Vigilada Mineducación

MÉTODOS DE MACHINE LEARNING CON ALGORITMOS DE CLÚSTER NO SUPERVISADOS, UNA ALTERNATIVA DE SEGMENTACIÓN DE LAS PYMES COLOMBIANAS PARA PLANTEAR ESTRATEGIAS DE ACUERDO CON SUS CONDICIONES ECONÓMICAS

DURLEY YALILE RAMÍREZ MENDOZA

Tesis presentada como requisito parcial para obtener el título de

Magíster en Administración Financiera

Asesor

César Augusto Orozco Echeverry, Mag.

UNIVERSIDAD EAFIT

ESCUELA DE ECONOMÍA Y FINANZAS

MAESTRÍA EN ADMINISTRACIÓN FINANCIERA – MAF

BOGOTÁ

2022

Dedicatoria

A mi pareja, Daren Rodríguez, quien siempre ha creído en mí. Gracias por su apoyo incondicional, paciencia y esfuerzo para la culminación de esta meta.

Agradecimientos

A mis padres, Felipe Ramírez y Teofilde Mendoza, quienes con su ejemplo me han demostrado el significado de la palabra esfuerzo y dedicación. A mi hermano Leonardo Ramírez por su colaboración durante el desarrollo de la maestría. Finalmente, a mi asesor de tesis, por su paciencia, disponibilidad, correcciones y confianza en esta investigación.

RESUMEN

Esta investigación creó una nueva alternativa de agrupación utilizando herramientas de machine learning como son los modelos K-means y el clúster aglomerativo, tomando como base la información financiera de 2016 a 2019 de 10.001 pymes colombianas. A partir de estos modelos, se originaron doce clústeres que tienen el 98,44% de los datos evaluados y se determinó que el modelo que presentó mejores resultados de agrupación fue el modelo aglomerativo, el cual genera los siguientes grupos principales: un primer grupo con márgenes negativos y un endeudamiento superior al 61%, un segundo grupo que inicia con un intervalo entre el -10% hasta el 40% de sus márgenes y un endeudamiento inferior al 60%, y un tercer grupo con márgenes positivos y un endeudamiento entre el 11 y 80%. Finalmente, desde estos grupos se crean estrategias de acuerdo con las condiciones económicas de cada uno de ellos.

Palabras clave: pymes, machine learning, algoritmos no supervisados, clúster, clúster K-means, clúster aglomerativo.

ABSTRACT

This research created a new grouping alternative using machine learning tools such as K-means and agglomerative clustering models, based on financial information from 2016 to 2019 of 10,001 Colombian SMEs. From these models twelve clusters originated that have 98.44% of the evaluated data and it was determined that the model that presented the best clustering result was the agglomerative model which generates the following main groups: a first group with negative margins and a debt exceeding 61%, a second group starting with a range between -10% to 40% of its margins and a debt below 60%, and a third group with positive margins and a debt between 11 and 80%. Finally, these groups create strategies according to the economic conditions of each of them.

Keywords: SMEs, machine learning, unsupervised algorithms, cluster, K-means clustering, agglomerative clustering.

Contenido

1. INTRODUCCIÓN 10

2. MARCO DE REFERENCIA 13

2.1 Pymes 13

2.2 Machine learning 14

3. METODOLOGÍA 17

3.1 Base de datos 18

3.2 Modelos de machine learning con algoritmos no supervisados 19

3.3 Análisis de los resultados de agrupación 32

4. RESULTADOS 36

4.1 Análisis de los resultados del modelo de componentes principales PCA 37

4.2 Análisis de los resultados obtenidos del modelo de machine learning K-means y aglomerativo 39

5. FORMULACIÓN DE ESTRATEGIAS 52

5.1 Estrategias generales 52

5.2 Estrategias específicas de acuerdo con la necesidad de cada clúster 53

6. CONCLUSIONES Y RECOMENDACIONES 54

7. REFERENCIAS 56

8. ANEXOS 60

Lista de figuras

Figura 1. Metodología desarrollo investigación 17
Figura 2. Depuración base de datos 19
Figura 3. Ejemplo del método PCA para un set de datos 21
Figura 4. Representación gráfica de los eigenvectores y eigenvalores 24
Figura 5. Ejemplo de código y librerías en Python para generar los componentes principales 24
Figura 6. Número de componentes principales vs. varianza explicada 25
Figura 7. Paso a paso del agrupamiento con el método K-means 27
Figura 8. Método del codo para identificar el número correcto de clústeres K 28
Figura 9. Ejemplo del Código en Python y librerías para el método K-means 29
Figura 10. Ejemplo de la formación de los grupos de clústeres de forma iterativa usando el método aglomerativo 30
Figura 11. Ejemplo del Código en Python y librerías para el método aglomerativo 31
Figura 12. Desarrollo del modelo 32
Figura 13. Agrupación con cifras vs. agrupación PCA 36
Figura 14. Modelo PCA 37
Figura 15. Correlación PCA con los indicadores 38
Figura 16. K-means, ingresos vs. clústeres 42
Figura 17. Aglomerativo ingresos vs. clústeres 42
Figura 18. K-means, margen bruto vs. clústeres 43
Figura 19. Aglomerativo, margen bruto vs. clústeres 43
Figura 20. K-means, margen operativo vs. clústeres 44
Figura 21. Aglomerativo, margen operativo vs. clústeres 44
Figura 22. K-mean, margen neto vs. clústeres 45
Figura 23. Aglomerativo, margen neto vs. clústeres 45
Figura 24. K-means, activo vs. clústeres 46
Figura 25. Aglomerativo, activo vs. clústeres 46
Figura 26. K-means, ROA vs. clústeres 47
Figura 27. Aglomerativo, ROA vs. clústeres 47
Figura 28. K-means, ROE vs. clústeres 47

Figura 29. Aglomerativo, ROE vs. clústeres.....47
Figura 30. K-means, ebitda vs. clústeres.....48
Figura 31. Aglomerativo, ebitda vs. clústeres.....48
Figura 32. K-means, índice de endeudamiento vs. clústeres 49
Figura 33. Aglomerativo, índice de endeudamiento vs. clústeres.....49

Lista de tablas

Tabla 1. Aporte de las pymes a la economía 13
Tabla 2. Evolución machine learning 14
Tabla 3. Conceptos 33
Tabla 4. Indicadores de solvencia 34
Tabla 5. Indicadores de rentabilidad 34
Tabla 6. Otros indicadores 35
Tabla 7. Número de componentes principales PCA y su acumulado de varianza 37
Tabla 8. Resultado clústeres-algoritmo K-means 39
Tabla 9. Resultado clústeres-algoritmo aglomerativo 40
Tabla 10. Direccionamiento subcapítulos y número de anexo 42
Tabla 11. Estrategias según resultados del modelo de machine learning aglomerativo 52

1. INTRODUCCIÓN

Las pymes colombianas tienen un papel fundamental en la economía del país y "representan más del 90% del sector productivo nacional" ("Mipymes representan más de 90%del sector productivo nacional y generan el 80% del empleo en Colombia", 2019). La ley 905 (República de Colombia, 2004) clasificó las pymes de acuerdo con su número de empleados y valor de activos. Además, el Ministerio de Comercio, Industria y Turismo (República de Colombia, 2019) las organiza de acuerdo con el valor de ingresos y sector económico; es decir, que son heterogéneas financieramente; estas segmentaciones no permiten evaluar las características financieras como cifras, márgenes e indicadores financieros de las pymes colombianas, sino que se enfocan en los sectores económicos y su comportamiento.

La implementación de herramientas como el machine learning (o inteligencia artificial) han venido creciendo en Colombia, pero no al ritmo en que lo hacen los países desarrollados, y esto lo menciona Molina (2020) en su publicación, donde deja en evidencia que los estudios, las investigaciones y la utilización de estas herramientas se realizan en países como Reino Unido, Estados Unidos y China, seguidos de países como Alemania, Francia e India.

De lo anterior, surge la oportunidad de diseñar agrupaciones de las pymes colombianas utilizando métodos de machine learning con algoritmos de clúster no supervisados desde sus cifras e indicadores financieros, con la finalidad de plantear estrategias acordes con sus condiciones económicas.

Por consiguiente, para realizar estas agrupaciones se tomó como base la información financiera de 10.001 pymes colombianas durante cuatro periodos consecutivos (2016, 2017, 2018 y 2019), calculando indicadores financieros tales como margen bruto, margen operativo, margen antes de impuestos, margen neto, rendimiento del activo ROA, rendimiento sobre la inversión ROE, ebitda y margen de ebitda, los cuales fueron reducidos mediante la técnica de análisis de componentes principales PCA utilizando solo los cinco primeros componentes que representan el 91% de varianza para la ejecución de los modelos K-means y clúster aglomerativo.

Desde las nuevas agrupaciones creadas por los métodos de machine learning se realiza un análisis de cada clúster tomando como referencia los indicadores financieros mencionados antes, y, a partir allí, se determina cuál es el modelo que genera mejores resultados de agrupación y se plantean estrategias de acuerdo con los resultados obtenidos.

La estructura de esta investigación se divide en cuatro capítulos. El primero es el marco de referencia, que hace un recuento mediante la revisión bibliográfica sobre las pymes colombianas, los orígenes del machine learning y la utilización de métodos de algoritmos de clúster no supervisados en pymes. El segundo capítulo contiene la metodología que explica cómo se obtuvieron y depuraron los datos. Además, expone todos los conceptos y las técnicas matemáticas y estadísticas del modelo de análisis de componentes PCA, los modelos de clúster K-means y clúster aglomerativo, y, finalmente, detalla conceptos y fórmulas de las cifras e indicadores financieros empleados. El tercer capítulo se encuentra enfocado en los resultados y análisis de los clústeres generados desde los modelos de machine learning y a la creación de

estrategias de acuerdo con los resultados obtenidos. El último corresponde a las conclusiones y recomendaciones generadas desde los resultados obtenidos de la presente investigación.

2. MARCO DE REFERENCIA

2.1 Pymes

La denominación pymes proviene de la abreviatura de pequeñas y medianas empresas, pero como lo indican Ueki *et al.* (2005) aún no tienen una estandarización internacional definida y normalmente se clasifican por el volumen de ventas y la cantidad de empleados, división que presenta diferencias entre países. Por ejemplo, Colombia clasifica las pymes de acuerdo con el valor de activos, número de empleados (República de Colombia, 2004), valor de ingresos operacionales y el sector económico al que pertenecen (República de Colombia, 2019), agrupaciones que hasta ahora son heterogéneas financieramente.

Es importante recalcar la importancia de las pymes en la economía de los países en desarrollo, ya que hacen un aporte significativo al Producto Interno Bruto (PIB) y a la generación del empleo formal (BID, 2017), como se observa en la Tabla 1.

Tabla 1. Aporte de las pymes a la economía

REGIÓN	APORTE AL PIB	APORTE AL EMPLEO FORMAL	FUENTE DE INFORMACIÓN
Escalas globales	60%	Más de la mitad	Estudio llamado *Evaluación del apoyo directo de la CII a las PYMES* (BID, 2017)
Países con bajo ingreso	61,5%	67%	
Países con mediano bajo ingreso	68,5%	67%	
Colombia	**35%**	**80%**	Ministerio de Trabajo: "Mipymes representan más de 90%del sector productivo nacional y generan el 80% del empleo en Colombia" (2019)

Fuente: elaboración propia con base en el BID (2017) y "Mipymes representan más de 90%del sector productivo nacional y generan el 80% del empleo en Colombia" (2019).

2.2 Machine learning

Marchine learning, o también llamado aprendizaje automático, es una subdivisión de la inteligencia artificial que permite que los algoritmos detecten patrones recurrentes (*patterns*) en un grupo de datos. Al encontrar estos patrones los algoritmos aprenden y mejoran los modelos, las predicciones estadísticas o cualquier tarea específica ("Machine Learning: definición, funcionamiento, usos", 2021). En la

Tabla 2 se detallan algunas de las principales investigaciones en términos de machine learning.

Tabla 2. Evolución del machine learning

AÑO	AUTOR	CONCEPTO	DETALLE
1943	Walter Pitts y Warren Sturgis McCulloch	Artículo "A Logical Calculus of the Ideas Immanent in Nervous Activity"	Vislumbraron al cerebro como un organismo computacional (López *et al.*, 2007)
1950	Alan Turing, matemático inglés	"Test de turing"	Un humano mantiene una conversación con una persona y una máquina al tiempo, sin tener conocimiento de cuál era la máquina, con el objetivo de determinar si una máquina puede generar respuestas humanas (López *et al.*, 2007)
1955	John McCarthy, científico y matemático estadounidense	Término "inteligencia artificial"	Introdujo el término de inteligencia artificial como la unión entre ciencia e ingeniería para crear máquinas inteligentes (Marsden, 2017)
1966	Charles Rosen, Nils Nilsson y Peter Hart	Shakey	Primer robot que podía razonar y realizar sus propias acciones (Marsden, 2017)
1997	IBM	Deep Blue	Computadora capaz de jugar ajedrez; la máquina derrota a Garry Kasparov, actual campeón de ajedrez (Marsden, 2017)
2002	iROBOT	Roomba	Primera aspiradora robótica (Marsden, 2017)
2006	APACHE	Hadoop	Primera Plataforma Big Data Open Source (Nalda, 2020)
2011	IBM	Watson	Computadora que responde preguntas en el concurso televisivo Jeopardy, y gana el primer lugar (Marsden, 2017)
2014	AMAZON	Alexa	Asistente de compras virtual que completa tareas por medio de interfaz de voz (Marsden, 2017).
2016	MICROSOFT	Tay	Chatbot que realiza comentarios racistas, ofensivos e incendiarios (Marsden, 2017)
2017	GOOGLE	IA AlphaGo	Vence al campeón de juego de mesa Go, Ke Jie, con 2*170 posiciones posibles (Marsden, 2017)

Fuente: elaboración propia con base en López *et al.* (2007), Marsden (2017) y Nalda (2020).

Ahora bien, es importante recordar que machine learning busca el aprendizaje autónomo de las máquinas y para ello se pueden emplear tres técnicas: 1) el aprendizaje supervisado, que consiste

en etiquetar los datos para indicar a la máquina qué patrones debe buscar; 2) el aprendizaje reforzado, que consiste en dejar que el algoritmo aprenda del error y así se logra un determinado objetivo ("Machine Learning: definición, funcionamiento, usos", 2021), y 3) algoritmos no supervisados, técnica que fue utilizada en el presente trabajo y se describe en la sección 3.2 "Modelos de machine learning con algoritmos no supervisados".

Los algoritmos no supervisados, y específicamente los modelos de agrupaciones por clúster, se han implementado en campos como la medicina para la detección de enfermades, en biología para la clasificación de plantas y animales, y en *marketing* para especificar las costumbres de compras de las personas (Ordoñez *et al.*, 2020).

Por otro lado, pese a que no se han evidenciado investigaciones previas donde se apliquen modelos de machine learning con algoritmos de clúster no supervisados para generar una alternativa de agrupación a las pymes, algunos autores han realizado estudios previos que recalcan la importancia de la implementación de los clústeres diferenciales en las pymes, y otros han generado dentro de su investigación clústeres demográficos. A continuación, se presentan algunos ejemplos de lo anteriormente descrito:

- López y Calderón (2006) realizaron una investigación que aborda los clústeres empresariales como una manera de promover la productividad en las regiones.
- Capó-Vicedo, Expósito-Langa y Masiá-Buades (2007) analizan la relevancia de las agrupaciones y las concentraciones demográficas de las pymes para que de esta manera las empresas aumenten su competitividad en una economía globalizada.

- Morales, Sanabria y Arias (2010) crean dinámicas de aglomeración para el sector del calzado, cuero y afines en Bogotá, tomando como base publicaciones realizadas en países desarrollados.
- Capa, Benítez y Rojas (2017) sistematizan un análisis de los clústeres como una alternativa en la competitividad de las pymes en Ecuador.

3. METODOLOGÍA

La figura 1 muestra un resumen de las metodologías tenidas en cuenta para el desarrollo, la ejecución y el análisis de la presente investigación, las cuales serán descritas en cada ítem de este capítulo.

