

TÍTULO

**Estructuras de costos de las empresas industriales y su
impacto en la rentabilidad: un estudio empírico a través de
herramientas de Machine Learning**

Autores:

Christian Kuster

Categoría: resultados de proyectos de investigación

*Este trabajo ha sido aprobado por la Comisión Técnica al sólo efecto de ser publicado en los
congresos del IAPUCo*

Resumen

El trabajo tiene como objetivo verificar si el nivel de ventas, combinado con las estructuras de costo y el tamaño empresarial, tiene efecto cuantificable y predecible en las utilidades. En base a esa premisa, se elaboran modelos matemáticos predictivos utilizando algoritmos de Machine Learning sobre una base de datos de estados contables de empresas industriales uruguayas durante un periodo de 8 años. Los resultados obtenidos indican que en la mayoría de los casos las herramientas de Machine Learning fueron ser eficaces para caracterizar a las empresas, encontrar patrones y proponer modelos predictivos de los costos y beneficios económicos, confirmando que junto al tamaño y los ingresos, éstos son factores centrales para explicar el signo y el monto de los beneficios.

1- Introducción

Este trabajo se propone analizar las estructuras de costos de las empresas industriales como forma de entender su desempeño, dada la importancia para la economía que tiene la marcha de las empresas. En particular se buscará demostrar que los costos de producción y ventas –mayormente variables-, los costos fijos y la utilidad –beneficio económico- son funciones directas del nivel de ventas con coeficientes identificables y confiables desde el punto de vista estadístico, pero dependen del tamaño de las empresas, presentando grandes diferencias de acuerdo a éste. Eso permite elaborar modelos matemáticos con alto poder predictivo, es decir que conocidas las ventas, el tamaño de los activos y la cantidad de personal, es posible predecir el monto de los costos de producción, los costos fijos y la utilidad de las empresas industriales uruguayas.

Para elaborar estos modelos, el trabajo se basa en dos herramientas de Machine Learning: el análisis de clúster mediante el algoritmo K-means y los modelos predictivos en base a regresión lineal por el algoritmo OLS (Ordinary Least Square). El mecanismo de clústerización permite separar a las empresas en grupos (clústers) con rasgos similares de forma avanzada. Una vez obtenidos esos clúster, se utilizan los modelos de regresión a la interna de cada uno de ellos, para buscar patrones en las estructuras de costos y utilidad en relación a las ventas, con fines predictivos.

Desde lo conceptual, el trabajo se aborda desde la Teoría General del Costo, donde encontramos todos los conceptos de costos variables o fijos y su relación con los ingresos, que nos permite plantear hipótesis de vinculación. En segundo lugar, se utilizan conceptos de la teoría de la firma, que estudia las conductas empresariales. Y en tercer lugar se utilizan conceptos de la teoría contable, tales como activos, resultados económicos y márgenes de utilidad, datos que provienen de los estados de resultados y estado de situación financiera de las empresas que conforman la base de datos. Este marco conceptual nos indica qué variables se van a analizar para explicar el desempeño, que conceptos debemos conocer para abordar el análisis y posibles respuestas a las preguntas.

Las hipótesis que se buscan verificar son las siguientes:

Hipótesis general: Es posible predecir el monto de los costos de producción y ventas, los costos fijos y la utilidad de una empresa uruguaya, a partir de conocer sus ventas y su tamaño.

De esta hipótesis general se derivan tres:

H1 Es posible predecir el monto de los **costos de producción variables** de las empresas industriales ya que guardan una relación robusta y estadísticamente significativa con los ingresos, mediante un modelo matemático, conocidos previamente el tamaño de sus activos, ventas y cantidad de empleados (Indicadores de tamaño)

H2 Es posible predecir el monto de los **costos fijos** a partir de los ingresos, conocidos previamente el tamaño de sus activos, ventas y cantidad de empleados (Indicadores de tamaño)

H3 Es posible predecir el **monto y signo de la utilidad de una empresa** a partir de sus ingresos, conocidos previamente el tamaño de sus activos, ventas y cantidad de empleados (Indicadores de tamaño)

Este trabajo se ordena de esta manera: La primera parte introduce el problema a estudiar y las razones para hacerlo. La segunda presenta el marco teórico y los

antecedentes en la literatura a partir de los cuales abordamos el análisis, definiendo qué variables vamos a analizar. La tercera detalla la metodología, la base de datos y la definición de las variables, especialmente la estructura de costos. Al final, la cuarta sección debate los resultados, verifica las hipótesis planteadas al inicio y define posibles líneas a seguir.

2- Marco teórico y antecedentes

En este trabajo analizaremos si es posible predecir la utilidad a partir de las ventas y de las estructuras de costo, partiendo de la base que son determinantes directas del beneficio económico que obtienen las firmas al vender, y que determinadas estructuras pueden provocar utilidades reducidas o negativas, y otras, son favorables al éxito. Al mismo tiempo revisaremos si el efecto de estas estructuras actúa en forma combinada con el tamaño de las firmas, dado por su volumen de ventas, sus activos o la cantidad de personal empleado. Nótese que las ventas, como variable económica, son indicadores tanto de nivel de actividad como de tamaño, al mismo tiempo.

La influencia del tamaño y los ingresos en la rentabilidad es materia de estudio de larga data hasta el presente, y se ha comprobado que influye, pero sin consenso en el signo del impacto. (Hernández y Cáceres, 2024; De la Cuadra et al, 2024; González et al, 2000)

Vasta literatura se concentra en el endeudamiento y pago de intereses, indicando que junto a baja rotación de inventarios y bajos márgenes, provocan reducción de rentabilidad. Por ejemplo el modelo planteado por Altman (1968), fue demostradamente eficaz para tomar cuenta de esas variables y predecir las quiebras utilizando un modelo matemático (Panchi et al, 2022).

Adicionalmente, es importante comprender el rol de las estructuras de costos, es decir, el peso porcentual de los costos variables y de los costos fijos en relación con los ingresos obtenidos. El concepto de costo variable, fijo y su relación con los niveles de actividad están muy bien analizados en las obras de Yardin (2012) y Cartier (2017). Comprender el funcionamiento de estas variables permite interpretar cómo se conforma la utilidad, cómo se llega al el punto de equilibrio empresarial o el margen de seguridad.

En este artículo se analizan realidades financieras de las empresas, en particular industriales. Las conductas empresariales son área de estudio de la teoría de la firma, existiendo antecedentes nacionales, regionales e internacionales. Uno de los aspectos estudiados por esta teoría, y que se recogen en este artículo, es la vinculación entre tamaño (size) y rentabilidad (profitability). Esta línea busca determinar si la decisión de crecer o expandirse afecta la rentabilidad. Sin embargo, y pese a que existen muchos trabajos previos, no muchos enfocan el problema desde los costos, aspecto frecuentemente soslayado. (Galindo, 2005)

La teoría de la firma contiene un aspecto fuertemente predictivo, es decir trata de explicar y predecir la conducta y los resultados económicos de las empresas. Se ha demostrado que esas explicaciones son multifactoriales: el desempeño resulta de una combinación de factores económicos, factores de gestión de recursos, características de los administradores, acceso a tecnología, participación en el mercado y otros. (Becker et al., 2010; Guiso y Rustichini, 2018).

Este aspecto predictivo de la teoría es el que se explota en este trabajo. Según las variables que se planteen como explicativas de la conducta, la teoría de la firma transita por caminos cualitativos y cuantitativos. La teoría económica de la firma, una de las más abordadas, cuenta con una base fundamentalmente neoclásica, y muy relacionada al análisis marginal, ya que busca las explicaciones del beneficio en base a la relación entre precio, volumen de producción, costos y utilidad. Es de corte básicamente internalista y no considera variables externas como la tecnología y las conductas de los grupos de interés. Este trabajo se aproxima especialmente a esta línea de análisis de la realidad económica, ya que no aborda la influencia del medio en la empresa, sino la gestión interna de recursos y su efecto en los resultados (Krugman y Wells, 2006; Favaro, 2013; Kammoun y Alcouffe, 2003; Dos Santos et al., 2015; Galindo Lucas, 2005).

