



Modelo predictivo de insolvencia empresarial para las Pymes en Colombia

Predictive model of business insolvency for SMEs in Colombia

CAMILO CASTRO ESPITIA

Proyecto de grado

ccastroe1@eafit.edu.co

Director y Co-director: Juan David Martínez Vargas y Lina María Sepúlveda Cano

jdmartinev@eafit.edu.co - lmsepulve@eafit.edu.co

UNIVERSIDAD EAFIT

ESCUELA DE CIENCIAS APLICADAS E INGENIERÍA

MAESTRÍA EN CIENCIAS DE LOS DATOS Y LA ANALÍTICA

MEDELLÍN

2024

CONTENIDO

INTRODUCCIÓN.....	5
PLANTEAMIENTO DEL PROBLEMA.....	6
JUSTIFICACIÓN.....	9
OBJETIVOS	10
MARCO TEÓRICO O MARCO CONCEPTUAL	11
DISEÑO METODOLÓGICO	18
DESARROLLO DEL TRABAJO	20
RESULTADOS	28
CONCLUSIONES	34
REFERENCIAS	36

RESUMEN

La insolvencia empresarial repercute económicamente en la empresa afectada y, de forma directa e indirecta, en sus proveedores de bienes y servicios. Esto termina afectando a la población en general, quienes son los consumidores de estos bienes y servicios. Por ende, es importante analizar y estudiar los posibles factores que llevan a una empresa a entrar en situación de insolvencia económica. De acuerdo con lo anterior, el presente trabajo tiene como objetivo predecir la insolvencia empresarial en el contexto colombiano para las Pymes. Asimismo, se hará uso de los datos financieros disponibles de la Superintendencia de Sociedades en el año 2021 y las empresas que entraron en situación de insolvencia económica en el año 2023. De esta manera, a través de indicadores financieros, será posible construir las variables explicativas de la insolvencia económica y lograr un modelo predictivo de insolvencia económica.

El presente trabajo se aborda de la siguiente manera: primero, se extrae la información de la Superintendencia de Sociedades para las pymes en el año 2021 y las empresas que entraron en insolvencia económica en 2023 para obtener nuestra variable objetivo. Después, con base en los datos obtenidos, se calculan los índices financieros necesarios para el presente estudio, utilizando los estados financieros de estas empresas correspondientes al año 2021. A continuación, se realiza el análisis exploratorio de datos para obtener un conocimiento más amplio del conjunto de datos y analizar temas de correlación entre variables, entre otros aspectos. Finalmente, se diseñan varios modelos de machine learning para determinar cuál es el más apropiado para el caso de estudio.

Finalmente, los resultados del presente trabajo nos indican que los modelos Random Forest y CatBoost son los más adecuados para predecir la insolvencia económica en Colombia, teniendo en cuenta que estos modelos son de gran

utilidad en variables categóricas y para efectos del trabajo la variable objetivo es categórica.

Palabras Clave: Insolvencia Económica, Pymes, Modelo Predictivo

ABSTRACT

Business insolvency has an economic impact on the affected company and, directly and indirectly, on its suppliers of goods and services. This ends up affecting the general population, who are the consumers of these goods and services. Therefore, it is important to analyze and study the possible factors that lead a company to enter a situation of economic insolvency. In accordance with the above, the objective of this work is to predict business insolvency in the Colombian context for SMEs. Likewise, use will be made of the financial data available from the Superintendency of Companies in the year 2021 and the companies that entered into a situation of economic insolvency in the year 2023. In this way, through financial indicators, it will be possible to construct the explanatory variables. of economic insolvency and achieve a predictive model of economic insolvency.

This work is approached as follows: first, information is extracted from the Superintendency of Companies for SMEs in 2021 and companies that entered economic insolvency in 2023 to obtain our objective variable. Then, based on the data obtained, the financial ratios necessary for this study are calculated, using the financial statements of these companies corresponding to the year 2021. Next, the exploratory data analysis is carried out to obtain a broader knowledge of the set. of data and analyze correlation issues between variables, among other aspects. Finally, several machine learning models are designed to determine which one is most appropriate for the case study.

Keywords: Economic Insolvency, SMEs, Predictive Model

INTRODUCCIÓN

Las pequeñas y medianas empresas son fundamentales en el tejido económico del país y representan el mayor porcentaje de empresas, siendo el 95% de las empresas del país para el 2023 (Ministerio de Comercio, Industria y Turismo, 2024). Por lo tanto, son un pilar tanto social como económico, ya que muchas familias dependen del sostenimiento y éxito de estas empresas. Si, por el contrario, estas empresas se ven obligadas a cerrar y entran en crisis, desencadenarían efectos negativos de gran impacto en la economía colombiana.

El hecho de que una empresa entre en un proceso de insolvencia económica, además de representar la incapacidad de la empresa para asumir sus obligaciones, implica una serie de costos y consecuencias, como el desempleo, que desencadenan problemas económicos y sociales, disminuyendo la calidad de vida de los habitantes mientras aumentan el desempleo y el cierre de empresas.

En adición a lo anterior, teniendo en cuenta el impacto que genera este fenómeno en la economía y en la sociedad, diversos autores han analizado la insolvencia económica desde distintas perspectivas. Por ejemplo, Pérez et al (2019) predicen la fragilidad de las empresas colombianas de caer en quiebra mediante el uso de modelos estadísticos, García (2023) estudia la insolvencia económica en el sector de construcción a través de modelos predictivos utilizando modelos Logit y Probit. Por otro lado, desde el ámbito financiero el cuál es objeto de este trabajo para el análisis de la insolvencia encontramos a Correa y Castaño (2019) los cuáles estudian la insolvencia económica a través de indicadores financieros utilizando la regresión logística y el algoritmo Boosting.

Es importante contar con insumos que permitan mitigar o reducir la insolvencia económica y su impacto en el país. Anticipando el riesgo de que las empresas entren en insolvencia económica, se puede prevenir esta situación y, de esta manera, minimizar los impactos que genera la insolvencia de estas empresas. Por ende, este trabajo busca proporcionar un modelo que permita predecir el riesgo de que una empresa entre en insolvencia económica de acuerdo con sus indicadores financieros. Para esto, se van a comparar tres modelos: *Random Forest*, *Isolation Forest* y *CatBoost*, se evalúan estos modelos y se ajustan sus parámetros. Lo anterior se basa en las variables financieras de las pequeñas y medianas empresas para el año 2021 y en cuáles de estas empresas entraron en insolvencia económica en el año 2023.

