



Vigilada Mineducación

**Predicción del precio de la energía eléctrica en Colombia mediante un
enfoque de *machine learning***

Prediction of electricity price in Colombia through a machine learning approach

Yeison José Villarreal Marimon

Luis Armando Flores San Martín

Tesis de grado

Paula María Almonacid Hurtado PhD

UNIVERSIDAD EAFIT

ESCUELA DE ECONOMÍA Y FINANZAS

MAESTRÍA EN ADMINISTRACIÓN FINANCIERA - MAF

MEDELLÍN

2023

CONTENIDO

INTRODUCCIÓN	8
1. PLANTEAMIENTO DEL PROBLEMA.....	10
2. JUSTIFICACIÓN.....	12
3. OBJETIVOS.....	14
3.1. OBJETIVO GENERAL	14
3.2. OBJETIVOS ESPECÍFICOS.....	14
4. MARCO TEÓRICO O MARCO CONCEPTUAL.....	16
4.1. ESTADO DEL ARTE.....	16
4.1.1. Predicción de precios en mercados internacionales	16
4.1.2. Predicción de precios en el mercado colombiano	20
4.2. MERCADO DE ENERGÍA ELÉCTRICA EN COLOMBIA.....	23
4.3. MODELOS PARA ESTIMACIÓN DEL PRECIO DE LA ENERGÍA.....	27
4.3.1. Regresión lineal y múltiple	27
4.3.2. Modelo de vector autorregresivo (VAR)	29
4.3.3. Modelo autorregresivo integrado de promedio móvil (ARIMA).....	30
4.3.4. Modelo autorregresivo integrado de media móvil con entrada exógena (ARIMAX).....	31
4.3.5. Modelo estacional autorregresivo integrado de media móvil exógena (SARIMAX)	31
5. DISEÑO METODOLÓGICO	33
6. DESARROLLO DEL TRABAJO.....	38
6.1. COMPRESIÓN Y CAPTURA DE LOS DATOS.....	38
6.2. PREPARACIÓN DE BASES DE DATOS.....	40
6.3. EXPLORACIÓN DE DATOS Y ESTADÍSTICA DESCRIPTIVA	41
6.4. MODELO DE REGRESIÓN LINEAL Y MÚLTIPLE.....	42
6.5. MODELO DE PREDICCIÓN VAR	45
6.6. MODELO ARIMA.....	46

6.7. MODELO ARIMAX	47
6.8. MODELO SARIMAX	49
7. RESULTADOS	51
7.1. RESULTADOS DE LA EXPLORACIÓN DE DATOS Y CORRELACIÓN..	51
7.2. EVALUACIÓN DE MODELOS.....	57
7.3. RESULTADOS DEL MODELO DE REGRESIÓN LINEAL Y MULTIPLE..	57
7.4. RESULTADOS DEL MODELO VAR.....	60
7.5. RESULTADOS DEL MODELO ARIMA.....	62
7.6. RESULTADOS DEL MODELO ARIMAX	65
7.6.1. ARIMAX – ARIMA (1,1,1) con variables exógenas Volumen Útil Diario expresado en términos de Energía (kWh) y el fenómeno ENSO.....	65
7.6.2. ARIMAX – ARIMA (2,1,1) con variables exógenas Volumen Útil Diario expresado en términos de Energía (kWh) y el fenómeno ENSO.....	67
7.6.3. ARIMAX – ARIMA (0,1,0) con variables exógenas Volumen Útil Diario expresado en términos de Energía (kWh) y del fenómeno ENSO.....	69
7.7. RESULTADOS MODELO SARIMAX.....	72
7.7.1. Modelo SARIMAX (1,1,1) x (1,1,1,12).....	72
7.7.2. Modelo SARIMAX (0,1,0) x (0,0,0,0).....	73
7.8. Resumen de los resultados de los modelos desarrollados	75
8. CONCLUSIONES	79
REFERENCIAS BIBLIOGRÁFICAS.....	83

LISTA DE FIGURAS

Figura 1 Estructura del mercado eléctrico en Colombia	25
Figura 2 Ciclo de vida del proceso de minería de datos	34
Figura 3 Datos históricos de las variables dependiente e independiente	54
Figura 4 Correlación de variables independientes y dependientes.....	55
Figura 5 Correlación del Valor kW con Volumen útil diario energía y fenómeno ENSO	56
Figura 6 Resultado gráfico de predicción del modelo VAR.....	62
Figura 7 Resultado gráfico de predicción del modelo ARIMA (1,1,1).....	64
Figura 8 Resultado gráfico de predicción del modelo ARIMAX con ARIMA orden (1,1,1)	67
Figura 9 Resultado gráfico de predicción del modelo ARIMAX con ARIMA orden (2,1,1)	69
Figura 10 Resultado gráfico de predicción del modelo ARIMAX con ARIMA orden (0,1,0)	71
Figura 11 Resultado gráfico del backtesting de predicción del modelo SARIMAX (1,1,1) x (1,1,1,12)	73
Figura 12 Resultado gráfico del backtesting de predicción del modelo SARIMAX (0,1,0) x (0,0,0,0)	74

LISTA DE TABLAS

Tabla 1 Estadística descriptiva	52
Tabla 2 Estadística descriptiva (continuación)	53
Tabla 3 Resultados de la prueba de la prueba de Dickey-Fuller	57
Tabla 4 Resultados del modelo de regresión lineal	58
Tabla 5 Resultados del modelo regresión múltiple	59
Tabla 6 Resultados del modelo regresión múltiple incluyendo la variable dummy	59
Tabla 7 Resultados de la evaluación del modelo de regresión múltiple.....	60
Tabla 8 Resultados de la evaluación del modelo VAR	61
Tabla 9 Resultados de la evaluación del modelo ARIMA (1,1,1)	63
Tabla 10 Resultados de la evaluación de modelo ARIMAX con ARIMA orden (1,1,1)	65
Tabla 11 Resultados de la evaluación de modelo ARIMAX con ARIMA orden (2,1,1)	68
Tabla 12 Resultados de la evaluación del modelo ARIMAX con ARIMA orden (0,1,0)	70
Tabla 13 Resultados evaluación de modelo SARIMAX (1,1,1)x(1,1,1,12).....	72
Tabla 14 Resultados de la evaluación de modelo SARIMAX (0,1,0) x (0,0,0,0)	74
Tabla 15 Resumen de la evaluación de resultados de los modelos desarrollados	75

RESUMEN

En esta investigación se desarrollan diferentes modelos de predicción basados en regresión, modelo VAR, modelo ARIMA, modelo ARIMAX y modelo SARIMAX, que fueron usados para estimar y predecir los precios de la energía en bolsa, y así obtener un valor aproximado de venta del kilovatio-hora, un insumo clave para calcular los ingresos en los modelos de valoración de proyectos de generación de energía eléctrica en Colombia. Lo anterior se realizó apoyado en los registros históricos de las bases de datos de XM, el análisis de la relación del precio histórico de la energía en bolsa para el período comprendido entre enero de 2000 y julio 2023, utilizando también las variables de aportes hídricos, vertimientos y reservas hídricas expresados en términos de energía y los posibles efectos de fenómenos climatológicos como *El Niño Southern Oscillation* (ENSO) que se presenta en el país. En la investigación se obtuvo como resultado que los valores del kilovatio se ven afectados por las temporadas de lluvias y en especial en la ocurrencia del Fenómeno de El Niño, durante el cual los precios aumentan incluso activando los precios de escasez del sistema, lo cual se ve marcado en los años 2015 y 2016. Finalmente, como resultado de la predicción, se encontró que todos los modelos siguen las tendencias de los comportamientos de los precios. A los modelos se le hicieron pruebas con diferentes horizontes, encontrando que el modelo para utilizar depende del horizonte de tiempo que necesite analizar el inversionista: corto plazo VAR, mediano plazo SARIMAX y largo plazo regresión múltiple.

Palabras clave: precio energía en Bolsa, ENSO, MEM, Modelos de regresión, VAR

ABSTRACT

In this research, numerous predictive models are developed, including regression models, VAR models, ARIMA models, ARIMAX models and SARIMAX models, which were further used to estimate and predict the electricity spot price, and therefore obtaining an approximate value for the sale of a kilowatt-hour, a critical input for calculating the revenues in the valuation models of electric power generations projects in Colombia. This was accomplished using the historical records from XM's databases, analyzing the relationship between the historical spot price for electricity in the frame of time from January 2000 to July 2023, other input variables were also considered such as hydrological contributions, hydrological discharges and hydrological reserves expressed in terms of energy, as well as the potential effects of climatological phenomena like the El Niño Southern Oscillation (ENSO) that occurs in the country. The results of the research indicate that the prices of the kilowatt-hour are affected by the rainy season and specially by the occurrence of the El Niño phenomena, during which prices increase triggering the scarcity price of the system, which can be observed in the years 2015 and 2016. Finally, as a result, all models follow the price behavior trends. The models were subjected to different time horizon tests, finding that the model to be used depends on the time horizon that the investor needs to analyze: VAR models for the short-term, SARIMAX models for the medium-term and multiple regression models for the long-term.

Keywords: electricity spot price, ENSO, MEM, regression models, VAR models

INTRODUCCIÓN

La energía eléctrica es uno de los servicios básicos más importantes para el ser humano. El consumo de energía eléctrica está presente en todos los aspectos de nuestras vidas y permite el desarrollo y mejora la calidad de vida de las sociedades. En la medida en que un país se desarrolla, crecen de forma proporcional el consumo y la demanda de energía eléctrica y, por ende, la necesidad de nuevas fuentes de generación para abastecer la creciente demanda. Lo anterior requiere de inversión para el desarrollo de nuevos proyectos de energía eléctrica.

Cuando un inversionista decide evaluar un proyecto de generación de energía eléctrica en Colombia, se encuentra con muchas variables por analizar, las cuales en algunos casos son muy difíciles de predecir o de conocer su comportamiento. Una de estas variables, que es importante para determinar la viabilidad financiera de los proyectos, es el precio de venta del kilovatio-hora, el cual finalmente determinará los ingresos que tendría el proyecto de generación de energía.

Para solucionar de antemano este problema a potenciales inversionistas, se construyó una base de datos con registros obtenidos de la plataforma Sinergox, de XM (2023, 2023a), de la cual se obtuvo información de las variables de aportes hídricos, vertimientos y volúmenes, en términos de energía y caudal; además, los precios *spot* (de corto plazo) del kilovatio-hora promedio diario, y de la National Oceanic and Atmospheric Administration (NOAA, 2023) se descargó el Índice Niño

Oceánico (ONI, por sus siglas en inglés *Oceanic Niño Index*), que nos indica cuándo se presenta el fenómeno Oscilación del Sur El Niño (ENSO, por sus siglas en inglés *El Niño Southern Oscillation*). Los anteriores datos se obtuvieron en una resolución diaria.

Con las bases de datos construidas, se analizó la significancia de las variables y la correlación entre ellas, para así seleccionar las variables que nos llevaran a construir modelos de predicción con los mejores resultados posibles.

Se construyeron varios modelos de predicción del valor del kilovatio-hora en la bolsa del mercado de energía de Colombia, por medio de modelos de regresión lineal y múltiple, modelo VAR, modelo ARIMA, modelos ARIMAX y modelos SARIMAX, los cuales mostraron buenos resultados en las tendencias y cuyos resultados se evaluaron con el RMSE y MAE.

Finalmente, con los modelos construidos se pudo obtener una predicción del precio del kilovatio en la bolsa de energía de Colombia, la cual le pueden ayudar a un inversionista a conocer las tendencias de los precios y disponer de una predicción del valor aproximado del kilovatio-hora en diferentes horizontes de tiempo, y así contar con una herramienta útil para planificación financiera.

1. PLANTEAMIENTO DEL PROBLEMA

Desde cuando entró a operar el mercado energético mayorista (MEM) en Colombia, en 1995, surgieron muchas oportunidades de inversión en el sistema eléctrico colombiano, debido que este sistema incluyó la participación del agente generador, comercializador, transmisor y distribuidor.

Los generadores de energía fueron los agentes del mercado que mostraron mayor interés por participar en esta oportunidad de inversión. Esto debido a las características de la competencia por el precio de venta de energía, que es conocido como subasta de sobre cerrado; es decir, en el que cada generador hace su oferta de precio para una generación establecida, y el administrador del mercado va asignando la generación según la demanda del país, siempre desde el menor precio ofertado.

Con este panorama, a los inversionistas generadores se le ha presentado siempre el problema de calcular sus ingresos operativos por medio de la venta de la energía, ya que es un mercado *spot* (de corto plazo) donde los precios están regido en un marco regulatorio, pero con una estimación particular donde surge el interrogante respecto a cuál será ese precio que podría ofertar para estar dentro del mercado que finalmente será despachado. Adicionalmente, en Colombia la mayor capacidad instalada es hidráulica; es decir, la temporada climática influye directamente en los precios. Cuando estamos en temporada de lluvia, hay mayores aportes hídricos a

los embalses, lo que se traduce en mayor oferta o capacidad de la producción de energía, y los precios que puede ofertar esta tecnología de generación son inferiores al precio que puede ofertar la tecnología de generación térmica.

Entendiendo esta problemática, en el presente trabajo de investigación se desarrollan modelos econométricos que ayudan a predecir el precio *spot* de la energía en el mercado mayorista, para que puedan ser utilizados por cualquier inversionista para calcular los ingresos operativos de un proyecto de generación de energía eléctrica en Colombia.

Por tanto, la pregunta de investigación que se busca responder en el presente trabajo es: ¿Cómo puede un inversionista del sector de la generación de energía eléctrica en Colombia obtener una mejor estimación del precio de venta de la energía (kilovatio-hora) en bolsa para utilizarlo como insumo clave en sus modelos financieros para proyectar sus ingresos?

2. JUSTIFICACIÓN

El comportamiento económico es uno de los factores determinantes en las proyecciones de la demanda de energía eléctrica en Colombia, y es utilizado como dato de entrada en la elaboración de los modelos de proyección y planeación del sector minero energético de Colombia (UPME, 2020).

Según la Unidad de Planeación Minero-Energética (UPME, 2023), la proyección de la demanda de energía eléctrica en Colombia ha venido mostrando señales de recuperación, producto de la reactivación económica que se ha venido produciendo desde 2021, luego del impacto de la pandemia del covid-19 ocurrida en 2020. En 2022, la demanda de energía creció 5,49% con respecto a 2020, y a marzo 2023 se observaba un crecimiento promedio mensual de la demanda de energía de 1,89% con respecto a 2022. La demanda de energía eléctrica es proyectada por la UPME considerando variables tales como la demanda histórica de energía, el indicador macroeconómico del producto interno bruto (PIB) histórico, las proyecciones del PIB futuras y el crecimiento poblacional, entre otros.