Dentro de la evidencia empírica en base a datos contables, se destacan los trabajos de Goddard y Becker. Goddard y Wilson (2005) estudian la relación entre el tamaño, el peso en el mercado que tiene la firma y su rentabilidad. Concluyen que a mayor tamaño, menor rentabilidad, pero en aquellas empresas de grandes dimensiones, con poder de mercado, ocurre lo contrario. Este antecedente es importante para interpretar los resultados obtenidos en esta investigación, ya que en Uruguay existe un fenómeno de concentración empresarial industrial bastante palpable, y esto se refleja en la rentabilidad.

Mientras muchas investigaciones definen el tamaño a partir de indicadores monetarios como los activos o el volumen de ventas, otros estudios adoptan medidas alternativas, como el número de empleados. Es el caso del trabajo de Becker et al. (2010), quienes estudiaron 109 empresas manufactureras estadounidenses mediante información financiera de Standard and Poor's. Los autores detectaron una relación clara entre tamaño y rentabilidad en solo 11 de las empresas analizadas, lo que sugiere que dicha relación no es generalizable.

Algunos enfoques exploran el vínculo entre el tamaño y el endeudamiento, donde la teoría de la agencia se aplica plenamente. Desde esta teoría se plantea que los conflictos entre propietarios y administradores pueden derivar en altos costos financieros, decisiones ineficientes como la distribución de utilidades en lugar de reinversión, y pérdida de oportunidades estratégicas, factores que, en conjunto, reducen la rentabilidad (Castaño, 1999).

Asimismo, Germán-Soto y Marín López (2023) argumentan que si bien las utilidades pueden fomentar el crecimiento, no se puede asegurar que el crecimiento, por sí mismo, conlleve una mejora en las ganancias.

Además de los aspectos financieros y estructurales, otros estudios se han enfocado en elementos cualitativos relacionados con la innovación, la iniciativa emprendedora, el desarrollo del capital humano y la gestión eficiente de los recursos disponibles. Estas aproximaciones, conocidas como la perspectiva basada en recursos (*Resource-Based View*), consideran que la rentabilidad puede incrementarse mediante ventajas competitivas derivadas de una combinación efectiva de recursos y capacidades empresariales, tangibles como intangibles. Por ejemplo, Barney (2001) y Winter (2003) destacan la importancia de las capacidades gerenciales, mientras que Slater y Olson (2002), Boulding (1952) y Favaro (2013) resaltan el papel que juega la calidad de los procesos internos en la generación de rentabilidad. Como el consumo de recursos es el elemento disparador de costos, la gestión de éstos determina la forma de las estructuras de costo, de ahí el motivo de su estudio.

En resumen, el análisis del desempeño financiero se comprende mejor desde un abordaje integral que contemple factores como el tamaño, así como los relacionados con estructuras internas y estrategias de gestión. Si bien no existe una relación consensuada entre tamaño y rentabilidad, la literatura coincide en que variables contextuales y estructurales inciden, sin acuerdo en el signo del impacto, en el desempeño económico de las empresas.

Los modelos predictivos y el fracaso empresarial: fundamentos, evolución y aplicaciones actuales

Paralelamente a las investigaciones sobre tamaño y rentabilidad empresarial, se ha consolidado una línea de estudio sobre la predicción de la quiebra o el fracaso empresarial (*business failure prediction*). Esta corriente se propone identificar, mediante herramientas cuantitativas, los ratios financieros que predicen si una empresa se encuentra en riesgo de fracasar o de tener éxito. El punto de partida de esta línea es el trabajo de Altman (1968), quien estudió una muestra de 66 empresas manufactureras estadounidenses, la mitad de las cuales quebraron entre los años 1946 y 1965. Su modelo, conocido como **Z Score**, se basó en identificar un conjunto de ratios clave que reflejan aspectos de la situación financiera de una empresa: actividad, solvencia, apalancamiento, liquidez y rentabilidad. El autor construyó una ecuación que los relaciona y arroja un valor para cada empresa, que permite predecir su probabilidad de quiebra (Altman, 1968; Pascale, 1998; Vázquez et al., 2015).

Este trabajo es una referencia metodológica central para esta investigación, que también se apoya en el análisis de estados contables. Además, el trabajo de Altman popularizó el uso del sistema Dupont, particularmente el cálculo del ROA (Rentabilidad sobre activos), indicador que también será aplicado en este estudio (Altman, 1968, p. 606; Pascale, 1998).

Dentro de la literatura reciente, Mojica Carreño (2023) subraya que el fracaso empresarial ha cobrado creciente relevancia por sus impactos en el entorno. Dentro de las causas comúnmente identificadas se destacan la falta de planificación, la competencia agresiva y el elevado nivel de endeudamiento.

El tamaño es un factor determinante de la performance empresarial. Los estudios más recientes coinciden en que las pequeñas y medianas empresas son especialmente vulnerables. Según Roitter y Federico (2022), la llamada "tasa de mortalidad" se ha incrementado tras la pandemia, mientras que Romero Espinosa (2013) ya había identificado patrones de fracaso frecuentes en este sector.

En síntesis, la predicción del fracaso empresarial es una línea que nos ayuda a definir el camino teórico y metodológico. Desde el modelo de Altman (1968) hasta las aplicaciones más recientes, esta línea de investigación ha demostrado ser útil no solo para anticipar escenarios de crisis, sino también para comprender los factores estructurales que afectan la supervivencia empresarial. Este enfoque financiero estructural es clave para este trabajo. Los costos fijos, en especial los financieros, emergen como una variable crítica, especialmente en el caso de las pequeñas y medianas empresas. En consecuencia, el enfoque predictivo basado en ratios contables se consolida como un recurso fundamental tanto para la investigación académica como para la toma de decisiones de gestión interna y externa. (Haro Sarango, 2021; Álvarez-Ferrer y Campa-Planas, 2020)

Pero si bien el desempeño empresarial ha sido vastamente estudiado, buscando explicaciones en el endeudamiento, la baja rotación de stocks, los costos asociados al crecimiento, al control o los problemas de agencia, la literatura no ha profundizado en el rol de los costos. Por eso este trabajo se plantea indagar si una de los motivos que impactan en la rentabilidad son los vinculados a ciertas estructuras no favorables. Yardin (2012) y Cartier (2017) abordan largamente el efecto de los costos fijos, los costos variables y los márgenes en la utilidad empresarial. Por estructura de costos se entiende en este trabajo el peso porcentual del total de costos variables, costos fijos y utilidad en los ingresos (ventas o facturación) de las firmas. Cada firma cuenta con su propia estructura de costos. En algunas los costos variables son reducidos, los costos fijos son elevados y la utilidad es reducida. En la industria exportadora es frecuente encontrar costos variables altos, márgenes reducidos y elevados costos fijos derivados de la gran cantidad de capital inmovilizado (maquinarias y materias primas).