El presente trabajo está organizado de la siguiente forma, primero se realiza una introducción al problema de estudio. Después, se contextualiza el problema y se explica por qué es importante investigarlo. En el estado del arte, se abordan los antecedentes de la ley de insolvencia económica en Colombia y los trabajos anteriores que también han investigado este tema. A continuación, se expone y se ejecuta la metodología propuesta para el desarrollo del trabajo, incluyendo el análisis exploratorio de los datos y todo lo relacionado con la preparación y el modelado de estos, para finalmente poder analizar y concluir a partir del modelo que presentó las mejores métricas.

PLANTEAMIENTO DEL PROBLEMA

Considerando la significativa presencia e impacto de las Pymes tanto a nivel nacional como internacional, como se discutió en el apartado anterior, desafíos como la insolvencia económica pueden conducir al cierre o liquidación de estas empresas, afectando toda la economía. La insolvencia económica se define como la incapacidad de las empresas para cumplir con sus obligaciones financieras, lo

que puede llevarlas a un régimen donde acuerden con sus acreedores formas de pago o enfrenten un proceso de liquidación a través de un procedimiento judicial. Además, un proceso de insolvencia económica afecta la liquidez de la empresa, lo que repercute negativamente en su desarrollo y crecimiento. La falta de liquidez impide a la empresa aprovechar oportunidades de expansión y crecimiento, además de incapacitarla para cumplir con sus obligaciones, lo que puede resultar en desempleo y reducción de operaciones. Asimismo, la insolvencia económica conlleva costos directos, como los pagos a los involucrados en el proceso, y costos indirectos, como los relacionados con oportunidades de expansión o desarrollo que la empresa no puede asumir.

Teniendo en cuenta que nos encontramos en un entorno en constante cambio, el cual está sujeto a amenazas externas, las crisis derivadas de estas amenazas tienen un impacto negativo en el desarrollo de las empresas y aumentan la probabilidad de su quiebra (León y Vargas, 2023, p. 2). Por lo tanto, es fundamental proporcionar mecanismos de reducción de riesgos para las empresas, especialmente para las pequeñas y medianas empresas, que suelen ser las más afectadas por estas crisis.

Adicionalmente, el panorama actual en Colombia no es tan alentador. A cierre del año 2023, más de 12,000 compañías entraron en proceso de liquidación y 308 anunciaron un estado de insolvencia (La República, 2023). Esto evidencia una situación preocupante, lo cual afecta el consumo y la economía en general debido al cierre de estas empresas. Por lo tanto, se destaca la importancia de realizar estudios que permitan prevenir y reducir la situación de insolvencia económica de las empresas, así como su impacto en la economía del país.

Por otro lado, en el primer semestre del año 2022, 181 compañías se acogieron al régimen de insolvencia empresarial (Portafolio, 2022). Este es un número considerable de empresas que se acogen al régimen de insolvencia económica, lo

cual repercute de manera negativa en toda la cadena de proveedores y consumidores.

Asimismo, para que una empresa perdure en el tiempo y genere valor para los accionistas, debe gozar de una buena salud financiera (García, F. 2023). Es por esto que existe la necesidad de desarrollar herramientas y modelos predictivos efectivos que permitan prever y mitigar los riesgos de insolvencia económica a través de los indicadores financieros de las empresas, los cuales se obtienen de sus estados financieros.

En adición a lo anterior, el informe de Insolvencia Colombiana e IFI (Instituciones Financieras Intermediarias) señala que, hasta septiembre del 2022, 2952 colombianos se declararon en quiebra, lo que representa un incremento en el número de solicitudes de hasta un 37,4% respecto al 2021 (Infobae, 2022). Como plantea González y Prieto (2018), muchas de estas pequeñas y medianas empresas inician sus operaciones sin haber realizado un análisis de mercado y financiero previo, lo que resulta en un alto nivel de discontinuidad para un sector el cual representa el 95% del total de compañías del país. Lo anterior refuerza la necesidad del desarrollo de modelos predictivos que puedan mitigar el impacto de la insolvencia económica y, de esta forma, reducir la situación de quiebra en la población colombiana.

Con base en todo lo anterior, el presente trabajo busca principalmente entrenar un modelo predictivo de la insolvencia económica. Por otro lado, también servir como insumo para el análisis y la reducción de la insolvencia económica en las pequeñas y medianas empresas colombianas. De esta manera, también se pretende mitigar los problemas económicos y sociales derivados de la insolvencia económica y el proceso que conlleva esta situación, lo cual repercute tanto en las empresas como en la sociedad en general. Para lograr este objetivo, se propone identificar el modelo de *machine learning* más adecuado para los indicadores

financieros calculados a partir de la información financiera de las Pymes para el año 2021.

En resumen, la insolvencia económica acarrea diversos problemas como la falta de liquidez en las empresas, el desempleo y la reducción de estas, entre otros. Además, el panorama en Colombia refleja una situación alarmante en cuanto al alto número de empresas que entran en insolvencia económica en el país. Por lo tanto, se vuelve imperativo realizar estudios centrados en este fenómeno para reducir su impacto en la economía nacional. Es así como surge la necesidad del presente trabajo, con el objetivo de estimar un modelo que permita predecir la probabilidad de que una empresa entre en insolvencia económica basándose en su situación financiera.

JUSTIFICACIÓN

León y Vargas (2023), citando a Roca Espinoza et al. (2020), señalan que es importante implementar estrategias financieras en situaciones de riesgo de este tipo, ya que corregir estos problemas rápidamente afecta la liquidez y no el capital (p. 9). Por lo tanto, contar con mecanismos que reduzcan el riesgo de las empresas de caer en insolvencia económica permite evitar la pérdida de capital, lo cual impacta de forma positiva los activos de la empresa y disminuye sus riesgos financieros.