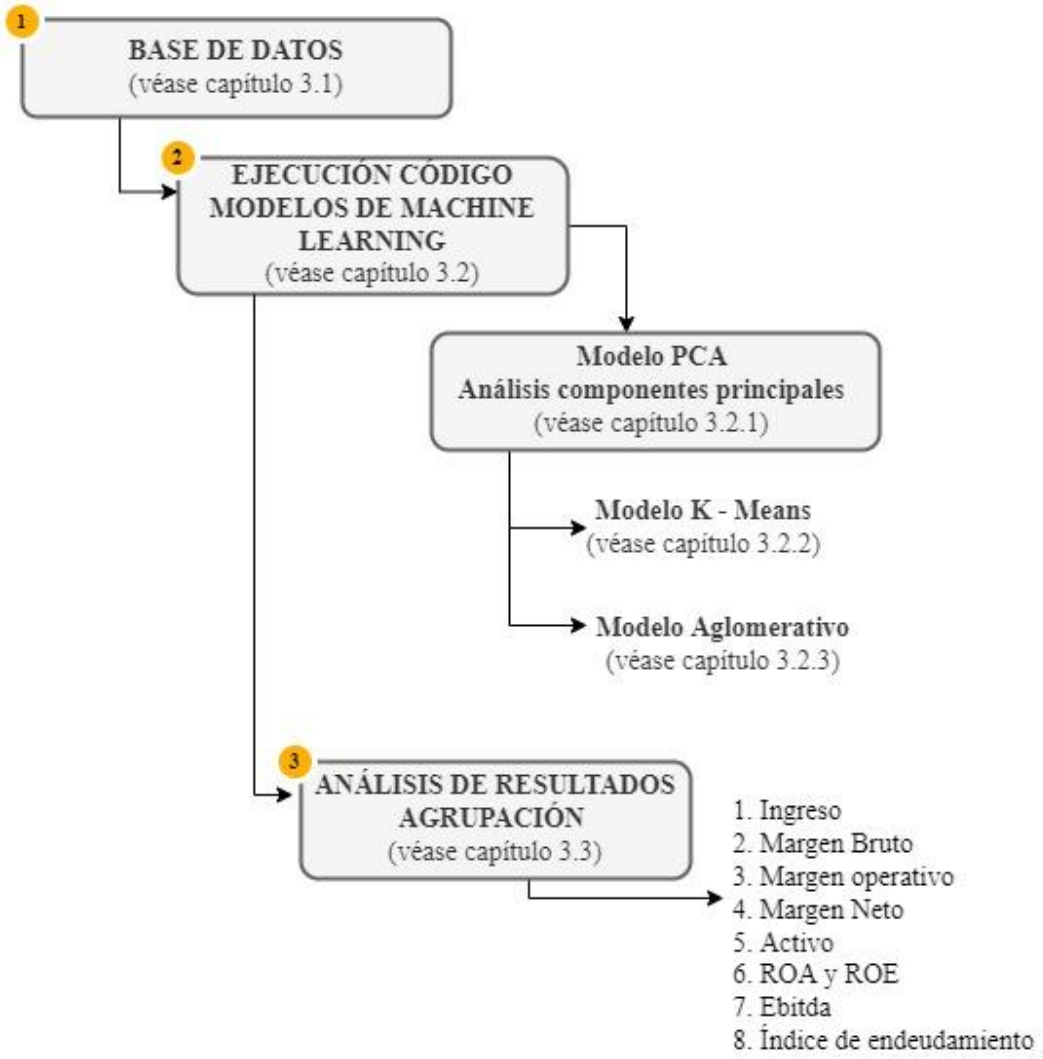

Figura 1. Metodología desarrollo investigación

Fuente: elaboración propia.

3.1 Base de datos

Los datos base son los estados financieros bajo las Normas Internacionales de Información Financiera (NIIF) que fueron descargados desde la página de la Superintendencia de Sociedades de reportes masivos, estados financieros clasificación pymes individuales para los años 2016, 2017, 2018 y 2019. El Ministerio de Hacienda y Crédito Público (República de Colombia, 2013) considera que los estados financieros individuales son aquellos que cumplen con las especificaciones de la NIC 1 o la NIC 34, y son presentados por una entidad cuyas inversiones no tengan la condición de controladora o negocio conjunto o asociada. La figura 2 describe las actividades que se deben realizar para obtener la base de datos principal para el desarrollo de los modelos de algoritmos de machine learning de clúster no supervisados.

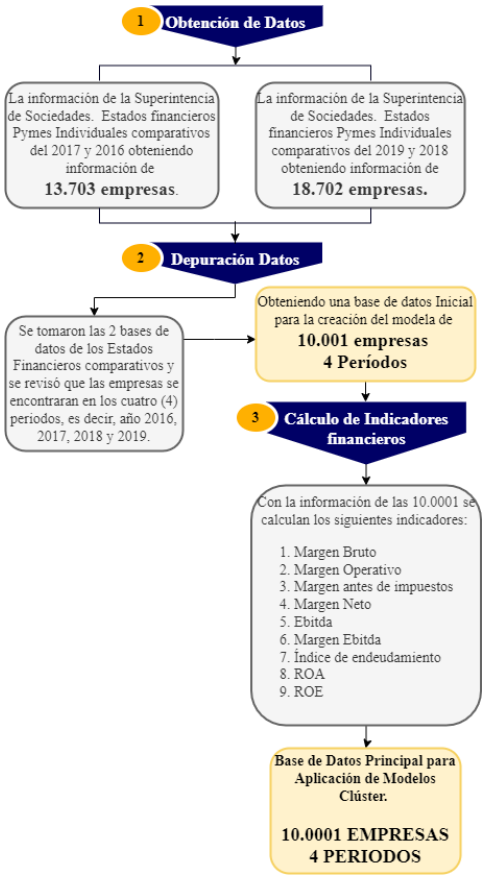

Figura 2. Depuración base de datos

Fuente: elaboración propia.

3.2 Modelos de machine learning con algoritmos no supervisados

Los modelos de machine learning con algoritmos no supervisados son una rama del aprendizaje automático utilizada para encontrar patrones ocultos y aprender la estructura subyacente en datos no etiquetados (Patel, 2019), es decir, contrario a herramientas de regresión en las que tenemos unos datos o variables X que buscan predecir una variable Y; en el aprendizaje no supervisado no hay una variable Y que se deba buscar, y el objetivo es encontrar patrones, tendencias o agrupaciones dentro de los datos rastreando características similares.

Dentro del aprendizaje no supervisado se usará el análisis de clúster, que se encarga de formar grupos diferentes dentro de los datos. Al usar algoritmos de agrupamiento se encuentra la estructura en los datos, de manera que los elementos del mismo clúster (o grupo) sean más similares entre sí que con los otros grupos generados (Roman, 2019). Así, si tenemos una base de datos con diferentes variables el aprendizaje no supervisado busca esos patrones comunes y agrupa los datos según las características similares de sus variables.

3.2.1 Método PCA (análisis de componentes principales)

El método de análisis de componentes principales es una de las técnicas más importantes de reducción lineal de dimensionalidad. El algoritmo busca crear una representación reducida de los datos, mientras se conserva la mayor cantidad de información posible.

El método aborda la correlación entre características. Si la correlación es muy alta entre un subconjunto de características, PCA combina las características altamente correlacionadas y representa estos datos con un número menor de características linealmente no correlacionadas. El algoritmo sigue realizando esta reducción de correlación, encontrando las direcciones de máxima variación en los datos originales de alta dimensión y proyectándolos en un espacio dimensional más pequeño. Estos nuevos componentes generados se conocen como componentes principales (Patel, 2019).

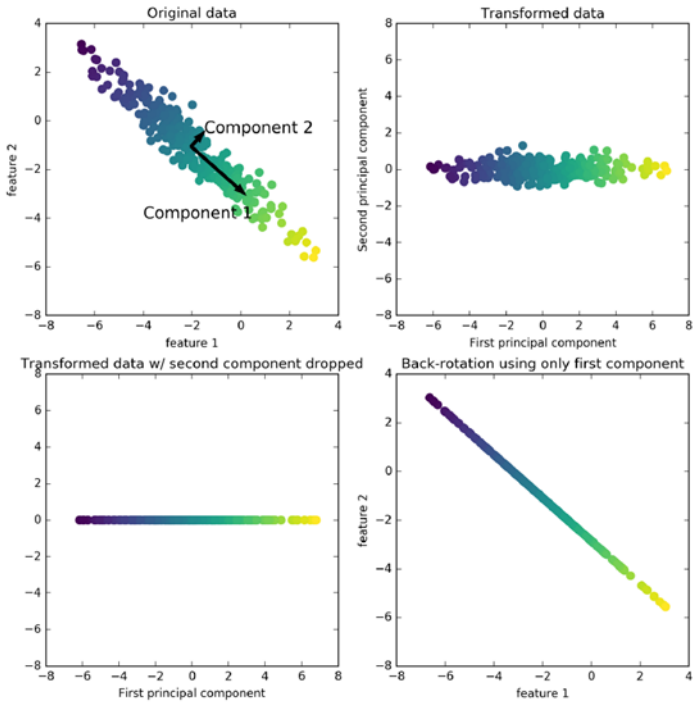

Figura 3. Ejemplo del método PCA para un set de datos

Fuente: Müller y Guido (2017, p. 141).

Según Müller y Guido (2017), en la figura 3 se describe la transformación de los datos usando componentes principales de la siguiente manera: la primera figura a la izquierda muestra los puntos de datos originales, con distintos colores para distinguirse entre ellos, pero teniendo en cuenta dos de sus variables. El algoritmo procede encontrando la dirección de máxima varianza, etiquetada como “Componente 1”. Esta es la dirección (o vector) en la que los datos contienen la mayor cantidad de información, o, en otras palabras, la dirección a lo largo de la cual las características son más correlacionadas entre sí. El algoritmo encuentra la dirección que contiene la mayor parte de la información siendo ortogonal (en ángulo recto) a la primera dirección. En dos dimensiones, solo hay una orientación posible que es en ángulo recto, pero en espacios de dimensiones superiores habría (infinitamente) muchas direcciones ortogonales.

Las direcciones encontradas usando este proceso se denominan componentes principales, ya que son las principales direcciones de varianza en los datos. En general, hay tantos componentes principales como características originales.

La segunda figura (arriba a la derecha) muestra los mismos datos, pero ahora girados para que el primer componente principal esté alienado con el eje X, y el segundo componente principal se alinee con el eje Y. Antes de la rotación, se restó la media de los datos, por lo que los datos transformados se centran alrededor de 0. En la representación rotada encontrada por PCA, los dos ejes no están correlacionados, lo que significa que la matriz de correlación de los datos en esta representación son 0 excepto por la diagonal.

En este ejemplo, podríamos mantener solo el primer componente principal, como se muestra en el tercer recuadro (abajo a la izquierda). Finalmente, podemos deshacer la rotación y volver a agregar la media a los datos como se muestra en el último recuadro (abajo a la derecha). Estos puntos están en la función original del espacio, pero conservamos solo la información contenida en el primer componente principal.

Esta transformación es utilizada para eliminar los efectos de ruido de los datos o visualizar qué parte de la información se retiene utilizando los componentes principales.

3.2.1.1 Transformación matemática usando componentes principales

Los eigenvectores (vectores propios) y los eigenvalores (valores propios) son necesarios cuando se realizan análisis de componentes principales. Los eigenvectores son los ejes (direcciones) a lo largo de los cuales la transformación actúa simplemente estirando, comprimiendo o volteando los datos; mientras que los eigenvalores generan los factores por los cuales se produce la compresión. Así, un eigenvector de una transformación lineal es un vector distinto de 0 cuya dirección no cambia cuando a esa transformación lineal se le aplica transformación (Dangeti, 2017).

Matemáticamente, A es una transformación lineal de un espacio vectorial y es un vector distinto de 0, entonces es vector propio de A si $A\vec{v}$ es un múltiplo escalar de $\lambda A\vec{v}$. La condición se puede describir como lo muestra la siguiente ecuación:

Ecuación 1:

$$A\vec{v} = \lambda A\vec{v}$$

En la ecuación, $\vec{v}$ es un eigenvector, A es una matriz cuadrada y λ es un escalar llamado eigenvalor. La dirección de un eigenvector permanece igual después que ha sido transformado por A y solo cambia su magnitud como es indicado por su eigenvalor, esto se hace multiplicando una matriz por uno de sus eigenvectores que es igual a escalar el eigenvector, lo cual es una representación compacta de la matriz original. En la figura 4 se representan gráficamente, en dos dimensiones, los eigenvectores y los eigenvalores.

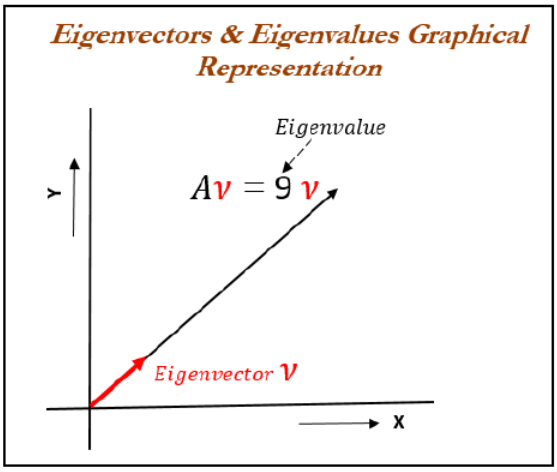

Figura 4. Representación gráfica de los eigenvectores y eigenvalores

Fuente: Dangeti (2017, p. 323).

Para generar los componentes principales se usan librerías y código en Python, como se describe en la figura 5.

```
from sklearn.preprocessing import StandardScaler
from sklearn.decomposition import PCA

scaler = StandardScaler(with_std=False)
x = scaler.fit_transform(df[['x1', 'x2']])

pca = PCA(n_components=1)
x_new = pca.fit_transform(x)
```

Figura 5. Ejemplo de código y librerías en Python para generar los componentes principales

Fuente: Amr (2020, p. 143).

El número de componentes principales se puede hallar gráficamente, como lo muestra la figura 6, donde cada componente principal aumenta el porcentaje de varianza explicada. En el ejemplo de la figura 6 con el método 1 con 10 componentes principales se explica alrededor del 60% de la varianza de los datos, y con el método 2 con 21 componentes principales se explica alrededor del 80% de la varianza de los datos. Por lo general, se toma la cantidad de componentes principales que expliquen más del 80% de la variabilidad de los datos.

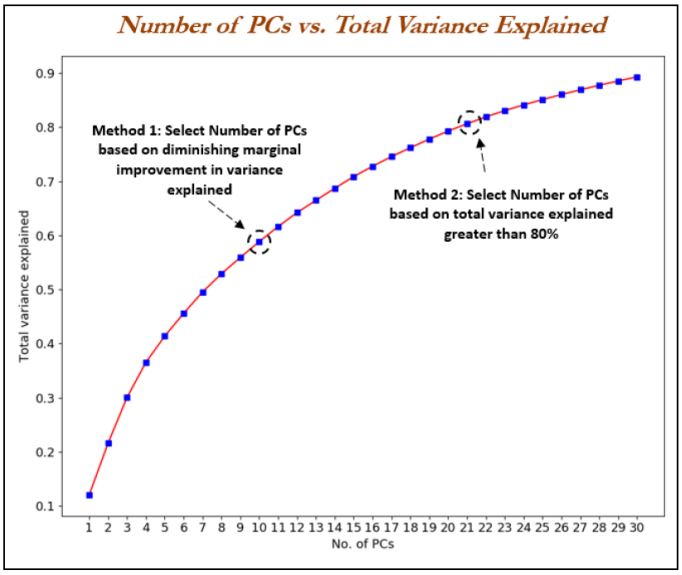

Figura 6. Número de componentes principales vs. varianza explicada

Fuente: Dangeti (2017, p. 337).

3.2.2 Método clúster K-means

El algoritmo K-means busca agrupar en clúster o grupos los puntos de datos que tienen una alta similitud entre ellos. En los términos del algoritmo, esta similitud se entiende como lo opuesto de la distancia entre puntos de datos. Cuanto más cerca estén los puntos de datos, serán más similares y con más probabilidades de pertenecer al mismo clúster (Roman, 2019).

El algoritmo de agrupamiento K-means es un proceso iterativo que mueve los centros de los agrupamientos o centroides a la posición media de sus puntos constituyentes, y reasigna instancias a sus clústeres más cercanos iterativamente hasta que no haya un cambio significativo en el número de clústeres posibles o número de iteraciones alcanzadas (Dangeti, 2017).

La función de costo de K-means está determinada por la distancia euclidiana (norma cuadrada) entre las observaciones pertenecientes a ese grupo con su respectivo valor centroide. Para

entender la ecuación, si tenemos un único clúster (K=1), entonces las distancias entre todas las observaciones son comparadas con la media de los datos. Si el número de clúster se incrementa a 2 (K=2), entonces se calculan dos medias y algunas de las observaciones se asignan al clúster 1 y otras observaciones se asignan al clúster 2 con base en la proximidad. Posteriormente, las distancias se calculan en funciones de costo aplicando la misma medida de distancia, pero por separado a sus centros de clúster (Dangeti, 2017).

