las series de tiempo al aplicar funciones de descomposición que permiten ampliar la información disponible al llevar los datos a altas dimensiones. (De Vitis, 2019)

Teniendo los datos en altas dimensiones, se pueden separar los diversos elementos que componen una señal, entre ellos los coeficientes que representan los elementos de alta y baja frecuencia de una señal.

A partir de los coeficientes resultantes se pueden aplicar técnicas de filtrado para reducir la cantidad de información redundante y extraer la información relevante (Rizvi et al., 2020a). Dichas técnicas de filtrado a su vez también pueden ser empleadas para la detección de datos anómalos al definir umbrales en los que los componentes de las señales pueden ser agrupados (Abbas et al., 2018).

Resulta propicio entonces considerar dichas técnicas, pues con el surgimiento del trading algorítmico, la velocidad en las que las negociaciones y las transacciones puede ser realizadas hoy en día no tienen precedente, por lo que contar con técnicas que den cuenta si un activo financiero ha sido objeto de manipulaciones,

siempre será relevante sobre todo cuando son billones de dólares los que pueden ser negociados en cuestión de segundos. (Rizvi et al., 2020a)

Las manipulaciones sobre activos financieros han puesto en apuros a los reguladores a nivel mundial, sobre todo en mercados emergentes, donde la falta de recursos económicos y tecnológicos, como de un ambiente regulatorio fuerte han hecho que sea muy complejo para los entes de control identificar y definir con precisión nuevos esquemas manipulativos. (Ergün et al., 2021)

Se ha realizado gran variedad de estudios empíricos con base en la aplicación de técnicas de aprendizaje de máquina, clamando una mejoría en las capacidades de detección de prácticas manipulativas en los mercados financieros. Sin embargo, la mayoría de las técnicas de aprendizaje de máquina empleadas se concentran en eventos de manipulación puntuales, dependen de información precisamente etiquetada (la cual no es fácil de conseguir y podría resultar costoso de hacer) o que han sido entrenados con información muy específica y casi siempre ajustada. (Rizvi et al., 2020b)

Por lo tanto, evaluar qué tan adecuado resulta la aplicación de un kernel para descomposición de señales, y de técnicas de aprendizaje de máquina en la caracterización y posterior detección de señales anómalas será pertinente, pues además de ir en línea con el propósito de los reguladores de mejorar en la mayor medida los estándares de negociación en los mercados; estas técnicas han probado ser eficientes y mejorar el resultado en la detección de anomalías. (Ngai et al., 2011)

OBJETIVOS

GENERAL

Evaluar la pertinencia de los kernel en series de tiempo para la detección de prácticas manipulativas en los mercados financieros con el fin de que se logre la elevación de los estándares de negociación y la ampliación de los campos de aplicación de dichas técnicas.

ESPECÍFICOS

- Realizar un diagnóstico de las técnicas más empleadas en la detección de prácticas manipulativas de mercado para el establecimiento de una línea base de comparación de la propuesta.
- Implementar los modelos de referencia seleccionados para la validación posterior de la eficacia del modelo propuesto.
- Implementar el modelo propuesto considerando los kernel más adecuados en la descomposición de señales y en la detección de señales anómalas.
- Definir las métricas que permitan la evaluación de la eficacia del modelo propuesto versus las técnicas más empleadas en la detección de prácticas manipulativas.

MARCO TEÓRICO O MARCO CONCEPTUAL

Los mercados financieros están soportados sobre plataformas robustas capaces de procesar millones de transacciones en un día, y miles en pequeñas fracciones de tiempo (Sridhar, 2020).

Estas transacciones son hechas por compradores y vendedores que publican sus intenciones de negociación a través de sistemas de negociación bursátil.

El cruce de dichas intenciones se da sin importar la identidad de las partes en dicha transacción, ya que la premisa bajo la que operan dichos mercados es la transparencia, el intercambio justo, la libre concurrencia y el principio de oportunidad (Decreto 1802, 2007).

Sin embargo, existen personas inescrupulosas que desean afectar los mercados y mover el precio de un activo financiero a su favor. Dichas afectaciones se conocen como “manipulación de mercado”. Las manipulaciones más allá de favorecer un interés personal, tiene otros efectos negativos que pueden llegar a tener consecuencias y efectos muy severos en los mercados.

Si bien los esquemas de manipulación de precios son los más comunes y los más extendidos, las prácticas manipulativas de mercado pueden ser categorizadas en tres tipos: Manipulación basada en Información, Manipulación basada en Hechos y Manipulación basada en Transacciones (Allen & Gale, 1992).

Las manipulaciones basadas en información ocurren cuando las personas de forma intencional divulgan de manera mal intencionada rumores o información falsa sobre el desempeño de una compañía para influenciar los precios bajo los que se negocian sus acciones.

Las manipulaciones basadas en hechos corresponden a decisiones tomadas por los administradores o ejecutivos de una compañía. Su intención es crear de forma artificiosa condiciones que alteren el potencial percibido por su compañía y por consiguiente su desempeño en bolsa. El ejemplo más común es la restricción de la oferta para aumentar los precios de sus productos y de esta manera aumentar el valor percibido de la compañía, aumentando su precio en bolsa.

Por último, se tienen las manipulaciones basadas en transacciones, éstas son propias de los mercados bursátiles y tienen relación directa con la compraventa de los activos financieros.

Esta última será el foco de la presente propuesta de investigación, pues corresponde al tipo de manipulación más común, la que mayor impacto ha tenido en los mercados ya que sus resultados pueden percibirse de inmediato y a su vez la que en teoría puede disfrazarse con mayor facilidad. En consecuencia, la cantidad de tipologías de este estilo es amplia.

Generalmente las manipulaciones basadas en transacciones son agrupadas en las siguientes categorías: Reversión de Tendencias, Manipulación del Volumen y Manipulación de Precios (Aitken et al., 2009).

La reversión de precios corresponde a los cambios súbitos y artificiosos en la tendencia de una acción. La manipulación del volumen corresponde a la afectación de este atributo para crear pánico o interés entre los inversionistas al presentar activos con niveles alarmantes o atractivos de negociación. Por último, la manipulación de precios corresponde a la interferencia de forma artificiosa de este atributo a través de operaciones malintencionadas o a la falsa intención de compraventa.

Como previamente se mencionó, la cantidad de tipologías de este tipo manipulación es amplia, sin embargo, entre las más comunes se tiene: *spoofing*, *pump & dump*, *quote stuffing*, *price fixing* y *high closing* (Aitken et al., 2009).

Todas son definidas a partir de anglicismos, ya que fueron descritas inicialmente en los mercados bursátiles estadounidenses. Sin embargo, a pesar de la variedad de esquemas y nombres llamativos, en la generalidad todas parten de la premisa de emplear medios artificiosos y contrarios a los principios de los mercados bursátiles para subir o bajar el precio de un activo.

Una vez el precio deseado fue logrado, el defraudador opera de forma contraria para beneficiarse del movimiento artificial en el precio.

Detectar dichas manipulaciones ha ganado gran interés recientemente, pues con los avances en las tecnologías de la información y en las capacidades de cómputo, se han dado grandes cambios en la naturaleza y el desempeño de los mercados financieros.

El surgimiento del trading de alta frecuencia ha acelerado el proceso de las negociaciones y ha facilitado el surgimiento y la consolidación de prácticas manipulativas (Aksenov et al., 2020).

Por fortuna en la misma vía, el advenimiento de dichas tecnologías y capacidades también ha facilitado la aplicación de diversas técnicas, capaces de hacer frente a los nuevos retos que los regulares vienen enfrentando para garantizar mercados transparentes y eficientes, proteger los intereses de todos los inversionistas y elevar los estándares de negociación.

Wang et al., (2019) ponen en manifiesto como especialmente en años recientes aquellas técnicas basadas en aprendizaje automático han demostrado mejorar los resultados en detección de anomalías particularmente aquellas basadas en clasificación y han logrado resolver problemas complejos en el tratamiento de información en temas de fraude financiero.

Entre los trabajos pioneros en esta materia tenemos la propuesta de Ferdousi & Maeda (2006), en el que proponen la búsqueda de datos anómalos en series de

tiempo a partir del análisis PGA (*'Peer Group Analysis'*). Si bien fue una primera aproximación muy interesante, queda en manifiesto como muchas técnicas que dependen de etiquetado de los sets de datos de entrenamiento, realmente pueden no ofrecer los mejores resultados y por el contrario puede ser costoso de implementar.

Por su parte Ögüt et al., (2009) consideraron como base para su propuesta de validación en la manipulación de precios, ANN (*'Artificial Neural Network'*) y SVM (*'Support Vector Machine'*), entre sus hallazgos encontraron que los modelos de minería de datos considerados se comportan mejor que las técnicas estadísticas tradicionales. Al respecto, Qu & Zhang (2016), confirman como los modelos lineales de clasificación, ofrecen resultados muy pobres si se validan contra modelos no lineales como GARCH (*'Generalized Autoregressive Conditional Heteroskedasticity'*), ANN y SVR (*'Support Vector Machine for Regression'*).

Song et al., (2012) propusieron técnicas basadas en CHMM (*'Coupled Hidden Markov Model'*), si bien lograron presentar resultados interesantes al respecto, es claro como este tipo de técnicas depende de un conocimiento avanzado del dominio de los datos, lo que hace que los resultados sean susceptibles a análisis ambiguos y que haya una alta probabilidad de falsos positivos, lo mismo ocurre con el método propuesto por Golmohammadi & Zaiane (2015), si bien sus resultados implementando CAD (*'Contextual Anomaly Detection'*), logran superar los resultados comparados contra kNN (*'k-Nearest Neighbor'*), la cantidad de falsos positivos es alta, pues generaliza el comportamiento de una serie respecto a un grupo.

Oregi et al., (2018) propusieron métodos basados en AML (*'Adversarial Machine Learning'*) los cuales probaron funcionar en casos muy específicos en los que deben cumplirse ciertas condiciones, las cuales fueron ajustadas como parte de la implementación de la propuesta. En general trabajos derivados de estos métodos comparten este tipo de conclusiones, lo que tal vez no los hace muy adecuados, pues entre otras cosas requiere de sets de datos previamente etiquetado; al respecto Leangarun et al., (2018) consideraron LSTM (*'Long Short-Term Memory'*) para estructurar bases para la generación de GANs (*'Generative Adversarial Networks'*) que no dependen de etiquetado previo, sin embargo, los resultados no fueron muy prometedores y proponen que considerar sets de datos más amplio u otro tipo de modelos podría mejorar los resultados.

Abbas et al., (2018) proponen un método basado en kernel de densidad, el cual demostró ser muy efectivo en la detección de manipulaciones considerando para ello dos tipos de patrones, estos resultados no son generalizables, al respecto Rizvi et al., (2020) lograron una mejora en este tipo de técnicas al generalizarlas, sus resultados fueron basados en set de datos manipulados, lo que puede comprometer los resultados, sin embargo proponen como evaluar otro tipo de funciones kernel para la descomposición y para el agrupamiento, podrían mejorar los resultados obtenidos.

En consideración, se propone ampliar sus resultados, teniendo en cuenta el uso de set de datos que no hayan sido previamente manipulados y la búsqueda de kernel que se ajusten de mejor manera a series de tiempo de activos financieros y de detección de datos anómalos.

Es importante mencionar que el interés en la aplicación de kernel para la detección de anomalías, no solo parte de ser una de las técnicas más recientes, sino también porque esta técnica ha mostrado ser eficiente en otro tipo de aplicaciones.

Al respecto, Rizvi et al., (2020) proponen como consideración adicional, el hecho de que las técnicas basadas en kernel, no dependen de set de datos muy específicos, lo que permiten que estén en capacidad de adaptarse en la detección de diversos esquemas manipulativos, lo que implica que sus resultados son generalizables independiente de las series de datos que se consideren.

Los kernel son funciones definidas positivas, que tienen como premisa el mapeo de un set de datos dado a un espacio de diferentes dimensiones (Hofmann et al., 2008). Se consideran como parte de los algoritmos empleados para el análisis y el reconocimiento de patrones, ya que en las dimensiones resultantes de la aplicación de un kernel pueden aplicarse diversas técnicas para la clasificación de los datos o la selección de características.

Es importante tener en consideración que para tales tareas otros tipos de kernel pueden ser aplicados, en este caso las funciones de transformación dadas definirán umbrales en el espacio de características generado que permita esta aplicación.

Los espacios de características corresponden a espacios n-dimensionales en el que un set de datos o un conjunto de señales existe. Lo anterior implica que el tratamiento de una señal o un set de datos debe ser realizado dentro de dicho espacio, ya que la señal hereda la dimensión de dicho espacio (De Vitis, 2019).

Lo anterior puede entenderse de manera sencilla viendo cada elemento del conjunto de datos como un punto en dicho espacio y las dimensiones como los atributos o características de tales puntos.

La intuición anterior es clave, pues las series de tiempo pueden entenderse como señales en dos dimensiones (tiempo-escala), que según el problema a tratar pueden ser multivariadas al considerar atributos adicionales que complementen la información de la señal primaria.

Sin embargo, en ocasiones la información a procesar puede no ser suficiente (incluso en el caso de las series multivariadas). Conviene considerar técnicas de preprocesamiento que permitan la generación de nuevas características derivadas de la combinación o la descomposición de las existentes para que el problema sea tratable (Hofmann et al., 2008).

En este punto es fundamental considerar que los espacios generados, aún con un amplio conocimiento del dominio del problema, pueden no ser interpretables, sin

embargo, los espacios n-dimensionales son derivados de los datos en bajas dimensiones y, por tanto, los análisis resultantes en el espacio n-dimensional se pueden interpolar a bajas dimensiones, y por tanto comprenderse de mejor manera (De Vitis, 2019).

Al respecto es necesario precisar que, al generarse espacios de mayores dimensiones, puede haber información que no aporte o que la dimensión obtenida impida tener una solución viable.

Por tanto, se sugiere sean aplicadas técnicas complementarias de preprocesamiento que disminuyan las dimensiones (De Vitis, 2019).

En el caso de las señales, incluso para reducción de dimensiones pueden considerarse técnicas de filtrado, las cuales son muy comunes particularmente en señales de audio.

Es importante resaltar que la aplicación de los conceptos anteriores se da en gran medida en la denominada 'matriz kernel'. La cual define las transformaciones que mapean el espacio original en el nuevo, es decir define como las operaciones en dicho espacio son realizadas.

Formalmente, es una matriz simétrica, semidefinida positiva, y esto es clave, pues es la base sobre la cual técnicas como PCA (*'Principal Component Analysis'*) pueden ser aplicadas (De Vitis, 2019), como también los procesos anteriormente descritos y que se conoce como "el truco kernel" y parte de un teorema que define que es posible realizar operaciones en el espacio de características de altas dimensiones implícito, sin la necesidad de mapear las coordenadas de los datos en dicho espacio, sino simplemente operar el producto punto entre todas las imágenes de todos los pares de datos en el espacio de características (Rizvi et al., 2020).

Considerando la noción de kernel expuesta anteriormente, queda en manifiesto la versatilidad de este tipo de funciones y no es de extrañar como pueden ser empleados no solo con múltiples aplicaciones, sino también como piezas claves en diversos pasos del análisis de señales, especialmente cuando la definición de dichos kernel se complementa con otros conceptos asociados a la transformación, el filtrado y la descomposición (Hofmann et al., 2008).

Al respecto, se puede considerar como, un kernel de convolución, es una matriz kernel que se mueve por el espacio, es decir que puede desplazarse por los elementos de una serie de tiempo y a partir de ventanas procesar los elementos de la serie y de esta manera evaluar o hallar patrones.

Una convolución es una integral que expresa la cantidad de 'solapamiento' que una función tiene sobre otra en la medida que ambas son cruzadas, y de esta manera 'combinadas' (Bracewell, 1986).

El caso más básico corresponde a la transformada de Fourier de una dimensión.

La transformada de Fourier permite caracterizar los componentes de frecuencia de una función y como contribuyen a los valores asociados a sus transformadas.

Lo anterior es útil en el caso de que, si la traducción hecha por un kernel es invariante, su transformada de Fourier facilitará el análisis con base en los componentes de frecuencia que puede extraer de una señal (Hofmann et al., 2008).

De manera sencilla, la transformada de Fourier aplica un kernel (que puede ser aplicado sobre otro kernel) para llevar una señal al espacio de frecuencias w por medio de una convolución.

En el caso de las series de tiempo se recomiendan las transformadas bidimensionales, pues estas se presentan en los espacios de tiempo-escala, lo que finalmente facilita encontrar la distribución espacial de una señal. Hecho que implica que se obtendrá más información al considerar también la dimensión de la escala y no solo la de frecuencia, como es el caso de la transformada de Fourier (Quintero-Montoya, 2019).

El anterior concepto introduce la noción de la Transformada Wavelet. Este tipo de técnicas permite que una señal sea analizada en términos del tiempo y la escala. Siendo la escala inversamente proporcional a la frecuencia e imita el proceso de la Transformada de Fourier, teniendo en cuenta que el mapeo generado por esta función permite tener una mejor representación de características locales dentro de la serie de tiempo (Byrne, 2014).

Por medio de la Transformada Wavelet, las señales son descompuestas por paquetes localizados por tiempo y frecuencia, ya que esta transformada puede entenderse como el producto interno de dos funciones: la función base y la transformada en términos de escala-tiempo considerando los valores descritos por la función base (Byrne, 2014).

Campo, Quintero-Montoya & Bastidas (2016), describen las wavelets como señales (o forma de una onda) de duración limitada y cuyo valor medio es cero. Suelen ser irregulares y asimétricas y existen tres tipos: discretas, continuas y estacionarias. Dichos tipos no solo varían en las bases matemáticas en las que se soportan, sino también en la cantidad de información que representan, los dominios que consideran para dicha representación y finalmente en sus aplicaciones.

Brevemente la Transformada Wavelet Discreta, puede ser interpretada como la implementación de dos filtros. En primera instancia un filtro de banda baja encargado de obtener las bajas frecuencias y en segunda instancia, un filtro de banda alta encargado de obtener las altas frecuencias en una señal. Este submuestreo puede realizarse de forma sucesiva sobre los elementos obtenidos por el filtro de banda ancha y de esta manera identificar los elementos principales presentes en una señal, y a su vez aumentar la dimensión de los datos.

Por su parte, la Transformada Wavelet Continua, consiste en la definición de una función madre sobre la que se aplican funciones de dilación y cambio de fase para

analizar la señal en todo su espectro y tener una representación más precisa de una señal.

Por último, se tiene la transformada wavelet estacionaria, que corresponde a una modificación de la transformada wavelet discreta en la que no ocurre submuestreo, al emplear esquemas de rellenado cero con base en operadores que alternan las secuencias de sustitución entre los componentes de banda alta y los de banda baja. Se emplea generalmente en casos en los que se no es fácil obtener características de una señal (Quintero-Montoya, 2018).

Sin importar, el tipo de wavelet que se implemente la dimensión de una señal incrementará, sin embargo, en el caso de señales dentro del dominio temporal como las señales económicas o electrónicas no tienen características obvias que resulten al incrementar dimensiones.

Resulta fundamental considerar el dominio en el que desea evaluarse la señal, siendo relativamente útil el dominio de escala que ofrece la transformada Wavelet (Quintero-Montoya, 2018).

Es necesario notar que considerar la cantidad de información disponible y el dominio, no solo aplica en el caso de la elección del tipo de Wavelet, sino también en el tipo de técnica que permita la selección de las características más relevantes o que mejor describan el fenómeno dentro del nuevo espacio mapeado para la señal.

Al respecto no hay un libro de recetas que asegure que un tipo de transformada sea más recomendado en un campo de aplicación específico, (aunque si para su dominio, ya sea temporal, frecuencial o espacial) pues en última instancia es el tipo de datos y el tipo de análisis que se hará sobre los mismos los que darán la pauta del tipo de transformada a ser elegida.

Sin embargo, es necesario precisar que en las aplicaciones económicas la cantidad de energía e información no siempre es fácilmente representable (Quintero-Montoya, 2019)

Como se mencionó previamente estas señales no tienen características obvias que resulten de incrementar dimensiones. Por tanto, a pesar de que en ciertos casos de aplicación los kernel por si solos están en la capacidad de describir patrones o identificar artefactos en una señal, resulta fundamental y necesario considerar como parte de la arquitectura de la solución formas de aprendizaje que faciliten la tarea de detección de prácticas manipulativas, sin depender de una lista de reglas de comportamiento como ocurre actualmente y que limitan la tarea de los reguladores (Abbas et al., 2018).

DISEÑO METODOLÓGICO

No existe un libro de recetas al emplear kernels en series de tiempo, que dé cuenta del tipo de transformada que debería ser considerada en cada aplicación particular, sin embargo, es precisamente parte del alcance de esta propuesta, determinar los kernel más adecuados para análisis de datos atípicos y patrones inusuales en series de tiempo de activos financieros.

La premisa para considerar este tipo de técnicas es que la secuencia de cambios bruscos en el comportamiento de los precios de un activo financiero dará cuenta si hubo una manipulación o no. Para tal fin, se tomarán series de tiempo en la que se conoce hubo manipulaciones y se hará el análisis de extracción de características, selección de características y la clasificación de los puntos obtenidos. A partir de dichos análisis serán seleccionados los kernel más adecuados para este tipo de datos y este tipo de problemas. Para tal validación se tendrán set de datos de entrenamiento, prueba y validación.

Posteriormente, se consideran diversas técnicas de aprendizaje de máquina que mejor caractericen y aprendan los puntos obtenidos mediante el kernel seleccionado, en la etapa anterior y de esta manera determinar si una nueva señal fue objeto de una manipulación o no.

En detalle, el proceso de análisis de las series de tiempo será el siguiente:

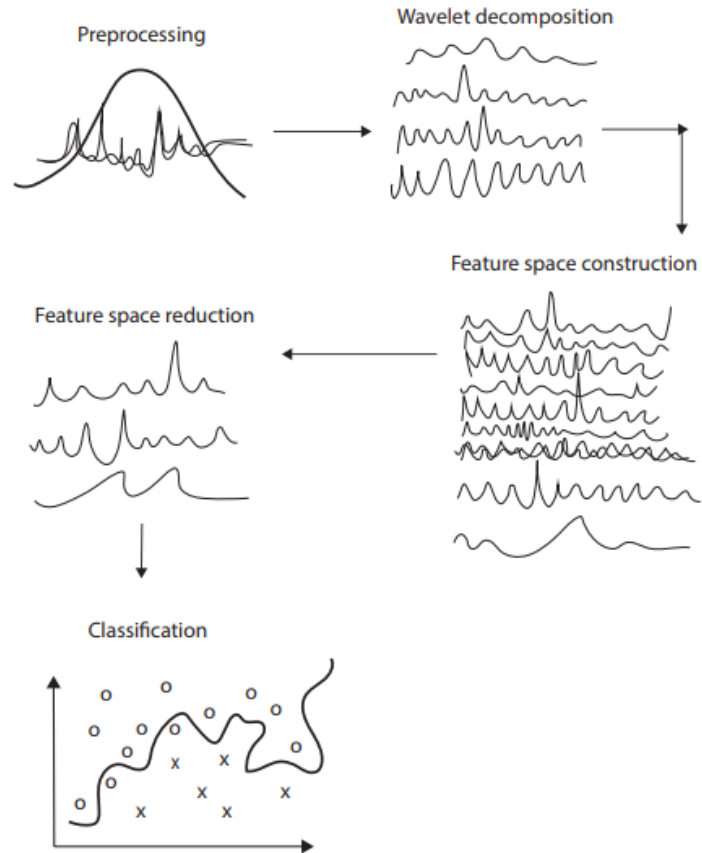
- Determinar la dimensión de los puntos del set de datos: Descripción básica de las características que componen cada dato y la definición de las ventanas para el análisis.
- Aumento de las dimensiones: Teniendo en consideración la naturaleza de las series de tiempo en cuestión, se aumentarán las dimensiones para determinar el espacio adecuado que permita separar los datos anómalos de los que no.

En esta etapa se considera la aplicación de las funciones de preprocesamiento, el análisis de la descomposición de la señal y la construcción del nuevo espacio de características.

- Selección de características: Se considerará esta etapa para facilitar el procesamiento de los datos, como también su entendimiento, ya que se busca eliminar ruido, características que aportan muy poca información y reducir el costo computacional.
- Clasificación: Por último, se hará la clasificación de los puntos de la serie de tiempo, para determinar los datos anómalos, considerando aquellas técnicas viables para los espacios generados, considerando en primera instancia técnicas de aprendizaje de máquina basadas en clasificación.

A modo de resumen la metodología que se propone para la extracción, selección y clasificación es la siguiente:

Figura 1. Framework para el aprendizaje de máquina mediante la aplicación de kernel



Fuente: Quintero-Montoya, 2018

Para validar los resultados de la clasificación se hará una comparación del desempeño diversos modelos clásicos, considerando para ello las medidas que mejor se ajusten a las características del problema tratado, teniendo como base los procedimientos que hacen parte de los esquemas de 'cross-validation'.

Plan de trabajo

La presente propuesta considera la recolección y análisis de información bibliográfica sobre el estado del arte y la información seminal de las diversas técnicas que serán consideradas, especialmente en el caso de los kernel para el procesamiento de señales.

Mediante la aplicación de conocimiento experto, se evaluará la pertinencia de las herramientas definidas, con el fin de proponer en caso de ser viable un modelo de detección de prácticas manipulativas en mercados financieros.

Equipo y roles

- Jose Daniel Herrera Ochoa (Investigador): Estudiante de la Maestría En Ciencias De Los Datos y La Analítica, con especialización en finanzas y diez años de experiencia en el sector financiero.
- Olga Lucia Quintero Montoya (Director): Investigador Senior del Departamento de Ciencias Matemáticas.
- Daniel Otero Gomez (Tutor): Estudiante Graduado del programa en Ingeniería Matemática, con conocimientos en aplicación de técnicas en aprendizaje de máquina.

Presupuesto

Referenciando una propuesta en 2018 realizada por un equipo de analítica, para realizar detección de datos atípicos en series de datos financieros, el costo de la solución estaría en quince millones de pesos colombianos. Se tiene en consideración incluir como parte de la propuesta el doble del valor inicial teniendo en cuenta la implementación de modelos de aprendizaje de máquina como parte del proceso.

Productos esperados

Como resultado de esta propuesta se tendrá la presentación de un informe de investigación, en el que conste la descripción de los kernels seleccionados, las técnicas de aprendizaje de máquina más adecuadas y las métricas para la evaluación de los resultados, de tal manera que se logre concluir sobre la viabilidad de este tipo de técnicas en la detección de prácticas manipulativas en mercados financieros.

Plan de gestión de datos

Los datos empleados en el desarrollo de esta propuesta son públicos y de libre acceso, ya que se encuentran disponibles para cualquier usuario con acceso a internet en los portales web de los reguladores de diversos mercados a nivel mundial, como también en las plataformas de diversos proveedores de precios e información.

En detalle, el tipo de datos considerados en el desarrollo de esta propuesta corresponde a series de tiempo que reflejan el comportamiento de los precios de negociación de un activo financiero, transacción por transacción durante una jornada dada (este tipo de tipo series de tiempo se conoce comúnmente en la jerga como '*tick by tick - level 1*'), en la que previamente los reguladores u otros agentes de mercado han evidenciado y descrito una práctica manipulativa de mercado.

Es importante destacar, que no se harán transformaciones y tratamientos particulares sobre los datos, más allá del procesamiento natural que tendrán en la aplicación de las técnicas consideradas, por lo que no se considera necesario emplear un repositorio externo para almacenar los datos ya que las fuentes son de libre acceso.

Aspectos Éticos

Al ser datos públicos y de libre acceso, el único aspecto ético a considerar es el respeto a los derechos de autor de la publicación, considerando para ello la respectiva referencia y la consideración al tipo de licencia de uso de estos conforme el autor de la publicación de los datos considere.

Los datos serán empleados para la construcción y validación de las técnicas descritas en esta propuesta. Se espera que con su uso pueda determinarse la viabilidad de los kernels como una herramienta adecuada para el objetivo de mejorar los estándares de negociación de los mercados financieros al lograr la detección acertada de prácticas manipulativas de mercado.

DESARROLLO

FÉNOMENOS DE MANIPULACIÓN EN MERCADOS FINANCIEROS

Los mercados en general desde su concepción, no solo se basan en los conceptos de oferta y demanda, sino que parten de la confianza y el acuerdo explícito entre las partes que intervienen en la negociación, que todo aquello que se considera en la misma viene dado desde la buena fe y se da en condiciones de mercado justas. (Allen & Galle, 1992)

Sin embargo, desde la creación de las primeras bolsas ha quedado manifiesta la posibilidad de que artificialmente se puede influir en los precios en los que se negocian los activos para obtener un beneficio particular (Allen & Galle, 1992).

Con el fin de proteger los intereses de todos los agentes que intervienen en las negociaciones y que se garantice transparencia, los mercados son regulados.

Entre uno de los principales focos de los reguladores de los mercados, es que las condiciones de los precios de los activos sean fiel reflejo de las condiciones financieras de una compañía y de su potencial de generación de ingresos, por tal motivo cualquier distorsión en los precios es sancionada.

De igual manera, se sancionan todas aquellas ofertas que no tengan un sustento real o que no sean dadas de buena fe con la intención de intercambiar la titularidad sobre un activo financiero. (STOCK Act, 1978)

En consideración, es válido afirmar que cualquier intento de un agente del mercado por aumentar o disminuir los precios de un activo para su propio beneficio es objeto de sanciones, y se tipifica como manipulación de mercado.

Las manipulaciones de mercado como fue mencionado previamente pueden tomar diversas formas. Su clasificación, parte de los diversos factores que inciden en la manipulación como, por ejemplo, la duración del esquema, la dirección y la selección de los precios y la materialización o no de las intenciones de negociación.

Son muchos los tipos de prácticas manipulativas que han sido bien tipificadas, sin embargo, para todos en general la evidencia de su ejecución en escenarios de negociación reales no ha sido bien documentada, en parte debido a que el avance tecnológico de los sistemas de negociación y de las capacidades de negociación - especialmente de grandes actores- excede las capacidades de detección de los reguladores (Lin, 2016) lo que implica que las manipulaciones generalmente no son fácil de reconocer, respecto a otras fallas de mercado y por tanto las sanciones al respecto son pocas (Aitken, Harris & Ji, 2009).

En consideración, serán detalladas prácticas que fueron tipificadas y sancionadas por reguladores y que por tanto están en capacidad de respaldar los hallazgos de la presente propuesta.

Pump and dump

Según (Lin, 2016) este esquema generalmente opera cuando un manipulador adquiere una posición en un instrumento financiero y luego artificialmente infla su valor a través de actos de promoción fraudulentos con el fin de vender su posición al precio inflado.

Generalmente los valores vuelven a sus valores regulares luego de dicha venta, pues como Yagemann et al., (2021) precisan, una vez el manipulador logra favorecerse del impulso que imprimió en el precio del instrumento financiero manipulado, no existen elementos o razones que sustenten el valor artificialmente incrementado y por esta razón regresa a los valores previos a la manipulación.

Frecuentemente los activos que suelen ser empleados para este tipo de fraude son acciones con precios bajos, poca liquidez y operadas en sistemas de registro según lo menciona la SEC (2006).

Wash trade

Según Cao et al., (2015) este esquema se refiere a las actividades ilegales que un operador del mercado emplea para incrementar los volúmenes de negociación de un activo para crear la falsa impresión de movimientos en el mercado.

Como la intención principalmente es crear la apariencia de liquidez, la diferencia en precios entre las compras y ventas son mínimas o inexistentes y, por tanto, se consideran operaciones nominales.

Sin embargo, Lin (2016) puntualiza que el objeto de la manipulación también pueden darse sobre el precio, pues luego de haber llamado la atención de otros inversionistas al aumentar, mantener o disminuir el valor de un activo con operaciones nominales, los manipuladores pueden efectuar operaciones posteriores que estén en línea con sus expectativas, ya sea para lucrarse con el resultado de los precios artificiosos o con el fin de afectar los intereses de otros actores sobre el activo manipulado.

Poop and scoop

Según el CFI (2015) esta práctica consiste en emplear medios fraudulentos de promoción para disminuir el valor de un activo financiero, adquirirlo en precios bajos y posteriormente venderlo una vez el mercado reaccione y haga retornar el precio de dicho activo a sus niveles habituales.

Si bien, claramente es la práctica opuesta al *pump and dump*. Según Lin (2016) es clave que, en la mayoría de los casos registrados, luego de que un manipulador haya logrado disminuir de manera deliberada los precios de un activo, también venderá su posición mediante promoción fraudulenta en sus niveles deseados, los cuales generalmente estarán cerca de los niveles originales de negociación.

De forma similar, como ocurre con el *pump and dump*. Acciones de bajo valor y poca liquidez, son las más susceptibles a este tipo de prácticas.

Spoofing

Según Lin (2016) en este esquema, las órdenes de negociación son ubicadas en las plataformas a precios por fuera de los niveles de mercado con el fin de 'generar pánico' entre los demás inversionistas y hacerlos pensar que el precio de dicho activo está subiendo o está bajando. En caso de que el movimiento resulte favorable para el manipulador éste rápidamente cancelará su orden y ubicará una orden contraria que se beneficia de su manipulación.

La SEC (2015) precisa que, en este esquema, el manipulador no tiene intención alguna de ejecutar sus órdenes manipulativas, ya que su único objetivo es inducir la intención de negociación de los demás agentes del mercado para beneficiarse del movimiento artificialmente creado.

Layering

Según Yagemann et al., (2021) esta práctica se considera un caso de *spoofing*, en el que son empleadas varias operaciones de manera sucesiva para disfrazar las intenciones manipulativas sobre el precio del activo.

La SEC (2015) precisa que, en esta categoría también se incluyen aquellas ocasiones en las que son varias las ordenes de mala fe que se emplean para influenciar los precios de los activos, pues de esta manera se ejerce mayor presión para motivar los demás actores del mercado en seguir el movimiento artificialmente creado. Una vez ejecutada las órdenes opuestas, dichas ordenes de mala fe serán canceladas.

HERRAMIENTAS PARA EL ANÁLISIS DE SEÑALES

El procesamiento de señales consiste en la preparación, análisis y extracción de información de una señal sin importar su fuente, por esta razón ha sido empleado en diversas áreas incluyendo telecomunicaciones, radares e ingeniería biomédica. (Boashash, 2015)

En el caso financiero, se ha usado frecuentemente para estudiar el comportamiento de los precios de un activo en un período de tiempo dado.

Los precios de un activo financiero son descritos a partir de series de tiempo finitas, en las que sus variaciones pueden ser comprendidas como 'señales' del comportamiento de cada activo en el mercado. (Quintero-Montoya, 2019)

Las señales financieras, como las demás señales del mundo real tienen una duración específica, períodos de aparición y una capacidad de representación

limitadas. Además, al ser variables en el tiempo se denominan como 'no estacionarias', ya que la aparición de una determinada frecuencia no es constante en el tiempo.

Lo anterior, debido a la compleja relación entre una gran variedad de factores externos que inciden en su comportamiento. (Nelson, 1991)

Es evidente como también en el caso financiero, el tiempo y la frecuencia están intrínsecamente relacionados (al igual que en otros fenómenos de la naturaleza). Sin embargo, éstas no son las únicas variables relacionadas, ni los únicos elementos de interés al analizar una señal.

Como tal uno de los objetivos del procesamiento de señales, es la extracción de toda la información posible de una señal y emplearla para una toma de decisiones efectiva. (Quintero-Montoya, 2019)

Por tal motivo, han sido desarrolladas gran variedad de herramientas que facilitan el procesamiento de una señal y por ende la comprensión de diversos fenómenos físicos, al considerar precisamente aquellos elementos diversos que intervienen en la generación, propagación y procesamiento de cada señal.

Las herramientas más básicas para el análisis de señales son aquellas que representan el dominio del tiempo (que no presentan claramente qué frecuencias están presentes en una señal) y las representaciones del dominio de frecuencias (que no muestran en que momentos del tiempo están presentes).

Si bien, de primera mano éstas representaciones pueden parecer no ser útiles para el caso de aplicación que se propone, no dejan de ser relevantes. Ya que, debido a la naturaleza altamente variable de las señales financieras, la cantidad de energía e información contenida en ellas no es fácilmente representable. (Quintero-Montoya, 2019)

Por lo anterior, considerar todas aquellas herramientas, técnicas o dominios de información que permitan la detección y la caracterización de una manipulación son útiles y necesarios.

Otras de las representaciones más comunes, además de las representaciones bidimensionales, (como en el caso de las representaciones del dominio del tiempo o de la frecuencia) son aquellas del dominio tiempo-escala-frecuencia, las cuales se consideran espacios más completos, pues están en capacidad de representar de manera más efectiva la información codificada en una señal. (Boashash, 2015)

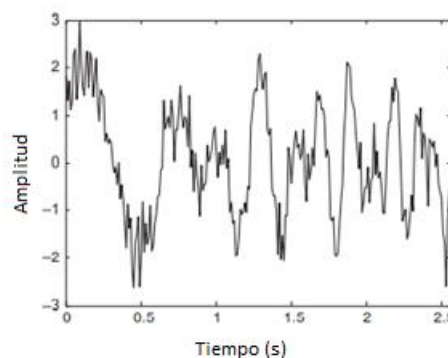
Sin embargo, tales dominios no son los únicos, y en general los dominios de los datos que cada herramienta está en capacidad de representar revelarán información sobre la presencia de ciertos componentes y de ciertos artefactos, que podrían estar en la señal.

La elección de cierto tipo de herramientas, técnicas o algoritmos dependerá del fenómeno que es estudiado y del tipo de señal analizada.

Por tanto, definir con precisión las características de la señal que se investiga es a menudo uno de los pasos más importantes en la resolución de procesos prácticos de ingeniería. (Boashash, 2015)

Una forma común de clasificación de una señal, parte de sus grados de complejidad, entendiéndose está a partir de su evolución temporal, la distribución de la energía presente y la presencia de ruidos y aleatoriedad. Según Boashash (2015), podríamos describir las señales financieras como señales aleatorias y ruidosas, en la que hay presencia de múltiples componentes cada uno con sus propios valores de amplitud y frecuencia variable en el tiempo, cuya suma ponderada representa la señal como tal, como puede apreciarse en la Figura 2:

Figura 2. Representación en el dominio temporal de una señal compleja



Fuente:
Boashash, 2015

Con base en la intuición anterior, podría afirmarse que conviene considerar aquellas representaciones que estén en capacidad de representar las señales en espacios más complejos.

No obstante, las señales financieras exhibirán comportamientos esperados dentro de períodos particulares del tiempo en respuesta a las relaciones económicas entre las diversas fuerzas que actúan en los mercados y sobre los activos en diferentes momentos y escalas del tiempo (Nelson, 1991).

Sin embargo, cuando un cambio brusco se introduce, desde su aparición y mientras sus efectos se perciban, las señales de mercado exhibirán oscilaciones y variaciones en sus regímenes de frecuencia. (Gallegati & Semmler, 2014)

Dichos efectos, pueden ser localizados en el tiempo y caracterizados a partir de diferentes escalas.

Lo anterior resulta clave, pues dichos cambios en el agregado tal vez no sean perceptibles, pero sí en un rango particular de escalas, ya sea en el corto o en el largo plazo.

No es de extrañar, que los términos “corto plazo” y “largo plazo” sean elementos centrales entre las complejas relaciones de las variables financieras. (Percival & Walden, 2000)

Por lo que considerar, la posibilidad de evaluar diversas escalas resulta fundamental.

Por lo anterior y teniendo en cuenta que las señales financieras son no-estacionarias, habrá herramientas que tal vez no sean capaces de lograr una buena representación de este tipo de señales. (Arts & van den Broek, 2022)

Por otra parte, al momento de considerar herramientas para el procesamiento de señales, es importante tener en mente el dominio de representación de la señal, pues cada dominio tiene sus propias ventajas y permitirá resolver preguntas específicas como: ¿en qué momento se ubica un componente en una señal? ¿cómo puedo ajustar el espectro para detectar algún componente o extraer algún artefacto en particular? (Varanis, Silva, Balthazar & Pederiva, 2021)

La caracterización de las manipulaciones es fundamental, por lo que considerar para tal fin, las mismas herramientas empleadas en el procesamiento de señales, va en línea con los objetivos planteados anteriormente como parte de este proceso y, por tanto, no serán descartadas las herramientas más básicas, pues de esta manera se puede describir de mejor manera el fenómeno de la manipulación y comprender de mejor manera su representación.

Espectro de potencia

De acuerdo con Stoica & Moses (2005) el espectro de potencia de una serie de tiempo describe la distribución de poder o la energía entre los diversos regímenes o componentes de frecuencia que conforman la señal de dicha serie.

Se considera para su cálculo, partir de un set finito de medidas de la señal por medio de métodos paramétricos o métodos no paramétricos. La clásica transformada de Fourier es el método comúnmente empleado por las ventajas computacionales que posee, que involucra dividir la serie tomando las frecuencias presentes en cada división y promediando dichas muestras. (Welch, 1967)

Se parte del principio de que una señal continua no finita puede ser analizada al descomponer las frecuencias presentes en funciones periódicas simples: senos y cosenos. (Harris, 1978)

Por practicidad, la señal es procesada en N muestras de igual tamaño, donde N es un número par y altamente compuesto. De manera que la transformada empleada,

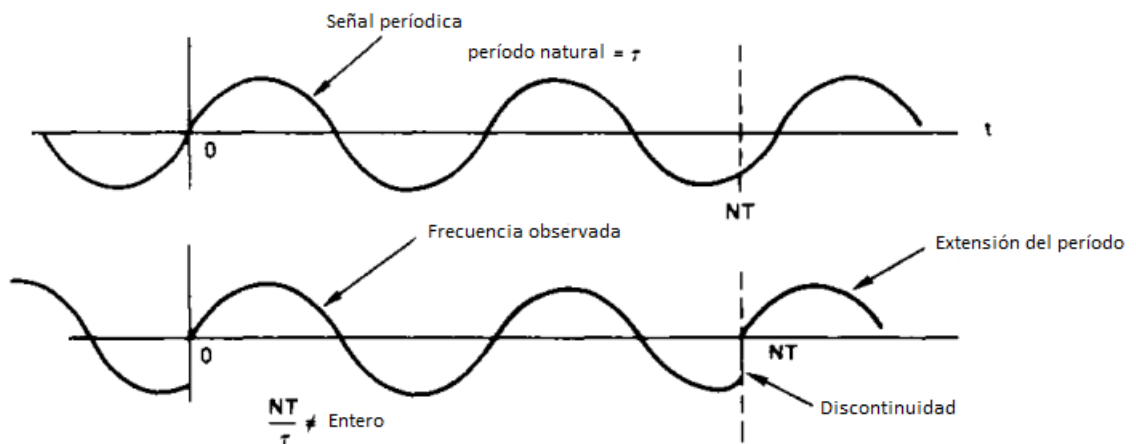
esté en capacidad de obtener las estimaciones armónicas de cada segmento y de esta manera construir el espectro de potencia. (Harris, 1978)

La selección de un intervalo finito puede determinarse mediante dos estrategias, la primera un submuestreo sobre puntos equidistantes en la señal o la segunda mediante ventanas.

Las ventanas son funciones cuyo valor es cero por fuera de un determinado intervalo. Normalmente son simétricas alrededor del punto medio de dicho intervalo, que generalmente es un punto máximo que disminuye su valor en los extremos. Al multiplicar una función por una ventana el producto también será cero por fuera de dicho intervalo, por tanto, la única parte que queda es donde ambas funciones se solapan, "la vista a través de la ventana". (Oppenheim, Schafer & Buck, 1999)

La idea al emplear una ventana es disminuir el efecto conocido como "filtrado espectral", ya que independiente de la base ortogonal seleccionada (muestral o continua) al final de cada intervalo, solo aquellas frecuencias que coinciden con la base serán proyectadas, las demás no, sin embargo, cuando la extensión periódica de una señal no se ajusta con su periodo natural está exhibe discontinuidades en los límites de cada observación. Estas discontinuidades son responsables de dejar marcas o filtrado sobre toda la base. (Harris, 1978)

Figura 3. Representación de una discontinuidad en una señal



Fuente:
Harris, 1978

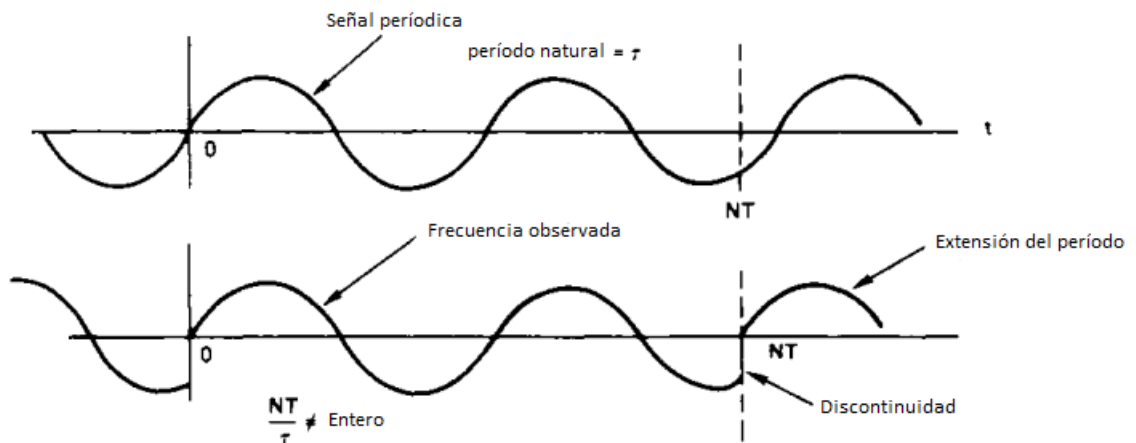
Es importante notar que la cantidad de filtrado en una señal puede ser influenciado por el período de muestreo, sin embargo, el filtrado no es causado por el muestreo, ya que el filtrado resultado del procesamiento de señales finitas. (Harris, 1978)

El uso de una determinada ventana impactará en varios de los atributos del procesador de señales, lo que incluye detectabilidad, resolución, rango dinámico, nivel de confianza y facilidad de implementación.

En general al procesar una señal se busca seleccionar aquellos mejores parámetros que permitan la identificación y la representación del fenómeno analizado.

Otro criterio que debe ser considerado al momento de elegir una ventana, es la capacidad de resolución. Al procesar una señal, es de interés identificar todos aquellos componentes de frecuencia. Sin embargo, en caso de que dos componentes estén muy cercanos y el ancho de la ventana es amplio, la identificación de ambos puede comprometerse y puede considerarse como un fenómeno individual cuando en realidad no lo es como puede apreciarse en la Figura 4.