La contribución marginal refleja principalmente la eficiencia comercial en la compra y venta de cada firma. Suele estar determinada por el mercado y las estrategias comerciales adoptadas, mientras que la participación de los costos fijos está generalmente vinculada al tamaño y rigidez de la estructura fija, es decir su fuerza laboral, equipos y edificios. En este contexto, el análisis de la estructura de costos se utiliza más como una herramienta de gestión que como un mecanismo de diagnóstico financiero. Las dificultades financieras de muchas empresas suelen interpretarse como consecuencias inevitables de factores externos, y no como efecto de una gestión inadecuada de los costos. Este enfoque es respaldado por Jánica et al (2023), quienes señalan que los modelos tradicionales como el de Altman, incluso con sus mejoras en los años 90, no lograban anticipar los factores precedentes al deterioro de los indicadores contables. Modelos más recientes, como el de Olshon y otros que incorporan inteligencia artificial, han comenzado a integrar tanto elementos internos como externos. No obstante, aún persiste una pregunta pendiente en la investigación: ¿Que peso tienen los costos fijos en las estructuras de las empresas con beneficio positivo? ¿Qué peso tiene el margen? Son preguntas que este trabajo se plantea.

Entonces, aunque el análisis de los costos se considera esencial para la gestión operativa interna, su vínculo con el desempeño financiero aún no se comprende completamente, por falta de datos empíricos. A pesar de los avances en modelos analíticos, sigue faltando identificar un patrón que relacione directamente las características de las estructuras de costos con los resultados económicos de las firmas uruguayas.

En materia de antecedentes metodológicos, cuando se trata de desempeño, el indicador más usado es la rentabilidad medida por el ROA, y en ocasiones, el margen o el EBITDA. Este trabajo, al atacar las estructuras de costos, se basa en la estructura interna del Estado de Resultados, fuente escasamente utilizada en trabajos similares. También será considerado el ROA, aunque no como variable a explicar, sino como rasgo o característica. Esta estrategia en base contable tiene limitaciones vinculadas a la interpretación de los datos. En efecto, por tratarse de datos contables, están condicionados por la forma en que la normativa legal ordena y cuantifica las variables contables, adicionalmente al riesgo de manipulación al que siempre se exponen los estados financieros. (Vázquez et al., 2015; Fernández y Gutiérrez, 2012).

Al mismo tiempo, se trata de una estrategia cuantitativa empírica, la cual ha sido utilizada en diversos estudios previos, aunque no siempre con resultados alineados. Este tipo de abordaje busca identificar patrones comunes en los estados financieros de las firmas. Un estudio importante es el realizado por Romero Espinosa et al. (2015)

quienes examinaron los estados contables de más de 17.000 empresas —8.891 pequeñas y 8.469 medianas— durante el período 2009-2013, utilizando información proveniente de la Superintendencia de Sociedades de Colombia. Según los autores, la mayoría de las empresas que enfrentaron la quiebra eran de menor tamaño y predominantemente industriales, lo cual sugiere una correlación negativa entre el tamaño empresarial y rentabilidad, donde el sector parece ser central, aspecto relevante para el enfoque de la presente investigación. En cuanto a la estructura de costos, se encontró que las empresas con buen estado financiero contaban con una mayor contribución marginal y realizaban más inversiones en áreas administrativas y comerciales, indicador de que daban mayor importancia a la información estratégica y a la planificación como parte de su gestión.

Otros trabajos similares que sientan bases para comprender al presente son:

- Tong y Saladrighes (2018). España. 12.865 balances. Período analizado 2009-2014. Regresiones sobre datos de panel. Impacto positivo del tamaño en la rentabilidad.
- Gschwandtner y Hirsch (2018). EE. UU. 409 industrias. Impacto positivo del tamaño en la rentabilidad.
- Melgarejo et al (2014). Colombia. 4.639 balances. Resultados: bajos niveles de ganancias, escaso márgenes, y peso muy fuerte de las remuneraciones en los ingresos.

En costos y análisis marginal, se cuenta con el trabajo de Laitinen (2011), donde se analizan los datos de 1.670 empresas, incluyendo 328 que quebraron. Se analizan la estructura de costos y se determina el punto de equilibrio. Luego determinan el Margen de Seguridad, para concluir que es un predictor confiable de la quiebra. Como el margen de seguridad depende directamente del punto de equilibrio y por tanto de los costos fijos y los costos variables, de ahí la relevancia de la gestión de costos en la rentabilidad y razón por la cual dirigimos este estudio hacia esa área.

Otro trabajo sobre costos es el de Merino Sánchez et al (2021), donde se estudia el papel de la existencia de una contabilidad de costos en el desempeño, como herramienta de control de gestión, pero no profundizan en la estructura de costos.

En síntesis, el tamaño empresarial, el sector económico y el nivel de endeudamiento influyen directamente en el desempeño económico de las firmas, con evidencias diversas en cuanto al signo del impacto. Al mismo tiempo, no hay muchos trabajos que aborden la problemática de los costos, la cual se relaciona a la gestión gerencial y su control. Eso motiva y da interés a este artículo.

El caso uruguayo

El primer autor en estudiar las conductas empresariales en Uruguay fue Ricardo Pascale. En sus trabajos (Pascale et al, 2013, Pascale, 2018), utiliza una base de datos muy similar a la de esta investigación, ya que proviene de DGI y tiene los mismos datos. Los trabajos de Pascale fueron importantes para definir las variables a medir: ROA y Resultados / Ventas, como indicadores de rentabilidad, Activos Totales como indicador de tamaño, Bienes de Cambio / Costo de Ventas, Activo Corrientes / Activos total y Pasivo / Activo, como indicador de endeudamiento. Como resultado general, el autor encuentra que el tamaño, ya sea medido por el activo o las ventas, impacta negativamente en la rentabilidad de las empresas industriales. Al igual que la mayoría de la literatura internacional revisada, Pascale no analiza el rol de los costos.

Entonces, dados estos conocimientos previos, tanto en lo nacional como internacional, es pertinente preguntarnos si la rentabilidad de las empresas industriales está determinada por los distintos esquemas de peso relativo de los costos en los ingresos en paralelo al tamaño y a las ventas. Si así fuera es posible plantearnos un modelo que permita predecir el beneficio económico en función de estos ingresos, del tamaño o de la propia estructura de los costos empresariales, separados en variables y fijos.

3- Metodología

Base de Datos y variables

Las firmas de la base son empresas industriales manufactureras de acuerdo a la codificación CIIU. La base de datos se conforma de los estados contables de estas firmas pertenecientes al grupo CEDE y Grandes Contribuyentes del Uruguay presentados ante la Dirección General Impositiva en el período 2010-2017 (8 años). Las variables utilizadas se muestran en la Tabla 1.

Tabla 1. Descripción de variables utilizadas

Variable	Definición
Activos	Valor contable de los activos, expresados en Millones de Dólares Americanos
Pasivos	Valor contable de los pasivos, expresados en Millones de Dólares Americanos
Ventas (V)	Valor contable de las ventas, expresadas en Millones de Dólares Americanos
Costo de Ventas (CV)	Valor contable del costo de los bienes vendidos, expresados en Millones de Dólares Americanos, mas costos de comercialización.
Costos Fijos (CF)	Valor contable de los costos fijos totales, expresados en Millones de Dólares Americanos
Utilidad	Utilidad = Ventas - Costos de Ventas - Costos Fijos, expresados en Millones de Dólares Americanos.
RCV	Razón de Costos Variables: Costo de Ventas/Ventas
RC	Razón de Contribución: (Ventas - Costo de Ventas)/Ventas
CF/Ventas	Costos Fijos / Ventas
Res/Ventas	Resultado / Ventas
ROA	Retorno sobre Activos (Resultado/Activos)
MS	Margen de Seguridad: (Ventas - Punto de Equilibrio) / Ventas
Equilibrio/Activos	Punto de Equilibrio / Activos
Punto de equilibrio	Nivel de Ventas para el cual el Resultado es igual a cero.
Edad	Cantidad de años en que permanece/ permaneció abierta la empresa.
Endeudamiento	Pasivos/ Activos
Personal ocupado	Cantidad de personas empleadas en cada año.