Ecuación 2:

$$J = \sum_{k=1}^{K} \sum_{i \in Ck} ||xi - \mu k||^2$$

A continuación, en la figura 7 se observa el paso a paso de la forma en que el algoritmo va separando cada clúster, primero asignando de forma aleatoria los centros de cada clúster (en este caso son tres clústeres —representados por triángulos—). El proceso iterativo comienza y, posteriormente, a cada dato (representado por círculos), se le asigna el clúster más cercano. Con ese nuevo grupo de datos con un clúster asignado se calcula el promedio de esos datos y el clúster se mueve a esa nueva posición. La posición del clúster se mueve hasta terminar el número de iteraciones o llegar a una posición que ya no varía, y entonces el algoritmo se detiene; para este caso se generan tres clústeres con un agrupamiento muy bien definido en colores verde, rojo y azul (Müller y Guido, 2017).

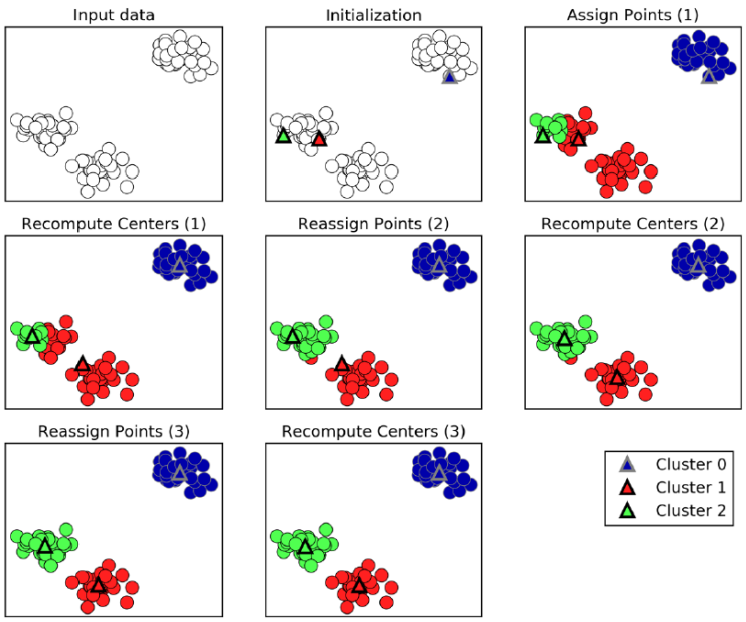

Figura 7. Paso a paso del agrupamiento con el método K-means

Fuente: Müller y Guido (2017, p. 169).

Según Roman (2019), los pasos para la ejecución del algoritmo son los siguientes:

- Primero, necesitamos elegir K, el número de clústeres que queremos que nos encuentren.
- Luego, el algoritmo seleccionará aleatoriamente los centroides de cada grupo.
- Se asignará cada punto de datos al centroide más cercano (utilizando la distancia euclídea).
- Se calculará la inercia del conglomerado.
- Los nuevos centroides se calcularán como la media de los puntos que pertenecen al centroide del paso anterior. En otras palabras, calculando el error cuadrático mínimo de los puntos de datos al centro de cada clúster, moviendo el centro hacia ese punto.
- Volver al paso 3.

3.2.2.1 Método del codo

Este método es usado para determinar el número óptimo de clústeres usando una interpretación gráfica. Se construye con los valores de la distancia euclidiana para diferentes números de K, donde K representa el número de clústeres. Si el valor de K se incrementa, la inercia o el porcentaje de distorsión de los datos tenderá a decrecer. En cada grupo se tendrán menos instancias constituyentes, y las instancias estarán más cerca de sus respectivos centroides. Sin embargo, las mejoras en la distorsión promedio disminuirán a medida que aumente K.

El valor de K en el que la mejora en la distorsión disminuye más se llama codo, como se observa en la figura 8; en este punto es donde deberíamos dejar de dividir los datos en más grupos (Dangeti, 2017).

Cuando se dispone de muchos datos se puede optar por generar una mayor cantidad de clústeres buscando ver valores anómalos o *outliers*, y comprender cuáles son las principales agrupaciones.

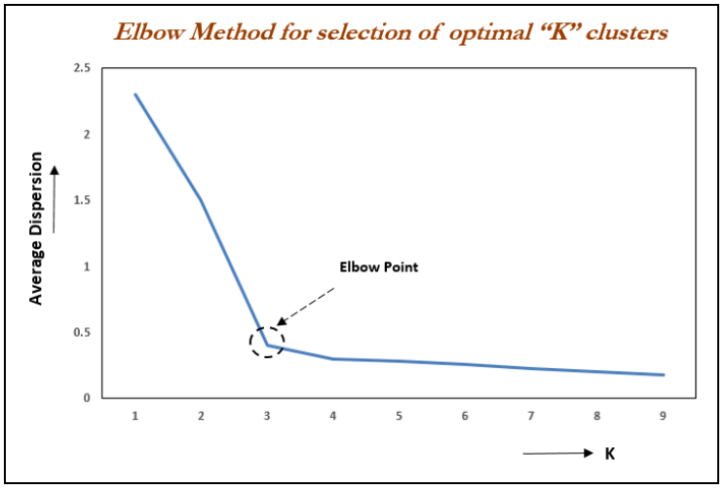

Figura 8. Método del codo para identificar el número correcto de clústeres K

Fuente: Dangeti (2017, p. 313).

El algoritmo de K-means se desarrolla usando librerías y el código en Python, como se describe en la **¡Error! No se encuentra el origen de la referencia.** 9.

```
from sklearn.cluster import KMeans
kmeans = KMeans(n_clusters=2, random_state=7)
x, y = df_blobs[['x1', 'x2']], df_blobs['y']
y_pred = kmeans.fit_predict(x)
```

Figura 9. Ejemplo del código en Python y librerías para el método K-means

Fuente: Amr (2020, p. 295).

3.2.3 Método clúster aglomerativo

Es un tipo de clúster jerárquico en el cual el agrupamiento se inicia con todas las observaciones separadas, cada una formando un clúster individual. Los clústeres se van combinado a medida que la estructura crece hasta converger en uno solo (Amat, 2020).

Hay varios criterios de vinculación (*linkage*) que especifican cómo es exactamente el clúster de mayor similitud medido. Esta medida siempre se define entre dos clústeres existentes. Las siguientes tres opciones se implementan en la librería scikit-learn (Müller y Guido, 2017).

- Ward

La opción predeterminada ward elige los dos grupos para fusionarlos de manera que la varianza dentro de todos los grupos es la que menos aumenta. Esto, a menudo conduce a grupos que son relativamente de igual tamaño.

- Average

La vinculación promedio fusiona los dos grupos que tienen la distancia promedio más pequeña entre todos sus puntos.

- Complete

El enlace completo (también conocido como enlace máximo) fusiona los dos grupos que tienen la menor distancia máxima entre sus puntos.

La figura 10 es un ejemplo donde inicialmente cada punto o dato es su propio clúster. Luego, en cada paso se fusionan los dos clústeres más cercanos. En los primeros cuatro pasos se seleccionan dos grupos de un solo punto y estos se unen en grupos de dos puntos. En el paso 5, uno de los grupos de dos puntos es extendido a un tercer punto, y así sucesivamente. En el paso 9 solo quedan tres clústeres. Como especificamos que estamos buscando tres grupos, el algoritmo se detiene (Müller y Guido, 2017).

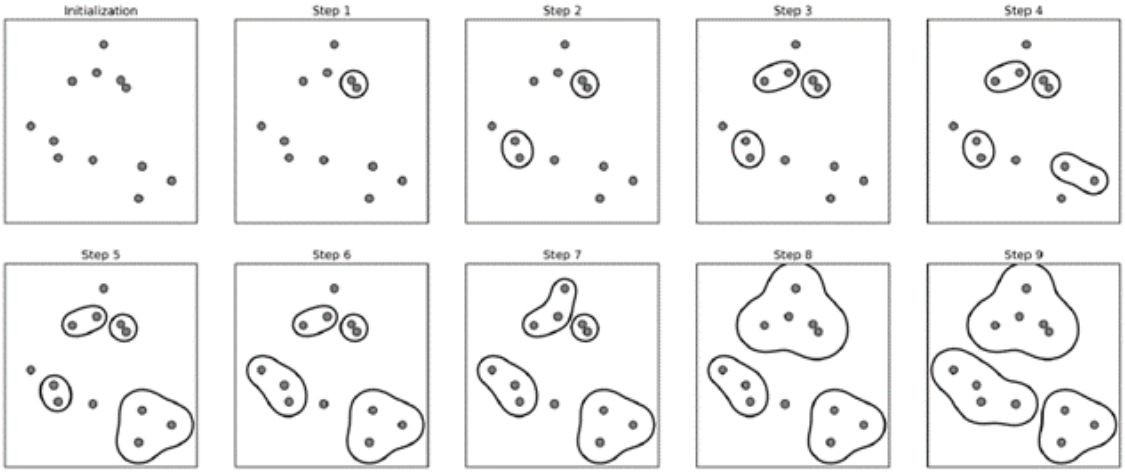

Figura 10. Ejemplo de la formación de los grupos de clústeres de forma iterativa usando el método aglomerativo

Fuente: Müller y Guido (2017, p. 183).

De acuerdo con Amat (2020), para aplicar el método del clúster aglomerativo el algoritmo realiza los siguientes pasos:

1. Considerar cada una de las *n* observaciones como un clúster individual, formando así la base del dendrograma (hojas).
2. Proceso iterativo hasta que todas las observaciones pertenecen a un único clúster: para que este proceso se realice, el algoritmo realiza los siguientes pasos:

   2.1 Calcular la distancia entre cada posible par de los *n* clústeres. El investigador debe determinar el tipo de medida empleada para cuantificar la similitud entre observaciones o grupos (distancia y *linkage*).

   2.2 Los dos clústeres más similares se fusionan, de forma que quedan *n*-1 clústeres.
3. Cortar la estructura de árbol generada (dendrograma) a una determinada altura para crear los clústeres finales (número de clúster).
4. El algoritmo de clúster aglomerativo se desarrolla usando librerías y el código en Python, como se describe en la figura 11.

```
from sklearn.cluster import AgglomerativeClustering

x, y = df_blobs[['x1', 'x2']], df_blobs['y']

agglo = AgglomerativeClustering(n_clusters=4)
y_pred = agglo.fit_predict(x)
```

Figura 11. Ejemplo del código en Python y librerías para el método aglomerativo

Fuente: Amr (2020, p. 302).

Por último, en la figura 12 se hace un resumen de los pasos que se deben realizar para la ejecución del código en Python de los modelos de machine learning con algoritmos de clúster no supervisados K-means y aglomerativo.

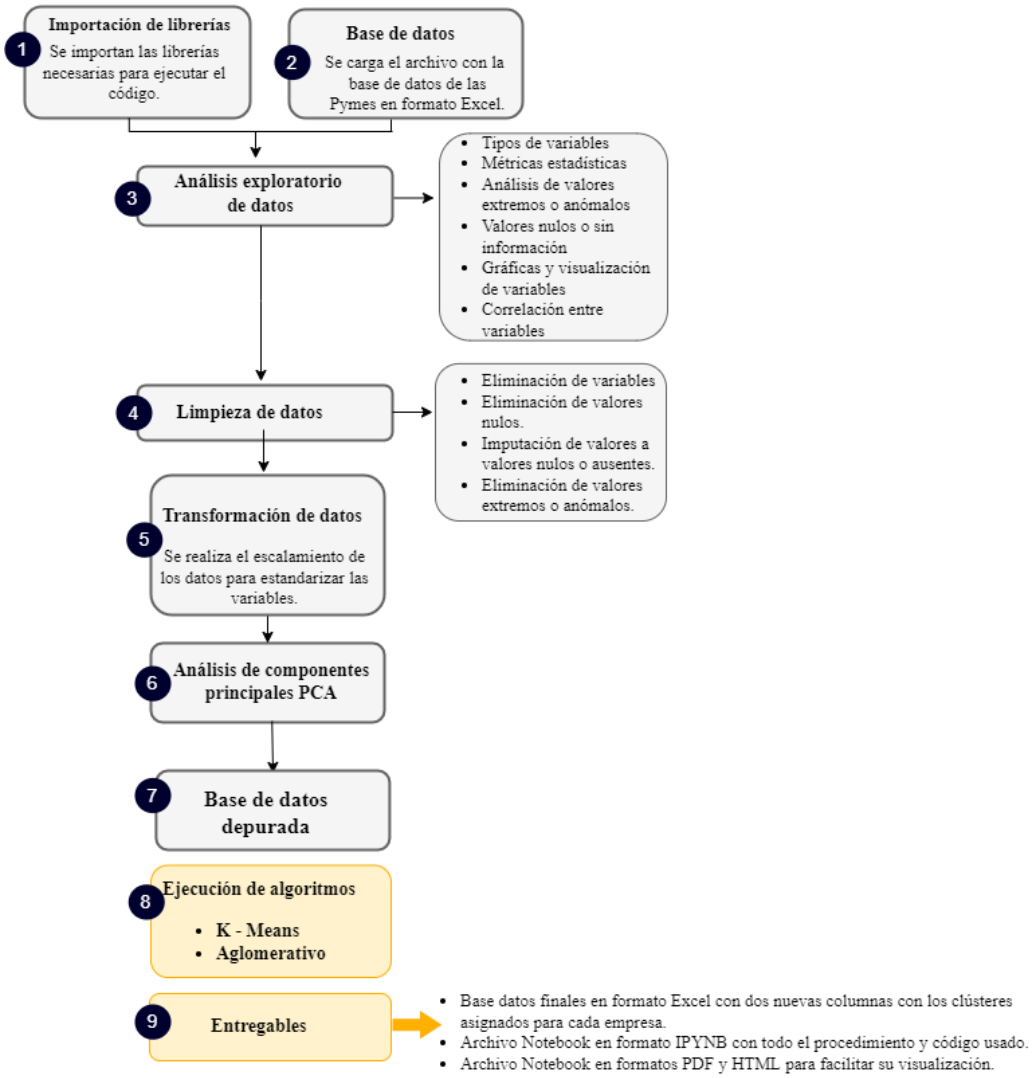

Figura 12. Desarrollo del modelo

Fuente: elaboración propia.

3.3 Análisis de los resultados de agrupación

Para realizar el análisis de las agrupaciones generadas desde los modelos K-means y el clúster aglomerativo, se tomaron las cifras financieras del estado de resultados, activo corriente, activo total, pasivo corriente y pasivo total. En la Tabla 3 **¡Error! No se encuentra el origen de la**

referencia.se hace un resumen de los conceptos mencionados, y los indicadores financieros están descritos en las Tabla 4, Tabla *5* y Tabla *6*.

Tabla 3. Conceptos

NOMBRE	CONCEPTO
Estado de resultado integral	Es uno de los principales estados financieros que presenta los resultados financieros de una organización durante un determinado periodo. Este estado se construye deduciendo de los ingresos, los costos de producción, los gastos operacionales y ventas, gastos financieros e impuestos (Ricra, 2014)
Activo	Es un recurso controlado por la empresa como resultado de sucesos pasados de los que se espera obtener beneficios económicos futuros ("NIIF para las pymes", s. f.)
Activo corriente	En el párrafo 4.5 de la NIIF para pymes IFRS Foundation ("Normas Internacionales de Información Financiera", 2009) se menciona cuando cumple las siguientes condiciones: se mantiene para negociación, se espera consumirlo en un periodo de doce meses
Pasivo	Es una obligación surgida de sucesos pasados, que durante la fecha del vencimiento se espera desprenderse de recursos que incorporan beneficios económicos ("NIIF para las pymes", s. f.)
Pasivo corriente	En el párrafo 4.7 de la NIIF para Pymes IFRS Foundation ("Normas Internacionales de Información Financiera", 2009) se menciona cuando cumple las siguientes condiciones: se espera cancelar dentro de los doce meses siguientes, se adquiere con el propósito de negociar
Patrimonio	Podría entenderse como el valor residual en los activos netos disminuidos todos sus pasivos ("NIIF para las pymes", s. f.)

Fuente: elaboración propia con base en "NIIF para las pymes" (s. f.), "Normas Internacionales de Información Financiera" (2009) y Ricra (2014).

3.3.1 Indicadores financieros

Con las cifras anteriormente mencionadas, se realizó el cálculo de los siguientes indicadores: 1) indicadores de endeudamiento o solvencia, que según Gitman y Zutter (2012) son indicadores que miden el riesgo, 2) indicadores de rentabilidad y 3) cálculo de ebitda y margen ebitda.

3.3.1.1 Indicadores, ratios o razones de solvencia o endeudamiento

Estos indicadores muestran la cantidad de dinero que tienen los terceros o socios en cuanto a los recursos de la empresa.