Figura 4. Resolución Espectral



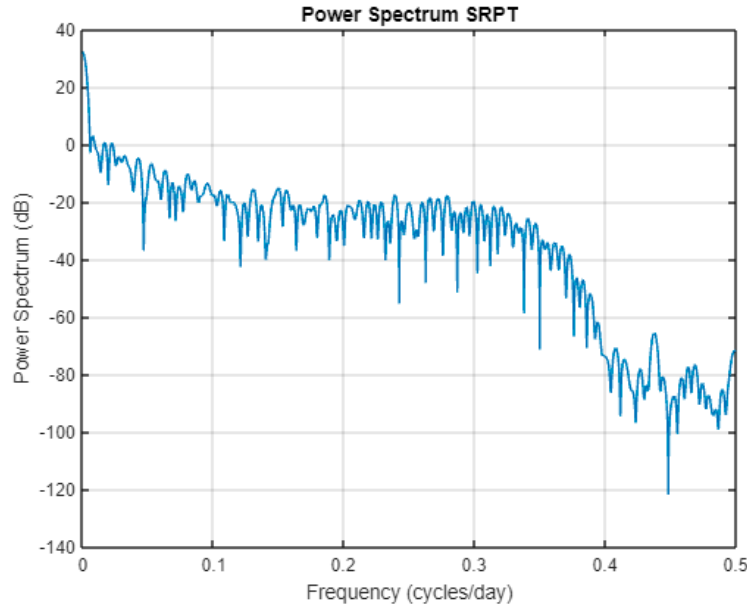
Fuente:
Harris, 1978

Se puede dividir cualquier señal, en intervalos tan cortos que la resolución mejore, sin embargo, cada componente detectado proyectará energía sobre todo el ancho de frecuencias. Por lo que el proceso de ventaneo siempre involucra un compromiso entre dos objetivos en conflicto: mejorar la resolución y disminuir el filtrado espectral. (Harris, 1978)

Una mejor resolución, implica un mayor filtrado. Mientras que un menor filtrado, implica una peor resolución. Por tanto, al momento de elegir una ventana conviene considerar que no solo el tamaño importa, sino su forma. Entre más ancha sea la forma del centro de la función, menor será la resolución y entre más atenuado estén sus extremos mejor se controlará la cantidad del filtrado.

Para su representación gráfica, generalmente se escala el espectro de manera que, si el contenido de frecuencia de un segmento es detectado, su amplitud pueda ser promediada y de esta manera obtener un mejor estimado de los componentes presentes en la señal. (Welch, 1967)

Figura 5. Representación gráfica del espectro de potencia



Según Welch (1967), mediante el uso de segmentos de tamaño K es posible evaluar el comportamiento en diversos momentos de la señal. Para dicha segmentación no solo propone la definición de la cantidad de ventanas, sino también su tamaño y su nivel de solapamiento.

En cada segmento se realiza la Transformada de Fourier y se realiza el promedio de todas las transformadas: $A_k(n) = \frac{1}{L} \sum_{j=0}^{L-1} X_k(j)W(j)e^{-2kij_n/L}$, donde $i = (-1)^{\frac{1}{2}}$, $X(j)$ corresponde a cada una de las muestras de una secuencia de una señal igual a $N - 1$, donde N es la cantidad de muestras. L es el tamaño es de los segmentos y K la cantidad de segmentos de manera que $(K - 1)D + L = N$.

Y finalmente se obtiene el periodograma de cada segmento mediante: $I_k(f_n) = \left(\frac{L}{U}\right) |A_k(n)|^2$, $k = 1, 2, \dots, K$, donde $f_n = \frac{n}{L}$, $n = 0, \dots, \frac{L}{2}$ y $U = \frac{1}{L} \sum_{j=0}^{L-1} W^2(j)$.

El Espectro se estima mediante el promedio de cada uno de los periodogramas: $\hat{P}(f_n) = \frac{1}{K} \sum_{k=1}^K I_k(f_n)$.

Y de esta manera obtener el resultado de la Figura 6 (a).

Espectrograma

De acuerdo con Fulop & Fitz (2006) un espectrograma es una representación de los componentes modulares o de amplitud separable de las frecuencias instantáneas variables en el tiempo de una señal cualquiera.

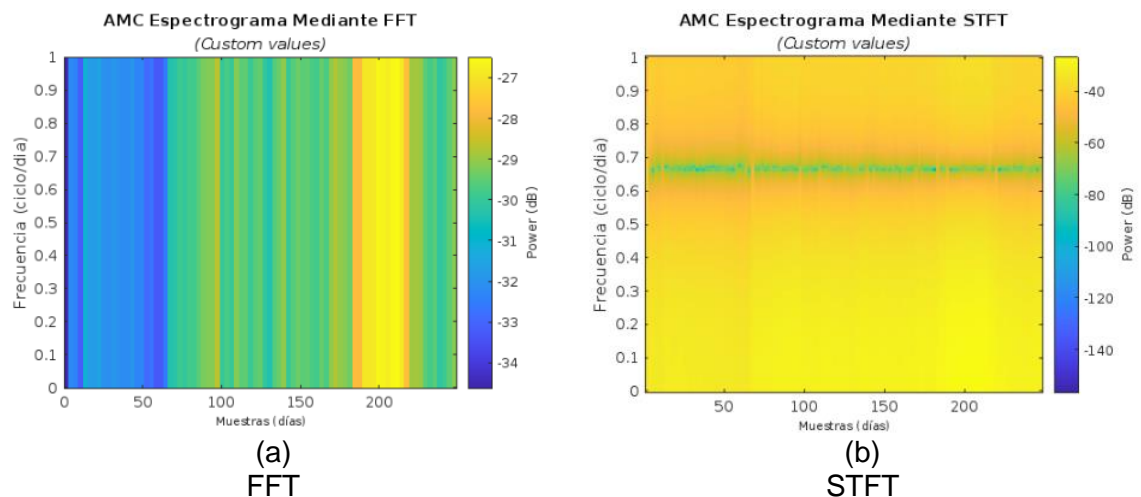
De forma más general puede describirse por sus siglas en inglés TCIF (*time-corrected instantaneous frequency*) -*spectrogram*, pues en pocas palabras constituye una representación visual del espectro de frecuencias de una señal conforme esta varía en el tiempo.

En la actualidad comúnmente se presenta como un gráfico logarítmico (en decibeles) de las magnitudes cuadradas de la Transformada Corta de Fourier.

Sin embargo, también puede ser generado a partir de un espectrofotómetro óptico, filtros de paso de banda o por transformadas wavelets (en cuyo caso se conoce más específicamente como escalograma) (Sejdic & Stankovic, 2008)

Entre las representaciones más comunes del espectrograma se tienen los mapas de calor, esto es una imagen que muestra la intensidad de las frecuencias detectadas variando en color o intensidad del brillo. Cuando estos son representados en 3D se conocen como gráficos en cascada. (Boashash, 2015)

Figura 6. Representación gráfica de un espectrograma



Generalmente en las herramientas de procesamiento de señales se representa de forma gráfica el espectro de potencia empleando la Transformada Corta de Fourier

en un mapa de calor, con el fin de determinar la frecuencia sinusoidal y el contenido de fase local en N segmentos de una señal. (Fulop & Fitz, 2006)

Según (Fulop & Fitz, 2006), mediante esta transformada, la señal es procesada haciendo uso de una ventana deslizante $g(n)$ de tamaño M , que recorre la señal en cada punto definido de acuerdo con el parámetro de solapamiento seleccionado y sobre el cual se calcula la Transformada Discreta de Fourier (DTF).

La ventana salta sobre la señal original a intervalos de R muestras, equivalentes a $L = M - R$ muestras de solapamiento entre segmentos contiguos. La mayoría de las funciones de ventana se reducen en los bordes para evitar el timbre espectral.

La DFT de cada segmento con ventana se agrega a una matriz de valores complejos que contiene la magnitud y la fase para cada punto en el tiempo y la frecuencia. La matriz STFT tiene k columnas que vienen dadas por: $k = \lfloor \frac{N_x - L}{M - L} \rfloor$. Donde N_x es el tamaño de la señal $x(n)$ y el símbolo $\lfloor \cdot \rfloor$ denota la función piso. El número de filas de la matriz es igual a N_{DFT} , donde el número de los puntos DFT para transformadas centradas de dos lados y un número impar cercano a $\frac{N_{DFT}}{2}$ para transformadas de un solo segmento de los valores reales de la señal.

La m -ésima columna de la matriz STFT $X(f) = [X_1(f) X_2(f) X_3(f) \cdots X_k(f)]$ contiene la DFT de los datos en ventana centrados en el tiempo mR : $X_m(f) = \sum_{n=-\infty}^{\infty} x(n) g(n - mr) e^{-j2\pi fn}$.

De esta manera se puede obtener información de la ubicación en el tiempo de cierta frecuencia y detectar picos o anomalías como se evidencia en la gráfica.

De manera similar a como ocurre con el cálculo del espectro de potencia, los segmentos pueden ser tan pequeños que en esencia se pueden considerar estacionarios en cada sección, y a su vez mejorar la resolución temporal, ya que las frecuencias estarán mejor localizadas en el tiempo.

Sin embargo, la estimación de la frecuencia puede verse comprometida por el filtrado espectral. Por tanto, segmentos más amplios proveen mejor resolución frecuencial, mientras que segmentos más cortos proveen mejor resolución temporal. (Harris, 1978)

Generalmente muchas herramientas de procesamiento de señales emplean hiperparámetros para encontrar un balance entre la resolución frecuencial y la resolución temporal. Estos normalmente se obtienen a partir de la información temporal de la señal la cual debe estar dada en segundos. Para este caso, una función convierte el resultado de muestras y este se redondea al entero más cercano, el cual será igual o menor a la duración de la señal. (Harris, 1978)

En caso de que una señal no cuente con información temporal o está no pueda ser ajustada el tamaño de los segmentos se determina a partir del tamaño de la señal y la cantidad de muestras que está posee, con base en el siguiente ratio: $\lceil \frac{N}{d} \rceil$, donde los símbolos $\lceil \cdot \rceil$ denotan la función techo, N es la longitud de la señal, y d es un divisor que depende de N , como puede apreciarse en la Figura 7.

Figura 7. Estimación de los segmentos de un espectrograma

| Longitud de la señal (N) | Divisor (d) | Longitud del segmento |
|-------------------------------|-----------------|--|
| 2 muestras– 63 muestras | 2 | 1 muestra– 32 muestras |
| 64 muestras– 255 muestras | 8 | 8 muestras– 32 muestras |
| 256 muestras– 2047 muestras | 8 | 32 muestras– 256 muestras |
| 2048 muestras– 4095 muestras | 16 | 128 muestras– 256 muestras |
| 4096 muestras– 8191 muestras | 32 | 128 muestras– 256 muestras |
| 8192 muestras– 16383 muestras | 64 | 128 muestras– 256 muestras |
| 16384 muestras– N muestras | 128 | 128 muestras– $\lceil N / 128 \rceil$ muestras |

Fuente:
Welch, 1967 (adapt.)

Una particularidad que fue desarrollada mediante el uso de ventanas en la transformada corta de Fourier y que tiene un fuerte impacto en la cantidad de segmentos que pueden resultar de procesar una señal, es el solapamiento.

Al solapar una ventana antes de otra se logra reducir la cantidad de filtrado espectral y mantener una buena resolución temporal de un fenómeno. (Welch, 1967)

Adicionalmente al solapar ventanas, se modifica el número de segmentos, todos aquellos elementos de un segmento solapado que queden por fuera de la señal son rellenados con ceros (o como se conoce en inglés *zero-padding*).

Este ajuste ofrece una gran ventaja, sin embargo, al aumentar el número de segmentos, se aumenta el tiempo de procesamiento al considerar mayor cantidad de ventanas. Por tanto, al momento de elegir la cantidad de solapamiento es necesario nuevamente considerar el tamaño de la señal y la cantidad de muestras disponibles, como puede apreciarse en la Figura 8.

Es importante tener en cuenta que cada segmento solapado es procesado mediante la misma función de ventaneo y posteriormente el espectro calculado es solapado con el resto para de esta manera formar la matriz del espectrograma.

Figura 8. Segmentos resultantes luego de solapar muestras

| Longitud de la señal (N) | Divisor (d) | Longitud del segmento |
|-------------------------------|-----------------|--|
| 2 muestras– 63 muestras | 2 | 1 muestra– 32 muestras |
| 64 muestras– 255 muestras | 8 | 8 muestras– 32 muestras |
| 256 muestras– 2047 muestras | 8 | 32 muestras– 256 muestras |
| 2048 muestras– 4095 muestras | 16 | 128 muestras– 256 muestras |
| 4096 muestras– 8191 muestras | 32 | 128 muestras– 256 muestras |
| 8192 muestras– 16383 muestras | 64 | 128 muestras– 256 muestras |
| 16384 muestras– N muestras | 128 | 128 muestras– $\lfloor N / 128 \rfloor$ muestras |

Fuente:
Welch, 1967 (adapt.)

Escalograma

Como fue mencionado previamente, los escalogramas corresponden a los espectrogramas generados a partir de la Transformada Wavelet. Esta transformada es una representación lineal de tiempo-frecuencia que conserva los cambios de frecuencia en su dimensión de tiempo.

La conservación de la dimensión de tiempo, en este tipo de análisis es posible ya que esta transformada emplea funciones llamadas ‘wavelet’ que descomponen la señal en múltiples escalas de manera localizada en el tiempo, al dilatar o contraer la función wavelet elegida antes de realizar una operación de convolución contra la señal. (Sejdic, Djurovic & Stankovic, 2008)

Lo anterior, permite que se represente la señal en un espacio de tiempo-frecuencia más amplio y con mayor precisión. Pues el fraccionamiento de las señales a múltiples escalas es tan fino, que permite cada parte de la señal ser analizada y clasificada en forma independiente en todo su espectro de potencia. (Arts & van den Broek, 2022)

Escalas altas corresponden a una vista global (no detallada) de la señal y las escalas bajas corresponden a una vista detallada. De manera similar, en términos de frecuencia, las frecuencias bajas (escalas altas) corresponden a información global de una señal (que generalmente abarca toda la señal), mientras que las frecuencias altas (escalas bajas) corresponden a información detallada del patrón oculto en una señal (que generalmente dura relativamente poco tiempo). (Sejdic, Djurovic & Stankovic, 2008)

En consideración al fenómeno que se esté estudiando y lo que se requiera hacer con la información disponible luego del procesamiento de la señal, es posible elegir wavelets entre dos grupos: discretas y continuas.

La diferencia entre ambos grupos radica en la forma en la que discretiza el parámetro de escala. La CWT discretiza de manera más fina, ya que generalmente la base en la que se establece la escala es una potencia fraccional de dos, $2^{\frac{1}{v}}$ donde v es un entero mayor a 1.

Este parámetro normalmente se conoce como ‘voces por octava’. La razón por la que se conoce de esta manera es que la escala se incrementa mediante una duplicación que requiere v escalas intermedias mediante octavas. Entre mayor sea v más fina será la discretización de la señal. Sin embargo, esto también aumenta el tiempo de cómputo requerido pues la transformada debe ser calculada en cada escala; valores típicos de v son 10, 12, 14, 16 y 32. (Lilly, 2017)

Por su parte, la transformada wavelet discreta, el parámetro de escala siempre se discretiza mediante potencias enteras de 2, 2^j , $j = 1, 2, 3, \dots$, de modo que el número de voces por octava es siempre 1. (Lilly, 2017)

Las nociones anteriores son fundamentales para comprender de mejor forma la información que representa el escalograma. Pues la transformada wavelet para efectos prácticos puede considerarse como un filtro de paso de banda de la señal de entrada. (Gallegati & Semmler, 2014)

A diferencia del filtrado de paso de banda de Fourier, el filtro de paso de banda de una transformada wavelet es inversamente proporcional a la escala, es decir el ancho de los filtros disminuye al aumentar la escala. (Daubechies, 1990)

En el filtro de paso de banda de una transformada wavelet las frecuencias bajas resultan a partir de wavelets de escala más amplia, mientras que las frecuencias más altas resultan de wavelets de una menor escala. (Lilly, 2017)

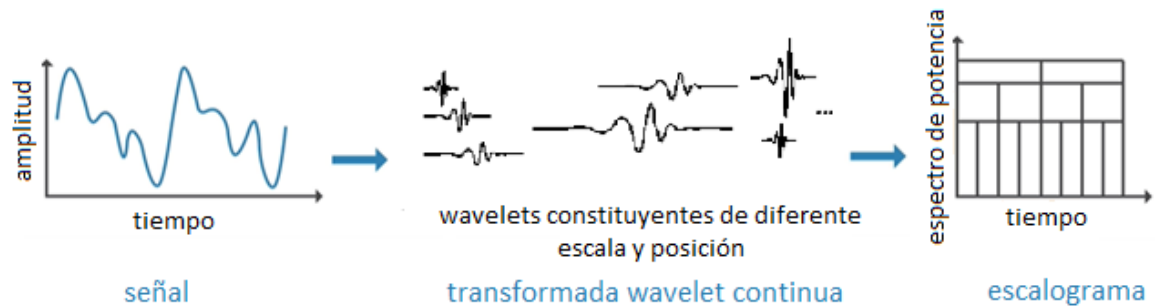
Debido a que la señal $f \in L^2(\mathbb{R})$ es descompuesta sobre un diccionario de wavelets trasladadas y dilatadas. $\psi_{u,s}(t) = \frac{1}{\sqrt{s}} \psi\left(\frac{t-u}{s}\right)$. Estos se construyen con base en el parámetro s sobre toda la secuencia exponencial $\{v^j\}_{j \in \mathbb{Z}}$, mientras mantiene el parámetro de traslación en u . Por su sencillez se elige $v = 2$ para facilitar su cálculo:

$\mathcal{D} = \left\{ \psi_{u,2^j}(t) = \frac{1}{\sqrt{2^j}} \psi\left(\frac{t-u}{2^j}\right) \right\}_{u \in \mathbb{R}, j \in \mathbb{Z}}$. El resultado de la transformada wavelet diádica

de $f \in L^2(\mathbb{R})$ está definida por: $Wf(u, 2^j) = \langle f, \psi_{u,2^j} \rangle = \int_{-\infty}^{\infty} f(t) \frac{1}{\sqrt{2^j}} \psi\left(\frac{t-u}{2^j}\right) dt = f * \bar{\psi}_{2^j}(u)$, con $\bar{\psi}_{2^j}(t) = \psi_{2^j}(-t) = \frac{1}{2^j} \psi\left(\frac{-t}{2^j}\right)$.

Considerando lo expuesto, el escalograma puede entenderse de forma sencilla en la Figura 9.

Figura 9. Representación gráfica del proceso de cálculo de un escalograma



Fuente:
Boashash, 2015 (adapt.)

Esta representación es importante, pues muchos procesos físicos revelan cambios interesantes en el largo plazo o en escalas espaciales más amplias que tal vez no son evidentes en escalas más cortas, y es a esto a lo que se denomina el factor de escala.

Entre más pequeño sea el factor, más comprimida estará la wavelet y entre más grande sea, más amplia será ésta (está relación inversa entre escala y frecuencia aplica para señales en general). (Mallat, 2009)

Mediante el proceso anterior se representa y se preserva de mejor manera la energía de una señal, pues estirar o estrechar el soporte de frecuencia de una wavelet, resulta en aumentos o incrementos proporcionales en la energía tope que es identificada. (Arts & van den Broek, 2022)

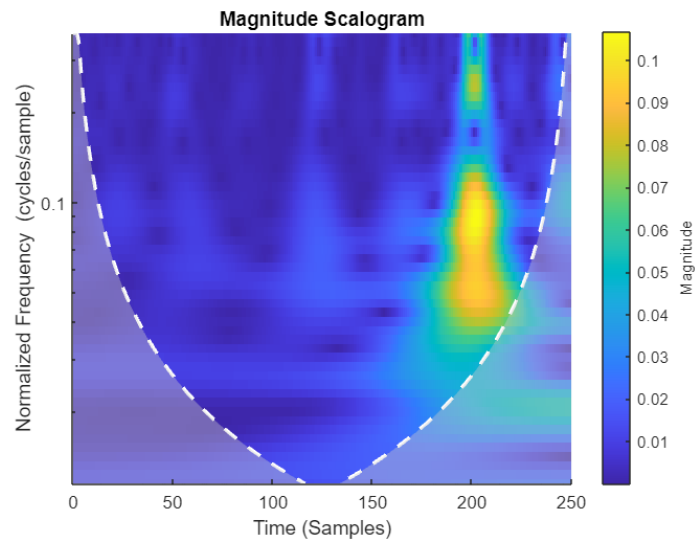
El resultado de la representación de la señal a diversas escalas puede visualizarse de manera sencilla mediante un mapa de calor como puede apreciarse en la Figura 10.

En el gráfico se ilustra lo que teóricamente ocurre cuando se amplía una wavelet, entre más se amplió la wavelet, más larga será la porción de la señal con la que será comparada y de esta manera menos finas serán las características medidas por los coeficientes de la wavelet (como ocurre con la transformada corta de Fourier), y esto se da, porque al estirar una wavelet en el tiempo hace que su soporte en el dominio de la frecuencia se reduzca. (Arts & van den Broek, 2022)

Sin embargo, al considerar una gran variedad de escalas y ventaneo de la señal, se obtiene una mejor representación del espectro de una frecuencia de una señal.

Además, esta representación de escala de tiempo no solo es una forma diferente de ver los datos, sino que también es una forma muy natural de representar señales de una gran cantidad de fenómenos naturales, pues facilita diferenciar los diferentes impulsos, componentes y frecuencias presentes en una señal. (Boashash, 2015).

Figura 10. Representación gráfica de la CWT mediante un escalograma



Sin embargo, como la transformada wavelet continua es una transformada redundante y los coeficientes dependen de la wavelet, no siempre resulta sencillo interpretar los resultados. Lo anterior debido a las finas escalas de procesamiento en el que los coeficientes cercanos suelen ser muy similares o incluso los mismos, debido a los efectos provocados por el solapamiento y el filtrado espectral.

La transformada wavelet discreta al respecto no ofrece una solución, pues los niveles y escalas con diversos tamaños por efecto del sub-muestreo no facilita tal labor.

Para tal fin se sugiere iniciar con una señal sencilla y una wavelet con una estructura simple. (Addison, 2018)

Por otra parte, cuando se interpreten los coeficientes de la transformada wavelet continua conviene considerar “el cono de influencia”. En el gráfico presentado previamente, puede apreciarse como la región delimitada por la línea blanca y el área sombreada sobre los ejes de tiempo y frecuencia. (Torrence & Compo, 1998)

Esta área representa sectores del gráfico en el que el espectro de potencia puede verse afectado por los 'efectos de borde'. Estos son efectos que surgen cuando las versiones estiradas de la wavelet se extienden más allá del intervalo de observación, lo que implica que todo aquello que se ubica en dicha área debe ser tratado con sospecha de que fue afectado por esta clase de eventos. (Taranenko & Rizun, 2022)

Es importante, tener en consideración que, dependiendo de la escala, los coeficientes podrían afectarse por puntos de la señal que no fueron considerados en dichas escalas. Además, se debe tener en cuenta, que no todas las wavelets tienen el mismo soporte y por tanto en diversas escalas los coeficientes se verán afectados. (Arts & van den Broek, 2022)

MANIPULACIONES FINANCIERAS USANDO KERNELS DE ANÁLISIS DE SEÑALES

Las diferentes fuerzas del mercado afectan las relaciones económicas durante diferentes períodos de tiempo. Los choques económicos están localizados en el tiempo y dentro de ese período exhiben oscilaciones de frecuencia variable. (Gallegati & Semmler, 2014)

Dichos choques pueden entenderse como 'singularidades', ya que son eventos puntuales que imponen una variación abrupta de los componentes de frecuencia en la señal.

Dichas singularidades o variaciones abruptas de frecuencia afectarán el régimen exhibido por la señal y a menudo se integrarán como parte de la información crítica que ésta contiene. (Boashash, 2016)

Dichas singularidades o eventos abruptos suelen ubicarse en los segmentos de alta frecuencia de las señales, ya que dichas singularidades son vistos como picos de energía. (Oppenheim, Schafer & Buck, 1999)

Entre mayor sea la respuesta transitoria (o cambio de fase), mayor será la energía registrada, ya que el área que describe un pico en una señal posee una amplitud infinitamente alta -matemáticamente hablando- y por tanto este es modulado con base en la escala de los diversos componentes sinusoidales que componen la señal. (Oppenheim, Schafer & Buck, 1999)

Detectar dichos picos en las frecuencias, generalmente puede considerarse una tarea trivial, pues normalmente el tamaño absoluto del choque en una serie de tiempo se verá reflejado en igual magnitud en sus frecuencias. (Hatemi, 2014)

Sin embargo, esto no siempre ocurre así. Previamente se hizo mención como es bien sabido en la literatura, como los diversos actores de un mercado responden de manera diferente ante diferentes eventos.

Por ejemplo, si los ingresos de una compañía incrementan un 5% esto tendrá diferentes consecuencias en términos de la magnitud en los precios comparado con el hecho de que dichos ingresos disminuyan ese mismo 5%. (Hatemi, 2014)

Por tanto, si la información asimétrica es una realidad de los mercados, entonces los impulsos asimétricos también lo deben ser. Una forma de evidenciarlo es como no hay un límite natural en el que el precio de una acción puede incrementar (teóricamente hablando), sin embargo, si existe un límite hasta el cual puede disminuir. (Hatemi, 2014)

Por ende, movimientos en los precios tendrán diferentes consecuencias en términos absolutos, independiente de la dirección del movimiento y su magnitud.

Basados en lo anterior, se puede concluir que estudiar las respuestas transitorias impresas en una señal financiera a causa de manipulaciones en los precios de un activo es un importante tema de estudio y que es necesario comprender, describir y caracterizar mediante las diversas herramientas para el procesamiento de señales dichas respuestas.

Si bien la teoría sugiere de manera particular cierto tipo de herramientas para el procesamiento de señales no estacionarias y altamente aleatorias como las que describen las series de los precios de los activos financieros. (Boashash, 2015)

Es fundamental comprender, describir y caracterizar como una respuesta transitoria en particular incide en el comportamiento del resto de la señal, y en general del fenómeno que se estudia, en este caso si un activo financiero ha sido objeto de una manipulación o no.

Wang, Yang & He (2022) para el análisis de acumulación de materia en clústeres galácticos; Gómez et al. (2017) para el análisis de emociones mediante impulsos cerebrales y Sierra-Sosa, Bastidas & Quintero-Montoya (2016) para la identificación de emociones en la voz, tuvieron en consideración la evaluación de diversas herramientas para entender de mejor manera el fenómeno de estudio y de esta forma elegir la representación más adecuada.

Incluso en algunos casos ciertas herramientas fueron propuestas como complemento de otras, para mejorar de esta manera las capacidades de detección y caracterización del fenómeno de estudio particular.

Es importante tener en cuenta, que el tipo de señal, el dominio de representación, las capacidades de resolución de los regímenes de frecuencia y las características del fenómeno como tal, fueron fundamentales al momento de elegir un tipo de herramienta sobre otra.

En consideración, se puede partir del hecho que las características importantes de una señal financiera pueden ser representados en dominios bidimensionales de tiempo-frecuencia.

Cada dominio o la combinación de ambos permitirá una caracterización particular de los elementos presentes en una señal, por ejemplo, las representaciones del dominio de frecuencia permitirán identificar el ancho de banda de cada componente de la señal; la representación del dominio del tiempo permitirá identificar la duración de una singularidad; mientras que la combinación de ambos permite identificar la energía presente en cada componente del espectro.

Debido a que los fenómenos que describen las series de tiempo financieras no son periódicos y que los términos como 'corto plazo' y 'largo plazo' son el eje central en el modelado de las diversas variables y factores que inciden en la conformación de los precios en las que se negocian los activos financieros, definir una ventana de tiempo adecuada que permita extraer y revelar información de las series de tiempo de los precios no es una tarea obvia, ni trivial. (Percival & Walden, 2014)

Al respecto, es necesario precisar que se hace referencia a las funciones de ventaneo que fueron definidas previamente. Sin embargo, definir el tamaño de la señal o la cantidad de muestras que serán procesadas también resulta clave para el problema de aplicación, pues las series de precios de los activos financieros al regirse por principios estocásticos, los análisis son altamente sensibles a la cantidad de información disponible. Especialmente cuando los procesos de variación de precios de las series son altamente aleatorios. (Nelson, 1991)

Lo anterior implica que se tendrá mayor cantidad de información sobre un fenómeno al tener mayor cantidad de muestras disponibles. Al respecto, Nelson (1991) detalla como al incluir los días de no negociación en cualquier análisis se afecta la percepción de cambio estimada en los precios.

Sin embargo, Hamilton & Susmel (1994) describen como un choque fuerte en los mercados afectará los regímenes de variación de precios durante un tiempo más amplio que choques más pequeños. Por tanto, es necesario considerar un balance que permita identificar con la información suficiente los fenómenos asociados a las manipulaciones, sin que los valores residuales de un evento posterior afecten la descripción de un evento nuevo y particular.

En consideración, es necesario precisar que el tamaño de la señal, la cantidad de muestras, el tamaño y la forma de la función de ventaneo, y la tasa de muestro, son conceptos relacionados que hacen parte de un mismo proceso.

Sin embargo, a pesar de su estrecha relación, cada concepto hace referencia a valores que aplican en etapas muy específicas del procesamiento de una señal, al punto, que variaciones en cualquiera de los parámetros afectará aquellos que dependan de estos.

Precisando, el tamaño de la señal ha referencia al período de análisis y a la cantidad de muestras recogidas en dicho período. Es decir, una señal que tenga una duración de un año puede tener 365 muestras si fue recogido el dato diario de los precios, puede tener 52 muestras en caso de que los datos hayan sido recogidos semanalmente y así sucesivamente, de acuerdo con el nivel de detalle que se requiera.

El tamaño de la ventana hace referencia a la cantidad de muestras que serán analizadas en dicha función. Mientras que su forma hace referencia a cómo cada elemento será operado en función del intervalo que abarca el tamaño de la ventana dado.

Por último, la tasa de muestreo corresponde a la cantidad de mediciones por unidad de tiempo que se realizan sobre un fenómeno, en general a mayor tasa de muestreo mejor la representación que se tiene de una señal, sin embargo, en este caso el concepto de tasa se emplea específicamente a la cantidad de muestras que serán tenidas en cuentas al procesar la señal, lo cual impactará la cantidad y la posición de las ventanas que serán evaluadas.

Por tanto, no solo la forma de la ventana permitirá la mejor identificación de ciertos componentes de frecuencia en una señal, sino también su tamaño y la cantidad de muestras consideradas tanto para el procesamiento de la señal, como en la constitución de esta.

Esta precisión se hace necesaria pues en algunos casos estos conceptos son considerados homónimos cuando no lo son, ya que el balance entre estos conceptos puede dar cuenta de una mejor descripción de la señal.

En general, los kernel han sido poco explorados en esta materia, debido a las restricciones expuestas, ya que en teoría la falta de capacidad en la representación del fenómeno y la complejidad de la señal ha hecho que por mucho tiempo sean considerados otro tipo de estrategias. (Quintero-Montoya, 2019)

Sin embargo, al descomponer una señal financiera para tratar de describir las relaciones entre sus componentes gráficamente se evidenció como las manipulaciones imprimen cambios abruptos en los regímenes de frecuencia, los cuales a su vez representan cambios energéticos abruptos en la señal .

Por lo anterior, revelar, extraer y caracterizar esos cambios energéticos, podría emplearse para determinar si un activo financiero fue objeto de manipulación, incluso de acuerdo con la caracterización de dichos cambios, determinar el tipo de la manipulación.

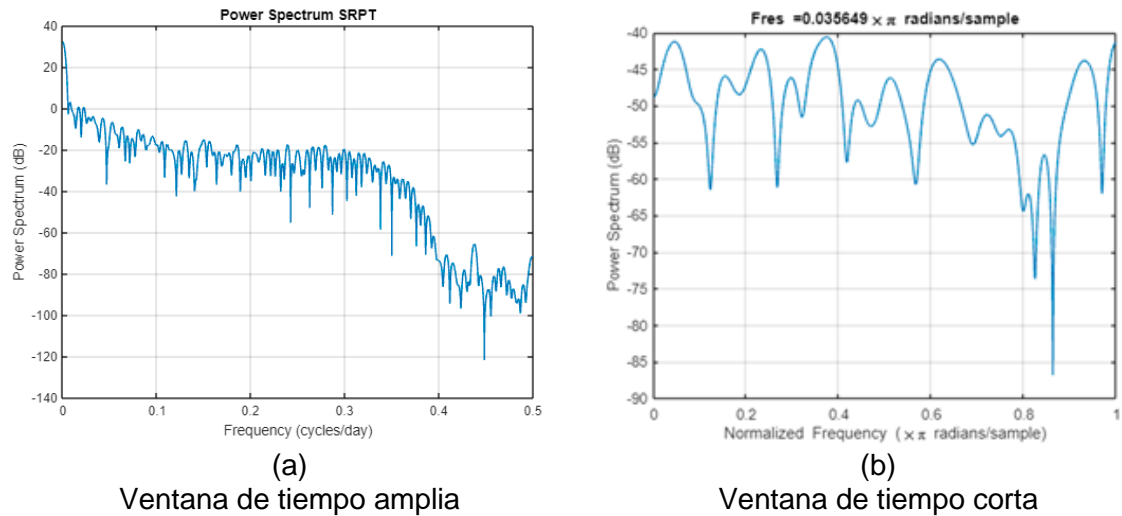
Por ejemplo, la Figura 11 (a) puede apreciarse el dominio de frecuencias de una acción que ha sufrido a una manipulación de tipo *pump and dump*.

La señal corresponde a los precios de cierre diario de la acción, sin embargo, en esta gráfica sólo se representa el comportamiento de la frecuencia que trae la señal y como hubo un cambio súbito en dicha frecuencia, el cual puede ser atribuido a la manipulación, ya que el cambio en las frecuencias es evidencia suficiente de que hubo una alteración de los precios.

Si bien esta gráfica no considera el dominio del tiempo, la señal corresponde a un año previo al inicio de la manipulación y un mes después de la misma. Si se hace el mismo análisis para la señal, un mes antes y un mes después del período de manipulación, la caída será más evidente como se puede ver en la Figura 11 (b).

En los espectrogramas se evidencia, como en presencia de manipulaciones se producen cambios energéticos abruptos al verse impactados los regímenes de frecuencia y por lo anterior, es fundamental considerar el ajuste del tamaño de las ventanas.

Figura 11. Espectro de Potencia de una acción con manipulación tipo *pump and dump*



Si bien puede parecer obvio dicho ajuste, parte del análisis de la señal y el fenómeno que está describiendo debe considerar cuáles son los tamaños de ventana más adecuados. Para ello han sido descrito incluso teoremas, que facilitan dicha elección.

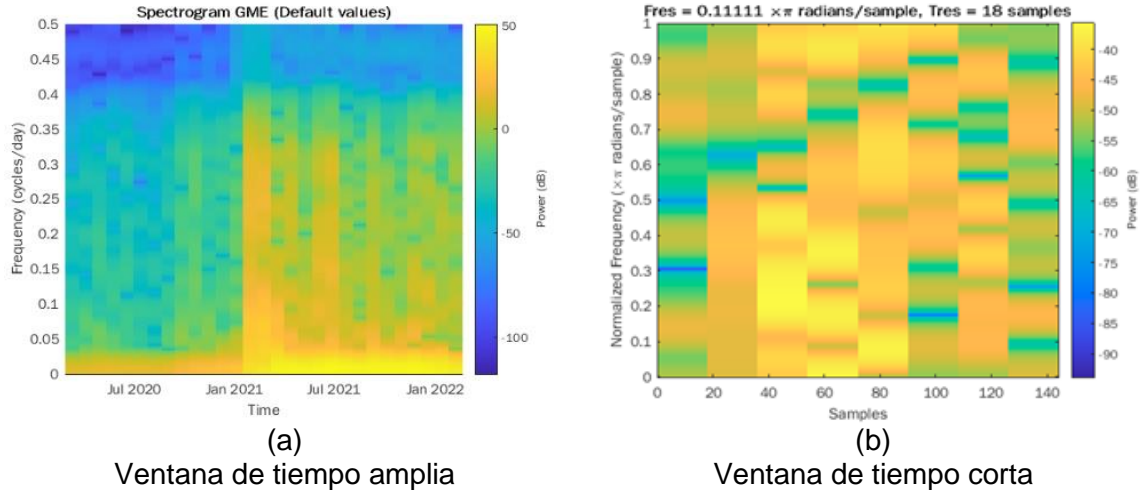
Sin embargo, dichos teoremas deben ir soportados en aquello que describe el fenómeno en sí, pues el objetivo a la final es encontrar los períodos que permitan recoger la información suficiente para estimar la aparición o no del fenómeno de manipulación, como también para poder analizar la señal como un todo, y por tanto son diversos los estudios enfocados en análisis empíricos de las señales como proponen Rizvi, et al. (2020b), Abbas, Belatreche & Bouridane (2018), Qu & Zhang (2016) y Golmohammadi & Zaiane (2015).

Con el fin de profundizar, respecto a los momentos en los que ese tipo de cambios abruptos se dieron, se pueden considerar los espectrogramas que incluyen el dominio del tiempo.

Generalmente los espectrogramas son producidos a partir de la transformada de tiempo corto de Fourier (o transformada corta de Fourier por sus siglas en inglés STFT – Short-Time Fourier Transform), conocida también como la transformada de Gabor. (Arts & van den Broek, 2022)

La STFT emplea una función de ventana, generalmente una ventana de Gabor (de allí su nombre alterno), ya que dicha ventana permite la descripción de la señal a partir de ondas cortas de ancho fijo localizadas tanto en tiempo, como en frecuencia lo que permite, generar mapas de calor como los que pueden apreciarse en la Figura 12 (b).

Figura 12. Espectrograma de una acción con manipulación tipo *pump and dump* con ventana de tiempo más corta



La Figura 12 (a) representa el mismo fenómeno de manipulación *pump and dump* de la Figura 11 y corresponde a los mismos períodos. Sin embargo, en este caso se puede dividir casi cualquier señal variable en el tiempo en intervalos de tiempo lo suficientemente cortos como para que la señal sea esencialmente estacionaria en cada sección. El análisis de tiempo-frecuencia se realiza generalmente en dichos intervalos, estimando su espectro a partir de ventanas deslizantes. (Harris, 1978)

No obstante, esto no siempre ocurre así, ya que como se mencionó previamente entre más pequeña sea el segmento de la ventana, menor será la capacidad de resolución de las frecuencias, por tanto, se debe considerar la compensación entre el tamaño de la ventana y la capacidad de resolución deseada, ya que a su vez ventanas muy grandes afectan la capacidad de ubicación de una señal en el tiempo.

Es muy grande la cantidad de parámetros que pueden ser ajustados en este tipo de kernels para obtener una mejor representación del fenómeno que se está estudiando. Sin embargo, dichos ajustes siempre deben considerar que debe haber un balance entre la exactitud del tiempo en la que se ubicará el fenómeno y la resolución de la frecuencia que la señal describe a partir del tamaño de la ventana elegido.

En consideración, los análisis de frecuencia efectuados mediante espectrogramas estarán restringidos a frecuencias con una longitud de onda cercana al ancho de la ventana. Adicionalmente, cortar la señal en segmentos de onda pequeños de ancho fijo, afecta las propiedades de la señal y en consecuencia el análisis puede verse comprometido. (Arts & van den Broek, 2022)

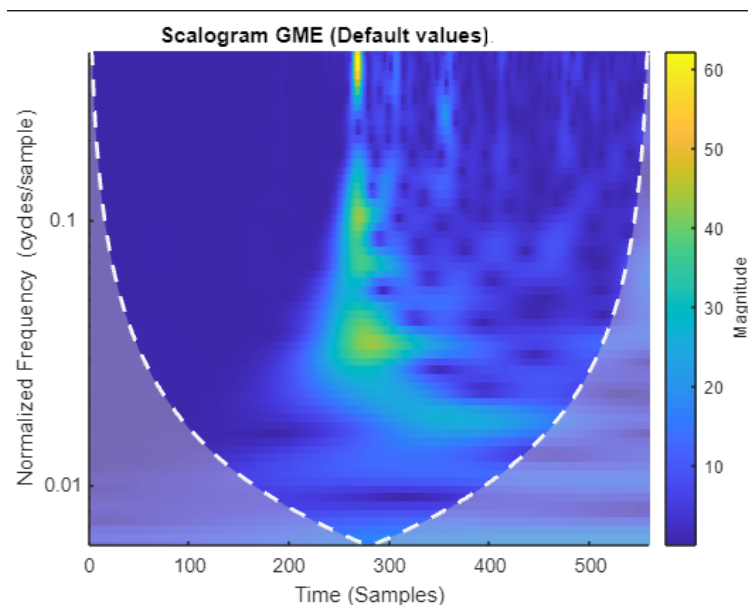
Para superar este inconveniente y no depender de una función de ventana, en su lugar se propuso el uso de funciones básicas que se dilatan y contraen a diversas escalas conocidas como 'wavelets'. (Mallat, 2009)

El uso de diversas escalas para representar la señal garantiza una alta resolución en todo el espectro de frecuencias, como también una alta resolución temporal ya que escalas finas de representación son incluidas. (Sejdic, Djurovic & Stankovic, 2008)

Como en el caso de los mapas, las escalas altas corresponden a una vista global no detallada (de la señal), y las escalas bajas corresponden a una vista detallada. De manera similar, en términos de frecuencia, las frecuencias bajas (escalas altas) corresponden a información global de una señal (que generalmente abarca toda la señal), mientras que las frecuencias altas (escalas bajas) corresponden a información detallada del patrón oculto en una señal (que generalmente dura relativamente poco tiempo). (Sejdic, Djurovic, & Stankovic, 2008)

El gráfico que representa este proceso se conoce como escalograma. En la Figura 13 puede apreciarse como la resolución espectral detalla en mejor medida los regímenes de frecuencia y detalla de manera precisa la ubicación temporal de cada evento de respuesta transitoria.

Figura 13. Escalograma de una acción con manipulación tipo *pump and dump* con ventana temporal a un año



Entre las ventajas que se obtiene al hacer uso de las wavelets, es que de esta manera el análisis puede realizarse de forma local en la señal, en contraposición a

las técnicas derivadas de la transformada de Fourier, en las que el análisis es inherentemente no local, debido a la naturaleza de las funciones trigonométricas que requieren ajustes en la señal para facilitar su procesamiento (submuestreos, rellenado de datos, etc.) (Sejdic, Djurovic, & Stankovic, 2008)

Sin embargo, el uso de las wavelets en diversas escalas implica una alta carga computacional, lo que restringe su uso en hardware de gama baja y en aplicaciones de tiempo real, ya que el tiempo de cómputo requiere que el tiempo de procesamiento logarítmico sea inferior al tiempo de duración de la señal. (Arts & van den Broek, 2022)

Cada representación tendrá sus ventajas y desventajas, sin embargo, como fue mencionado previamente, se pueden considerar las diversas herramientas como análisis complementarios que ayudarán a entender de mejor manera como se comporta una señal en presencia de manipulaciones.

Al considerar cada herramienta de forma particular, se podrá responder preguntas específicas como: ¿cuándo hay un componente de frecuencia particular en mi señal? ¿cómo puedo mejorar la resolución de tiempo o la resolución de frecuencia? ¿cómo puedo pulir el espectro de un elemento de la señal o extraer un elemento en particular? ¿cómo puedo identificar intermitencias en el contenido de frecuencia? (Varanis, Silva, Balthazar & Pederiva, 2021)

Teniendo lo anterior claro, y conociendo el fenómeno se podrá describir de mejor manera aquello que ocurre con la señal en presencia de manipulaciones.

Un detalle en particular al momento de considerar cada una de las representaciones, es que muchas señales del mundo real tienen un contenido de oscilación lento que ocurre en escalas largas, mientras que los eventos de alta frecuencia tienden a ser abruptos o transitorios, como se ha mencionado previamente.

Sin embargo, si fuera natural que los eventos de alta frecuencia tuvieran una duración prolongada, entonces no sería apropiado utilizar una wavelet, ya que se tendría una resolución de frecuencia más pobre sin ganar resolución de tiempo. (Zhang, 2019)

No obstante, este no suele ser el caso. El sistema auditivo humano funciona de esta manera; tenemos una localización de frecuencia mucho mejor en frecuencias más bajas y una mejor localización de tiempo en frecuencias altas. (Arts & van den Broek, 2022)

En el ejemplo de la Figura 13, se evidencia como la manipulación perturba los regímenes de frecuencia considerados normales para dicha señal. Además, puede apreciarse el llamado ‘cono de influencia’ (las líneas blancas con guiones) en la que los coeficientes de estimación son confiables (Torrence & Compo, 1998).

Contar con diversas escalas, hace que las representaciones mediante este kernel no dependan de una gran diversidad de parámetros en contra posición a las herramientas que hacen uso de la transformada de Fourier.

Sin embargo, es importante considerar que la elección del tipo de wavelet si incide de forma directa en la representación de la señal y, por tanto, deben ser elegidos aquellas wavelets que permitan el procesamiento de señales no estacionarias, altamente aleatorias, con singularidades y cuya frecuencia instantánea crecer rápido como es el caso de las series de precios de un activo financiero.

Para evaluar la pertinencia de cada herramienta deben ser consideradas todas las prácticas manipulativas previamente tipificadas y en cada caso, evaluar la capacidad de cada técnica para respaldar los hallazgos empíricos obtenidos como parte de la presente propuesta.

Por cada acción, fueron consideradas dos series de control, la primera el índice de precios del sector al que pertenece el activo, y la segunda otra compañía de dicho sector que no ha sufrido manipulaciones.

El objetivo de incluir información de control es facilitar las tareas de evaluación y validación de los resultados obtenidos con cada configuración de parámetros en cada herramienta.

Sin importar el tipo de práctica, para cada serie se define el mismo número de muestras diarias, en este caso quinientos registros por serie. Dicha cantidad corresponde a un período de dos años. A saber, un año antes y un año después de que iniciara la manipulación.

En algunos casos la duración del esquema manipulativo es hasta de un año, por lo que el margen de tiempo posterior a la manipulación es relativamente corto. En otros, cuando el esquema manipulativo se dio cerca de la emisión del activo, el tiempo posterior de validación fue superior a un año para completar las quinientas muestras diarias.

Es necesario precisar que no fueron incluidos como parte de las muestras los días en los que las bolsas están cerradas, pues la duplicidad artificial que supone considerar festivos y fines de semana, introduce ruido. Ya que se duplican precios y por tanto se afecta la resolución del espectro frecuencial. Este hecho impacta directamente en las capacidades de detección y aprendizaje.

Para todo el proceso de validación de las señales, y con el fin de procurar una mejor comprensión y descripción del fenómeno, cada señal debe ser tratada a partir del número de muestras (en este caso quinientas como se mencionó previamente) y no a partir de la escala temporal.

Lo anterior debido a que al tratar cada día como una muestra y no en su escala temporal, facilitará la generalización y la interpretabilidad de los parámetros al evitarse la necesidad de hacer conversiones innecesarias a Hertz.

Lo anterior, puede parecer simple, sin embargo, debe ponerse en consideración que las series de tiempo de los precios de los activos financieros son discontinuas en cuanto la tasa de muestreo no es homogénea, debido a que no fueron incluidos los días no bursátiles.

Además, la mayoría de las herramientas de procesamiento, requieren que los parámetros de procesamiento estén dados en Hertz, pues asumen que las señales provienen de procesos continuos con gran cantidad de muestras en intervalos muy cortos de tiempo, por tanto, al tener valores diarios de los precios, implicaría hacer cálculos en grandes magnitudes cuando finalmente la tasa de muestreo de los precios es nominalmente diaria.

Otra alternativa sería realizar submuestreos entre los valores diarios que afectarían el espectro de potencias al aumentar en forma proporcional el filtrado espectral.

Sin embargo, como consideración adicional las series de precios financieros en términos de señales ofrecen la magnitud del diferencial entre un precio y otro, no la frecuencia de aparición de un precio.

Al respecto, es necesario precisar que considerar la frecuencia de aparición de un precio por sí solo aportará poca información, pues desconoce el factor temporal de estas señales.

Si bien para términos de procesamiento la información temporal de las series de precios no es relevante, sí lo es para la descripción y comprensión del fenómeno, pues se debe respetar la secuencia en la que los precios fueron registrados.

Primero, porque dicha secuencia representa el comportamiento del activo en los mercados. Y segundo, las manipulaciones tienen un principio y un fin, por lo que requieren ser ubicadas en el tiempo.