Con base en estimaciones del Departamento Administrativo Nacional de Estadística (DANE), el Banco Mundial y el Fondo Monetario Internacional (FMI) para Colombia, se contempla un crecimiento esperado de la economía del 3,5% para el período 2025-2037. Como resultados de dichas proyecciones, se observa que para el

periodo 2023-2037 la demanda de energía eléctrica en el escenario medio tendrá un crecimiento entre el 1,65% y el 2,99% (UPME, 2023).

Para hacerle frente al desarrollo económico, el país requerirá nuevas fuentes de generación para abastecer la creciente demanda de energía. Estas nuevas fuentes de generación de energía conllevan el desarrollo de proyectos con altos presupuestos de inversión, los cuales en algún momento deben otorgarle al inversionista una rentabilidad esperada. En este sentido, con el presente trabajo de investigación se desarrollaron modelos econométricos de predicción para la variable del precio de la energía en bolsa, los cuales posteriormente podrán ser utilizados por los inversionistas como un insumo para modelar los ingresos en sus modelos financieros.

La presente investigación deja un aporte significativo a la sociedad y a los desarrolladores de proyectos que deseen invertir en Colombia. De esta manera, el país podrá contar con una seguridad energética que respalde el desarrollo económico proyectado para el país.

3. OBJETIVOS

3.1. OBJETIVO GENERAL

Diseñar un modelo de predicción para estimar del precio del kilovatio-hora de la bolsa del mercado eléctrico en Colombia, que sirva como insumo para calcular los ingresos en los modelos de valoración para proyectos de generación de energía renovable en un horizonte de los próximos tres años.

3.2. OBJETIVOS ESPECÍFICOS

- Crear una base de datos integral para la predicción del precio en bolsa del kilovatio-hora en el mercado eléctrico de Colombia, considerando las variables identificadas en la literatura y las reportadas en el portal de indicadores de XM.
- Realizar un análisis estadístico de las variables incluidas en la base de datos.
- Llevar a cabo un análisis descriptivo univariado y multivariado, para comprender las características de cada variable y sus relaciones, por medio de métodos visuales efectivos para ilustrar los patrones y conexiones entre las variables.
- Desarrollar y aplicar modelos de predicción específicamente adaptados a las peculiaridades de los datos.

- Implementar técnicas avanzadas de modelado que consideren las particularidades del mercado eléctrico y las dinámicas de precios.
- Evaluar de manera rigurosa el rendimiento de los modelos de predicción en términos de su capacidad para predecir los precios energéticos.
- Comparar los modelos a través de métricas apropiadas y seleccionar el más adecuado, considerando su precisión y su eficacia en la predicción.

4. MARCO TEÓRICO O MARCO CONCEPTUAL

4.1. ESTADO DEL ARTE

En la revisión de literatura, se identificó que en últimos años se han desarrollado diferentes metodologías que tienen como finalidad pronosticar el precio *spot* de la energía eléctrica. Los diversos métodos que han sido implementados se analizaron a nivel de mercados internacionales y del mercado de Colombia, con diferentes niveles de complejidad y de variables que ayudan a robustecer la predicción.

4.1.1. Predicción de precios en mercados internacionales

En los mercados de energía eléctrica de distintos países se encontraron una serie de investigaciones que estudiaron no sólo los precios del kilovatio-hora, sino cómo predecirlos, en las que se relacionaron datos históricos de los precios y otras variables tales como demanda de energía, disponibilidad de recursos y tipos de tecnologías de la matriz de generación, entre otras.

Tehrani y otros (2022), desarrollaron un modelo de predicción de corto plazo para las próximas 24 horas del precio *spot* de la energía, en los mercados de Francia, Alemania, Italia, España, Reino Unido y Dinamarca, tomando como insumos la oferta y la demanda de energía, la disponibilidad de recursos de generación, los

costos de combustible y las condiciones climatológicas. En dicho estudio, utilizaron modelos ARIMA y diferentes versiones de GARCH, para incluir los efectos de la heteroscedasticidad de los datos. Como resultado, concluyen que los modelos Garch tienen un mejor desempeño que los modelos ARIMA.

Schöniger y Morawetz (2022), en su investigación *What comes down must go up: Why fluctuating renewable energy does not necessarily increase electricity spot price variance in Europe* analizan en una investigación el efecto de la varianza del precio *spot* de la energía en Europa, en la que se consideró la integración de energías renovables no convencionales a la matriz energética y se analiza la forma de la función de oferta, los precios *spot* históricos y la generación de fuentes de energía no convencionales sobre las cuales se aplican métodos de estadísticas descriptivas y modelos de regresión. De los modelos de panel y modelo de regresión, se concluye que un alto volumen de aportes de fuentes de energía no convencionales incrementa la varianza en los precios, lo cual invita a los reguladores a revisar sus políticas de seguridad energética.

Con un modelo un poco más robusto, Shah y otros (2021) desarrollaron un modelo de predicción de los precios de energía en bolsa del mercado eléctrico de Italia, después de tratar los picos de precios de la serie de datos, utilizando como insumo principal para el análisis el precio histórico de la energía. Como metodología, el estudio realiza un tratamiento de datos extremos con las técnicas de umbrales de precio fijos (TFP, por sus siglas en inglés *total factor productivity*), filtro de desviación

estándar, filtro recursivo, filtro de ventana móvil (por sus siglas en inglés MFP) y filtro de precio porcentual, que luego fueron reemplazados utilizando cuatro metodologías para tratamiento de datos atípicos, tales como el reemplazo de los precios por la media, por la mediana, por valor umbral y por enfoque de esquema de amortiguación. La serie de precios, libre de datos extremos, se divide en los componentes determinísticos y estocásticos, los cuales se modelan utilizando los modelos autorregresivo (AR), autorregresivo no paramétrico (NPAR), autorregresivo de media móvil (ARMA) y autorregresivo vectorial (VAR), tomando como medida de desempeño el error medio absoluto porcentual (MAPE) y el error medio absoluto (MAE). Como resultado, los valores más bajos de MAPE y MAE son obtenidos de los modelos VAR utilizando el filtro de ventana móvil en precios (MFP) para los datos atípicos y reemplazando dichos valores con la técnica del valor umbral.

La investigación *A hybrid short-term electricity price forecasting framework: Cuckoo search-based feature selection with singular spectrum analysis and SVM*, de Zhang y otros (2019), a pesar de haber sido desarrollada hace algunos años, reviste gran importancia por su aplicación de modelos con buen comportamiento en la predicción al mercado del sur de Gales, que para el análisis toman como insumo los precios históricos de la energía. Como metodología de pronóstico, utilizaron una estrategia de pronóstico con un método de selección de características híbridas (HFS) incluyendo un análisis de espectro singular (SSA), un algoritmo de búsqueda del

Cuckoo Search (CS) y una triangulación basada en los coeficientes de correlación de Pearson, Spearman y Kendall.

El modelo del estudio de Zhang y otros (2019) se compara con modelos de referencia tales como SARIMA, ARMA, SVM, ANN y BPNN, de donde se concluye que el modelo propuesto por método de selección de características híbridas (HFS) tiene un mejor desempeño y mayor precisión al predecir en comparación con los modelos de referencia.

Monteiro y otros (2018) llevan a cabo la investigación *New probabilistic 19rice forecasting models: Application to the Iberian electricity market*, Haga clic o pulse aquí para escribir texto. para predecir los precios de la energía eléctrica en un horizonte de corto plazo (*next-day*), para el mercado de la península Ibérica. Como insumos del modelo consideraron los precios históricos de la energía eléctrica, la demanda y generación del día previo, la demanda y generación de la semana previa, los pronósticos de la demanda, los aportes de la generación eólica y los pronósticos climatológicos para el día siguiente. Como metodología utilizaron modelos de pronóstico de precio probabilístico (PPFM), utilizando el enfoque del estimador de densidad de núcleo Naradaraya-Watson (NW-KDE). Los resultados demuestran que los valores del error medio absoluto (MAE) mejoran al incorporarle más variables al modelo. El mejor modelo de predicción obtuvo un MAE de 5,55 euro/MWh y un indicador de fiabilidad de 85,64%.

Otro estudio importante por la integración de redes neuronales es *Composite forecasting approach, application for next-day electricity price forecasting*, realizado por Mirakyan y otros (2017), quienes tratan de predecir los precios de la energía eléctrica en el mercado *spot* para los mercados de Alemania y Austria, que parte de la base de los precios históricos de la energía eléctrica. Para el modelamiento, utilizaron un procesos de pronóstico compuesto (CF) con regresión de vectores de soporte (SVR), red neuronal artificial (ANN) y regresión *ridge* (RR), tomando como medida de rendimiento el error cuadrático medio inverso de la Raíz (IRMSE). De los métodos individuales, el SVR fue el que mostró el mejor desempeño; sin embargo, la combinación de métodos individuales de predicción puede generar un modelo más robusto que la aplicación de un solo método. Para dicho estudio, el proceso de pronóstico compuesto (CF) mejoró la precisión en un 22%.

4.1.2. Predicción de precios en el mercado colombiano

Para el caso del mercado de energía de Colombia, también se han desarrollado estudios e investigaciones que tienen como objetivo predecir el precio *spot* de la energía, en los que se ha explorado la incorporación del fenómeno ENSO, de gran importancia debido a que un gran aporte de energía a la matriz de generación de energía en Colombia procede de la tecnología hidroeléctrica. Adicionalmente, se ha explorado la relación de otras variables que han ayudado a robustecer los modelos.

La investigación *Does climate impact the relationship between the energy price and the stock market? The Colombian case*, de Villa y otros (2023) es considerada importante, ya que estudió la relación entre el precio de la energía eléctrica y los mercados financieros en Colombia, utilizando como variables de interés el fenómeno climático ENSO, el precio histórico de la energía en el mercado *spot* y el Índice General de la Bolsa de Valores de Colombia (IGBX).

La metodología utilizada por Villa y otros (2023) fue la prueba de contagio para examinar correlaciones y el análisis de ondículas para series temporales bivariados (CWA), específicamente un análisis espectral a través de la transformada de ondícula continua (CWT). El análisis reveló una correlación fuerte en el mercado durante el Fenómeno de El Niño que se presentó entre 2015 y 2016, y se determinó que durante este fenómeno se incrementa el precio de la energía en bolsa, mientras que, en contraste, en presencia del Fenómeno de la Niña ocurre lo contrario.

Otra investigación que se destaca es *Proyección de corto plazo para el precio de bolsa de energía en el mercado colombiano*, efectuada por Urbano y González (2022), en la cual desarrollan un modelo de proyección de corto plazo del precio de la energía en bolsa, tomando para ello como insumos la demanda de energía, los vertimientos, los aportes y reservas expresados en términos de energía, los aportes de generación de fuentes de energía no convencionales, hidroeléctricos y termoeléctricos, y el precio histórico de la energía en bolsa. Como metodología, analizaron cuatro tipos de modelo: modelos autorregresivos de media móvil

(ARIMA), modelos Prophet, modelos de media móvil integrada autorregresiva estacional con regresores exógenos (SARIMAX) y redes neuronales perceptor multicapa (LSTM, por sus siglas en inglés *long-short term memory*) tomando como métricas de desempeño el error medio cuadrático (MSE), la varianza explicada y el error máximo. Dichos investigadores concluyen que los modelos de series de tiempo y redes neuronales tienen una gran capacidad de predicción y ayudan a entender la autocorrelación con las variables explicativas, y el precio de energía en bolsa se considera un *commodity* de corto plazo.

Un poco más antiguo, pero relevante para la investigación, es el estudio *Pronóstico del precio de la energía en Colombia utilizando modelos ARIMA con IGARCH*, de Muñoz y otros (2017), [Haga clic o pulse aquí para escribir texto](#). en el cual se pronostica el precio de la energía en bolsa para horizonte de tiempo de un mes en el mercado colombiano, tomando como insumo el precio histórico de la energía en bolsa durante el período 2004-2016. Como metodologías para el pronóstico del precio de energía, se utilizó el modelo autorregresivo de media móvil (ARIMA) y el de autorregresión generalizada condicional heterocedástica (GARCH). El resultado fue un modelo de pronóstico robusto, que permitió pronosticar el precio para la energía de enero de 2017 con un desfase del 4% con respecto al precio real de energía en bolsa para dicho mes.

A las anteriores investigaciones se suma *Modelamiento de los precios de la energía en bolsa en Colombia incorporando el efecto del ENSO*, desarrollada por Galindo

(2017) incorpora los efectos climáticos de ENSO, con el fin de modelar los precios de la energía en bolsa de Colombia mediante la utilización de modelos autorregresivos de media móvil (ARIMA), considerando tres escenarios de intensidad para el Fenómeno de El Niño. Como resultado, el modelo propuesto tiene un buen ajuste a la serie histórica de precios, encontrando que, con un 95% de confianza, el próximo Fenómeno de El Niño tendrá una intensidad media y llegará en un máximo de 967 días contados a partir del 30 de abril del 2016.

Finalmente, entre estos estudios, se destaca *La hidrología como predictor del comportamiento del precio de energía en bolsa*, de Barrientos y Toro (2016), que tiene como finalidad analizar el impacto de la hidrología como predictor del comportamiento de los precios de la energía en Colombia, tomando como insumos los precios históricos de la energía en bolsa, los aportes hídricos y la demanda del mercado. Como metodologías, utilizaron modelos autorregresivos de retardo distribuido (ARDL) y vectores autorregresivos (VAR), mediante los cuales se concluye que son modelos muy efectivos para analizar la relación entre la hidrología y el precio de energía en bolsa, siendo la hidrología un predictor del comportamiento del precio de la energía para el mercado de Colombia.

4.2. MERCADO DE ENERGÍA ELÉCTRICA EN COLOMBIA

En Colombia, el suministro de energía eléctrica por medio de empresas inicia en el siglo XIX, y al irse desarrollando, con el tiempo se le fueron sumando inversionistas;

sin embargo, en los años ochenta del siglo XX, debido al crecimiento del país y a la poca eficiencia en la planeación de la demanda, entre otros factores, comenzó una crisis que llevó al gran racionamiento nacional que tuvo lugar entre 1991 y 1992 (Bello y Beltrán, 2013).

Luego de esta crisis, el mercado de energía eléctrica comenzó su reestructuración con la *Constitución Política de Colombia* (Asamblea Nacional Constituyente, 1991), en la que se consideraron algunas pautas para el suministro de servicios públicos domiciliarios; sin embargo, con las Leyes 142 y 143 de 1994, se estructura el mercado con la creación de la Comisión de Energía y Gas (CREG), encargada del desarrollo del marco regulatorio, y en 1995 entra en funcionamiento el Mercado Energético Mayorista (Galindo, 2017).

El Mercado Energético Mayorista en Colombia, desde su creación se ha venido desarrollando con agentes activos en el mercado y la inclusión de inversionistas, lo cual hoy en día mantiene una estructura con agentes del mercado, que producen, transmiten y comercializan la energía eléctrica a los consumidores finales.