Fuente: elaboración propia

El trabajo utiliza el nivel de activos, ventas y la cantidad de personal ocupado como variables proxy de tamaño. El desempeño económico se mide a través de los ratios *ROA (en términos porcentuales)* y, *Resultado del ejercicio (en términos absolutos – dólares)*. Las variables financieras adyacentes son la edad de la empresa, el margen de seguridad, el punto de equilibrio y la capacidad o potencialidad de la empresa para alcanzar el punto de equilibrio, el endeudamiento y el peso de los Costos fijos en las Ventas.

El Costo de Ventas representa el costo de materias primas, remuneraciones de fábrica y otros costos de fabricación tales como energía combustible y mantenimiento, más costos de comercialización tales como comisiones por venta. En este trabajo se realiza un supuesto importante, pero fundamentado y que se entiende razonable: se asume que el costo de lo vendido es variable. Como se podrá apreciar párrafos adelante, este rubro demuestra tener una variabilidad muy elevada con respecto a las ventas, aunque derivado de la normativa contable vigente en Uruguay puede potencialmente incluir algunos costos fijos de fabricación. Los Costos Fijos se asume que son otros costos no incluidos en el Costo de Ventas y de naturaleza independiente del nivel de actividad, de ahí su carácter de fijos.

La Razón de Costos Variables (RCV) representa el peso porcentual del costo de ventas en las ventas. La razón de Contribución (RC) refleja el porcentaje del margen (ventas – costo de ventas) en las ventas.

El punto de equilibrio se obtiene dividiendo los Costos Fijos (CF) / Razón de Contribución, y representa el monto de ventas en que el resultado es igual a cero.

El Margen de Seguridad resulta del cociente (Ventas – Punto de Equilibrio)/ Ventas y mide la distancia de las ventas actuales respecto su punto de equilibrio.

El ratio: Punto de Equilibrio / Activos representa las veces que se debe rotar el activo para llegar al equilibrio. Representa en tanto la potencialidad o facilidad de la firma para lograrlo, ya sea por buenos niveles de RC o bajos niveles de CF.

La base de datos está limpia de casos atípicos y se compone de empresas cuyos activos no superan 200 millones de dólares:

2010: 1568 balances
2011: 1583 balances
2012: 1563 balances
2013: 1596 balances
2014: 1523 balances
2015: 1594 balances
2016: 1613 balances
2017: 1564 balances

Totaliza 12.604 estados contables. Solo 1086 firmas tienen datos completos en los 8 años. El resto son empresas que han sido comenzado actividades luego de 2011, o han cerrado entre 2010 y 2017. Se trabaja de esta forma, porque trabajar con solo 1086 firmas con datos completos, implicaría introducir un sesgo de supervivencia, perdiendo, entre otras cosas, el rastro de las que han sido cerradas, probablemente por no lograr un adecuado desempeño. En una investigación que se centra justamente en la performance, no sería un camino metodológico aconsejable. Luego de la depuración, quedan 12.582 observaciones (balances) en 8 años.

Para procesar la base de datos se utilizó el software *Dataiku Free Edition*, una de las herramientas de Machine Learning de acceso libre, en dos fases:

- En primer lugar se aplicó un algoritmo de clústerización (K-means) que permitió agrupar a las empresas en 3 conglomerados (clústers) muy parecidos entre sí con rasgos comunes, y muy distintos a las de los otros grupos.
- En segundo lugar, para cada clúster se aplicó una herramienta predictiva que posibilitó plantear un modelo matemático que relaciona las ventas con los costos y la utilidad.

Machine Learning es una rama de la Inteligencia Artificial, que utiliza el aprendizaje automático sobre los datos y mediante algoritmos detecta patrones recurrentes. Encontrados estos patrones, los algoritmos vuelven a aplicarse en varias fases de entrenamiento y testeo, lo cual posibilita mejorar los modelos tanto de clasificación o predicción. Este aprendizaje se puede realizar de tres formas: 1) algoritmos con aprendizaje supervisado, donde se le indican a la máquina que patrones y en qué datos debe buscar, 2) el aprendizaje reforzado, donde el algoritmo elabora un modelo en la fase de entrenamiento y lo prueba en la fase de test, para luego mejorar el modelo y 3) algoritmos no supervisados, como el de clúster, utilizados para identificar grupos de entidades similares. (Ramírez Mendoza, D. Y. 2022)

Los algoritmos no supervisados utilizan el aprendizaje automático para encontrar patrones ocultos a partir de lo que se llama datos no etiquetados. En una regresión lineal, la cual también utilizaremos en una segunda fase, tenemos una serie de datos que forman parte de una variable explicativa X , (y otras) y una variable Y a explicar en función de X . Pero en el aprendizaje no supervisado no hay una variable Y que se deba buscar, sino solo encontrar patrones, tendencias, grupos con rasgos comunes dentro de un conjunto grande de datos en función de caracteres similares en variables que se le van indicando al modelo (Ramírez Mendoza, 2022).

Dentro de los algoritmos de clúster, uno de los más conocidos es el K-means. A partir de un conjunto de datos de empresas el algoritmo K-means fija al azar a 2 o más que se le indique de ellas, que se denominan “centroides” o semillas. Cada empresa de la base tiene un valor de activos, de ventas, de utilidad, que son asimilables a sus coordenadas en un plano. Si consideramos dos variables (*activo* y *ventas*), a cada empresa le corresponde un punto en el plano de coordenadas cartesianas *Activo-Ventas*. Otra empresa estará en otro punto y así sucesivamente. Si introducimos una tercera variable, estamos en un espacio tridimensional, donde cada empresa es un punto en el espacio. Acto seguido el algoritmo mide las distancias entre los centroides (que son puntos o empresas) y el resto de las empresas (puntos). Para medir esa distancia utiliza la distancia euclídea en ese espacio de coordenadas. Luego, para cada empresa de la muestra, determina la distancia que tiene con cada centroide, y la asigna al que esté más cerca, formando un aglomerado de empresas alrededor de cada centroide. Forma entonces dos aglomerados (1 y 2) y calcula los valores promedio de cada variable (activo, ventas, utilidad). Con esos promedios, fija dos nuevos centroides (uno para el clúster 1 y otro para el 2) y repite los pasos anteriores, formando de nuevo el clúster 1 y el 2. Cuando las empresas no cambian de clúster, se considera que el algoritmo converge y se detiene el proceso. (Ramírez Mendoza, 2022)

Resultado de ese proceso de aglomeración, el algoritmo define grupos de empresas que tienen patrones comunes, son parecidas entre sí y distintas al de las que están en otros grupos. Estos grupos, o conglomerados, se denominan clúster.

4- Resultados y discusión

Para el caso de este trabajo el algoritmo identificó 3 clúster y un grupo de empresas atípicas, cuyas características se exponen en la tabla 2.

Tabla 2. Características de cada clúster

	Rentabilidad		Estructura de costos							Rasgos operativos				
Cluster	Activos (Millones USD)	ROA	Ventas (Millones USD)	Costo de Ventas (Millones USD)	Costos Fijos (Millones USD)	Utilidad (Millones USD)	RCV (%)	RC (%)	CF/V (%)	Res/Ventas (%)	MS (%)	VeAct (%)	Edad	Empleados
0	2,38	0,10	3,23	2,49	0,57	0,16	0,75	0,25	0,20	0,05	0,17	0,54	31,09	32,02
1	1,63	0,11	1,86	0,82	0,88	0,16	0,42	0,58	0,50	0,07	0,12	2,34	31,54	27,08
2	34,53	0,07	39,67	31,36	6,57	1,74	0,77	0,23	0,19	0,05	0,04	0,36	43,27	237,81
Promedio	6,28	0,10	8,25	6,44	1,46	0,35	0,66	0,35	0,30		0,15	1,07	32,43	53,63

Fuente: elaboración propia.

Caracterización del clúster 0.