Tabla 4. Indicadores de solvencia

	INDICADOR	CONCEPTO
Solvencia o endeudamiento	**Índice endeudamiento** Total, de pasivos / Total de activos	Indica el porcentaje de los recursos financiados por los acreedores. Si el cociente es bajo, indica que la empresa tiene poco financiamiento de terceros para la obtención de ganancias, y si es alto, indica que la empresa tiene un alto apalancamiento y nivel de endeudamiento (Ricra, 2014)

Fuente: elaboración propia con base en Estrada y Monsalve (2014) y Ricra (2014).

3.3.1.2 Indicadores, ratios o razones de rentabilidad

Estos indicadores evalúan la eficiencia de generar ganancias a lo largo de los años, sea con recursos propios o de acreedores.

Tabla 5. Indicadores de rentabilidad

	INDICADOR	CONCEPTO
Rentabilidad	**Margen bruto** Utilidad bruta / Ventas	Representa el porcentaje de peso que queda después de descontarle a las ventas los costos de producción. Entre más alto significa mayor eficiencia en costos asociados al producto
	Margen operativo Utilidad operativa / Ventas	Representa el porcentaje de peso que queda después de descontarle a las ventas los costos de producción, los gastos administrativos y las ventas
	Margen neto Utilidad neta / Ventas	Relaciona la utilidad neta con las ventas. Es una medida exacta, ya que a los ingresos se les descuentan todos los costos y gastos que la organización ha tenido durante un periodo determinado (Ricra, 2014)
	Rendimiento sobre el activo (ROA) Utilidad neta / Activos	Este ratio mide la capacidad de generar ganancias con los activos disponibles (Gitman y Zutter, 2012)
	Rendimiento sobre el patrimonio (ROE) Utilidad neta / Patrimonio	Este ratio mide el rendimiento generado sobre el capital en acciones de los socios de una empresa (Gitman y Zutter, 2012)

Fuente: elaboración propia con base en Gitman y Zutter (2012) y Ricra (2014).

Tabla 6. Otros indicadores

INDICADOR	CONCEPTO
Ebitda Utilidad operativa + depreciaciones + amortizaciones + provisiones + deterioro	El ebitda, llamado así por sus cifras en inglés (earnings before interest, taxes, depreciation, and amortization), indica la salida efectiva de dinero, es decir, en el cálculo del ebitda no se tienen en cuenta los costos no efectivos
Margen ebitda Ebitda / Ventas	Es la relación entre el cálculo del ebitda con las ventas

Fuente: elaboración propia con base en Allen, Myers y Brealey (2010).

4. RESULTADOS

La ejecución del algoritmo de los modelos de machine learning de clúster no supervisados K-means y aglomerativo se puede observar en el siguiente enlace: https://github.com/DurleyRM/Codigo_tesis/wiki/Codigo; allí se encuentran la base de datos, la metodología aplicada y el código utilizado en formato notebook, PDF y HTML.

La utilización del modelo de análisis de componentes principales PCA se tomó debido a que el resultado de las agrupaciones por las cifras e indicadores financieros no generaban una división marcada en cada uno de los clústeres; además, al usar la metodología PCA se definió cada clúster de manera más específica y concisa, como se puede observar en la figura 13.

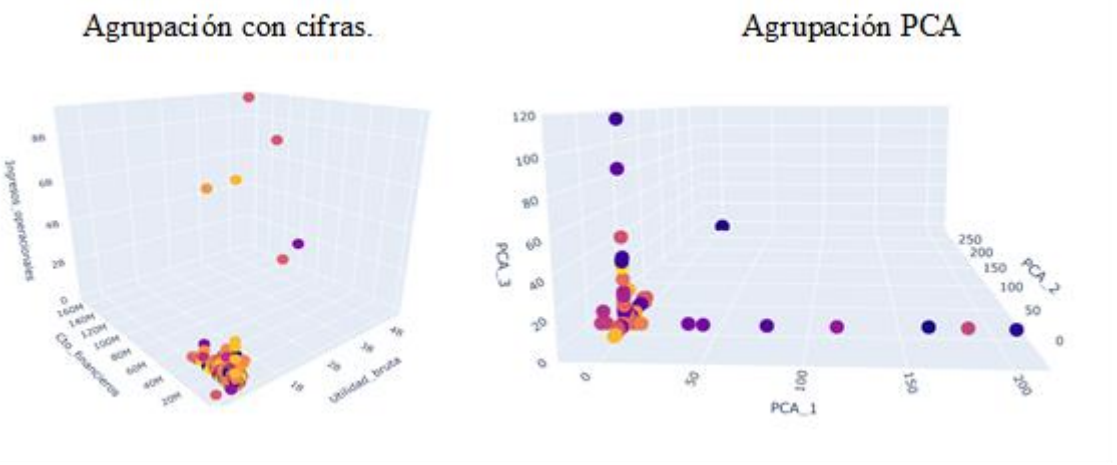

Figura 13. Agrupación con cifras vs. agrupación PCA

Fuente: elaboración propia.

4.1 Análisis de los resultados del modelo de componentes principales PCA

Para la elaboración del modelo de análisis de componentes principales PCA se utilizaron los indicadores financieros que han sido descritos en la sección 3.3 “Análisis de los resultados de agrupación” y las tablas 4, 5 y 6. Los nueve indicadores financieros son las variables del modelo que fueron reducidos a cinco componentes principales, ya que explican alrededor del 91% de la variabilidad o varianza. El resultado obtenido se muestra en la tabla 7 y la figura 14.

Tabla 7. Número de componentes principales PCA y su acumulado de varianza

NÚMERO COMPONENTES	%	% ACUMULADO
Primer componente	35%	35%
Segundo componente	26%	61%
Tercer componente	11%	72%
Cuarto componente	11%	83%
Quinto componente	8%	91%

Fuente: elaboración propia.

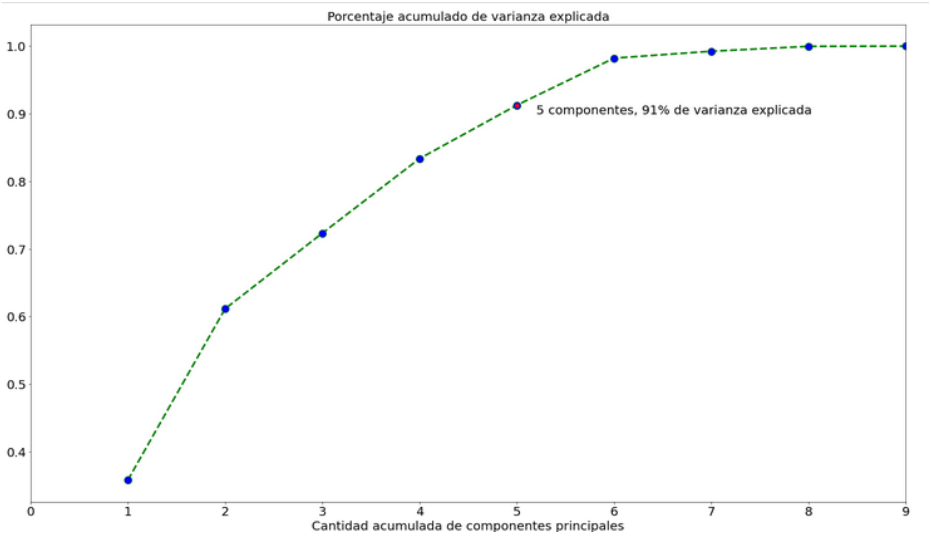

Figura 14. Modelo PCA

Fuente: elaboración propia.

Por lo tanto, para la aplicación de los modelos K-means y aglomerativo se utilizaron como variables los cinco componentes principales descritos anteriormente, y cada componente es explicado con la correlación de cada una de las variables, o, para el caso de estudio, los indicadores financieros, como se puede observar en la figura 15.

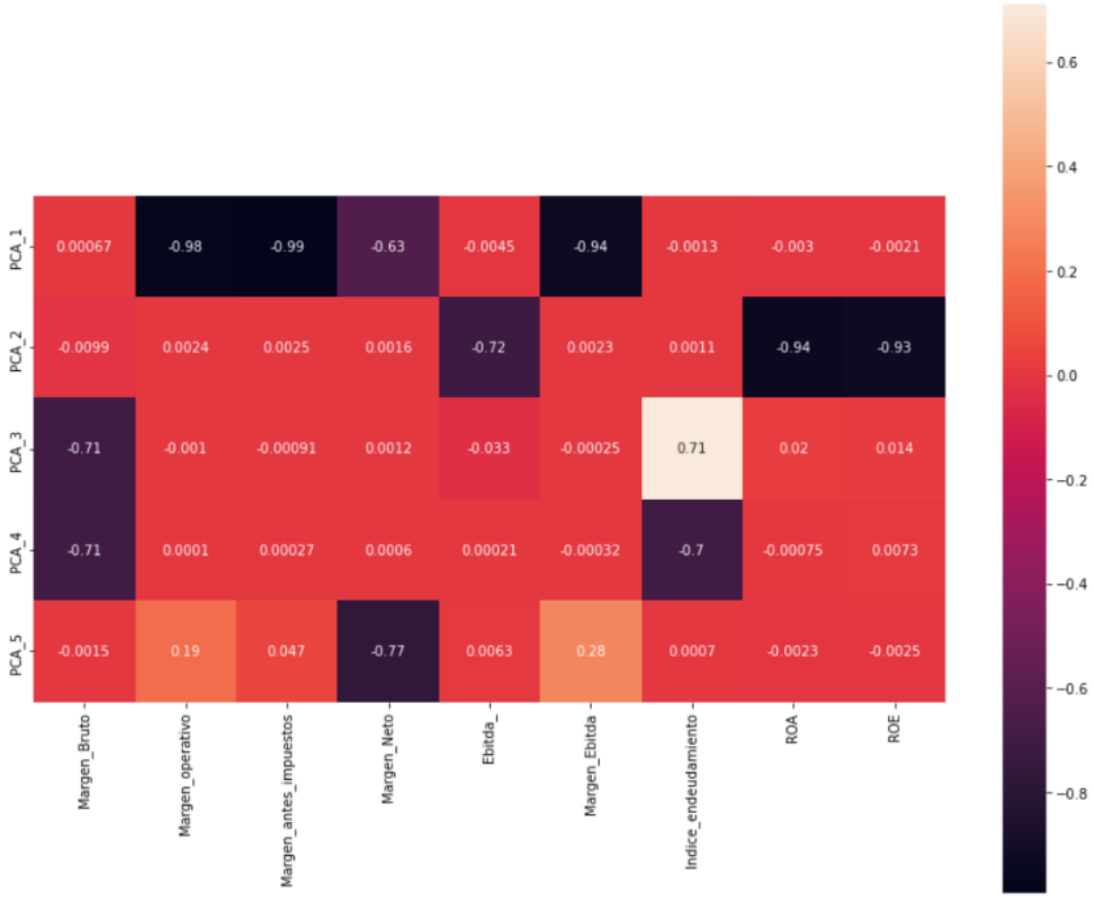

Figura 15. Correlación PCA con los indicadores

Fuente: elaboración propia.

Con la figura 15 se puede concluir que existe predominantemente una correlación negativa, es decir, entre más alto el PCA más bajo el valor del indicador. Los porcentajes de correlación entre los PCA y los indicadores son los siguientes:

- PCA1 se correlaciona con el margen operativo en -0,98, con margen antes de impuestos en -0,99, -0,63 con margen neto y -0,94 con margen ebitda.
- PCA2 se correlaciona con ebitda en -0,72, -0,94 y -0,93 con ROA y con ROE.
- PCA3 se correlaciona con margen bruto en -0,71 y 0,71 con índice de endeudamiento.
- PCA4 se correlaciona con margen bruto en -0,71 y -0,7 con índice de endeudamiento.
- PCA5 se correlaciona con margen neto en -0,77.

4.2 Análisis de los resultados obtenidos del modelo de machine learning K-means y aglomerativo

En las Tabla 8 y Tabla *9* se hace un resumen de la conformación de los cincuenta clústeres generados por los modelos K-means y aglomerativo en cuanto a cantidad de datos para cada uno de ellos.

Tabla 8. Resultado clústeres - Algoritmo K-means

Clúster	**0**	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**	**12**	**13**	**14**	**15**	**16**	**17**	**18**	**19**
N.º de datos	528	1	1	1	4	1	1	2	1	1	12	3.169	2	11	11	9	1	7.584	1	3

Clúster	**20**	**21**	**22**	**23**	**24**	**25**	**26**	**27**	**28**	**29**	**30**	**31**	**32**	**33**	**34**	**35**	**36**	**37**	**38**	**39**
N.º de datos	1	122	58	2	1	7.582	1	5	3	386	23	1	7	1.953	2	480	29	37	3	4.698

Clúster	**40**	**41**	**42**	**43**	**44**	**45**	**46**	**47**	**48**	**49**
N.º de datos	1	3	5.022	4.200	15	119	2	2.250	125	1.529

Fuente: elaboración propia.

En la tabla 8 se observa que el algoritmo realiza una agrupación de manera heterogénea, además de esto, genera doce clústeres principales, los cuales tienen el 98,44% de los datos distribuidos de la siguiente manera:

1. Clúster 17 = 7.584 que equivale al 18,96% de los datos

2. Clúster 25 = 7.582 que equivale al 18,95% de los datos
3. Clúster 42 = 5.022 que equivale al 12,55% de los datos
4. Clúster 39 = 4.698 que equivale al 11,74% de los datos
5. Clúster 43 = 4.200 que equivale al 10,50% de los datos
6. Clúster 11 = 3.169 que equivale al 7,92% de los datos
7. Clúster 47 = 2.250 que equivale al 5,62% de los datos
8. Clúster 33 = 1.953 que equivale al 4,88% de los datos
9. Clúster 49 = 1.529 que equivale al 3,82% de los datos
10. Clúster 0 = 528 que equivale al 1,32% de los datos
11. Clúster 35 = 480 que equivale al 1,20% de los datos
12. Clúster 29 = 386 que equivale al 0,96% de los datos

Total 39.381, igual al 98,44% de los datos

A continuación, en la tabla 9 se pueden evidenciar los resultados obtenidos desde el método de agrupación aglomerativo, que es similar a K-means, ya que presenta la misma cantidad de grupos principales y estos tienen el 98,44% de los datos.

Tabla 9. Resultado clústeres-algoritmo aglomerativo

Agglom.	**0**	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**	**12**	**13**	**14**	**15**	**16**	**17**	**18**	**19**
N.º de datos	4.347	5.932	237	303	86	2	9	5	8	5.463	13	1.138	6	3	12	1	825	15	7.852	2

Agglom.	**20**	**21**	**22**	**23**	**24**	**25**	**26**	**27**	**28**	**29**	**30**	**31**	**32**	**33**	**34**	**35**	**36**	**37**	**38**	**39**
N.º de datos	5.901	11	1.240	3	1.717	2	1	1	616	1	1	1	1	1	22	1	6	4.044	5	1

Agglom.	**40**	**41**	**42**	**43**	**44**	**45**	**46**	**47**	**48**	**49**
N.º de datos	1	100	3	28	2	1	29	1	1	3

Fuente: elaboración propia.

Los doce grupos principales están conformados así:

1.	Clúster 18	=	7.852 que equivale al 19,6% de los datos
2.	Clúster 1	=	5.932 que equivale al 14,8% de los datos
3.	Clúster 20	=	5.901 que equivale al 14,8% de los datos
4.	Clúster 9	=	5.463 que equivale al 13,7% de los datos
5.	Clúster 0	=	4.347 que equivale al 10,9% de los datos
6.	Clúster 37	=	4.044 que equivale al 10,1% de los datos
7.	Clúster 24	=	1.717 que equivale al 4,3% de los datos
8.	Clúster 22	=	1.240 que equivale al 3,1% de los datos
9.	Clúster 11	=	1.138 que equivale al 2,8% de los datos
10.	Clúster 16	=	825 que equivale al 2,1% de los datos
11.	Clúster 28	=	616 que equivale al 1,5% de los datos
12.	Clúster 3	=	303 que equivale al 0,8% de los datos
	Total		39.378, igual al 98,44% de los datos

4.2.1 Análisis de los resultados generados desde los modelos K-means y aglomerativo versus cifras e indicadores financieros

Durante el desarrollo de este subcapítulo, se presentan los resultados de segmentación obtenidos desde el análisis de los doce clústeres principales generados desde los modelos y la comparación de cada clúster versus algunas de las cifras e indicadores financieros. La tabla 10 hace un direccionamiento del subcapítulo y número de anexo donde se explica de manera detallada cada uno de los resultados.