La clasificación de los agentes que estructuran el mercado eléctrico en Colombia se presenta a continuación en la figura 1.

Figura 1

Estructura del mercado eléctrico en Colombia



Nota. Diagrama tomado de *Estructura del mercado* (XM, 2023b).

Según XM (2023b):

Los agentes del mercado son los encargados de producir, llevar y vender la energía al usuario final. Se clasifican en generadores, transmisores, distribuidores, comercializadores y administradores, según el rol que desempeñan.

Generadores. Son los encargados de producir la energía por medio de centrales hidráulicas, térmicas y eólicas.

Transmisores. Son los encargados de transportar largas distancias la energía desde las centrales eléctricas hasta las subestaciones de transformación a través de redes que operan a tensiones iguales o superiores a 220 kV.

Distribuidores. Son los encargados de llevar la energía hasta el consumidor final a través de redes que operan a tensiones inferiores a 220 kV.

Comercializadores. Son los encargados de la compra de energía eléctrica en el mercado mayorista y su venta a usuarios finales.

Administrador. El administrador del mercado, XM, es el encargado de:

- Registrar las fronteras, es decir, los sistemas de medida de consumo de energía, su ubicación y su representante.
- Liquidar y facturar los intercambios de energía resultantes entre los agentes generadores y comercializadores del mercado, que venden y compran en la bolsa de energía.
- Recaudar el dinero producto de las transacciones en bolsa, las Transacciones Internacionales de Electricidad y los servicios por transmisión nacional y regional para entregarlos a los agentes transmisores y distribuidores por el uso de sus redes. (párrs. 1-5)

Con respecto a los precios de la energía eléctrica en el MEM, el procedimiento se establece por medio de subasta de sobre cerrado, ordenando los precios ofertados de menor a mayor, y se fija por medio de la demanda; es decir, hasta que se

satisfaga la demanda, tomando los precios de menor a mayor. Esto con el fin de garantizar la libre competencia y se manifiesten las condiciones del mercado (Galindo, 2017).

4.3. MODELOS PARA ESTIMACIÓN DEL PRECIO DE LA ENERGÍA

Para la estimación de los precios de energía eléctrica en Colombia, se han utilizado diversos modelos econométricos autorregresivos que pueden llegar a una buena aproximación.

A continuación, explicamos cada uno de los modelos utilizados en la presente investigación.

4.3.1. Regresión lineal y múltiple

El modelo de regresión lineal, en su expresión más simple, se puede observar en la ecuación, en donde el modelo trata de predecir de manera cuantitativa la variable Y a partir de una única variable explicativa X . La calidad del modelo de regresión lineal es normalmente evaluada utilizando el R-Cuadrado, el cual toma valores entre 0 y 1, y equivale a la proporción de la variabilidad que es explicada por el modelo para la variable Y (James y otros, 2023).

$$Y = \beta_0 + \beta_1 \times X + \epsilon \quad (1)$$

donde:

β_0 = coeficiente de intercepción

β_1 = coeficiente de pendiente

X = variable independiente

ϵ = término del error

Cuando la intención es tener más de una variable explicativa en el modelo, se puede extender el modelo de regresión lineal a un modelo de regresión múltiple, el cual toma la forma general expresada en la ecuación. En los modelos de regresión múltiple, los coeficientes de asociación de las variables independientes se deben calcular utilizando métodos de álgebra matricial. En el proceso de desarrollo de estos modelos, también se deben analizar los *p-value* de cada variable independiente, para determinar si dichas variables son estadísticamente significativas. Al igual que en los modelos de regresión lineal, la calidad del modelo de regresión múltiple se puede analizar evaluando el R-Cuadrado; sin embargo, es usual que en los modelos de regresión múltiple se puedan adicionar variables *dummy*, con la finalidad de mejorar el ajuste del modelo (James y otros, 2023).

$$Y = \beta_0 + \beta_1 \times X_1 + \dots + \beta_p \times X_p + \epsilon \quad (2)$$

donde:

β_0 = coeficiente de intercepción

$\beta_1, \beta_2, \dots, \beta_p$ = coeficientes de asociación de variables independientes

X_1, X_2, \dots, X_p = variables independientes

ϵ = término del error

4.3.2. Modelo de vector autorregresivo (VAR)

El modelo VAR, utilizado en la econometría con mayor regularidad desde la década de 1980, es una regresión de más de una variable dependiente, que representa una combinación entre modelos univariados y modelos con varias ecuaciones.

El modelo VAR, expresado de forma simple, es donde solo existen dos variables: y_{1t} y y_{2t} , donde los valores dependen de diferentes combinaciones, que se puede apreciar en las ecuaciones (3) y (4):

$$y_{1t} = \beta_{10} + \beta_{11}Y_{t-1} + \dots + \beta_{1k}y_{1t-k} + \alpha_{11}y_{2t-1} + \dots + \alpha_{1k}y_{2t-k} + \mu_{1t} \quad (3)$$

$$y_{2t} = \beta_{20} + \beta_{21}Y_{2t-1} + \dots + \beta_{2k}y_{2t-k} + \alpha_{21}y_{1t-1} + \dots + \alpha_{2k}y_{1t-k} + \mu_{2t} \quad (4)$$

Al modelo VAR, en lugar de dos variables, también se le pueden agregar un número y_{gt} de variables (y_{1t} , y_{2t} , y_{3t} , \dots , y_{gt}), cada una de las cuales tiene una ecuación y una combinación.

Una forma compacta del modelo VAR es la que expresa Brooks (2019): “Por ejemplo, consideremos el caso anterior donde $k = 1$, de modo que cada variable depende solo de los valores inmediatamente anteriores de y_{1t} y y_{2t} , además de un término de error” (p. 411).

Esto podría escribirse de la siguiente manera:

$$y_{1t} = \beta_{10} + \beta_{11}Y_{t-1} + \alpha_{11}y_{2t-1} + \mu_{1t} \quad (5)$$

$$y_{2t} = \beta_{20} + \beta_{21}Y_{2t-1} + \alpha_{21}y_{1t-1} + \mu_{2t} \quad (6)$$

O de la siguiente manera:

$$\begin{pmatrix} y_{1t} \\ y_{2t} \end{pmatrix} = \begin{pmatrix} \beta_{10} \\ \beta_{20} \end{pmatrix} + \begin{pmatrix} \beta_{11} & \alpha_{11} \\ \alpha_{21} & \beta_{21} \end{pmatrix} \begin{pmatrix} y_{1t-1} \\ y_{2t-1} \end{pmatrix} + \begin{pmatrix} \mu_{1t} \\ \mu_{2t} \end{pmatrix} \quad (7)$$

).

4.3.3. Modelo autorregresivo integrado de promedio móvil (ARIMA)

El modelo autorregresivo integrado de promedio móvil (ARIMA, por sus siglas en inglés *autoregressive integrated moving average*) se define por medio de un promedio móvil integrado autorregresivo, que proviene de una ecuación lineal que finalmente se utiliza para describir y pronosticar datos en una serie de tiempo. Para crear la ecuación, se deben tener en cuenta sus tres características: la autorregresión, basada en datos históricos; la integración, que trata de la tendencia que traigan los datos; y el promedio móvil, en términos de ecuación de error basados en los datos históricos (Urbano y González, 2022).

Para el ARIMA se debe identificar si los datos tienen: estacionalidad, cantidad de valores atípicos variación de los datos sobre la media (Urbano y González, 2022).

4.3.4. Modelo autorregresivo integrado de media móvil con entrada exógena (ARIMAX)

Según Avellaneda y otros (2012), los modelos autorregresivo integrado de media móvil con entrada exógena (ARIMAX, por sus siglas en inglés *autoregressive integrated moving average with exogeneous input*) tienen tres partes básicas:

La primera es auto regresiva que relaciona la serie con observaciones de periodos anteriores de ella misma. La segunda parte son medias móviles (MA) que relaciona la serie como una función de una sucesión de errores correspondientes a periodos anteriores ponderados. La tercera parte (X) son las observaciones anteriores de una serie exógena, que en este caso es un indicador técnico de los ocho que comúnmente se utilizan en la economía para estimar los precios. (p. 28) Haga clic o pulse aquí para escribir texto.

4.3.5. Modelo estacional autorregresivo integrado de media móvil exógena (SARIMAX)

El modelo estacional autorregresivo integrado de media móvil exógena (SARIMAX, por sus siglas en inglés *seasonal autoregressive integrated moving average with exogenous regressors*) presenta unas modificaciones a los modelos ya definidos.

De acuerdo con Fathi y otros (2019):

La simulación de series temporales más comúnmente utilizados son los conocido como los modelos ARIMA. Para convertirse en el modelo SARIMAX, estos modelos pueden ser extendidos para tener en cuenta tanto una o más series temporales exógenas como un componente estacional (p, d, q)(P, D, Q)s, junto con variables exógenas. (p. 828)

La expresión matemática de forma general se presenta en la ecuación.

$$\varphi_p(B)\Phi_P(B^s)(1 - B)^d(1 - B^s)^D Y_t = c + X_t\beta + \theta_q(B)\Theta_Q(B^s)\varepsilon_t \quad (8)$$

5. DISEÑO METODOLÓGICO

Los modelos de pronóstico son amplios, y en la literatura encontramos una diversidad de estos; sin embargo, los modelos, cuanto más amplios, más complejos se vuelven, y en algunos casos, para elaborarlos se requiere un alto consumo de recursos de procesamiento.

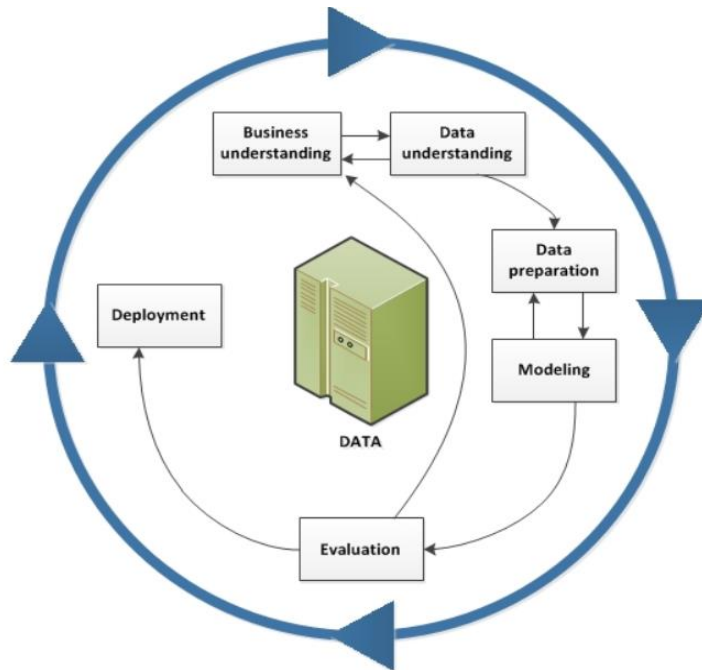
Teniendo en cuenta que para este estudio se utilizan datos históricos observados, algunos de los cuales fueron clasificados como serie de tiempo, se debe hacer en principio una caracterización de los datos. Para ello, se revisa si existen estacionalidades apoyadas de gráficos, luego se verifican tendencias, pruebas de correlación, se elige longitudes de rezago basados en conceptos empíricos y se realizan algunas pruebas de hipótesis (Muñoz y otros, 2017).

Teniendo en cuenta la finalidad del presente estudio, que tiene un enfoque de *machine learning*, se selecciona la metodología de proceso estándar inter-industrias para minería de datos (CRISP-DM, por sus siglas en inglés *cross-industry standard process for data mining*), según lo define IBM (2021): “Es un método probado para orientar sus trabajos de minería de datos” (párr. 1)

El esquema de la metodología CRISP-DM se aprecia en la Figura 2.

Figura 2

Ciclo de vida del proceso de minería de datos



Nota. Fases o pasos de metodología CRISP-DM. Diagrama tomado de Conceptos básicos de ayuda de CRISP-DM. Figura 1. Ciclo de vida de minería de datos. *Guía de CRISP-DM de IBM SPSS Modeler* (IBM, 2021).

Esta metodología incluye seis fases o tareas necesarias, para el desarrollo de un proyecto, y explica las relaciones entre ellas, tal como se describe a continuación. Aclarando que para nuestra investigación el paso uno se adapta en comprensión del problema.

- 1. Compresión del problema.** En esta fase, se plantea el problema que se va a resolver y se define con claridad cuáles son los datos y recursos disponibles para

resolver el problema; además, se plantean los objetivos, para proceder a identificar los datos que realmente son o serían efectivos para lograr lo que se tiene (minería de datos); es decir, en este paso se deja definida la base de datos que se utilizará en el problema ya comprendido (IBM, 2021).

2. Comprensión de los datos. En esta fase, inicialmente se estudian la base de datos disponibles y luego se pasa al análisis, con la ayuda de tablas, gráficos y herramientas estadísticas que ayuden a obtener datos de buena calidad y a hacer una caracterización específica de estos; se realiza una verificación de la calidad de los datos, para detectar errores, valores perdidos o cualquier tipo de incoherencia que puedan sesgar o llevar a dificultar el análisis, y se identifican datos que no estén registrados, errores tipográficos, datos con mediciones incorrectas, medidas o unidades incorrectas o errores en el significado aparente de un campo o definición del campo (IBM, 2021).

3. Preparación de los datos. Luego de tener un análisis preliminar de los datos, se inicia la preparación de la base de datos, la cual tendrá en cuenta los siguientes aspectos:

- Unión de las distintas fuentes de datos, para que sólo quede una base de datos concatenada.
- Si es necesario, se selecciona una muestra de los datos; normalmente, para el análisis se utilizan todos los datos disponibles.
- Donde es necesario, se crean nuevos atributos en la base de datos; para

casos como el de esta investigación, se produce por lo menos una nueva base de datos con los datos que se predicen.

- Clasificación de datos, para los modelos que se vayan a utilizar.
- Depuración de la base de datos, de registros con errores o en blanco.
- División de datos para calibración y validación del modelo (IBM, 2021).

4. Modelo. En esta fase, se utiliza la base de datos que se estuvo preparando en las fases anteriores, la cual se le incorpora a las herramientas analíticas o modelos, que en esta investigación son regresión lineal y múltiple, VAR, ARIMA, ARIMAX y SARIMAX. En el modelado, se hacen varias iteraciones o corridas del modelo, que permitirán ir ajustando parámetros, calibrando y validando el modelo. Por dicha razón, en esta fase, en cada una de las iteraciones se regresa a la fase de preparación de datos, para obtener un mejor ajuste o predicción para nuestra investigación (IBM, 2021).