Por lo que ajustar la cantidad de muestras, hacer conversiones de tasas y consolidar los datos mediante subtotales en esta etapa, dificulta la comprensión del fenómeno y los parámetros que serán evaluados.

Espectro de potencia

Previamente se ha hecho mención de que el espectro de potencia no es una herramienta recomendada para el procesamiento de señales financieras.

En primer lugar, debido a que no representa el dominio del tiempo. Dimensión fundamental para poder ubicar el momento en el que se presentó alguna singularidad en la señal.

Y, en segundo lugar, porque el espectro es generado mediante operaciones derivadas de la Transformada de Fourier. Dicha transformada asume estacionalidad, y por tanto no procesa de manera confiable señales resultantes de procesos estocásticos como lo son las señales financieras.

Sin embargo, es necesario, tener en cuenta que independiente de su origen, una señal necesita ser procesada para facilitar la generación, la transformación, la extracción y la interpretación de información que ésta lleva. (Arts & van den Broek, 2022)

En consideración, entender como interactúan los diversos componentes de una señal en términos de su frecuencia, detectabilidad de tonos, la capacidad de resolución, la identificación de patrones. Ayudará a entender de mejor manera cómo una señal financiera se ve afectada a causa de manipulaciones. Además, facilitará la estimación de los parámetros que permitan procesar las características previamente mencionadas. (Harris, 1978)

Por lo anterior, se hizo el cálculo del espectro de potencia haciendo uso de la Transformada Corta de Fourier con una tasa muestral diaria (es decir los precios de cada día fueron procesados uno a uno) para cada activo y sus respectivas series de control.

Se tuvo en consideración una ventana de Kaiser para el procesamiento de las muestras, ya que esta ventana es mucho más sencilla de procesar debido a que solo requiere de dos parámetros, el tamaño: $M + 1$, donde M corresponde al tamaño de la ventana, el cual no será mayor al tamaño de la señal y la forma: β , donde β es el factor de atenuación de lóbulos laterales.

Al atenuar los lóbulos laterales de la ventana, se ajusta la amplitud del lóbulo central. (Oppenheim, Schafer & Buck, 1999)

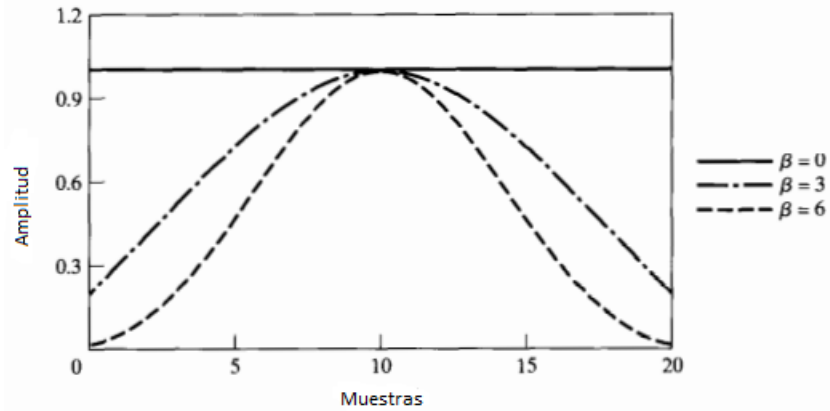
Es necesario recordar, que los lóbulos en una ventana se definen para controlar la cantidad de filtrado espectral que será considerado como parte del procesamiento de la señal al atenuar las discontinuidades en las señales causadas por el muestreo y las propiedades intrínsecas que constituyen la señal (impulsos, picos, etc).

Un factor $\beta = 0$, implica que la ventana no tendrá lóbulos y para efectos prácticos será una ventana rectangular. Un factor $\beta = 40$ hará tan amplio el lóbulo principal que permitirá que esencialmente se capture toda la energía espectral representable. (Oppenheim, Schafer & Buck, 1999)

Muchas herramientas de procesamiento incluyen como parámetro anexo un factor de filtrado espectral: ℓ , el cual atenúa aún más el tamaño de los lóbulos laterales mediante la siguiente operación $\beta = 40(1 - \ell)$.

En consideración, se debe validar dicho parámetro de filtrado, iniciando con un $\ell = 0$, lo que implica que no se atenuará el filtrado espectral y por consiguiente se tendrá un $\beta = 40$, pasando por diversos factores de atenuación hasta llegar a $\ell = 1$, donde $\beta = 0$, para tener una idea de los eventos de la ventana con diversos valores de β como puede apreciarse en la Figura 14.

Figura 14. Representación de la ventana de Kaiser con diversos valores de Beta



Fuente: Mallat, 2009

Como resultado de la validación previa se evidencia que un factor de $\ell = 0.25$, es idónea, ya que la relación entre resolución frecuencial y filtrado en este factor permite mostrar de manera más pronunciada ciertos picos que pueden asociarse a las prácticas manipulativas.

Respecto al tamaño de la ventana, la función considera la tasa de muestreo. Al no disponer de la información temporal, el análisis se hará muestra por muestra, por lo que cada registro de precios fue procesado.

Espectrograma

Los espectrogramas han sido empleados en gran variedad de aplicaciones, básicamente porque dan solución a uno de los principales problemas de la representación del espectro de potencia de una señal: contar con la dimensión del tiempo.

Previamente se ha hecho mención sobre la importancia de tener en consideración los dominios de representación del fenómeno que se está estudiando, pues cada dominio ofrece características particulares de una señal que explicarán a mayor profundidad tal fenómeno.

Lo anterior es clave, pues si bien esta herramienta también emplea diversas versiones de la Transformada de Fourier para su cálculo, y por tal motivo podría no considerarse una técnica adecuada (Boashash, 2015), profundizar en la relación de los diversos componentes de una señal, la detectabilidad de tonos, la capacidad de resolución de una señal y la identificación de patrones con relación al dominio del tiempo es necesario.

Por otra parte, es necesario tener en consideración que para el cálculo del espectrograma la señal es dividida en segmentos. Segmentos más largos ofrecen mejor resolución frecuencial y segmentos más cortos ofrecen mejor resolución temporal. (Fulop & Fitz, 2006)

Muchos sistemas de procesamiento controlan estos valores mediante parámetros que se encargan de ajustar el tamaño de la ventana que tengan implementada por defecto. Generalmente cuando estos valores no son dados, se procura encontrar un buen balance entre la resolución temporal y la resolución frecuencial, basado en el tamaño de la señal y la cantidad de muestras. (Fulop & Fitz, 2006)

Por tanto, si luego de ajustar este balance se ofrece una buena representación del fenómeno de manipulación, podría considerarse como una herramienta útil en la detección, particularmente por la facilidad de cómputo que ofrecen los procedimientos derivados de la Transformada de Fourier. (Fulop & Fitz, 2006)

En caso contrario, como se mencionó previamente, profundizar en la relación de los diversos componentes de la señal que permita caracterizar de mejor manera la manipulación es necesario.

La separación de datos agregados en diferentes escalas de tiempo y el análisis profundo del espectro de frecuencias es una poderosa herramienta para el análisis de datos financieros. Ya que las diferentes fuerzas del mercado afectan las relaciones económicas durante diferentes períodos de tiempo.

Los shocks económicos están localizados en el tiempo y dentro de ese período exhiben oscilaciones de frecuencia variable. (Gallegati & Semmler, 2014)

Y los espectrogramas en su mejor capacidad de resolución en cada dominio facilitarán tal entendimiento y posterior caracterización.

En consideración, se considera cada uno de los activos con sus respectivas series de control y se evalúa el espectrograma mediante dos técnicas: la Transformada Corta de Fourier (STFT) y la Transformada Rápida de Fourier (abreviada FFT por sus siglas en inglés *Fast Fourier Transform*).

La FFT es un algoritmo que permite calcular la Transformada de Fourier Discreta (abreviada DFT por sus siglas en inglés *Discrete Fourier Transform*) la cual al ser discretizada como indica su nombre, permite la representación del dominio temporal (Welch, 1967).

La particularidad de este algoritmo es que emplea técnicas de factorización y recursividad para procesar las frecuencias de acuerdo con la definición dada para la resolución temporal.

Al respecto es necesario mencionar que el tamaño de la señal y la cantidad de muestras son elementos determinantes al momento de definir la resolución temporal y a su vez las ventanas de procesamiento.

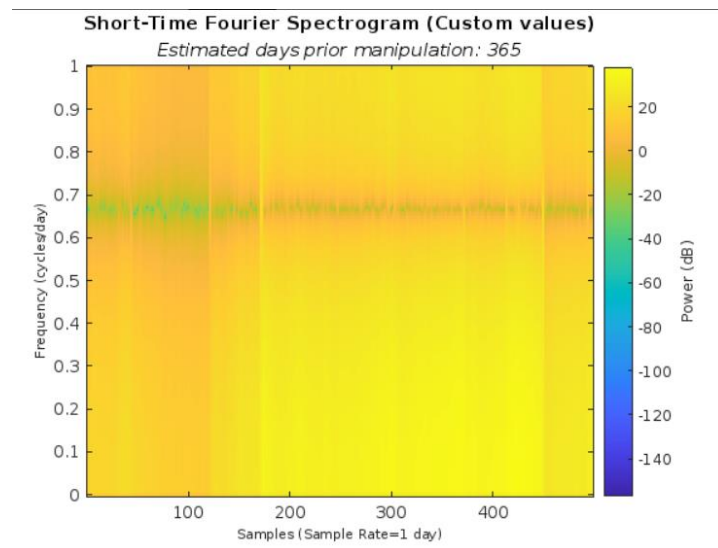
En relación y como previamente fue mencionado, elegir un tamaño de ventana corto ayuda a obtener una buena resolución temporal a expensas de la resolución frecuencial. Por el contrario, elegir una ventana más grande ayuda a obtener una buena resolución frecuencial a expensas de la resolución temporal.

Al respecto es necesario precisar, que una vez se elige un tamaño de ventana, este permanece fijo durante todo el análisis. Por tanto, si una ventana en particular está en capacidad de estimar los componentes de frecuencia que se esperan ver en la señal, se debe considerar tal información para elegir el tamaño de ventana para el análisis. (Harris, 1978)

A diferencia del cálculo del espectro de potencia, en el que cada muestra diaria pudo ser procesada en forma individual, en el caso del espectrograma el tamaño de la ventana elegida para procesar las muestras disponibles incide de forma directa en los detalles que se hacen visibles en el mapa de calor conforme acaba de ser mencionado.

Esto implica que, si se tomaran ventanas diarias o muy cortas, la resolución temporal se conservaría, sin embargo, la resolución frecuencial se vería comprometida, como puede apreciarse en la Figura 15.

Figura 15. Espectrograma basado en STFT de una acción con manipulación tipo *wash trade* con ventana de 3 días



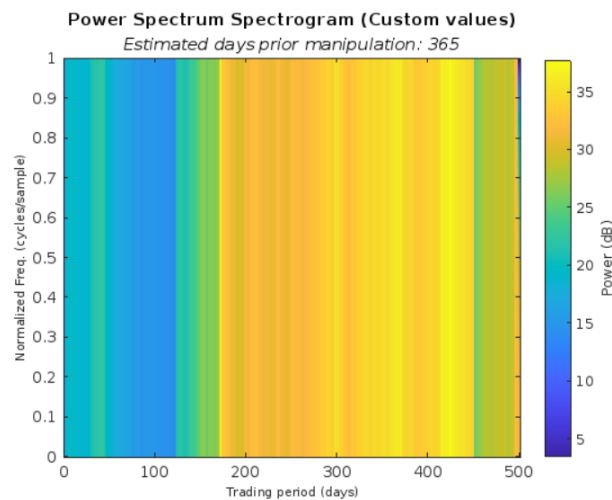
A veces este fenómeno se denomina ancho de banda de ruido equivalente, porque es proporcional a la potencia promedio que registrará cada ventana cuando la señal de entrada contiene un componente de ruido aleatorio (o es solo ruido aleatorio). (Oppenheim, Schafer & Buck, 1999)

En este gráfico claramente se ve un piso de ruido plano, causado por este efecto. La altura del ruido de fondo es proporcional al ancho de banda de la ventana y su relación con el filtrado espectral. Por lo tanto, dos funciones de ventana diferentes pueden producir diferentes ruidos de fondo. (Oppenheim, Schafer & Buck, 1999)

Lo cual no facilita su generalización y en algunos casos de aplicación puede ser considerado como un concepto subjetivo y ambiguo.

Un fenómeno similar puede apreciarse con el espectrograma basado en el espectro de potencia, en el que el detalle y la presencia de otros componentes de frecuencia se pierden, como puede apreciarse en la Figura 16.

Figura 16. Espectrograma basado en FFT de una acción con manipulación tipo *wash trade* con ventana de 3 días



En ambos casos se empleó una ventana rectangular a intervalos de tres días (muestras) y un solapamiento de un día (muestra). De esta manera se recoge la información de los saltos dados por las disparidades de los días de no bursatilidad, como también las razones de cambio que se dan a causa de las variaciones diarias en los precios.

Con el resultado obtenido, es posible identificar como se distribuye la energía, y validar que efectivamente las manipulaciones corresponden a eventos de alta energía y por tanto la evaluación del espectrograma resulta útil para la caracterización de una señal en presencia de una manipulación.

En caso de que se desee mejorar la resolución espectral, se debe considerar que la capacidad de resolución temporal se perdería debido a que las ventanas serán de mayor tamaño, en ciertos casos cuando el tamaño de las ventanas es grande, los análisis espectrales se harían sobre todo el conjunto de datos. (Harris, 1978)

Por otra parte, tomar ventanas más grandes podría derivar en que ciertos componentes de frecuencia no puedan ser resueltos, ya que por el proceso de ventaneo ciertos tonos podrían quedar solapados al estar incluidos en el mismo rango de ventana.

Adicionalmente, el hecho de que se empleen ventanas fijas implica que podría haber cortes que podrían afectar el régimen frecuencial de las manipulaciones al dividir la secuencia y agrupar cada segmento por separado con otros regímenes diferentes presentes en la señal.

Al respecto, Boashash (2015) precisa que ningún tamaño de ventana individual puede resolver todo el contenido de frecuencia de dichas señales, debido a la no estacionalidad de las series de datos financieras.

Por lo anterior, se puso en consideración encontrar un balance entre la resolución temporal y la resolución frecuencial con el fin de evidenciar, si este ajuste permite describir de forma suficiente la información de ambos dominios.

Para tal fin, y como se mencionó previamente se debe considerar la longitud de la señal completa para elegir la longitud de los segmentos. La resolución temporal se establece a partir de $\lceil \frac{N}{d} \rceil$ muestras, donde los símbolos $\lceil \cdot \rceil$ denotan la función de techo, N es la longitud de la señal y d es un divisor que depende de N , como puede apreciarse en la Figura 7, de la sección sobre el Espectro de Potencia.

De igual manera que con el Espectro de Potencia, al especificar la superposición entre segmentos contiguos, cambia el número de segmentos. Los segmentos que se extienden más allá del punto final de la señal se rellenan con ceros.

Los instantes de tiempo o el número de la muestra corresponden a los centros de los segmentos. (Welch, 1967)

Por último, es necesario precisar que debido a la STFT está en capacidad de determinar la frecuencia sinusoidal y el contenido de fase de las secciones locales de una señal, mientras que la FFT solo representa el espectro de potencia a partir de las ventanas definidas. (Fulop & Fitz, 2006) Conviene evaluar el balance entre ambos dominios sólo a partir de la STFT ya que el contenido de fase permite el cálculo de la densidad de poder espectral (se considera la distribución de la energía entre las frecuencias, no solo su poder instantáneo) (Harris, 1978) y de esta manera lograr una mejor representación incluso en escalas más pequeñas, como puede apreciarse en la Figura 15.

Por lo anterior, en el análisis de las señales de los precios con manipulaciones se procuró un equilibrio entre la resolución de componentes en la señal con frecuencia similar (alta resolución/sensibilidad) y la resolución de componentes con frecuencias diferentes (alto rango dinámico).

Los conceptos de resolución y rango dinámico tienden a ser algo subjetivos, dependiendo del problema que se esté tratando. Sin embargo, ambos conceptos

están altamente correlacionados con el filtrado espectral que es cuantificable. (Oppenheim, Schafer & Buck, 1999)

Por lo general, el filtrado espectral se expresa como un ancho de banda equivalente, B . Ya que distribuye los segmentos en una forma rectangular con una altura igual al máximo espectral y un ancho B . (Harris, 1978)

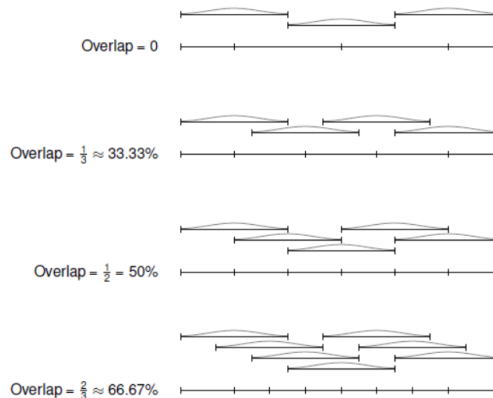
Según Oppenheim, Schafer & Buck (1999), cuanto mayor es el filtrado espectral, mayor es el ancho de banda. En consideración para cada caso se consideran valores intermedios para las ventanas, ya que capturan en la mejor medida ese balance entra la detección de los diversos componentes en la señal y su resolución temporal.

Lo anterior es relevante, pues una vez que se establecen la longitud y la superposición del segmento, la cantidad de segmentos y sus ubicaciones de borde permanecen fijos y son iguales para toda la señal.

Cuando los segmentos tienen un 0% de superposición, cada segmento se centra en el momento real en que ocurrió. Cuando la superposición es distinta de cero, la alineación del eje de tiempo del espectrograma con el eje de dominio de tiempo tiene el efecto de que el primer y último intervalo de se alarga. Todos los demás intervalos de tiempo tienen la misma duración. En otras palabras, el centro de cada segmento, excepto el primero y el último, corresponde al tiempo real de ocurrencia. (Oppenheim, Schafer & Buck, 1999)

Por tanto, valores intermedios están en capacidad de ajustarse mejor en las ventanas y de esta manera recoger en la mejor medida los componentes del espectro como puede verificarse en la Figura 17.

Figura 17. Diversos porcentajes de solapamiento para las ventanas



Fuente:
(Welch, 1967)

Espectrograma

La Transformada Wavelet supera el inconveniente de la Transformada de Tiempo Corto de Fourier de no lograr una buena resolución en el dominio tiempo-frecuencia al no depender de una función de ventana. En su lugar, utiliza una familia de funciones básicas que se dilatan y contraen a diversas escalas para representar la señal, lo que garantiza una alta resolución en todo el espectro de frecuencias. (Arts & van den Broek, 2022)

Escalas altas corresponden una vista global (no detallada) de la señal, mientras que las escalas bajas corresponden a una vista muy detallada. De manera similar, en términos de frecuencia, las frecuencias bajas (escalas altas) corresponden a información global de una señal (que generalmente abarca toda la señal), mientras que las frecuencias altas (escalas bajas) corresponden a información detallada del patrón oculto en una señal (que generalmente dura relativamente poco tiempo). (Sejdic, Djurovic & Stankovic, 2008)

Por tanto, resulta útil para la detección de prácticas manipulativas, pues dichos patrones, como se ha venido mencionando pueden ser detectados y aprendidos con mayor facilidad mediante estrategias de aprendizaje automático. Considerando incluso la posibilidad de emplear para tal fin aprendizaje de imágenes como estrategia de aprendizaje como lo hicieron Castrillón, Quintero-Montoya, & Sierra-Sosa (2018), al tratar los escalogramas como una imagen de tamaño fijo.

Si bien considerar la mejor resolución para facilitar cualquier tarea de aprendizaje es muy útil, es necesario considerar que una gran cantidad de escalas implica una alta carga computacional, con la CWT, se pasa de N muestras para una señal de longitud N a una matriz de M -por- N coeficientes, donde M es igual al número de escalas. (Zhang, 2019)

El CWT es una transformación altamente redundante. Hay una superposición significativa entre las wavelets en cada escala y entre escalas. Por tal razón los recursos computacionales necesarios para calcular la CWT y almacenar los coeficientes son mucho mayores que los de la DWT.

Lo anterior se debe en parte a los efectos del sub-muestro y al uso de octavas con valor igual a 1 en la DWT para su discretización. Sin embargo, esto que implica que las escalas son más amplias, en contraposición al caso más detallado de la CWT.

Más allá del tamaño de las escalas y su efecto en el nivel de detalle que se puede obtener mediante una wavelet continua o una wavelet discreta. Las diferencias en los procesos de discretización de la escala dan como resultado ventajas y desventajas para las dos clases de transformadas wavelets en casos de aplicación muy específicos:

- La DWT está en capacidad de comprimir una señal, pues las características más importantes de muchas señales son capturadas en un subconjunto

específico de coeficientes, debido a que muchos de ellos son cercanos a cero y pueden ser descartados. Por su parte la CWT, se considera una transformada altamente redundante, pues hay una superposición entre las wavelets en cada escala y entre escalas. (Addison, 2018)

- La DWT es equivalente a un filtro de sub-bandas de paso bajo y de paso alto, que no requiere de la definición explícita de una wavelet, ya que el resultado del filtro de paso bajo se filtra de forma iterativa hasta llegar a una sub-banda con octavas muy pequeñas, aplicando el mismo esquema de filtros de paso bajo y de paso alto en cada nivel. Por tal motivo, la salida es sub-muestreada (es decir se reduce la cantidad de muestras). Este no es el caso de la CWT, pues la wavelet es explícitamente definida y operada contra la señal en cada segmento y escala. (Addison, 2018)
- La DWT no es invariante al cambio de fase, debido a que reduce la cantidad de muestras. Esto implica que un cambio en la señal de entrada no se manifiesta como un cambio equivalente en los coeficientes calculados en todos los niveles, sino que causará una realineación significativa por escala. Por su parte la CWT si es invariante al cambio de fase, lo que implica que la energía se mantiene y es bien representada en todos los niveles. (Addison, 2018)
- La transformada wavelet discreta proporciona una reconstrucción perfecta de la señal al calcular su inversa. Mientras que la transformada wavelet continua no lo hace, pues el cálculo de su inversa es una operación numérica menos estable debido al efecto de las escalas solapadas. (Addison, 2018)
- El muestreo más fino de escalas de la CWT generalmente da como resultado un análisis de señal de mayor fidelidad, lo que facilita que la localización de picos en una señal, como también la caracterización del comportamiento oscilatorio, cosa que la DWT no permite.

En consideración sobre las ventajas de cada clase, para el presente caso de estudio se favorece el uso de la Transformada Wavelet Continúa en la representación del escalograma, pues al ser explícitamente definida y operada contra la señal en cada segmento y escala, su capacidad de resolución en el dominio tiempo-frecuencia es muy fina lo que garantiza una excelente localización temporal sin sacrificar la resolución espectral como sí ocurre en el caso del espectrograma.

La DWT proporciona una resolución de escala deficiente, por lo que no se pueden detectar algunas características significativas de una señal. (Addison, 2018)

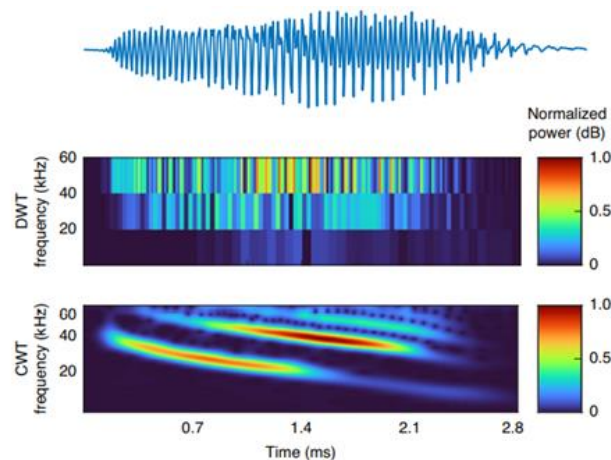
En segundo lugar, como la CWT es invariante al cambio de fase, la energía se mantiene entre octavas, y de esta manera se contribuye a una buena localización y representación de la potencia en todos los niveles, ya que la energía total en el dominio de la escala se conserva antes cambios en la señal. (Addison, 2018)

De forma similar ocurre con el muestreo más fino que la CWT brinda. A diferencia del DWT, la CWT permite que los parámetros de escala y traducción cambien continuamente, lo que lo hace invariable y redundante en la traducción. La redundancia garantiza que la CWT pueda proporcionar resultados de alta resolución, mucho más fáciles de interpretar que los obtenidos con la DWT. (Wang, Yang & He, 2022) sin saltos abruptos o cortes introducidos por el computo de la señal.

Siendo esto último de gran importancia pues es lo que facilita que la CWT haga una adecuada localización de picos y una detalla caracterización de comportamientos oscilatorios que pueden estar presentes en la señal. Siendo estos precisamente los elementos que considerar para la detección y descripción de manipulaciones en activos financieros.

Lo anterior puede apreciarse con facilidad en la Figura 18, en la que se representa una señal emitida por un sonar con una duración inferior a 3 milisegundos y un rango de frecuencia de 0 a 60 kHz.

Figura 18. Comparación del espectrograma generado mediante CWT y la DWT



Fuente:
Arts & van den Broek, 2022

Claramente se nota como la CWT logra un detalle fino que representa las frecuencias, su ubicación y la energía concentrada durante toda la duración de la señal, mientras que la DWT presenta un resultado similar al obtenido con el espectrograma.

Esto se logra gracias a que la CWT compara la señal original con versiones comprimidas o estiradas, y con cambios de fase de la wavelet elegida. (Paniagua-Castrillón & Quintero-Montoya, 2018)

Estirar o comprimir una señal se conoce como dilación o ampliación y se corresponden con la noción física de escala. Esto es fundamental, pues al comparar una señal en varias escalas y posiciones, se obtiene una función de dos variables (lo que implica que la representación de una señal de un solo canal -1 Dimensión- es redundante)

Adicionalmente, si la wavelet incluye valores complejos, la transformada wavelet continua tendrá valores complejos de escala y posición, lo cual es útil para identificar contenido de fase de una señal (información útil para comparar señales que resulta a partir de correlacionar las frecuencias de ambas señales). (Percival & Walden, 2000)

Por su parte, los cambios de fase corresponden simplemente al hecho adelantar o retrasar el punto de partida sobre el cual la wavelet procesará toda la señal.

Sin embargo, no solo los valores de escala y posición afectan los coeficientes de la transformada, sino también, la elección de la wavelet como tal, ya que su forma al igual que en casos anteriores redundará en la capacidad de detección de ciertos artefactos y la precisión de la resolución frecuencial a causa de los efectos provocados por el filtrado espectral.

Deben favorecerse entonces aquellas wavelets que mejor representen transiciones abruptas de frecuencia en una señal, pues estas transiciones también pueden dar cuenta si efectivamente una serie de precios de activos financieros fue objeto de alguna práctica manipulativa.

Estas transiciones abruptas, resultan en coeficientes de alto valor absoluto (ya que en muchos casos el signo no importa mucho, ya que este depende si la fase de la señal coincide con la fase de la wavelet). (Addison, 2018)

Para este objetivo, se debe favorecer el uso de wavelets que permitan capturar en mayor medida los cambios de fase, pues este factor incide en que las wavelets continuas sean una buena herramienta para detectar discontinuidades o singularidades a diferencia de las wavelets discretas. (Addison, 2018)

Sin embargo, debe tenerse en consideración que entender los efectos que dichas discontinuidades producen en la señal es igualmente relevante.

Previamente con ayuda de los espectrogramas se evidenció como las manipulaciones se consideran eventos de alta energía y por tanto detectar tales eventos en presencia de discontinuidades o cambios abruptos en la señal hace parte de la caracterización de las prácticas manipulativas.

Por otra parte, comprender los comportamientos oscilatorios resultantes de dichos picos y los efectos posteriores en la señal, también hacen parte de los elementos a considerar para la detección y descripción de las prácticas manipulativas.

En consideración, los tipos de wavelets que recoge en mejor medida los criterios antes expuestos son las wavelets analíticas. Estas son wavelets con espectros unilaterales y un valor complejo en el dominio del tiempo. (Lilly, 2017)

Se dice que una función es analítica si su transformada de Fourier es cero para frecuencias negativas. Las funciones analíticas son necesariamente complejas, sin embargo, son caracterizadas por su parte real, debido a que se basan casi exclusivamente en funciones que disminuyen su valor rápidamente. (Mallat, 2009)

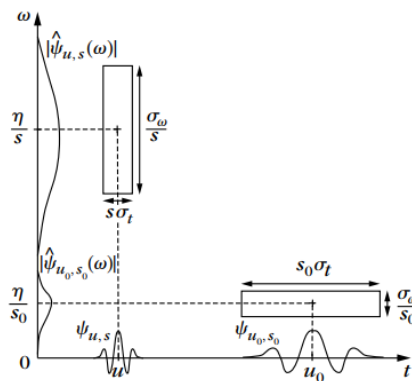
Si $\psi(t)$ es una función analítica que decrece rápidamente en el tiempo, entonces su transformada de Fourier $\psi \cdot (\omega)$ es una función que disminuye rápidamente en frecuencia y es pequeña fuera de algún intervalo $\alpha < \omega < \beta$, donde $0 < \alpha < \beta$. (Lilly, 2017)

Vale la pena aclarar que precisamente la parte analítica de una señal es la inversa de la transformada de Fourier. De esta manera se define una densidad de energía de frecuencia de tiempo local. (Mallat, 2009)

La resolución de la densidad espectral depende de la dispersión de tiempo-frecuencia de cada versión estirada, comprimida y trasladada de la wavelet (conocidas como átomos de la wavelet). (Mallat, 2009)

Debido a que la dilatación se logra a partir de un factor racional s , y las frecuencias negativas son cero. El segmento procesado para las frecuencias permanece igual a t en todas las escalas, afectando únicamente la capacidad de resolución y frecuencia según se ilustra en la Figura 19.

Figura 19. Representación de la capacidad de resolución y el soporte frecuencial de las wavelets analíticas



Fuente:
Mallat, 2009

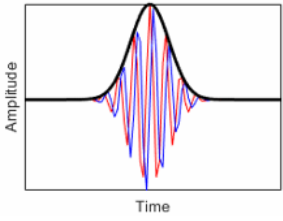
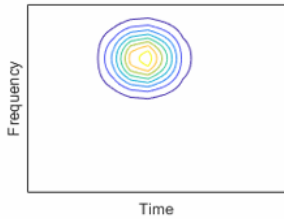
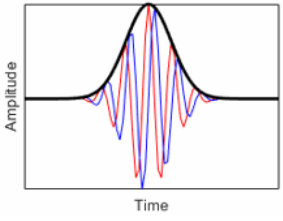
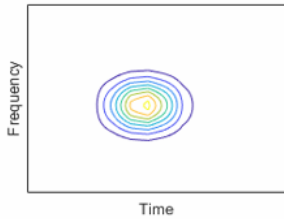
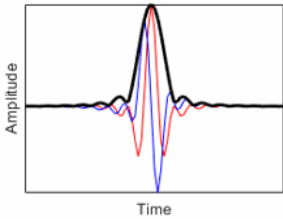
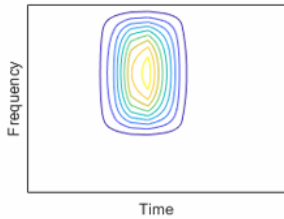
En parte lo anterior explica, porque las wavelets continuas son invariantes al cambio de fase, pero susceptibles a ocasionar alteraciones al momento de calcular la transformada inversa para reconstruir la señal. Sin embargo, a su vez estas propiedades son las que brindan la capacidad de medir la evolución en el tiempo de los cambios abruptos en la frecuencia y de capturar el comportamiento ondulatorio en una señal.

Un beneficio adicional que ofrecen las wavelets analíticas es que, al ser complejas, los coeficientes que resultan mediante esta transformada también lo son, por tanto, están en capacidad de proporcionar información de fase y amplitud de la señal que se analiza. (Lilly, 2017)

Ya que, al conservar los centros de las frecuencias en todas las escalas, se puede separar con facilidad los componentes de amplitud y de fase. (Mallat, 2009)

Entre las wavelets analíticas muchos sistemas de procesamiento de señales favorecen el uso de las wavelets descritas en la Figura 20.

Figura 20. Wavelets Analíticas Recomendadas

| Wavelet | Principal Característica | Dominio del Tiempo | Dominio Tiempo-Frecuencia |
|--|---|--|---|
| Wavelet Morse Generalizada | Hace uso de los parámetros β y γ para ajustar la dispersión de tiempo frecuencia de los átomos de la wavelet |  |  |
| Wavelet Gabor (Wavelet Morlet Analítica) | Iguala la varianza en tiempo y frecuencia para procurar un buen balance en la resolución en ambos dominios |  |  |
| Bump Wavelet | Aplica una variación más amplia en el tiempo y una variación más estrecha en la frecuencia |  |  |

Fuente:
The MathWorks, Inc., n.d.

Estas wavelets favorecen un buen balance en la representación conjunta del tiempo-frecuencia, y procuran cumplir el objetivo de lograr una buena resolución espectral sin sacrificar la resolución temporal.

En caso de que se desee mejorar la resolución de alguno de los dominios en particular, cada una de las wavelets descritas será la wavelet recomendada para tal labor.

En el caso de la wavelet Morse generalizada, se puede variar los parámetros β y γ para recrear otras wavelets de uso común. En las gráficas de la Figura 20 los gráficos de contorno muestran la dispersión de las wavelets en el tiempo y la frecuencia, para dar una idea de los resultados que se pueden obtener con cada una. (Y en el dominio del tiempo, las líneas roja y azul son las partes real e imaginaria, respectivamente, de la wavelet) (The MathWorks, Inc., n.d.)

Las wavelets de Morse generalizadas son una familia de wavelets analíticas exactas. Es decir, tienen expresión analítica tanto para la forma de la región de concentración como para la concentración de energía fraccional, ya que los parámetros β y γ son libres, pues el decaimiento en frecuencias más altas y en frecuencias más bajas desde el centro de la banda de paso se puede especificar de forma independiente. (Lilly & Olhede, 2008)

Estos dos parámetros producen una caracterización natural de una función wavelet analítica, además de permitirle adoptar una amplia gama de formas que aún no se han explorado por completo. De hecho, esta familia abarca la mayoría de las otras wavelets analíticas populares. (Lilly & Olhede, 2008)

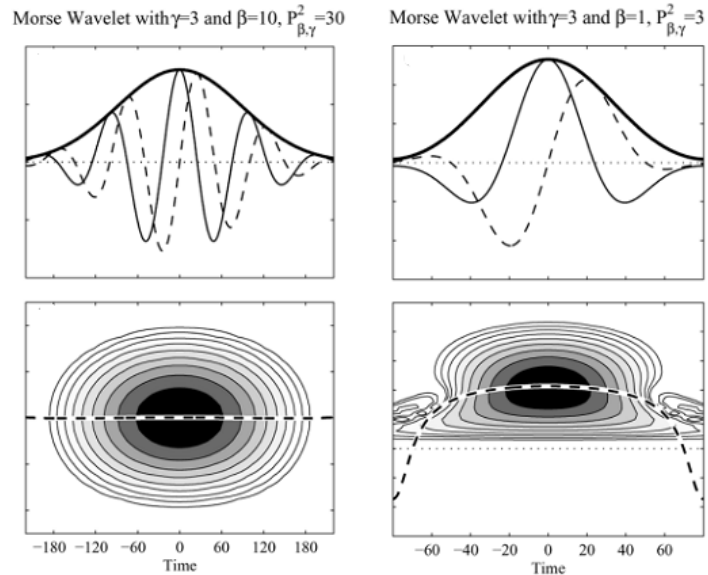
Estas wavelets, además de ser una buena solución a un problema conjunto de localización de tiempo/frecuencia, ofrecen ventajas adicionales, por ejemplo, las wavelets aproximadas y las wavelets reales, imprimen filtrado espectral en frecuencias negativas, las cuales pueden dar como resultado una variación espuria de la fase de transformación. (Lilly & Olhede, 2008)

Otra ventaja adicional, es que estas wavelets también están en capacidad de reflejar la frecuencia instantánea (una cantidad que refleja el contenido de frecuencia variable en el tiempo de una señal oscilatoria modulada) sin afectar significativamente la resolución temporal, en caso de favorecer la resolución frecuencial y viceversa. (Lilly & Olhede, 2008) Como puede verse en la Figura 20.

Lo anterior, debido a que es capaz de mantener de manera constante a diversas escalas las frecuencias instantáneas (al estar concentrada en frecuencias positivas). (Lilly & Olhede, 2008)

Y esto es importante para el análisis de funciones fuertemente moduladas o aleatorias, donde se requiere que las wavelets sean estrechas en el tiempo para coincidir con la escala de tiempo de modulación. (Lilly & Olhede, 2008)

Figura 21. Detalle de como la Wavelet Morse mantiene un buen balance entre resolución frecuencial y temporal



Fuente:
Lilly & Olhede, 2008

Teniendo en consideración las propiedades previamente definidas y sus beneficios en garantizar una alta resolución de tiempo/frecuencia con un buen balance entre ambas y un sesgo minimizado para analizar señales oscilatorias, deben ser consideradas estas wavelets como parte del análisis de las manipulaciones en activos financieros, pues es necesario evaluar la concentración de alta energía y los cambios abruptos en la señal conservando una buena resolución temporal.

Finalmente, deben definirse los mejores parámetros que favorezcan la representación del fenómeno, para ello se puede partir de los análisis propuestos por Lilly & Olhede (2008).

En sus investigaciones ellos no solo describieron y ampliaron la caracterización y los análisis sobre esta familia de wavelets que fueron propuestos en la investigación seminal de Olhede & Walden (2002), sino que también propusieron otros criterios de aplicación y selección que vale la pena considerar.

La estructura de una wavelet se puede describir en términos de sus momentos o cumulantes, tanto los momentos wavelet en el dominio de la frecuencia como los momentos de energía. Técnicamente, estos se denominan “momentos formales” y “cumulantes formales”, ya que la wavelet en el dominio de la frecuencia no está normalizada como una función de densidad de probabilidad. (Lilly & Olhede, 2008)

Cuando los cumulantes se expanden en términos del dominio de frecuencia de la wavelet, sólo los cumulantes impares se registran cuando se quita la parte imaginaria. Dicha cantidad es una medida adimensional de la curvatura de la frecuencia instantánea evaluada en el centro de la wavelet, cuya interpretación es una función de densidad de probabilidad cuya asimetría en el dominio de la frecuencia corresponde a la curvatura de la frecuencia instantánea. (Lilly & Olhede, 2008)

El comportamiento de estas medidas es recogido por los parámetros de la wavelet donde γ caracteriza la simetría de la wavelet y β es visto como un parámetro de descomposición o compacidad de la wavelet, la cual permite definir el producto del ancho de la banda temporal de la wavelet mediante la siguiente expresión $P^2 = \beta\gamma$. (Lilly & Olhede, 2008)

De esta manera se puede determinar la forma de la wavelet y afectar el comportamiento de la transformada. El parámetro gamma de wavelet Morse, γ , controla la simetría de la wavelet en el tiempo a través de la asimetría demodulada, mientras que la raíz cuadrada del producto tiempo-ancho de banda, P , es proporcional a la duración de la wavelet en el tiempo. La duración determina cuántas oscilaciones pueden caber en la ventana central de la wavelet en el dominio del tiempo en su frecuencia máxima. (Lilly & Olhede, 2008)

El comportamiento de estas medidas para algunos valores de gama y beta pueden apreciarse en la Figura 22.

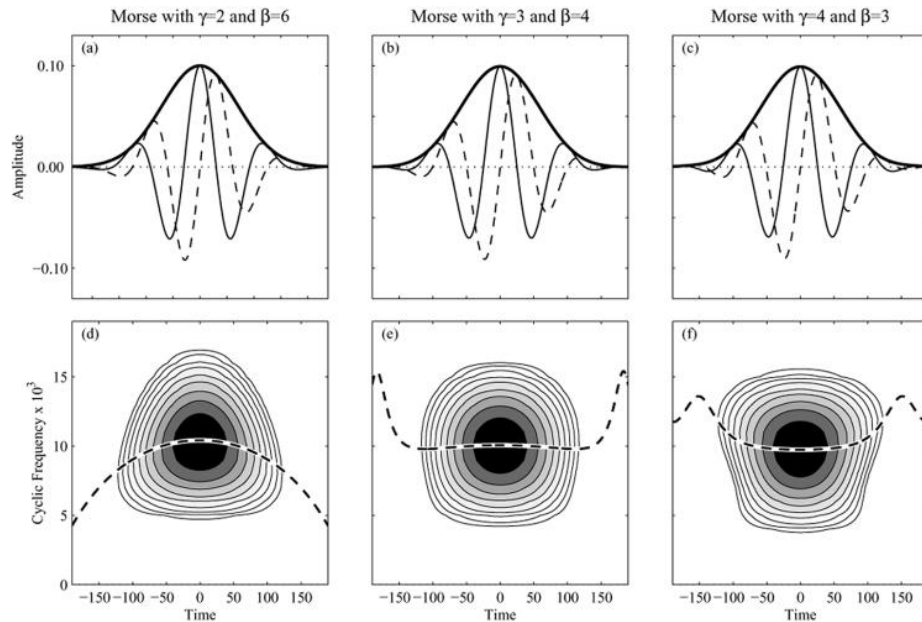
Estos gráficos muestran cómo los diferentes valores de simetría y ancho de banda de tiempo afectan la forma de una wavelet Morse.

Los anchos de banda de tiempo más largos amplían la porción central de la wavelet y aumentan la tasa de decaimiento de tiempo largo. El aumento de la simetría amplía la envoltura de la wavelet, pero no afecta el decaimiento a largo plazo. (Lilly & Olhede, 2008)

Para valores de simetría menores o iguales a 3, la disminución del tiempo aumenta a medida que aumenta el ancho de banda de tiempo. Para una simetría mayor o igual a 3, la reducción del ancho de banda de tiempo hace que la wavelet sea menos simétrica. A medida que aumentan tanto la simetría como el ancho de banda temporal, la wavelet oscila más en el tiempo y se estrecha en frecuencia. Un ancho de banda de tiempo muy pequeño y valores de simetría grandes producen lóbulos laterales en el dominio del tiempo y asimetría en el dominio de la frecuencia no deseados. (Lilly & Olhede, 2008)

Lo anterior implica, que la existencia de una correspondencia única entre la escala y la frecuencia requiere simetría sobre la frecuencia máxima, medida por la asimetría en el dominio de la frecuencia, y también requiere igualdad entre la media y la moda del módulo cuadrático de la wavelet en el dominio de la frecuencia. (Lilly & Olhede, 2008)

Figura 22. Comportamiento de la wavelet Morse con diversos parámetros



Fuente:
Lilly & Olhede, 2008

Por lo tanto, con una propiedad de orden inferior mantenida fija, como la duración de la wavelet en proporción a su período, elegir una wavelet que tenga un alto grado de simetría tanto en el dominio del tiempo como en el dominio de la frecuencia conduce a un buen desempeño para el análisis de señales oscilatorias. (Lilly & Olhede, 2008)

Con un valor de gamma fijo, incrementando el producto de la banda de tiempo P^2 se reduce la frecuencia del filtro wavelet, mientras aumenta el ancho de la parte central del filtro en el tiempo. Aumentando también el número de oscilaciones de la wavelet debajo de la parte central del filtro. (Olhede & Walden, 2002) Y de esta manera lograr una excelente resolución frecuencial, sin sacrificar la resolución temporal.

En consideración, para el cálculo de la Transformada Wavelet Continua de cada activo manipulado, se debe emplear la wavelet analítica de Morse con el parámetro de simetría, gamma (γ) igual a 3 y el producto tiempo-ancho de banda igual a 60. La implementación de la CWT empleando 10 voces por octava y las escalas mínima y máxima se determinaron automáticamente en función de la dispersión de energía de la wavelet en frecuencia y tiempo de cada serie de tiempo evaluada.

El cono de influencia, y la frecuencia además son descritos en términos de ciclos por muestra.

Para la resolución temporal, se puede considerar cada muestra de manera individual, lo que implica una resolución temporal precisa, que permite un amplio detalle en el espectro frecuencial representado en términos de la densidad de energía de la dimensión conjunta tiempo-frecuencia.

Como resultado de procesar la señal mediante estos parámetros, se obtiene una matriz tridimensional, donde los índices corresponden a los centros de frecuencias detectados y a la resolución temporal definida.

Las escalas son definidas por coeficiente complejos que representan la menor área temporal disponible a partir de los máximos de frecuencia calculados para la wavelet obtenidos a partir de la frecuencia instantánea detectada en la señal. (Lilly & Olhede, 2008)

Lo que implican un sesgo minimizado para analizar señales aleatorias. Al respecto, al procesar una señal mediante una CWT y un escalograma, se obtiene también el cono de influencia.

Como fue mencionado previamente, este corresponde a la línea punteada del escalograma. A medida que disminuye la frecuencia, aumenta el ancho de los coeficientes de la CWT en el tiempo que son distintos de cero y están centrados en el impulso de cada muestra analizada. Por el contrario, a medida que aumenta la frecuencia, el ancho de los coeficientes de CWT que son distintos de cero disminuye y se centra cada vez más en el impulso. (Torrence & Compo, 1998)

Las frecuencias bajas corresponden a wavelets de escala más larga, mientras que las frecuencias más altas corresponden a wavelets de escala más corta. (Lilly, 2017)

El efecto del impulso persiste más tiempo con wavelets más largas. En otras palabras, cuanto más larga sea la wavelet, mayor será la duración de la influencia de la señal. Por tanto, el límite en el que este pulso centrado está en capacidad de recoger información de la señal es lo que se conoce como el cono de influencia de la wavelet. (Torrence & Compo, 1998)

VENTANEO DE LAS SEÑALES

Luego de evaluar la pertinencia de cada herramienta para la detección y la caracterización de prácticas manipulativas, es importante reconocer que emplear técnicas para el procesamiento de señales efectivamente es una poderosa herramienta para el análisis de datos financieros y en este caso particular una herramienta adecuada para la detección y caracterización de prácticas manipulativas.

Entre las principales ventajas que ofrece el procesamiento de señales para el análisis de prácticas manipulativas, está la separación de datos agregados en diferentes escalas, pues de esta manera se obtiene una perspectiva interesante sobre como diversas fuerzas en el mercado afectan las relaciones económicas en diferentes períodos de tiempo. (Gallegati & Semmler, 2014)

Al respecto, previamente fue mencionado como choques en los mercados afectarán el comportamiento de una señal y pueden ser entendidos como singularidades, ya que imponen una variación abrupta de los componentes de frecuencia de la señal, no solo como un salto abrupto tanto en la energía que ésta representa, como en su propia magnitud, sino también exhibiendo oscilaciones sobre el resto del espectro.

En las diversas herramientas pudo validarse como ante eventos de manipulación, los regímenes de frecuencia de las series de tiempo financiera se ven afectados.

Dicha afectación se pudo evidenciar como cambios drásticos de energía, saltos abruptos en la magnitud del régimen frecuencial exhibido y oscilaciones en la frecuencia que previamente no estaban presentes.

Cada herramienta de acuerdo con sus dominios de representación, el balance que ofrece entre la resolución temporal y la resolución frecuencial, la posibilidad de detectar una respuesta transitoria o variaciones abruptas y la concentración de energía aportó al entendimiento del problema.

En el Espectro de Potencia pudo evidenciarse como las manipulaciones efectivamente afectan el espectro, al registrar picos que se hacen muy notorios en segmentos de tiempo corto.

El Espectrograma permitió evidenciar a expensas de la resolución temporal, como las manipulaciones se evidencian como cambios abruptos en el régimen de frecuencia de la señal, como también a partir de concentraciones altas de energía.