Contiene al 61% de las empresas, donde predominan las del rubro alimentos y bebidas, seguidas por las del sector químico, farmacéutico y textiles. La mayoría tienen más de 20 años de trayectoria y el 81% han presentado ganancias en sus estados contables. Se trata de firmas **medianas** (2.38 Millones de USD en activos y 3.23 Millones de USD en ventas promedio), con alta rentabilidad (ROA 10%), bajo margen sobre ventas (25%) y Costos Fijos reducidos en términos comparativos (20% sobre ventas). Su cantidad de personal ronda los 32 empleados. En promedio, los costos variables (75% s/Ventas) pesan más que los Fijos (25%).

Caracterización del clúster 1.

Contiene al 30% de las empresas, donde los rubros están muy distribuidos entre las del sector alimentos y bebidas, papelería, químicos y otros. Se trata de firmas de igual edad que las del clúster 0. Son firmas **pequeñas** (1.63 Millones de USD en activos y 1.86 Millones de USD en ventas promedio), con alta rentabilidad (ROA 11%) levemente superior al 0, alto margen sobre ventas y Costos Fijos bastante importantes en su peso relativo a los ingresos. Necesitan vender 2.34 veces su activo para llegar al equilibrio, ratio superior al clúster 0 que necesitaba vender 0.54 de su activo. De hecho, su margen de seguridad es menor y también muy reducido, ya que alcanza con que sus ventas bajen un 12% para que entren en la zona de pérdidas. En promedio, los costos variables y fijos están equilibrados.

Caracterización del clúster 2.

Representando apenas el 6% del total, se trata de empresas **muy grandes**, tanto en activos (34 millones de dólares promedio) como en personal ocupado (237 empleados). También se trata de firmas más antiguas (43 años). Si bien predominan alimentos y bebidas, hay presencia de frigoríficos y algunas farmacéuticas. Financieramente, son menos rentables (ROA 7%). Tienen un alto nivel de ventas pero sus márgenes son muy escasos (23%). Sus costos fijos podrían considerarse normales. Sus resultados medidos en dólares son elevados, pero en relación a las ventas o al capital invertido, son reducidos. Se trata de empresas muy riesgosas, ya que su facturación actual apenas supera su punto de equilibrio en un 4%. Sus costos variables pesan en promedio un 77% y los fijos un 19%.

Clúster de empresas atípicas.

El algoritmo K –means detectó un cierto grupo reducido de empresas extremadamente grandes (76 millones de dólares en activos y 112 en ventas, con 390 empleados promedio) con rentabilidad negativa. A lo largo del periodo 2010-2017 fueron aproximadamente 40 empresas, un 3% del total. Podría decirse que llevan al extremo los rasgos de las empresas del clúster 2, poco margen sobre ventas y elevado peso de los costos fijos en su estructura. En materia de edad, son en promedio más antiguas que el resto (43 años). Este tipo de empresas en el contexto de una base de datos, son las que se denominan *outliers* o casos atípicos, y son generalmente separadas.

En resumen, se trata de 3 clúster de empresas muy diferentes en lo que refiere a sus costos y a su tamaño. Hay un grupo de empresas (Clúster 3) muy grandes, de larga trayectoria, cuya relación de contribución marginal sobre ventas promedia el 23% y los costos fijos pesan 19% en promedio. Sus ventas actuales están apenas por encima de su punto de equilibrio. Al mismo tiempo hay otro grupo (Clúster 1) de empresas mucho más pequeñas, con alta relación de margen sobre ventas (58%), alto peso de los costos fijos en relación a sus ventas (50%) y ventas muy superiores a su punto de equilibrio. Al mismo tiempo, son las más rentables.

Por último, la mayoría de las empresas (61 de cada 100) conforman el clúster 0, de tamaño mediano, pero con estructura de costos muy similar a las más grandes (Clúster 2).

Análisis predictivo

La división de las firmas por clúster es fundamental para establecer modelos predictivos de la rentabilidad a partir de sus estructuras de costo. Esta técnica de segmentación y posterior elaboración de modelos predictivos por segmento no es novedosa, sino que ya fue utilizada por Rios y Priamo (2023) y Naoui et al (2020). Lo que se busca es mejorar la comprensión del universo de firmas que actúan en la industria, partiendo de la base de que sus rasgos económicos básicos condicionan sus comportamientos futuros.

Como se planteó en el capítulo metodológico, se elaboraron modelos matemáticos predictivos para cada clúster, obtenidos por el método de regresión lineal, aplicando la herramienta Dataiku. Esta metodología, como casi todas las de Machine Learning, tiene el énfasis puesto en la capacidad predictiva. El primer paso es indicar que variable se desea explicar –variable dependiente- a partir de ciertas variables explicativas –independientes-. En el caso de este trabajo, se plantea explicar la

variable Utilidad, a partir de las Ventas. También se plantea explicar el monto de los Costos Variables y Costos Fijos de la misma forma. En definitiva se busca plantear un modelo que permita predecir la utilidad, los costos variables y los costos fijos, a partir de las ventas, pero teniendo en cuenta el tamaño, medido éste ya sea por los activos o el número de personal empleado.

En concreto, para cada clúster por separado se procedió a correr la regresión:

- De la Utilidad en función de las Ventas
- De los Costos Variables en función de las Ventas
- De los Costos Fijos en función de las Ventas

Como resultado se obtienen finalmente los coeficientes de impacto de la estructura de costos (V-CV-CF) en la Utilidad. **Es claro que la ecuación V-CV-CF es igual a la Utilidad y es claro que la utilidad, los costos variables y fijos dependen y se correlacionan con las Ventas. Lo que este trabajo busca como aporte es encontrar los coeficientes que la relacionan.**

En efecto, la **utilidad** puede modelarse como una función de las **ventas**, del tipo:

$$Utilidad = f(x)$$

Donde:

- Y= Utilidad: utilidad (definida en tabla 1)
- x: nivel de ventas (definida en tabla 1)

En términos contables se puede modelar una relación lineal del tipo:

$$Utilidad = ax - b$$

Donde *a* es el coeficiente que relaciona a la utilidad con las ventas, más un término fijo (o independiente) *b*, que es el valor de la utilidad si las ventas son iguales a cero, es decir una ecuación de la recta.

Los resultados se pueden apreciar en la tabla 1.

Tabla 3. Resultados de las regresiones

<u>Cluster 0</u>	Pearson	P Valor	R2	Coefficiente (a)	Comp. Fijo (b)	Función (Ecuación) $Y = aX + b$
Utilidad = <i>f</i> (Ventas)	0,42	<1e-4	0,175	0,0433	0,0172	Utilidad = 0,0433 x Ventas + 0,0172
Costos Variables = <i>f</i> (Ventas)	0,988	<1e-4	0,976	0,8028	-0,097	Costos Var= 0,8028 x Ventas + -0,097
Costos Fijos = <i>f</i> (Ventas)	0,733	<1e-4	0,614	0,1529	0,081	Costos Fijos = 0,1529 x Ventas + 0,081
<u>Cluster 1</u>						
Utilidad = <i>f</i> (Ventas)	0,609	<1e-4	0,368	0,086	0	Utilidad = 0,086 x Ventas + 0
Costos Variables = <i>f</i> (Ventas)	0,961	<1e-4	0,917	0,451	-0,0171	Costos Var= 0,451 x Ventas + -0,0171
Costos Fijos = <i>f</i> (Ventas)	0,95	<1e-4	0,903	0,4627	0,0171	Costos Fijos = 0,4627 x Ventas + 0,0171
<u>Cluster 2</u>						
Utilidad = <i>f</i> (Ventas)	0,09079	0,0027	0,0119	0,0178	1,035	Utilidad = 0,0178 x Ventas + 1,035
Costos Variables = <i>f</i> (Ventas)	0,9585	<1e-4	0,918	0,9081	-4,6	Costos Var= 0,9081 x Ventas + -4,6
Costos Fijos = <i>f</i> (Ventas)	0,3293	<1e-4	0,095	0,0735	3,6	Costos Fijos = 0,0735 x Ventas + 3,6

Fuente: elaboración propia.