El tejido económico empresarial colombiano está compuesto en gran medida por Pymes. Según el Ministerio de Comercio, Industria y Turismo, entre enero y marzo de 2023, las microempresas representaron el 95,3% del tejido empresarial del país, mientras que las pequeñas empresas representaron el 3,5%, y las medianas y grandes el 0,9% y 0,3% del total nacional, respectivamente (Ministerio de Comercio, Industria y Turismo, 2023). En un contexto internacional, la ONU indica que las micro, pequeñas y medianas empresas representan el 90% de las

empresas del mundo, entre el 60% y el 70% del empleo y el 50% del PIB mundial (ONU, 2023).

Por otro lado, la detección temprana en base a la probabilidad de que una empresa entre en insolvencia económica puede ser fundamental para el diseño de políticas públicas destinadas a mitigar este problema, reduciendo así su efecto negativo en el empleo y el bienestar social (Hernández, L. 2022). Teniendo en cuenta que las Pymes son la columna vertebral de la economía colombiana y mundial, es necesario realizar estudios que nos permitan prever, evitar y mitigar la situación de insolvencia económica en las empresas colombianas, ya que esto repercute tanto económicamente como socialmente en el país.

En adición a lo anterior, es importante realizar estudios que sirvan de insumo para el análisis de la insolvencia económica, examinándola a través de los indicadores financieros de la empresa, ya que, como señalan Pairetty y Cruz (2021), estos indicadores evalúan la situación y el desempeño económico y financiero de una empresa con respecto a la realidad, lo que permite llevar a cabo una gestión óptima y eficiente de sus recursos. Complementando lo anterior, los factores que determinan la quiebra empresarial son complejos, ya que se presentan en un contexto económico muy dinámico. Sin embargo, uno de los factores de mayor importancia es el riesgo financiero (Penagos y Muñoz, 2015).

OBJETIVOS

Objetivo general: Desarrollar un modelo de machine learning para predecir la insolvencia económica en las pymes con base a sus indicadores financieros utilizando algoritmos de aprendizaje de máquina

Objetivos específicos:

- Construir los indicadores financieros a partir de los estados financieros de las pymes para el año 2021 para entrenar el modelo de aprendizaje de máquina.
- Explorar el Dataset mediante un Análisis Exploratorio de Datos (EDA) para identificar las relaciones de correlación entre las variables, y llevar a cabo el proceso de transformación y limpieza de los datos con el fin de prepararlos para su posterior procesamiento.
- Evaluar y comparar modelos predictivos a través de métricas de evaluación basadas en los datos para determinar cuál de ellos se ajusta mejor al presente estudio.

MARCO TEÓRICO O MARCO CONCEPTUAL

En una economía de mercado, el fracaso de una empresa puede desencadenar un efecto contagio que afecta negativamente al resto de las empresas con las que tenía relación, ya que estas pueden perder un cliente o proveedor, lo que a su vez puede incrementar el desempleo y afectar al consumo en general (García et al, 2016). Asimismo, a medida que aumentan los fracasos empresariales, estos afectan de forma directa al sistema económico y pueden generar un efecto dominó negativo en la actividad económica.

Por otro lado, debido a la creciente turbulencia del entorno económico mundial y a los constantes cambios sociales y tecnológicos, la predicción oportuna de la quiebra empresarial se ha convertido en un desafío para los economistas, dado el impacto que tiene en todos los agentes relacionados con la empresa (Contreras, 2015). A esto se suma el entorno dinámico en el que operan las empresas, el cual puede ser afectado por diversos factores tanto endógenos como exógenos, lo que

subraya la importancia de disponer de herramientas para anticipar estas situaciones de insolvencia. Algunos países han implementado medidas para evitar que las empresas se liquiden directamente, brindando la posibilidad de presentar planes de reestructuración que permitan su continuidad, junto con incentivos y exoneraciones para aquellas que se acojan a estas leyes (García, 2023). En el caso de Colombia, se cuenta con la Ley 1116 de 2006, conocida como Ley de Insolvencia Empresarial.

Antecedentes de la Ley de Insolvencia Empresarial en Colombia:

En su primera etapa, se creó el Decreto 750 de 1940 (Colombia, Presidencia de la República, 1940), el cual enfatizaba la protección a los acreedores y la sanción al deudor. Este decreto empleaba un régimen severo con el deudor con el fin de garantizar la seguridad del crédito, inhabilitando comercialmente al deudor y castigando penalmente todo acto de negligencia o engaño.

Después, se comenzó a buscar la protección de la empresa, abandonando la postura severa contra el deudor. Sin embargo, debido a las dificultades económicas que atravesaba el país en los años 80, se presentaron grandes deficiencias en esta normativa, lo que llevó al desarrollo de proyectos para reformar dichos decretos.

Bajo la Constitución de 1991, se consolidó la protección de la empresa como finalidad de la ley, reconociendo a la empresa como fuente de empleo y explotación económica. A pesar de su importancia, la crisis económica de finales de los años 90 dejó un gran número de empresas en insolvencia. Ante esta situación, se hizo necesario un proceso transitorio más ágil para la recuperación económica. Por esta razón, se expidió otra ley cuyo objetivo era alcanzar un convenio con los acreedores, con el fin de mantener la empresa y realizar los ajustes necesarios para superar la crisis sin necesidad de ser liquidada.

Finalmente, se consolidó la normativa del régimen de insolvencia, priorizando la protección de la empresa, con la Ley 1116 de 2006. Esta ley tiene como objetivo "la protección del crédito y la recuperación y conservación de la empresa como unidad de explotación económica y fuente generadora de empleo, a través de los procesos de reorganización y de liquidación judicial, siempre bajo el criterio de agregación de valor" (Colombia, Presidencia de la República, 2006), con el fin de contar con un régimen permanente y unificado.