El Escalograma permitió identificar las oscilaciones en la frecuencia a causa de los efectos causados por la manipulación en diversas escalas.

Sin embargo, tanto en términos de la representación, como en la capacidad de explicar de manera más holística el fenómeno, las técnicas tradicionales de análisis espectral tienen varias deficiencias. Por ejemplo, la transformada de Fourier no puede analizar y representar con precisión una señal que tiene componentes no

periódicos (Boashash, 2016), como es el caso de las señales financieras al ser fenómenos estocásticos.

Al respecto, la transformada de Fourier, es inherentemente no local debido a la naturaleza de las funciones trigonométricas de rellenar el espacio y a la dependencia de la definición de límites fijos para los rangos de análisis de la señal, lo que no solo restringe y afecta la resolución temporal, sino que también dificulta la identificación de eventos transitorios. (Sejdic, Djurovic & Stankovic, 2008).

Por lo anterior, es necesario precisar que otra gran debilidad del análisis espectral tradicional, es su incapacidad para proporcionar información en la dimensión del tiempo del contenido de frecuencia de las señales no estacionarias (Boashash, 2016), como las de los precios de los activos financieros.

Asimismo, su implementación en el Espectro de Potencia carece de una representación del dominio del tiempo. Al respecto, el Espectrograma tampoco ofrece una mejor solución a este inconveniente, pues debe sacrificar la resolución temporal para lograr una buena resolución frecuencial y viceversa, lo que condiciona su uso.

Por otra parte, las señales transitorias pueden ser una buena fuente de información sobre el estado de un sistema que no está disponible durante los estados frecuenciales estables. Se sabe que los métodos basados en la transformada de Fourier son inadecuados para representar este tipo de señales, ya que los eventos transitorios difícilmente pueden aproximarse mediante senos y cosenos. (Boashash, 2016) Además, de ser altamente susceptibles a los efectos del filtrado espectral, a causa de los efectos del submuestreo de la señal.

Controlar la cantidad de filtrado espectral es fundamental al momento de evaluar manipulaciones, pues discontinuidades en la señal, son responsables de dejar marcas o filtrado sobre toda la base. (Harris, 1978) Y en caso de no controlar el filtrado a causa del submuestreo, se imprime un ruido adicional sobre el espectro de potencia.

En ambos casos las herramientas derivadas de la Transformada de Fourier no son las más adecuadas, pues el balance entre resolución temporal y resolución frecuencial, el balance entre filtrado espectral y resolución frecuencial y el balance entre escala y capacidad de resolución de picos; afectan enormemente los resultados, como pudo evidenciarse.

Por tanto, para la detección y representación de prácticas manipulativas en activos financieros, las herramientas basadas en wavelets se proponen como mejor alternativa de representación, pues son capaces de superar las deficiencias de las técnicas tradicionales de análisis espectral.

Estas herramientas como pudo evidenciarse están en capacidad de representar las señales en un espacio de frecuencia más detallado sin sacrificar la ubicación en la dimensión temporal de los artefactos hallados.

Incluso permite una discriminación más directa y precisa entre las condiciones normales y anómalas de la frecuencia a causa de las manipulaciones, al registrar las singularidades como picos de energía, debido a su notable capacidad para mantener una resolución óptima tanto en el espacio como en la escala. (Mallat, 2009)

Además, permite representar los efectos en el espectro de frecuencia introducidos por las discontinuidades que las manipulaciones generan, al hacer evidente su marca en el espectro.

En particular, se confirmó como el Escalograma en uso de una wavelet continua, logra una muy buena representación del régimen frecuencial y su densidad espectral, manteniendo una resolución temporal en pleno detalle de esta dimensión.

Además, se confirmó que está en capacidad de representar eventos transitorios en la señal como eventos de alta energía, como también las marcas que estos dejan en el resto del espectro. Ya que, al emplear gran cantidad de versiones estiradas, comprimidas y con fase ajustada de una wavelet es posible realizar un análisis multi-resolución.

Este tipo de análisis, según previamente fue expuesto favorece la detección de prácticas manipulativas.

También, es necesario precisar que escalas bajas favorecen la detección de singularidades en una señal y escalas más altas favorecen la detección de sus efectos en el resto de la señal, y esto es posible con el uso de una misma Transformada Wavelet Continúa.

En consideración, al evaluar diversos tipos de wavelet continuas se favoreció el uso de las wavelets analíticas particularmente de la wavelet Morse, ya que está en capacidad de mantener un buen balance general en los diversos objetivos de identificación y representación frecuencial de una señal previamente expuestos, sin sacrificar unos a expensas del otros, debido a su forma.

Sin embargo, para que la resolución frecuencial permita evidenciar singularidades en las señales de los precios de los activos financieros y sus efectos en la señal, es importante contar con suficiente historia previa a la manipulación, ya que los eventos transitorios se hacen más evidentes en la medida que el espectro de potencia tenga suficientes muestras que permitan describir de mejor manera el comportamiento frecuencial de los precios de un activo financiero.

Lo anterior es fundamental, pues al definir el régimen frecuencial considerado 'normal' para cada serie, las anomalías se hacen evidentes, pues se hacen aún más notorios los choques que causan las manipulaciones respecto a los regímenes de frecuencia previos a dicha manipulación.

Es importante tener en cuenta que, en las etapas de análisis y caracterización, fue necesario contar con información posterior a la manipulación, pues dicha

información facilita la comprensión de los cambios energéticos producidos por las prácticas manipulativas que fueron consideradas como parte de la presente investigación.

De manera particular los efectos de la discontinuidad en el régimen frecuencial, los picos evidenciados en la densidad espectral, y las variaciones en el resto del espectro de frecuencia, se hacen más notorias en escalas más globales y surgen justo después de que dan inicio los esquemas manipulativos.

La identificación de dichos artefactos es clave para la detección de las prácticas manipulativas, pues en su gran mayoría los defraudadores buscan que pasen desapercibidas, pues es el mismo mercado el encargado de corregir la dirección de los precios con base en la percepción de valor que se tiene sobre un activo.

Los efectos de una manipulación serán notorios, mientras el esquema aún esté activo, como también mientras el mercado no logre estabilizar los precios de acuerdo con la equivalencia del valor de un activo en libros o sus promedios históricos.

Es por esta misma razón, que en ausencia de manipulaciones los precios de los activos se mantienen 'relativamente' estables y los cambios son tenues, opuestos a los cambios bruscos que las prácticas manipulativas imprimen en los activos financieros.

En resumen, como en cualquier análisis de señales, fue necesario considerar el balance entre una mejor resolución de tiempo y una mejor resolución de frecuencia y por tanto se incluyó información posterior a la manipulación y se consideró un rango de tiempo amplio.

Sin embargo, como se busca caracterizar anomalías en las señales que obedezcan a prácticas manipulativas, el ajuste deberá favorecer la resolución de frecuencia, ya que precisamente son dichos cambios y regímenes alterados los que describen una manipulación de los precios de un activo financiero.

Al respecto es necesario precisar, que si bien, contar con la dimensión del tiempo es fundamental, no se requiere detallar con exactitud el momento en el que ocurre la manipulación, pues la información financiera de nivel 2 (es decir el detalle intradía de las operaciones realizadas durante la sesión de negociación) es la que permitirá caracterizar de forma plena el detalle en que el defraudador fraguó el esquema manipulativo.

Por otra parte, al respecto de la precisión del tiempo es necesario mencionar que el escalograma mediante el procesamiento de la transformada wavelet continua ofrece una muy buena resolución temporal gracias a su alta sensibilidad (capacidad de resolución temporal y de eventos transitorios). Además, en este caso al considerar la secuencia de muestras hace que se pueda tener un alto rango dinámico (capacidad de resolver diversos componentes de frecuencia) sin sacrificar la resolución temporal.

Por lo anterior, para esta herramienta es fundamental considerar aquello que haga mejor la resolución espectral y no la temporal. Especialmente cuando el filtrado espectral puede interferir con la capacidad de la transformada de distinguir en términos de la frecuencia dos componentes diferentes en la señal. (Oppenheim, Schafer & Buck, 1999)

La interferencia causada por el filtrado espectral puede darse de dos maneras: la primera el filtrado espectral de un componente más fuerte puede oscurecer la presencia de un componente más débil, y la segunda dos componentes con frecuencias similares y una misma densidad podrían no ser resueltos. (Oppenheim, Schafer & Buck, 1999)

Por tanto, es importante controlar este parámetro, ya que el filtrado espectral generado por el procesamiento de la señal puede introducir ruido indeseado sobre el filtrado espectral que naturalmente las manipulaciones imprimen sobre el espectro.

Previamente se mencionó como las manipulaciones se pudieron caracterizar como respuestas transitorias (o cambios de fase). Además, se mencionó como las respuestas transitorias generan cortes abruptos en el régimen exhibido por la señal.

Dichos cortes abruptos no se ajustan con el período natural de observación, causando “desbordamientos” que dejan marcas o filtrado sobre toda la base. (Harris, 1978)

Como pudo validarse, estas marcas persisten mientras el esquema de la manipulación esté vigente o mientras sus efectos no hayan sido absorbidos por el mercado.

El tiempo de duración de la señal o la cantidad de muestras que se consideran en el análisis es clave. De forma intuitiva se podría considerar como a mayor cantidad de muestras, mejor resolución.

Lo anterior, es cierto. Sin embargo, es importante notar que, si bien la cantidad de filtrado en una señal puede ser influenciado por el período de muestreo, este no es la causa del filtrado, ya que el filtrado que resulta del procesamiento de la señal se presenta al trabajar con señales finitas (Harris, 1978)

Al respecto, es común que se haga de manifiesta la pregunta ¿qué tiene entonces de continuo el análisis mediante wavelets continuas? Pues bien, la respuesta está en las capacidades multi-resolución de las wavelets.

A diferencia de las wavelets discretas, las wavelets continuas permiten que los parámetros de escala y de fase cambien continuamente, lo que lo hace menos invariante al filtrado como también que sea una transformada redundante. (Wang, Yang & He, 2022)

La redundancia garantiza que la transformada wavelet continua pueda proporcionar resultados de alta resolución, mucho más fáciles de interpretar que los obtenidos con la transformada wavelet discreta. (Wang, Yang & He, 2022) Ya que la transformada wavelet continua está en capacidad de tener la misma resolución que la señal original en cada banda de frecuencias. (Addison, 2018)

Lo anterior, podría indicar que existe entonces un límite que garantiza que la cantidad de muestras, los tamaños de los segmentos y el tamaño de las ventanas es suficiente para identificar las manipulaciones.

Sin embargo, es un hecho como algunos casos el comportamiento de los precios de un activo se retrasa, adelanta o coincide con el de otros activos o indicadores en diversos momentos del tiempo.

En parte, es por lo anterior que diferentes actores en los mercados financieros evalúan el comportamiento de los precios de los activos en diversas escalas. Términos como "corto plazo" y "largo plazo" son fundamentales para modelar las complejas relaciones entre las variables financieras.

Lo anterior es precisamente una razón fundamental para que el análisis mediante wavelets, particularmente mediante wavelets continuas sea útil.

Previamente se mencionó como el uso de wavelets continuas, ofrece en un mismo análisis una vista global (no detallada) de la señal (largo plazo) y una vista detallada de la señal (corto plazo) en la que se pueden descubrir patrones ocultos en la señal, que generalmente son de corta duración. (Sejdic, Djurovic & Stankovic, 2008)

El mosaico de diversas escalas que produce la transformada wavelet continua con la wavelet Morse, es útil porque está en capacidad de representar el contenido de oscilación lenta que ocurre y es evidente en escalas largas, como también eventos abruptos o transitorios que ocurren en corto tiempo y, por tanto, en algunos casos solo son evidentes en escalas cortas.

Además, está en capacidad de discriminarlos, ya que los representa en la manera adecuada: contenidos de oscilación lenta, en bajas frecuencias, y eventos transitorios en frecuencias altas.

Por lo anterior, se refuerza la elección de una wavelet continua analítica como la wavelet Morse, ya que es una herramienta que permite evidenciar de manera general los diversos efectos que diferentes tipos de manipulaciones a diversas escalas imprimen en la señal.

Es importante recalcar, que se pueden notar como algunos tipos de kernel describen de mejor manera unas prácticas que otras, lo mismo que ciertos parámetros favorecen el ajuste de los diversos balances de resolución de componentes en la señal, como se expuso previamente, para dar respuesta a los diversos objetivos del procesamiento de una señal.

Al respecto, dichos hallazgos son útiles para ajustar los parámetros en cada caso. Sin embargo, estos ajustes pueden ser muy particulares, mientras que los coeficientes que resultan de aplicar con CWT con la Wavelet Morse, por si solos y con parámetros que garanticen un buen balance general son más generalizables.

Los coeficientes de esta wavelet están en capacidad de mostrar los picos donde los componentes sinusoidales se activan y desactivan, como también de recoger la energía de tales eventos, además de localizar temporalmente los defectos en las señales.

Por tanto, una alternativa del ajuste del filtrado mediante los parámetros hallados es emplear técnicas de aprendizaje automático que estén en capacidad de detectar en los coeficientes los elementos antes expuestos. Evitando de esta manera la necesidad de ajustar parámetros particulares en cada señal para lograr resolver su contenido frecuencial particular.

Sin embargo, para que los coeficientes estén en capacidad de describir de mejor manera los fenómenos que resultan de las manipulaciones, aún es necesario ajustar un parámetro que no depende de la capacidad de resolución espectral de la herramienta como tal, sino de la señal misma.

Es necesario determinar para las series de precios de los activos financieros la cantidad de muestras, los tamaños de los segmentos de procesamiento y el tamaño de las ventanas de análisis que sean suficientes para identificar las manipulaciones en general.

Ya que de manera natural la información disponible de la señal limitará la capacidad de discriminar una práctica manipulativa a pesar de las capacidades de multi-resolución de la transformada wavelet continua.

No tener la información suficiente para validar cambios en un régimen, eventos transitorios y picos de energía, hará inviable la detección.

Previamente se mostró como a mayor cantidad de muestras mejor la resolución de las frecuencias, sin embargo, la mejoría se hizo evidente cuando se incluyó información posterior a la manipulación.

Por tanto, una forma de capturar la expectativa de un precio con pocas muestras, buena resolución y sin requerir muchas muestras posteriores es mediante el cálculo de la volatilidad de los precios.

Generalmente la volatilidad es calculada como la desviación estándar del retorno de los precios, por lo que se considera como la incertidumbre que rodea su movimiento potencial. (Kissell, 2020) Valores altos de volatilidad implican, mayor incertidumbre, mientras que valores bajos, implican una menor incertidumbre. (Nelson, 1991)

Anomalías en los precios afectan dichas expectativas. Movimientos súbitos en los precios per se afectarán el valor de la desviación media de los datos, incidiendo en el valor de la volatilidad. Por lo que variaciones abruptas en las expectativas producirán singularidades las cuales darán cuenta de manipulaciones en los precios de un activo.

Además, manipulando el valor de la volatilidad se influye en las expectativas que perciben los demás inversionistas, por lo que detectar alteraciones en los regímenes de las expectativas en los precios es una tarea igualmente necesaria (lo cual es especialmente relevante en el caso de las manipulaciones de largo plazo).

Por tanto, sin importar la duración del esquema manipulativo cambios en la expectativa a futuro de un precio está en capacidad de producir singularidades o alteraciones en los regímenes las cuales están asociadas a manipulaciones en los precios como ya previamente se ha visto.

Por lo anterior, conviene continuar con muestras diarias de volatilidad, es decir, serán consideradas las muestras de forma individual, ya que se garantiza que sin importar la escala temporal se detecten las singularidades con la mayor cantidad de información disponible.

Por su parte, es fundamental elegir una escala temporal adecuada, ya que la cantidad de datos históricos incide en las dinámicas que pueden ser detectadas.

No es lo mismo recoger las variaciones de toda la historia de una acción, a los últimos 5 días o las del último mes. Por tanto, precisar las ventanas de volatilidad supone nuevamente una búsqueda del rango que ofrezca la mejor resolución que permita caracterizar las singularidades.

Una ventaja adicional, de emplear la volatilidad es que, al ser una razón de cambio entre los precios, se eliminan los efectos de la magnitud y los datos se estandarizan, lo que facilita la generalización de los parámetros.

Generalmente para su cálculo se emplea las diferencias logarítmicas, ya que de esta manera se precisa que tan fuerte o que tan débil fue el movimiento entre los precios.

Por ejemplo, puede que parezca un cambio bajo, que el precio de un activo que antes costaba 10 unidades ahora cuesta 20 unidades, sin embargo, dicho cambio en términos porcentuales implica un aumento del doble en su valor de precio. Aun así, puede ser considerada una variación esperada de acuerdo con su historia.

Por lo anterior, cualquier choque que afecte dicho comportamiento esperado de forma abrupta podría estar asociado a una manipulación.

Es pertinente hacer énfasis en el hecho de mediante este cálculo se obtiene más información que al solo considerar los precios de los activos, ya que estos sólo describirán las medidas de magnitud de los precios, además lo harán en una escala limitada, al considerar la diferencia entre los precios la diferencia es de uno a uno,

mientras que al considerar la diferencia en volatilidad la comparación se hace contra un parámetro que comprime el comportamiento de los datos dentro de una ventana temporal dada.

Este efecto tiene una ventaja adicional y es que suaviza las fluctuaciones de plazos cortos, y resalta tendencias o ciclos de largo plazo. (Arce, 2005)

Por otra parte, diferencias en los precios no están en capacidad de incluir dinámicas particulares en los mercados, por ejemplo, la caída en los precios de un activo tiende a incrementar de manera subsecuente la volatilidad más de lo que lo haría un incremento en los precios de igual magnitud, o como períodos de alta volatilidad en ciertos índices pueden estar asociados a recesiones económicas (Hamilton & Susmel, 1994)

Lo anterior, como se mencionó previamente ocurre principalmente porque los actores económicos responden de manera diferente a malas noticias en comparación a las buenas. (Hatemi, 2014)

Se puede concluir que el cálculo de la volatilidad está en capacidad de calibrar la sensibilidad de los precios sin depender de la magnitud, como también revelar información adicional que la mera amplitud entre el valor de un precio y otro no ofrece.

Al respecto según lo propuesto por Nelson (1991), la volatilidad además está en capacidad de explicar la relación entre el nivel de riesgo de un activo y el retorno requerido, la asimetría entre los retornos positivos y negativos de los precios, la persistencia de choques y la presencia de anomalías.

En consideración a lo expuesto resulta clave emplear la volatilidad de los precios para la detección de prácticas manipulativas sobre activos financieros.

Sin embargo, aún es necesario definir la mejor ventana de volatilidad, pues la cantidad de muestras que se consideran puede comprometer la capacidad de resolución de ciertos eventos y en particular todos aquellos asociados a las manipulaciones. Como fue mencionado, nuevamente se debe considerar un balance entre capacidad de compresión y capacidad de resolución.

De acuerdo con Müller et al. (1997) la volatilidad no solo depende del período en el que se mide, sino también de la resolución temporal con la que se calcula. De acuerdo con sus hallazgos, la volatilidad medida con mayor resolución temporal contiene información que no es cubierta por la menor resolución temporal y viceversa. Confirmando lo que gráficamente se evidencia con las prácticas manipulativas al hacer uso de las técnicas de procesamiento de señales.

Sin embargo, según Vințe & Ausloos (2022) para intervalos de tiempo cortos a medianos y ventanas móviles estrechas, la estimación de volatilidad puede capturar perspectivas más esclarecedoras de un mercado. Según sus hallazgos, el riesgo de volatilidad muestra niveles más bajos en intervalos de tiempo más cortos: más del

90% más bajo para un período de 30 días y más del 75% más bajo para un período de 60 días.

Por lo anterior, se propone considerar intervalos de 5, 10, 20, 30 y 60 días de volatilidad, pues es necesario evaluar diversos rangos, ya que como se ha mencionado los actores de un mercado responden de manera diferente ante diversos sucesos condicionados por las visiones de corto y largo plazo.

Y es por lo anterior, que es necesario considerar diversos intervalos de volatilidad para elegir aquel que de manera más general permite representar los efectos de las manipulaciones en las series de datos financieras, particularmente cuando se ha demostrado que información asimétrica es una realidad de los mercados.

Previamente se explicó que una forma de evidenciar la asimetría en los mercados es considerar el hecho de que no hay un límite natural en el que el precio de una acción puede incrementar (teóricamente hablando), sin embargo, si existe un límite hasta el cual puede disminuir (cero). (Hatemi, 2014)

Lo anterior implica entonces la existencia de los impulsos asimétricos que estarían en capacidad de describir cambios fuertes en los regímenes de volatilidad y de esta manera detectar si hubo anomalías que puedan asociarse a las manipulaciones.

Al respecto, es importante caracterizar y determinar cambios abruptos en los regímenes de volatilidad cortos, pues en períodos más amplios las manipulaciones podrían pasar desapercibidas o no ser caracterizados en forma plena. Entre más amplias sean las ventanas, las diferencias y los cambios en la volatilidad de los datos comenzarán a ser considerados 'normales' por el efecto del suavizado de la media en el cálculo de la volatilidad.

Cuando hay un aumento en la volatilidad, el precio de un valor también se moverá más de lo normal. En este momento, existe la expectativa de que algo cambiará o ha cambiado. (Hamilton & Susmel, 1994)

Como ocurre cuando se presentan manipulaciones en una señal. Por otro lado, si la volatilidad está cayendo, significa que la incertidumbre está disminuyendo, por tanto, los efectos de la manipulación no serán tan evidentes, al punto que los retornos de los precios pueden llegar a ser como antes. (Hamilton & Susmel, 1994)

Considerando lo anterior, se hace aún más relevante controlar el filtrado espectral. Pues sin importar que los precios se estabilicen en niveles previos a la manipulación, las marcas en el espectro inducidas por los cambios abruptos serán visibles en el resto de la señal.

Por tanto, gana relevancia considerar la diferencia logarítmica como base para el cálculo de la volatilidad. Ya que, a diferencia de otros métodos como el cálculo de betas, las variaciones en las expectativas de los precios dependerán del propio activo y de la manipulación específica a la que puede estar sujeta el activo evaluado.

Al calcular la volatilidad propia del activo, se comprime la información intrínseca del comportamiento de su precio y no de valores de referencia externos como el precio de otros activos o índices económicos, necesarios para el cálculo de la volatilidad mediante técnicas como las betas.

Al respecto, emplear índices de referencia para el cálculo de la volatilidad con el fin de detectar prácticas manipulativas, puede presentar inconvenientes adicionales, ya que dichos índices de igual manera están sujetos a manipulación como en el caso de Navinder Sarao en 2010 (United States v. Navinder Singh Sarao, 2022) y de esta manera distorsionar los valores propios de la manipulación de los activos evaluados.

Además, para el presente caso de estudio considerar la volatilidad en los precios ofrece otras ventajas adicionales.

Primero, las wavelets están en capacidad de descomponer datos de series temporales en diferentes escalas y pueden revelar relaciones que no son obvias en los datos agregados (como es el caso de la volatilidad). (Gallegati & Semmler, 2014)

Además, a menudo es posible aprovechar las propiedades de los coeficientes calculados mediante wavelet para derivar estimadores basados en escalas para la varianza. (Gallegati & Semmler, 2014)

Y de esta manera profundizar en el entendimiento de los efectos causados por las manipulaciones.

Por último, considerar la volatilidad de las series de precios de los activos financieros, es una forma de estandarización sencilla de las señales que no se ve afectado por el valor nominal de cada activo que sea evaluado.

AJUSTE Y SELECCIÓN DE PARÁMETROS PARA LOS KERNEL

Con el fin de determinar el rango de volatilidad que permita de forma general establecer si una acción fue objeto o no de una manipulación, conviene partir de las validaciones previas hechas sobre las series de los precios.

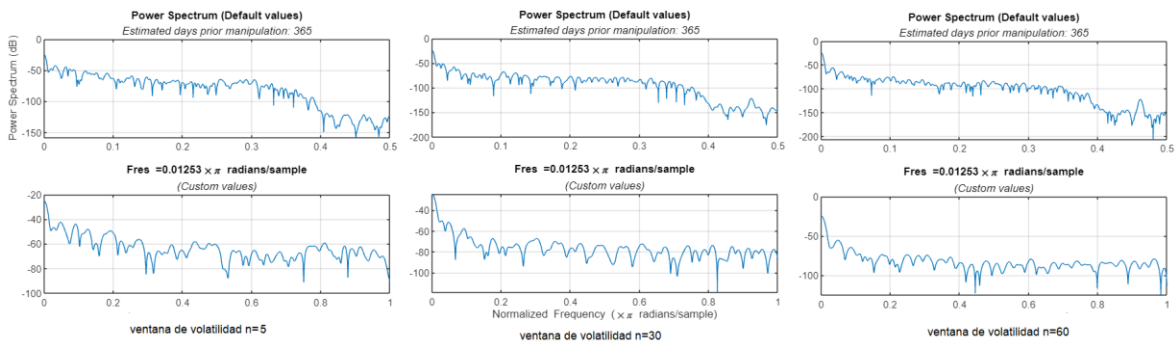
Para todas las series de volatilidad, se considera el mismo número de datos, en el que se tiene información de un año previo a la manipulación y al menos un mes posterior a dicho evento. En caso de que la manipulación se haya dado cerca a la emisión, el tiempo se ajusta para cumplir el tamaño de la matriz de tal manera que las quinientas muestras por serie sea completada. Generalmente se abarca un rango de dos años.

Las ventanas móviles consideradas para todos los intervalos, es diaria. Es decir, se consideran las muestras de forma individual, pues de esta manera se logra capturar en mejor medida los detalles de una manipulación, al comprimir toda la información disponible en un período, para luego validar como se altera ante innovaciones en los datos.

Conviene entonces considerar el análisis realizado con las series de precios sobre el Espectro de Potencia, procurando entender para el caso de la volatilidad como interactúan los diversos componentes de una señal en términos de su frecuencia, detectabilidad de tonos, la capacidad de resolución y en caso particular la aparición de patrones, ya que estos podrían surgir por el efecto de suavizado que la normalización con medias móviles induce en los datos.

Tomando una ventana kaiser con factor $\ell = 0.25$, y un tamaño de ventana igual a 1 (es decir analizando cada muestra) como en los precios, el resultado no difiere mucho, debido a la falta de representación del dominio del tiempo.

Figura 23. Espectro de potencia de una acción con manipulación tipo *pump and dump* en diversas ventanas de volatilidad



Sin embargo, como puede apreciarse en la Figura 23 son evidentes ciertas mejoras, especialmente si se comparan con los parámetros por defecto de la herramienta de procesamiento (ventana Kaiser con $\ell = 0$ y $\beta = 40$) presentados en la fila superior de la figura.

Es evidente como la presencia de ruido disminuye de forma drástica al elegir una ventana que favorece una resolución temporal diaria y con un balance entre resolución frecuencial y temporal, pues de esta manera se limitan los efectos del filtrado espectral el cual se concentra en bajas escalas (en este caso las menos a -100 db).

Por otra parte, es evidente como algunos picos y caídas en frecuencia se suavizan en la medida que aumenta la ventana de volatilidad, lo que hace aún más evidente la caída que representa el inicio de la manipulación y sus efectos en plazos más amplios.

Se evidencia que con ventanas mensuales se logra un buen balance, gracias a que se logran discriminar bien los segmentos y se evidencian con buena resolución los regímenes y los picos de frecuencia.

Lo anterior es relevante, pues a partir de 60 días los picos se consolidan y este efecto será más notorio en la medida que la escala crece, mientras que escalas más pequeñas como la de 5 días los regímenes son más sensibles a las variaciones, lo que implica que se pierde capacidad de separar los efectos en la señal a causa de la manipulación de los eventos del día a día.

Además, es relevante destacar como en el Espectro de Potencia se nota como al disminuir la resolución temporal de la ventana se restringe la aparición de artefactos en bajas frecuencias los cuales no están asociados al fenómeno y deben ser filtrados en las demás herramientas.

En general, al evaluar todas las acciones en el Espectro de Potencia, se logró estandarizar los resultados, definir una base para el filtrado de ciertas frecuencias (aquellas por debajo de -100 db) y la propuesta de un rango de volatilidad mensual.

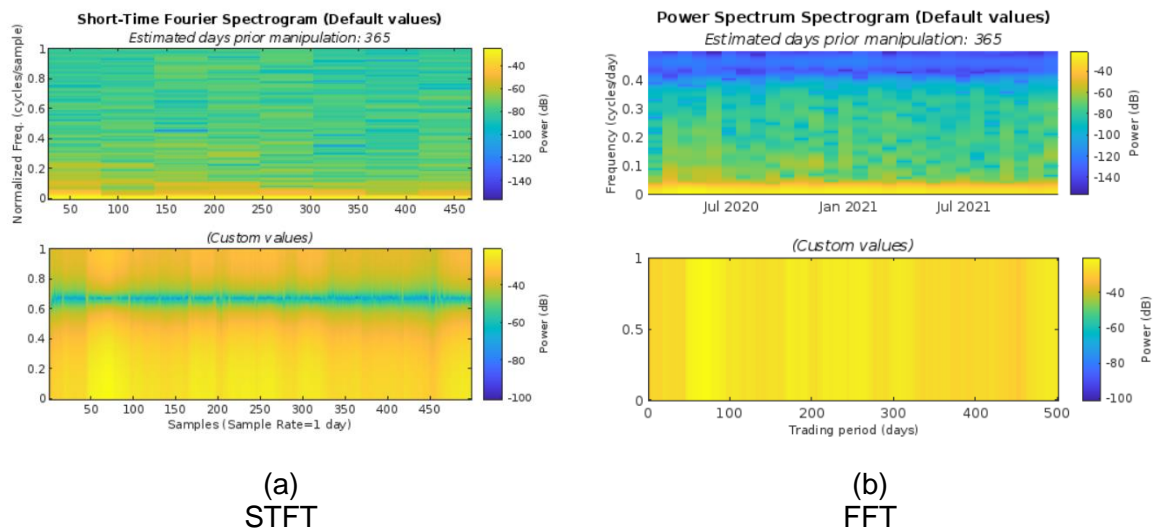
Los resultados obtenidos concuerdan con las evaluaciones previas, particularmente se evidencia el mismo comportamiento de la señal en la Figura 23.

Al analizar el espectro de potencia, no es posible identificar un patrón y en este caso particular su búsqueda es compleja y de existir estaría limitado a escalas muy específicas.

Mientras, que al evaluar las series de volatilidad en los Espectrogramas fue posible caracterizar las manipulaciones como cambios abruptos en la señal (a varios niveles de frecuencia), con registros de alta energía y densidad, la cual permite tener una idea de cómo se distribuyen las frecuencias en presencia de manipulaciones.

Para la evaluación de las series de volatilidad en los Espectrogramas, se tuvo en cuenta los resultados obtenidos con los precios. Con una ventana rectangular a intervalos de 3 muestras y un solapamiento de 1 muestra, se evidencia como para todas las acciones y sin importar el tipo de manipulación es más notoria una discontinuidad en la línea de ancho de banda de ruido al emplear la volatilidad como puede verse en el ejemplo de la Figura 24 (a).

Figura 24. Espectrograma basado en STFT y FFT de una acción con manipulación tipo *pump and dump* con volatilidad de 30 días



Sin embargo, visualmente no fue el caso para el espectrograma mediante la Transformada Rápida de Fourier (FFT), pues en este caso por los efectos de suavizado en la señal hace que los regímenes sean más estables y la presencia de otros componentes de frecuencia sean más difusos como puede apreciarse en la Figura 24 (b).

Al respecto, es necesario precisar que, si bien puede filtrarse aún más la energía para hacer más evidente ciertos picos, de igual manera se pierde aún más capacidad de representación del espectro de la señal, ya que otros efectos que son más notorios en escalas más balanceadas (ventanas con 128 muestras y 0 solapamiento) o en el espectrograma presentado por la STFT dejarían de estar presentes en el análisis.

Lo anterior, implica que al disminuir aún más el umbral de poder mínimo se perderá información y la posibilidad de detectar de mejor forma los cambios abruptos en los regímenes de frecuencia. En contrapeso mantener el umbral alto, dificulta la resolución de los picos de alta energía.

Modificar los parámetros para mejorar la capacidad de resolución frecuencial, afectará enormemente la capacidad de resolución temporal.

Adicionalmente, en la medida que aumenta la ventana de volatilidad los efectos de suavizado son aún más notorios, sin embargo, la ganancia en resolución es más evidente al emplear una ventana más amplia de tiempo posterior a la manipulación, en consideración la ventana mensual ofrece de nuevo un mejor balance, ya que logra una buena representación con períodos más cortos.

Es importante destacar que mediante el espectrograma es viable detectar la existencia de cambios abruptos en los estimados de los regímenes de volatilidad y la persistencia de regímenes implícitos en ella, y como estos pueden permanecer durante largos períodos (Hamilton & Susmel, 1994).

Pero la resolución frecuencial no es la mejor, y en caso de mejorarse podría comprometerse la resolución temporal como previamente se ha mencionado.

En general los espectrogramas permiten identificar como se distribuye la energía, caracterizando las manipulaciones como eventos de alta densidad y singularidades que afectan los regímenes de frecuencia. Lo cual no solo es útil para aportar en la caracterización de la señal en los escalogramas, sino también se puede considerar para emplear estos mapas para aprendizaje de imágenes como hicieron Gomez et al. (2017) y Ariza-Jiménez, Villa, & Quintero-Montoya (2019).

Sin embargo, el escalograma se sobrepone a las dificultades antes expuestas del espectrograma, ya que sin sacrificar resolución temporal es posible obtener información en diversas escalas de manera más detallada.

Al igual que en el caso de los precios, fue posible registrar los eventos de alta frecuencia asociados a los cambios abruptos en la señal. También fue posible detectar las oscilaciones de menor frecuencia producidas por dichos cambios abruptos con una resolución mayor para ambos casos.

Es importante recordar, que lo anterior se logra principalmente debido a que los efectos de filtrado espectral producidos por ventanas pequeñas empleadas en el Espectrograma, se disipan en el Escalograma mediante el mosaico de escalas que produce la CWT al emplear diversas versiones de la función wavelet. (Daubechies, 1990)

El análisis Wavelet permite el uso de intervalos de tiempo prolongados en los que se tendrá información de baja frecuencia más precisa y regiones más cortas en las que se tendrá también información de alta frecuencia. (Daubechies, 1990)

Mientras que en el caso de la STFT o la FFT el tamaño de la ventana debe ser constante, lo que afecta la capacidad de resolución de los diversos artefactos que hacen parte de una señal, ya que ciertos efectos podrían quedar por fuera de la ventana o verse opacados por el filtrado espectral, como ya se ha mencionado.

Es importante, tener en consideración que la CWT y el escalograma permiten representar la información disponible en la señal que caracterizan de mejor manera una manipulación.

Por su parte, el resultado obtenido mediante la DWT es similar al resultado que se obtiene mediante el espectrograma donde la detección de cambios abruptos en la señal se hace compleja y no es caracterizado el comportamiento oscilatorio que producen las singularidades en una señal.

Al respecto, es necesario precisar como los eventos transitorios en los datos de series temporales suelen ser uno de los eventos más informativos. (Smith, 2007)

Y el escalograma con la CWT está en capacidad de presentar la mayor cantidad de información, ya que no solo los detecta, sino que también representa sus efectos en la señal, describe los cambios en magnitud en diversas escalas y los ubica muy bien en el tiempo.

Esto es relevante cuando, al considerarse la volatilidad los eventos transitorios en la frecuencia son aún más evidentes y abruptos. Por tanto, conviene considerar los parámetros de los precios es adecuado.

Además, como previamente se mencionó como las wavelets Morse están en capacidad de lograr un buen análisis de señales moduladas (aleatorias) y discontinuidades localizadas (Olhede & Walden, 2002) debido a la dilación que este tipo de wavelets logran a partir de la definición de un factor racional s y a que las frecuencias negativas se hagan cero.

Por la anterior, el segmento procesado para todas las frecuencias permanece igual a t en todas las escalas, lo que a su vez permite que se mantenga una densidad de energía de tiempo local. (Mallat, 2009) Lo que restringe la aparición de filtrado espectral a causa del procesamiento de la señal, por lo que el filtrado que se evidencia será causado precisamente por las singularidades.

Un valor de gamma (γ) igual a 3 y un producto tiempo-ancho de banda igual a 60, empleando 10 voces por octava asegura un nivel de simetría adecuado, ya que dichos valores garantizan la simetría sobre la frecuencia máxima detectada, además de que facilita que el valor mínimo y máximo en cada escala sean determinados sin restricciones de rangos.

Las unidades de medida no cambian tanto la frecuencia como el cono de influencia se describen en términos de ciclos por muestra, teniendo en consideración que las series de volatilidad no están descritas en escala temporal, sino por cantidad de muestras, que para el caso particular son 1000 por acción manipulada, si se consideran los datos del activo de referencia.

Las validaciones efectuadas en los kernel, muestra como la volatilidad en diversas ventanas de tiempo, no solo permite identificar los cambios abruptos tanto en frecuencia como en escala, mejorando la resolución del espectro de potencia en

términos de los componentes asociados a las anomalías respecto a los precios, sino que respeta la localización temporal de los mismos.

En los escalogramas los picos de energía que producen los eventos manipulativos en una señal pueden ser asociados con el momento (muestra) que lo origina, por lo que haría viable identificar las operaciones que produjeron el evento manipulativo.

Sin embargo, a pesar de la buena resolución del escalograma, para ventanas de volatilidad pequeñas se requiere una mayor cantidad de muestras para lograr una buena resolución de las manipulaciones que permita diferenciarlas del efecto breve y corto que imprime los cambios aleatorios generales de cualquier señal financiera.

Para períodos más largos, dicho efecto se atenúa, lo que permite mejorar la resolución de singularidades empleando pocos datos, tanto en el tamaño de la ventana de volatilidad como en cantidad de muestras necesarias.

Por lo anterior, se propone emplear una volatilidad diaria con ventana móvil de 30 días, ya que esta ventana logra una buena resolución frecuencial y temporal de las singularidades causadas por las manipulaciones tanto de corto, como de largo plazo en diferentes escalas sin sacrificar información.

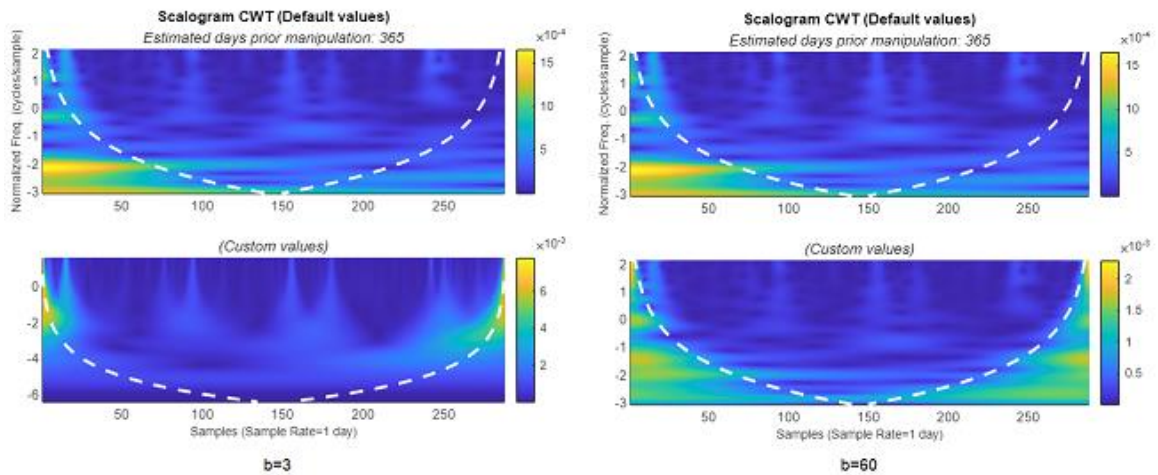
Esto ocurre en gran medida, porque la ventana de 30 días logra atenuar los efectos causados por los cambios aleatorios que se dan en las dinámicas normales de todos los mercados.

Por otra parte, teniendo en cuenta que sí bien el fenómeno admite una cantidad de muestras inferior a los 30 días previos a la manipulación, serán empleados 365 días, ya que ofrecen mejor resolución de los eventos asociados a las manipulaciones en una señal, lo cual es útil para el entrenamiento posterior de los algoritmos de aprendizaje.

Hacer uso de los escalogramas y de la volatilidad ofrece ventajas adicionales, ya que se logra capturar la singularidad que marca el inicio de la manipulación, como también sus efectos en el resto del espectro a diversas escalas con una gran resolución frecuencial, sin afectar la resolución temporal.

Sin embargo, para que lo anterior fuera posible es necesario respetar los valores previamente descritos para la wavelet, ya que entre mayor el ancho de banda mejor la capacidad de resolución de los efectos de las singularidades en el resto del espectro como puede apreciarse en la Figura 25.

Figura 25. Escalograma basado en CWT de una acción con manipulación *tipo wash trade* con diversos valores de Beta



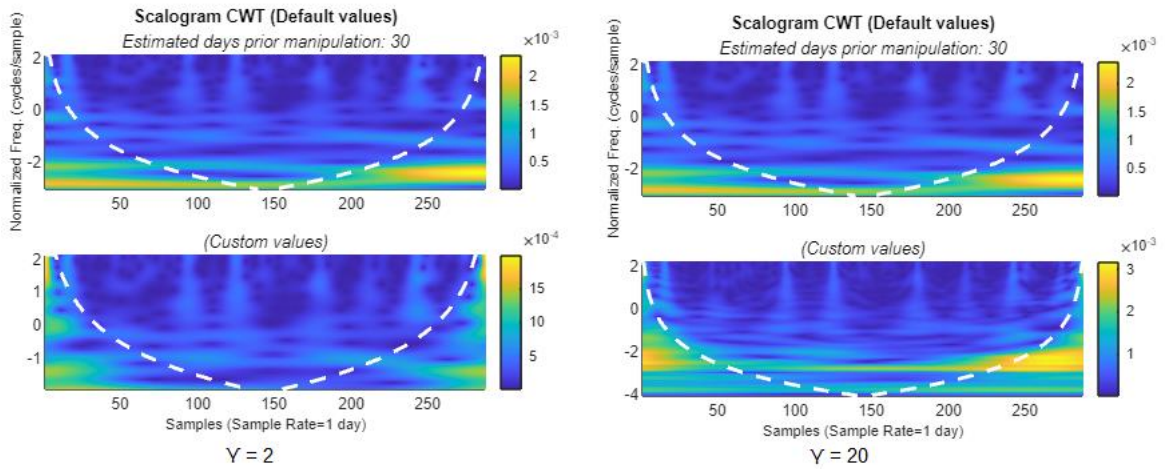
Esto es posible, ya que la wavelet morse con un valor de gamma (γ) igual a 3, logra un alto grado de simetría tanto en el dominio del tiempo como en el dominio de la frecuencia lo que conduce a un buen desempeño para el análisis de señales oscilatorias. (Lilly & Olhede, 2008)

Es precisamente la simetría lo que permite que se mantenga una densidad de energía en tiempo local, gracias a que un buen factor de simetría permite controlar mediante los factores de escala dados por las octavas, la invarianza del cambio de fase como puede apreciarse en la Figura 26.

Mantener una buena simetría en la wavelet en el caso de señales aleatorias, es fundamental pues, no solo de esta manera se gana precisión temporal (se logra mayor simetría cuando el tiempo de decaimiento de la wavelet es mayor), sino que al mantener simetría en la señal cuando las frecuencias negativas se hacen cero, se restringe aún más el filtrado espectral lo que mejora las capacidades de resolución frecuencial y temporal del efecto de las manipulaciones sobre las señales financieras.

La existencia de una correspondencia única entre la escala y la frecuencia requiere simetría sobre la frecuencia máxima, medida por la asimetría en el dominio de la frecuencia, y también igualdad entre la media y la moda del módulo cuadrático de la wavelet en el dominio de la frecuencia.

Figura 26. Escalograma basado en CWT de una acción con manipulación *tipo wash trade* con diversos valores Gamma



Por lo anterior, con una propiedad de orden inferior mantenida fija, como la duración de la wavelet en proporción a su período, es decir una wavelet con alto grado de simetría tanto en el dominio del tiempo como en el dominio de la frecuencia, es posible que se tenga un buen desempeño en el análisis de señales oscilatorias. (Lilly & Olhede, 2008)

Es claro como valores de simetría grandes hacen que la wavelet oscile más en el tiempo y se estreche en frecuencia, como previamente fue mencionado.

Lo anterior, es fundamental pues a medida que disminuye la frecuencia, aumenta el ancho de los coeficientes de la CWT en el tiempo que son distintos de cero y están centrados en los impulsos que se detecten en la señal.

Por el contrario, a medida que aumenta la frecuencia, el ancho de los coeficientes de la CWT que son distintos de cero disminuye y se centra cada vez más en los impulsos detectados. (Lilly & Olhede, 2008)

Por lo que se garantiza que, sin importar la escala, los cambios abruptos y los efectos sobre el resto de la señal estarán bien ubicados en el tiempo y permitirá conocer los efectos sobre el resto de la señal.

Conviene pues considerar wavelets con alto nivel de simetría y un ancho de banda amplio ya que no solo se logra la detección de impulsos en una señal, sino que logra una buena resolución de estos, ya que entre mayor el ancho de banda mejor la capacidad de resolución de los efectos de las singularidades

Los impulsos o singularidades en una señal financiera generalmente se asocian con las singularidades causadas por las manipulaciones, pues corresponden a cambios

en los regímenes de frecuencia. Entre más fuerte el impulso, más probable que se deba a una manipulación.

El efecto del impulso persiste más tiempo con wavelets más largas. En otras palabras, cuanto más larga sea la wavelet, mayor será la duración de la influencia del impulso en el procesamiento. (Lilly & Olhede, 2008)

Estirar o encoger la wavelet da como resultado que la wavelet "vea" más o menos señal y es precisamente a lo que se conoce como el cono de influencia de la wavelet. (Lilly & Olhede, 2008)

Para una wavelet centrada en un cierto punto en el tiempo, las regiones en cada extremo indican el "cono de influencia" de toda la señal, donde los efectos de borde se vuelven importantes. (Torrence & Compo, 1998)

De acuerdo con Mallat (2009) una transformada wavelet puede centrarse en estructuras de señal localizadas con un procedimiento de zoom que reduce progresivamente el parámetro de escala, lo que facilita que singularidades o estructuras irregulares pueden ser detectadas en diversos niveles, según estos elementos afecten o se evidencien en un segmento de frecuencia particular.

La amplitud de la transformada wavelet a través de las escalas está relacionada con la regularidad de la señal local, por tanto, las singularidades y los límites de una señal se detectan a partir de los máximos locales de la transformada wavelet en múltiples escalas. Estos máximos definen un soporte espacial-escala geométrico a partir del cual se recuperan aproximaciones de señal. Las singularidades no aisladas aparecen en señales altamente irregulares como multifractales. (Mallat, 2009)

La transformada wavelet aprovecha las autosimilitudes multifractales para calcular la distribución de sus singularidades. Este espectro de singularidad caracteriza las propiedades multifractales, por lo que cualquier elemento que esté por fuera de la distribución debe emplearse con cautela, ya que tal vez no hace parte de la señal y por tanto a este fenómeno se conoce como "efecto de borde" (Mallat, 2009)

Se debe tener en cuenta que los efectos de borde son producto del procesamiento de señales finitas y no cíclicas, por lo que en la medida que la wavelet procese información por fuera de un intervalo definido se creará un borde que afecta todo el espectro. (Arts & van den Broek, 2022)

Como previamente se mencionó, este efecto se vuelve más evidente con wavelets más grandes (es decir, en frecuencias más bajas) y de ahí su forma cónica. (Arts & van den Broek, 2022)

No existe una regla matemáticamente precisa para determinar la extensión del cono de influencia en cada escala. (Torrence & Compo, 1998) Sin embargo, existen varias estrategias para lidiar con estos efectos.