Tabla 10. Direccionamiento, subcapítulos y número de anexo

NOMBRE INDICADOR	NÚMERO DE SUBCAPÍTULO DESCRITO	NÚMERO DE ANEXOS
Ingresos	4.2.1.1	1 y 2
Margen bruto	4.2.1.2	3 y 4
Margen operativo	4.2.1.3	5 y 6
Margen neto	4.2.1.4	7 y 8
Activo	4.2.1.5	9 y 10
ROA y ROE	4.2.1.6	11, 12, 13 y 14
Ebitda	4.2.1.7	15 y 16
Índice de endeudamiento	4.2.1.8	17 y 18

Fuente: elaboración propia.

4.2.1.1 Ingresos versus clústeres

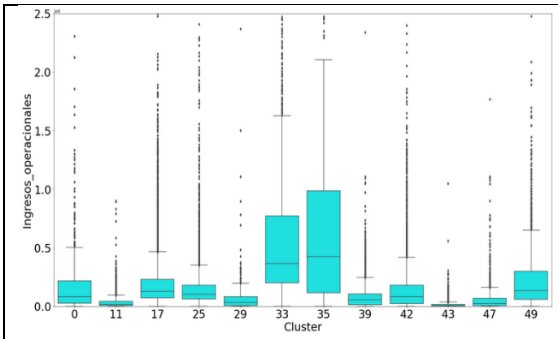

Figura 16. K-means, ingresos vs. clústeres

Fuente: elaboración propia.

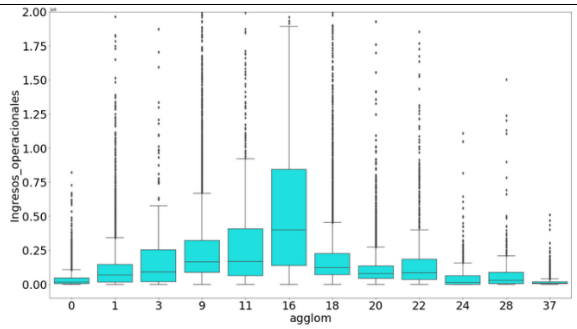

Figura 17. Aglomerativo ingresos vs. clústeres

Fuente: elaboración propia.

- Grupo 1: pymes con intervalos de ingresos entre 0 a 6.000 millones.

  Clústeres: 43, 29, 11 y 47.
- Grupo 2: pymes con intervalos de ingresos que van desde 0 a 40.000 millones.

  Clústeres: 39, 25, 17, 42, 49 y 0.

- Grupo 1: pymes con intervalos de ingresos entre 0 a 4.000 millones

  Clústeres: 37 y 0.
- Grupo 2: pymes con intervalos de ingresos que van desde 0 a 40.000 millones.

  Clústeres: 24, 28, 20, 1, 22, 3, 9 y 18.

• Grupos 3: pymes con intervalos de ingresos entre 0 a 100.000 millones. Clústeres: 35 y 33.	• Grupos 3: pymes con intervalos de ingresos entre 0 a 75.0000 millones. Clústeres: 16.

4.2.1.2 Margen bruto versus clústeres

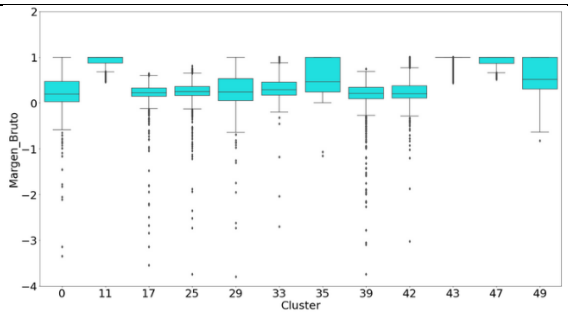

Figura 18. K-means, margen bruto vs. clústeres

Fuente: elaboración propia.

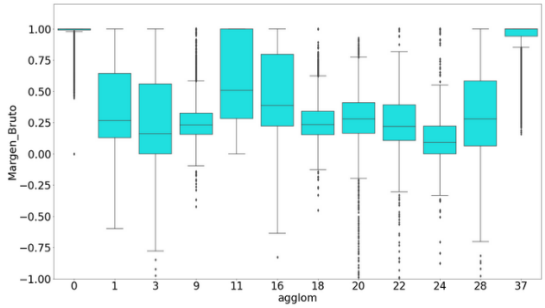

Figura 19. Aglomerativo, margen bruto vs. clústeres

Fuente: elaboración propia.

• Grupo 1: margen bruto que inicia intervalos negativos que van entre inferiores a 0 hasta el 50%. Clústeres: 29, 39 y 0. • Grupo 2: margen bruto con intervalos positivos que van desde 11 a 50%. Clústeres: 17, 42, 25 y 33. • Grupo 3: margen bruto con intervalos entre 21 a 100%. Clústeres: 49 y 35. • Grupo 4: margen bruto con un único intervalo entre 91 a 100%	• Grupo 1: margen bruto que inicia intervalos negativos que van entre inferiores a 0 hasta el 40%. Clústeres: 24, 28 y 22. • Grupo 2: margen bruto con intervalos positivos que van desde 1 a 60%. Clústeres: 22 y 1. • Grupo 3: margen bruto con intervalos entre 11 a 100%. Clústeres: 18, 9, 20, 16 y 11. • Grupo 4: margen bruto con un único intervalo entre 91 a 100%

Clústeres: 43.	Clústeres: 0, 37 y 3.

4.2.1.3 Margen operativo versus clústeres

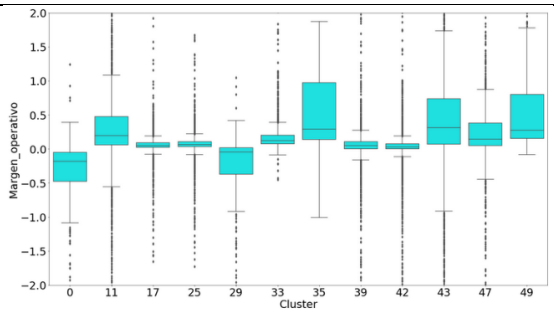

Figura 20. K-means, margen operativo vs. clústeres

Fuente: elaboración propia.

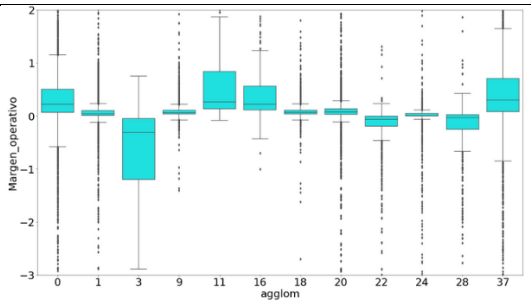

Figura 21. Aglomerativo, margen operativo vs. clústeres

Fuente: elaboración propia.

- Grupo 1: margen operativo inferior al 0%.

  Clúster: 0.

- Grupo 2: margen operativo que inicia intervalos negativos que van entre inferiores a 0 hasta el 40%.

  Clústeres: 29, 11, 43, 39 y 42.

- Grupo 3: margen operativo con intervalos positivos que van desde 1 a 40%.

  Clústeres: 17, 25, 33 y 47.

- Grupo 4: margen operativo con intervalos más amplios que van entre 11 a 100%.

- Grupo 1: margen operativo inferior al 0%.

  Clúster: 3.

- Grupo 2: margen operativo que inicia intervalos negativos que van entre inferiores a 0 hasta el 50%.

  Clústeres: 24, 22, 28, 37, 0 y 1.

- Grupo 3: margen operativo con intervalos positivos que van desde 1 a 40%.

  Clústeres: 0, 18, 20 y 6.

- Grupo 4: margen operativo con intervalos más amplios que van entre 21 a 100%.

Clústeres: 49 y 35.	Clústeres: 11.

4.2.1.4 Margen neto versus clústeres

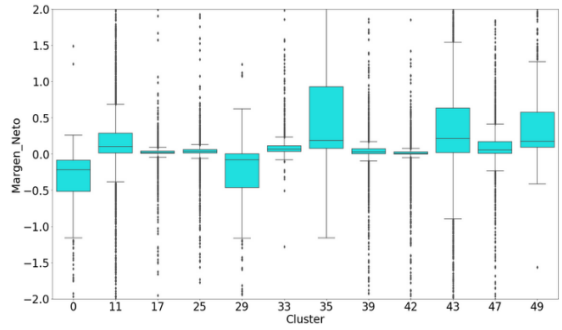

Figura 22. K-means, margen neto vs. clústeres

Fuente: elaboración propia.

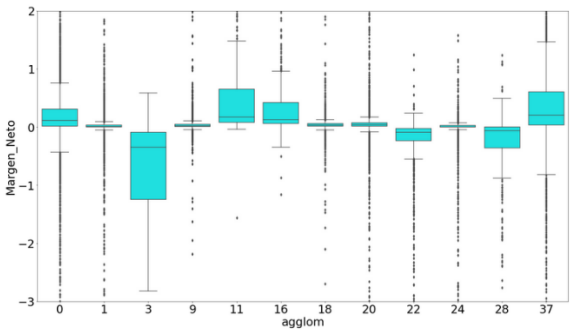

Figura 23. Aglomerativo, margen neto vs. clústeres

Fuente: elaboración propia.

- Grupo 1: margen neto negativo inferior al 0%.

  Clústeres: 0 y 29.
- Grupo 2: margen neto con intervalos que van entre -1 a 10%.

  Clústeres: 43, 11, 47, 42, 39, 17 y 25.
- Grupo 3: margen neto con intervalos positivos amplios que van entre 1 a 100%.

  Clústeres: 33, 49 y 35.

- Grupo 1: margen neto negativo inferior al 0%.

  Clústeres: 3 y 28.
- Grupo 2: margen neto que inicia con un intervalo inferior a 0 hasta el 40%.

  Clústeres: 24, 1, 22, 0, 37 y 20.
- Grupo 3: margen neto con intervalos positivos amplios que van entre 1 a 30%.

  Clústeres: 9, 18, 16 y 11.

4.2.1.5 Activo versus clústeres

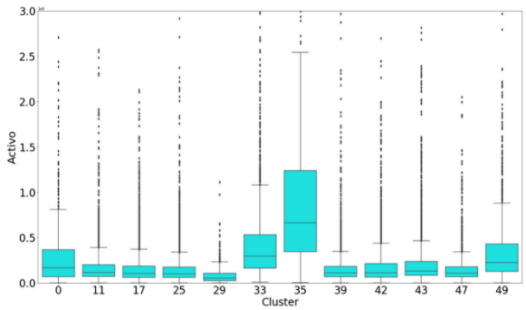

Figura 24. K-means, activo vs. clústeres

Fuente: elaboración propia.

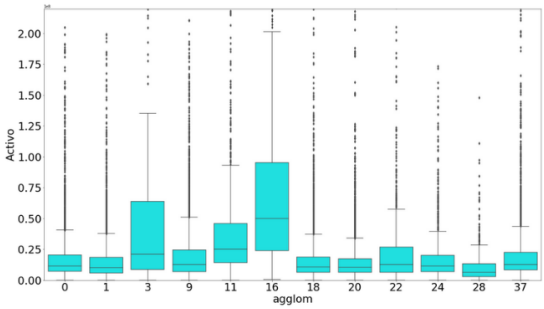

Figura 25. Aglomerativo, activo vs. clústeres

Fuente: elaboración propia.

- Grupo 1: activos entre 0 a 50.000 millones, pero con concentración de datos de más del 40% en todos los clústeres entre 0 a 10.000 millones.

  Clústeres: 29, 25, 17, 39, 47, 42, 11, 43, 0, 49 y 33.

- Grupo 2: activos con intervalos superiores al 50%.

  Clústeres: 35.

- Grupo 1: activos entre 0 a 40.000 millones, pero con concentración de datos de más del 40% en todos los clústeres entre 0 a 10.000 millones.

  Clústeres: 28, 20, 1, 18, 24, 22, 0, 9, 37, 11 y 3.

- Grupo 2: activos neto que inician con un intervalo mayor o igual al 10%, hasta más de 100.0000 millones.

  Clústeres: 16.

4.2.1.6 Rentabilidad sobre el activo y sobre el patrimonio (ROA-ROE) versus clústeres

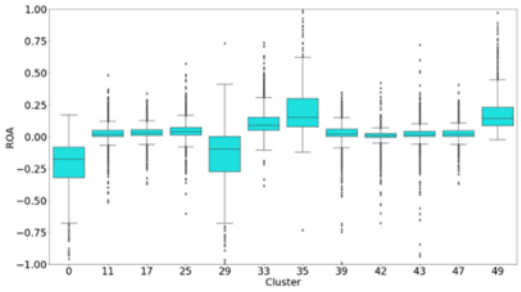

Figura 26. K-means, ROA vs. clústeres

Fuente: elaboración propia.

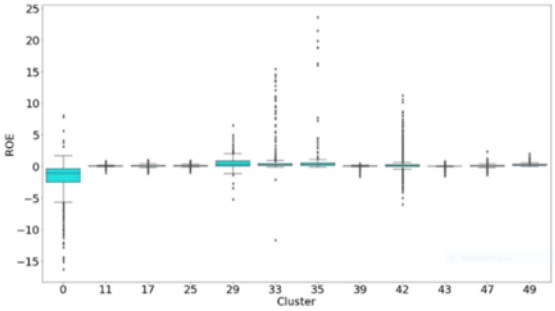

Figura 27. K-means, ROE vs. clústeres

Fuente: elaboración propia.

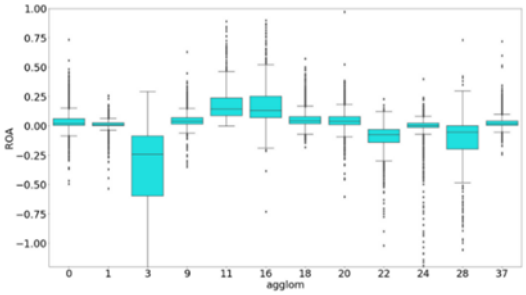

Figura 28. Aglomerativo, ROA vs. clústeres

Fuente: elaboración propia.

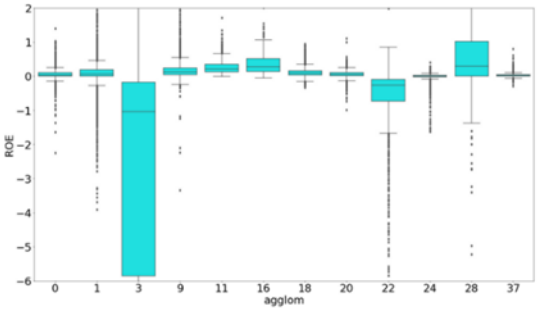

Figura 29. Aglomerativo, ROE vs. clústeres

Fuente: elaboración propia.

La fórmula de ROA y ROE va ligada a la utilidad neta, así que se crean los mismos tres grupos principales obtenidos en el análisis de margen neto tanto para K-means como para aglomerativo (véase anexos 12, 13, 14 y 15).

4.2.1.7 Ebitda versus clústeres

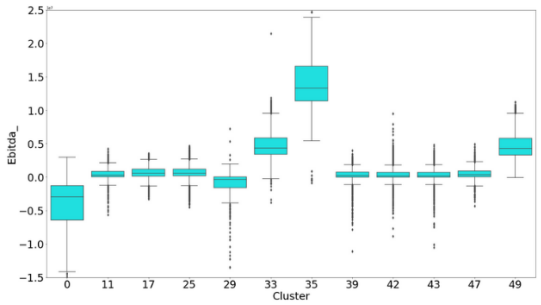

Figura 30. K-means, ebitda vs. clústeres

Fuente: elaboración propia.

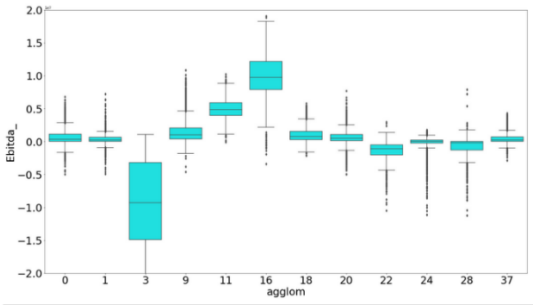

Figura 31. Aglomerativo, ebitda vs. clústeres

Fuente: elaboración propia.

- Grupo 1: ebitda negativo inferior al 0.

  Clústeres: 0.

- Grupo 2: ebitda que inicia con un intervalo negativo de -999 a 2.000 millones. Pero es importante mencionar que más de 50% de los datos tienen ebitda de 0 hasta 1.000 millones.

  Clústeres: 43, 47, 11, 25, 42, 39 y 29.

- Grupo 3: ebitda, entre 2.001 a 7.000 millones.

  Clústeres: 33 y 49.

- Grupo 4: ebitda que va entre 8.001 a 18.000 millones.

  Clústeres: 35.

- Grupo 1: ebitda negativo inferior al 0.

  Clústeres: 3 y 22.