Antecedentes:

Como se mencionó anteriormente, la necesidad de contar con herramientas para mitigar el riesgo de que una empresa caiga en insolvencia económica es evidente, y diversos estudios han abordado este problema. En Colombia, Salomón (2022) propuso medir la precisión de los modelos de predicción de insolvencia económica más utilizados en los cinco sectores más importantes del país afectados por la pandemia de Covid-19. Los resultados de este estudio muestran una predicción del 62% en el sector de manufactura, mientras que para el sector agropecuario solo tuvo una predicción del 33%

El uso de modelos de machine learning y métodos estadísticos es de gran relevancia para estudios de este tipo. Por ejemplo, Pérez, Lopera y Vásquez (2019) plantean predecir la fragilidad de las empresas colombianas mediante el uso de modelos estadísticos. En este trabajo, emplean diferentes modelos para medir la probabilidad de riesgo de quiebra empresarial e identificar el modelo que mejor se ajusta al problema. Entre los modelos seleccionados por los autores, el logístico mixto demostró tener el mejor desempeño para predecir la fragilidad empresarial con un porcentaje de clasificación correcta del 69,23%. Además, García (2023) lleva a cabo un estudio sobre modelos predictivos de insolvencia en el sector de la construcción en Colombia durante los períodos pre y pospandémico. Este estudio utiliza los modelos Logit y Probit. Finalmente, se

concluye que el modelo Logit fue el que mejor se ajustó para predecir la insolvencia económica en el sector de la construcción en Colombia.

Por otro lado, Correa y Castaño (2019) estudian la insolvencia empresarial en Colombia a través de indicadores financieros de liquidez, rentabilidad y endeudamiento, y contrastan los resultados de la regresión logística con el algoritmo de *boosting*. Concluyen que los indicadores financieros permiten pronosticar la insolvencia empresarial, aunque se debe recurrir a metodologías como el algoritmo de *boosting* que consideren la asimetría de la información.

Finalmente, como se ha presentado anteriormente, existen diversos métodos estadísticos y modelos de machine learning que se han utilizado en diferentes estudios para predecir la insolvencia económica en las empresas. La necesidad de implementar estas herramientas se hace cada vez más apremiante para mejorar la efectividad de la predicción y obtener un análisis más robusto de las causas de la insolvencia económica. Esto permitirá mitigar el impacto económico causado por este fenómeno. A continuación, se explicarán los modelos de clasificación que se utilizarán para el presente trabajo.

Random Forest:

Este algoritmo fue introducido por Leo Breiman en 2001 y se basa en la idea de "*Bagging*", que consiste en entrenar múltiples árboles de decisión a partir de múltiples subconjuntos de datos utilizando muestreo con reemplazo, con la finalidad de combinar sus predicciones para obtener un solo modelo mucho más robusto que cada árbol por separado.

Cada árbol se crea en dos etapas: primero, se genera un número considerable de árboles de decisión, cada uno con un subconjunto aleatorio de variables; después, cada árbol crece hasta su máxima extensión (Zuñiga, 2020, p. 2).

Las ventajas de este modelo según Canovas et al (2017) son:

- Pueden usarse para clasificación o predicción: En el primer caso, cada árbol “vota” por una clase y el resultado del modelo es la clase con mayor número de “votos” en todos los árboles, de forma que cada nueva observación se presenta a cada uno de los árboles y se asigna a la clase más “votada”. En el segundo caso, el resultado del modelo es el promedio de las salidas de todos los árboles.
- Es un modelo más simple de entrenar a comparación de otros modelos más complejos, pero tiene un rendimiento similar.
- Desempeño eficiente y certero en bases de datos grandes.
- Mantiene su precisión con proporciones grandes de datos perdidos.

Por otro lado, sus desventajas son:

- La visualización gráfica de los resultados puede ser difícil de interpretar.
- Las predicciones no son de naturaleza continua y no puede predecir mas allá del rango de valores del conjunto de datos usado para entrenar el modelo.

Las ventajas de este modelo lo hacen un modelo muy usado en clasificación de imágenes, otorgamientos de créditos, detección de fraude, finanzas, entre otros.

Catboost:

Este algoritmo se basa en la idea de "*Boosting*". Según Cano (2023), fue desarrollado por Prokhorenkova et al. en 2018 y fue diseñado para resolver los problemas identificados en modelos anteriores de *Boosting*, como *XGBoost*.

Catboost se basa en los árboles de decisión de *Gradient Boosting*. En la fase de entrenamiento, se construyen consecutivos conjuntos de árboles de decisión, los cuales dependen de los parámetros establecidos inicialmente. Este modelo utiliza

un detector que, cuando se activa, deja de construir árboles y así evita el sobreajuste.

El algoritmo *Catboost* incluye las siguientes etapas (Cano, 2023):

- Cálculo preliminar del número de participantes.
- Opcionalmente, transformar las variables categóricas a numéricas utilizando *Target Statistics* en lugar de *One Hot Encoding*.
- Opcionalmente, transformar las variables de texto en numéricas.
- Elegir la estructura del árbol.
- Calcular los valores en cada hoja.

Las ventajas de este modelo son:

- Manejo eficiente de variables categóricas.
- Evita el sobreajuste de manera efectiva.
- Alta precisión en tareas de clasificación y regresión.

Por otro lado, sus principales desventajas radican en:

- Requiere más tiempo de entrenamiento comparado con otros algoritmos.
- La complejidad del modelo puede dificultar su interpretación.

Finalmente, este modelo se aplica en sistemas de recomendación, Análisis del comportamiento del cliente, modelos de riesgo crediticio, predicción de ventas y análisis de mercado, entre otros.

Isolation Forest:

Este algoritmo se utiliza para la identificación y el tratamiento de *outliers*. Debido a su estructura basada en árboles, tiene buen desempeño con grandes volúmenes de datos, donde su enfoque principal es el aislamiento de las características anómalas (Rodríguez, 2023). Este modelo aprovecha propiedades de las

anomalías, como que son la minoría y tienen valores de atributos muy diferentes a los de las observaciones normales.

Es decir, las anomalías son pocas y diferentes, lo que las hace más susceptibles de aislamiento que los puntos normales. Debido a esta susceptibilidad al aislamiento, las anomalías se aíslan más cerca de la raíz del árbol, mientras que los puntos normales se aíslan en el extremo más profundo del árbol (Romero, 2023). Por lo tanto, se construyen varios árboles y las anomalías son aquellas que tienen una profundidad corta en dichos árboles.

Las ventajas de este modelo son:

- Eficiente en términos de tiempo y espacio.
- No requiere un conjunto de datos etiquetado.
- Capaz de manejar grandes conjuntos de datos y alta dimensionalidad.