En todos los casos el error estándar en cada variable resultó muy pequeño, así como también el componente fijo (intercepto).

Clúster 0

En este grupo, la Utilidad está correlacionada con las Ventas de manera moderada, de acuerdo al Coeficiente de Correlación de Pearson, que se ubica en 0.42. El modelo matemático planteado que predice la Utilidad en forma directa a partir de las ventas es:

$$Utilidad (Y) = 0.0433 Ventas (x) + 0.0172$$

Esto implica que la Utilidad está vinculada a las ventas a través de un coeficiente de 0.0433, más un término independiente de 0.0172. Este modelo presenta un p valor menor a 0.05 y un R2 de 0.175. Esto quiere decir que la relación existe, no es casual, el ajuste del modelo a la realidad de las empresas es moderado, pero existen ciertamente otros factores que están impactando en la Utilidad y el modelo no los recoge, lo cual afecta su valor predictivo.

Si las ventas ascienden a \$100, la utilidad promedio de las firmas de este clúster será de \$ 4.33 más \$ 0.017, pero la probabilidad de acierto no es alta. Se puede dar constancia del signo o tendencia. Ciertamente las ventas influyen positivamente en la utilidad, pero no basta con conocer su monto para estimar la utilidad de forma exacta.

Ahora, con respecto a los **costos variables**, el modelo planteado es estadísticamente significativo y muy preciso para estimarlos a partir de conocidas las ventas. El coeficiente que los vincula es 0.8028. Se puede afirmar con mucha seguridad que en las firmas de este clúster, los costos variables representan el 80.28% de las ventas, menos \$ 0.097, siendo muy reducido el monto independiente (intercepto).

$$Costos Variables = 0.8028 Ventas -0.097$$

Algo parecido ocurre con los **costos fijos**, los cuales representan el 15.29% de los ingresos, mas \$ 0.081 (intercepto):

$$Costos Fijos = 0.1529 Ventas +0.081$$

Con esos tres modelos mirados en conjunto es posible plantear esta ecuación representativa de la estructura de costos:

$$Utilidad = Ventas - 0.8028 Ventas (Costos Variables) - 0.1529 Ventas (Costos Fijos)$$

la cual puede predecir la Utilidad, los costos variables y los costos fijos de las firmas del clúster, es decir empresas medianas que se acerquen a un tamaño de 2.38 millones de dólares en activos y 31 empleados.

La estructura de costos:

$$Costos Variables: 0.8028 Ventas + Costos fijos: 0.1529 Ventas$$

Explica la conformación de la utilidad y se puede cuantificar con esos coeficientes.

En términos monetarios, si una firma de este clúster factura \$ 100, sus costos variables ascenderán a \$80, sus costos fijos a \$15 y su utilidad será aproximadamente \$ 4-5. Validado el modelo contra los datos reales, mostró un **error promedio de 6%** sobre ventas, que se estima reducido.

Si bien el modelo que predice la utilidad directamente como 0.0433 de las ventas no tiene un R2 alto y por tanto no tiene poder predictivo, las relaciones de los costos fijos y variables (0.8028 y 0.1529) si cuentan con R2 altos , por tanto es posible predecir la utilidad a partir de los costos como:

$$Utilidad = Ventas (1 - 0.8028 - 0.1529) = Ventas \times 0.04$$

En la validación contra datos reales, el modelo presenta un error promedio en la predicción de la utilidad equivalente al 5 % de las ventas, con baja variabilidad, lo cual indica una capacidad predictiva relativamente robusta en relación con el volumen de ventas

Clúster 1

En el caso de este clúster, conformado por las firmas más pequeñas de la muestra, los modelos matemáticos son más robustos. La ecuación que predice la utilidad directamente a partir de las ventas es:

$$Utilidad (Y) = 0.086 Ventas (x) + 0.0$$

Cuyo R2 es 0.368 y p valor es menor a 0.05, lo cual indica que la relación positiva existe, no es casual, y el ajuste de la recta de regresión es mejor que en el caso anterior. Esto permite concluir que el modelo tiene un mejor ajuste a la realidad de las firmas y su poder predictivo es bueno.

En cuanto a los modelos que predicen los montos de Costos Variables y Fijos a partir de las ventas, los indicadores son aún mejores (R2 superiores a 0.90). De igual manera que en el cluster 0, se puede plantear la estructura de costos en esta ecuación:

$$Utilidad = Ventas - 0.451 Ventas (Costos Variables) - 0.4627 Ventas (Costos Fijos)$$

El error promedio del modelo en la predicción de la utilidad representa en promedio un 6 % de las ventas, con una dispersión muy reducida, lo que indica un nivel de precisión adecuado en términos relativos.

Como se puede observar, la estructura de costos es totalmente distinta en las firmas de este clúster. El peso de los costos variables y fijos está distribuido casi en partes iguales, al contrario del clúster 0 donde los costos variables representaban el 80% de las ventas y los costos fijos un 15%. La utilidad es mayor, situándose en un 8.60 % de los ingresos. Tanto los costos variables como los fijos son mucho más sencillos de predecir en este caso. Un dato a destacar es que los costos fijos son sensiblemente más altos que en los otros clúster, pero su comportamiento es altamente predecible.

Clúster 2

En tercer lugar se ubican las empresas de grandes dimensiones. En estos casos, las herramientas de Machine Learning utilizadas no permiten construir modelos predictivos eficientes. Si bien hay un impacto positivo desde las ventas hacia la utilidad, no se puede concluir en qué términos. La excepción está únicamente en el caso de los costos variables, que en este clúster tienen un peso muy importante (90% de los ingresos son costos variables) y por tanto la Razón de Contribución es reducida. Son empresas que venden a precios apenas superiores a sus costos variables.

Resumen comparativo

Si se combinan los resultados obtenidos por el algoritmo de clústerización con los resultados de los modelos matemáticos predictivos, se obtienen los resultados que se expone en la tabla 4:

Tabla 4. Resumen

Clúster	Tamaño	ROA	Estructuras de costos			
			Ventas	Costos Variables	Costos Fijos	Utilidad
			Coefficiente	Coefficiente	Coefficiente	Coefficiente
Clúster 0	Medianas	0,10	1	0,80	0,15	0,04
Clúster 1	Pequeñas - medianas	0,11	1	0,45	0,46	0,09
Clúster 2	Grandes	0,07	1	0,91	0,07	0,02

Fuente: Elaboración propia.

La primera conclusión que se destaca es que un tamaño cercano a pequeño o mediano, más una partición igualitaria de costos fijos y variables está relacionado a mayores niveles de rentabilidad. El clúster 1 muestra un ROA promedio de 11%, y Utilidad porcentual sobre Ventas de 9%. Cada cluster tiene su estructura de costos propia y característica. ¿Se puede plantear un modelo con adecuado poder predictivo de la Utilidad, conocidas las ventas y el tamaño? La respuesta es afirmativa en tanto la firma a estudiar tenga los rasgos del cluster 0 y 1.

Ahora, cuando se trata de empresas de grandes dimensiones, no es posible concluir lo mismo. Salvo que los costos variables son muy gravitantes y la rentabilidad es más reducida, no es mucho más lo que se puede predicar acerca de este tipo de empresas.