- Grupo 2: ebitda que inicia con un intervalo negativo de -999 a 3.000 millones. Pero es importante mencionar que más de 40% de los datos tienen ebitda de 0 hasta 1.000 millones.

  Clústeres: 28, 24, 1, 20, 37, 0, 18 y 9.

- Grupo 3: ebitda, entre 1 a 3.000 millones.

  Clústeres: 18 y 9.

- Grupo 4: ebitda que va entre 3.001 a 20.000 millones.

  Clústeres: 11 y 16.

4.2.1.8 Índice de endeudamiento versus clústeres

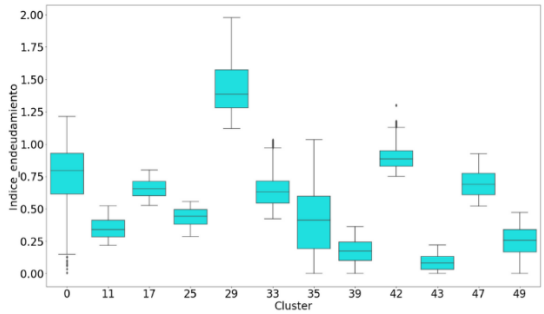

Figura 32. K-means, índice de endeudamiento vs. clústeres

Fuente: elaboración propia.

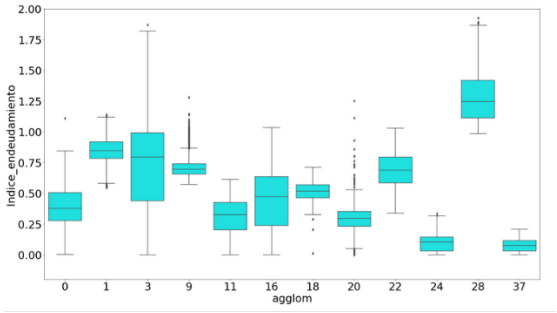

Figura 33. Aglomerativo, índice de endeudamiento vs. clústeres

Fuente: elaboración propia.

- Grupo 1: pymes con índice de endeudamiento del 0 al 40%.

  Clústeres: 43, 39 y 49.
- Grupo 2: pymes con índice de endeudamiento entre 20 a 80%.

  Clústeres: 11, 25, 33, 47, 35 y 17.
- Grupo 3: pymes con índice de endeudamiento entre 80 a más del 100%.

  Clústeres: 0, 42 y 29.

- Grupo 1: pymes con índice de endeudamiento del 0 al 50%.

  Clústeres: 37, 24 y 11.
- Grupo 2: pymes con índice de endeudamiento entre 11 a 70%.

  Clústeres: 16, 20, 0 y 18.
- Grupo 3: pymes con índice de endeudamiento entre 60 a más del 100%.

  Clústeres: 22, 9, 1, 3 y 28

Según la sección 4.2.1 “Análisis de los resultados generados desde los modelos K-means y aglomerativo versus cifras e indicadores financieros” se puede concluir:

- A pesar de que los dos modelos generan divisiones similares en cada uno de sus clústeres, se puede observar que el modelo aglomerativo realiza agrupaciones con menores *ouliers* y

menor dispersión en cada uno de sus intervalos, por lo tanto, sobre el análisis realizado en el modelo aglomerativo se presentará un informe de los resultados y sobre este se plantearán estrategias de acuerdo con las condiciones económicas de los clústeres de las pymes colombianas.

Con los datos de las pymes colombianas durante los cuatro periodos evaluados y los clústeres generados desde el método aglomerativo, se observa la creación de tres grupos importantes con características diferenciales, que serán descritos a continuación:

- Grupo 1: en este grupo están las pymes con márgenes negativos, endeudamiento superior al 61% y un ebitda inferior a 1.000 millones.

  Los clústeres que pertenecen a este grupo son: 3, 28, 22 y 1 que tienen en total el 20,55% de los datos de los doce clústeres principales de las pymes colombianas.
- Grupo 2: en este grupo se encuentran las pymes que inician con un -10% hasta el 40% de sus márgenes y un endeudamiento inferior al 60%, endeudamiento superior hasta el 60% y un ebitda inferior a 1.000 millones.

  Los clústeres que pertenecen a este grupo son: 24, 37, 0 y 20, que tienen en total el 40,65% de los datos de los doce clústeres principales de las pymes colombianas.
- Grupo 3: a este grupo pertenecen las pymes con márgenes positivos, con un endeudamiento entre el 11 y 80% y un ebitda entre 0 a 20.000 millones.

  Los clústeres que pertenecen a este grupo son: 16, 11, 9 y 18, que tienen en total el 38,7% de los datos de los doce clústeres principales de las pymes colombianas.

Es importante mencionar que el valor de ingresos y activos de los clústeres generados no presentan ninguna dispersión específica frente a los tres grupos principales evaluados en los párrafos anteriores. Es decir, el valor de los activos en los tres grupos presenta intervalos diversos que van desde 0 a 20.000, 30.000 y 40.000. Con respecto a los ingresos, se puede evidenciar algo similar, ya que los intervalos de los grupos pueden ir de 0 a 10.000, 20.000, 30.000, 40.0000 o más.

4.2.2 Evaluación de número de periodos de cada empresa en los doce clústeres principales K-means y aglomerativo

Para concluir el análisis de los doce clústeres principales para los dos modelos, se evaluó el número de periodos en los que aparece cada empresa en cada uno de los clústeres.

Al revisar los datos, podemos notar que únicamente en un clúster 40% para K-means y 36% para aglomerativo de los datos aparecen los cuatro periodos, es decir, se puede decir que las cifras financieras de las pymes tienen variaciones significativas y volatilidad, lo cual genera que estas no permanezcan dentro de la misma agrupación o clústeres en sus periodos (véase anexos 19 y 20)

4.2.3 Clústeres restantes generados desde el modelo K-means y aglomerativo

Con respecto a los 38 clústeres pendientes que corresponden al 1,56% de los datos, son valores aislados, y generan agrupaciones pequeñas donde la agrupación que contiene más datos es de 125 para K-means y 237 para aglomerativo (véase anexos 22 y 23). Y para algunas de ellas generan agrupaciones de un solo dato, por lo cual se optó por analizar en esta investigación los 12 clústeres principales que corresponden al 98,44% de los datos.

5. FORMULACIÓN DE ESTRATEGIAS

Teniendo en cuenta los resultados generados desde el modelo de machine learning aglomerativo, se formularon estrategias generales para todos los grupos; en la tabla 11 se muestran estrategias específicas de acuerdo con la necesidad de cada segmentación generada por el modelo.

5.1 Estrategias generales

- Crear accesos para la implementación de nuevas tecnologías mediante la evaluación de nuevos proyectos, revisando su viabilidad y fuentes de financiación, ya sea mediante capital o acceso a crédito.
- Educación organizacional, donde la gerencia o socios de las pymes colombianas creen un gobierno corporativo con planes y pensamientos estratégicos.
- Educación financiera y estratégica, donde las pymes colombianas vean la importancia de tener indicadores de gestión, presupuestos definidos, flujos de caja y planes de inversión y expansión en corto, mediano y largo plazo.
- La TRM es una tasa fluctuante en el país que afecta directamente el costo de las materias primas o productos de las pymes en Colombia, por lo tanto, es importante crear una línea especial para pymes que permita reducir el riesgo cambiario, ya sea mediante la implementación de *forwards*, coberturas, opciones, *swaps*, entre otros.

5.2 Estrategias específicas de acuerdo con la necesidad de cada clúster

Tabla 11. Estrategias según resultados del modelo de machine learning aglomerativo

GRUPO 1 **Márgenes negativos** **Endeudamiento superior al 61%** **N.º datos = 8.091 (20,55%)**	**GRUPO 2** **Intervalos de márgenes negativos** **Endeudamiento inferior al 60%** **N.º datos = 16.009 (40,65%)**	**GRUPO 3** **Márgenes positivos** **Endeudamiento entre el 11 y 80%** **N.º datos = 15.278 (38,7%)**
• Estudios de la estructura de costos y gastos para que a partir de estos se genere un plan de acción de acuerdo con las necesidades de cada pyme y permita crear eficiencias, ya sea por medio de economías de escala u otro método. • Estos clústeres presentan un endeudamiento superior al 61%, por lo anterior, es importante impartir educación financiera donde se traten conceptos como por qué adquirir deuda, cómo cubrir el pago de deuda, flujos de caja que permitan ver la viabilidad de la deuda y tener afectaciones en el capital de trabajo. • Restructuración operativa y financiera para encontrar eficiencias y generar utilidades	• Estudios de la estructura de costos y gastos para que a partir de estos se genere un plan de acción de acuerdo con las necesidades de cada pyme y permita crear eficiencias, ya sea por medio de economías de escala u otro método. • Restructuración operativa para encontrar eficiencias y generar utilidades.	• Acompañamiento para crear eficiencias, ya sea por automatización de procesos e innovaciones de productos o servicios. • Educación financiera para estudiar si el endeudameinto es es el óptimo. • Crear planes de expansión y de *marketing*.

Fuente: elaboración propia.

6. CONCLUSIONES Y RECOMENDACIONES

- El resultado de la agrupación de las 10.001 pymes colombianas evaluadas, correspondientes a los periodos 2016, 2017, 2018 y 2019, y de los métodos empleados (K-means y clúster aglomerativo), es la generación de doce clústeres principales que tienen el 98,44% de los datos. Al evaluar estos doce clústeres se concluye que el método clúster aglomerativo presenta mejores resultados de agrupación, ya que en sus clústeres tiene menos *outliers* y menos dispersión en sus intervalos.

- La agrupación de los datos de las pymes colombianas evaluadas con el método de clúster aglomerativo genera tres grupos diferenciales. Un primer grupo donde se encuentran el 20,65% de los datos de las pymes, las cuales tienen márgenes negativos y un endeudamiento superior al 61%. Un segundo grupo que equivale al 40, 65% de los datos de pymes, que inician con un intervalo entre el -10% hasta el 40% de sus márgenes y un endeudamiento inferior al 60%. Un tercer grupo con el 38,7% de los datos de las pymes, que tienen márgenes positivos y un endeudamiento entre el 11 y 80%.

- De los tres grupos principales se generaron las siguientes estrategias: estudio de la estructura de costos, educación financiera, planes de marketing, automatización de procesos, generación de innovaciones y métodos de expansión y planificación

financiera y tributaria. Estas estrategias van dirigidas a los cuatro grupos de acuerdo con la condición económica de cada uno de ellos.

- Se recomienda, con esta metodología, crear actualizaciones que permitan entender el comportamiento financiero de las pymes y crear estrategias teniendo en cuenta la necesidad de estas.

- Otra recomendación es la de utilizar otros métodos de agrupamiento no supervisado como DBSCAN, y evaluar si produce mejores resultados de agrupación con respecto al método aglomerativo.

.

7. REFERENCIAS

Allen, F., Myers, S. y Brealey, R. (2010). *Principios de finanzas corporativas*. McGraw-Hill.

Amat, J. (2020). Clustering con Python. https://www.cienciadedatos.net/documentos/py20-clustering-con-python.html

Amr, T. (2020). *Hands-On Machine Learning with scikit-learn and scientific Python Toolkits : A practical guide to implementing supervised and unsupervised machine learning algorithms in Python.* Packt Publishing.

Banco Interamericano de Desarrollo (BID) (2017). *Evaluación del Apoyo Directo de la CII a las PYME*. Banco Iberoamericano para el Desarrollo.

Capa, L., Benítez, R. y Rojas, W. (2017). El clúster: una alternativa para la competitividad de las pymes de banano orgánico en Ecuador. *Conference Proceedings UTMACH*, *1*(1), 138-144. https://investigacion.utmachala.edu.ec/proceedings/index.php/utmach/article/view/214.

Capó-Vicedo, J., Expósito-Langa, M. y Masiá-Buades, E. (2007). La importancia de los clústers para la competitividad de las pyme en una economía global. *Revista Eure 33*(98), 119- 133. https://www.scielo.cl/scielo.php?script=sci_arttext&pid=S0250-71612007000100007.

Dangeti, P. (2017). *Statistics for machine learning : build supervised, unsupervised, and reinforcement learning models using both Python and R.* Packt Publishing.

Estrada, V. y Monsalve, S. M. (2014). *Diagnóstico financiero de la empresa Procopal S.A.* [tesis de grado]. Universidad de Medellín, Medellín, Colombia.

Gitman, L. J. y Zutter, C. J. (2012). *Principios de administración financiera.* Pearson Educación.

López, L. M. y Calderón, G. (2006). Análisis de las dinámicas culturales al interior de un clúster empresarial. Estudios Gerenciales, (99), 13-37. http://www.scielo.org.co/pdf/eg/v22n99/v22n99a01.pdf.

López, L. O., Aldape, M. y Yáñez, C. (2007). *Neurona artificial de McCulloch & Pitts*. Instituto Politécnico Nacional.

Machine Learning: definición, funcionamiento, usos (2021). *DataScientest*. https://datascientest.com/es/machine-learning-definicion-funcionamiento-usos.

Marsden, P. (21 de agosto de 2017). Artificial intelligence timeline infographic. Froma Eliza to Tay and beyond. https://digitalwellbeing.org/artificial-intelligence-timeline-infographic-from-eliza-to-tay-and-beyond/.

"Mipymes representan más de 90%del sector productivo nacional y generan el 80% del empleo en Colombia": ministra Alicia Arango (26 de septiembre de 2019). https://www.mintrabajo.gov.co/web/guest/prensa/comunicados/2019/septiembre/mipymes-representan-mas-de-90-del-sector-productivo-nacional-y-generan-el-80-del-empleo-en-colombia-ministra-alicia-arango.

Molina, J. E. (2020). Análisis bibliométrico del uso de machine learning en finanzas a través de un modelo K-Means. http://ediciones.ascolfa.edu.co/index.php/eficiencia/article/view/104/109.

Morales, M. E., Sanabria, S. E. y Arias, M. A. (2010). Acumulación de conocimiento, innovación y competitividad en aglomeraciones empresariales. *Revista Facultad de Ciencias Económicas: Investigación y Reflexión*, *18*(2), 19-53. http://www.scielo.org.co/pdf/rfce/v18n2/v18n2a03.pdf.

Müller, A. C. y Guido, S. (2017). *Introduction to machine learning with Python: A guide for data scientists*. O'Reilly Media.

Nalda, V. (29 de septiembre de 2020). Machine learning: los orígenes y la evolución. https://www.futurespace.es/machine-learning-los-origenes-y-la-evolucion/#:~:text=Estamos%20en%20los%20inicios%20del,File%20System'%20(GFS).

NIIF para las PYMES (s. f.). https://www.nicniif.org/home/normas/niif-para-las-pymes.html.

Normas Internacionales de Información Financiera (2009). https://www2.deloitte.com/co/es/pages/ifrs_niif/normas-internacionales-de-la-informacion-financiera-niif---ifrs-.html.

Ordoñez, C., Ordonez, A., Méndez, C. y Ordoñez, H. (2020). Evaluación e implementación de técnicas de clustering para un sistema de recuperación de documentos judiciales. *Revista Ibérica de Sistemas e Tecnologias de Informacao*, *33*(1), 141-151.

Patel, A. (2019). *Hands-On unsupervised learning using Python: How to build applied Machine Learning Solutions from unlabeled data*. O'Reilly Media.

República de Colombia (2 de agosto de 2004). Congreso de Colombia. Ley 905 de 2004. "Por medio de la cual se modifica la Ley 590 de 2000 sobre promoción del desarrollo de la micro, pequeña y mediana empresa colombiana y se dictan otras disposiciones". https://minciencias.gov.co/node/297.

República de Colombia (29 de agosto de 2013). Ministerio de Hacienda y Crédito Público. Decreto 1851 de 2013. file:///C:/Users/USER/Downloads/dec1851_13.pdf.

República de Colombia (5 de junio de 2019). Ministerio de Comercio Industria y Turismo. Decreto 957 de 2019.

https://www.mincit.gov.co/normatividad/decretos/2019/decreto-957-por-el-cual-se-adiciona-el-capitulo-13.

Ricra, M. (2014). Análisis financiero en las empresas. *Actualidad Empresarial. Revista de Investigación y Negocios*, 1-40.

Roman, V. (12 de junio de 2019). Aprendizaje no supervisado en Machine Learning: agrupación. https://medium.com/datos-y-ciencia/aprendizaje-no-supervisado-en-machine-learning-agrupaci%C3%B3n-bb8f25813edc.

Ueki, Y., Tsuji, M. y Cárcamo, R. (2005). *Tecnología de la Información y las Comunicaciones (TIC) para el fomento de las pymes exportadoras en América Latina y Asia oriental*. Comisión Económica para América Latina y el Caribe (CEPAL).