5. Evaluación del modelo. Luego de terminar la fase del modelado, incluyendo las iteraciones y ajustes, se lleva a cabo un comparativo de los modelos utilizados con sus resultados, y se debe evaluar cuáles de ellos reflejan con mayor ajuste la realidad, según lo definido en los objetivos y apoyado en criterios estadísticos tales como error medio cuadrático y error medio cuadrático medio de la raíz, y herramientas gráficas que dejen notar las tendencias (IBM, 2021).

6. Virtual o despliegue. En esta fase, se definen las conclusiones de la investigación, se determina qué tan buenos fueron los modelos aplicados y en qué condiciones o limitaciones se pueden utilizar los modelos para que arrojen buenos resultados; adicionalmente, se propone qué posibles investigaciones resultarán a partir de este estudio, con base en los resultados obtenidos (IBM, 2021).

Estas fases de la metodología le ayudarán a la presente investigación a cumplir los objetivos planteados y a llegar a unas conclusiones de importancia en la predicción del precio del kilovatio en el mercado de energía en Colombia (IBM, 2021).

6. DESARROLLO DEL TRABAJO

Para desarrollar el presente trabajo, se utilizó la metodología CRISP-DM; sin embargo, en este capítulo detallamos los pasos que se utilizaron para obtener los resultados buscados en esta investigación.

6.1. COMPRENSIÓN Y CAPTURA DE LOS DATOS

Luego de entender el problema que se buscaba resolver, tal como se describe en la sección del planteamiento del problema, el primer paso de esta investigación fue la captura de datos para el análisis, los cuales fueron tomados de las siguientes fuentes:

- Plataforma Sinergox, de XM (XM, 2023a)
- National Oceanic and Atmospheric Administration (NOAA,2023)

De la base de datos Sinergox, de XM (2023a), se recolectaron datos para el período de análisis, comprendido entre el 1 de enero de 2000 y el 31 de julio de 2023, y se tomaron en cuentas las variables *Aporte Hídricos*, expresados en porcentaje y en términos de energía (kWh), *Aportes en Caudal*, expresados en metros cúbicos por segundo (m^3/s), *Volumen Útil Diario*, expresado en millones de metros cúbicos (Mm^3) y en términos de energía (kWh) y porcentaje (%), *Volumen Total*, expresado en millones de metros cúbicos (Mm^3) y en términos de energía (kWh) y porcentaje

(%), y *Volumen de Vertimientos*, expresado en miles de metros cúbicos (miles m³) y en términos de energía (kWh). De la base de datos de NOAA (2023), se tomaron los datos del Índice Oceánico de El Niño (ONI, por sus siglas en inglés *Oceanic Niño Index*) para el período de análisis.

Una vez obtenidos los datos crudos, se inició un proceso para organizar las dos bases de datos que posteriormente se cargaron a Python, con la finalidad de desarrollar los modelos de predicción del precio de la energía en bolsa.

Las bases de datos cargadas en Python: Base de Datos Precio del Kilovatio-hora y Base de Datos Variables Explicativas, se describen a continuación.

- **Base de Datos Precio del Kilovatio-hora.** La primera base de datos generada contiene tanto la información de las fechas del período de análisis como los precios de energía en bolsa, expresados en pesos colombianos (COP) por kilovatio hora.
- **Base de Datos Variables Explicativas.** La segunda base de datos generada contiene la información de las fechas del período de análisis y los datos de las variables Aportes Hídricos (%), Aportes en Energía (kWh), Aportes en Caudal (m³/s), Volumen Útil Diario (Mm³), Volumen Útil Diario Energía (kWh), Volumen Útil Diario (%), Volumen Total (Mm³), Volumen Total Energía (kWh), Volumen Total (%),

Vertimientos Volumen (miles m³), Vertimientos Energía (kWh) y el ONI, donde un valor de 1 se interpreta como la ocurrencia del Fenómeno de El Niño. Es importante resaltar que la resolución para ambas bases de datos es diaria, por lo cual se obtuvieron 8613 observaciones.

A cada una de las bases de datos se le realizó una exploración de los datos, a los cuales se le hizo un análisis de estadística descriptiva y de las respectivas tendencias.

6.2. PREPARACIÓN DE BASES DE DATOS

Con las bases de datos del precio del kilovatio-hora y las variables explicativas, se procedió a cargarlos en Python, y mediante un proceso de concatenación de ambas bases de datos se generó una sola base de datos, la cual es posteriormente utilizada para desarrollar los modelos de predicción. Sobre esta nueva base de datos se procedió a realizar una revisión de los datos vacíos, en la que se encontró que las series de datos para las variables de Vertimientos (miles m³) y Vertimientos Energía (kWh) contenían 2986 datos vacíos, razón por la cual se decidió eliminar dichas variables de la base de datos antes de realizar el análisis. Es importante resaltar que los vertimientos de caudales son producto de un exceso de aportes hídricos, y que la base de datos utilizada para el análisis mantiene los datos de las variables asociadas a los aportes hídricos.

6.3. EXPLORACIÓN DE DATOS Y ESTADÍSTICA DESCRIPTIVA

Con las bases de datos definidas, se inició un proceso de análisis exploratorio de las variables, que comenzó con una estadística descriptiva, donde se observa un incremento notorio del precio del kilovatio-hora en bolsa en el período 2015-2016, el cual coincide con un período de ocurrencia del Fenómeno de El Niño fuerte que se presentó en Colombia, y que conlleva una clara reducción de los aportes hídricos y de volúmenes útiles. Para el período 2023, existía la amenaza de ocurrencia del Fenómeno de El Niño, y en el período mayo-julio de 2023 se presentó un incremento del precio de la energía en bolsa.

Posteriormente, se procedió a elaborar una matriz de correlaciones, en la que existen correlaciones inversas entre el precio de la energía en bolsa y los aportes hídricos, aportes de caudal, volumen útil diario expresado en términos de energía y porcentaje y volumen total expresado en términos de energía. Lo anterior, da indicios de que el precio de energía en bolsa puede ser susceptible a la hidrología del país, y también que la generación hidráulica tiene un alto componente de aporte a la matriz energética de Colombia, tal como se muestra en el informe sectorial elaborado por Corficolombiana (2023), en el que se indica que para 2022 el 86,66% de la generación del país provino de la energía hidráulica.

Luego de generar la matriz correlación, para desarrollar los modelos de Regresión Múltiple, VAR, ARIMA, ARIMAX y SARIMAX se tomaron las variables con la mayor

correlación negativa y positiva. Las variables tomadas fueron Volumen útil Diario en kWh y fenómeno ENSO.

Finalmente, a los datos seleccionados se le aplicó la prueba de Dickey-Fuller, para establecer la estacionalidad de los datos (Mahadeva y Robinson, 2009). Para esta prueba, la hipótesis nula considera que la serie de tiempo no es estacionaria. Es útil aclarar que esta prueba fue aplicada sobre las series de tiempo de las variables Valor del kW, Volumen Útil Diario expresado en términos de Energía (kWh) y fenómeno ENSO.

Como resultado de la prueba Dickey-Fuller, se rechaza la hipótesis nula para las series de tiempo de las variables, debido a que el *p-value* de la prueba es menor al nivel de significancia de 0,05, lo que indica que las series de datos son estacionarias. Debido a lo anterior, se decide no aplicar ningún tipo de transformación sobre los datos.

6.4. MODELO DE REGRESIÓN LINEAL Y MÚLTIPLE

Una vez finalizada la exploración de datos y estadística descriptiva, se construyó un modelo de regresión lineal y múltiple. Para desarrollar este modelo, se utilizó la base de datos concatenada con resolución diaria, la cual contiene la información de las variables Valor del kW (COP), Aportes en Energía (kWh), Aportes en Caudal (m³/s), Volumen Útil diario (Mm³), Volumen Útil diario en términos de Energía (kWh),

Volumen Total (Mm³), Volumen Total en Energía (kWh) y fenómeno ENSO, para las fechas comprendidas en el período de análisis.

Como primer paso, se analiza la regresión lineal entre el Valor del kWh, la cual en el modelo de regresión será la variable dependiente, y cada una de las variables independientes de la base de datos.

La expresión matemática general de cada uno de los modelos se observa en la ecuación (9).

$$Y = \beta_0 + \beta_1 \times X \quad (9)$$

Donde:

Y = valor del kWh (COP/kWh)

X = variable independiente según cada escenario

β_0 = coeficiente de intercepción

β_1 = coeficiente de pendiente

Posteriormente, se procedió a generar el modelo de regresión múltiple tomando en cuenta las variables independientes más relevantes de la base de datos, con la finalidad explicar la variable del valor del kWh.

La expresión matemática del modelo se encuentra en la ecuación (10).

$$Y = \beta_0 + \beta_1 \times X_1 + \beta_2 \times X_2 \quad (10)$$

Donde:

Y = valor del KWh (COP/kWh)

X_1 = volumen útil diario en energía (kWh)

X_2 = fenómeno ENSO

β_0 = coeficiente de intercepción

β_1, β_2 = coeficientes de variables independientes

Una vez construido el modelo de regresión múltiple, se procedió a analizar los residuales, a los cuales se les aplicó la prueba de Jarque Bera, con el fin de determinar si los residuales siguen una distribución normal. Como resultado de la prueba, se rechaza la hipótesis nula; sin embargo, debido al gran volumen de datos, se consideró que no es un inconveniente para el modelo. Para mejorar las estimaciones, y considerando que en 2015 y 2016 ocurrió en Colombia un fuerte Fenómeno de El Niño, que alteró significativamente los precios de la energía en bolsa, se procedió a construir una variable *dummy* para los datos comprendidos entre el 15 de septiembre de 2015 y el 14 de abril de 2016.

La expresión matemática del modelo se presenta en la ecuación (11).

$$Y = \beta_0 + \beta_1 \times X_1 + \beta_2 \times X_2 + \beta_3 \times X_3 \quad (11)$$

Donde:

Y = valor del KWh (COP/kWh)

X_1 = volumen útil diario en energía (kWh)

X_2 = fenómeno ENSO

X_3 = variable Dummy β_0 = coeficiente de intercepción

$\beta_1, \beta_2, \beta_3$ = coeficientes de variables independientes

6.5. MODELO DE PREDICCIÓN VAR

Habiendo desarrollado el modelo de regresión múltiple, se procede a elaborar el modelo de predicción VAR. Como primer paso, se dividió la serie de tiempo: primero, en un conjunto de datos para entrenamiento, compuesto por 8598 observaciones, y un segundo conjunto, de 15 observaciones que se utilizaron para la prueba de pronóstico del modelo.

Como siguiente paso, se aplica la prueba de Causalidad de Granger, la cual tiene como finalidad determinar si una serie de tiempo de una variable puede predecir la serie de tiempo de otra variable. En la prueba de Causalidad de Granger, la hipótesis nula indica que los valores pasados de una variable X no causan Granger en la variable Y . Los resultados de la prueba de Causalidad de Granger nos muestran que, a excepción del fenómeno ENSO, se puede rechazar la hipótesis nula. Por lo tanto, los valores pasados de las variables causan Granger en las otras variables.

Una vez aplicada la prueba de Causalidad de Granger, se procede a analizar los residuales, a los cuales también se les aplica la prueba de Dickey-Fuller. Como resultado, se rechaza la hipótesis nula, que indica que los residuales también tienen un comportamiento estacionario. Una vez verificado lo anterior, se procede a correr el modelo de predicción del precio de la energía y se compara contra el precio real de la energía para los últimos 15 días, comprendidos entre el 17 y el 31 de julio de 2023.

6.6. MODELO ARIMA

Una vez desarrollados los modelos de regresión múltiple y VAR, se procede a construir el modelo de predicción ARIMA. Para construir el modelo, también se dividió la serie de tiempo; primero, en un conjunto de datos para entrenamiento, compuesto por 8598 observaciones, y segundo, en un conjunto de 15 observaciones que se utilizaron para la prueba de pronóstico del modelo.

Una vez dividida la serie de tiempo, se procede a hacer una descomposición estacionaria (STL, por sus siglas en inglés *seasonal decomposition of time serie*) de la serie de tiempo del Valor del kW, tanto aditiva como multiplicativa, con el fin de graficar los resultados para analizar las tendencias, estacionariedad y residuales. Una vez revisado lo anterior, se aplica la función de autocorrelación y autocorrelación parcial para el Valor del KW, para luego calcular los cambios

porcentuales de las variables del Valor del kW, el Volumen Útil Diario expresado en términos de Energía (kWh) y el fenómeno ENSO.

Luego se procede a aplicarle la prueba Dickey Fuller Aumentada (ADF) a la serie de retornos porcentuales del Valor del KW, con el fin de determinar la estacionariedad de los datos. Como resultado, se puede rechazar la hipótesis nula debido a que el *p-value* es menor que 0,05, lo que indica que la serie de retornos porcentuales del Valor del KWh es estacionaria, y se procede a aplicar la función de autocorrelación y autocorrelación parcial a los retornos del Valor del KW.

Una vez comprobado lo anterior, se procede a correr el modelo ARIMA, en el cual se define que este modelo se ajustará a la serie de tiempo de entrenamiento de la variable del Valor del kW. El modelo ARIMA contempla un componente autorregresivo (AR) de orden 1, un componente de integración (I) de orden 1 y un componente de media móvil (MA) de orden 1, quedando como orden del modelo: (1,1,1), que para efectos de evitar cualquier ambigüedad se escribirá siempre entre paréntesis.

6.7. MODELO ARIMAX

El modelo ARIMAX desarrollado se diferencia del ARIMA en la inclusión de variables exógenas en el modelo. Mientras que en el modelo ARIMA la predicción de la variable Valor del kW se basó en función de sus propios valores históricos, en el

modelo ARIMAX se contempla la inclusión de variables exógenas e independientes. En este caso, la inclusión de las variables Volumen Útil Diario, expresado en términos de Energía (kWh), y el fenómeno ENSO. Al igual que en el procedimiento seguido en la construcción del modelo ARIMA, aquí se dividió la serie de tiempo en un conjunto de datos para entrenamiento y un conjunto de datos para la prueba de pronóstico del modelo, para el período comprendido entre el 17 y el 31 de julio de 2023.

En el proceso de generación del modelo, se inició con la creación de un modelo ARIMA de orden (1,1,1), con la inclusión de la variable exógena Volumen Útil Diario expresado en términos de Energía (kWh). Esto con el fin de modelar cómo los cambios en esta variable exógena afectan a la variable a pronosticar Valor del kW. Una vez desarrollado el proceso anterior, se ajusta el modelo ARIMA de orden (1,1,1) con la inclusión de dos variables exógenas: Volumen Útil Diario expresado en términos de Energía (kWh) y fenómeno ENSO, siempre manteniendo como objetivo analizar el efecto sobre la variable Valor del kW a predecir.