En las empresas medianas del clúster 0, promedialmente más grandes que el 1, es posible construir modelos predictivos pero su potencial no es tan robusto. Se puede afirmar que los costos variables ascienden con bastante certeza al 80% de las ventas y los costos fijos a un 15%, pero la influencia directa de las ventas en la rentabilidad no puede mostrarse a través de un modelo predictivo confiable, básicamente porque la ecuación de la recta que representa a las variables no se ajusta demasiado bien a la realidad (R^2 bajo). De todas formas, se puede predecir de manera adecuada la utilidad a partir de la estructura de costos, es decir a partir de predecir los costos variables y fijos a partir de las ventas, que si cuentan con R^2 altos.

Y finalmente en las empresas de grandes dimensiones es muy difícil concluir. Este resultado, en relación a que estas empresas tienen conductas particulares, se alinea a los resultados de Goddard et al (2005). Si esto tiene relación con el fenómeno de concentración empresarial, queda para futuras investigaciones. También debe aclararse que las conclusiones y modelos predictivos de este trabajo tienen limitaciones propias de modelos lineales univariados, que excluyen los efectos de variables externas a las firmas y se concentran en lo predictivo. También excluyen el efecto de otras variables que siendo internas juegan sin duda en el desempeño, como la tecnología y los recursos.

5- Conclusiones

Este trabajo se propuso analizar las estructuras de costos de las empresas industriales y demostrar que los costos variables, los costos fijos y la utilidad son funciones directas del nivel de ventas con vínculos claros dados por coeficientes identificables y confiables desde el punto de vista estadístico, pero dependientes del tamaño de cada firma y de su estructura de costos. Para lograrlo, la estrategia metodológica es doble: clusterización y análisis predictivo. Sobre una base de datos de Estados Contables de firmas industriales, se procedió primero a formar grupos de empresas similares mediante clústerización por algoritmo K means. Este proceso permitió identificar tres grupos de firmas con características muy diferentes entre sí.

Dentro de cada grupo, fue posible elaborar modelos matemáticos con poder predictivo en dos de los tres grupos de firmas. En esos dos clústeres, conocidas las ventas, el tamaño de los activos y la cantidad de personal, fue posible predecir el monto de los costos de producción, los costos fijos y la utilidad, con reducido margen de error. Estas empresas son de dimensiones pequeñas y medianas (30 empleados en promedio). En el tercer grupo, formado por empresas notoriamente más grandes, fue imposible plantear un modelo predictivo, salvo para relacionar los costos variables a las ventas, no pudiendo identificar patrones de comportamiento de la utilidad y los costos fijos.

En líneas generales, en las empresas medianas (cl 0), la estructura de costos tiene un peso importante de costos variables (cercano a 80% respecto a las ventas), un reducido peso de los costos fijos (15%), y está asociada a rentabilidades positivas (10% ROA). Estas ganancias pueden predecirse como el 4-5% de los ingresos, con adecuada confianza, en función de las ventas. Tanto los niveles de costos fijos como variables pueden predecirse con adecuada fiabilidad también, en base a las ventas.

En las empresas más pequeñas (cl 1), la estructura de costos presenta una partición equilibrada de costos variables y fijos (predecibles con alta confianza los dos), siendo la utilidad predecible, con altos niveles de confianza, y estimable como el 9% de los ingresos, así como también los niveles de costos fijos y variables. Estas son las empresas más rentables, ya sea esta rentabilidad medida por el ROA (11%) así como porcentaje sobre ventas (9%).

Finalmente, en el clúster de las firmas más grandes, no fue posible identificar patrones que relacionen a los costos con la utilidad y por tanto no se pueden plantear modelos predictivos. Los efectos o consecuencias financieras de sus conductas son difíciles de prever, aumentando el nivel de incertidumbre respecto a las grandes inversiones industriales, tan necesarias para las economías nacionales. Sin embargo se pudo

observar que estas firmas tienen rasgos particulares que las distinguen y tienen que ver con el tamaño y la rentabilidad. En efecto, son en promedio 15 veces más grandes que el resto y su rentabilidad se ubica en torno a un 7% de ROA, frente a un 10-11% de las restantes. Eso podría alinearse a la teoría de que el tamaño impacta negativamente en la renta.

La hipótesis H1 se verifica en todos los clúster: los costos variables son una función de las ventas con coeficientes ciertos y confiables en cada uno de ellos: 80% de las ventas en las empresas medianas del clúster 0, 45% en el clúster 1 y 90% en el 2, grupo de firmas muy grandes. La hipótesis H2 solo se verifica en los clúster 0 y 1: Los Costos Fijos son una función de las ventas con coeficientes un poco menos ajustados, lo cual es esperable dada su naturaleza de fijos, pero distintos en cada clúster (15% y 40%). La hipótesis H3 solo se verifica en los clúster 0 y 1: el monto de la utilidad es una función de las ventas, con coeficientes de mediana confianza, posible de predecir, con un coeficiente de 4% en el clúster 0 y 9% en el 1.

En suma, las herramientas de Machine Learning han demostrado ser eficaces para caracterizar a las empresas, establecer patrones y plantear modelos predictivos de los costos y beneficios económicos, que nos permite conocer la realidad empresarial. El tamaño, los ingresos y las estructuras de costos de las firmas son factores que contribuyen a explicar, en la mayoría de los casos, la conformación de la utilidad y su impacto puede ser cuantificable con estas herramientas.

6- Referencias

Altman, E. (1968). Financial ratios, discriminant analysis and the prediction of corporate bankruptcy. *The Journal of Finance*, 23(4), 589-609. <https://doi.org/10.2307/2978933>

Álvarez-Ferrer, A. y Campa-Planas, F. (2020). La predicción del fracaso empresarial en el sector hotelero. *Cuadernos de Turismo*, (45), 33-59. <https://doi.org/10.6018/turismo.426031>

Barney, J. B. (2001). Resource-based theories of competitive advantage: A ten-year retrospective on the resource-based view. *Journal of Management*, 27(6), 643-650. [https://doi.org/10.1016/S0149-2063\(01\)00115-5](https://doi.org/10.1016/S0149-2063(01)00115-5)

Becker, J. R., Kaen, F. R., Etebari, A. y Baumann, H. (2010). Employees, Firm Size and Profitability of US Manufacturing Industries. *Investment Management and Financial Innovations*, 7(2), 119-132.

Boulding, K. E. (1952). Implications for general economics of more realistic theories of the firm. *The American Economic Review*, 42(2), 35-44.

Castaño, G. A. (1999). Teoría de la agencia y sus aplicaciones. *Decisión Administrativa*, (1), 7-12.

Cartier, E. (2017). *Apuntes para una teoría general del costo*. Editorial La Ley.

De la Cuadra, Y. M. E., Bocca, D. A. M., Ortiz, O. G. V., & de la Cuadra, S. J. V. (2024). Factores Determinantes de la Rentabilidad de las Microempresas. *Ciencia Latina Revista Científica Multidisciplinar*, 8(6), 76-91.

Dos Santos, J. G. C., Calíope, T. S. y Coelho, A. C. (2015). Teorias da Firma como fundamento para formulação de teorias contábeis. *Revista de Educação e Pesquisa em Contabilidade (REPeC)*, 9(1), 101-116. <https://doi.org/10.17524/repec.v9i1.1182>

Favaro, D. (2013). Enfoques de la teoría de la firma y su vinculación con el cambio tecnológico y la innovación. *Revista Cultura Económica*, 31(85), 51-70.