Para las wavelets Morse, Lilly (2017) utiliza el concepto de "huella wavelet", que es el intervalo de tiempo que abarca aproximadamente el 95% de la energía de la wavelet. Lilly delinea el cono de influencia sumando la mitad de la huella de wavelet al comienzo del intervalo de observación y restando la mitad de la huella del final del intervalo en cada escala.

Aquí está claro que el cono de influencia para los límites extremos del intervalo de observación se extiende dentro del intervalo en un grado que depende de la escala de la wavelet. Por lo tanto, los coeficientes de la wavelet dentro del intervalo de observación pueden verse afectados por los datos que la wavelet procesa en los límites de la señal. (Lilly, 2017)

Este último es un caso particular, pues cuando las wavelets procesan información al inicio o al final de la señal, la frecuencia pierde sentido ya que no es posible determinar cómo procedería una señal más allá de estos límites. (Arts & van den Broek, 2022)

De igual forma al caso general, existen diversos métodos para compensar el hecho de que los coeficientes de wavelet cerca del principio y el final de los datos se vean afectados por las wavelets que se extienden fuera del límite. Generalmente se ofrece la opción de tratar los límites reflejando la señal de forma simétrica o extendiéndola periódicamente. (Lilly, 2017)

Conviene para este caso, emplear el método definido por Lilly (2017) para la wavelet Morse y extenderla periódicamente.

Sin embargo, los efectos de borde al final no afectan la detección de las manipulaciones pues extender o no de forma simétrica la señal para el fenómeno de las manipulaciones describirá en escalas diferentes los mismos fenómenos.

La energía de la manipulación presentada se encuentra en el caso de la extensión simétrica distribuida en escalas más largas, mientras que si se extiende periódicamente se representan los picos de manera más natural y lo demás como sus efectos en bajas escalas.

Lo anterior pudo comprobarse al emplear información posterior a la manipulación ya que las singularidades se seguían identificando en todas las escalas con una resolución frecuencial muy similar, respetando siempre el tiempo.

Es necesario precisar que incluir más voces por octava no hace que la resolución mejore, por el contrario, el tiempo de procesamiento aumenta de forma significativa. Por tanto, el valor de 10 es un valor suficiente para lograr una buena resolución de los fenómenos asociados a las manipulaciones.

RECONOCIMIENTO DE PRÁCTICAS MANIPULATIVAS USANDO MÁQUINAS DE APRENDIZAJE

Es claro como la volatilidad aporta una mejor resolución de las manipulaciones mediante el uso de escalogramas generados a partir de una CWT definida mediante una wavelet Morse.

Es posible localizar las frecuencias en el tiempo con un alto nivel de precisión, particularmente aquellas asociadas a las singularidades en la señal, las cuales se identifican de forma clara.

Además, es posible representar con un buen nivel de detalle en un mismo gráfico los efectos derivados de las singularidades en diferentes escalas:

- Escala pequeña $a \Rightarrow$ Wavelet comprimida \Rightarrow Detalles que cambian rápidamente \Rightarrow Alta frecuencia $\omega \Rightarrow$ Eventos transitorios \Rightarrow Cambios abruptos en los precios.
- Escala larga $a \Rightarrow$ Wavelet estirada \Rightarrow Características generales que cambian lentamente \Rightarrow Baja frecuencia $\omega \Rightarrow$ Eventos oscilatorios \Rightarrow Respuesta ante modificaciones en el régimen de precios.

Lo anterior es fundamental, pues como previamente se ha mencionado las singularidades causadas por las manipulaciones generan efectos adicionales en el resto del espectro a diversas escalas sin importar la duración y el tipo de esquema manipulativo.

La representación gráfica que ofrece el escalograma con los parámetros elegidos, facilita la detección de los diversos efectos que las manipulaciones imprimen en una señal financiera con un balance de pérdida de información mínimo para ambos casos a diferencia de los resultados obtenidos mediante otros tipos de kernels y representaciones, como previamente se ha demostrado.

Sin embargo, para que esta validación sea escalable conviene considerar técnicas de aprendizaje automático para facilitar los procesos de detección de prácticas manipulativas.

Al respecto, Addison (2018) comenta como la CWT puede ser una técnica particularmente valiosa para iniciar e impulsar el desarrollo de soluciones algorítmicas al proporcionar una visualización rápida de los componentes subyacentes de una señal.

Independientemente de si forma parte del algoritmo final, Addison (2018) agrega que la CWT puede ayudar a comprender las características de la señal, proporcionar información sobre la viabilidad de la tarea subyacente y facilitar el desarrollo de estrategias de procesamiento de señales para atacar el problema a resolver.

En sus palabras, Addison considera valido tratar la CWT como una 'herramienta de preprocesamiento visual', donde son diversas las posibilidades que pueden ser

evaluadas para su implementación, partiendo desde la aplicación de filtros, considerando el uso de doble transformada, hasta llegar al aprendizaje de máquina; pasando por la búsqueda de patrones, el aprendizaje de imágenes y llegando a los algoritmos de clasificación.

Cada posibilidad ofrece sus ventajas y desventajas, a priori es válido mencionar que cada una al igual que los kernel, ofrece visiones particulares del fenómeno o incluso complementarias, al considerar los resultados obtenidos mediante distintas técnicas como una base para la selección de los mejores parámetros en otras. Este es el caso, por ejemplo, de la aplicación de filtros.

Para considerar la aplicación de filtros como estrategia de detección de prácticas manipulativas se requiere que estos sean definidos de acuerdo con la práctica manipulativa y a la duración de ésta. No existe un filtro que sea general, ya que para todos los componentes presentes en la señal la frecuencia está cambiando continuamente (Boashash, 2015)

En las representaciones de cada kernel pudo apreciarse como a pesar de que las prácticas manipulativas coinciden en los efectos causados en la señal, estos se diferencian en intensidad, efecto en el largo plazo a causa de la manipulación y la densidad espectral.

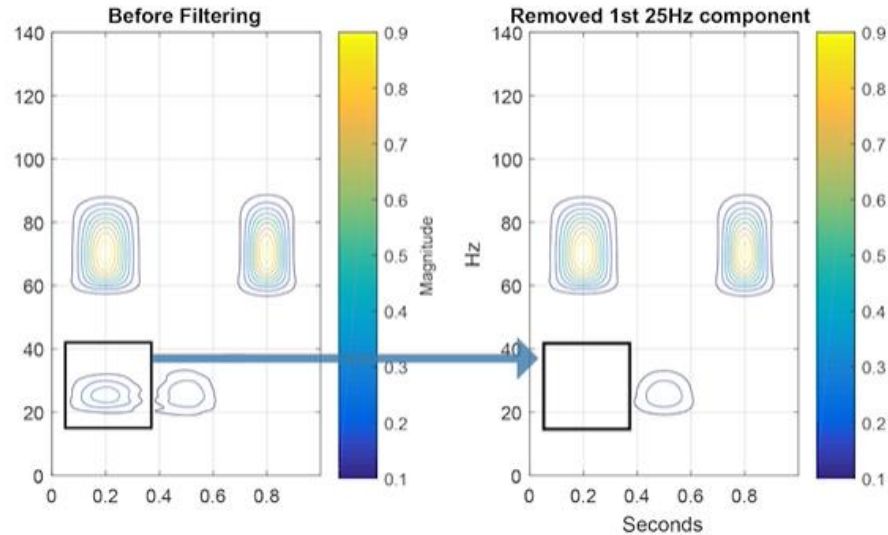
Idealmente, se desea una respuesta de tiempo prolongado con soporte de frecuencia estrecha en frecuencias bajas y un soporte de tiempo corto con soporte de frecuencia más amplio en frecuencias altas. Sin embargo, como previamente se ha visto, no hay un solo filtro sobre toda la señal que permita fácilmente separar la información relacionada con las manipulaciones de la que no, por lo que se requieren filtros específicos para cada segmento del espectro.

No obstante, si pueden emplearse filtros que mejoren la densidad espectral al restringir el ruido aditivo, cuyas características no cambian con el tiempo (Boashash, 2015) y en cierta medida mejorar también el ruido multiplicativo al punto que no entorpezca las tareas de detección de prácticas manipulativas, como hicieron Paniagua-Castrillon & Quintero-Montoya (2018), donde artefactos de baja frecuencia generaban ruido que escondían características relevantes para el procesamiento de la señal, los cuales pudieron ser disminuidos gracias a la aplicación de diversas técnicas de filtrado sobre la señal y otras estrategias de post-procesamiento.

Restringir la aparición de ciertos elementos en la señal, puede mejorar la resolución del escalograma procesado como se representa en la Figura 27.

Al respecto, otras técnicas basadas en wavelets han sido consideradas y podría resultar más beneficioso, como la correlación cruzada de wavelet y la bicoherencia de wavelet (Wang, Yang & He, 2022) que tal vez permitan obtener una nueva “visión” útil del contenido de la señal (Addison, 2018).

Figura 27. Representación del filtrado de eventos reflectivos en una señal



Fuente:
The MathWorks, Inc., n.d.

También hay disponible una amplia gama de métodos innovadores para manipular la información de la señal contenida en la representación CWT, incluidos métodos de módulo máximo para identificar singularidades, seguimiento de crestas para el análisis modal, reasignación y sincronización para resolver aún más los componentes de la señal en el dominio wavelet, ejecutando el arquetipo wavelet (RWA) para buscar componentes de señal repetitivos dentro de señales ruidosas y SWFD para permitir el análisis los componentes no estacionarios en señales previamente clasificadas como no estacionarias. (Addison, 2018)

Incluso se han propuesto técnicas de procesamiento de imagen teniendo en cuenta las particularidades de las estructuras de los espectros de poder, ya que al considerar las representaciones del procesamiento mediante kernels de una señal se pueden extraer y reconocer varias características de las señales, como reconocen Sierra-Sossa, Bastidas & Quintero-Montoya (2016).

En su caso particular, Sierra-Sossa, Bastidas & Quintero-Montoya (2016) emplearon métodos de correlación y filtrado multi-escala para sintetizar variaciones en el color, la textura y la forma. Además, los operadores de morfología matemática como la erosión y la dilatación fueron empleados para reducir el ruido del espectrograma, lo que, combinado con el filtrado no lineal bidimensional, permitió el reconocimiento de los componentes de señales que estaban tratando de ubicar.

Es importante tener en consideración al momento de emplear aprendizaje de imágenes, las cantidades abrumadoras de datos que las señales transportan, en los que la información relevante suele ser más difícil de encontrar que una aguja en un pajar. (Mallat, 2009)

Conviene entonces considerar técnicas cuyo procesamiento sea más rápido y sencillo, como en una representación escasa donde pocos coeficientes revelan la información que se busca. (Mallat, 2009)

Sin embargo, como previamente se ha mencionado, si bien representaciones escasas como los espectrogramas dan pistas útiles, no muestran la naturaleza transitoria de la señal, es decir, la localización precisa de la señal en el tiempo. (Boashash, 2015)

Al respecto, Tripura & Chakraborty (2022) precisan como una de las principales deficiencias del espectrograma es que las funciones básicas en las FFT generalmente están localizadas en frecuencia sin resolución espacial. Por lo tanto, las FFT no son adecuadas para estudiar el comportamiento espacial de ninguna señal o imagen y, como consecuencia, el rendimiento de los espectrogramas se ve obstaculizado en condiciones de contorno complejas como las que se pudieron apreciar en los espectrogramas de las señales financieras con manipulación.

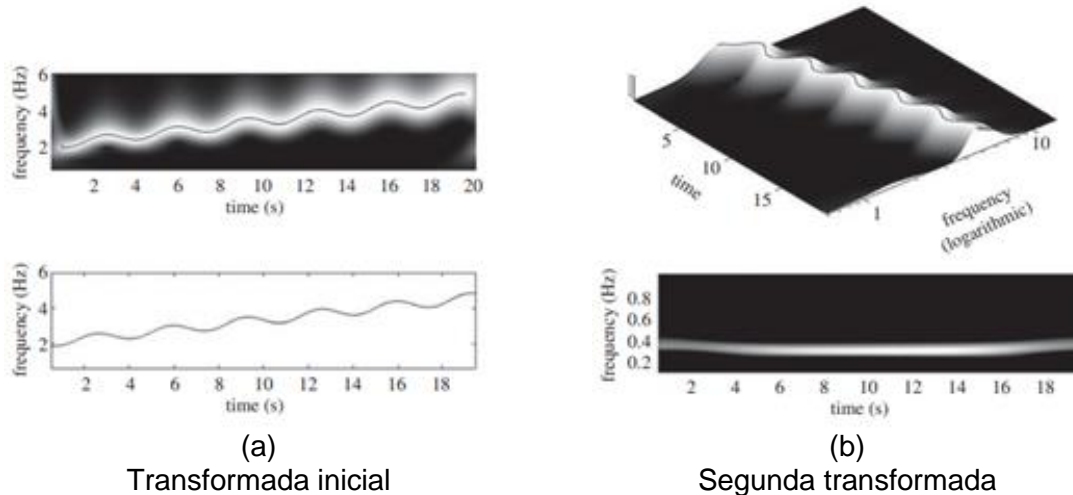
Este no fue el caso de Sierra-Sossa, Bastidas & Quintero-Montoya (2016) donde, aunque con ligeros cambios en las distribuciones de densidad espectral espacial (inducidos por el muestreo de voz y los parámetros de la ventana STFT) su señal analizada retrata una estructura particular reconocible, lo que fue útil para la identificación de emociones a partir de la voz.

Como alternativa Tripura & Chakraborty (2022), proponen el uso de las wavelets, ya que están localizadas tanto en el espacio como en la frecuencia. Dado que las wavelets tienen información espacial, pueden manejar mejor la señal con discontinuidad y/o picos y, por lo tanto, pueden aprender mejor los patrones en las imágenes en comparación con las FFT.

Sin embargo, como los diferentes componentes de la señal observada en el dominio de tiempo-frecuencia varían de forma significativa entre tipos de prácticas y duración de los esquemas, lo que indica que las relaciones tiempo-frecuencia para cada componente deben conocerse para identificar la señal con precisión. (Tripura & Chakraborty, 2022)

Sin embargo, considerar el aprendizaje mediante imágenes para este caso, puede ser útil y tiene otras ventajas. De acuerdo con Addison (2018), la modulación de frecuencia representada en Figura 28 (a) puede analizarse con mayor profundidad realizando una segunda transformada wavelet de esta nueva señal, en los que la frecuencia de modulación de alrededor de 0.3 Hz se manifiesta como una nueva banda a esta frecuencia en esta transformación posterior, como se aprecia en la Figura 28 (b).

Figura 28. Representación de las diversas vistas de un escalograma



Fuente:
Addison, 2018 (adapt.)

Según Addison (2018) este método de desacoplamiento de características de wavelets secundarias (SWFD por sus siglas en inglés) es útil para extraer información de modulación de señales no estacionarias. Además, vale la pena señalar que la cresta no solo contiene información de frecuencia instantánea, sino que también la altura de la cresta desde el plano de transformación contiene información sobre la intensidad del componente de la señal. Por lo tanto, también se puede analizar las modulaciones de amplitud proyectando la cresta horizontalmente (no se muestra en la Figura 28).

Es de tener en consideración, que el caso que se muestra en la Figura 28 es relativamente simple y no contiene ruido en aras de representar las posibilidades de que existen al emplear kernels para procesar señales y sus diferentes representaciones de acuerdo con el tipo de señal analizada, sin embargo, debido a la diferencia en la intensidad de los componentes entre manipulaciones, como a las diversas amplitudes de acuerdo a la magnitud de las singularidades y a las diversas ventanas de tiempo en las que las manipulaciones se pueden dar; implica que se requiera una gran cantidad de eventos de referencia para lograr un resultado generalizable.

Adicionalmente, es fundamental tener en consideración que el filtrado eficaz mediante wavelets de señales reales no es el mejor sin haber determinado previamente su forma. De acuerdo con Tarenko & Rizun (2022), la forma de una señal real está relacionada con su espectro wavelet.

Para el análisis de forma, a menudo se utiliza los escalogramas mediante wavelets continuas, sin embargo, su desventaja radica en la complejidad de analizar una imagen de color borrosa, pues, en algunos casos no hay estructuras uniformes que permitan identificar de forma clara las variaciones de frecuencia que a menudo contienen información crítica sobre la señal y el proceso estudiado en las aplicaciones. Sin embargo, vale la pena aclarar que su magnitud si es bien representada. (Boashash, 2015)

Adicionalmente, como ya se ha visto, una señal real con una gran cantidad de ruido aditivo como las señales financieras, distorsiona fuertemente el escalograma en comparación con lo que se obtendría mediante una señal pura, lo que dificulta el tratamiento de los escalogramas mediante técnicas de aprendizaje de imágenes.

Para solucionar este problema Tarenko & Rizun (2022) proponen la realización de análisis comparativos de los coeficientes con los que se generan los escalogramas y emplear para ello diversas funciones que permitan describir de mejor manera el fenómeno y concluyen que los métodos de aprendizaje mediante imágenes son particularmente útiles para dividir la señal del ruido y para el caso de señales complejas proponen el análisis sobre los coeficientes que resultan a partir de la aplicación de las transformadas.

Idealmente, se desea una respuesta de tiempo prolongado con soporte de frecuencia estrecha en frecuencias bajas y un soporte de tiempo corto con soporte de frecuencia más amplio en frecuencias altas. Los coeficientes de la transformada wavelet proporcionan exactamente eso y, por lo tanto, considerar técnicas de aprendizaje sobre dichos valores resulta especialmente útil.

Considerando lo que señala Quintero-Montoya (2018) sobre el uso de estrategias de procesamiento de señales y aprendizaje de máquina, un amplio rango de métodos innovadores ha sido desarrollado para manipular la información contenida en una señal y revelada por la representación de la CWT, sin embargo en relación con el poder de discernimiento de los seres humanos en materia de clasificación, los desarrollos de algoritmos con kernels, lineales y no lineales, han permitido mezclas interesantes de aproximaciones estadísticas y probabilísticas y la consolidación del aprendizaje estadístico que facilita a las máquinas diferenciar clases de datos.

Por lo anterior, conviene emplear técnicas de clasificación, particularmente técnicas de agrupamiento, ya que están en capacidad de implementar estrategias que permite diferenciar aquellos valores asociados a manipulaciones de los que no, y de esta manera lograr una detección escalable, ya que los coeficientes wavelet son traducidos en cada escala en la misma proporción y sin modificaciones adicionales (Addison, 2018). Así, los picos y los datos anómalos pueden ser diferenciados de los datos asociados al comportamiento esperado de los precios de los activos.

De esta manera la detección de las prácticas manipulativas es generalizable, ya que la clasificación parte de la base del comportamiento que describen los coeficientes

en el hiperplano que resulta al procesar la señal mediante la CWT. Adicionalmente, emplear los coeficientes facilita la ubicación temporal de las manipulaciones, pues al final los coeficientes son una función de escala o de frecuencia y tiempo. (Mallat, 2009)

Al emplear técnicas de agrupamiento, es posible separar mejor los datos y profundizar en la comprensión de los espacios generados por el procesamiento de la señal mediante la CWT, especialmente cuando no es fácilmente detectable una estructura estable o patrón que sea producido por las manipulaciones.

El factor de la ventana de tiempo juega un papel fundamental en el diseño y ejecución de las manipulaciones, lo mismo que los niveles de los montos y los valores de precios negociados.

En caso de que se requiera, el uso de filtros permite un detalle más profundo de los componentes asociados a una manipulación dentro de una señal financiera, ya que se puede limitar la presencia de otros componentes en la señal que son “esperados”.

Al respecto, los coeficientes describen los picos y los efectos en bandas bajas en las señales a partir de valores altos en el espectro de poder, por lo anterior la identificación de cambios abruptos en la señal por el kernel de procesamiento de la señal es indispensable.

En consideración, pueden filtrarse elementos en las bandas intermedias los cuales pueden ser caracterizados a partir del análisis del escalograma y los coeficientes de la matriz kernel que resulta de procesar cada señal.

Sin embargo, se espera que esto no sea necesario ya que a partir del aprendizaje del mapeo no lineal entre los dos espacios funcionales (el espacio del que proviene la señal y el espacio en el que resulta luego del procesamiento) se caractericen los coeficientes de dichas bandas intermedias. Según Tripura & Chakraborty (2022) esto ofrece el aprendizaje de los detalles más finos de las dependencias paramétricas en la solución de problemas complejos, como lo es el caso de las señales financieras.

Según Mallat (2009) adaptar representaciones dispersas a las propiedades de la señal y derivar operadores de procesamiento eficientes es, una estrategia de supervivencia necesaria.

Tarea que se logra con facilidad al realizar aprendizaje sobre los coeficientes, sin embargo, entender los coeficientes puede ser retador, especialmente cuando la CWT es una transformada redundante y los coeficientes dependen de la wavelet elegida. (Percival & Walden, 2000)

Según Mallat (2009), ya sea a partir de distribuciones o cantidades físicas específicas, los coeficientes pueden interpretarse como la medida del flujo de energía a través de una ventana espectral en un intervalo de tiempo definido.

Entre más grande sea la ventana, más probable será que los coeficientes correspondan a una medida real de energía física. Esta debe ser lo suficientemente grande como para satisfacer la relación de incertidumbre de las desviaciones estándar de cada componente de la señal.

Sin embargo, lo anterior aplica particularmente en el caso del Espectrograma pues la representación obtenida depende explícitamente del tamaño de ventana elegido. Al considerar la CWT, se tendrá una matriz de Na por N , donde Na es el número de escalas y N el número de muestras que trae la señal.

Para el caso puntual se tendrían 500 muestras a partir del uso de una sola señal. Tener la ventaja de no solo ampliar el espacio de características, sino la cantidad de muestras es una ventaja adicional que ofrece el uso de las transformadas para el procesamiento de señales, particularmente de las transformadas wavelet y en especial la continua.

La escala máxima y mínima de acuerdo con Lilly & Olhede (2012) se determinan automáticamente en función de la dispersión de energía de la wavelet en frecuencia y tiempo, mediante los siguientes parámetros:

- Escala mínima: se presenta en la frecuencia más larga $s_0 = \frac{\omega'_x}{\pi}$, donde los π radianes equivalen al 10% de la frecuencia más larga
- Escala máxima: las escalas de la transformada wavelet son potencias de 2, denotadas por $s_0 \left(2^{\frac{1}{NV}}\right)^j$. Donde NV es el número de voces por octava, y j representa los rangos desde 0 hasta la escala más larga, la cual no será mayor a la desviación estándar de las frecuencias de acuerdo con la cantidad de muestras $\frac{N}{2\sigma_t}$.

La función que genera la CWT emplea normalización $L1$. De acuerdo con Lilly & Olhede (2012), con esta normalización si se tienen componentes oscilatorios de igual amplitud en a diferentes escalas, se tendrá la misma magnitud en la transformada.

El uso de la normalización $L1$ muestra una representación más precisa de la señal, pues logra preservar la energía que se pierde al implementar la transformada de forma numéricamente al emplear un mismo valor para todas las amplitudes de las frecuencias $\frac{1}{\sqrt{s}}$. (Lilly & Olhede, 2012)

Es importante recordar que los coeficientes de la CWT se obtienen calculando el producto interno de la señal con la wavelet de análisis desplazada y escalada e integrando el resultado. Por lo que los signos de los coeficientes dependerán de la coincidencia entre el período de oscilación de la señal y el de la wavelet. (Mallat, 2009)

En escalas en las que la oscilación en la wavelet ocurre en una escala mucho mayor o menor que el período de la onda sinusoidal, se obtienen coeficientes CWT cercanos a cero. (Mallat, 2009)

Lo anterior implica, que las transiciones abruptas en las señales dan como resultado coeficientes de wavelet con valores absolutos grandes, por tanto, para localizar con mayor precisión estos picos, se recomienda emplear los coeficientes de menor escala (frecuencia más alta), pues son estos los que facilitan la detección de cambios transitorios en los datos. (Lilly, 2017)

Sin embargo, con el fin de definir y seleccionar las mejores características que permitan caracterizar si una acción fue objeto de una manipulación o no, es necesario considerar todas las escalas. Finalmente, la presencia de valores altos en los coeficientes para una frecuencia en particular dará cuenta si hubo o no una manipulación en una señal financiera.

Es de tener en consideración que, al emplear wavelets analíticas como la Morse, los coeficientes que se obtendrán son de valor complejo, cuyas transformadas de Fourier se admiten solo en el eje real positivo. (Lilly & Olhede, 2009)

Por lo general, una señal analítica se define por la diferencia de la parte real de la señal original con su valor medio, y una parte imaginaria obtenida mediante el uso de una transformada integral de la señal real. (Castrillon, Montoya-Quintero & Sierra-Sossa, 2018)

Por tanto, para el aprendizaje se emplea solo la parte real de cada coeficiente, teniendo en consideración además que como previamente se ha mencionado la parte imaginaria, solo incluye información de fase.

Respecto a los efectos de largo plazo de las manipulaciones financieras, Hamilton & Susmel (1994) hacen mención que la persistencia de regímenes de volatilidad pueden llegar a durar años, y que regímenes de altísima volatilidad suelen estar asociados a recesiones económicas, por tanto la presencia de coeficientes de alto valor relativamente estables en bajas frecuencias en diferentes momentos del tiempo, darán cuenta de este tipo de elementos en la señal y pueden ser caracterizados por los algoritmos de aprendizaje.

Lo anterior es posible, ya que por lo general las señales financieras son no estacionarias (Hamilton & Susmel, 1994), lo que significa que la frecuencia de la señal depende del tiempo, y esta dependencia temporal es equivalente a la función de autocorrelación de la señal. (Boashash, 2015)

Lo anterior, fácilmente puede entenderse gracias a que la Densidad Espectral de Potencia (PSD por sus siglas en inglés) debería verse afectada debido a que cada término de ruido modula un componente cuya frecuencia varía con el tiempo, y esto no ocurre a causa de la manipulación, como en el caso del movimiento de las hélices de un helicóptero, donde su PSD permanece relativamente constante solo en los momentos en los que la nave está fija en el aire sin desplazarse. (Boashash, 2015)

Al respecto, Mallat (2009) precisa que, si la frecuencia de un tono en una señal tiene otros componentes de alta energía en diversas frecuencias, este será diferenciable si los componentes no han sido atenuados por la presencia de regímenes, y por tanto es posible la identificación de los eventos transitorios y los efectos en el largo plazo que producen las manipulaciones en una señal.

Sin embargo, es necesario hacer énfasis que hasta el momento no se logra identificar un patrón que defina de manera particular una manipulación en los escalogramas.

Lo anterior aplica en el caso de que se deseen emplear técnicas alternativas en la caracterización de los eventos asociados a las manipulaciones, pues no son generalizables al requerir una mayor cantidad de muestras o al depender fuertemente de los parámetros de entrada.

Por lo que el problema se aborda desde el comportamiento que los coeficientes wavelet describen.

De forma adicional, es de tener en consideración que las relaciones que puede un algoritmo de clasificación puedan detectar para la evaluación del comportamiento de un activo dado, conviene emplear otros estadísticos y otros elementos asociados a la señal, se pueda explicar el fenómeno a partir de variables más sencillas de interpretar, con menos características y eventualmente con una mejoría en el desempeño de clasificación del modelo a elegir.

Como precisan Tripura & Chakraborty (2022), los coeficientes wavelet se asocian con una frecuencia característica y con información espacial particular. Sin embargo, al haber considerado wavelets simétricas, se logra mantener una correspondencia única entre la escala y la frecuencia, lo que permite que bajo los parámetros previamente establecidos se logre estandarizar los rangos de frecuencias que serán evaluados y de esta manera entender los valores de coeficientes esperados según sea el caso.

Contar con wavelets simétricas tiene ventajas adicionales, ya que facilitan la aplicación de técnicas de ingeniería de características estadísticas, como también estadísticos espectrales los cuáles serán descritos más adelante.

Lo anterior, es posible debido a que la existencia de una correspondencia única entre la escala y la frecuencia requiere simetría sobre la frecuencia máxima, medida por la asimetría en el dominio de la frecuencia, y también igualdad entre la media y la moda del módulo cuadrático de la wavelet en el dominio de la frecuencia, por lo que las desviaciones sobre los valores esperados darían cuenta de anomalías en los coeficientes que son evaluados.

La correspondencia entre escala y frecuencia permite, que el aprendizaje de las relaciones entre los coeficientes se pueda abordar de diversas formas.

Por ejemplo, se puede aprender el comportamiento esperado para un activo en una fecha dada o se puede aprender el comportamiento de los coeficientes a partir de las diversas frecuencias detectadas.

Este último acercamiento, es mucho más generalizable y permite tratar cada muestra de manera individual, lo que permite que se tengan mayor cantidad de elementos para el entrenamiento.

Inicialmente, se considera el conjunto de frecuencias detectadas como las características que describirán el fenómeno, y posteriormente serán consideradas las variables resultantes del proceso de ingeniería de características.

Por otra parte, es necesario tener en cuenta que previamente se definió que el tamaño de la ventana (y por ende el muestreo de la señal) fuera diario, por lo que al clasificar una muestra como anómala se puede conocer la fecha en la que inició u ocurrió la manipulación.

Es fundamental recordar que en presencia de manipulaciones se presenta filtrado espectral, ya que se alteran los regímenes de frecuencia de forma abrupta. Por tanto, los efectos de la manipulación podrán ser detectados en varias muestras, particularmente en aquellas cercanas al evento de manipulación.

Debido a que los coeficientes de CWT tienen la misma resolución de tiempo que los datos, es posible detectar cambios transitorios en los datos que podrían asociarse a manipulaciones al exceder los valores esperados en un rango de frecuencia particular.

Otras características que también pueden ser relevantes para la identificación, detección y clasificación de señales y artefactos presentes en ellas, pueden incluir: concentración de frecuencias, complejidad de la señal (como una medida de entropía) y el flujo tiempo-escala. (Boashash, 2015)

Acercamientos con Aprendizaje de Máquina

Según Paniagua-Castrillón, Miranda, Garcés & Quintero-Montoya (2021), en ciencias de la computación, el aprendizaje puede verse como un proceso multidimensional de comprensión de fenómenos o la repetición competente de las tareas relacionadas con tales fenómenos. En el mundo real, la medición de dichos procesos está directamente relacionada con los datos. Los datos se convierten en una fuente muy importante de Información que debe ser transformada en Conocimiento, que debidamente formalizado es Inteligencia.

En consideración, al procesar los datos representados por los coeficientes, podemos generar conocimiento relacionado con los fenómenos de manipulación y al formalizarlo obtener inteligencia sobre los esquemas o elementos que permiten diferenciar una acción que ha sido manipulada de la que no.

La definición propuesta por Paniagua-Castrillón, Miranda, Garcés & Quintero-Montoya (2021), nos ofrece además un detalle adicional, y es como la repetición competente de las tareas relacionadas con la obtención de datos, también está relacionado con las tareas de aprendizaje, y es precisamente la consideración de ambos enfoques lo que llevara la pauta en la propuesta.

En primera instancia se busca la generalización de los procesos de detección y la escalabilidad en las tareas de validación de las acciones, como se ha mencionado, sin embargo, es fundamental entender las relaciones entre los datos, y como este entendimiento favorece la selección de los métodos, técnicas y estrategias más adecuadas que faciliten las tareas de detección.

De acuerdo con Quintero-Montoya (2018), las ramas principales del aprendizaje automático son el aprendizaje supervisado y el aprendizaje no supervisado. Según acota, en el aprendizaje supervisado, el sujeto se encuentra en una predicción precisa, mientras que, en el aprendizaje no supervisado, el objetivo es encontrar descripciones compactas de los datos.

El interés en ambos casos es la implementación de métodos que generalicen bien en presencia de datos no vistos, de forma que una nueva entrada, puede ser clasificada con la mayor precisión posible, o pueda ser descrita de manera adecuada a partir de las relaciones identificadas con el resto del conjunto de datos.

Teniendo en consideración que el objetivo del presente estudio es la identificación de activos que han sido objeto de manipulación. Se consideran aquellas técnicas de clasificación que permitan categorizar las muestras como anómalas o no anómalas.

Sin embargo, para que esto sea viable es necesario comprender las relaciones entre los coeficientes y las diversas medidas a las que pueden ser sometidas para elegir

de esta manera aquellas que mejor representen y expliquen el fenómeno de la manipulación.

Para tal fin, es importante iniciar el análisis a partir de técnicas de no supervisión con el fin de tener un acercamiento a las descripciones de las relaciones que pueden ofrecer los datos. Posteriormente, con el entendimiento de las relaciones detectadas profundizar en aquellas relaciones y elementos que faciliten la clasificación de los activos entre aquellos que han sido manipulados de los que no.

El análisis por clústeres es de las técnicas más populares en el aprendizaje no supervisado. Su objetivo es encontrar agrupaciones naturales en los datos o patrones en la información. Se conoce también como análisis de segmentación o análisis de taxonomías, ya que estas técnicas dividen los datos en grupos o conglomerados a partir de nociones de distancia o de semejanza.

En consideración los clústeres se forman de tal manera que los objetos en el mismo clúster son similares y los objetos en diferentes clústeres son distintos. (Landau, Leese, Stahl & Everitt, 2011).

De acuerdo con el tipo de análisis, el tipo de variables, el comportamiento de los datos y el objetivo del análisis se sugiere de manera particular algunos algoritmos sobre otros. (Arthur & Vassilvitskii, 2007)

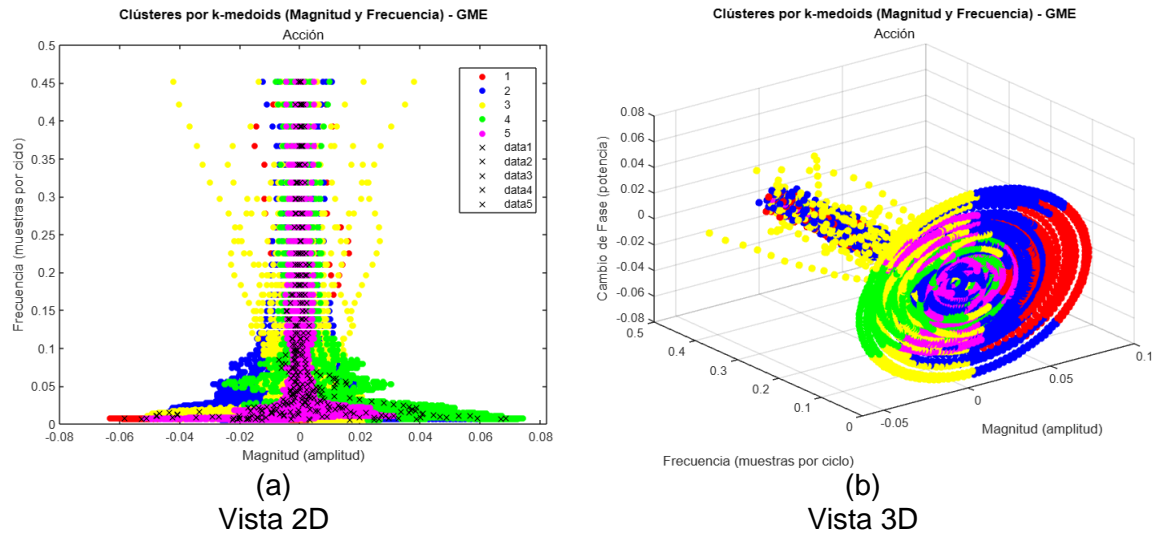
Para el caso particular se descarta el uso de k -medias y k -medianas, a pesar de ser los algoritmos más populares debido a su facilidad de uso e interpretabilidad. Ya que según precisan Raykov, Boukouvalas, Baig, & Little (2016), todas las variables tienen igual importancia en cada clúster por tal motivo tienden a ser de tamaños similares. Adicionalmente, considerando que estas técnicas parten de la noción de una medida de distancia sobre un centro suelen tener comportamientos esféricos en la distribución de los datos en cada clúster, lo que no favorece la evaluación de muestras con comportamientos dispersos o aleatorios como puede apreciarse en la Figura 29.

Si bien puede parecer obvio dicho ajuste, parte del análisis de la señal y el fenómeno

Considerando el comportamiento de los datos presentado en las gráficas, el análisis será efectuado mediante algoritmos de agrupamiento espectral los cuales son recomendados para datos con este tipo de distribuciones.

La ventaja que ofrecen sobre k -medias y k -medianas, es que no sólo tienen en consideración la distancia, sino también que tan similares son los puntos, entre sí y no respecto a un centroide.

Figura 29. Representación de clústeres mediante *k*-medoids



Se evalúan las técnicas DBSCAN o Agrupamiento espacial basado en la densidad en aplicaciones con ruido y Grafos Espectrales mediante la matriz de Laplace. Ambos métodos se recomiendan cuando los clústeres no tienen una forma definida y/o los datos tienen comportamientos aleatorios. (Von Luxburg, 2007)

De acuerdo con los algoritmos que implementa, cada técnica sugiere métodos para detectar la cantidad recomendada de clústeres con base en la distribución de los datos.

Sin embargo, en ambos casos se emplea la matriz vectorizada de los coeficientes en el que se tendrá para cada coeficiente la frecuencia en la que fue detectada y la concatenación de las muestras se hará en orden ascendente.

En el caso de grafos espectrales mediante la matriz de Laplace, se hará uso de los valores propios de dicha matriz, con base en una semilla de 5 clústeres para dicha estimación.

Lo anterior considerando que según Shi & Malik (2000) la cantidad de valores propios cercanos o iguales a cero, es un buen indicador de la cantidad de componentes conectados en un grafo de semejanza y, por lo tanto, es una buena estimación de la cantidad de clústeres en los datos.

Ya que la matriz de Laplace es una forma de representar un grafo de semejanza, que modela las relaciones de los puntos en la vecindad local de manera no dirigida (Ng, Jordan & Weiss, 2001). Partiendo de la premisa de que los datos son llevados a una dimensión más baja donde los clústeres quedan separados de manera más amplia.

En el caso de DBSCAN, es debe considerar un parámetro ϵ que define un radio alrededor de cada punto en el que se evalúan si otros puntos están incluidos o no, con el fin de efectuar la pertenencia a un clúster en particular.

Una estrategia para estimar un valor para ϵ es generar un grafo de k -distancias para los datos de entrada X . Para cada punto en X , se encuentra la distancia al k -ésimo punto más cercano graficando los puntos ordenados contra esta distancia. El gráfico contiene una rodilla. La distancia que corresponde a la rodilla es generalmente una buena opción para ϵ , porque es la región donde los puntos comienzan a disminuir en territorio atípico (ruido). (Ester, Kriegel, Sander & Xiaowei, 1996) Como puede verse en la Figura 30.

Figura 30. Grafo de k -distancias para la estimación clústeres en DBSCAN



Este gráfico requiere, primero que se ordenen las distancias de forma ascendente por pares más pequeñas a partir de los puntos mínimos de cada observación en X . (Ester, Kriegel, Sander & Xiaowei, 1996)

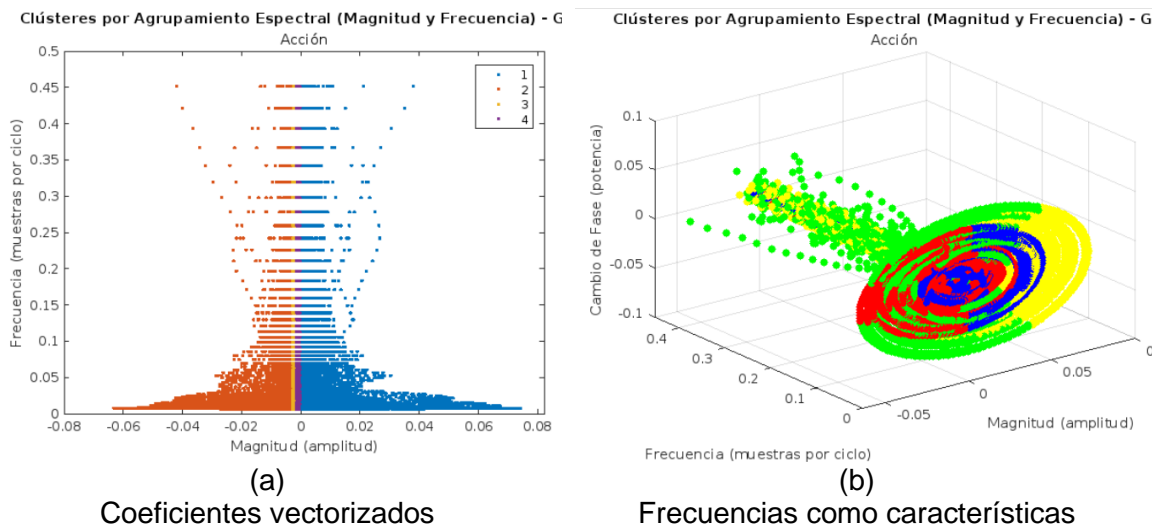
En este caso el doblez de la rodilla se da en 0.0002 y se sugiere como el valor de ϵ en la estimación de los puntos en el vecindario. Para la cantidad de puntos se sugiere emplear mínimo la cantidad de características mas uno, y es por esta razón que se emplearon las primeras 63 distancias, que corresponden a las 63 frecuencias detectadas para la acción en el ejemplo.

Luego de procesar para cada acción manipulada y los índices de referencia se tiene que los valores de clústeres más comunes aplicando la metodología de Grafos Espectrales mediante la matriz de Laplace fue de 3 y 4 clústeres. Mientras que mediante dbscan en cada caso fueron solo 2.

Considerando la distribución de los puntos de la Figura 31, pueden considerarse 3 clústeres para segmentar los coeficientes de acuerdo con el rango de distribución de energía por frecuencia y uno adicional para aquellos coeficientes con valor cero.

En la Figura 31(a) destaca como se crea un segmento para los coeficientes con dicho igual a cero. Al emplearlo en las muestras se evidencia que de forma análoga aquellas muestras cuyos coeficientes en su gran mayoría son cercanos a cero son clasificados en este mismo clúster.

Figura 31. Clústeres por agrupamiento espectral mediante la matriz de Laplace



Un detalle adicional para destacar con la selección de los clústeres está relacionado con el hecho de que se evidencia la estructura circular de las muestras y como éstas son agrupadas conforme a la cantidad de energía detectada de forma general en cada muestra.

Además, que en bajas frecuencias la energía suele ser mayor que en frecuencias más altas y no es común encontrar niveles altos de energía en tales frecuencias.

Al respecto es necesario precisar como una onda que no ha sido expuesta a un cambio fuerte en la energía registrada tendrá un comportamiento sinusoidal, mientras que en caso contrario se tendrán trayectorias atípicas como en la Figura 31 (b).

Selección de Características

Teniendo en consideración que, si bien evaluar los coeficientes de forma individual aportó información sobre la relación entre escalas y frecuencias. Procesar una gran cantidad de coeficientes hace inviable considerar características adicionales que puedan ayudar a explicar los resultados obtenidos con los modelos.

Sin embargo, al generar tantas características como sea posible, la dimensionalidad del problema aumenta, pero el tamaño del conjunto de datos etiquetado sigue siendo el mismo. (Quintero-Montoya, 2018)

Con lo que se pueden lograr capacidades interesantes de clasificación con una cantidad de datos suficiente para el entrenamiento de los modelos.

Al respecto es necesario precisar como el uso de la transformada wavelet ofrece que la cantidad de datos etiquetados aumente. Quintero-Montoya (2018), propone el siguiente ejemplo: suponga que debe determinar 18 clases de señales de audio y tiene 500 muestras de 5 segundos de duración cada una. Si la base de datos está equilibrada, debe tener aproximadamente 27 muestras etiquetadas para cada clase.

Una base de datos equilibrada se refiere a tener el mismo tamaño de muestra para que cada clase aprenda.

Si el proceso de ventanas se realiza teniendo en cuenta las características de la señal, se pueden tener 50 ventanas para cada muestra. Ahora la base de datos etiquetada se eleva a 1350 muestras para cada clase. Significa que el tamaño del espacio de entrenamiento S es de 24300 puntos de datos y es aproximadamente la longitud de la ventana.

El vector real de muestras de una señal $f(t)$, se conoce como $f(kT)$ donde T es el tiempo de muestro. Las secuencias generalmente no son muy útiles para clasificación (Quintero-Montoya, 2018). Sin embargo, bajo el supuesto de que algunas características relevantes deben proporcionar información para la construcción de modelos y, tal vez, otras son simplemente innecesarias, surge la selección de características.

Según Quintero-Montoya (2018) no supervisión, no es del todo cierto porque, de una forma u otra, es fundamental tener unos criterios claros para elegir los mejores modelos o más aún los más adecuados, más allá de los criterios formales, basados en cuantificación de índices, etcétera.

Agrega además que, al definir un espacio de datos es fundamental asignar cada dimensión de acuerdo con la naturaleza del problema, antes de ir directo a “moler” los datos y aprender cualquier kernel que los represente.

Ahora bien, Quintero-Montoya (2018) agrega que se puede decir que todo resultado depende, no sólo de la pregunta a responder, sino también de los procedimientos para encontrar una respuesta. El modelado y el aprendizaje de los datos en espacios de alta dimensión, y posteriormente reducidos a bajas dimensiones, pueden no ser los resultados esperados por el tomador de decisiones. Finalmente, no querrán conocer la información subyacente, querrán ver lo que la red tiene que decir.

En consideración, teniendo en cuenta que los coeficientes de las señales de tiempo financieras no tienen características obvias para ser usadas, no solo resulta importante conocer las relaciones de energía entre frecuencias, sino también resulta importante incluir y considerar todas aquellas características que puedan ofrecernos informacional adicional que permita explicar las relaciones en los datos.

Según Quintero-Montoya (2018) hay una lista muy diversa de características que pueden ser extraídas. Dependiendo del problema, el dominio del tiempo, el dominio de la frecuencia, el dominio espacial, las estadísticas, los datos espectrales y otros serán muy útiles. Sugiere los siguientes, como un buen conjunto de características para extraer de la descomposición de una señal:

La media, representa la energía directa de la señal definida como:

$$\bar{x} = 1/N \sum_{n=1}^N x_n ;$$

- La desviación estándar, es la derivada del segundo momento y la media de x, y representa la energía alterna de una señal y está definida por:

$$\hat{\sigma} = \sqrt{\frac{1}{N} \sum_{n=1}^N (x_n - \bar{x})^2};$$

- La curtosis, es un indicador de no-estacionalidad en una señal, se usa principalmente para la detección de artefactos en señales continuas. Puede ser estimado como: $K = \frac{\frac{1}{N} \sum_{n=1}^N (x_n - \bar{x})^4}{\hat{\sigma}^4};$
- La frecuencia mediana, es una estimación de la mediana en el espectro de poder. Esta característica representa la distribución de la energía por sus componentes espectrales (potencia). Esta dada por $\tilde{f} = \frac{1}{2} \sum_{w=1}^W P_w$; donde P es la densidad del espectro de energía y W el tamaño de la transformada de x;
- La frecuencia media, es una característica que está asociada a la frecuencia central de una señal. Está definida por $\bar{f} = \frac{\sum_{w=1}^W f_w P_w}{\sum_{w=1}^W P_w}$;
- El valor máximo espectral, el valor máximo espectral representa el valor más alto de energía de una señal y está definido por: $P_{max} = \max P_w.$

Para el caso de aplicación se tiene en cuenta las medidas que pueden derivarse del cálculo de las características frecuencias generadas por la CWT como:

- la entropía, entendida como la dispersión frente a valores centrales y extremos;
- en la frecuencia (no solo la magnitud, sino también entendida como qué tan común es el dato);
- relacionadas con la energía (la magnitud de los coeficientes);
- y estadísticos espectrales complementarios.