Por otro lado, sus desventajas son:

- Sensible a la selección de hiperparámetros.
- Menos efectivo en conjuntos de datos con mucho ruido.

Este tipo de modelos tiene aplicaciones en detección de fraudes financieros, seguridad informática, detección de intrusos, monitorización de sistemas y redes, entre otros.

Finalmente, para la sintonización de los mejores hiperparámetros se hizo uso de un método de validación cruzada conocido como GridSearchCV, el cual realiza una búsqueda en cuadrícula sobre el conjunto de datos. Inicialmente, se establecen los criterios para la búsqueda de los mejores parámetros. Para el presente trabajo, se empleó la validación cruzada de 5 pliegues ($cv=5$). La métrica por optimizar en este caso es el F1-Score y se parametriza $n_jobs=-1$ para hacer uso de todos los núcleos del procesador, acelerando el proceso y entrenando el

modelo con los resultados arrojados por este método. Por otro lado, es importante mencionar la razón por la cuál no se usaron redes neuronales para efectos de este trabajo y se debe principalmente a los requisitos computacionales que demandan este tipo de modelos y a la distribución de nuestra variable objetivo (la cuál por su naturalidad y en el caso real, va a ser mucho mayor el número de empresas no insolventes que las empresas insolventes).

Las ventajas de este método son:

- Explora de manera exhaustiva el espacio de hiperparámetros.
- Mediante validación cruzada evalúa la robustez del modelo.
- Puede mejorar significativamente el rendimiento del modelo seleccionando la mejor combinación de los parámetros.

Por otro lado, sus desventajas son:

- Puede ser computacionalmente costoso, especialmente con grandes conjuntos de datos o muchas combinaciones de hiperparámetros.
- Técnicas como RandomizedSearchCV u optimización bayesiana pueden ser más eficientes para grandes espacios de búsqueda.

DISEÑO METODOLÓGICO

Para el desarrollo del presente trabajo, se empleó la metodología CRISP-DM (*Cross Industry Standard Process for Data Mining*), la cual fue creada en 1999 por la entonces Comunidad Económica Europea, hoy la Unión Europea.

Adicionalmente, es la metodología de referencia más utilizada en el desarrollo de proyectos de minería de datos. Entre sus ventajas se destaca la posibilidad de replicación de proyectos, su independencia en la industria, su neutralidad con respecto a las herramientas y su enfoque en las situaciones de negocio (García, G. 2018).

Sus etapas son: Comprensión del negocio, Comprensión de los datos, Preparación de los datos, Modelado, Evaluación e Implementación.

- **Comprensión del negocio:** Se presenta la introducción al problema de la insolvencia económica y el método que se empleará en el trabajo. Además, se explica por qué la insolvencia económica tiene efectos negativos en el país y se describe la situación actual de este fenómeno. Posteriormente, se establecen los objetivos a cumplir en el trabajo, así como los antecedentes legislativos y los trabajos previos relacionados con el tema.
- **Comprensión de los datos:** Se realiza un análisis exploratorio de los datos obtenidos en la página web de la Superintendencia de Sociedades para las Pymes en el año 2021. Se analiza la distribución de las diferentes variables del dataset para observar el comportamiento de estos datos en el país, incluyendo la visualización del número de empresas por departamento, entre otros aspectos.
- **Preparación de los datos:** Después de realizar el análisis exploratorio de los datos, se analiza la correlación de las variables con respecto a la variable objetivo. De esta forma, se determina qué variables se podrían eliminar del modelo, manteniendo solo aquellas que realmente expliquen el comportamiento de la variable objetivo.
- **Modelado:** Se entrenan los modelos CatBoost, Random Forest e Isolation Forest para determinar cuál de estos es el más adecuado para el estudio, sintonizando los mejores parámetros para cada uno de ellos basándose en aquellos que maximicen el F1-Score.
- **Evaluación:** Se evalúa el rendimiento de los modelos seleccionados mediante el *F1-Score*, así como a través del análisis de sus matrices de confusión.

- **Implementación:** Después de haber entrenado y evaluado los modelos, se elige el modelo más adecuado en base a su *F1-Score* para predecir la insolvencia económica.

DESARROLLO DEL TRABAJO

Análisis Exploratorio de los Datos:

El conjunto de datos que se analizará posteriormente consta de un total de 21,990 empresas, cada una con sus respectivos indicadores financieros. En total, el conjunto de datos está compuesto por 15 variables. La insolvencia económica servirá como la variable objetivo en el análisis. A continuación, observamos como se trataron los datos atípicos para la limpieza del conjunto de datos.

Inicialmente, se optó por eliminar los valores nulos del dataset los cuáles estaban presentes en dos variables. Se imputan estos valores debido a que para el posterior cálculo de los indicadores financieros, estos valores Nulos nos generarían errores de cálculo en los índices. Así mismo, se cambio el tipo de dato de la columna “Pasivos Corrientes Totales” para su adecuado procesamiento

Figura 1. Limpieza de valores nulos

```
Tipo societario                                0
Departamento de la dirección del domicilio    0
Efectivo y equivalentes al efectivo (CashAndCashEquivalents)  0
Cuentas comerciales por cobrar y otras cuentas por cobrar corrientes (TradeAndOtherCurrentReceivables)  908
Activos corrientes totales (CurrentAssets)      0
Total de activos (Assets)                      0
Pasivos corrientes totales (CurrentLiabilities) 222
Total pasivos (Liabilities)                    0
Patrimonio total (Equity)                      0
Total de patrimonio y pasivos (EquityAndLiabilities) 0
Utilidades                                    0
Insolvencia Económica                         0
Patrimonio                                    0
Utilidad Neta                                 0
dtype: int64

df = df.dropna(subset=['Cuentas comerciales por cobrar y otras cuentas por cobrar corrientes (TradeAndOtherCurrentReceivables)'])
df = df.dropna(subset=['Pasivos corrientes totales (CurrentLiabilities)'])
df['Pasivos corrientes totales (CurrentLiabilities)'] = df['Pasivos corrientes totales (CurrentLiabilities)'].astype('int64')
```

Después de limpiar las variables para el correcto cálculo de los índices, procedemos a definir y calcular cada uno de los indicadores financieros que se usarán como variables predictoras de la insolvencia económica (figura 2)