Después se generaron diversos modelos ARIMAX, en los cuales se incluyen las variables exógenas anteriormente mencionadas, y se modifica el orden de los componentes autorregresivo, de integración y promedio móvil. Posteriormente se evalúan los modelos ARIMAX con ARIMA orden (1,1,1), ARIMA orden (2,1,1) y ARIMA orden (0,1,0), todos contemplando las variables exógenas del Volumen Útil Diario expresado en términos de Energía (kWh) y el fenómeno ENSO.

6.8. MODELO SARIMAX

Habiendo desarrollado los modelos de regresión, VAR, ARIMA y ARIMAX se procede a construir el modelo de predicción SARIMAX. A diferencia de los modelos anteriores, el modelo SARIMAX incorpora componentes estacionales. En un procedimiento similar al realizado en la construcción de los modelos anteriores, se procede a dividir la serie de datos en un conjunto para entrenamiento y un conjunto para pruebas de pronóstico. La fecha de corte de la serie de tiempo se define como el 1 de enero de 2022, obteniendo de esta manera 8036 observaciones para entrenamiento y 577 observaciones para la prueba de pronóstico.

Como siguiente paso en el desarrollo del modelo, se aplican pruebas de estacionariedad a la variable del Valor del kW utilizando las pruebas Dickey Fuller Aumentada (ADF) y Kwiatkowski-Phillips-Schmidt-Shi (KPSS). Como resultado de las pruebas ADF y KPSS, se puede decir que la serie de datos del Valor del kW es estacionaria; sin embargo, en las serie de diferenciación de orden 1 y 2 la prueba KPSS indica que son serie no estacionarias.

Una vez verificado lo anterior, se procede a generar gráficas de las autocorrelaciones, las autocorrelaciones parciales y a la descomposición de las

series del Valor del kW y la diferenciación de orden 1, con el fin de analizar la tendencia, estacionalidad y residuos.

Como siguiente paso, se procede a la creación del modelo SARIMAX, el cual se entrenará con el conjunto de entrenamiento de la variable Valor del kW, y en el cual se define un orden (1,1,1) para los componentes autorregresivos, integración y media móvil, y un orden (1,1,1,12) para los componentes estacionales. Este modelo funciona como modelo base; sin embargo, mediante un proceso de *backtesting* y *grid search*, se buscan diferentes combinaciones de hiperparámetros de los componentes estacionales, con el fin de buscar el modelo SARIMAX más preciso para reducir el Error Medio Absoluto (MAE) de las predicciones.

7. RESULTADOS

7.1. RESULTADOS DE LA EXPLORACIÓN DE DATOS Y CORRELACIÓN

Para la exploración de los datos, lo primero que se hizo fue graficar las series disponibles, analizar cuál era su tendencia y qué datos podrían estar distintos al del comportamiento histórico. Luego se obtuvo la estadística descriptiva de los datos, para conocerlos mejor. Se encontró que se contaba con una variable dependiente y 10 independientes, cada una con 8613 datos, que abarcaban desde enero de 2000 hasta julio de 2023. Adicionalmente, se resalta la desviación estándar obtenida del valor del kilovatio de \$143, la cual será analizada más adelante en el presente documento.

En las gráficas se destaca el evento de El Niño que ocurrió entre 2015 y 2016, donde el valor del kilovatio también se elevó, incluso rompiendo el comportamiento histórico de los datos. Es aquí, en esta franja de tiempo, donde se llegan a obtener valores de \$1942 por kilovatio, y por regulación del mercado eléctrico se activan las condiciones del precio de escasez de la energía, evento que en los últimos años solo se ha presentado en esta ocasión. También se puede observar que el Fenómeno de El Niño ocurrido en esta misma franja de tiempo ha sido el más prolongado de los años analizados.

Otro aspecto para resaltar es que entre mayo y julio de 2023 los valores del kilovatio también habían comenzado a subir, lo que también se relaciona con el fenómeno ENSO, el cual también marca la formación del Fenómeno de El Niño.

A continuación, se presentan la **Tabla 1** y tabla 2, que contiene los resultados de la estadística descriptiva de cada una de las variables, y la **Figura 3**, donde se grafican los datos históricos que se utilizaron para elaborar los modelos de predicción del valor del kilovatio en el mercado de energía eléctrica colombiano.

Tabla 1

Estadística descriptiva

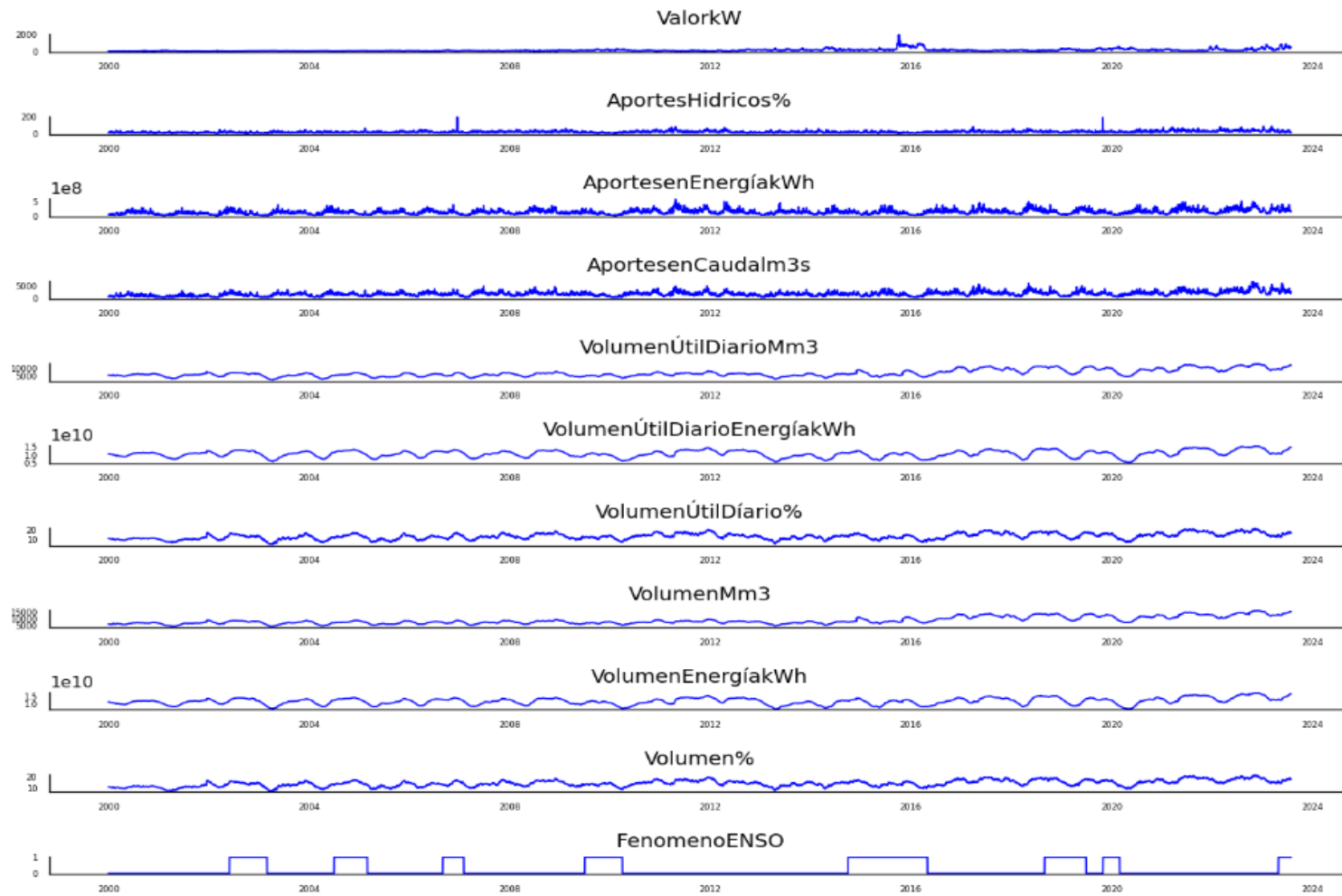
	Valor kW	Aportes hídricos%	Aportes en Energía kWh	Aportes en Caudal m ³ s	Volumen Útil Diario Mm ³
count	8613	8613	8613	8613	8613
mean	146,29	26,20	1,47E+08	1804,66	6664,16
std	143,08	10,78	7,44E+07	852,09	2178,59
min	28,84	8,45	2,53E+07	217,96	2229,87
25%	67,30	18,89	9,37E+07	1195,85	5156,06
50%	100,64	24,03	1,35E+08	1681,83	6101,58
75%	169,70	30,97	1,85E+08	2244,95	7821,77
max	1942,69	199,44	5,91E+08	6765,19	12677,13

Tabla 2*Estadística descriptiva (continuación)*

	Volumen Útil Diario Energía kWh	Volumen Útil Diario%	Volumen Mm ³	Volumen Energía kWh	Volumen %	Fenó- meno ENSO
count	8613	8613,00	8613,00	8613,00	8613,00	8613
mean	10745550000	13,37	8713,65	11988570000	14,61	-
std	2097450000	3,00	2647,75	2176855000	2,69	-
min	5405008000	5,11	4051,06	6907038000	7,84	-
25%	9344523000	11,07	6812,06	10471960000	12,53	-
50%	10933790000	13,33	7898,94	12092510000	14,61	-
75%	12382670000	15,56	10497,06	13602940000	16,51	-
max	15649800000	21,51	16260,15	17662670000	21,89	-

Figura 3

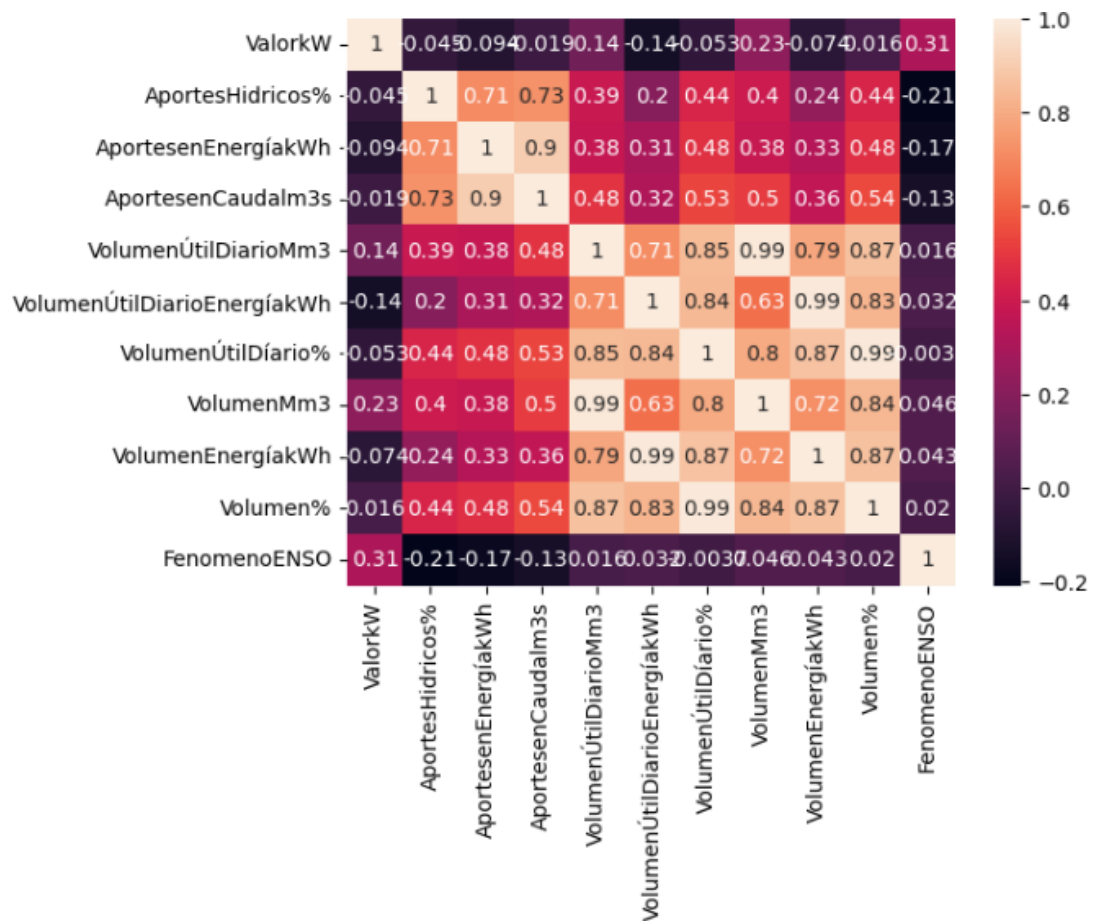
Datos históricos de las variables dependiente e independiente



Luego de analizar los datos, se hizo un análisis de correlación de las variables independientes con respecto a la dependiente, las cuales marcaron correlaciones negativas y positivas, tal como se muestra en la Figura 4.

Figura 4

Correlación de variables independientes y dependientes



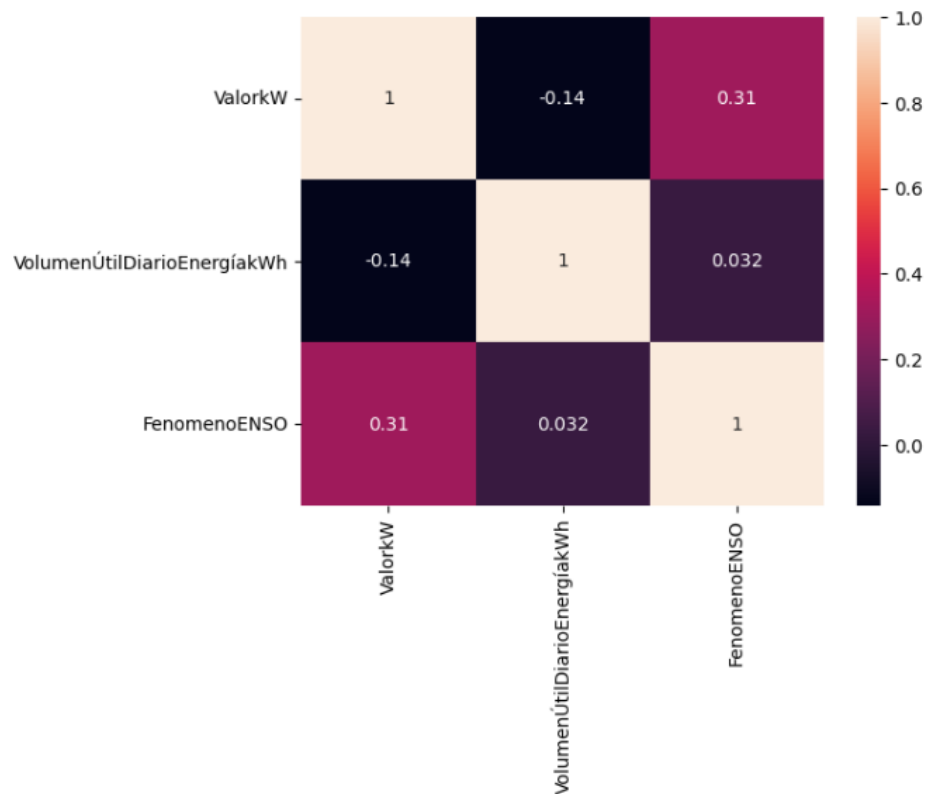
De la anterior correlación se tomaron las variables con mayor magnitud negativa y positiva, resultando el fenómeno ENSO con 0,31 (correlación positiva) y Volumen

Útil Diario Energía kWh con -0,14 (correlación negativa). Por tanto, se utilizan solo estas variables para la construcción de los modelos de predicción.