El objetivo es generar un espacio de características que sea suficiente, no solo para clasificar las muestras de las acciones en anómalas y no anómalas, sino que permita explicar el motivo de la clasificación a partir de variables que den cuenta de qué tan disperso o inusual es un dato. Según Quintero-Montoya (2018), la extracción de características permite la imitación del procedimiento humano en definición de las características más diferenciables.

Lo anterior, implica que las variables generadas deben aportar y explicar el fenómeno, pues de otra manera su inclusión podría degradar el desempeño del modelo. Se necesita un espacio que no sea tan grande que haga el problema muy complejo de resolver, y no muy pequeño que pueda llevar a una capacidad pobre de aprendizaje. (Quintero-Montoya, 2018).

Es importante considerar que las distancias en los espacios de altas dimensiones no son necesariamente fáciles de comprender y generalmente la distancia entre los puntos en altas dimensiones puede estar más cercana de lo esperado para dos clases particulares. Esto se evidencia en la Figura 31, donde hay concentración de alta energía en altas y en bajas frecuencias.

Al respecto Quintero-Montoya (2018) precisa, que intentar generar una separación lineal o no lineal en espacios de alta dimensión a través de clasificadores puede requerir un gran esfuerzo computacional y, si las características no son relevantes, puede provocar un sobreajuste en el aprendizaje de los datos.

La selección de características es una técnica necesaria para eliminar información no deseada, y reducir el espacio de características para disminuir los costos computacionales, permitiendo mejores resultados de clasificación.

La idea principal que subyace a la selección de características es encontrar un conjunto adecuado de características (dimensiones) útiles en la clasificación, que sean identificables y de fácil descripción que permitan establecer criterios para buenos procesos de toma de decisiones.

Sin embargo, independientemente del método utilizado para su generación, una vez obtenido el espacio de alta dimensión, se debe encontrar la manera de reducirlo a las características más relevantes. (Quintero-Montoya, 2018).

Por tanto, para cada muestra además de las frecuencias, serán generadas las características previamente descritas y mediante métodos de filtrado, serán seleccionadas las variables que mejor poder descriptivo tengan.

Los métodos de filtrado se basan en las propiedades estadísticas de los predictores y su relación con la respuesta. Se aplican antes de entrenar un modelo y, por tanto, pueden ser aplicados sin importar su tipo. Generalmente son usados en combinación con los métodos de envoltura o los métodos embebidos. (Hopf & Reifenrath, 2021)

Como los datos no tienen un comportamiento monótonico según se evidencia en la Figura 31(b), no es viable emplear los métodos de correlación para la selección de variables, por tanto, en el caso particular de aplicación deben ser considerados los test de estadísticos de independencia. (Herwartz & Maxand, 2020)

La idea principal detrás de este tipo de test, es que los valores de dos características independientes deberían distribuirse de una manera determinada cuando son considerados juntos. De esta manera es posible calcular la probabilidad o p -valor de que dos características son independientes analizando su distribución. (Herwartz & Maxand, 2020)

El p -valor para este tipo de pruebas se obtiene para describir la probabilidad de obtener un estadístico- F que indique si dos características no están relacionadas. (Tabachnick & Fidell, 2013)

El estadístico- F , puede ser descrito brevemente como la relación entre varianzas de un grupo. Por tanto, requiere que una variable sea categórica o represente una respuesta. (Tabachnick & Fidell, 2013)

Por tanto, se requiere contar con un etiquetado previo o con un conjunto de categorías definidas. Se opta por evaluar el poder predictivo de las variables que permitan definir si una muestra pertenece a una acción que ha sido manipulada, de una que no.

Según Guyon & Elisseeff (2003), en caso de que la probabilidad esté bajo un umbral (generalmente 0.05 o 0.01) se puede concluir que las características son estadísticamente dependientes.

En consideración, un p -valor por debajo de los umbrales indicará que ambas características están fuertemente relacionadas y que, por tanto, ambas añadirán suficiente valor predictivo a la otra. Es decir que dicha variable es relevante para clasificar si una muestra en particular pertenece a una acción que ha sido manipulada o pertenece a una acción que no lo ha sido.

El test Chi-cuadrado, es de los test más comunes por su sencillez de cálculo e interpretabilidad: diferencias grandes entre los valores observados y esperados producen un estadístico chi-cuadrado más grande.

Valores de Chi-cuadrado más pequeños, implican que los conteos reales están cerca de los conteos esperados en la mayoría de los casos, y existe una relación débil entre las características. (Greenwood & Nikulin, 1996)

Mediante $-\log(p)$ se puede crear un puntaje, donde valores altos indican que el predictor es importante. Sin embargo, es precisamente un p -valor pequeño para el Chi-cuadrado lo que indica que la variable predictora depende de la variable de respuesta y , por lo tanto, es una característica realmente importante. (Shreffler & Huecker, 2023)

Si hay suficientes observaciones, ambas características satisfacen uno o más supuestos sobre sus distribuciones (p. ej., normalidad), y si no hay otros sesgos (p. ej., sesgos de recolección/medición, diseño experimental deficiente), el p -valor es un buen estimador de la probabilidad de la independencia. (Guyon & Elisseeff, 2003)

Conviene por tanto incluir, algún análisis adicional, que ofrezca una visión complementaria del fenómeno y que no dependa de tantos supuestos.

A partir del análisis de la dispersión de las variables relacionadas, se puede obtener información muy relevante sobre las capacidades predictivas de una variable, como se ha mencionado previamente.

Un test derivado de la varianza es el Análisis de la Varianza o el ANOVA test, en el que se puede obtener un p -valor que describe la probabilidad de obtener un estadístico- F si dos características no están relacionadas. Como antes, si la probabilidad cae por debajo de un umbral determinado (normalmente 0.05 o 0.01), se puede concluir que las características son estadísticamente dependientes. (Tabachnick & Fidell, 2013)

En consideración un p -valor bajo puede entenderse como que, al estar las variables relacionadas, una añade un valor predictivo significativo a la otra. Sin embargo, es importante reiterar que el p -valor no es lo mismo que la probabilidad de que las variables sean independientes. No obstante, ambos conceptos están íntimamente relacionados. (Shreffler & Huecker, 2023)

Por tanto, conviene a la par emplear el análisis de la varianza por sí sólo, pues se tendrá información relevante sobre el aporte de una característica al modelo. Lo anterior, debido a que las variables con baja varianza tienden a tener menor poder predictivo, que aquellas con alta varianza. (Shreffler & Huecker, 2023)

Además, que no dependen de la consideración de una variable respuesta como si requieren otros métodos de filtrado como los test estadísticos de independencia, los métodos de envoltura o los métodos embebidos. (Tabachnick & Fidell, 2013)

En el caso del análisis de la varianza, para seleccionar los predictores, se puede definir un umbral de varianza mínimo para excluir aquellos predictores con poca variabilidad, ya que es poco probable que ayuden a explicar las asignaciones que realice un modelo.

Otra alternativa, consiste en definir un umbral de varianza total en la que sean seleccionadas aquellas variables que logren dicho umbral.

Para que las validaciones previas sean viables, se requiere que las variables sean escaladas, por lo que se realiza mediante: $X^i \rightarrow \frac{X^i - E[X^i]}{\text{Rango}[X^i]}$, donde X^i es la i -ésima variable.

Finalmente, se sugiere calcular la proporción de la varianza de cada variable sobre la varianza total del modelo y ordenar los valores de mayor a menor, para aplicar los umbrales de varianza según sea el caso. Gráficamente, puede apreciarse en la Figura 32.

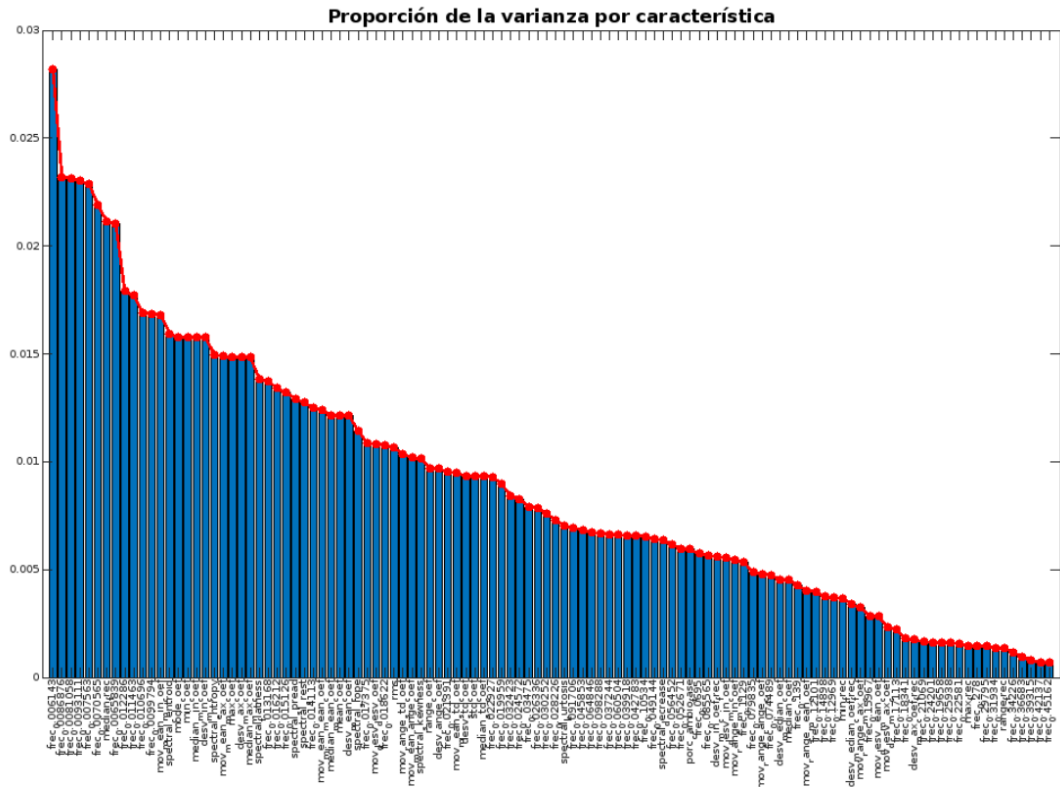
De acuerdo con Shreffler & Huecker (2023), elegir un umbral puede ser complicado en la práctica, pero idealmente un buen umbral separará las características de alta varianza de una o más características de baja varianza para que éstas últimas sean descartadas.

Es importante tener en consideración, que el rendimiento del modelo puede verse afectado si se descarta demasiada varianza. Por tanto, si un grupo de características de baja varianza constituye colectivamente una gran proporción de la varianza total, o si las características de varianza más baja aún constituyen una proporción significativa de la varianza total, se deben considerar otros métodos de selección de características. (Shreffler & Huecker, 2023)

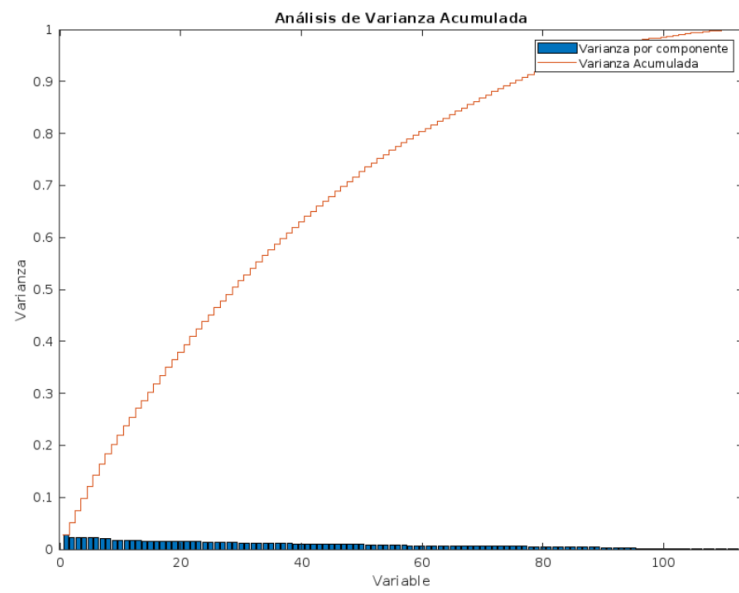
En consideración, y como previamente se había mencionado conviene realizar validaciones adicionales mediante los métodos de envoltura y los métodos embebidos. Ambos ocurren como parte del entrenamiento de un modelo:

- En el caso de los métodos de envoltura se entrenan diversos modelos usando diferentes combinaciones de características y seleccionando aquellas que fueron empleadas en el modelo con el mejor desempeño; (Guyon & Elisseeff, 2003)
- Por su parte, los métodos embebidos, enfatizan solo las características más descriptivas del set de datos. (Ferri, Pudil, Hatef, & Kittler, 1994)

Figura 32. Proporción de la varianza por característica



(a)
Coeficientes vectorizados



(b)
acumulada

El método de envoltura por excelencia hace uso de algoritmos de búsqueda secuencial, que agregan o eliminan características de un subconjunto candidato mientras evalúa el criterio de clasificación. Dado que una comparación exhaustiva del valor del criterio en los $2n$ subconjuntos de un conjunto de datos de n características generalmente no es factible (dependiendo del tamaño de n y el costo de las llamadas objetivas se mueve en una sola dirección: hacia adelante o hacia atrás. (Guyon & Elisseeff, 2003)

Para el caso particular no se considera la búsqueda hacia atrás ya que el set de datos al tener una varianza relativamente similar entre variables genera poca innovación en el modelo. Esto no solo hace que tarde más en converger, sino que restringe la selección de variables al buscar el modelo más parsimonioso.

En la Figura 32, la caída de los datos menos relevantes mediante la varianza se da al final de la clasificación, dando a entender que una gran cantidad de las variables calculadas aportan en la explicación del fenómeno y por lo que por un lado se conserva una varianza adecuada para el modelo, sin embargo, es conveniente el emplear validaciones adicionales para seleccionar las características más representativas.

Por tanto, se considera también los métodos embebidos, en este caso será empleado un modelo basado en árboles, ya que estos test son útiles en seleccionar la mejor interacción que permita dividir el set de datos mediante las variables más relevantes. (Guyon & Elisseeff, 2003)

Existen diversos tipos de validaciones, sin embargo, en el caso de las implementaciones con árboles la definición de la relevancia de una variable viene dada por la suma de los cambios en el error cuadrático medio (MSE en inglés) debido a divisiones en cada predictor y dividiendo la suma por el número de nodos de la rama. (Guyon & Elisseeff, 2003)

Teniendo en cuenta el tamaño del set de datos se propone un árbol medio (máximo con total de 20 divisiones) en donde las variables con estimaciones más altas corresponderán a los predictores más relevantes mediante este método.

En el Anexo C, se encuentra un detalle de las variables seleccionadas mediante cada método. Teniendo en consideración provista por cada una, se definen como relevantes los siguientes predictores para el modelo:

- Como parte del set de datos fueron considerados 62 valores de frecuencia en una banda entre los 0.0065 hasta los 0.3 ciclos por **Frecuencias:** freq_0_0081058, freq_0_017375, freq_0_021391, freq_0_014113, freq_0_028226, freq_0_018622, freq_0_015126, freq_0_0086876, freq_0_007563, freq_0_0070565, freq_0_0093111, freq_0_032423,

frec_0_013168, frec_0_030252, frec_0_026336, frec_0_0065839,
frec_0_045853;

- **Medias móviles:** mov_range_std_coef, mov_range_max_coef, mov_desv_max_coef, mov_desv_desv_coef, mov_range_range_coef, mov_desv_mean_coef, mov_range_min_coef, mov_desv_min_coef, mov_range_mean_coef, mov_mean_min_coef, mov_mean_max_coef, mov_mean_std_coef, mov_mean_range_coef;
- **Dispersión del valor estadístico:** median_min_coef, median_max_coef, median_std_coef, median_mean_coef;
- **Tendencia central:** min_coef, mode_coef, rms, max_coef;
- **Relacionados con el espectro de potencia:** desv_median_coef_frec;
- **Estadísticos espectrales:** spectral_slope

Resultan 40 características de un total de 128 que fueron evaluadas.

muestra, de los cuales se eligieron 17 valores ubicados en un segmento de baja frecuencia entre los 0.007 y los 0.045 ciclos por muestra, confirmando los supuestos previos sobre su relevancia en la descripción de señales con eventos transitorios.

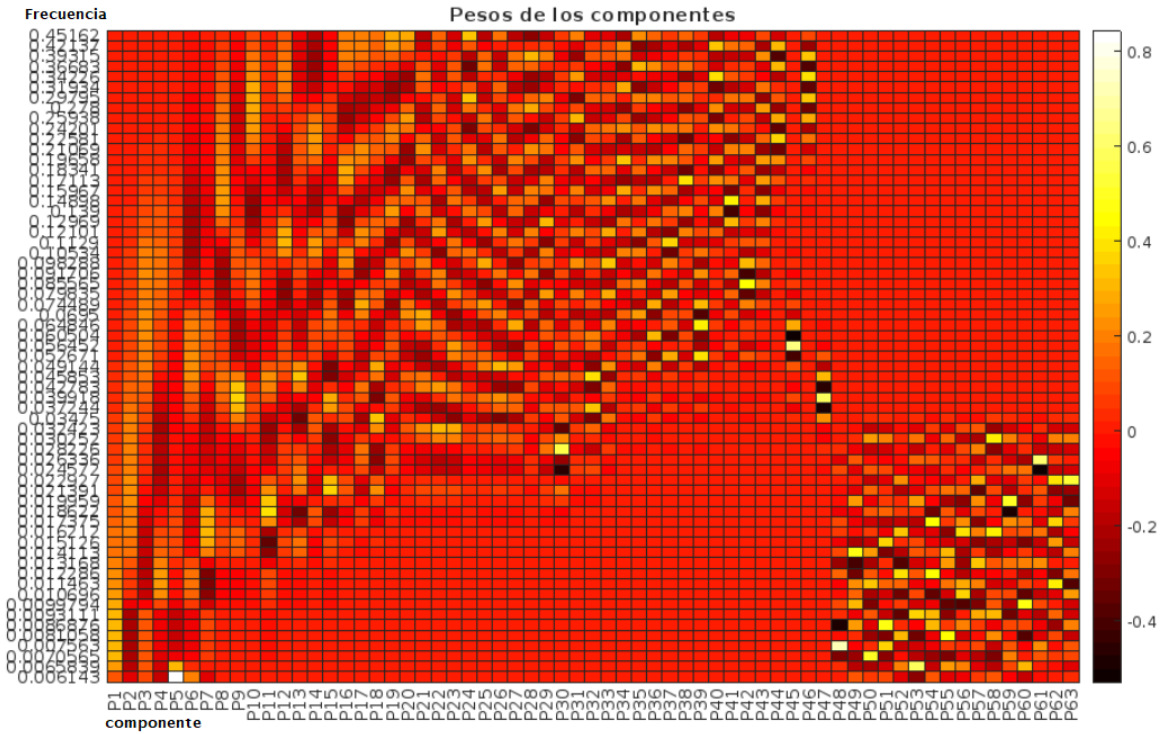
La selección de las variables se hizo con base en la aparición de las variables mediante Chi-cuadrado en al menos la mitad de las pruebas. Luego se eligieron de este primer filtro aquellas variables que al menos hayan sido seleccionadas una vez en alguno de los métodos embebidos o de envoltura. Ya que se confirma que realmente las variables elegidas aportan a la descripción del modelo y por tanto tienen algún poder predictivo.

Estos criterios pueden ser aún más restrictivos y de esta manera disminuir aún más la cantidad de variables con las que el modelo puede ser entrenado. Por ejemplo, partir de aquellas variables presentes solo en el análisis consolidado y/o al menos en algún otro método.

Sin embargo, no existen reglas satisfactorias para descubrir una función de base wavelet óptima para la transformada wavelet. Al respecto Zhang, He, Yang, Zhang & Wu (2022) comentan como una función wavelet de base óptima se selecciona en función del cálculo de la puntuación de distancia y la puntuación media de sus características, en dónde las características extraídas por la función de base de wavelet óptima se consideran las mejores características para representar las señales elegidas en particular.

En la Figura 33 mediante PCA se sugiere que hay una cantidad de frecuencias que explica mejor el fenómeno que otras como puede apreciarse.

Figura 33. Análisis PCA sobre las características frecuenciales



Sin embargo, PCA no se considera como parte del proceso, es recomendada especialmente para procesos de visualización del comportamiento de los puntos en otros espacios y para set de datos anchos o de gran dimensión, es decir aquellos en los que el número de características es comparable o incluso mayor que el número de observaciones. (Ariza-Jiménez, Villa & Quintero-Montoya, 2019)

Adicionalmente, al reducir la dimensionalidad de los datos, no se garantiza una representación precisa de la estructura del grupo original. (Ariza-Jiménez, Villa & Quintero-Montoya, 2019)

Sin embargo, los vectores propios que corresponden a los valores propios más grandes (los componentes principales) corresponden a una gran fracción de la varianza inicial, es decir, la que corresponde a los datos originales. Los primeros vectores propios generalmente se interpretan en términos del comportamiento físico a gran escala de un sistema particular; y conservará las características de varianza más importantes. (Jorgensen, Kang, Song & Tian, 2019)

Confirmando de esta manera, que, en su mayoría, los componentes de baja frecuencia se utilizan como características porque representan lo esencial de la señal original. (Zhang, He, Yang, Zhang & Wu, 2022)

Finalmente, el objetivo principal es obtener información sobre el por qué ciertas características son importantes y de qué manera podrían explicar el fenómeno o la clasificación).

Exploración de modelos de clasificación

Al generar el espacio de características, el objetivo es asociar los resultados numéricos con la física real del fenómeno que se investiga. Se debe tener en cuenta el hecho de que cada dimensión en su conjunto de datos puede ser una característica, no implica reducción de dimensión. Las características pueden ser temporales, espaciales, frecuenciales o en el dominio de escala a través de la transformada Wavelet, y cada uno de ellos podría ser una propiedad del propio sistema. (Quintero-Montoya, 2018)

Sin embargo, en general la extracción de características es extremadamente importante para el diagnóstico de fallas intrínsecas, y las distintas características son útiles para representar diferentes fallas. (Zhang, He, Yang, Zhang & Wu, 2022)

En el contexto fallas se entenderían como los efectos que producen las manipulaciones sobre una señal.

Por lo anterior, a partir de las cuarenta variables que seleccionadas se consideran diversos modelos de clasificación para elegir aquel modelo que logre los mejores resultados, teniendo en consideración su capacidad de explicar las clasificaciones de tal manera que sea viable una adecuada descripción del fenómeno.

Según Hutter, Hoos & Leyton-Brown (2010), al momento de elegir un algoritmo debe considerarse un balance entre diversas características intrínsecas que cada uno posee, tales como:

- velocidad de entrenamiento,
- uso de memoria,
- precisión predictiva en nuevos datos,
- transparencia o interpretabilidad, es decir, con qué facilidad puede comprender las razones por las que un algoritmo hace sus predicciones.

Canbek, Sagiroglu, Temizel & Baykal (2017) sugieren iniciar con modelos que sean rápidos y fáciles de interpretar. Ya que en caso de que no logren buenas predicciones, podrán ser considerados otros clasificadores con mayor flexibilidad.

Según Hutter, Hoos & Leyton-Brown (2010), para controlar la flexibilidad, se deben evaluar las restricciones y los límites particulares de cada tipo de clasificador. Y para evitar el sobreajuste, se debe buscar el modelo de menor flexibilidad que proporcione un nivel suficiente de precisión.

Tabla 1. Análisis de diversos modelos de clasificación

| Clasificador | Soporta Múlticlase | Soporta Categorías | Velocidad Predicción | Uso de Memoria | Inter-pretabilidad |
|-----------------------|---------------------------------------|--------------------|---|--|---|
| Árboles de decisión | Si | Si | Rápido | Poco | Fácil |
| Análisis de regresión | No. Combina varias funciones binarias | Si | Rápido | Poco para lineal. Largo para cuadrática | Fácil |
| SVM | No. Combina varios SVM binarios | Si | Medio para lineal. Rápido otros | Medio para lineal. Los otros, medio multiclase, largo binario | Fácil para lineal. Complejo para otros kernel |
| Náive Bayes | Si | Si | Medio para distribuciones simples. Lento para kernel y altas dim. | Poco para distribuciones simples. Medio para kernel y altas dim. | Fácil |
| Vecino más cercano | Si | Si | Lento para cúbico, medio para el resto | Medio | Difícil |
| Kernel | Si | Si | Rápido a medio de acuerdo con los kernel elegidos | Poco a mucho de acuerdo con los kernel | Fácil para lineal. Complejo para otros kernel |
| Ensamble | Si | Si | Rápido a medio de acuerdo con los algoritmos | Poco a mucho de acuerdo con los algoritmos | Difícil |

Nota. Las siguientes listas definen los criterios empleados en la clasificación de cada característica de los modelos:

- Velocidad: Rápido (0.01 segundos), medio (1 segundo) y lento (100 segundos);
- Memoria: Pequeño (1 MB), medio (4 MB) y grande (100 MB).

Los resultados fueron basados en un análisis de muchos conjuntos de datos en los que fueron considerados hasta 7000 observaciones, 80 predictores y 50 clases.

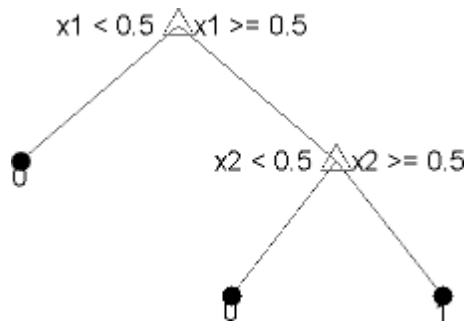
Fuente: Hutter, Hoos & Leyton-Brown (2010)

Para el caso particular, teniendo en cuenta que se definió un modelo de clasificación binario en el que se busca determinar si una muestra hace parte o no de una acción que ha sido manipulada. Que las clases se encuentran balanceadas al evaluar acciones de referencia y que la cantidad de muestras es suficiente son evaluados los modelos propuestos en la Tabla 1, a excepción de los modelos de ensamble.

Los árboles de decisión o árboles de clasificación predicen respuestas a los datos a partir de la definición de un nodo raíz (comienzo) hasta un nodo hoja. El nodo hoja contiene la respuesta, que normalmente son nominales, como 'verdadero' o 'falso'. (Breiman, Friedman, Olshen & Stone, 1984)

En la Figura 34, puede apreciarse un ejemplo en el que se tiene un árbol que realiza sus predicciones basado en dos predictores x_1 y x_2 . Para predecir, parte del nodo inicial, representado por un triángulo. La primera decisión se da sobre si x_1 es menor que 0.5. Si la condición se cumple, sigue la rama izquierda clasificando el dato como tipo 0. Si se da el caso contrario se clasifica el dato como tipo 1.

Figura 34. Ejemplo de árbol de clasificación



Sin embargo, si x_1 excede 0.5, entonces sigue la rama derecha y en esta división el árbol pregunta si x_2 es menor que 0.5. Si esto ocurre sigue la rama izquierda clasificando el dato como tipo 0. Si se da el caso contrario se clasifica el dato como tipo 1.

El realizar preguntas binarias, los datos pueden irse dividiendo en la medida que se detecte aún suficiente variabilidad en los resultados luego de aplicar los criterios seleccionados, por tal motivo generalmente los árboles suelen presentar mucho sobre-ajuste ya que pueden dividir los datos hasta que la variabilidad sea mínima. (Loh, 2002)

Existen diversos métodos no sólo para seleccionar la mejor cantidad de divisiones, sino también para seleccionar las variables más apropiadas en cada división. Sin embargo, considerando su sencillez de implementación conviene evaluar árboles en diversos tamaños para tener una base de cuál deberían ser los valores recomendados para el problema en cuestión.

En consideración se define un árbol fino, es decir con una gran cantidad de divisiones (100), un árbol medio (20 divisiones) y un árbol ancho sólo con 4

divisiones. En cada caso fueron empleados las cuarenta características y no se realiza ningún proceso adicional de selección de variables.

Para la selección de las divisiones en cada nodo se emplea el índice de diversidad de Gini: $1 - \sum_i p^2(i)$.

Este índice consiste en la suma sobre las i -clases en cada nodo y la $p(i)$, que corresponde a la fracción observada de clases con clase i que llega a dicho nodo. Un nodo con una sola clase (un nodo puro) tiene índice de Gini 0; de lo contrario, el índice de Gini es positivo. Entonces, el índice de Gini es una medida de la impureza del nodo. (Nembrini, König & Wright, 2018)

Esta misma medida se emplea en los procesos de división de cada nodo. De acuerdo con Loh (2002), el procedimiento es el siguiente: Se calcula el ponderado de impureza del nodo t , i_t . Luego se estima la probabilidad de que una observación esté en el nodo t mediante: $P(T) = 1 - \sum_{j \in T} w_j$.

w_j es el peso de una observación en j y T es el set de todos los índices de las observaciones en el nodo t . En caso de que no se especifiquen, pesos para todos: $w_j = \frac{1}{n}$, donde n es el tamaño de la muestra.

De esta manera al maximizar la ganancia de impurezas (ΔI) sobre todos los candidatos en un nodo, se puede elegir entre los candidatos aquellos que tengan la ganancia más alta de impurezas.

La variable elegida se corta en el punto en el que se maximiza la ganancia de impureza.

Sin embargo, según Ghane & Braga-Neto (2021) los árboles son propensos a verse afectados por el error de sustitución. El error de sustitución es la diferencia entre los datos de entrenamiento de respuesta y las predicciones que hace el árbol de la respuesta en función de los datos de entrenamiento de entrada. Si el error de sustitución es alto, no puede esperarse que las predicciones del árbol sean buenas. Sin embargo, tener un error de reposición bajo no garantiza buenas predicciones para nuevos datos. El error de sustitución es a menudo una estimación demasiado optimista del error predictivo en los datos nuevos.

En consideración una clase de modelos que pone esta situación es el clasificador naïve de Bayes. Según Webb (2010) este clasificador está diseñado para ser usado cuando los predictores son independientes entre sí dentro de cada clase, pero parece funcionar bien en la práctica incluso cuando la suposición de independencia no es válida. Clasifica los datos en dos pasos:

- **Paso de entrenamiento:** utilizando los datos de entrenamiento, el método estima los parámetros de una distribución de probabilidad, asumiendo que los predictores son condicionalmente independientes dada la clase.
- **Paso de predicción:** para cualquier dato de prueba no visto, el método calcula la probabilidad posterior de que esa muestra pertenezca a cada clase. A continuación, el método clasifica los datos de prueba según la mayor probabilidad posterior.

La suposición de independencia condicional de clase simplifica enormemente el paso de entrenamiento, ya que puede estimar la densidad condicional de clase unidimensional para cada predictor individualmente. Si bien la independencia condicional de clase entre los predictores no es cierta en general, la investigación muestra que esta suposición optimista funciona bien en la práctica. Esta suposición de independencia condicional de clase de los predictores permite que el clasificador *náive* de Bayes estime los parámetros necesarios para una clasificación precisa mientras usa menos datos de entrenamiento que muchos otros clasificadores. Esto lo hace particularmente efectivo para conjuntos de datos que contienen muchos predictores.

El paso de entrenamiento en la clasificación *náive* de Bayes se basa en estimar $P(X|Y)$, la probabilidad o densidad de probabilidad de los predictores X dada la clase Y . (Webb, 2010) Generalmente se hace uso de una función de distribución gaussiana, sin embargo, también puede ser usado funciones kernel.

Como la validación se hace variable por variable, según sea el caso para un par en particular puede aplicarse un método diferente, pues en el caso de que un par de predictores tenga un comportamiento normal, el clasificador de Bayes estima una distribución normal separada para cada clase calculando la media y la desviación estándar de los datos de entrenamiento en esa clase. (Webb, 2010)

Por su parte, la distribución kernel es apropiada cuando los predictores tienen una distribución continua, es útil particularmente en los casos en que la distribución de un predictor pueda estar sesgado o tener múltiples picos o modas. (Webb, 2010)

Requiere más tiempo de computación y más memoria que la distribución normal. Para cada predictor que modele con una distribución de kernel, el clasificador *náive* de Bayes calcula una estimación de densidad de kernel separada para cada clase en función de los datos de entrenamiento para esa clase. Por defecto, el kernel se recomienda el kernel normal y que el clasificador selecciona un ancho de banda automáticamente con base en la información disponible para cada clase y predictor.

Los clasificadores naïve de Bayes asignan observaciones a la clase más probable (en otras palabras, la regla de decisión máxima a posteriori).

Explícitamente, el algoritmo sigue estos pasos:

- Estime las densidades de los predictores dentro de cada clase.
- Modelar probabilidades posteriores según la regla de Bayes. Es decir, para todo $k = 1, \dots, K$, donde la variable aleatoria será índice de la clase de una observación, los X serán los predictores de una observación y la probabilidad posterior viene dada por: $\pi(Y = k)$.
- Por último, la observación se clasifica mediante la selección de probabilidad posterior máximo de una observación para cada clase.

En cada caso, se pueden usar una media ponderada en cada predictor, donde el peso, puede ser calculado de forma empírica, sin embargo, lo anterior requiere que las frecuencias relativas en cada clase sean representativas, para lograr una buena capacidad predictiva.

Al respecto, Manning, Raghavan & Schütze, (2008), consideran que naïve Bayes puede funcionar bien cuando hay una cantidad significativa de datos faltantes al predecir los valores faltantes de las distribuciones de probabilidad generadas. Precisamente por obviar los supuestos de independencia.

Lo que puede resultar complejo de explicar, a pesar de la sencillez del cálculo y la estimación.

Por su parte, si bien los modelos de regresión consideran un valor continuo de acuerdo con las relaciones expresadas mediante las variables con las que se define el modelo. En el caso de la regresión logística el objetivo es estimar variables de respuesta binarias como sí, no, o verdadero o falso, todas las cuales pueden codificarse como cero o uno.

Lo anterior es posible, ya que se la regresión logística, ajusta una función logística, conocida como la función sigmoide. Esta curva va desde 0 a 1 y estima la probabilidad de que se elijan las clases en función del comportamiento de los predictores. Por lo que la tarea en este caso es encontrar los coeficientes de la ecuación para construir la función sigmoide. (Kleinbaum & Klein, 2010)

Es importante tener en consideración, que adicionalmente debe definirse un umbral de probabilidad. Si la probabilidad al evaluar una muestra es mayor que dicho umbral se dice que el valor pertenece a la clase codificada en 1, de lo contrario no. (Nagelkerke, 1991)

Las regresiones logísticas trabajan bien cuando los datos son muy limpios y las relaciones no son demasiado complejas. Debido a que es rápido de entrenar, se considera como un buen punto de referencia inicial para evaluar la precisión de otros modelos.

En consideración será implementado un modelo generalizado. Un modelo de regresión lineal generalizado es una clase especial de modelos no lineales que describen una relación no lineal entre una respuesta y los predictores. (McConway, 2003)

Un modelo de regresión lineal generalizada tiene características generalizadas de un modelo de regresión lineal. La variable de respuesta sigue una distribución normal, binomial, Poisson, gamma o gaussiana inversa con parámetros que incluyen la respuesta media μ . Una función de enlace f define la relación entre μ y la combinación lineal de predictores. (McConway, 2003)

Teniendo en consideración que los valores corresponden a cero o una al tener un modelo de clasificación binario, que puede ser representado como 0 u 1, se recomienda emplear la distribución binomial para la ejecución del modelo. Por lo que la función de enlace estaría dada por: $\log\left(\frac{\mu}{1-\mu}\right) = X_b$. (Nagelkerke, 1991)

A partir, de esta función se calcula la probabilidad $f(\mu)$, y la inversa calcula la media $g(X_b) = \mu$, y de esta manera estaría definido el umbral.

Teniendo en consideración que no se detecta una relación fuerte entre los predictores, se sugiere emplear un constructor lineal en el que se tenga para cada una de las 40 características un intercepto y un término lineal para cada predictor y de esta manera realizar las estimaciones.

A veces, las relaciones entre los datos puede ser demasiado complicados para una fórmula matemática simple. En casos como este, usar un modelo de clasificación conocido como k -vecinos más cercanos, o KNN, puede ser un buen enfoque.

Los modelos de vecinos más cercanos asumen que cosas similares existen muy cerca, o, en otras palabras, se agrupan cerca unas de otras.

En este tipo de algoritmos resulta clave la definición de una noción de distancia o una métrica de distancia y la cantidad k de vecinos más cercanos (de allí su popular nombre).

Esto se conoce como un mecanismo de votación por mayoría. Un modelo KNN para hacer una nueva predicción requiere hacer referencia a todos los datos existentes en lugar de ejecutarlos a través de una ecuación matemática, como ocurre en el caso de la regresión logística, o de una regla de división como en el caso de los

árboles. Por lo tanto, los modelos KNN pueden ser computacionalmente costosos para grandes conjuntos de datos. (Hajizadeh, Aghagolzadeh & Ezoji, 2022)

Además, debe tener en cuenta el valor correcto de k . Un valor de 1 podría dar lugar a predicciones menos sólidas frente al ruido o los valores atípicos. Los valores más grandes de k producirán predicciones más estables debido a la votación por mayoría. Pero eventualmente, un valor muy grande de k hará predicciones menos precisas ya que se vuelve difícil capturar el comportamiento complejo. (Hajizadeh, Aghagolzadeh & Ezoji, 2022)

Existen diversos métodos para encontrar el valor más apropiado según un conjunto de datos en particular. Sin embargo, de forma similar a cómo se realizó con árboles se usan cantidades estándar de vecinos y las distancias: euclídea y cúbica, al ser de las más comunes y la distancia del coseno considerando el comportamiento sinusoidal de los datos.

En general, el modelo de clasificación KNN se encuentra entre los más fáciles de comprender e interpretar y puede ser bastante preciso. (Ali, Neagu & Trundle, 2019)

La principal desventaja de KNN es que se vuelve significativamente más lento a medida que aumenta el volumen de datos. Esto puede convertirlo en una opción poco práctica en entornos en los que las predicciones deben realizarse rápidamente o en los que existen limitaciones de memoria estrictas, ya que todos los datos deben estar disponibles al realizar una predicción. (Ali, Neagu & Trundle, 2019)

Otro algoritmo con características muy particular son las Máquinas de Soporte Vectorial o SVM (por sus siglas en inglés). El objetivo mediante esta clase de algoritmo es determinar una línea que de forma más adecuada divida los datos en dos, asignando una clase a cada rango en la que serán clasificadas las nuevas observaciones.

Para determinar esa línea, el algoritmo procura maximizar la distancia entre ambas clases y por tanto crea un buffer para cada clase. El espacio que se crea entre ambas clases se denomina margen. (Cristianini & Ricci, 2008)

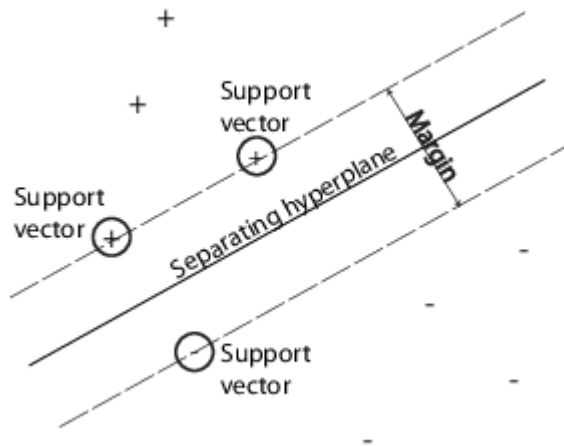
Para encontrar la línea que maximiza el margen, el algoritmo SVM primero encuentra los puntos más cercanos a la línea de ambas clases. Estos puntos se denominan vectores de soporte. (Cristianini & Ricci, 2008)

Por lo tanto, el algoritmo SVM intenta encontrar un límite de decisión de tal manera que la separación entre las dos clases sea lo más amplia posible. En este caso bidimensional, ese límite de decisión corresponde a una línea. Pero este límite se conoce generalmente como un hiperplano, que es aplicable en dimensiones superiores. (Cristianini & Ricci, 2008)

En resumen, una máquina de vectores de soporte es un clasificador que encuentra un hiperplano óptimo que maximiza el margen entre dos clases. En ejemplos reales, normalmente es imposible encontrar un hiperplano que separe perfectamente las dos clases. Un punto dentro del margen, pero clasificado correctamente, se denomina error de margen. Un punto en el lado equivocado del límite de separación es un error de clasificación.

El error total es la suma del error de margen y el error de clasificación como puede apreciarse en la Figura 35.

Figura 35. Ejemplo de árbol de clasificación



Fuente: Cristianini & Ricci (2008)

Considerando a Cristianini & Shawe-Taylor (2000), los datos para entrenamiento corresponden a un set de puntos (vectores) x_j sobre sus categorías y_j . Para una dimensión d , $x_j \in R^d$, y $y_j = \pm 1$. La ecuación del hiperplano viene dada por: $f(x) = x' \beta + b = 0$. Donde cada valor de la ecuación son números reales.

El siguiente problema define el mejor hiperplano de separación (es decir, el límite de decisión). Encuentre β y b que minimicen $\|\beta\|$ tal que para todos los puntos de datos (x_j, y_j) : $y_j f(x) \geq 1$.

Los vectores de soporte están en x_j en la frontera, donde $y_j f(x) = 1$. Sin embargo, algunos problemas de clasificación binaria no tienen un hiperplano simple como criterio de separación útil. Para esos problemas, existe una variante del enfoque matemático que conserva casi toda la simplicidad de un hiperplano de separación SVM. Este enfoque utiliza estos resultados de la teoría de la reproducción de kernel:

Según Kecman, Huang & Vogt (2005), existe una clase de funciones $G(x_1, x_2)$ con la siguiente propiedad. Hay un espacio lineal S y una función φ mapeando x a S tal que: $G(x_1, x_2) = \langle \varphi(x_1), \varphi(x_2) \rangle$.

El producto punto toma lugar en el espacio S . La clase de funciones más comunes y empleadas es la polinomial, en la que para un entero positivo p , se tiene $G(x_1, x_2) = (1 + x_1'x_2)^p$.

El enfoque matemático que utiliza kernels en SVM se basa en el método computacional de los hiperplanos. Todos los cálculos para la clasificación de hiperplanos no usan más que productos escalares. Por lo tanto, los kernels no lineales pueden usar cálculos y algoritmos de solución idénticos y obtener clasificadores que no son lineales. Los clasificadores resultantes son hipersuperficies en algún espacio S , pero el espacio S no tiene que ser identificado o examinado.

Sin embargo, cuando los datos no pueden ser separados por una línea recta o un hiperplano. Se puede usar un método kernel, que proyecte los datos en una dimensión adicional.

En lugar de una línea de decisión, ahora hay una superficie de decisión que separa los puntos. Este concepto puede generalizarse a dimensiones superiores mediante un método kernel, que asigne los datos en un espacio dimensional superior, donde los datos son linealmente separables. (Zhang, 2010)

Lineal es el más común, pero otras opciones incluyen polinomial, función de base radial y particularmente gaussiana. Cada una de estas funciones tiene sus propias características y su propia expresión.

El método kernel es una verdadera fortaleza de SVM, ya que permite manejar datos no lineales de manera eficiente. Sin embargo, para que esto sea posible, es necesario elegir la función que mejor se ajuste para evitar incrementar los tiempos de manera drástica.

En consideración se hará uso de un kernel lineal, uno cuadrático y uno cubico de manera similar a los métodos previos.

Por último, en caso de que no se desee asumir un kernel con una forma predeterminada, existen métodos kernel que asignan datos en un espacio de baja dimensión a un espacio de alta dimensión, para luego ajustar un modelo lineal en el espacio de alta dimensión minimizando una función objetivo regularizada. (Rahimi & Recht, 2008)

El modelo lineal en el espacio de alta dimensión es equivalente al modelo con kernel gaussiano en el espacio de baja dimensión. Para los modelos de clasificación generalmente se sugiere SVM o los modelos de regresión logística. (Rahimi & Recht, 2008)

La idea detrás de este ajuste es que las características de un modelo que no son linealmente separables en un espacio de baja dimensión pueden ser separables en el espacio expandido de alta dimensión. Para el cálculo del hiperplano de clasificación, sólo se emplean variaciones del producto punto. Sin embargo, se puede obtener un modelo de clasificación no lineal reemplazando el producto escalar $x_1 x_2'$ con una función kernel no lineal: $G(x_1, x_2) = \langle \varphi(x_1), \varphi(x_2) \rangle$. Donde x_i es la i -ésima observación (vector fila) y $\varphi(x_i)$ es una transformación que asigna x_i a un espacio de alta dimensión (este es el “truco del kernel”). (Rahimi & Recht, 2008)

Sin embargo, evaluar $G(x_1, x_2)$ (matriz de Gram) para cada par de observaciones es computacionalmente costoso para un conjunto de datos grande (aquellos donde n es grande).

El esquema de expansión de características aleatorias encuentra una transformación aleatoria de tal manera que la función que sustituye el producto escalar se aproxime al kernel gaussiano. Esto es: $G(x_1, x_2) = \langle \varphi(x_1), \varphi(x_2) \rangle \approx T(X_1)T(X_2)'$, donde $T(X_1)$ mapea x en (\mathbb{R}^m) . (Rahimi & Recht, 2008)

De acuerdo con Cui, Loureiro, Krzakala, & Zdeborová (2022), el parámetro de escala en la que el kernel mapeara los datos, puede calcularse mediante un procedimiento heurístico que usa submuestreo, a partir de definir un parámetro de penalización adicional conocido como “la restricción de caja” (C).

Con este parámetro se controla que la sanción máxima impuesta a las observaciones que violen el margen en el caso de que se haya usado un algoritmo de SVM, y ayuda a prevenir el sobreajuste (regularización), en el caso de las regresiones logarítmicas. Sin embargo, aumentar la restricción de la caja puede llevar a tiempos de entrenamiento más largos. (Cui, Loureiro, Krzakala, & Zdeborová, 2022)

Lo anterior ocurre, ya que dicho termino está relacionado por $C = \frac{1}{\lambda n}$, donde n es el número de observaciones, y (λ) la fuerza del término de regularización, el cual puede tomarse por defecto como $\frac{1}{n}$, donde n es nuevamente el número de observaciones. Y por tanto se sugiere de forma particular cuando el problema emplea grandes volúmenes de datos, donde el umbral del factor es bajo dada la cantidad de información disponible. (Rahimi & Recht, 2008)

Luego de definir los modelos, es necesario definir el método de validación para medir los niveles de precisión en cada uno.

Canbek, Sagiroglu, Temizel & Baykal (2017) describen tres métodos de validación:

- error de resustitución, previamente fue definido como la diferencia entre las variables respuesta de los datos de entrenamiento y las predicciones que hace basados en los datos de entrada. Lo que indica que una pérdida baja indica una mejor clasificación. Lo anterior viene dado por: $m_j = y'_j f(x_j)$, que corresponde a la es la puntuación para clasificar la observación j en la clase correspondiente a y_j . Los valores positivos de m_j indican una clasificación correcta y no contribuyen mucho a la pérdida promedio. Los valores negativos de m_j indican una clasificación incorrecta y contribuyen significativamente a la pérdida promedio.
- validación cruzada o *cross-validation*, es un proceso cuya premisa principal es evaluar diversos subconjuntos de datos entre sí con el fin de realizar una validación más exhaustiva y diversa de los datos. El proceso general es definido por Liu & Tang (2009) de la siguiente manera:
 1. Aleatoriamente se dividen los datos en k sets
 2. Para cada set, se reserva 1 como set de validación y el modelo se entrena empleando los demás $k - 1$ sets.
 3. Se repiten los pasos 1 y 2, unas k veces dejando por fuera en cada iteración una de las k particiones para entrenamiento
 4. Se combinan las estadísticas de generalización en cada iteración (o *fold* como se denomina en inglés)
- La cantidad de *folds* para usar en un modelo de *cross-validation*, especificado debe ser valor entero positivo mayor que 1.
- error de exclusión u *out-of-bag*, este método es particularmente útil para modelos de ensamble y para modelos multiclase, particularmente aquellos que requieren entrenamientos binomiales entre cada clase como SVM o Näive Bayes. Cada modelo es entrenado con un subconjunto seleccionado al azar, en el que se admite reemplazo, por lo que una muestra puede repetirse varias veces incluso en una misma muestra o no aparecer en ninguna.