Figura 2. Indicadores financieros

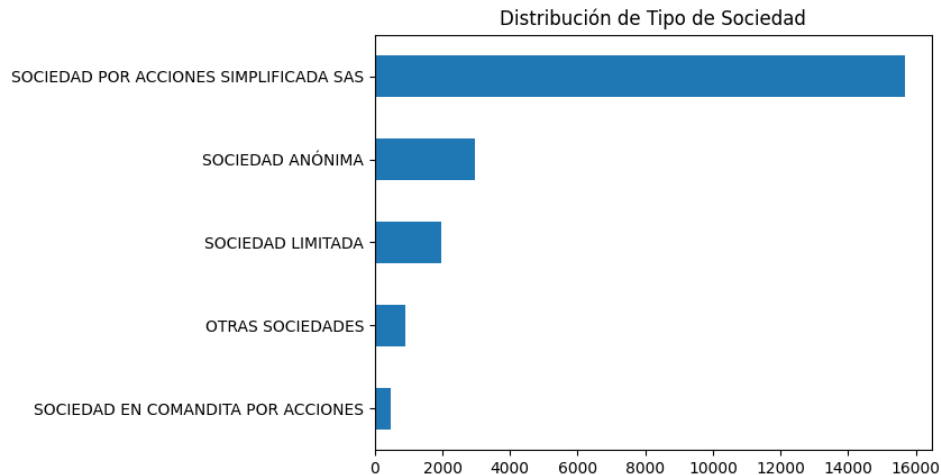
```
df['Ventas'] = df['Efectivo y equivalentes al efectivo (CashAndCashEquivalents)'] + df['Cuentas comerciales por cobrar y otras cuentas por cobrar corrientes']
df['Razon Corriente'] = df['Activos corrientes totales (CurrentAssets)'] / df['Pasivos corrientes totales (CurrentLiabilities)']
df['Rentabilidad del activo'] = df['Utilidades'] / df['Pasivos corrientes totales (CurrentLiabilities)']
df['Rentabilidad del Patrimonio'] = df['Utilidades'] / df['Total de activos (Assets)']
df['Nivel de Endeudamiento'] = df['Total pasivos (Liabilities)'] / ((df['Total pasivos (Liabilities)'] + df['Patrimonio total (Equity)']))
df['Apalancamiento Total'] = df['Total pasivos (Liabilities)'] / df['Patrimonio total (Equity)']
df['Relacion Activos y Pasivos'] = df['Total pasivos (Liabilities)'] / df['Total de activos (Assets)']
df['Rotacion Activos'] = df['Ventas'] / df['Total de activos (Assets)']
df['Relacion Activo Corriente y Total Activos'] = df['Activos corrientes totales (CurrentAssets)'] / df['Total de activos (Assets)']
df['Margen Neto'] = df['Utilidades'] / df['Ventas']
df['Apalancamiento a Corto Plazo'] = df['Pasivos corrientes totales (CurrentLiabilities)'] / df['Patrimonio total (Equity)']
```

Adicionalmente, podemos observar en la Figura 3 un resumen de la estadística descriptiva de las variables predictoras.

Figura 3. Descripción estadística de las variables predictoras

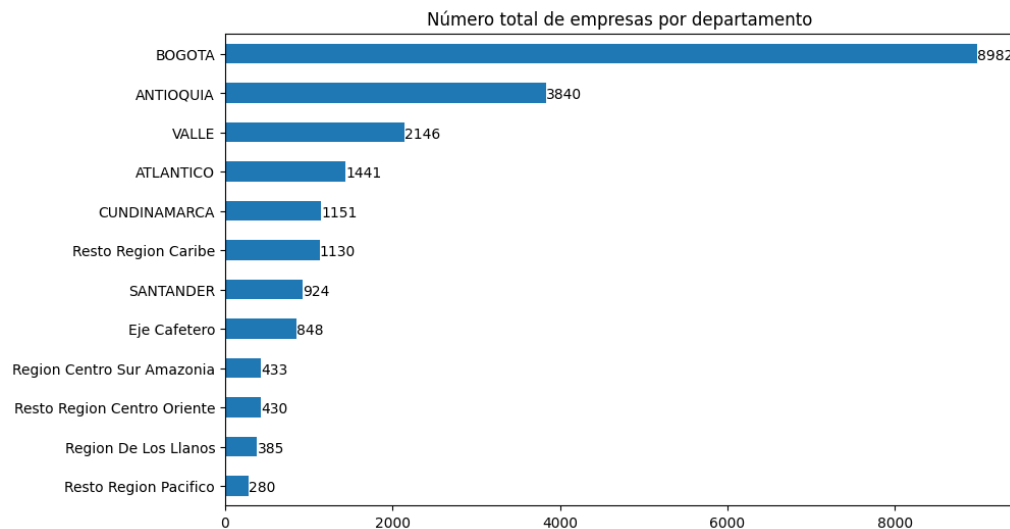
	Patrimonio	Utilidad Neta	Razon Corriente	Rentabilidad del activo	Rentabilidad del Patrimonio	Nivel de Endeudamiento	Apalancamiento Total	Relacion Activos y Pasivos	Rotacion Activos	Relacion Activo Corriente y Total Activos	Apalancamiento a Corto Plazo
count	21990.000000	21990.000000	2.199000e+04	21990.000000	21990.000000	21990.000000	21990.000000	21990.000000	21990.000000	21990.000000	21990.000000
mean	0.959573	0.786949	3.172026e+02	46.588616	0.009402	0.542933	5.415038	0.542933	0.352883	0.567317	3.429950
std	0.196964	0.409473	1.343517e+04	4405.719319	1.492860	1.772304	423.077288	1.772304	0.273107	0.329010	340.050386
min	0.000000	0.000000	0.000000e+00	-15770.474576	-140.998159	0.000002	-24792.971429	0.000002	0.000000	0.000000	-8277.939614
25%	1.000000	1.000000	1.259258e+00	0.009688	0.001937	0.243005	0.251778	0.243005	0.118028	0.262213	0.083546
50%	1.000000	1.000000	2.102172e+00	0.130878	0.026277	0.473657	0.783997	0.473657	0.301086	0.617008	0.392319
75%	1.000000	1.000000	4.636803e+00	0.458480	0.076007	0.696346	1.931110	0.696346	0.539903	0.876039	1.096897
max	1.000000	1.000000	1.524104e+06	613475.000000	80.995192	164.257620	48390.000000	164.257620	1.000000	1.000000	48390.000000

Figura 4. Distribución de empresas por Tipo de Sociedad



En la Figura 4, se puede observar claramente que el conjunto de datos está compuesto en su mayoría por empresas clasificadas como "Sociedad por Acciones Simplificadas (S.A.S.)", siendo esta categoría la que abarca la mayor parte del número de empresas estudiadas. Por otro lado, la categoría con menos empresas en el estudio es "Sociedad en Comandita por Acciones". Además, en la Figura 5 se muestran las empresas que componen el conjunto de datos por región.