A continuación, se presenta la Figura 5, con la correlación de estas variables.

Figura 5

Correlación del Valor kW con Volumen útil diario energía y fenómeno ENSO



Con las variables definidas, se procede a realizar la prueba de Dickey-Fuller, donde se verificó que los datos no eran estacionarios, resultando el *p-value* por debajo de 0,05. Por tanto, no se les hizo transformación para construir los modelos (tabla 3).

Tabla 3*Resultados de la prueba de la prueba de Dickey-Fuller*

Variable	<i>p-value</i>
Valor kW	0,000015
Volumen Útil Diario en Energía kWh	0,000000
Fenómeno ENSO	0,000285

7.2. EVALUACIÓN DE MODELOS

En esta sección, se procede a evaluar los resultados de los modelos desarrollados en la investigación, los cuales se evalúan tomando como parámetros de medición el Error Cuadrático Medio de la Raíz (RMSE) y el Error Medio Absoluto (MAE).

7.3. RESULTADOS DEL MODELO DE REGRESIÓN LINEAL Y MULTIPLE

En la tabla 4, se aprecian los resultados de cada uno de los modelos de regresión lineal, de los cuales podemos concluir que ninguna de las variables independientes por sí sola logra explicar la variabilidad de la variable dependiente, la cual, en todos los casos, es el Valor del kW. Lo anterior se ve reflejado en los bajos valores de R-Cuadrado que tienen los modelos de regresión lineal.

Tabla 4*Resultados del modelo de regresión lineal*

Modelo	Variable dependiente	Variable independiente	R-Cuadrado
1	Valor kW	Aportes en Energía kWh	0,009
2	Valor kW	Volumen Útil Diario Mm ³	0,020
3	Valor kW	Volumen Útil Diario en Energía kWh	0,020
4	Valor kW	Fenómeno ENSO	0,096
5	Valor kW	Volumen Total Mm ³	0,051

Los resultados de la regresión múltiple se pueden observar a continuación en la tabla 5, en donde el valor del R-Cuadrado es de 0,119, siendo este valor mayor al valor de los R-Cuadrado de los modelos de regresión lineal, lo cual nos indica que bajo este modelo de regresión múltiple las variables independientes del Volumen Útil Diario en Energía (kWh) y fenómeno ENSO logran explicar un 11,9% de la variabilidad del Valor del KW. En este modelo de regresión múltiple, también se identifica que las variables independientes son estadísticamente significativas, al tener un *p-value* menor a 0,05.

Tabla 5*Resultados del modelo regresión múltiple*

Modelo	Variable dependiente	VARIABLES INDEPENDIENTES	R-Cuadrado
1	Valor kW	Volumen Útil Diario en Energía kWh	0,119
		Fenómeno ENSO	

En la tabla 6, se pueden observar los resultados de la regresión múltiple luego de incluir la variable *dummy* para los datos comprendidos entre el 15 de septiembre de 2015 y el 14 de abril de 2016, los cuales corresponden a las fechas de ocurrencia de un Fenómeno de El Niño fuerte presentado en ese período en Colombia. Se observa que la inclusión de la variable *dummy* mejoró el resultado del R-Cuadrado, el cual resulta en 0,452. En este modelo de regresión múltiple, todas las variables independientes, incluyendo la variable *dummy*, son estadísticamente significativas, al tener un *p-value* menor a 0,05.

Tabla 6*Resultados del modelo regresión múltiple incluyendo la variable dummy*

Modelo	Variable dependiente	VARIABLES INDEPENDIENTES	R-Cuadrado
1	Valor kW	Volumen Útil Diario en Energía kWh	0,452
		Fenómeno ENSO	

La evaluación del modelo se realiza sobre el último modelo de regresión múltiple con la inclusión de la variable *dummy*, donde se obtuvo como resultado de la predicción un RMSE de 105,93 COP/kWh y un MAE de 69,37 COP/kWh. Los valores obtenidos del RMSE y el MAE son comparados contra el promedio del precio real de la energía en bolsa para el período de análisis. El RMSE y el MAE representan un error de un 72,4 % y un 47,4%, respectivamente. Los resultados de la evaluación del modelo de regresión múltiple con la inclusión de la variable *dummy* se resumen a continuación en la Tabla 7.

Tabla 7

Resultados de la evaluación del modelo de regresión múltiple

Parámetro	Resultado (COP/kWh)	Error porcentual (%)
RMSE	105,93	72,41%
MAE	69,37	47,42%

7.4. RESULTADOS DEL MODELO VAR

En la evaluación del modelo VAR, se obtuvo como resultado un RMSE de 85,01 COP/kWh y un MAE de 78,16 COP/kWh. Los valores obtenidos del RMSE y del MAE, al compararlos contra el promedio del precio real de la energía en bolsa entre el 17 y el 31 de julio de 2023, representan un error de un 17,75% y un 16,32%, respectivamente.

En la Tabla 8, se muestran los resultados del modelo VAR. El ValorkWPred corresponde al pronóstico del modelo VAR, y el Valor kW, al valor real de la energía en bolsa, para el período comprendido entre el 17 y el 31 de julio de 2023.

Tabla 8

Resultados de la evaluación del modelo VAR

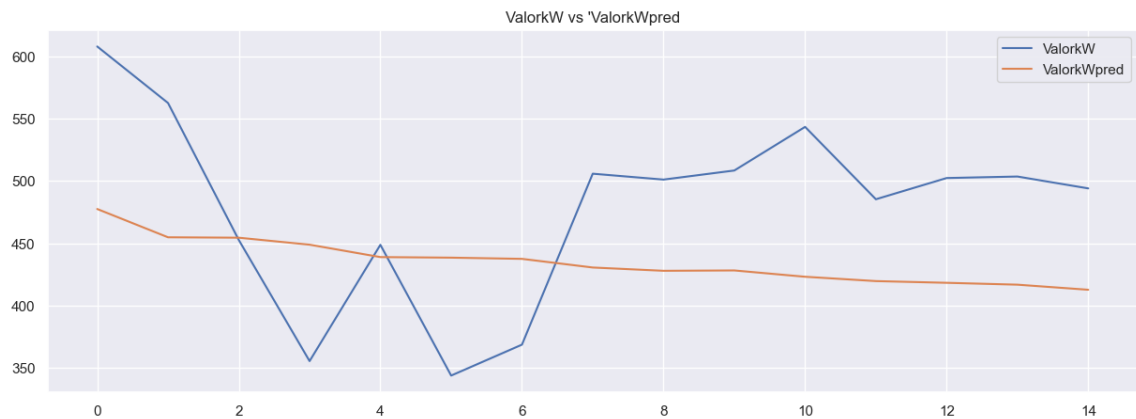
Fecha	ValorkWPred (COP/kWh)	Valor kW (COP/kWh)
31/07/2023	412,71	493,97
30/07/2023	416,78	503,48
29/07/2023	418,28	502,28
28/07/2023	419,65	485,26
27/07/2023	423,13	543,28
26/07/2023	428,19	508,31
25/07/2023	427,91	501,03
24/07/2023	430,54	505,75
23/07/2023	437,48	368,69
22/07/2023	438,45	343,91
21/07/2023	438,9	448,74
20/07/2023	448,84	355,52
19/07/2023	454,44	452,61
18/07/2023	454,77	562,45
17/07/2023	477,45	607,74

Parámetro	Resultado (COP/kWh)	Error porcentual (%)
RMSE	85,01	17,75%
MAE	78,16	16,32%

En la Figura 6, se observa el resultado gráfico del modelo de predicción VAR comparado contra los precios reales de la energía en bolsa.

Figura 6

Resultado grafico de predicción del modelo VAR



7.5. RESULTADOS DEL MODELO ARIMA

En la evaluación del modelo ARIMA (1,1,1), se obtuvo como resultado un RMSE de 153,27 COP/kWh y un MAE de 168,68 COP/kWh. Los valores obtenidos del RMSE y del MAE, al compararlos con el promedio del precio real de la energía en bolsa

entre el 17 y 31 de julio de 2023, representan un error de un 32,01% y un 35,22%, respectivamente.

En la Tabla 9, se muestran los resultados del modelo ARIMA (1,1,1), en donde el ValorkWPred corresponde al pronóstico del modelo ARIMA(1,1,1), y el Valor kW corresponde al valor real de la energía en bolsa, para las fechas comprendidas entre el 17 y el 31 de julio de 2023.

Tabla 9

Resultados de la evaluación del modelo ARIMA (1,1,1)

Fecha	ValorkWPred (COP/kWh)	Valor kW (COP/kWh)
31/07/2023	627,59	493,97
30/07/2023	628,44	503,48
29/07/2023	629,14	502,28
28/07/2023	628,44	485,26
27/07/2023	628,26	543,28
26/07/2023	628,51	508,31
25/07/2023	628,55	501,03
24/07/2023	628,14	505,75
23/07/2023	629,53	368,69
22/07/2023	630,10	343,91
21/07/2023	632,21	448,74

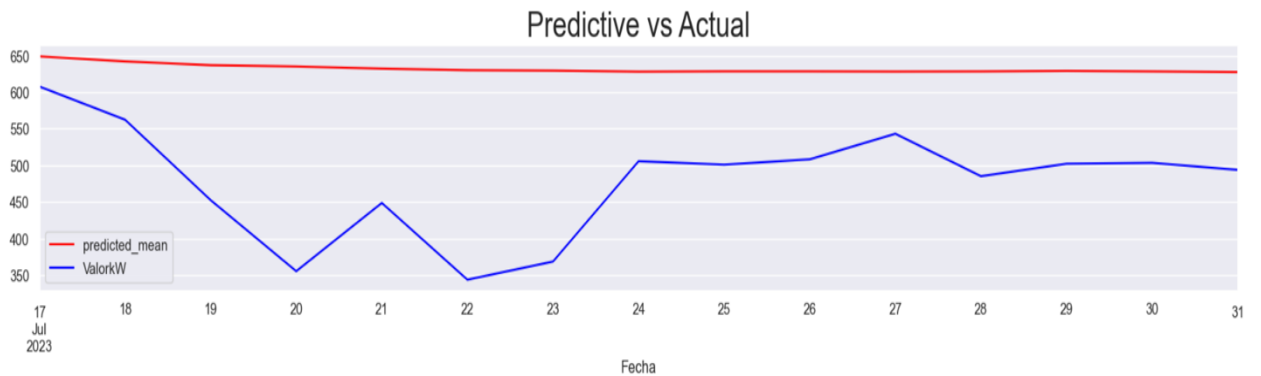
Fecha	ValorkWPred (COP/kWh)	Valor kW (COP/kWh)
20/07/2023	635,17	355,52
19/07/2023	636,92	452,61
18/07/2023	642,07	562,45
17/07/2023	648,97	607,74

Parámetro	Resultado (COP/kWh)	Error Porcentual (%)
RMSE	168,68	35,22%
MAE	153,27	32,01%

En la Figura 7, se observa el resultado gráfico del modelo de predicción ARIMA (1,1,1) comparado contra los precios reales de la energía en bolsa.

Figura 7

Resultado gráfico de predicción del modelo ARIMA (1,1,1)



7.6. RESULTADOS DEL MODELO ARIMAX

7.6.1. ARIMAX – ARIMA (1,1,1) con variables exógenas Volumen Útil Diario expresado en términos de Energía (kWh) y el fenómeno ENSO

En la evaluación del primer modelo ARIMAX con orden ARIMA (1,1,1), se obtuvo como resultado un RMSE de 153,27 COP/kWh y un MAE de 168,68 COP/kWh. Los valores obtenidos del RMSE y el MAE, al compararlos con el promedio del precio real de la energía en bolsa entre el 17 y 31 de julio de 2023, representan un error de un 32,01% y un 35,22%, respectivamente.

En la Tabla 10, se muestran los resultados del primer modelo ARIMAX con orden ARIMA (1,1,1), en donde ValorkWPred corresponde al pronóstico del modelo ARIMAX (1,1,1), y el Valor kW corresponde al valor real de la energía en bolsa para las fechas comprendidas entre el 17 y el 31 de julio de 2023.

Tabla 10

Resultados de la evaluación de modelo ARIMAX con ARIMA orden (1,1,1)

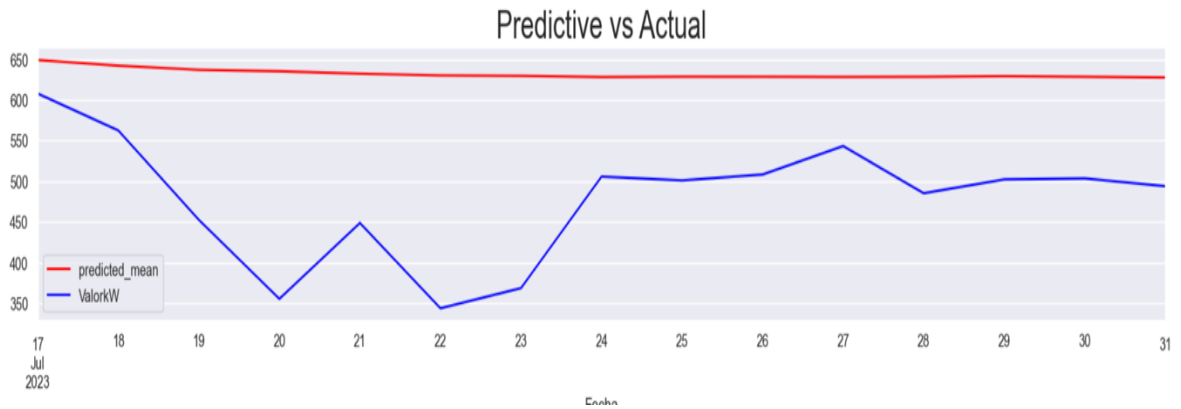
Fecha	ValorkWPred (COP/kWh)	Valor kW (COP/kWh)
17/07/23	648,97	607,74
18/07/23	642,07	562,45
19/07/23	636,92	452,61

Fecha	ValorkWPred (COP/kWh)	Valor kW (COP/kWh)
20/07/23	635,17	355,52
21/07/23	632,21	448,74
22/07/23	630,10	343,91
23/07/23	629,53	368,69
24/07/23	628,14	505,75
25/07/23	628,55	501,03
26/07/23	628,51	508,31
27/07/23	628,26	543,28
28/07/23	628,44	485,26
29/07/23	629,14	502,28
30/07/23	628,44	503,48
31/07/23	627,59	493,97
Parámetro	Resultado (COP/kWh)	Error Porcentual (%)
RMSE	168,68	35,22%
MAE	153,27	32,01%

En la Figura 8, se observa el resultado gráfico del modelo de predicción ARIMAX (1,1,1) comparado contra los precios reales de la energía en bolsa.