Por último, los resultados de cada modelo se combinan mediante un mecanismo de votación. Liu & Tang (2009)

La calificación de cada método se realiza mediante una función de pérdida la cual se sugiere se elija conforme los modelos a evaluar o el tipo de datos.

Este método, no se recomienda para tamaños de muestra pequeños, y modelos con desbalanceo de clases. Si bien esto puede solucionarse mediante matrices de costos o probabilidades sesgadas, hace que sea compleja su interpretación. (Freund, 2009)

Según, Quintero-Montoya (2018). La validación cruzada es una estrategia muy difundida por su sencillez y su (aparente) universalidad. Existen muchos resultados sobre el rendimiento de la selección de modelos de los procedimientos de validación cruzada.

Arlot & Celisse, (2010) hacen énfasis en qué esto es posible debido al énfasis que se logra al distinguir declaraciones empíricas, de las declaraciones rigurosos de los resultados teóricos.

De forma adicional, emplear esta técnica favorece la selección de diversos tipos de muestras y no de secuencias, que llevarían a sobreajustes en los datos.

Por último, en algunos casos Hutter, Hoos & Leyton-Brown (2010) sugieren realizar procesos adicionales de optimización, como la calibración de hiper-parámetros, el balanceo de clases, o un filtrado adicional de características, ya que es natural luego de obtener un buen resultado buscar obtener la mayor precisión, la mejor velocidad o usar menos memoria).

Entrenamiento de modelos

Previamente, se mencionó como el uso de los mismos datos para pruebas y el entrenamiento, no protegerá el modelo contra sobreajuste en los datos. Además, que el uso de funciones complejas para procesar datos conducirá a modelos más complejos que pueden ser más difíciles de interpretar y tomar más tiempo para entrenar y hacer predicciones.

Es decir, que el modelo captura patrones que no son reales y que no volverán a aparecer de la misma forma en el futuro. Por lo que, al evaluar, nuevos datos el modelo aumentará su tasa de error.

Cuando un modelo presenta estas condiciones se dice que el modelo tiene una gran varianza. Por otro lado, sino logra capturar patrones y tendencias complejos también se pueden registrar errores de gran magnitud. Este efecto se denomina subajuste y se dice que el modelo tiene un alto sesgo. (Bashir, Montañez, Sehra, Sandoval-Segura & Lauw, 2020)

En el caso particular, las clases se encuentran bien balanceadas por lo que modelos como regresión logística o SVM no deberían tener mayores dificultades para ser ejecutados y evaluados. Lo anterior es fundamental, pues como se acaba de mencionar, si los datos tienen un desbalanceo de clase, es decir una gran mayoría de los datos son negativos, entonces un modelo que predice que casi todos los puntos de datos, si no todos, serán negativos, tendrá alta precisión e incluso alta exactitud, sin embargo, otros indicadores complementarios sufrirán. (Leevy, Khoshgoftaar, Bauder & Seliya, 2018)

Es fundamental, considerar diversas métricas al evaluar un modelo, de tal manera que se facilite la explicación y el conocimiento del comportamiento de las predicciones.

De igual forma y a la par de cómo evitar el sobreajuste, también es necesario considerar la complejidad del modelo. Según Bashir, Montañez, Sehra, Sandoval-Segura & Lauw (2020), se sugiere comenzar con un modelo simple e ir aumentando su complejidad de tal manera que se obtenga un error menor con un ajuste suficiente en los datos.

Lo anterior ocurre. Ya que hay un punto en el que a medida que aumenta la complejidad ocurre un quiebre y el error comienza a aumentar, dando como resultado peores predicciones en presencia de datos adicionales.

Precisamente contar con datos adicionales, resulta clave en la evaluación de los modelos. Previamente, se mencionó como es necesario más de un solo conjunto de datos para esta tarea.

Cuando se extraen datos adicionales, no solo se hace referencia a los datos para entrenamiento y sus posibles combinaciones, sino también a la extracción de datos para validación. Este conjunto de datos se usará para evaluar los resultados después de entrenar y validar el modelo final.

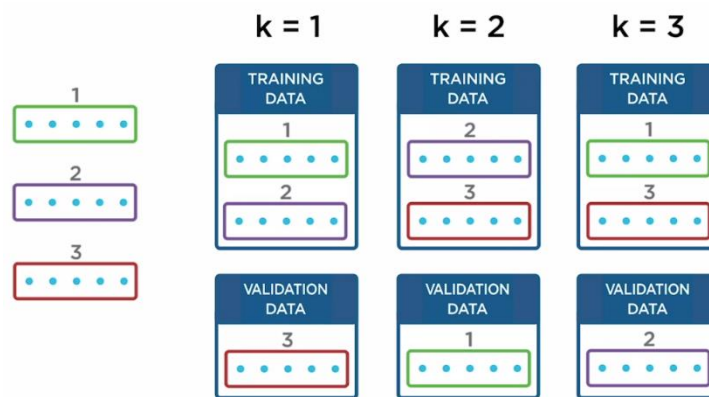
Según Hutter, Hoos & Leyton-Brown (2010), en general al momento de aplicar técnicas de aprendizaje automático se necesitan tres conjuntos diferentes de datos: el conjunto de datos de entrenamiento, (que suele ser el conjunto de datos más grande), los datos de validación y los datos de prueba, que suelen ser conjuntos mucho más pequeños, y se requieren respectivamente para determinar los niveles de clasificación del modelo entrenado y la capacidad de predicción frente a datos nuevos.

La primera división se da al momento de separar los datos de validación de los datos de entrenamiento. De forma empírica una división típica conserva el 20% de los datos para validación y el 80% restante para entrenamiento. Esta sencilla técnica se conoce como “retención” (o *hold-out* en inglés), como también el conjunto de reserva para datos de prueba. (Hoos & Leyton-Brown, 2010).

Posteriormente en el caso de los datos de entrenamiento se realiza una nueva división, en la que se aplica nuevamente la misma regla empírica, un 80% de los datos para el entrenamiento del modelo y unos 20% restantes para validación.

En el caso de los datos de entrenamiento, de manera adicional, se suele emplear una técnica conocida como validación cruzada mediante k -iteraciones (o k -folds), debe recordarse que los datos se dividen en varios subconjuntos llamados “pliegues”. Si se elige por ejemplo un $k = 3$, se tendrá algo como en la Figura 36.

Figura 36. Ejemplo de k -fold cross-validation



Siguiendo el ejemplo, cada subconjunto tendrá la misma cantidad de datos. En la primera iteración, se elige un subconjunto para la validación mientras que los otros dos se usan para el entrenamiento. Después de calcular el rendimiento del modelo en los subconjuntos restantes, se utiliza un subconjunto diferente como datos de validación en la siguiente iteración. Los conjuntos de entrenamiento también cambian en consecuencia. Este proceso se repite hasta que todos los subconjuntos se hayan utilizado como datos de validación una vez. Finalmente, se calcula el rendimiento medio de todos los pliegues. (Liu & Tang, 2009)

Según Liu & Tang (2009) mediante o k -folds, se pueden brindar resultados más precisos, especialmente en conjuntos de datos pequeños. En consideración para la propuesta, se considera un conjunto de datos de validación seleccionado mediante técnicas de retención y selecciones dirigidas. Mientras que los modelos serán entrenados mediante un o k -folds cross-validation de 10 iteraciones.

Si bien, para algunos modelos los tiempos de procesamiento, se verán drásticamente incrementados, es una forma de asegurar que el modelo está en capacidad de generalizar sin importar el tipo de práctica que será evaluada.

Según Bejani & Ghatee (2021), los modelos simples suelen ser más fáciles de interpretar y más rápidos de entrenar. Por lo general, también hacen predicciones más rápido y requieren menos recursos computacionales. Sin embargo, los modelos simples pueden no capturar patrones y tendencias más complejos en los datos.

Es precisamente, a este tipo de falla a la que se conoce como sobreajuste (o *underfitting* en inglés).

Bejani & Ghatee (2021), precisan que los modelos más complejos, como las regresiones con polinomios de orden superior, los árboles de decisión o los modelos SVM con kernel no lineales, pueden capturar estas tendencias, pero es posible que no sean capaces de generalizar bien en presencia de datos nuevos. Este tipo de falla se conoce como sobreajuste (u *overfitting* en inglés).

Es precisamente por lo anterior, que contar con datos de validación es importante. Ya que permiten guiar el proceso de modelado de tal manera que se logre un nivel adecuado de complejidad, sin que se sacrifique capacidad de predicción en presencia de datos nuevos.

Evaluación de modelos

Previamente, se mencionó la importancia de no sólo considerar la precisión del modelo, sino también indicadores y métricas adicionales que den cuenta de la calidad general de las predicciones que realiza.

Partiendo de la precisión del modelo, ésta métrica se define con base en la capacidad de que el modelo clasifique una muestra o un individuo en la categoría en la que realmente pertenece.

En el caso binario, se tiene una frontera de decisión que separa los datos en dos clases: una puede ser denominada como positiva ('No Manipulación') y la otra como negativa ('Manipulación').

Se predice que los puntos de datos en un lado del límite de decisión serán positivos y los puntos de datos en el otro lado se predice que serán negativos. (Yacoub, 2020)

Si un punto considerado positivo es clasificado como tal se define como "Verdadero Positivo" o TP (*True Positive*, por sus siglas en inglés). Por su parte, un punto negativo clasificado como positivo se define como "*False Positive*" o FP (por sus siglas en inglés).

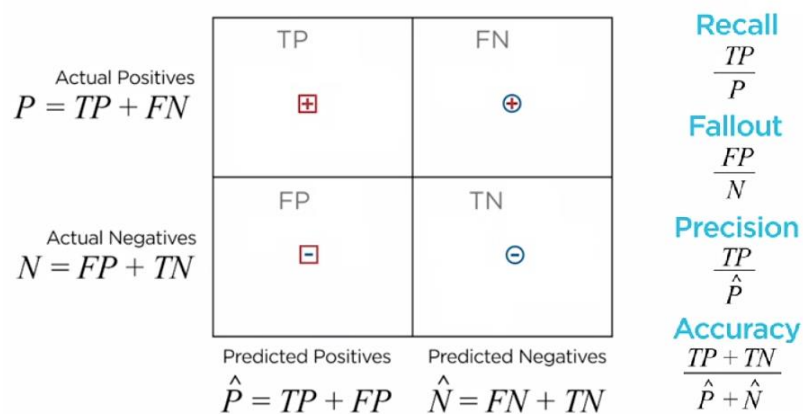
De forma análoga, un punto positivo calificado como negativo, se define como “Falso Negativo” o FN (*False Negative*, por sus siglas en inglés). Mientras que un punto negativo clasificado como tal, se conoce como “Verdadero Negativo” o TN.

Según Yacouby (2020), una forma común de evaluar estas cantidades de manera conjunta es mediante una matriz de confusión:

- Los valores reales se agrupan por fila y los valores pronosticados se agrupan por columna.
- Cada cuadrante de la matriz de confusión representa el número correspondiente de verdaderos positivos, verdaderos negativos, falsos positivos y falsos negativos. (Estas cuatro cantidades forman la base de todos los métodos de evaluación de clasificación usados para clasificadores binarios).
- El total de positivos reales es igual a la suma de verdaderos positivos y falsos negativos. De manera similar, el total de negativos reales es igual a los falsos positivos más los verdaderos negativos.
- El total de positivos previstos es igual a los verdaderos positivos mas los falsos positivos. El total de negativos predichos es igual a los falsos negativos mas los verdaderos negativos.
- A partir de las variables anteriores se calcula la precisión y los indicadores adicionales: exactitud (o *accuracy* en inglés), exhaustividad/sensibilidad (o *recall* en inglés) y especificidad/decaimiento (o *fallout* en inglés).

En la Figura 37, puede apreciarse el detalle de la matriz de confusión y las relaciones para el cálculo de los indicadores.

Figura 37. Matriz de confusión y estadísticos de desempeño



Fuente: (Yacouby, 2020) (adapt.)

Antes de considerar las compensaciones entre ellos, conviene primero definirlos:

- *Recall*: dice qué parte de una clase dada está correctamente identificada por un modelo. Aquí, esa es la relación entre los datos correctamente clasificados como positivos y el total de datos positivos: $\frac{\text{true positives}}{\text{all positives}} = \frac{TP}{TP+FN} = \frac{TP}{P}$,
- *Fallout*: dice cuántas falsas alarmas genera el modelo para una clase dada. Para la clasificación binaria, esta es la relación entre los datos clasificados incorrectamente como positivos y el total de datos negativos: $\frac{\text{false negatives}}{\text{all negatives}} = \frac{FP}{TN+FP} = \frac{FP}{N}$,
- *Precision*: representa la fracción de clasificaciones de una clase dada que fueron correctas, que es la relación entre los datos positivos clasificados correctamente y el total de datos clasificados como positivos: $\frac{\text{true positives}}{\text{positives predicted}} = \frac{TP}{TP+FP} = \frac{TP}{\hat{P}}$,
- *Accuracy*: es la tasa general en la que un modelo clasifica correctamente los datos. Es la fracción de todas las predicciones correctas del total de todas las predicciones. En otras palabras, con qué frecuencia el modelo acertó en sus predicciones en general: $\frac{\text{true predictions}}{\text{total predictions}} = \frac{TP+TN}{\hat{P}+\hat{N}}$.

Yacouby (2020), precisa además que, si la clasificación es perfecta, la exactitud, la precisión y la sensibilidad serán uno y el decaimiento será cero. Sin embargo, lograr esto es un escenario ideal, por el contrario, será necesario considerar ponderar según el contexto las condiciones óptimas para la selección del mejor modelo.

En el caso particular de aplicación conviene mantener bajo el nivel de decaimiento, pues la idea es mantener la cantidad de falsas alarmas en niveles que realmente faciliten las tareas de detección y hagan escalable la solución. En otras palabras, evitar que la cantidad de alertas que surjan no sea gestionable.

Es de tener en cuenta, que siempre se debe procurar no afectar mucho la sensibilidad del modelo de tal manera que llegue a ser tan baja abriendo la posibilidad de que deje pasar muestras que debieron ser clasificadas negativas, o puntualmente en el caso de aplicación como “Manipulación”.

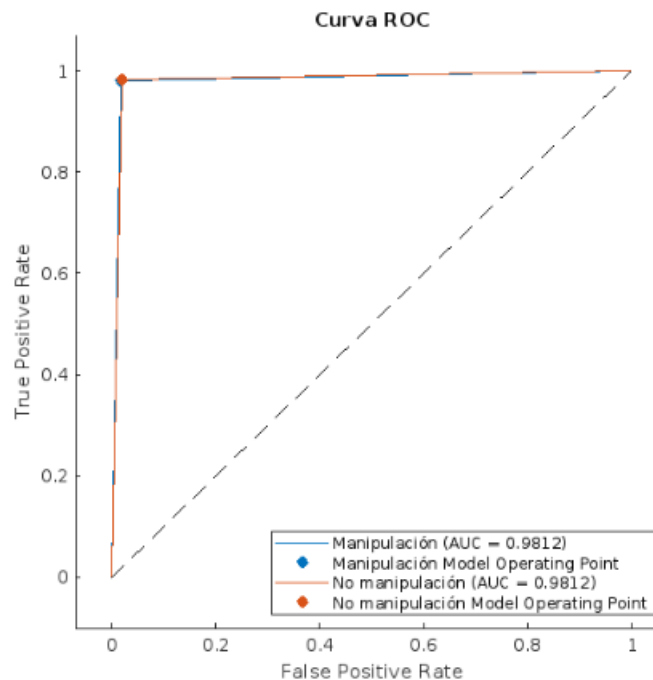
Según Yacouby (2020), en escenarios de la vida real, rara vez habrá un límite de decisión que separe perfectamente las clases positivas y negativas. Por lo tanto, es inherente a la construcción de un modelo aceptable, considerar que es más beneficioso para un modelo en términos de filtro de alertas o conversar un nivel alto *accuracy*.

Hay métodos como los propuestos por Bellmann et al. (2020) para definir los niveles de *accuracy* más adecuados para un modelo mediante técnicas de aprendizaje de máquina; diversos métodos de optimización de parámetros, o entre los más comunes por su sencillez es la construcción de la curva ROC.

Según Davis & Goadrich (2006), ROC es una sigla en inglés para Capacidad Operativa de un Receptor, y es un término que proviene de la ingeniería de radar - *donde se desarrolló y utilizó por primera vez*-. La curva ROC representa el *recall* y el *fallout* de un modelo en función de un umbral paramétrico que varía de 0 a 1, moviendo el límite de decisión a través del rango de datos.

Sobre los rangos evaluados de la función se calcula el área bajo la curva (AUC, por sus siglas en inglés) que también varía de 0 a 1, donde los mejores modelos están más cerca de uno, como puede verse en el ejemplo de la Figura 38.

Figura 38. Ejemplo de curva ROC para modelo KNN Fino ($k = 1$)



Sin embargo, este indicador es recomendado en procesos posteriores de optimización y calibración de los mejores parámetros para el modelo más prometedor.

En la Tabla 2, puede apreciarse el detalle de los modelos evaluados y sus niveles de precisión y decaimiento expresados por clase de manera desagregada en FNR (*False Negative Rate*).

Tabla 2. Desempeño de los modelos de clasificación

| Tipo de Modelo | (Entrenamiento) | (Validación) | (Entrenamiento) | | | |
|------------------------------------|-----------------|--------------|-----------------|---------|---------|---------|
| | % Precisión | % Precisión | TPR (0) | FNR (0) | TPR (1) | FNR (1) |
| Árbol Fino | 97.41 | 97.58 | 96.60 | 3.40 | 98.20 | 1.80 |
| Árbol Medio | 94.84 | 95.64 | 90.00 | 10.00 | 99.20 | 0.80 |
| Árbol Ancho | 92.09 | 92.45 | 84.00 | 16.00 | 99.40 | 0.60 |
| Regresión Log Binaria | 92.26 | 92.12 | 90.60 | 9.40 | 93.80 | 6.20 |
| Näive Bayes Gaussiano | 86.91 | 87.24 | 74.00 | 26.00 | 98.60 | 1.40 |
| Näive Bayes con Kernel Gaussiano | 79.00 | 81.26 | 57.00 | 43.00 | 98.90 | 1.10 |
| SVM con Kernel Lineal | 92.61 | 92.92 | 87.70 | 12.30 | 97.10 | 2.90 |
| SVM con Kernel Cuadrático | 88.20 | 95.09 | 96.20 | 3.80 | 81.00 | 19.00 |
| SVM con Kernel Cúbico | 58.92 | 49.65 | 79.80 | 20.20 | 40.10 | 59.90 |
| KNN, k=1, dist=euclidea | 97.78 | 98.13 | 97.00 | 3.00 | 98.50 | 1.50 |
| KNN, k=10, dist=euclidea | 95.69 | 96.66 | 92.40 | 7.60 | 98.70 | 1.30 |
| KNN, k=100, dist=euclidea | 89.45 | 90.43 | 78.70 | 21.30 | 99.20 | 0.80 |
| KNN, k=10, dist=coseno | 95.48 | 96.44 | 91.80 | 8.20 | 98.80 | 1.20 |
| KNN, k=10, dist=cúbica | 94.77 | 95.78 | 90.50 | 9.50 | 98.70 | 1.30 |
| Kernel mediante SVM | 91.81 | 91.75 | 90.20 | 9.80 | 93.30 | 6.70 |
| Kernel mediante regresión log. | 89.97 | 89.11 | 90.10 | 9.90 | 89.80 | 10.20 |
| KNN para selección de param. | 98.96 | 99.23 | 98.70 | 1.30 | 99.20 | 0.80 |
| KNN k=1, dist=manhattan, caract=9 | 90.69 | 91.09 | 89.50 | 10.50 | 91.80 | 8.20 |
| KNN k=1, dist=manhattan, caract=16 | 99.41 | 99.49 | 99.30 | 0.70 | 99.50 | 0.50 |
| KNN k=1, dist=euclidea, caract=16 | 99.49 | 99.71 | 99.40 | 0.55 | 99.50 | 0.01 |

Nota. Para los datos de entrenamiento se tuvo un total de 10911 observaciones y se realizó validación cruzada mediante $k = 10$.

Para los datos de validación se tuvo un total de 2727 observaciones correspondientes al 20% de las observaciones del conjunto de datos.

Las clases respuesta fueron asignada de la siguiente manera: 0 corresponde a 'Manipulación', y 1 corresponde a 'No manipulación'.

Los primeros 16 modelos emplearon 40 predictores, los últimos 4 modelos corresponden a las validaciones de optimización, siendo el último modelo la versión óptima definitiva.

Considerando que los modelos evaluados presentaron un buen *accuracy* en general, se elegirá el modelo cuyo *fallout* sea bajo, y que en general bajo los supuestos que maneja sea un clasificador adecuado para el problema que se está tratando.

Por lo anterior, se elige el modelo KNN con $k = 1$ y distancia euclídea como el mejor modelo, ya que sus niveles de *accuracy* (98.13%) y *fallout* (2.25%) son los más adecuados.

Además, categorizar un nuevo punto en función de su distancia a los puntos de un conjunto de datos de entrenamiento puede ser una forma simple, pero efectiva de clasificar nuevos puntos.

Y en el caso de aplicación es pertinente, ya que, a partir del entendimiento de la distribución de la energía, su intensidad, las frecuencias detectadas y los efectos sobre todo el espectro, se pudo comprobar que hay marcadas diferencias entre las muestras de una acción que ha sido manipulada de una que no.

Son muchas las métricas que pueden ser empleadas para determinar la distancia entre un par de puntos, sin embargo, estos están atados a la forma en la que se distribuyen los datos y al tipo de problema que se está tratando. (Quintero, 2018)

Por lo que si la distancia euclídea siendo de las métricas de distancia más básica, logra explicar y diferenciar dos puntos es un parámetro suficiente para el caso de aplicación en cuestión.

Por otra parte, es importante tener en cuenta que la selección de los modelos se hizo considerando cuarenta características, sin embargo, según Gonzalez & Miikkulainen (2019), el uso de un conjunto de datos más pequeño tiene múltiples beneficios como: requisitos de memoria más bajos, tiempos de entrenamiento más rápidos y mejor interpretabilidad.

Por tanto, mediante métodos de regularización, se considerará la cantidad mínima de características que garanticen que el modelo no disminuirá sus niveles de *accuracy*.

Los métodos de regularización parten de la premisa de restringir los parámetros del modelo para evitar el sobreajuste. Por ejemplo, en el caso de los árboles de decisión, la forma más sencilla de regularizar es cortando ramas al reducir la profundidad del árbol (como pasar de un árbol fino a un árbol ancho), por ejemplo.

Guyon & Elisseeff (2003), sugieren que la función de selección de parámetros debe elegirse según el problema y el tipo de datos de que describen los predictores.

De forma adicional, sugieren algunos métodos según el tipo de algoritmo. En este caso al haber elegido KNN como el mejor modelo debido a su alto *accuracy* y su bajo *fallout*, se recomienda el uso de un algoritmo de ranqueo para evaluar la importancia de los predictores y se sugiere el algoritmo "ReliefF". (Robnik-Sikonja & Kononenko, 2003)

El algoritmo penaliza los predictores que dan valores diferentes a los vecinos de la misma clase y recompensa los predictores que dan valores diferentes a los vecinos de diferentes clases.

De acuerdo con Robnik-Sikonja & Kononenko (2003), ReliefF primero establece los pesos de todos los predictores W_j en 0. Luego, el selecciona iterativamente una observación aleatoria x_r ; encuentra las k observaciones más cercanas a x_r para cada clase y para cada vecino más cercano x_q ; actualiza todos los pesos para los predictores F_j de la siguiente manera:

- Si x_r y x_q son de la misma clase: $W_j^i = W_j^{i-1} - \frac{\Delta_j(x_r, x_q)}{m} \cdot d_{rq}$.
- Si x_r y x_q son de diferente clase: $W_j^i = W_j^{i-1} + \frac{p_{y_q}}{1-p_{y_r}} \cdot \frac{\Delta_j(x_r, x_q)}{m} \cdot d_{rq}$.

Donde, W_j^i es el peso del predictor F_j en la i -ésima iteración. p_{y_r} , es la probabilidad posterior de la clase a la que x_r pertenece, y p_{y_q} es la probabilidad posterior de la clase a la que x_q pertenece. m es el número de iteraciones definidas. $\Delta_j(x_r, x_q)$ es la diferencia entre el valor del predictor F_j entre las observaciones x_r y x_q . Sea x_{rj} el valor del j -ésimo predictor para la observación x_r , y sea x_{qj} el valor del j -ésimo predictor para la observación x_q . Sea $\Delta_j(x_r, x_q)$ para el caso de variables continuas (como las del problema de aplicación): $\Delta_j(x_r, x_q) = \frac{|x_{rj} - x_{qj}|}{\max(F_j) - \min(F_j)}$. Y sea d_{rq} una

función de distancia de la forma: $d_{rq} = \frac{\tilde{d}_{rq}}{\sum_{i=1}^k \tilde{d}_{rq}}$ La distancia es objeto de escalamiento: $\tilde{d}_{rq} = e^{-(rank(r,q)/sigma)^2}$, donde $rank(r, q)$ es la posición de la q -ésima observación entre los vecinos más cercanos de la r -ésima observación, ordenadas por distancia.

Donde generalmente se asigna un valor de sigma tan pequeño, que esencialmente todos los vecinos tengan la misma influencia, es decir se elija con base en los k vecinos definidos previamente en la función.

Como parte de los procesos de optimización y calibración del modelo final, según Saito & Yasui (2019), es necesario encontrar valores apropiados de los hiperparámetros para estimar los parámetros del modelo.

Precisamente, un buen ejemplo es k , en el modelo elegido. Previamente los valores establecidos se hicieron con base en valores por defecto sugerido, ya que al ser un hiperparámetro, no es un valor que genere el modelo de clasificación automáticamente.

La distancia como tal, también puede ser catalogada como un hiper-parámetro y también puede ser incluido en el proceso de optimización o ajuste de hiper-parámetros.

Según Li et al. (2018), conceptualmente, el ajuste de hiper-parámetros es solo un ciclo de optimización sobre el entrenamiento del modelo para buscar el conjunto de hiper-parámetros que conducen al error más bajo.

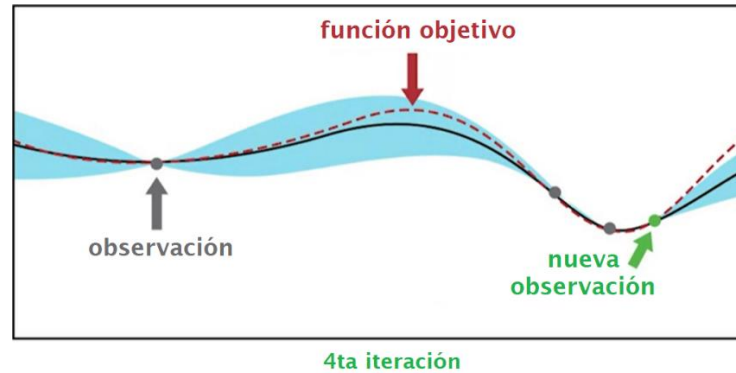
Sin embargo, hacer uso de los datos de validación para este proceso, puede ser tedioso, ya que debe realizarse de manera manual, además de convertirse en un proceso de ensayo y error. (Li et al., 2018)

Un método eficiente para elegir la combinación más adecuada para seleccionar los mejores parámetros es a partir del desempeño de combinaciones pasadas. Un método que cumple este objetivo es la optimización bayesiana.

Según Kim & Choi (2019), con este enfoque, el algoritmo realiza un seguimiento de los resultados de evaluaciones previas, que utiliza para formar una representación probabilística del rendimiento del modelo. Lo anterior se logra, empleando una función objetivo como principal evaluador de combinaciones de hiper-parámetros y una aproximación de esta función objetivo llamada sustituto como puede verse en la Figura 39.

El algoritmo de optimización bayesiano intenta minimizar una función objetivo escalar $f(x)$ para x en un dominio acotado. La función puede ser determinista o estocástica, lo que significa que puede devolver resultados diferentes cuando se evalúa en el mismo punto x . Los componentes de x pueden ser reales continuos, enteros o categóricos, es decir, un conjunto discreto de nombres. (Snoek, Larochelle, & Adams, 2012)

Figura 39. Ejemplo del proceso de optimización de Bayes



Fuente: Kim & Choi, 2019 (adapt.)

Según Snoek, Larochelle, & Adams (2012), los elementos clave en la minimización son:

- Un modelo de proceso gaussiano de $f(x)$.
- Un procedimiento de actualización bayesiano para modificar el modelo de proceso gaussiano en cada nueva evaluación de $f(x)$.
- Una función de adquisición $a(x)$ (un kernel basado en el modelo de proceso gaussiano de $f(x)$ que maximiza para determinar el siguiente punto x para la evaluación).

Las funciones de adquisición evalúan la “bondad” de un punto x en base a la función de distribución posterior Q . Cuando existen restricciones acopladas, incluida la restricción Error, las funciones de adquisición modifican su estimación de “bondad” multiplicando por una estimación de la probabilidad de que se satisfagan las restricciones, para llegar a la función de adquisición.

Para el caso puntual, se tiene en cuenta la modificación de la bondad de ajuste, a partir de la estimación de que se está sobreexplotando un área de validación.

Según Bull (2011), para entender la sobreexplotación, sea $\sigma^F(x)$ la desviación estándar de la función objetivo posterior en x . Sea σ la desviación estándar posterior del ruido aditivo, tal que: $\sigma_Q^2(x) = \sigma_F^2(x) + \sigma^2$. Defina $t\sigma$ como el parámetro que controla el equilibrio entre exploración y explotación, cuando se requiere un balance medio se sugiere que su valor sea 0.5. Luego de cada iteración, se debe evaluar si el siguiente punto x satisface: $\sigma_F(x) < t\sigma$.

Si es así, el algoritmo declara que x está sobreexplotando. Entonces la función de adquisición modifica su función kernel multiplicando θ por el número de iteraciones. Esta modificación eleva la varianza σ_Q para puntos entre observaciones.

Luego genera un nuevo punto basado en la nueva función del kernel ajustada. Si el nuevo punto x vuelve a estar sobreexplotado, la función de adquisición multiplica θ por un factor adicional de 10 y vuelve a intentarlo. Se continúa así hasta cinco veces, intentando generar un punto x que no esté sobreexplotado. El algoritmo acepta la nueva x como el siguiente punto, y se detendrá una vez aplique su criterio de parada.

Bull (2011) sugiere considerar al menos treinta iteraciones como una regla empírica en el que se valide una cantidad considerable de veces los mejores clasificadores y no se caiga en un óptimo local.

Procurando precisamente evitar caer en mínimos locales, y definir un criterio de búsqueda más amplio es que se plantea el procedimiento previo.

Luego de aplicarlo, se define que el tamaño óptimo para k es 1, como puede verse en la Figura 39, donde los puntos contienen etiquetas de los mejores modelos identificados hasta el momento. No debe ser tratarse como si fuera un resultado específico de cada iteración. (Chou & Li, 2004)

El algoritmo selecciona los hiper-parámetros que producen el menor error, generalmente suele coincidir con el menor error de clasificación, pero no siempre es el caso, como se puede ver en la Figura 40.

Figura 40. Gráfico de error de clasificación



Fuente: Chou & Li, 2004 (adapt.)

Por otra parte, según Robnik-Sikonja & Kononenko (2003), cuando $k = 1$ no se recomienda aplicar el algoritmo “ReliefF”, debido a que las estimaciones no pueden ser confiables debido a ruido en los datos. Sugieren iniciar con un valor $k = 10$, e investigar la estabilidad y confianza de los ránkines y los pesos que algoritmo arroja conforme se evalúan diversos valores de k .

Considerando, que se acaba de definir que el valor más adecuado para k es 1, se valida la estabilidad de las variables elegidas entre $k = 1$ y $k = 3$.

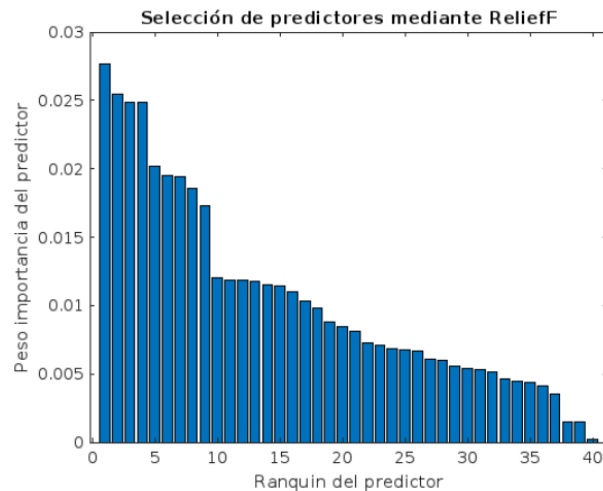
La selección final de los campos elegido puede visualizarse en la Tabla 3 y la Figura 40. Para el detalle de todos los campos ver Anexo D.

Tabla 3. Selección de las características en modelo KNN mediante el algoritmo “ReliefF”

| Predictor | Ranquin | Peso |
|-----------------------|---------|--------|
| 'mov_mean_range_coef' | 1 | 0.0274 |
| 'mov_mean_min_coef' | 2 | 0.0251 |
| 'min_coef' | 3 | 0.0246 |
| 'mode_coef' | 4 | 0.0246 |
| 'mov_mean_std_coef' | 5 | 0.0200 |
| 'mov_mean_max_coef' | 6 | 0.0192 |
| 'max_coef' | 7 | 0.0191 |
| 'rms' | 8 | 0.0184 |
| 'spectral_slope' | 9 | 0.0172 |
| 'frec_0_0093111' | 10 | 0.0117 |
| 'frec_0_0086876' | 11 | 0.0116 |
| 'frec_0_0081058' | 12 | 0.0115 |
| 'median_min_coef' | 13 | 0.0113 |
| 'frec_0_007563' | 14 | 0.0113 |
| 'frec_0_0065839' | 15 | 0.0110 |
| 'frec_0_0070565' | 16 | 0.0107 |

Nota. Los resultados mediante los 3 valores de k coincidieron en los valores de tabla como las 9 variables más relevantes del predictor. El valor del peso corresponde al promedio de los 3 valores de k para cada predictor (sin embargo, fue muy semejante en cada caso)

Figura 41. Selección de predictores mediante algoritmo ReliefF



Es de tener en cuenta que haciendo menos complejo el modelo al elegir las variables más representativas (sólo nueve predictores) se retira un peso de importancia del 50% que aporta el resto de predictores del set de datos. (Ver Tabla 2, Modelo: KNN k=1, dist=manhattan, caract=9)

Por lo que conviene, considerar la regla empírica del codo y hacer el corte en la selección de características justo en el punto en el que el aporte de las variables se estabiliza.

En este caso, dicha estabilización ocurre para pesos menores a 0.01. De esta manera se captura hasta un 70% del peso de importancia de los predictores.

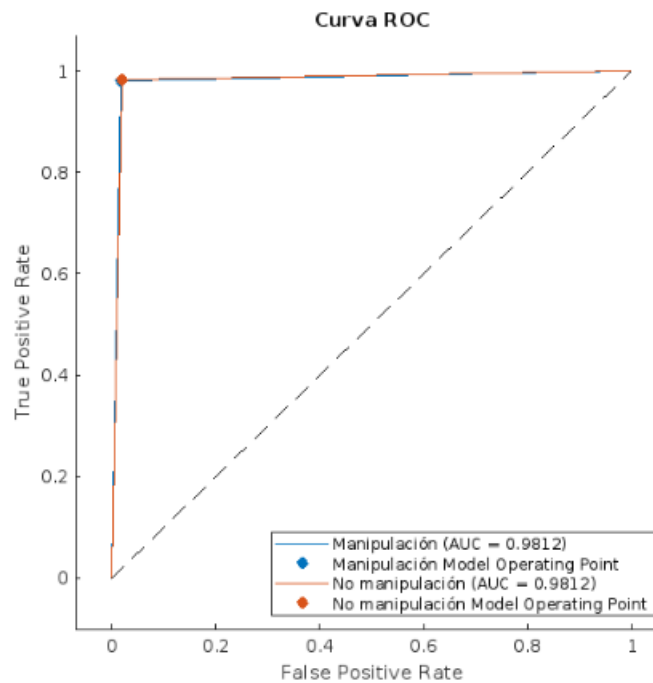
Según Robnik-Sikonja & Kononenko (2003), al seleccionar sólo aquellas características con ranquin más alto, se reduce el sesgo en las métricas de validación, como pudo apreciarse en los resultados. Se evidenció de igual manera un leve sobreajuste que fue solucionado al emplear menos variables.

Por su parte, los métodos de optimización bayesiana son eficientes porque eligen los siguientes valores de hiper-parámetro de manera informada y realizan menos evaluaciones de funciones. (Kim & Choi, 2019)

Además, la velocidad de predicción es mucho más rápida. El tamaño exportado del modelo es significativamente más pequeño y el error de clasificación mejoró muy levemente como puede verse en las últimas filas de la Tabla 2.

De igual forma el valor ROC mejoró respecto a la referencia de la Figura 38, ya que como puede apreciarse en la Figura 42 los valores son aún más cercanos a 1. Lo que indica que es un mejor modelo.

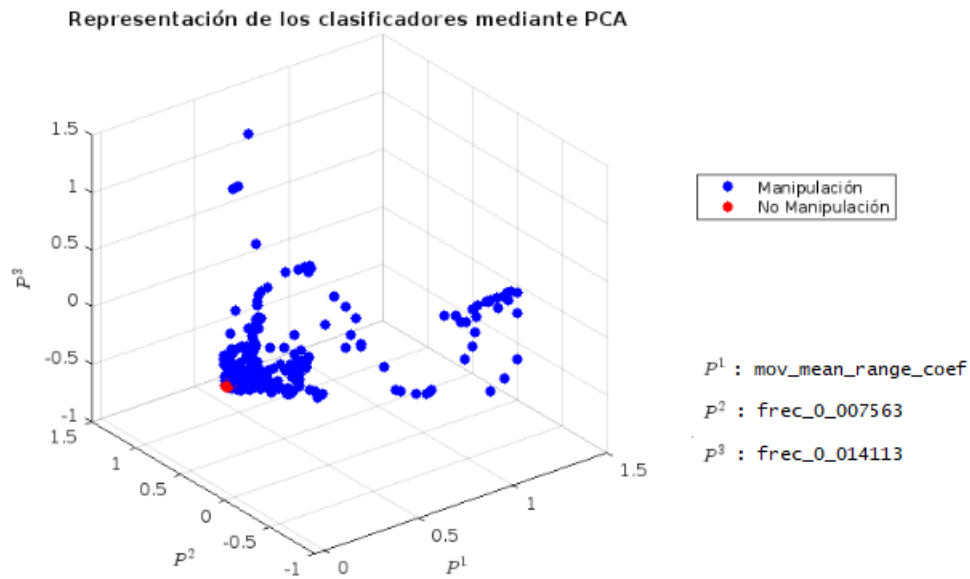
Figura 42. Curva ROC para modelo KNN Optimizado



Teniendo en cuenta que con solo catorce de las características inicialmente propuestas se logra un mejor desempeño, el problema se sigue considerando de alta dimensión, por lo que la visualización de los resultados de clasificación en sí misma no es una tarea trivial (Ariza-Jiménez, Villa & Quintero-Montoya, 2019).

Sin embargo, previamente se presentó como el análisis de componentes principales ofrece una alternativa de visualización viable, por lo que, aplicando nuevamente este proceso sobre los predictores finales elegidos, es posible apreciar en la Figura 43 las principales características que describen el fenómeno y la clasificación de las muestras de validación.

Figura 43. Representación de los clasificadores más relevantes para el modelo KNN Optimizado



Consideraciones Finales

Según Mallat (2006), las señales transportan cantidades abrumadoras de datos en los que la información relevante suele ser difícil de encontrar, sin embargo, al realizar tareas de procesamiento de la señal, es posible representar la señal mediante coeficientes más dispersos que de forma más sencilla están en capacidad de revelar la información buscada.

Mediante la Transformada Wavelet Continúa (CWT) usando una wavelet Morse, es posible revelar la información que se buscaba en una señal que permite dar cuenta si un activo financiero fue objeto de manipulación o no, ya que con solo evaluar las relaciones entre las frecuencias detectadas es viable generar una diferenciación considerando para ello niveles de energía.

El espectrograma ofrece una vista muy particular del fenómeno, y como se mencionó previamente por sí sola constituye una herramienta de visualización interesante (Addison, 2018).

Sin embargo, empleando los coeficientes resultantes del procesamiento de la señal mediante esta transformada fue posible de forma generalizada: adaptar, derivar y seleccionar características adicionales que permitieran describir y profundizar de mejor manera las relaciones en términos de energía en las frecuencias detectadas.

Lo anterior es posible según Mallat (2006), ya que las señales transitorias resultan en una buena fuente de información respecto a los estados considerados estables de un sistema, de hecho, para localizar con mayor precisión estos picos, recomienda emplear los coeficientes de menor escala (frecuencia más alta), pues son estos los que facilitan la detección de cambios transitorios en los datos como previamente fue mencionado.

De hecho, se detectó que en presencia de anomalías no se registran un segmento de bajas frecuencias que en los activos sin manipulación si se presentan, ya que al ser más estables concentran energía en bajo valor en otros rangos.

Sin embargo, como previamente se mencionó se tendría una gran cantidad de información vacía por lo que para los procesos de aprendizaje dichas frecuencias no son consideradas relevantes.

De hecho, según Zhang et. al (2022), no existen reglas satisfactorias para descubrir una función de base wavelet óptima para la transformada wavelet, que dé cuenta de manera general la diferencia en ambos fenómenos, por lo que contar con el subconjunto elegido es suficiente.

Considerando lo anterior, es importante destacar que finalmente como se evidencia en la Figura 43 el problema puede definirse como un problema de detección de energía, ya que es evidente la cantidad de energía que se registra en una señal financiera en presencia de manipulaciones.

Si bien desde el escalograma en muchos casos son visibles las caídas, no se logra identificar un patrón, además a simple vista en algunos casos no es sencillo determinar si una acción pudo ser manipulada de otra que no.

Por tanto, considerar técnicas de aprendizaje automático, especialmente técnicas de agrupamiento es relevante, no solo por la escalabilidad que brinda a los procesos de detección, sino que ofrece herramientas y alternativas que permiten separar de manera informada los registros de energía que caracterizan una señal que ha sido manipulada de una que no.

Por otra parte, la caída en el espectro de potencia podría ser empleada como filtro de frecuencias antes o después de aplicar la transformada wavelet, ya que es viable ajustar aún más los parámetros que permiten generalizar el fenómeno, como también generar características adicionales en caso de requerirse en presencia de eventos en los que el modelo llegase a disminuir sustancialmente sus niveles de precisión y decaimiento.

Para ventanas de volatilidad pequeñas se requiere mayor cantidad de muestras para lograr una buena resolución del efecto de los cambios aleatorios generales de cualquier señal financiera de aquellos que se puedan asociar con manipulaciones y cambios abruptos en la señal. Mientras que períodos más amplios logran atenuar dicho efecto y mejorar la resolución de singularidades tanto en volatilidad como en cantidad de muestras previas necesarias.

Lo anterior es necesario, ya que dichos ajustes son útiles a la hora de generar y definir las características que facilitan las tareas de clasificación.

Finalmente, un buen preprocesamiento (que considere los parámetros más adecuados) para efectuar el análisis de la señal lleva a un modelo simple que permite explicar el fenómeno que se pretende comprender, en este caso si una acción fue manipulada o no.

CONCLUSIONES

Como respuesta a los objetivos de la presente propuesta, respecto a la elección de un modelo que estuviera en capacidad de indicar si una acción fue manipulada o no, fueron elegidos los modelos de aprendizaje automático.

Dicha elección partió de la necesidad de hacer escalable las tareas de detección y contar con herramientas que por sí mismas estén en capacidad de indicar si una acción fue manipulada o no.

Si bien dicho objetivo se cumplió, el uso de este tipo de herramientas debe ser desmitificado, ya que el uso de dichas técnicas requiere un minucioso análisis matemático y estadístico que debe partir del entendimiento de los datos. Para el caso puntual comprender primero las señales financieras y su comportamiento, especialmente en presencia de manipulaciones.

Al respecto, las técnicas de aprendizaje automático dependen de que la información que será empleada en los procesos de entrenamiento sea la adecuada, ya que de lo contrario los resultados no serán los mejores y sería inviable la justificación de los resultados arrojados por el modelo elegido.

Partir de la definición de un modelo simple, que sea fácil de explicar permite que se obtenga un conocimiento más amplio de lo que ocurre a una acción en presencia de una manipulación. De esta manera se evita caer en una visión utilitarista de los modelos de aprendizaje, ya que generalmente no contribuyen a crear conocimiento del sistema.

Por ejemplo, comprender que los componentes de baja frecuencia son claves en la determinación de si una acción fue manipulada o no, es evidencia de ello.

Lo anterior fue fundamental en el entendimiento del problema, ya que es precisamente en estas frecuencias donde principalmente se evidencian los efectos producidos por prácticas manipulativas en una señal financiera.

Al respecto, se pudo evidenciar como las manipulaciones sobre los activos financieros afectan sus regímenes de volatilidad y provocan picos sobre la señal.

En este caso fue necesario emplear una transformada que estuviera en capacidad de detectar picos transitorios, y por lo anterior se eligió la Transformada Wavelet Continua como mejor transformada para representar los componentes de frecuencia de una señal financiera.

Es de destacar que la CWT permite que los parámetros de escala y fase cambien continuamente, lo que lo hace invariante en eventos transitorios y redundante en la traducción.

Según Addison (2018), la redundancia garantiza que la CWT pueda proporcionar resultados de alta resolución, mucho más fáciles de interpretar que los obtenidos mediante otras técnicas como la DWT.

Lo anterior, no solo conlleva a considerar la Transformada Wavelet Continúa como mejor herramienta de visualización, sino que se tuvo en cuenta también su capacidad para detectar eventos transitorios en una señal.

Los eventos transitorios se evidencian generalmente como picos en la señal, y lo anterior es relevante, pues dichos picos son buena fuente de información de un sistema, ya que pueden dar cuenta si dicho sistema cambió de estado temporalmente de manera abrupta.

En el caso puntual, esto pudo ser evidenciados, ya que generalmente los cambios abruptos en el régimen de volatilidad esperado para un activo en particular se evidenciaban como picos y coincidían cada vez con el inicio del esquema definido.

Considerando lo anterior, se evidenció mediante el uso de las transformadas que a partir de picos y cambios abruptos en los regímenes de volatilidad se puede dar cuenta de que una acción puede estar siendo objeto de manipulación.

Adicionalmente, se pudo identificar y validar que dichos picos se registran como eventos de alta energía, por lo que según Mallat (2006) los métodos basados en la transformada de Fourier son inadecuados para representar este tipo de señales, ya que el evento transitorio difícilmente puede aproximarse mediante senos y cosenos.