8. ANEXOS

Anexo 1

Clúster	N.º datos	INGRESOS K-MEANS (millones de pesos COP)		
		% Datos	Intervalos (COP)	Concentración datos (COP)
43	4.200	87,50%	0 a 3.000 m	61,05% = 0 a 1.000 m
29	386	68,91%	0 a 6.000 m	23,32% = 0 a 1.000 m
11	3.169	74,28%	0 a 4.000 m	39,73% = 0 a 1.000 m
47	2.250	83,16%	0 a 10.000 m	32,13% = 0 a 2.000 m
39	4.698	76,03%	0 a 20.000 m	56,56% = 0 a 10.000 m
25	7.582	87,22%	0 a 30.000 m	46,56% = 0 a 10.000 m
17	7.584	80,92%	0 a 30.000 m	37,59% = 0 a 10.000 m
42	5.022	84,95%	0 a 30.000 m	48,09% = 0 a 10.000 m
49	1.529	8,02%	0 a 40.000 m	39,76% = 0 a 10.000 m
0	528	82,20%	0 a 40.000 m	50,19% = 0 a 10.000 m
35	480	75%	<=100.000 m	19,58% = 0 a 10.000 m
33	1.953	82,64%	<=100.000 m	16,64% = 20.000 a 30.000 m

Anexo 2

Clúster	N.º datos	INGRESOS AGLOMERATIVO (millones de pesos COP)		
		% Datos	Intervalos (COP)	Concentración datos (COP)
37	4.044	85,58%	0 a 3.000 m	59,08% = 0 a 1.000 m
0	4.347	72,81%	0 a 4.000 m	72,81% = 0 a 4.000 m
24	1.717	85,38%	0 a 10.000 m	37,57% = 0
28	825	78,90%	0 a 10.000 m	33,93% = 0 a 1.000 m
20	5901	82,32%	0 a 20.000 m	57,91% = 0 a 10.000 m
1	5.932	77,34%	0 a 20.000 m	56,7% = 0 a 10.000 m
22	1.240	73,71%	0 a 20.000 m	52,42% = 0 a 10.000 m
3	303	76,90%	0 a 30.000 m	41,91% = 0 a 10.000 m
9	5.463	71,96%	0 a 30.000 m	27,48% = 10.000 a 20.000 m
18	7852	81,57%	0 a 30.000 m	38,42% = 10.000 a 20.000 m
11	1.138	74,17%	0 a 40.000 m	36,64% = 0 a 10.000 m
16	825	69,09%	< =75.000 m	17,33% = 0 a 10.00 m

Anexo 3

Clúster	N.º datos	MARGEN BRUTO K-MEANS (%)		
		% Datos	Intervalos (%)	Concentración datos (%)
29	386	69,69%	7,77% <-10% y 61,92% = -9% a 40%	15,03% = 20 a 30%
0	528	77,27%	12,50% <-10% y 64,77% = -9% a 50%	17,23% = 10 a 20%
39	4.698	92,32%	-10 a 50%	21,16% = 20 a 30%
17	7.584	82,89%	0 a 40%	30,42% = 10 a 20%
42	5.022	67,94%	0 a 40%	25,43% = 10 a 20%
25	7.582	83,85%	10 a 50%	28,45% = 20 a 30%
33	1.953	73,48%	10 a 50%	25,91% = 10 a 20%
49	1.529	87,44%	20 a 100%	29,37% = 90 a 100%
35	480	95,21%	20 a 100%	31,67% = 90 a 100%
11	3.169	86,94%	70 a 100%	73,27% = 90 a 100%
47	2.250	96,31%	70 a 100%	72,31% = 90 a 100%
43	4.200	84,62%	90 a 100%	84,62% = 90 a 100%

Anexo 4

Clúster	N.º datos	MARGEN BRUTO AGLOMERATIVO (%)		
		% Datos	Intervalos (%)	Concentración datos (%)
24	1.717	84,45%	-9 a 30%	39,72% = -9% a 0
28	825	65,26%	-9 a 50%	13,80% = -9 a 0%
22	1.240	71,61%	1 a 50%	21,13% = 11 a 20%
1	5.932	66,25%	1 a 60%	21,02% = 11 a 20%
18	7852	73,97%	11 a 40%	29,25% = 11 a 20%
9	5.463	75,22%	11 a 40%	29,91% = 11 a 20%
20	5901	80,83%	11 a 60%	22,05% = 21 a 30%
16	825	94,55%	11 a 90%	22,30% = 91 a 100%
11	1.138	85,68%	21 a 100%	30,76% = 91 a 100%
0	4.347	83,57%	91 a 100%	83,57% = 91 a 100%
37	4.044	76,76%	91 a 100%	76,76% = 91 a 100%
3	303	68,65%	53,47% = -10 a 40% y 15,18% = 91 a 100%	15,18% = 91 a 100%

Anexo 5

Clúster	N.º datos	MARGEN OPERATIVO K-MEANS		
		% Datos	Intervalos (%)	Concentración datos (%)
0	528	89,58%	< = 0%	65,91% < = -10%
29	386	91,45%	41,16% < -10% y 50,26% = -9 a 10%	41,191% < = -10%
11	3.169	66,27%	12,56% <-10% y 53,51% = 1 a 40%	17,17% = 11 a 20%
43	4.200	52,93%	16,14% <-10% y 36,79% = 1 a 40%	16,14% < = -10%
39	4.698	84,21%	-11 a 20%	42,64% = 0% a 10%
42	5.022	85,56%	-11 a 20%	56,17% = 0% a 10%
17	7.584	87,10%	1 a 20%	70,61% = 1 a 10%
25	7.582	84,82%	1 a 20%	62,28% = 1 a 10%
33	1.953	87,30%	1 a 30%	37,99% = 11 a 20%
47	2.250	63,42%	1 a 40%	26,27% = 1 a 10%
49	1.529	81,82%	11 a 100%	23,87% = 11 a 20%
35	480	75,76%	11 a 100%	21,04% = 11 a 20%

Anexo 6

Clúster	N.º datos	MARGEN OPERATIVO AGLOMERATIVO		
		% Datos	Intervalos (%)	Concentración datos (%)
3	303	91,41%	<= 0%	69,97% = <-10%
24	1.717	87,89%	14,68% < = -10 y 73,21% = -9 a 10%	44,96% = -9 a 0%
22	1.240	82,66%	40,65% < = -10 y 42,04% = -9 a 10%	42,02% = -9 a 0%
28	825	90,26%	36,69% < = -10 y 5357% = -9 a 10%	36,69% < = -10%
37	4.044	60,95%	13,58% < = -10 y 47,38% = 1 a 50%	13,58% < = -10%
0	4.347	71,38%	13% < = -10 y 58,39% = 1 a 50%	16,06% = 11 a 20%
1	5.932	82,50%	-11 a 20%	51,75% = -9% a 5%
9	5.463	89,62%	1 a 20%	68,37% = 1 a 10%
18	7852	89,00%	1 a 20%	65,61% = 1 a 10%
20	5901	75,31%	1 a 20%	51,18% = 1 a 10%
16	825	66,67%	1 a 40%	25,58% = 11 a 20%
11	1.138	85,68%	21 a 100%	30,76% = 91 a 100%

Anexo 7

Clúster	N.º Datos	MARGEN NETO K-MEANS		
		% Datos	Intervalos (%)	Concentración datos (%)
0	528	94,13%	< = 0%	73,30% < = -10%
29	386	76,68%	< = 0%	47,67% < = -10%
43	4.200	60,00%	18,14% <-10% y 41,86% = 1 a 40%	18,14% < = -10%
11	3.169	70,21%	15,75% <-10% y 54,47% = 1 a 30%	28,56% = 0% a 10%
47	2.250	70,27%	12,84% <-10% y 57,42% = 1 a 20%	42,89% = 0 a 10%
42	5.022	84,49%	-11 a 4%	68,38% < = 2%
39	4.698	77,71%	-11 a 10%	31,16% < =0%
17	7.584	93,13%	-11 a 10%	65,84% < = 3%
25	7.582	87,88%	-1 a 10%	68,33% < = 5%
33	1.953	87,05%	1 a 20%	68,61% = 0 a 10%
35	480	76,46%	1 a 100%	28,33% = 0 a 10%
49	1.529	80,51%	1 a 100%	24,98% = 10% a 20%

Anexo 8

Clúster	N.º datos	MARGEN NETO AGLOMERATIVO		
		% Datos	Intervalos (%)	Concentración datos (%)
3	303	93,07%	< = 0%	74,59%< =-10%
28	825	75,65%	< = 0%	43,67% = < = -10%
24	1.717	75,42%	-9 a 3%	44,85% = -10% 0
1	5.932	82,87%	-9 a 10%	70,14% = -10 a 4%
22	1.240	89,76%	47,26% < = -10% y 42,50% = 1 a 10%	47,26% < = -10%
0	4.347	74,19%	16,54% < = -10% y 57,65% = 1 a 40%	25,51% = 1 a 10%
37	4.044	60,98%	15,45% < = -10% y 45,52% = 1 a 40%	15,75% = 1 a 10%
20	5901	88,85%	-9 a 20%	66,94 = 1 a 10%
9	5.463	73,90%	1 a 6%	73,90% = 1 a 6%
18	7852	75,66%	1 a 6%	14,06% = 1 a 2%
16	825	66,79%	1 a 30%	38,18% = 1 a 10%
11	1138	62,83%	1 a 30%	31,99% = 1 a 10%

Anexo 9

Clúster	N.º datos	ACTIVOS K-MEANS		
		% Datos	Intervalos (%)	Concentración datos (%)
29	386	88,60%	0 a 20.000 m	73,06% = 0 a 10.000 m
25	7.582	80,35%	0 a 20.000 m	50,24% = 0 a 10.000 m
17	7.584	78,10%	0 a 20.000 m	49,29% = 0 a 10.000 m
39	4.698	78,95%	0 a 20.000 m	44,72% = 0 a 10.000 m
47	2.250	89,82%	0 a 30.000 m	46,80 = 0 a 10.000 m
42	5.022	84,35%	0 a 30.000 m	45,86% = 0 a 10.000 m
11	3.169	85,64%	0 a 30.000 m	41,84% = 0 a 10.000 m
43	4.200	82,17%	0 a 30.000 m	35,02% = 0 a 10.000 m
0	528	88,77%	0 a 50.000 m	37,12% = 0 a 10.000 m
49	1.529	79,33%	0 a 50.000 m	28,78% = 10.001 a 20.000 m
33	1.953	73,09%	0 a 50.000 m	22,53% = 10.001 a 20.000 m
35	480	54,58%	> = 50.000 m	33,75% > = 100.000 m

Anexo 10

Clúster	N.º datos	ACTIVOS AGLOMERATIVO		
		% Datos	Intervalos (%)	Concentración datos (%)
28	825	86,85%	0 a 20.000 m	49,84% = 0 a 3.000
20	5901	80,43%	0 a 20.000 m	48,64% = 0 a 10.000 m
1	5.932	87,81%	0 a 30.000 m	49,26% = 0 a 10.000 m
18	7852	88,86%	0 a 30.000 m	46,87% = 0 a 10.000 m
24	1.717	84,51%	0 a 30.000 m	41,53% = 0 a 10.000 m
22	1.240	78,95%	0 a 30.000 m	41,45% = 0 a 10.000 m
0	4.347	85,16%	0 a 30.000 m	41,27% = 0 a 10.000 m
9	5.463	80,58%	0 a 30.000 m	39,72% = 0 a 10.000 m
37	4.044	83,68%	0 a 30.000 m	36,40% = 10.001 a 20.000 m
3	303	67,00%	0 a 40.000 m	27,06% = 0 a 10.000 m
11	1.138	70,39%	0 a 40.000 m	26,98% = 10.001 a 20.000 m
16	825	50,06%	> = 10.000 m	13,82% = 10.001 a 20.000

Anexo 11

Clúster	N.º datos	ROA K-MEANS		
		% Datos	Intervalos (%)	Concentración datos (%)
0	528	94,13%	< = 0%	71,21% < = -10%
29	386	76,17%	< = 0%	49,22% < = -10%
17	7.584	78,09%	-11 a 10%	78,09 = 1 a 10%
47	2.250	89,96%	-11 a 10%	71,16% = 1 a 10%
43	4.200	89,76%	-11 a 10%	68,83% = 1 a 10%
11	3.169	87,91%	-11 a 10%	68,38% = 1 a 10%
42	5.022	94,24%	-11 a 10%	66,51% = 1 a 10%
39	4.698	85,63%	-11 a 10%	61,96% = 1 a 10%
25	7.582	89,99%	-11 a 20%	73,53% = 1 a 10%
33	1.953	82,95%	1 a 20%	54,63% = 1 a 10%
49	1.529	92,94%	1 a 30%	36,76 = 11 a 20%
35	480	73,96%	1 a 30%	34,38% = 1 a 10%

Anexo 12

Clúster	N.º datos	ROE K-MEANS		
		% Datos	Intervalos (%)	Concentración datos (%)
0	528	87,69%	<= -10%	87,69% < = -10%
47	2.250	80,71%	8,62% < = -10% y 72,09% = -9 a 30%	42,93% = 1 a 10%
29	386	65,80%	10,88% < = -10% y 54,92% = 0 a 50%	14,77% = 1 a 10%
42	5.022	69,51%	19,79% < = -10% y 49,72% = 0 a 30%	24,89% = 1 a 10%
43	4.200	88,21%	-9 a 10%	67,57% = 1 a 10%
11	3.169	77,60%	-9 a 10%	59,70% = 1 a 10%
39	4.698	79,52%	-9 a 10%	56,46% = 1 a 10%
25	7.582	86,69%	-9 a 20%	51,64% = 1 a 10%
17	7.584	86,84%	-9 a 30%	42,34% = 1 a 10%
49	1.529	85,35%	1 a 40%	32,83% = 11 a 20%
33	1.953	77,88%	1 a 50%	23,25% = 11 a 20%
35	480	78,75%	1 a 60%	23,96% = 11 a 20%

Anexo 13

Clúster	N.º datos	ROA AGLOMERATIVO		
		% Datos	Intervalos (%)	Concentración datos (%)
3	303	97,03%	< = 0%	72,94% < = -10%
22	1.240	89,68%	< = 0%	51,13%< = -10%
28	825	95,78%	39,61% < = -10% y 56,17% = 1 a 10%	39,61%< = -10%
1	5.932	77,82%	-9 a 3%	77,82% = -9 a 3%
24	1.717	70,30%	-9 a 3%	40,77% = -9 a 0%
0	4.347	83,74%	-9 a 10%	57,23% = -9 a 3%
37	4.044	91,59%	-9 a 10%	74,36% = -9 a 4%
9	5.463	80,21%	1 a 10%	68,68% = 1 a 7%
18	7852	89,43%	1 a 20%	71,80% = 1 a 7%
20	5901	83,17%	1 a 20%	70,01% = 1 a 7%
16	825	78,67%	1 a 30%	36,73% = 1 a 10%
11	1.138	86,12%	1 a 30%	35,15% = 11 a 20%

Anexo 14

Clúster	N.º datos	ROE AGLOMERATIVO		
		% Datos	Intervalos (%)	Concentración datos (%)
22	1.240	74,35%	< = -10%	74,35%< = -10%
3	303	81,85%	< = 0%	77,89% < = -10%
24	1.717	71,93%	-9 a 4%	12,52% = 1 a 1%
37	4.044	90,38%	-9 a 10%	72,43% = -9 a 4%
0	4.347	83,74%	-9 a 20%	52,84% = 1 a 10%
1	5.932	77,01%	-9 a 30%	33,88% = 1 a 10%
20	5901	79,17%	1 a 20%	57,04% = 1 a 10%
9	5.463	73,77%	1 a 30%	35,07% = 1 a 10%
18	7852	85,32%	1 a 30%	47,40% = 1 a 10%
11	1.138	80,40%	1 a 40%	29,44% = 11 a 20%
16	825	73,09%	1 a 50%	22,67% = 11 a 20%
28	825	70,62%	44,81% < = 20% y 25,81% > = 100%	25,81% > = 100%

Anexo 15

Clúster	N.º datos	EBITDA K – MEANS		
		% Datos	Intervalos (%)	Concentración datos (%)
0	528	96,21%	< = 0	79,36% = < -1.000 m
43	4.200	100,00%	-999 a 1.000 m	73,90% = 0 a 1.000 m
47	2.250	90,44%	-999 a 2.000 m	55,87% = 0 a 1.000 m
17	7.584	80,81%	-999 a 2.000 m	55,38% = 0 a 1.000 m
11	3.169	90,85%	-999 a 2.000 m	54,62% = 0 a 1.000 m
25	7.582	88,26%	-999 a 2.000 m	54,37% = 0 a 1.000 m
42	5.022	88,26%	-999 a 2.000 m	51,29% = 0 a 1.000 m
39	4.698	91,15%	-999 a 2.000 m	50,57% = 0 a 1.000 m
29	386	97,15%	-999 a 1.000 m	41,97% = 0 a 1.000 m
33	1.953	80,54%	2.001 a 7.000 m	25,86% = 3.001 a 4.000 m
49	1.529	81,43%	2.001 a 7.000 m	24,98% = 3.001 a 4.000 m
35	480	82,08%	8.001 a 18.000 m	13,96% = 11.001 a 12.000 m