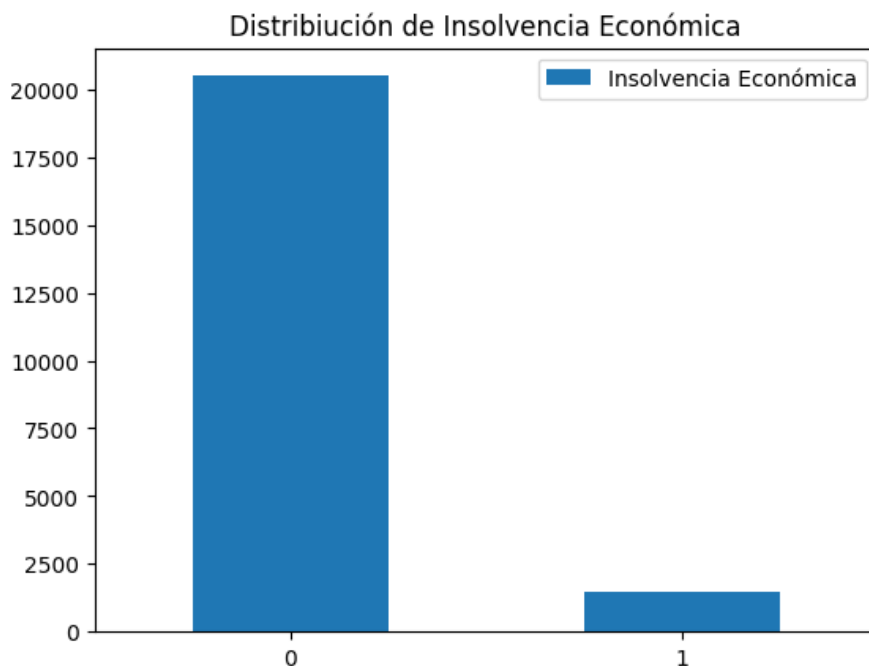
Figura 5. Cantidad de empresas por Región



Como era de esperarse, la mayoría de las pymes se encuentran en las ciudades más grandes y con mayor actividad económica. Bogotá es la ciudad con más empresas en el estudio (8,982 empresas), seguida de Antioquia (3,840 empresas) y el Valle del Cauca (2,146 empresas). Estas tres regiones concentran la mayor parte de las empresas del país. Por otro lado, el "Resto de la Región Pacífico", que comprende Cauca, Chocó y Nariño, presenta la menor actividad empresarial, con 280 empresas.

Para analizar más a fondo la variable objetivo "Insolvencia Económica", que indica 0 si la empresa no entró en insolvencia económica y 1 si efectivamente entró en un proceso de insolvencia económica, visualizamos la distribución de esta variable para entender el comportamiento presente en nuestro conjunto de datos (Figura 3).

Figura 6. Distribución de las empresas respecto a la variable objetivo



En la gráfica anterior, se observa una clara diferencia en las etiquetas de la variable objetivo, con un gran número de empresas etiquetadas como "0" y pocas empresas etiquetadas como "1". Aunque existe este evidente desbalance en los datos, esto refleja la realidad de las empresas, ya que en una economía que no esté en crisis y tenga un comportamiento "normal", es lógico encontrar muchas más empresas que no entran en un proceso de insolvencia económica en comparación con aquellas que sí lo hacen. Si la situación fuera al contrario, reflejaría una crisis y una posible caída económica del país, ya que implicaría que prácticamente todas las empresas estarían entrando en quiebra. Por lo tanto, dicho desbalance refleja la realidad económica de un país.

Por otro lado, posterior al cálculo de los diferentes indicadores financieros (como la razón corriente, rentabilidad del activo, margen neto, etc.) para

entrenar los modelos a partir de las variables financieras de las empresas, se calculó la matriz de correlación de las variables para de esta manera observar las variables que tengan mayor correlación entre sí y en base a esto depurar las variables teniendo en cuenta que una correlación alta entre las variables indica que explican el mismo comportamiento. Para el caso del presente trabajo, se consideran a depurar las variables con una correlación mayor a 0.9. Es importante mencionar que, teniendo en cuenta que el conjunto de datos está compuesto por varias variables continuas y siendo la variable objetivo una variable categórica, se optó por calcular también el coeficiente eta para estas variables ya que este coeficiente es mayormente usado para analizar la relación entre variables continuas y variables categóricas. Sin embargo, observamos que el coeficiente eta nos muestra valores similares e incluso inferiores de los registrados con el coeficiente Pearson (Figura 7 y 8).