Y por tal razón, más allá de las capacidades de representación de la señal, la CWT fue elegida por su capacidad de detectar artefactos útiles en la detección de prácticas manipulativas, (en este caso los eventos transitorios identificados en las señales).

Lo anterior, refuerza la idea de que la clave en la implementación de soluciones de aprendizaje automático es observar y comprender los datos, (para este caso puntual, entender las señales y su comportamiento), de tal manera que el conocimiento experto se nutra de las posibilidades que ofrecen las técnicas de aprendizaje automático.

Por otra parte, comprender que mediante las técnicas de espectrografía es posible la generación de espacios y características fue crucial en la elección de los modelos de aprendizaje como mejor alternativa para hacer más escalable la solución, ya que es posible conocer los componentes subyacentes de una señal y de esta manera dar solución al problema planteado.

Según Addison (2018) independientemente de si forma parte del algoritmo final, la CWT puede ayudar a comprender las características de la señal, proporcionar información sobre la viabilidad de la tarea subyacente y facilitar el desarrollo de estrategias de procesamiento de señales para atacar el problema a resolver.

Por lo que usar la CWT solo como “herramienta de procesamiento visual” es un enfoque necesario.

Sin embargo, no se descarta el uso de las demás transformadas, ya que mediante su uso es posible obtener información adicional e incluso la generación de características adicionales para ser usadas en los modelos de aprendizaje automático.

Es precisamente, esta clase de conocimientos los que se obtienen cuando se desmitifica el uso de los algoritmos de aprendizaje automático y se asumen como una herramienta que facilita la generación de nuevo conocimiento del fenómeno.

Según Mallat (2006) lo anterior es una estrategia de supervivencia necesaria, ya que adaptar diversas representaciones y derivar operadores resulta clave en la clasificación de una señal. La concentración de frecuencias, la complejidad de la señal (entendida como una medida de entropía) y el flujo tiempo-escala, son ejemplos de tales operadores.

En resumen, con el proceso llevado a cabo se demostró que, para abordar un problema no trivial, es necesario reflexionar sobre la información disponible y las respuestas que se desean dar. Resulta clave analizar, procesar y entender el problema, aplicando un conocimiento previo y posteriormente un conocimiento experto para de esta manera lograr una solución más inteligente.

Es importante tener en cuenta, que además de la elección de la transformada y la definición de operadores, la elección de la Wavelet y sus parámetros tiene un fuerte efecto sobre la detección de los elementos previamente mencionados y la escalabilidad de la solución.

Resulta fundamental estandarizar tanto las entradas y las salidas, para con el fin de tener criterios comparables para elegir la mejor alternativa disponible. Adicionalmente, de esta manera se puede lograr un nivel de generalización útil en la detección de anomalías en señales financieras.

Lo anterior, facilitó el uso de modelos de aprendizaje automático. Al respecto, según Cristianini & Shawe-Taylor (2000), la selección de kernels adecuados y parámetros de kernel asociados es una tarea importante cuando se utilizan métodos de kernel para regresión, clasificación y reducción de dimensionalidad.

Existen diversas posibilidades para elegir la función kernel, como también diversos momentos en los que este tipo de técnicas resulta útil y necesario. Entre los mayores beneficios que ofrecen para este tipo de problemas, está la generación de espacios de características más profundos y a la vez la implementación de modelos más simples, al aprender directamente de los datos un mapeo no lineal de dos espacios funcionales. (Tripura & Chakraborty, 2022)

Facilitando las tareas de clasificación, al adaptarse con relativa sencillez a las dinámicas complejas de los datos. Sin embargo, en algunos casos hace compleja la generalización del problema.

Si bien, este no fue el caso de la presente propuesta, es un criterio importante que considerar, especialmente cuando de forma explícita se puso de manifiesto la necesidad de contar con un modelo que generalizará la detección de prácticas manipulativas, lo cual fue posible con un alto porcentaje de precisión mediante un modelo KNN con $k = 1$ y distancia euclídea.

Por otra parte, la necesidad de contar con modelos más simples presenta beneficios adicionales, ya que es viable estar en producción rápidamente debido a las grandes ventajas previamente expuestas en términos de tiempos de ejecución y recursos.

Adicionalmente, es viable extrapolar, la solución a otro tipo de señales aleatorias que requieran una clasificación binaria, debido a que los kernel wavelet no son lineales y, por tanto, en segmentos diferenciables de tiempo es capaz de dividir los datos en altas y bajas frecuencias sin importar el tipo de señal.

Sin embargo, como se acaba de mencionar partir del conocimiento del problema es fundamental no solo para facilitar el ciclo de aprendizaje, sino también para que haya apropiación del conocimiento que puede obtenerse al analizar y comprender los datos del fenómeno que se desea modelar.

Finalmente, la intención de considerar las señales financieras como series de tiempo y emplear diversas clases de kernel para la detección de prácticas manipulativas surgió del deseo de brindar un vistazo de las oportunidades interdisciplinarias y generar ideas para otras vías de análisis de problemas del día a día.

REFERENCIAS

- Abbas, B., Belatreche, A., & Bouridane, A. (2018). Stock price manipulation detection using empirical mode decomposition-based Kernel density estimation clustering method. *Proceedings of SAI Intelligent Systems Conference* (pp. 851-866). Springer, Cham.
- Addison, P. S. (2018). Introduction to redundancy rules: the continuous wavelet transform comes of age. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, 376(2126), 2017-025-8.
- Aitken, M. J., Harris, F. H., & Ji, S. (2009). Trade-based manipulation and market efficiency: A cross-market comparison. *22nd Australasian Finance and Banking Conference* (p. 18).
- Aksenov, A., Grebenschikova, E., & Fayzrakhmanov, R. (2020). Front-running Model in the Stock Market. *2020 2nd International Conference on Control Systems, Mathematical Modeling, Automation and Energy Efficiency (SUMMA)* (pp. 681-686). IEEE.
- Ali, N., Neagu, D., & Trundle, P. (2019). Evaluation of k-nearest neighbour classifier performance for heterogeneous data sets. *SN Applied Sciences*, 1, Article 1559. <https://doi.org/10.1007/s42452-019-1356-9>
- Allen, F., & Gale, D. (1992). Stock-price manipulation. *The Review of Financial Studies*, 5(3), 503-529.
- AMV (2022). *Reglamento de AMV*. Autorregulador del Mercado de Valores (de Colombia) <https://amvcolombia.org.co/wp-content/uploads/2022/04/2.-Reglamento-AMV-Asesoría-002.pdf>
- Arce, G. R. (2005). *Nonlinear signal processing: a statistical approach*. John Wiley & Sons.
- Ariza-Jiménez, L., Villa, L. F., & Quintero-Montoya, O. L. (2019). Memberships Networks for High-Dimensional Fuzzy Clustering Visualization. *Workshop on Engineering Applications* (pp. 263-273). Springer, Cham.
- Arlot, S., & Celisse, A. (2010). A survey of cross-validation procedures for model selection. *Statistics Surveys*, 4, 40-79. <https://doi.org/10.1214/09-SS054>
- Arthur, D., & Vassilvitskii, S. (2007). K-means++ the advantages of careful seeding. *Proceedings of the eighteenth annual ACM-SIAM symposium on Discrete algorithms* (pp. 1027-1035).

- Arts, L., & van den Broek, E. (2022). The fast continuous wavelet transformation (fCWT) for real-time, high-quality, noise-resistant time–frequency analysis. *Nature Computational Science*, 2(1), 47-58.
- Bashir, D., Montañez, G. D., Sehra, S., Sandoval Segura, P., & Lauw, J. (2020). An Information-Theoretic Perspective on Overfitting and Underfitting. *AI 2020: Advances in Artificial Intelligence* (pp. 347-358).
- Bejani, M. M., & Ghatee, M. (2021). A systematic review on overfitting control shallow and deep neural networks. *Artificial Intelligence Review*, 54(12), 6391-6438
- Bellmann, P., Hihn, H., Braun, D. A., & Schwenker, F. (2020). Binary Classification: Counterbalancing Class Imbalance by Applying Regression Models in Combination with One-Sided Label Shifts. arXiv preprint arXiv:2011.14764
- Boashash, B. (2015). Time-frequency signal analysis and processing: a comprehensive reference. *Academic press*.
- Bracewell, R. N. (1986). *The Fourier transform and its applications* (Vol. 31999, pp. 267-272). New York: McGraw-Hill.
- Breiman, L., J. H. Friedman, R. A. Olshen, and C. J. Stone. Classification and Regression Trees. Boca Raton, FL: *Chapman & Hall*, 1984.
- Bull, A. D. (2011) Convergence rates of efficient global optimization algorithms. arXiv preprint <https://arxiv.org/abs/1101.3501v3>
- Byrne, C. L. (2014). *Signal Processing: a mathematical approach*. CRC Press.
- Cao, Y., Li, Y., Coleman, S., Belatreche, A., & McGinnity, T. M. (2015). Detecting wash trade in financial market using digraphs and dynamic programming. *IEEE transactions on neural networks and learning systems*, 27(11), 2351-2363.
- Campo, D., Quintero-Montoya, O. L., & Bastidas, M. (2016, April). Multiresolution analysis (discrete wavelet transform) through Daubechies family for emotion recognition in speech. *Journal of Physics: Conference Series* (Vol. 705, No. 1, p. 012034). IOP Publishing.
- Canbek, G., Sagiroglu, S., Temizel, T. T., & Baykal, N. (2017, October). Binary classification performance measures/metrics: A comprehensive visualized roadmap to gain new insights. *2017 International Conference on Computer Science and Engineering (UBMK)* (pp. 821-826). IEEE.

- CFI (2015). Poop and Scoop. Corporate Finance Institute. <https://corporatefinanceinstitute.com/resources/knowledge/trading-investing/poop-and-scoop/>
- Chou, W., & Li, L. (2004). A minimum classification error (MCE) framework for generalized linear classifier in machine learning for text categorization/retrieval. *2004 International Conference on Machine Learning and Applications, 2004. Proceedings.* (pp. 26-33). IEEE.
- Cristianini, N., & Shawe-Taylor, J. (2000). *An Introduction to Support Vector Machines and Other Kernel-Based Learning Methods*. Cambridge, UK: Cambridge University Press
- Cristianini, N., & Ricci, E. (2008). Support vector machines¹. *M.-Y. Kao (Ed.), Encyclopedia of algorithms* (pp. 928-932). Springer. https://doi.org/10.1007/978-0-387-30162-4_415
- Cui, H., Loureiro, B., Krzakala, F., & Zdeborová, L. (2022). Error rates for kernel classification under source and capacity conditions. arXiv preprint arXiv:2201.12655.
- Davis, J., and M. Goadrich. (2006). The Relationship Between Precision-Recall and ROC Curves. *Proceedings of ICML '06*, pp. 233–240.
- De Vitis, A. C. (2019). *Kernel methods for high dimensional data analysis* (Doctoral dissertation, Université Côte d'Azur).
- Daubechies, I. (1990). The wavelet transform, time-frequency localization and signal analysis. *IEEE Transactions on Information Theory*, 36(5), 961–1005. <https://doi.org/10.1109/18.57199>
- Decreto 1802 de 2007. Por el cual se dictan disposiciones sobre operaciones preacordadas. 24 de mayo de 2007. D.O. No. 46638
- Doshi-Velez, F., & Kim, B. (2017). Towards a rigorous science of interpretable machine learning. arXiv preprint arXiv:1702.08608.
- Ergün, H. O., Yalaman, A., Manahov, V., & Zhang, H. (2021). Stock market manipulation in an emerging market of Turkey: how do market participants select stocks for manipulation?. *Applied Economics Letters*, 28(5), 354-358.

- Ferdousi, Z., & Maeda, A. (2006, April). Unsupervised outlier detection in time series data. *22nd International Conference on Data Engineering Workshops (ICDEW'06)* (pp. x121-x121). IEEE.
- Ferri, F. J., Pudil, P., Hatf, M., & Kittler, J. (1994). Comparative study of techniques for large-scale feature selection. *Machine intelligence and pattern recognition* (Vol. 16, pp. 403-413). North-Holland.
- Freund, Y. (2009). A more robust boosting algorithm.
arXiv:0905.2138v1
- Fulop, S. A., & Fitz, K. (2006). Algorithms for computing the time-corrected instantaneous frequency (reassigned) spectrogram, with applications. *The Journal of the Acoustical Society of America*, 119(1), 360-371.
- Gallegati, M., & Semmler, W. (Eds.). (2014). *Wavelet applications in economics and finance (Vol. 20)*. Springer International Publishing.
<https://doi.org/10.1007/978-3-319-07061-2>
- Ghane, P., & Braga-Neto, U. (2021). Generalized resubstitution for classification error estimation.
arXiv preprint arXiv:2110.12285.
- Greenwood, P.E.; Nikulin, M.S. (1996). A guide to chi-squared testing. New York: Wiley. ISBN 0-471-55779-X.
- Golmohammadi, K., & Zaiane, O. R. (2012, August). Data mining applications for fraud detection in securities market. *2012 European Intelligence and Security Informatics Conference* (pp. 107-114). IEEE.
- Golmohammadi, K., & Zaiane, O. R. (2015). Time series contextual anomaly detection for detecting market manipulation in stock market. *2015 IEEE International Conference on Data Science and Advanced Analytics (DSAA)* (pp. 1-10). IEEE.
- Gomez, A., Quintero-Montoya, O. L., López, N., Castro, J., Villa, L., & Mejía, G. (2017). An approach to emotion recognition in single-channel EEG signals using stationary wavelet transform. *VII Latin American Congress on Biomedical Engineering CLAIB 2016*, Bucaramanga, Santander, Colombia, October 26th-28th, 2016 (pp. 654-657). Springer, Singapore.
- Gonzalez, S., & Miikkulainen, R. (2019). Improved Training Speed, Accuracy, and Data Utilization Through Loss Function Optimization. *2020 IEEE Congress on Evolutionary Computation (CEC)*, 1-8.
<https://doi.org/10.1109/CEC48606.2020.9185777>.

- Guyon, I., & Elisseeff, A. (2003). An introduction to variable and feature selection. *Journal of machine learning research*, 3(Mar), 1157-1182.
- Hajizadeh, R., Aghagolzadeh, A., & Ezoji, M. (2022). Mutual neighborhood and modified majority voting based KNN classifier for multi-categories classification. *Pattern Analysis and Applications*, 25, 773-793.
<https://doi.org/10.1007/s10044-022-01069-0>
- Hamilton, J. D., & Susmel, R. (1994). Autoregressive conditional heteroskedasticity and changes in regime. *Journal of econometrics*, 64(1-2), 307-333.
- Harris, F. J. (1978). On the use of windows for harmonic analysis with the discrete Fourier transform. *Proceedings of the IEEE*, 66(1), 51-83.
- Hatemi-j, A. (2014). Asymmetric generalized impulse responses with an application in finance. *Economic Modelling*, 36, 18-22.
- Herwartz, H., & Maxand, S. (2020). Nonparametric tests for independence: a review and comparative simulation study with an application to malnutrition data in India. *Statistical papers*, 61, 2175-2201.
- Hofmann, T., Schölkopf, B., & Smola, A. J. (2008). Kernel methods in machine learning. *The annals of statistics*, 36(3), 1171-1220.
- Hogg, R. V., and J. Ledolter. *Engineering Statistics*. MacMillan Publishing Company, 1987.
- Hopf, K., & Reifenrath, S. (2021). Filter Methods for Feature Selection in Supervised Machine Learning Applications-Review and Benchmark.
 arXiv preprint arXiv:2111.12140.
- Hutter, F., Hoos, H. H., & Leyton-Brown, K. (2010). Tradeoffs in the empirical evaluation of competing algorithm designs. *Annals of Mathematics and Artificial Intelligence*, 60(1-2), 65-89.
- Jorgensen, P. E., Kang, S., Song, M. S., & Tian, F. (2019). Dimension reduction and kernel principal component analysis.
 arXiv preprint arXiv:1906.06451.
- Kecman, V., Huang, T. M., & Vogt, M. (2005). Iterative single data algorithm for training kernel machines from huge data sets: Theory and performance. *Support vector machines: Theory and Applications*, 255-274.

- Khanna, N., & Sonti, R. (2004). Value creating stock manipulation: feedback effect of stock prices on firm value. *Journal of financial markets*, 7(3), 237-270.
- Kim, J., & Choi, S. (2019). On Local Optimizers of Acquisition Functions in Bayesian Optimization. , 675-690.
https://doi.org/10.1007/978-3-030-67661-2_40.
- Kissell, R. L. (2020). *Algorithmic trading methods: Applications using advanced statistics, optimization, and machine learning techniques*. Academic Press.
- Kleinbaum, D. G., & Klein, M. (2010). *Logistic regression: A self-learning text*. Springer.
<https://doi.org/10.1007/978-1-4419-1742-3>
- Landau, S., Leese, M., Stahl, D., & Everitt, B. S. (2011). *Cluster analysis*. John Wiley & Sons.
- Leangarun, T., Tangamchit, P., & Thajchayapong, S. (2018). Stock price manipulation detection using generative adversarial networks. *2018 IEEE Symposium Series on Computational Intelligence (SSCI)* (pp. 2104-2111). IEEE.
- Li, A., Wu, J., & Liu, Z. (2017). Market manipulation detection based on classification methods. *Procedia Computer Science*, 122, 788-795.
- Li, L., Jamieson, K., Rostamizadeh, A., Gonina, E., Hardt, M., Recht, B., & Talwalkar, A. (2018). *Massively Parallel Hyperparameter Tuning*. ArXiv, abs/1810.05934.
- Lilly, J. M., & Olhede, S. C. (2008). Higher-order properties of analytic wavelets. *IEEE Transactions on Signal Processing*, 57(1), 146-160.
- Lilly, J. M., & Olhede, S. C. (2012). Generalized Morse wavelets as a superfamily of analytic wavelets. *IEEE Transactions on Signal Processing*, 60(11), 6036-6041.
- Lilly, J. M. (2017). Element analysis: a wavelet-based method for analysing time-localized events in noisy time series. *Proceedings of the Royal Society A: Mathematical, Physical and Engineering Sciences*, 473(2200), 20160776.
- Lin, T. C. (2016). *The new market manipulation*. Emory LJ, 66, 1253.
- Liu, H., Refaeilzadeh, P., & Tang, L. (2009). Cross-validation. In L. Liu & M. T. Özsu (Eds.), *Encyclopedia of database systems* (pp. 532-538). Springer US.
https://doi.org/10.1007/978-0-387-39940-9_565

- Loh, W.Y. (2002). Regression Trees with Unbiased Variable Selection and Interaction Detection. *Statistica Sinica*, Vol. 12, pp. 361–386.
- Mallat, S. (2009) *A Wavelet Tour of Signal Processing (Third Edition)*. Academic Press. ISBN 9780123743701
<https://doi.org/10.1016/B978-0-12-374370-1.00004-5>.
- Manning, C. D., Raghavan P., & Schütze., M. (2008) *Introduction to Information Retrieval*. NY: Cambridge University Press
- Markham, J. W. (2015). *Law enforcement and the history of financial market manipulation*. Routledge.
- McConway, K. (2003). Generalized linear models with applications in engineering and the sciences. *Journal of the Royal Statistical Society Series D: The Statistician*, 52(2), 245-246.
<https://doi.org/10.1111/1467-9884.t01-4-00356>
- Müller, U. A., Dacorogna, M. M., Davé, R. D., Olsen, R. B., Pictet, O. V., & Von Weizsäcker, J. E. (1997). Volatilities of different time resolutions—analyzing the dynamics of market components. *Journal of Empirical Finance*, 4(2-3), 213-239.
- Nagelkerke, N. J. (1991). A note on a general definition of the coefficient of determination. *biometrika*, 78(3), 691-692.
- Nelson, D. B. (1991). Conditional heteroskedasticity in asset returns: A new approach. *Econometrica: Journal of the econometric society*, 347-370.
- Nembrini, S., König, I. R., & Wright, M. N. (2018). The revival of the Gini importance? *Bioinformatics*, 34(21), 3711-3718.
<https://doi.org/10.1093/bioinformatics/bty373>
- Ng, A., Jordan, M., & Weiss, Y. (2001). On spectral clustering: Analysis and an algorithm. *Advances in neural information processing systems*, 14.
- Ngai, E. W., Hu, Y., Wong, Y. H., Chen, Y., & Sun, X. (2011). The application of data mining techniques in financial fraud detection: A classification framework and an academic review of literature. *Decision support systems*, 50(3), 559-569.
- Olhede, S. C., & Walden, A. T. (2002). Generalized morse wavelets. *IEEE Transactions on Signal Processing*, 50(11), 2661-2670.
- Oregi, I., Del Ser, J., Perez, A., & Lozano, J. A. (2018). Adversarial sample crafting for time series classification with elastic similarity measures. *International*

- Symposium on Intelligent and Distributed Computing* (pp. 26-39). Springer, Cham.
- Oppenheim, Alan V., Schafer, Ronald W., Buck, John R. (1999). "7.2". *Discrete-time signal processing (2nd ed.)*. Upper Saddle River, N.J.: Prentice Hall. pp. 465–478.
- Ögüt, H., Doganay, M. M., & Aktas, R. (2009). Detecting stock-price manipulation in an emerging market: The case of Turkey. *Expert Systems with Applications*, 36(9), 11944-11949.
- Paniagua-Castrillón, J. G., & Quintero-Montoya, O. L. (2018). Application of the continuous wavelet transform in the extraction of directional data on RTM imaging condition wavefields. *CT&F-Ciencia, Tecnología y Futuro*, 8(2), 65-73.
- Paniagua-Castrillón, J. G., Quintero-Montoya, O. L. & Sierra-Sosa, D. (2018). Laguerre-Gauss filters in reverse time migration image reconstruction. *Brazilian Journal of Geophysics*, 35(2), 81-93.
- Paniagua-Castrillón, J. G., Miranda, E. S., Garcés, J. J., & Quintero-Montoya, O. L. (2021). Laguerre Gauss kernel for COVID-19 medical decision making from chest Tomography. *International Journal of Artificial Intelligence*, 193-205.
- Percival, D. B., & Walden, A. T. (2000). *Wavelet methods for time series analysis (Vol. 4)*. Cambridge university press.
- Pudjihartono, N., Fadason, T., Kempa-Liehr, A. W., & O'Sullivan, J. M. (2022). A Review of Feature Selection Methods for Machine Learning-Based Disease Risk Prediction. *Frontiers in Bioinformatics*, 2, 927312.
<https://doi.org/10.3389/fbinf.2022.927312>
- Qu, H., & Zhang, Y. (2016). A new kernel of support vector regression for forecasting high-frequency stock returns. *Mathematical Problems in Engineering*, 2016.
- Quintero-Montoya, O. L. (2018). *Machine Intelligence for Human Decision Making*. Editorial Artes y Letras.
- Quintero-Montoya, O. L. (2019). The possibilities of artificial intelligence development. *Tecnologías disruptivas del proceso de globalización*, 196.
- Quirk, T.J. (2011). *One-Way Analysis of Variance (ANOVA)*.
https://doi.org/10.1007/978-1-4614-3725-3_8
- Rahimi, A., & Recht, B. (2007). Random features for large-scale kernel machines. *Advances in neural information processing systems*, 20.

- Raykov, Y. P., Boukouvalas, A., Baig, F., & Little, M. A. (2016). What to do when K-means clustering fails: a simple yet principled alternative algorithm. *PLoS one*, 11(9), e0162259.
- Rizvi, B., Belatreche, A., Bouridane, A., & Mistry, K. (2020a). Stock Price Manipulation Detection based on Autoencoder Learning of Stock Trades Affinity. *2020 International Joint Conference on Neural Networks (IJCNN)* (pp. 1-8). IEEE.
- Rizvi, B., Belatreche, A., Bouridane, A., & Watson, I. (2020b). Detection of Stock Price Manipulation Using Kernel Based Principal Component Analysis and Multivariate Density Estimation. *IEEE Access*, 8, 135989-136003.
- Robnik-Sikonja, M., and I. Kononenko. (2003). Theoretical and empirical analysis of ReliefF and RReliefF. *Machine Learning*, 53, 23–69.
- Saito, Y., & Yasui, S. (2019). Counterfactual Cross-Validation: Effective Causal Model Selection from Observational Data. ArXiv, abs/1909.05299.
- Sejdic, E., Djurovic, I., & Stankovic, L. (2008). Quantitative performance analysis of scalogram as instantaneous frequency estimator. *IEEE Transactions on Signal Processing*, 56(8), 3837-3845.
- SEC (2006, agosto 07). Pump and Dump Schemes. U.S. Securities and Exchange Commission.
https://www.sec.gov/rss/your_money/pump_and_dump.htm
- SEC (2015, septiembre 04). Defining Spoofing. U.S. Securities and Exchange Commission.
<https://www.sec.gov/comments/sr-bats-2015-57/bats201557-3.pdf>
- Shi, J., & Malik, J. (2000). Normalized cuts and image segmentation. *IEEE Transactions on pattern analysis and machine intelligence*, 22(8), 888-905.
- Shreffler, J., & Huecker, M. R. (2023). Hypothesis testing, P values, confidence intervals, and significance. *StatPearls [Internet]*. StatPearls Publishing.
- Sierra-Sosa, D., Bastidas, M., & Quintero-Montoya, O. L. (2016). Double fourier analysis for emotion identification in voiced speech. *Journal of Physics: Conference Series* (Vol. 705, No. 1, p. 012035). IOP Publishing.
- Smith, J. O. (2007). *Transient and steady-state signals. Mathematics of the discrete Fourier transform (DFT) with audio applications (2nd ed.)*. W3K Publishing.
https://ccrma.stanford.edu/~jos/fp/Transient_Steady_State_Signals.html

- Snoek, J., Larochelle, H., & Adams, R. P. (2012). Practical Bayesian Optimization of Machine Learning Algorithms. ArXiv, abs/1206.2944.
- Song, Y., Cao, L., Wu, X., Wei, G., Ye, W., & Ding, W. (2012). Coupled behavior analysis for capturing coupling relationships in group-based market manipulations. *Proceedings of the 18th ACM SIGKDD international conference on Knowledge discovery and data mining* (pp. 976-984).
- Sridhar, S., Mootha, S., & Subramanian, S. (2020). Detection of Market Manipulation using Ensemble Neural Networks. *2020 International Conference on Intelligent Systems and Computer Vision (ISCV)* (pp. 1-8). IEEE.
- STOCK Act, 7 U.S.C. § 6c (1978). <https://www.law.cornell.edu/uscode/text/7/6c>
- Stoica, P., & Moses, R. L. (2005). *Spectral analysis of signals* (Vol. 452, pp. 21-26). Upper Saddle River, NJ: Pearson Prentice Hall.
- Tabachnick, B. G., & Fidell, L. S. (2013). *Using Multivariate Statistics, 6th ed.* Boston: Allyn and Bacon.
- Taranenko, Y., & Rizun, N. (2022). Wavelet filtering of signals without using model functions. *Radioelectronics and Communications Systems*, 65(2), 96-109.
- The MathWorks, Inc. (n.d.). Continuous and Discrete Wavelet Transforms. MathWorks. <https://www.mathworks.com/help/wavelet/gs/continuous-and-discrete-wavelet-transforms.html>
- Torrence, C., & Compo, G. P. (1998). A practical guide to wavelet analysis. *Bulletin of the American Meteorological society*, 79(1), 61-78.
- Tripura, T., & Chakraborty, S. (2022). Wavelet neural operator: a neural operator for parametric partial differential equations. arXiv preprint arXiv:2205.02191.
- United States v. Navinder Singh Sarao. The United States Department of Justice. (2022, April 30). <https://www.justice.gov/criminal-vns/united-states-v-navinder-singh-sarao>
- Varanis, M., Silva, A. L., Balthazar, J. M., & Pederiva, R. (2021). A tutorial review on time-frequency analysis of non-stationary vibration signals with nonlinear dynamics applications. *Brazilian Journal of Physics*, 51(6), 859–877. <https://doi.org/10.1007/s13538-020-00842-y>

- Von Luxburg, U. (2007). A tutorial on spectral clustering. *Statistics and computing*, 17, 395-416.
- Vințe, C., & Ausloos, M. (2022). The Cross-Sectional Intrinsic Entropy—A Comprehensive Stock Market Volatility Estimator. *Entropy*, 24(5), 623.
- Wang, Y., Yang, H. Y., & He, P. (2022). Continuous Wavelet Analysis of Matter Clustering Using the Gaussian-derived Wavelet. *The Astrophysical Journal*, 934(1), 77.
- Wang, Q., Xu, W., Huang, X., & Yang, K. (2019). Enhancing intraday stock price manipulation detection by leveraging recurrent neural networks with ensemble learning. *Neurocomputing*, 347, 46-58.
- Webb, G. I. (2010). Naïve Bayes. C. Sammut & G. I. Webb (Eds.), *Encyclopedia of machine learning* (pp. 768-776). Springer.
https://doi.org/10.1007/978-0-387-30164-8_576
- Welch, P. (1967). The use of fast Fourier transform for the estimation of power spectra: a method based on time averaging over short, modified periodograms. *IEEE Transactions on audio and electroacoustics*, 15(2), 70-73.
- Yacouby, R. (2020). Probabilistic Extension of Precision, Recall, and F1 Score for More Thorough Evaluation of Classification Models. *Proceedings of the 1st Workshop on Evaluation and Comparison of NLP Systems* (pp. 69-78)
- Yagemann, C., Chung, P. H., Uzun, E., Ragam, S., Saltaformaggio, B., & Lee, W. (2021). Modeling Large-Scale Manipulation in Open Stock Markets. *IEEE Security & Privacy*, 19(6), 58-65.
- Zhang, C., He, Y., Yang, T., Zhang, B., & Wu, J. (2022). An analog circuit fault diagnosis approach based on improved wavelet transform and MKELM. *Circuits, Systems, and Signal Processing*, 1-32.
- Zhang, X. (2010). Kernel methods. C. Sammut & G. I. Webb (Eds.), *Encyclopedia of machine learning* (pp. 577-582). Springer.
https://doi.org/10.1007/978-0-387-30164-8_430
- Zhang, Z. (2019). Spectral and time-frequency analysis. *EEG Signal Processing and feature extraction*, 89-116.

ANEXOS

ANEXO A – FUENTES Y SCRIPTS

Los *scripts* empleados para la validación de las técnicas y la ejecución de los modelos se encuentran en el siguiente repositorio en GitHub:

https://github.com/jherre35/deteccion_practicas_manipulativas.git

Cada *script* se relaciona con cada una de las etapas de la siguiente propuesta, y un script a modo de resumen en el que se implementación la solución. En cada uno se describen los pasos y las etapas consideradas en la validación de cada etapa del proceso de análisis.

Por último, se incluye las series de tiempo de referencia y demás archivos misceláneos empleados como parte de la solución.

ANEXO B – SELECCIÓN DE CLÚSTERES

| Spectral Cluster | | | |
|------------------|-----------|--------|----------|
| acción | clústeres | índice | clusters |
| GME | 4 | CNK | 0 |
| AMC | 4 | CHWY | 2 |
| CURR | 3 | NVS | 4 |
| NOK | 3 | MSI | 5 |
| ADNC | 1 | DLB | 3 |
| SRPT | 2 | NVO | 2 |
| DRNG | 2 | VRNT | 3 |
| PLPL | 3 | NVO | 0 |
| PZOO | 3 | NVO | 2 |
| SOX | 3 | INTC | 2 |
| ECEZ | 2 | WTRG | 1 |
| VRSYF | 4 | ABT | 3 |
| BRZL | 3 | WMT | 2 |
| CPMD | 2 | GILD | 4 |
| ROYL | 5 | CVX | 3 |
| ARYC | 5 | NVO | 2 |
| LGND | 1 | NVO | 0 |
| VMRI | 0 | MSI | 0 |
| CSSE | 3 | CNK | 3 |
| HELE | 3 | CHWY | 3 |
| RARE | 3 | CVX | 2 |
| TV | 3 | CNK | 5 |
| AIG | 2 | MET | 5 |

| dbscan | | | | | |
|--------|---------|-----------|--------|---------|-----------|
| acción | épsilon | clústeres | índice | épsilon | clústeres |
| GME | 0.00025 | 2 | CNK | 0.0003 | 2 |
| AMC | 0.00025 | 2 | CHWY | 0.001 | 2 |
| CURR | 0.00025 | 2 | NVS | 0.0002 | 2 |
| NOK | 0.0005 | 2 | MSI | 0.0005 | 2 |
| ADNC | 0.001 | 2 | DLB | 0.0005 | 2 |
| SRPT | 0.0002 | 2 | NVO | 0.005 | 2 |
| DRNG | 0.0002 | 2 | VRNT | 0.002 | 2 |
| PLPL | 0.00025 | 2 | NVO | 0.001 | 2 |
| PZOO | 0.0005 | 2 | NVO | 0.0005 | 2 |
| SOX | 0.0001 | 2 | INTC | 0.0003 | 2 |
| ECEZ | 0.0001 | 2 | WTRG | 0.0002 | 2 |
| VRSYF | 0.0003 | 2 | ABT | 0.00001 | 2 |
| BRZL | 0.0003 | 2 | WMT | 0.0001 | 2 |
| CPMD | 0.0002 | 2 | GILD | 0.0003 | 2 |
| ROYL | 0.0001 | 2 | CVX | 0.0015 | 2 |
| ARYC | 0.00025 | 2 | NVO | 0.004 | 2 |
| LGND | 0.0005 | 2 | NVO | 0.0005 | 2 |
| VMRI | 0.0008 | 2 | MSI | 0.0002 | 2 |
| CSSE | 0.001 | 2 | CNK | 0.001 | 2 |
| HELE | 0.0005 | 2 | CHWY | 0.0005 | 2 |
| RARE | 0.0005 | 2 | CVX | 0.0005 | 2 |
| TV | 0.0001 | 2 | CNK | 0.0005 | 2 |
| AIG | 0.00025 | 2 | MET | 0.0005 | 2 |

ANEXO C – SELECCIÓN INICIAL DE CARACTERÍSTICAS

| SELECCIÓN DE CARACTERÍSTICAS | | Método | | | | | | |
|------------------------------|-------------|------------------|-----------|----------|----------------------|--------------------|---------------------|---------------|
| Predicador | ch2 | embedido_forward | envoltura | mixto | porcentaje_acumulada | varianza_acumulada | varianza_individual | Total general |
| frec_0_0081058 | 24 | | | 1 | 1 | 49 | 49 | 124 |
| frec_0_017375 | 24 | | | | | 49 | 49 | 122 |
| frec_0_021391 | 24 | | | | | 49 | 49 | 122 |
| mov_range_std_coef | 24 | | | | | 49 | 49 | 122 |
| frec_0_014113 | 24 | | | | | 49 | 49 | 122 |
| mov_range_max_coef | 24 | | | | | 49 | 49 | 122 |
| frec_0_028226 | 24 | | | | | 49 | 49 | 122 |
| mov_desv_max_coef | 24 | | 1 | | | 49 | 49 | 123 |
| frec_0_018622 | 24 | | | | | 49 | 49 | 122 |
| mov_desv_desv_coef | 24 | | | | | 49 | 49 | 122 |
| frec_0_015126 | 24 | | | | | 49 | 49 | 122 |
| median_min_coef | 24 | | 1 | 1 | | 49 | 49 | 124 |
| frec_0_0086876 | 24 | | | 1 | 1 | 49 | 49 | 124 |
| median_coef | 24 | | 4 | | | 49 | 49 | 126 |
| frec_0_007563 | 24 | | | | 1 | 49 | 49 | 124 |
| frec_0_0070565 | 24 | | | | 1 | 49 | 49 | 124 |
| min_coef | 23 | | | 1 | | 49 | 49 | 122 |
| mov_range_range_coef | 23 | | | | | 49 | 49 | 121 |
| mov_desv_mean_coef | 23 | | | | | 49 | 49 | 121 |
| mov_range_min_coef | 23 | | | | | 49 | 49 | 121 |
| frec_0_0093111 | 23 | | | 1 | 1 | 49 | 49 | 123 |
| median_max_coef | 23 | | 1 | 1 | | 49 | 49 | 123 |
| mov_desv_min_coef | 23 | | | | | 49 | 49 | 121 |
| mov_range_mean_coef | 23 | | 1 | | | 49 | 49 | 122 |
| spectral_slope | 23 | | | | | 49 | 49 | 121 |
| frec_0_032423 | 23 | | | | | 49 | 49 | 121 |
| frec_0_013168 | 23 | | | | | 49 | 49 | 121 |
| frec_0_030252 | 23 | | | | | 49 | 49 | 121 |
| mode_coef | 23 | | | | | 49 | 49 | 121 |
| mov_mean_min_coef | 23 | | 2 | 2 | | 49 | 49 | 125 |
| median_std_coef | 23 | | 1 | | | 49 | 49 | 122 |
| frec_0_026336 | 23 | | | 1 | | 49 | 49 | 122 |
| median_mean_coef | 22 | | | | | 49 | 49 | 120 |
| frec_0_0065839 | 22 | | 2 | 1 | | 49 | 49 | 123 |
| frec_0_012286 | 22 | | | | | 49 | 49 | 120 |
| mov_mean_max_coef | 22 | | 2 | | | 49 | 49 | 122 |
| frec_0_022927 | 22 | | | | | 49 | 49 | 120 |
| frec_0_019959 | 22 | | | | | 49 | 49 | 120 |
| max_coef | 22 | | 5 | 3 | | 49 | 49 | 128 |
| frec_0_0695 | 22 | | | | | 49 | 49 | 120 |
| rms | 22 | | 1 | 2 | | 49 | 49 | 123 |
| frec_0_045853 | 22 | | | 1 | | 49 | 49 | 121 |
| frec_0_024572 | 22 | | | | | 49 | 49 | 120 |
| frec_0_016212 | 22 | | | | | 49 | 49 | 120 |
| frec_0_042783 | 22 | | | | | 49 | 49 | 120 |
| frec_0_079835 | 22 | | | | | 49 | 49 | 120 |
| frec_0_074489 | 22 | | | | | 49 | 49 | 120 |
| frec_0_085565 | 22 | | | | | 49 | 49 | 120 |
| frec_0_011463 | 21 | | | | | 49 | 49 | 119 |
| mov_mean_std_coef | 21 | | 1 | 1 | | 49 | 49 | 121 |
| frec_0_0099794 | 21 | | | | | 49 | 49 | 119 |
| frec_0_03475 | 21 | | | | | 49 | 49 | 119 |
| mean_coef | 21 | | | | | 49 | 49 | 119 |
| std_coef | 21 | | | | | 49 | 49 | 119 |
| frec_0_010696 | 21 | | | | | 49 | 49 | 119 |
| frec_0_10534 | 21 | | | | | 49 | 49 | 119 |
| frec_0_052671 | 21 | | | | | 49 | 49 | 119 |
| frec_0_037244 | 21 | | | | | 49 | 49 | 119 |
| mov_mean_mean_coef | 21 | | | | | 49 | 49 | 119 |
| mov_mean_range_coef | 20 | | 9 | 7 | | 49 | 49 | 134 |
| frec_0_1129 | 20 | | | | | 49 | 49 | 118 |
| frec_0_049144 | 20 | | | | | 49 | 49 | 118 |
| frec_0_060504 | 20 | | | | | 49 | 49 | 118 |
| frec_0_056452 | 20 | | | | | 49 | 49 | 118 |
| frec_0_039918 | 20 | | | | | 49 | 49 | 118 |
| frec_0_064846 | 20 | | | | | 49 | 49 | 118 |
| range_coef | 20 | | 1 | | | 49 | 49 | 119 |
| frec_0_12101 | 19 | | | | | 49 | 49 | 117 |
| frec_0_091706 | 19 | | | | | 49 | 49 | 117 |
| frec_0_098288 | 19 | | | | | 49 | 49 | 117 |
| frec_0_15967 | 19 | | | | | 49 | 49 | 117 |
| desv_median_coef_frec | 19 | | 2 | 13 | | 49 | 49 | 132 |
| spectral_centroid | 19 | | | | | 49 | 49 | 117 |
| frec_0_12969 | 18 | | | | | 49 | 49 | 116 |
| frec_0_17113 | 18 | | | | | 49 | 49 | 116 |
| frec_0_18341 | 17 | | | | | 49 | 49 | 115 |
| max_frec | 17 | | | | | 49 | 49 | 115 |
| min_frec | 17 | | | | | 49 | 49 | 115 |
| frec_0_21069 | 17 | | | | | 49 | 49 | 115 |
| frec_0_19658 | 17 | | | | | 49 | 49 | 115 |
| frec_0_29795 | 16 | | | | | 49 | 49 | 114 |
| frec_0_14898 | 16 | | 1 | | | 49 | 49 | 115 |
| frec_0_25938 | 16 | | | | | 49 | 49 | 114 |
| frec_0_22581 | 16 | | | | | 49 | 49 | 114 |
| frec_0_139 | 15 | | | | | 49 | 49 | 113 |
| frec_0_278 | 14 | | | | | 49 | 49 | 112 |
| spectral_crest | 14 | | | 1 | | 49 | 49 | 113 |
| median_frec | 13 | | | | | 49 | 49 | 111 |
| spectral_spread | 13 | | | | | 49 | 49 | 111 |
| spectral_kurtosis | 13 | | | 1 | | 49 | 49 | 112 |
| spectral_flatness | 11 | | | | | 49 | 49 | 109 |
| frec_0_34226 | 11 | | | | | 49 | 49 | 109 |
| frec_0_31934 | 11 | | | | | 49 | 49 | 109 |
| spectral_skewness | 11 | | | | | 49 | 49 | 109 |
| porc_cambio_fase | 11 | | 1 | | | 49 | 49 | 110 |
| frec_0_24201 | 10 | | | | | 49 | 49 | 108 |
| frec_0_45162 | 9 | | | | | 49 | 49 | 107 |
| desv_max_coef_frec | 9 | | | | | 49 | 49 | 107 |
| spectral_entropy | 9 | | | | | 49 | 49 | 107 |
| range_frec | 8 | | | | | 49 | 49 | 106 |
| frec_0_36683 | 8 | | | | | 49 | 49 | 106 |
| frec_0_42137 | 7 | | | | | 49 | 49 | 105 |
| desv_min_coef_frec | 6 | | | 1 | | 49 | 49 | 105 |
| spectral_decrease | 6 | | | | | 49 | 49 | 104 |
| frec_0_39315 | 4 | | | | | 49 | 49 | 102 |
| desv_std_coef | 4 | | | 1 | | 49 | 49 | 103 |
| desv_min_coef | 3 | | | | 1 | 49 | 49 | 102 |
| desv_range_coef | 3 | | | | 1 | 49 | 49 | 102 |
| desv_mean_coef | 3 | | | | | 49 | 49 | 101 |
| desv_max_coef | 2 | | | | | 49 | 49 | 100 |
| Total general | 2022 | 36 | 40 | 5 | 5 | 5390 | 5390 | 12888 |

ANEXO D – SELECCIÓN FINAL DE CARACTERÍSTICAS

| Vecinos | k=3 | | k=2 | | k=1 | |
|------------------------|---------|--------|---------|---------|---------|--------|
| | Ranquin | Pesos | Ranquin | Pesos | Ranquin | Pesos |
| 'mov_mean_range_coef' | 1 | 0.0274 | 1 | 0.0276 | 1 | 0.0276 |
| 'mov_mean_min_coef' | 2 | 0.0251 | 2 | 0.0253 | 2 | 0.0254 |
| 'min_coef' | 3 | 0.0246 | 3 | 0.0249 | 3 | 0.0248 |
| 'mode_coef' | 4 | 0.0246 | 4 | 0.0249 | 4 | 0.0248 |
| 'mov_mean_std_coef' | 5 | 0.0200 | 5 | 0.0201 | 5 | 0.0201 |
| 'mov_mean_max_coef' | 6 | 0.0192 | 6 | 0.0195 | 6 | 0.0195 |
| 'max_coef' | 7 | 0.0191 | 7 | 0.0194 | 7 | 0.0195 |
| 'rms' | 8 | 0.0184 | 8 | 0.0185 | 8 | 0.0185 |
| 'spectral_slope' | 9 | 0.0172 | 9 | 0.0173 | 9 | 0.0173 |
| 'frec_0_0093111' | 10 | 0.0117 | 10 | 0.0120 | 10 | 0.0120 |
| 'frec_0_0086876' | 11 | 0.0116 | 11 | 0.0118 | 11 | 0.0119 |
| 'frec_0_0081058' | 12 | 0.0115 | 12 | 0.0117 | 13 | 0.0118 |
| 'median_min_coef' | 13 | 0.0113 | 13 | 0.0116 | 12 | 0.0118 |
| 'frec_0_007563' | 14 | 0.0113 | 14 | 0.0115 | 14 | 0.0115 |
| 'frec_0_0065839' | 15 | 0.0110 | 15 | 0.0113 | 15 | 0.0114 |
| 'frec_0_0070565' | 16 | 0.0107 | 16 | 0.0109 | 16 | 0.0110 |
| 'mov_range_max_coef' | 17 | 0.0096 | 17 | 0.0099 | 17 | 0.0103 |
| 'frec_0_013168' | 18 | 0.0095 | 18 | 0.0097 | 18 | 0.0098 |
| 'median_max_coef' | 19 | 0.0084 | 19 | 0.0087 | 19 | 0.0088 |
| 'frec_0_015126' | 20 | 0.0082 | 20 | 0.0084 | 20 | 0.0085 |
| 'frec_0_014113' | 21 | 0.0079 | 21 | 0.0081 | 21 | 0.0081 |
| 'median_mean_coef' | 22 | 0.0071 | 22 | 0.0072 | 22 | 0.0073 |
| 'frec_0_018622' | 23 | 0.0068 | 23 | 0.0070 | 23 | 0.0071 |
| 'mov_desv_max_coef' | 24 | 0.0066 | 25 | 0.0067 | 25 | 0.0067 |
| 'frec_0_017375' | 25 | 0.0065 | 24 | 0.0068 | 24 | 0.0068 |
| 'median_std_coef' | 26 | 0.0064 | 26 | 0.0066 | 26 | 0.0067 |
| 'frec_0_032423' | 27 | 0.0057 | 27 | 0.0060 | 27 | 0.0061 |
| 'frec_0_021391' | 28 | 0.0057 | 28 | 0.0059 | 28 | 0.0060 |
| 'frec_0_030252' | 29 | 0.0051 | 29 | 0.0053 | 30 | 0.0054 |
| 'frec_0_026336' | 30 | 0.0050 | 30 | 0.0052 | 31 | 0.0053 |
| 'frec_0_028226' | 31 | 0.0049 | 32 | 0.0051 | 32 | 0.0051 |
| 'mov_range_range_coef' | 32 | 0.0048 | 31 | 0.0052 | 29 | 0.0056 |
| 'mov_range_std_coef' | 33 | 0.0041 | 33 | 0.0043 | 33 | 0.0046 |
| 'mov_desv_desv_coef' | 34 | 0.0040 | 35 | 0.0041 | 35 | 0.0044 |
| 'frec_0_045853' | 35 | 0.0039 | 34 | 0.0043 | 34 | 0.0045 |
| 'mov_range_min_coef' | 36 | 0.0037 | 36 | 0.0039 | 36 | 0.0041 |
| 'mov_desv_min_coef' | 37 | 0.0033 | 37 | 0.0035 | 37 | 0.0035 |
| 'mov_desv_mean_coef' | 38 | 0.0009 | 38 | 0.0011 | 38 | 0.0015 |
| 'mov_range_mean_coef' | 39 | 0.0008 | 39 | 0.0011 | 39 | 0.0015 |
| 'desv_median_coef_fre' | 40 | 0.0001 | 40 | -0.0002 | 40 | 0.0002 |