Figura 8

Resultado gráfico de predicción del modelo ARIMAX con ARIMA orden (1,1,1)



7.6.2. ARIMAX – ARIMA (2,1,1) con variables exógenas Volumen Útil Diario expresado en términos de Energía (kWh) y el fenómeno ENSO

En la evaluación del segundo modelo ARIMAX con orden ARIMA (2,1,1), se obtuvo como resultado un RMSE de 160,00 COP/kWh y un MAE de 144,12 COP/kWh. Los valores obtenidos del RMSE y el MAE, al compararlos con el promedio del precio real de la energía en bolsa entre el 17 y el 31 de julio de 2023, representan un error de un 33,41% y un 30,10%, respectivamente.

En la Tabla 11, se muestran los resultados del segundo modelo ARIMAX con orden ARIMA (2,1,1), en donde ValorkWPred corresponde al pronóstico del modelo ARIMAX (2,1,1), y el Valor kW corresponde al valor real de la energía en bolsa para las fechas comprendidas entre el 17 y el 31 de julio de 2023.

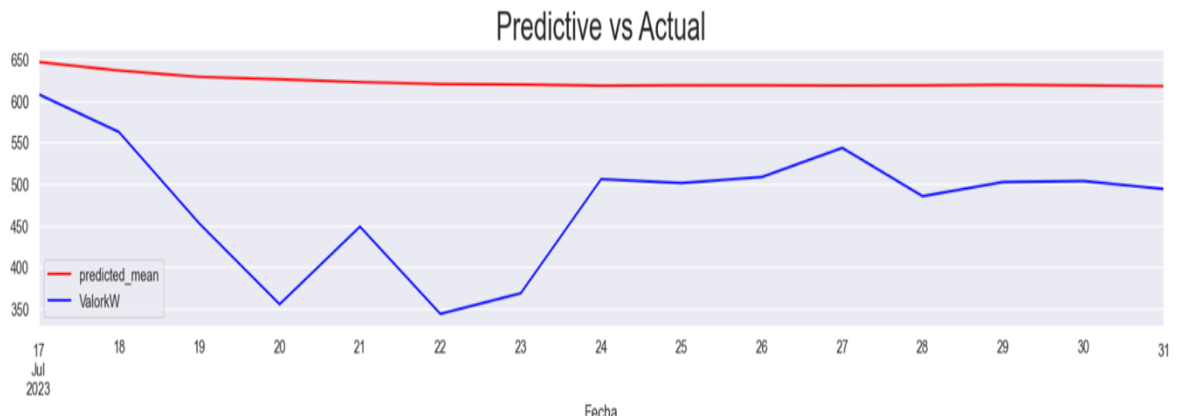
Tabla 11*Resultados de la evaluación de modelo ARIMAX con ARIMA orden (2, 1, 1)*

Fecha	ValorkWPred (COP/kWh)	Valor kW (COP/kWh)
17/07/23	646,45	607,74
18/07/23	636,20	562,45
19/07/23	628,52	452,61
20/07/23	625,64	355,52
21/07/23	622,27	448,74
22/07/23	620,03	343,91
23/07/23	619,44	368,69
24/07/23	618,04	505,75
25/07/23	618,45	501,03
26/07/23	618,41	508,31
27/07/23	618,17	543,28
28/07/23	618,35	485,26
29/07/23	619,05	502,28
30/07/23	618,35	503,48
31/07/23	617,49	493,97
Parámetro	Resultado (COP/kWh)	Error Porcentual (%)
RMSE	160,00	33,41%
MAE	144,12	30,10%

En la Figura 9, se observa el resultado gráfico del modelo de predicción ARIMAX (2,1,1) comparado contra los precios reales de la energía en bolsa.

Figura 9

Resultado gráfico de predicción del modelo ARIMAX con ARIMA orden (2,1,1)



7.6.3. ARIMAX – ARIMA (0,1,0) con variables exógenas Volumen Útil Diario expresado en términos de Energía (kWh) y del fenómeno ENSO

En la evaluación del tercer modelo ARIMAX con orden ARIMA (0,1,0), se obtuvo como resultado un RMSE de 171,50 COP/kWh y un MAE de 158,31 COP/kWh. Los valores obtenidos del RMSE y MAE, al compararlos contra el promedio del precio real de la energía en bolsa entre el 17 y el 31 de julio de 2023, representan un error de un 35,81% y un 33,06%, respectivamente.

En la Tabla 12, se muestran los resultados del tercer modelo ARIMAX con orden ARIMA (0,1,0), en donde ValorkWPred corresponde al pronóstico del modelo ARIMAX (0,1,0), y el Valor kW corresponde al valor real de la energía en bolsa para las fechas comprendidas entre el 17 y el 31 de julio de 2023.

Tabla 12

Resultados de la evaluación del modelo ARIMAX con ARIMA orden (0,1,0)

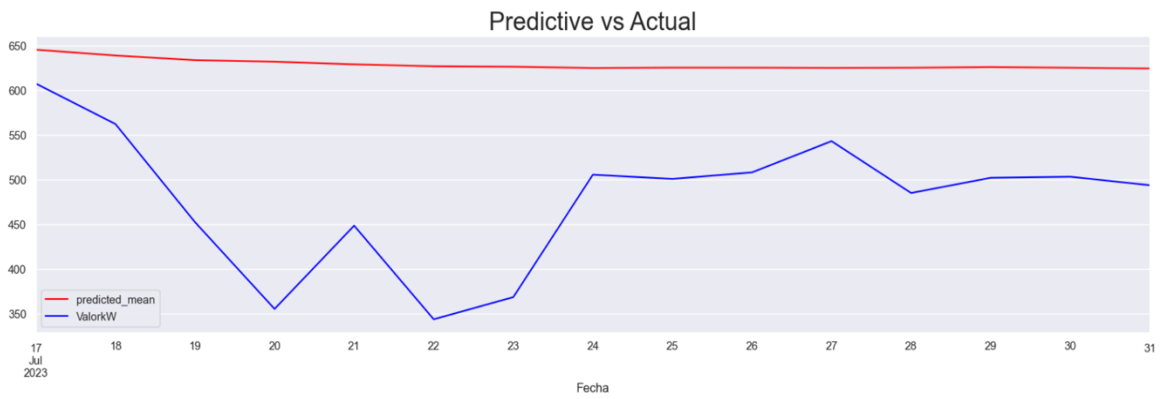
Fecha	ValorkWPred (COP/kWh)	Valor kW (COP/kWh)
17/07/23	645,66	607,74
18/07/23	639,14	562,45
19/07/23	633,95	452,61
20/07/23	632,21	355,52
21/07/23	629,24	448,74
22/07/23	627,13	343,91
23/07/23	626,57	368,69
24/07/23	625,18	505,75
25/07/23	625,59	501,03
26/07/23	625,55	508,31
27/07/23	625,30	543,28
28/07/23	625,48	485,26
29/07/23	626,18	502,28
30/07/23	625,48	503,48

Fecha	ValorkWPred (COP/kWh)	Valor kW (COP/kWh)
31/07/23	624,62	493,97
Parámetro	Resultado (COP/kWh)	Error Porcentual (%)
RMSE	171,50	35,81%
MAE	158,31	33,06%

En la Figura 10, se observa el resultado gráfico del modelo de predicción ARIMAX (0,1,0) comparado contra los precios reales de la energía en bolsa.

Figura 10

Resultado gráfico de predicción del modelo ARIMAX con ARIMA orden (0,1,0)



7.7. RESULTADOS MODELO SARIMAX

7.7.1. Modelo SARIMAX (1,1,1) x (1,1,1,12)

En la evaluación del modelo SARIMAX (1,1,1) x (1,1,1,12), se obtuvo como resultado un MAE de 88,46 COP/kWh. Los datos para el pronóstico se tomaron a partir del período que va del 1 de enero de 2022 al 31 de julio de 2023. Los valores obtenidos del MAE, al compararlos contra el promedio del precio real de la energía en bolsa entre el 1 de enero de 2022 y el 31 de julio de 2023, representan un error del 28,22%.

En la Tabla 13, se muestran los resultados del modelo SARIMAX (1,1,1) x (1,1,1,12).

Tabla 13

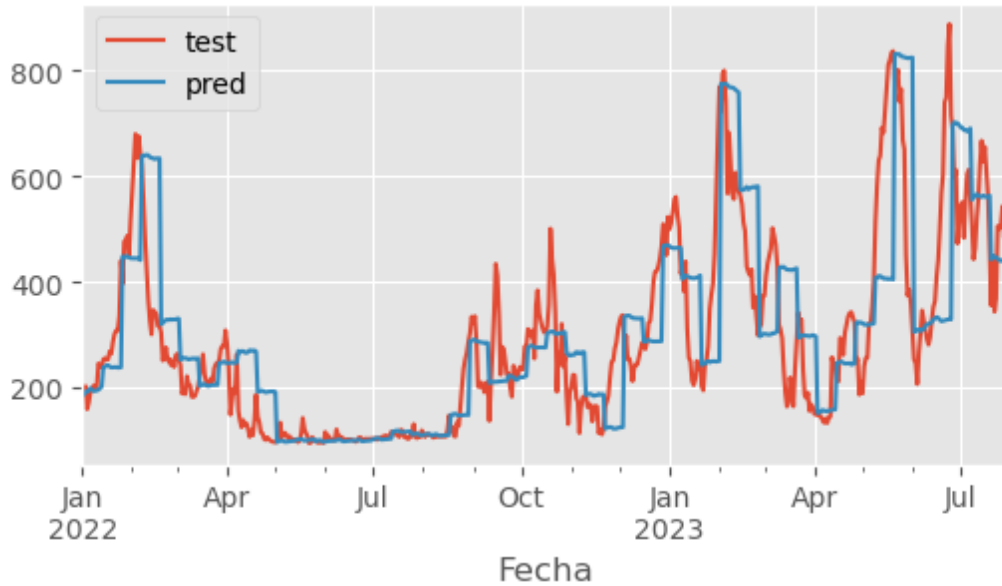
Resultados evaluación de modelo SARIMAX (1,1,1)x(1,1,1,12)

Conjunto de datos para entrenamiento, del 01/01/2000 al 31/12/2021		
Conjunto de datos para prueba de pronóstico, del 01/01/2022 al 31/07/2023		
Parámetro	Resultado (COP/kWh)	Error Porcentual (%)
MAE	82,46	28,22%

En la Figura 11, se observa el resultado gráfico del *backtesting* realizado sobre la predicción del precio de la energía en bolsa contra el precio real para las fechas comprendidas entre el 1 de enero de 2022 y el 31 de julio de 2023.

Figura 11

Resultado gráfico del backtesting de predicción del modelo SARIMAX (1,1,1) x (1,1,1,12)



7.7.2. Modelo SARIMAX (0,1,0) x (0,0,0,0)

Luego de hacer un *backtesting* y un *grid search* para encontrar la mejor combinación de hiperparámetros, se obtiene un modelo SARIMAX de orden (0,1,0) x (0,0,0,0), en el cual se obtuvo como resultado un MAE de 111,95 COP/kWh. Los datos para el pronóstico se tomaron a partir del período 1 de enero 2023 al 31 de julio de 2023. En la Tabla 14, se muestran los resultados del modelo SARIMAX (0,1,0) x (0,0,0,0).

Tabla 14

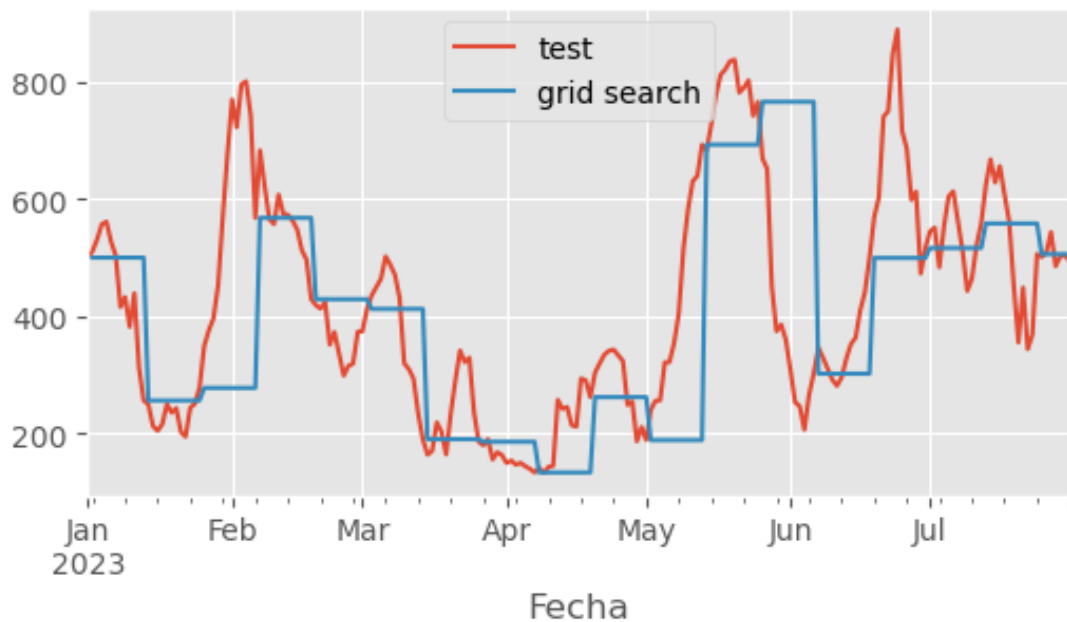
Resultados de la evaluación de modelo SARIMAX (0,1,0) x (0,0,0,0)

Conjunto de datos para entrenamiento, del 01/01/2000 al 31/12/2022		
Conjunto de datos para prueba de pronóstico, del 01/01/2023 al 31/07/2023		
Parámetro	Resultado (COP/kWh)	Error Porcentual (%)
MAE	111,96	26,42%

En la figura 12, se observa el resultado gráfico del *backtesting* realizado sobre la predicción del precio de la energía en bolsa contra el precio real para las fechas comprendidas entre el 1 de enero y el 31 de julio de 2023.