Fernández, M. y Gutiérrez, F. (2012). Variables y modelos para la identificación y predicción del fracaso empresarial: revisión de la investigación empírica reciente. *Revista de Contabilidad-Spanish Accounting Review*, 15(1), 7-58. [https://doi.org/10.1016/S1138-4891\(12\)70037-7](https://doi.org/10.1016/S1138-4891(12)70037-7)

Galindo Lucas, A. (2005). *El tamaño empresarial como factor de diversidad*. Eudem.net. <http://www.eumed.net/libros/2005/agl3/index.htm>

Germán-Soto, V. y Marines López, A. (2023). Modelando crecimiento y rentabilidad empresarial de la industria de Coahuila. *Ensayos. Revista de Economía*, 42(1), 1-32. <https://doi.org/10.29105/ensayos42.1-1>

Goddard, J., Tavakoli, M. y Wilson, J. O. (2005). Determinants of profitability in European manufacturing and services: evidence from a dynamic panel model. *Applied Financial Economics*, 15(18), 1269-1282. <https://doi.org/10.1080/09603100500387139>

González Gómez, J. I. G., Hernández García, M. d. C. y Rodríguez Ferrer, T. (2000). Contribución del tamaño y el sector en la explicación de la rentabilidad empresarial. *Revista Española de Financiación y Contabilidad*, 29(106), 903-930.

Gschwandtner, A. y Hirsch, S. (2018). What drives firm profitability? A comparison of the US and EU food processing industry. *The Manchester School*, 86(3), 390-416. <https://doi.org/10.1111/manc.12201>

Guiso, L. y Rustichini, A. (2018). Understanding the size and profitability of firms: The role of a biological factor. *Research in Economics*, 72(1), 65-85. <https://doi.org/10.1016/j.rie.2017.04.006>

Haro Sarango, A. F. (2021). La estructura financiera y el fracaso empresarial: una apreciación a las grandes empresas de pesca y acuicultura. *Revista Ciencia Multidisciplinaria CUNORI*, 5(1), 1-16. <https://doi.org/10.36314/cunori.v5i1.148>

Hernández Jiménez, A., & Cáceda Quispe, M. (2024). Factores que influyen en la rentabilidad de las micro y pequeñas empresas en tiempos de contingencia. *Revista de Métodos Cuantitativos para la Economía y la Empresa*, 37, 1-23. <https://doi.org/10.46661/rev.metodoscuant.econ.empresa.7796>

Jánica, F., Hernández-Fernández, L., Escobar Castillo, A. y Velandia Pacheco, G. (2023). Factores que explican, median y moderan el fracaso empresarial: Revisión de publicaciones indexadas en Scopus (2015-2022). *Revista de Ciencias Sociales*, 29(2), 73-95. <https://doi.org/10.31876/rcs.v29i2.39963>

Jhon Rios, L. Priamo. (2023). Aplicación de técnicas de Data Analytics: Clústering y Regresión Lineal Múltiple, para la segmentación de la oferta y proyección de ciclos inmobiliarios en el mercado de oficinas prime. Tesis de Maestría en Ingeniería Industrial. *Pontificia Universidad Católica del Perú*

Kammoun, S. y Alcouffe, A. (2003). Enfoque económico de las competencias de la firma. Hacia una síntesis de las teorías neoinstitucionales y evolucionistas. *Boletín Cinterfor: Boletín Técnico Interamericano de Formación Profesional*, (154), 35-66.

Krugman, P. R. y Wells, R. (2006). *Introducción a la economía: microeconomía*. Reverté

Laitinen, E. K. (2011). Extension of break-even analysis for payment default prediction: evidence from small firms. *Investment Management and Financial Innovations*, 8(4), 96-108.

Melgarejo, Z., Vera, M. y Mora, E. (2014). Diferencias de desempeño empresarial de pequeñas y medianas empresas clasificadas según la estructura de la propiedad del capital, caso colombiano. *Suma de Negocios*, 5(12), 76-84. [https://doi.org/10.1016/S2215-910X\(14\)70029-1](https://doi.org/10.1016/S2215-910X(14)70029-1)

Merino Sánchez, L. E., Fonseca Esparza, R. V. y Rodríguez Rodríguez, P. J. (2021). La Contabilidad de Costos, los Sistemas de Control y su relación con la Rentabilidad Empresarial. *Revista Científica FIPCAEC (Fomento de la investigación y publicación científico-técnica multidisciplinaria)*, 6(3), 122-149. <https://www.fipcaec.com/index.php/fipcaec/article/view/459>

Mojica Carreño, K. J. (2023). *Fracaso empresarial. El abismo financiero: una mala gestión empresarial conduce al colapso económico y la quiebra*. Universidad Militar Nueva Granada, Facultad de Ciencias Económicas, Programa de Alta Gerencia. Colombia.

Naoui, M. A., Lejdel, B., & Ayad, M. (2020). Usando el algoritmo K-means para la curva de regresión en un gran sistema de datos para el entorno empresarial. *Revista Cubana de Ciencias Informáticas*, 14(2), 34-48.

Panchi, P. A. M., Panchi, D. F. M. y Tapia, C. E. F. (2022). Modelo de predicción de quiebra Z2 de altman de análisis multivariable en empresas del sector inmobiliario de la provincia de Pichincha: Altman's Z2 bankruptcy prediction model of multivariate analysis in real estate companies in the province of pichincha. *REVISTA CIENTÍFICA ECOCIENCIA*, 9(2), 53-76. <https://doi.org/10.21855/ecociencia.92.643>

Pascale, R. (1998). *Decisiones Financieras*. (6ª ed.). Ediciones Macchi.

Pascale, R. (2017). Vinculación entre tamaño y rentabilidad. Evidencia empírica en las empresas industriales manufactureras en Uruguay. *Revista de Investigación en Modelos Financieros*, 2, 39-54.

Pascale, R. (2018). Rasgos estilizados de las finanzas empresariales: industria manufacturera y comercio en el Uruguay, 2010-2016. *Revista de Derecho Comercial*, (10), 81-106.

Pascale, R., Blanco, A. y Gari, J. (2013). *Determinantes de la rentabilidad: Un modelo para las empresas en el Uruguay* [Conferencia]. XXX Conferencia Interamericana de Contabilidad, Montevideo, Uruguay.

Ramírez Mendoza, D. Y. (2022). *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*. UNIVERSIDAD EAFIT ESCUELA DE ECONOMÍA Y FINANZAS MAESTRÍA EN ADMINISTRACIÓN FINANCIERA – MAF BOGOTÁ 2022.

Roitter, S. y Federico, F. (2022). *El desarrollo territorial en debate: la digitalización y la sustentabilidad ambiental como desafíos y oportunidades para la transformación productiva* [Ponencia]. 27° Reunión Anual Red Pymes Mercosur. Universidad Nacional de Río Negro. Bariloche, Argentina.

Romero Espinosa, F. (2013). Variables financieras determinantes del fracaso empresarial para la pequeña y mediana empresa en Colombia: análisis bajo modelo Logit. *Pensamiento & Gestión*, (34), 235-277.

Romero Espinosa, F., Melgarejo Molina, Z. A. y Vera Colina, M. A. (2015). Fracaso empresarial de las pequeñas y medianas empresas (pymes) en Colombia. *Suma de negocios*, 6(13), 29-41. <https://doi.org/10.1016/j.sumneg.2015.08.003>

Slater, S. F. y Olson, E. M. (2002). A fresh look at industry and market analysis. *Business Horizons*, 45(1), 15-22. [https://doi.org/10.1016/S0007-6813\(02\)80005-2](https://doi.org/10.1016/S0007-6813(02)80005-2)

Tong, Y. y Saladríguez, R. (2018). The predictability of financial, accounting-based, and industrial factors on the success of newly incorporated Spanish firms. *Intangible Capital*, 14(1), 127-145. <https://doi.org/10.3926/ic.1106>

Vázquez, M., Irimia, A. y Blanco, A. (2015). La metodología de los Rough Sets como técnica de preprocesamiento de datos: Una aplicación a las quiebras de microempresas familiares. *Rect@*, 16(1), 1-12.

Winter, S. G. (2003). Understanding dynamic capabilities. *Strategic Management Journal*, 24(10), 991-995. <https://doi.org/10.1002/smj.318>

Yardin, A. (2012). *El Análisis Marginal*. Editorial Buyatti. Argentina.