Anexo 16

Clúster	N.º datos	EBITDA AGLOMERATIVO		
		% Datos	Intervalos (%)	Concentración datos (%)
3	303	97,03%	< = 0 m	84,49% < = -1.000 m
22	1.240	97,26%	< = 0 m	53,23% < = -1.00 m
28	825	68.67%	-999 a 1.000 m	44,64% = -999 a 0 m
24	1.717	81,95%	-999 a 1.000 m	42,52% = -999 a 0 m
1	5.932	86,01%	-999 a 1.000 m	63,20% = 0 a 1.000 m
20	5901	88,76%	-999 a 2.000 m	53,87% = 0 a 1.000 m
37	4.044	92,95%	-999 a 2.000 m	58,31% = 0 a 1.000 m
0	4.347	83,94%	-999 a 3.000 m	48,38% = 0 a 1.000 m
18	7852	83,99%	1 a 3.000 m	51,80% = 0 a 1.000 m
9	5.463	78,64%	1 a 3.000 m	42,96% = 0 a 1.000 m
11	1.138	83,22%	3.001 a 7.000 m	27,68% = 4.000 a 5.000 m
16	825	83,64%	7.001 a 20.000 m	14,67% = 8.000 a 9.000 m

Anexo 17

Clúster	N.º datos	ÍNDICE DE DEUDAMIENTO K – MEANS		
		% Datos	Intervalos (%)	Concentración datos (%)
43	4.200	95,95%	0 a 20%	59,31% = 0 a 10%
39	4.698	97,53%	0 a 30%	39,95% = 21 a 30%
49	1.529	90,91%	0 a 40%	31,59% = 21 a 30%
35	480	63,30%	0 a 50%	15% = 41 a 50%
11	3.169	95,46%	21 a 50%	39,07% = 41 a 50%
25	7.582	100%	31 a 60%	44,34% = 41 a 50%
33	1.953	77,93%	41 a 80%	30,62% = 51 a 60%
47	2.250	82,31%	51 a 80%	30,27% = 61 a 70%
17	7.584	99,93%	51 a 80%	45% = 61 a 70%
0	528	81,63%	51 a 100%	26,14% = 91 a 100%
42	5.022	91,14%	81 a 100%	43,93% = 81 a 90%
29	386	100,00%	> 100%	90,93% > 100%

Anexo 18

Clúster	N.º datos	ÍNDICE DE DEUDAMIENTO AGLOMERATIVO		
		% Datos	Intervalos (%)	Concentración datos (%)
37	4.044	99,53%	0 a 10%	65,95% = 0 a 10%
24	1.717	93,59%	0 a 20%	46,3% = 0 a 10%
11	1.138	60,73%	11 a 60%	25,04% = 31 a 40%
16	825	83,41%	11 a 60%	15,52% = 51 a 60%
20	5901	76,24%	21 a 40%	38,74% = 31 a 40%
0	4.347	82,22%	21 a 60%	24,61% = 31 a 40%
18	7852	93,95%	41 a 70%	46,64% = 51 a 60%
22	1.240	69,84%	61 a 100%	23,55% = 71 a 80%
9	5.463	88,38%	61 a 80%	52,52% = 61 a 70%
1	5.932	91,81%	71 a 100%	37,19% = 81 a 90%
3	303	63,04%	71 a > = 100%	28,38% = 91 a 100%
28	825	99,19%	> 100%	99,19% >100%

Anexo 19

Clúster	4 periodos		3 periodos		2 periodos		1 periodo		N.º de empresas	N.º de datos
	Cant.	%	Cant.	%	Cant.	%	Cant.	%		
43	612	41%	289	19%	287	19%	311	21%	1.499	4.200
42	584	29%	391	19%	464	23%	585	29%	2.024	5.022
39	551	28%	345	18%	394	20%	671	34%	1.961	4.698
17	739	23%	662	20%	785	24%	1.072	33%	3.258	7.584
25	729	22%	674	21%	796	24%	1.052	32%	3.251	7.582
47	226	22%	157	15%	245	24%	385	38%	1.013	2.250
11	296	21%	254	18%	341	24%	541	38%	1.432	3.169
29	25	11%	22	10%	48	22%	124	57%	219	386
33	93	8%	145	13%	233	20%	680	59%	1.151	1.953
49	57	6%	98	10%	193	20%	621	64%	969	1.529
0	1	0%	12	3%	62	14%	364	83%	439	528
35	14	4%	21	6%	62	19%	237	71%	334	480

Anexo 20

Clúster	4 periodos		3 periodos		2 periodos		1 periodos		N.º de empresas	N.º de datos
	Cantidad	%	Cantidad	%	Cantidad	%	Cantidad	%		
37	549	36%	289	19%	304	20%	373	25%	1.515	4.044
0	490	27%	309	17%	437	24%	586	32%	1.822	4.347
1	645	26%	482	20%	594	24%	718	29%	2.439	5.932
20	546	21%	501	19%	622	24%	970	37%	2.639	5.901
18	665	19%	745	21%	895	26%	1167	34%	3.472	7.852
24	131	15%	120	13%	186	21%	461	51%	898	1.717
28	45	14%	36	11%	80	24%	168	51%	329	616
9	352	13%	484	18%	711	26%	1181	43%	2.728	5.463
16	25	5%	46	8%	109	20%	369	67%	549	825
11	28	4%	57	7%	147	19%	561	71%	793	1.138
3	5	2%	9	4%	28	12%	200	83%	242	303
22	11	1%	64	7%	180	20%	644	72%	899	1.240

Anexo 21

Clúster	N.º datos	N.º emp.	Ing. operacionales	M. Bruto	M. operativo	M. Neto	Activo	Ebitda	M. ebitda	Índ. Deuda	ROA	ROE
1	1	1	1	0%	-60944000%	-60944000%	17.000.392	-665.020	-66502000%	85%	-4%	-24%
2	1	1	15.663.111	14%	-6361%	-6362%	2.200.343	-996.407.589	-6361%	82%	-45287%	-249378%
3	1	1	1	100%	-47593200%	-1783310800%	58.404.436	-475.932	-47593200%	32%	-31%	-45%
5	1	1	7.099	-69095%	-89438%	-91512%	66.235.194	-6.349.205	-89438%	17%	-10%	-12%
6	1	1	330	-22696%	-33099%	-32169%	16.225.812	-109.226	-33099%	100%	-1%	-689%
8	1	1	1	100%	-23401500%	-23404500%	16.375.591	-452.337	-45233700%	6%	-1%	-2%
9	1	1	40.545	75%	2%	1%	1.180	628	2%	15948%	29%	0%
16	1	1	1	100%	-31290100%	-35097600%	16.903.801	-759.578	-75957800%	2%	-2%	-2%
18	1	1	386.633	42%	-99%	-135%	789.619	-601.129	-155%	3332%	-66%	2%
20	1	1	79.484.241	70%	70%	50%	371.158	32.077.734	40%	3034%	10631%	-362%
31	1	1	500	-15264%	-28300%	-40411%	7.092.841	-141.502	-28300%	72%	-3%	-10%
40	1	1	0	0%	0%	0%	120.623	-3.871.173	0%	3130%	-3211%	106%
24	1	1	1	100%	-70442800%	-70442800%	29.850.518	-704.428	-70442800%	54%	-2%	-5%
26	1	1	500	-29903%	-12171%	-72281%	8.693.087	-60.856	-12171%	35%	-4%	-6%
4	4	1	0	0%	0%	0%	37.455	-1.994	0%	6588%	-50%	1%
7	2	2	11.321.476	59%	-14%	-17%	13.248.169	-571.118.763	-8047%	69%	-10%	-32%
12	2	2	6	100%	-14822127%	-16195618%	13.704.000	-816.089	-14822127%	54%	-7%	320%
19	3	3	36.037	-8685%	4395%	-6722%	26.061.112	-449.204	4281%	62%	-1%	-4%
27	5	4	3.284.879.097	52%	700%	580%	1.884.486.833	193.618.397	700%	106%	664%	-227%
28	3	3	1.565.474.870	-337%	-365%	-383%	112.762.993	-592.329.491	-2337%	44%	-1886%	-6050%
34	2	2	1	100%	-3568650%	-910000%	7.826.572	-57.687	-5768650%	42%	0%	-1%
38	3	3	1.630.902	29%	66%	-13%	26.935.716	-2.568.212	53%	100%	-2%	-37227%
41	3	3	2.162.573.685	31%	75%	67%	2.646.030.528	327.402.308	75%	24%	43%	61%
46	2	1	0	0%	0%	0%	15.373	-3.599	0%	5961%	0%	0%
10	12	6	144.248	-3%	-144%	-144%	391.423	-98.711	-147%	1366%	-68%	5%
13	11	8	527.473.804	37%	138%	135%	125.268.071	52.132.018	137%	74%	111%	6385%
14	11	10	549.665	-3830%	-3626%	-5025%	65.700.981	-13.003.091	-4384%	36%	-24%	-28%
15	9	10	8.368.928	-10%	-20611%	-20616%	32.532.994	-49.239.807	-20664%	66%	-360%	-10678%
23	2	2	1	100%	4459350%	1076850%	11.063.752	44.594	4459350%	29%	0%	0%
30	23	23	39.961.637	35%	-49569%	-49668%	61.474.630	-36.090.879	-49659%	105%	-87%	-3325%
32	7	5	487.671	30%	-1160%	-1331%	529.085	-113.358	-1163%	2069%	-58%	3%
36	29	26	324.479	-875%	-5247%	-7255%	36.570.566	-1.497.277	-5596%	78%	-8%	-24%
44	15	11	705.987	19%	-353%	-354%	558.049	-396.643	-355%	852%	-229%	29%
21	122	91	109.734.713	61%	4983%	4322%	208.331.626	30.207.961	4990%	39%	27%	288%
22	58	34	3.386.742	33%	-2828%	-2034%	2.426.775	-1.481.412	-2832%	447%	-85%	26%
37	37	32	327.874.439	45%	2464%	2398%	361.872.019	51.704.307	2464%	47%	23%	1627%
45	119	103	31.430.918	19%	-451%	-768%	67.426.622	-15.488.062	-510%	67%	-65%	-689%
48	125	72	5.104.850	31%	-365%	-275%	8.601.721	-2.333.535	-385%	246%	-44%	31%

Anexo 22

Clúster	N.º datos	N.º empresas	Ing. operacionales	M. Bruto	M. operativo	M. Neto	Activo	Ebitda	M. ebitda	Índ. Deuda	ROA	ROE
2	237	169	107.246.903	56,8%	2882,3%	25,3345311	164.808.975	26.082.802	2886,9%	39,93%	29,03%	187,43%
4	86	50	3.127.780	23,0%	-2051,5%	-15,2795484	3.819.939	-2.279.906	-2058,5%	404,21%	-86,27%	30,22%
5	2	2	6	100,0%	-14822127,3%	-161956,182	13.704.000	-816.089	-14822127,3%	53,78%	-6,85%	320,13%
6	9	9	8.368.928	-9,8%	-20610,6%	-206,161297	32.532.994	-49.239.807	-20664,4%	65,74%	-360,11%	-10678,24%
7	5	5	43	100,0%	-1899074,3%	-18719,3367	14.478.442	-569.232	-2779074,3%	41,37%	-2,77%	3,25%
8	8	8	746.921	-3552,2%	-2322,2%	-25,3868004	86.418.860	-17.848.767	-3364,2%	27,49%	-31,99%	-35,13%
10	13	7	133.186	2,8%	-754,9%	-8,37332348	365.317	-93.917	-757,8%	1396,83%	-68,42%	5,03%
12	6	2	0	0,0%	0,0%	0	30.094	-2.529	0,0%	6378,90%	-33,58%	0,51%
13	3	3	1.565.474.870	-337,4%	-365,5%	-3,82948898	112.762.993	-592.329.491	-2337,4%	44,42%	-1885,86%	-6050,01%
14	12	10	252.954	-1413,8%	-6184,7%	-58,6305435	32.009.902	-1.247.974	-6282,3%	83,98%	-5,96%	-51,51%
15	1	1	79.484.241	70,4%	70,0%	0,49640731	371.158	32.077.734	40,4%	3034,12%	10630,66%	-362,31%
17	15	11	705.987	19,1%	-352,8%	-3,54053425	558.049	-396.643	-354,7%	852,43%	-229,38%	28,85%
19	2	2	11.321.476	58,8%	-13,9%	-0,17141599	13.248.169	-571.118.763	-8046,7%	68,95%	-10,06%	-32,26%
21	11	10	527.473.804	37,3%	138,1%	1,35467505	125.268.071	52.132.018	137,4%	74,01%	110,88%	6384,64%
23	3	3	36.037	-8685,3%	4395,5%	-67,2235303	26.061.112	-449.204	4281,5%	61,53%	-1,29%	-3,77%
25	2	2	1	100,0%	4459350,0%	10768,5	11.063.752	44.594	4459350,0%	29,05%	0,11%	0,18%
26	1	1	1	100,0%	-70442800,0%	-704428	29.850.518	-704.428	-70442800,0%	54,32%	-2,36%	-5,17%
27	1	1	7.099	-69095,4%	-89438,0%	-915,120439	66.235.194	-6.349.205	-89438,0%	16,86%	-9,81%	-11,80%
29	1	1	1	100,0%	-23401500,0%	-234045	16.375.591	-452.337	-45233700,0%	5,78%	-1,43%	-1,52%
30	1	1	500	-15264,0%	-28300,4%	-404,114	7.092.841	-141.502	-28300,4%	72,17%	-2,85%	-10,24%
31	1	1	1	100,0%	-47593200,0%	-17833108	58.404.436	-475.932	-47593200,0%	32,24%	-30,53%	-45,06%
32	1	1	330	-22696,4%	-33098,8%	-321,687879	16.225.812	-109.226	-33098,8%	99,91%	-0,65%	-688,71%
33	1	1	1	100,0%	-31290100,0%	-350976	16.903.801	-759.578	-75957800,0%	1,56%	-2,08%	-2,11%
34	22	20	41.709.271	34,2%	-51758,7%	-518,750614	64.161.010	-36.751.026	-51851,5%	92,86%	-58,31%	-3487,73%
35	1	1	500	-29902,6%	-12171,2%	-722,812	8.693.087	-60.856	-12171,2%	35,39%	-4,16%	-6,43%
36	6	5	568.875	23,1%	-5,9%	-0,25766459	608.593	-126.187	-9,4%	2119,02%	-54,63%	2,88%
38	5	4	3.284.879.097	52,2%	699,6%	5,80250225	1.884.486.833	193.618.397	699,6%	105,77%	663,60%	-226,85%
39	1	1	40.545	75,2%	1,5%	0,0084844	1.180	628	1,5%	15947,80%	29,15%	-0,18%
40	1	1	386.633	41,7%	-98,6%	-1,34689227	789.619	-601.129	-155,5%	3332,36%	-65,95%	2,04%
41	100	62	4.893.918	43,5%	-270,2%	-1,46710567	7.897.192	-1.075.408	-290,5%	224,89%	-28,56%	22,22%
42	3	3	2.162.573.685	30,6%	75,3%	0,67055449	2.646.030.528	327.402.308	75,3%	24,39%	42,55%	61,45%
43	28	26	370.631.373	38,2%	1125,0%	10,4039709	399.611.411	51.110.449	1151,6%	52,19%	20,42%	2132,84%
44	2	2	32.007	-5619,9%	-9314,9%	-161,055774	8.572.542	-29.134	-9314,9%	44,43%	-1,65%	-5,01%
45	1	1	1	0,0%	-60944000,0%	-609440	17.000.392	-665.020	-66502000,0%	85,04%	-3,58%	-23,96%
46	29	29	553.032	-514,4%	-3922,1%	-59,3044164	28.842.802	-2.406.474	-4253,6%	79,12%	-15,71%	9,03%
47	1	1	15.663.111	14,3%	-6361,5%	-63,6185093	2.200.343	-996.407.589	-6361,5%	81,84%	-45286,75%	-249377,79%
48	1	1	0	0,0%	0,0%	0	120.623	-3.871.173	0,0%	3130,03%	-3211,23%	105,98%
49	3	1	1.630.902	29,4%	65,9%	-0,13257544	26.935.716	-2.568.212	52,7%	99,99%	-1,78%	-37226,68%