Figura 12

Resultado gráfico del backtesting de predicción del modelo SARIMAX (0,1,0) x (0,0,0,0)



7.8. Resumen de los resultados de los modelos desarrollados

En la Tabla 15, se presentan los resultados de los ocho modelos desarrollados para predecir el precio de la energía en bolsa en el mercado de Colombia, los cuales se evaluaron en términos de su RMSE y MAE.

Tabla 15

Resumen de la evaluación de resultados de los modelos desarrollados

Modelo	Regresión Múltiple		VAR		ARIMA	
Parámetro	COP/kWh	Error Porcentual	COP/kWh	Error Porcentual	COP/kWh	Error Porcentual
RMSE	105,93	72,41%	85,01	17,75%	168,68	35,22%
MAE	69,37	47,42%	78,16	16,32%	153,27	32,01%
Modelo	ARIMAX - ARIMA (1,1,1)		ARIMAX - ARIMA (2,1,1)		ARIMAX - ARIMA (0,1,0)	
Parámetro	COP/kWh	Error Porcentual	COP/kWh	Error Porcentual	COP/kWh	Error Porcentual
RMSE	168,68	35,22%	160,00	33,41%	171,50	35,81%
MAE	153,27	32,01%	144,12	30,10%	158,31	33,06%

Modelo	SARIMAX		SARIMAX	
	(1,1,1)x(1,1,1,12)		(0,1,0)x(0,0,0,0)	
Parámetro	COP/kWh	Error	COP/kWh	Error
		Porcentual		Porcentual
MAE	82,46	28,22%	111,96	26,42%

Finalmente, para la selección del modelo más adecuado en términos de su predicción y eficacia, se procede a analizar diferentes horizontes de tiempo para la predicción, considerando diversos enfoques para que los inversionistas en el sector de la generación de energía en Colombia tengan elementos que les permitan usar estos modelos de forma correcta.

- Para un enfoque de planeación financiera de corto plazo, en este caso, días, el modelo que presenta los mejores resultados es el VAR. En términos de evaluación, este modelo presenta un RMSE de 85,01 COP/kWh y un MAE de 78,16 COP/kWh, los cuales, al compararlos contra los precios reales del KW en las fechas de análisis del modelo, representan un error de un 17,75% y un 16,32%, respectivamente.

La planeación financiera de corto plazo es sumamente importante para los inversionistas, ya que les permitirá una mejor gestión de su liquidez, con el fin

obtener proyecciones de flujo de caja que les permitan cumplir sus obligaciones de corto plazo.

- Desde una visión de mediano plazo, en este caso, meses, el modelo que presenta los mejores resultados es SARIMAX (0,1,0)x(0,0,0,0). En términos de evaluación, este modelo presenta un MAE de 111,96 COP/kWh, el cual, al compararlo contra los precios reales del KW en las fechas de análisis del modelo, representa un error del 26,4%.

La planeación financiera de mediano plazo es de gran importancia para los inversionistas, ya que les permitirá tener un mejor control de su flujo de caja de cara a los presupuesto, la planificación tributaria y la dirección estratégica del negocio para sus proyectos e inversiones de mediano y largo plazo.

- Para una proyección de largo plazo, en este caso, un horizonte mayor a un año, el modelo con los mejores resultados es el de la Regresión Múltiple incluyendo la variable dummy para los períodos del Fenómeno del Niño fuerte ocurrido entre 2015 y 2016. Este modelo presenta un R-Cuadrado de 0,452 y, en términos de evaluación, presenta un RMSE de 105,93 COP/kWh y un MAE de 69,37 COP/kWh al compararlos contra los precios reales del KW para todo el período de análisis.

La planeación financiera de largo plazo es vital para los inversionistas, ya que les permite definir su visión estratégica de crecimiento futuro.

Cabe destacar que los modelos ARIMA y ARIMAX, aunque no tuvieron los mejores resultados frente a los otros modelos analizados, indican que se pueden utilizar teniendo claro que en la predicción presenta error entre un 30% y un 35%.

8. CONCLUSIONES

Los modelos econométricos de predicción constituyen una potente herramienta para los objetivos del presente trabajo de investigación, el cual tiene como finalidad estimar el precio de la energía en bolsa en el mercado de Colombia. Dicha estimación del precio de la energía servirá como insumo para que los inversionistas del sector puedan proyectar sus ingresos y así poder hacer una planificación financiera efectiva.

Para crear los modelos de predicción, se ha generado una base de datos integral y sólida que ha funcionado como punto de partida para el desarrollo de los modelos de predicción del precio de la energía en la bolsa de Colombia. Para la construcción de dicha base de datos, se han tomado insumos de fuentes oficiales tales como la base de datos Sinergox de XM (2023a) y la NOAA (2023).

Con la base de datos construida, se ejecutó un análisis estadístico descriptivo, en el cual, mediante la aplicación de técnicas estadísticas, se logró identificar el alto impacto del factor climático sobre el precio de la energía en bolsa, y que se puede evidenciar claramente en la Figura 3, donde se observa un incremento del precio del kW en los años 2015 y 2016, como producto del fuerte Fenómeno de El Niño que tuvo lugar en Colombia en ese período. Asimismo, que es un fenómeno climatológico que afecta el mercado debido a la alta participación de generación hidroeléctrica en la matriz de generación de Colombia. Según el informe de Figueroa

Castro, A.C, y Mojica, J. L. Informe Perspectiva Sectorial – Energía. Actualidad del sector energético colombiano. Corficolombiana (2023) a 2022 la hidroeléctrica representó el 86,66% de la generación de energía del país.

La serie de datos cuenta con 8613 datos, que resultan de un espacio de tiempo entre enero de 2000 y julio de 2023 y de la eliminación de datos faltantes. Se resalta que la desviación estándar del valor de kW es de \$143,08 y que su dato máximo, de \$1.942,69, se presentó durante el Fenómeno de El Niño más extremo que se ha registrado en los últimos 23 años.

Luego de aplicar la correlación entre las variables, finalmente se toman las variables independientes con las correlaciones más fuertes con las variables dependientes, las cuales son: con correlación negativa, el Volumen útil diario en energía kWh, y correlación positiva, el fenómeno ENSO. Esto es coherente, ya que, a mayor volumen útil diario en los embalses, menor es el valor del kW; sin embargo, si se presenta el Fenómeno de El Niño, aumenta el valor del kW.

El presente trabajo de investigación arrojó como resultado el desarrollo de ocho modelos de predicción, mediante la aplicación de diferentes técnicas econométricas, así de como la utilización del sistema de programación Python, por medio del cual se generaron modelos de regresión múltiple, VAR, ARIMA, ARIMAX con diferentes órdenes en sus componentes de autorregresión, integración y media móvil, así como modelos SARIMAX con distintos órdenes en sus componentes

ARIMA y estacionales, los cuales se optimizaron mediante la implementación de procedimientos de backtesting y grid search.

Los modelos desarrollados se evaluaron por medio de los parámetros RMSE y MAE, para determinar la capacidad de cada modelo para predecir el Valor del kW de la energía en bolsa. Podemos decir que, por su función de cálculo, el RMSE castiga de manera más fuerte los errores grandes producto de datos atípicos, mientras que el MAE trata todos los errores de la misma manera. En vista de que la base de datos cuenta con valores del kW alejados hasta en 13,5 desviaciones estándar de la media, producto de la ocurrencia de Fenómeno de El Niño fuerte en el período 2015 y 2016, en todos los modelos el RMSE arroja errores superiores al MAE.

De la selección del modelo que mejor se ajuste para la predicción del valor del kW, el inversionista debe tener en cuenta el horizonte de tiempo que desee analizar para su valoración, indicando que, para la planeación a corto plazo, se ajusta mejor el modelo VAR; para la planeación a mediano plazo, el SARIMAX, y para la planeación a largo plazo, la Regresión Múltiple.

De los resultados obtenidos en los modelos Regresión Múltiple, VAR y SARIMAX, se concluye que, si suponemos que los datos tienen un comportamiento como una distribución normal, todos los modelos se encuentran dentro de una desviación

estándar de los datos; es decir, que la probabilidad de ocurrencia es del 68%, lo cual indica que tienen un buen comportamiento en la predicción.

La presente investigación permite obtener modelos de predicción del precio de la energía en bolsa de Colombia, los cuales se pueden ajustar a las necesidades de planificación financiera de los inversionistas del sector en diferentes horizontes de tiempo.

Como recomendaciones para futuros trabajos de investigación en este campo, se propone explorar otras técnicas de predicción que permitan mejorar el desempeño de los modelos de largo plazo, con la incorporación de técnicas tales como redes neuronales, máquinas de soporte vectorial (SVM) o modelos de regresión *ridge*.

REFERENCIAS BIBLIOGRÁFICAS

- ARIMA con IGARCH. *Revista de Economía del Rosario*, 20(1), 127-161.
<https://doi.org/10.12804/revistas.urosario.edu.co/economia/a.6152>
- Asamblea Nacional Constituyente (1991). Constitución Política de Colombia. *Gaceta Constitucional*, 114. <https://www.suin-juriscol.gov.co/viewDocument.asp?ruta=Constitucion/1687988>
- Avellaneda G., J. A., Ochoa R., C. M., y Figueroa García, J. C. (2012). Comparación entre un sistema neuro difuso auto organizado y un modelo ARIMAX en la predicción de series económicas volátiles. *Ingeniería*, 17(12), 26-34.
<https://dialnet.unirioja.es/servlet/articulo?codigo=4688810>
- Barrientos Marín, J., y Toro Martínez, M. (2016). La hidrología como predictor del comportamiento del precio de energía en bolsa. *Perfil de Coyuntura Económica*, 25, 25-140. <https://doi.org/10.17533/udea.pece.n25a07>
- Bello-Rodríguez, S. P., y Beltrán-Ahumada, R. B. (2013). Caracterización y pronóstico del precio spot de la energía eléctrica en Colombia. *Revista de la Maestría en Derecho Económico*, 6(6), 293-316.
<https://revistas.javeriana.edu.co/index.php/revmaescom/article/view/7174>
- Brooks, C. (2019). *Introductory Econometrics for Finance* (4th ed.). Cambridge University Press. <https://doi.org/10.1017/9781108524872>
- Fathi, M. M., Awadallah, A. G., Abdelbaki, A. M., & Haggag, M. (2019). A new Budyko framework extension using time series SARIMAX model. *Journal of Hydrology*, 570, 827-838. <https://doi.org/10.1016/j.jhydrol.2019.01.037>

- Figueroa Castro, A.C, y Mojica, J. L (28 de febrero, 2023). Informe Perspectiva Sectorial – Energía. Actualidad del sector energético colombiano. Corficolombiana. <https://cutt.ly/vwQLCyJM>
- Galindo Ortiz, A. F. (2017). *Modelamiento de los precios de la energía en bolsa en Colombia incorporando el efecto del ENSO* [trabajo de Grado, Universidad de los Andes]. Repositorio Institucional. <http://hdl.handle.net/1992/13633>
- IBM (17 de agosto, 2021). Conceptos básicos de ayuda de CRISP-DM. *Guía de CRISP-DM de IBM SPSS Modeler*. <https://cutt.ly/kwQLTEvy>
- James, G., Witten, D., Hastie, T., Tibshirani, R., & Taylor, J. (2023). *An Introduction to Statistical Learning - with Applications in Python*. Springer.
- Mahadeva, L., y Robinson, P. (2009). *Ensayo 76. Prueba de raíz para ayudar a la construcción de un modelo*. CEMLA. <https://www.cemla.org/PDF/ensayos/pub-en-76.pdf>
- Mirakyan, A., Meyer-Renschhausen, M., & Koch, A. (2017). Composite forecasting approach, application for next-day electricity price forecasting. *Energy Economics*, 66, 228-237. <https://doi.org/10.1016/j.eneco.2017.06.020>
- Monteiro, C., Ramirez-Rosado, I. J., Fernandez-Jimenez, L. A., & Ribeiro, M. (2018). New probabilistic price forecasting models: Application to the Iberian electricity market. *International Journal of Electrical Power and Energy Systems*, 103, 483-496. <https://doi.org/10.1016/j.ijepes.2018.06.005>
- Muñoz-Santiago, A., Urquijo-Vanstrahlengs, J., Castro-Otero, A., y Lombana, J. (2017). Pronóstico del precio de la energía en Colombia utilizando modelos

- National Oceanic and Atmospheric Administration – NOAA (July 10, 2023). Cold & Warm Episodes by Season. *Index ONI*. https://origin.cpc.ncep.noaa.gov/products/analysis_monitoring/ensostuff/ONI_v5.php
- Schöniger, F., & Morawetz, U. B. (2022). What comes down must go up: Why fluctuating renewable energy does not necessarily increase electricity spot price variance in Europe. *Energy Economics*, 111, 106069. <https://doi.org/10.1016/j.eneco.2022.106069>
- Shah, I., Akbar, S., Saba, T., Ali, S., & Rehman, A. (2021). Short-Term Forecasting for the Electricity Spot Prices with Extreme Values Treatment. *IEEE Access*, 9, 105451–105462. <https://doi.org/10.1109/ACCESS.2021.3100076>
- Tehrani, S., Juan, J., & Caro, E. (2022). Electricity Spot Price Modeling and Forecasting in European Markets. *Energies*, 15(16), 5980. <https://doi.org/10.3390/en15165980>
- Unidad de Planeación Minero-Energética – UPME (2020). *Plan de expansión de referencia generación – transmisión 2020 – 2034*. (Vol. 1). Ministerio de Minas y Energía. <https://cutt.ly/9wQZezl9>
- Unidad de Planeación Minero-Energética – UPME (2023). *Proyección de la demanda de energía eléctrica y potencia máxima 2023-2037*. Ministerio de Minas y Energía. <https://shorturl.at/ft035>
- Urbano Buriticá, S. S., y González Pérez, (2022). *Proyección de corto plazo para el precio de bolsa de energía en el mercado colombiano* [tesis de Maestría,

Universidad de los Andes]. Repositorio Institucional.

<https://repositorio.uniandes.edu.co/handle/1992/63441>

Villa-Loaiza, C., Taype-Huaman, I., Benavides-Franco, J., Buenaventura-Vera, G., & Carabalí-Mosquera, J. (2023). Does climate impact the relationship between the energy price and the stock market? The Colombian case.

Applied Energy, 336. <https://doi.org/10.1016/j.apenergy.2023.120800>

XM (25 de julio, 2023a). Históricos. *Sinergox*.

<https://sinergox.xm.com.co/hdrlg/Paginas/Historicos/Historicos.aspx>

XM (25 de julio, 2023b). *Estructura del mercado*.

<https://www.xm.com.co/transacciones/registros/registro-agentes-y-contactos/estructura-del-mercado>

Zhang, X., Wang, J., & Gao, Y. (2019). A hybrid short-term electricity price forecasting framework: Cuckoo search-based feature selection with singular spectrum analysis and SVM. *Energy Economics*, 81, 899-913.

<https://doi.org/10.1016/j.eneco.2019.05.026>