Figura 7. Coeficiente eta

	Variable Continua	Coeficiente Eta
0	Razon Corriente	0.001023
1	Rentabilidad del activo	0.002528
2	Rentabilidad del Patrimonio	0.020382
3	Nivel de Endeudamiento	0.078848
4	Apalancamiento Total	0.000036
5	Relacion Activos y Pasivos	0.078848
6	Rotacion Activos	0.038821
7	Relacion Activo Corriente y Total Activos	0.029526

Figura 8. Selección y depuración de variables en base a su correlación

```
# Obtener la matriz de correlación
corr_matrix = df.corr()

# Crear una lista para almacenar las columnas a eliminar
columns_to_drop = []

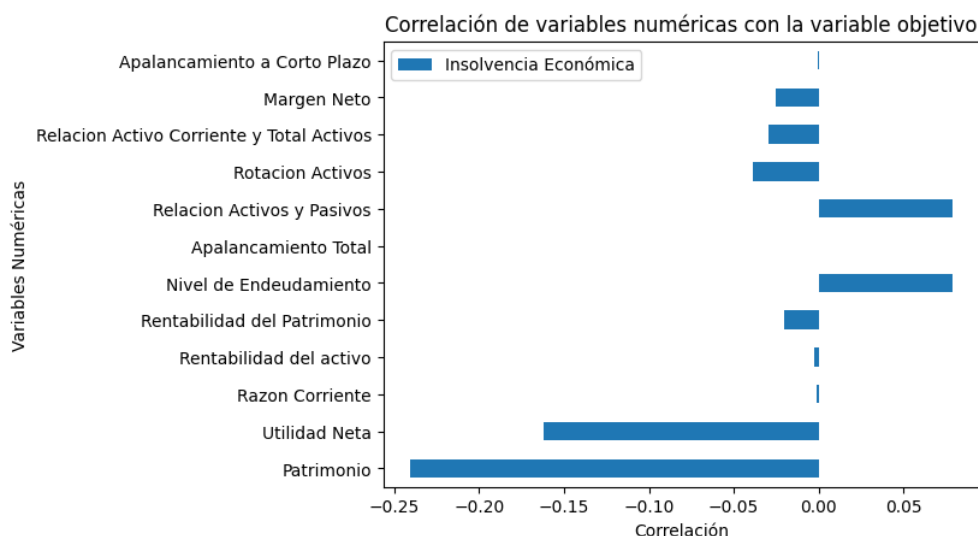
# Iterar sobre la matriz de correlación
for i in range(len(corr_matrix.columns)):
    for j in range(i+1, len(corr_matrix.columns)):
        if abs(corr_matrix.iloc[i, j]) > 0.9:
            column_name = corr_matrix.columns[j]
            columns_to_drop.append(column_name)

# Eliminar las columnas del DataFrame
df.drop(columns=columns_to_drop, inplace=True)

# Imprimir el DataFrame actualizado
df.head(5)
```

Después de observar la distribución de la variable objetivo y la correlación de las variables, procedemos a analizar la correlación existente entre las variables independientes, las cuales fueron previamente depuradas para seleccionar las que aportan significativamente al comportamiento de la variable objetivo. (Figura 9).

Figura 9. Correlación de las variables y la variable dependiente.



La gráfica anterior demuestra que, aunque ninguna variable tiene una correlación significativamente alta con la variable dependiente, ya que difícilmente se aproximan a ± 0.3 , se evidencia que las variables con mayor correlación respecto a la variable dependiente son "Patrimonio" y "Utilidad Neta", presentando una correlación negativa respecto a estas dos últimas. Adicionalmente, las variables que tienen una correlación positiva con la variable dependiente registran una correlación muy baja. Por ejemplo, La variable "Relación Activos y Pasivos" refleja una correlación positiva con la insolvencia económica. Teniendo en cuenta el cálculo de este indicador el cuál se muestra en apartados anteriores, se deduce desde el estudio del negocio que el aumento de este indicador financiero aumenta la probabilidad de insolvencia económica, teniendo en cuenta que este índice aumenta por un aumento en el numerador (en este caso el total de pasivos) por lo cual a mayor cantidad de pasivos en una empresa (deudas) es mayor la probabilidad de que esta entre en un proceso de insolvencia económica. Por lo que la realidad del negocio se asemeja a los resultados obtenidos en la correlación de las variables.

RESULTADOS

Entrenamiento de los modelos y resultados:

Posterior al análisis y la limpieza de los datos para obtener un conjunto óptimo que pueda ser procesado y evaluado, se seleccionaron tres modelos para ser entrenados. La elección del mejor modelo se basó en su F1-Score. Los hiperparámetros de cada modelo se ajustaron mediante validación cruzada (usando el método GridSearchCV) para entrenarlos con los parámetros que mejor se adaptaran y ofrecieran la mejor métrica de evaluación. Teniendo como variables explicativas del modelo, los índices financieros calculados en la Figura 2.

Es importante mencionar que, para el entrenamiento de los modelos se dividió la base de datos en subconjuntos de entrenamiento y prueba, para los datos de prueba se formaron subconjuntos compuestos por el 20% de los datos, calculando el F1 Score para estos conjuntos de prueba teniendo en cuenta la mejor combinación de hiperparametros.

Figura 10. Validación cruzada para el modelo CatBoost

```
# Definir los hiperparámetros a sintonizar
param_grid = {
    'iterations': [50, 100, 200],
    'learning_rate': [0.01, 0.05, 0.1],
    'depth': [6, 8, 10],
    'l2_leaf_reg': [1, 5, 10],
    'random_strength': [1, 5, 10]
}

# Definir el modelo CatBoost
model = CatBoostClassifier(random_state=42)

# Realizar la búsqueda de hiperparámetros
grid_search = GridSearchCV(model, param_grid, cv=5, scoring='f1_weighted', verbose=2, n_jobs=-1)
grid_search.fit(X_train, y_train)

# Imprimir los mejores hiperparámetros
print('Mejores hiperparámetros:', grid_search.best_params_)
```

Para la validación cruzada del modelo CatBoost y de los demás modelos entrenados en el trabajo, se sintonizaron los hiperparámetros que se visualizan en la figura 10 y se almacenaron en un diccionario definido como param_grid. Se realizaron 5 divisiones en el conjunto de datos (cv=5) para validar cada combinación de hiperparámetros. Estas combinaciones se evaluaron con base en el F1-Score Ponderado (f1_weighted), el cual tiene en cuenta el desbalance de la variable objetivo. Además, se utilizaron todos los núcleos disponibles de la máquina en paralelo (n_jobs=-1).

Figura 11. Validación cruzada para el modelo Random Forest

```
param_grid = {
    'n_estimators': [10, 50, 100, 200],
    'max_depth': [None, 10, 20, 30],
    'min_samples_split': [2, 5, 10],
    'min_samples_leaf': [1, 2, 4],
    'bootstrap': [True, False]
}

# Definir el modelo Random Forest
model = RandomForestClassifier(random_state=42)

# Realizar la búsqueda de hiperparámetros
grid_search = GridSearchCV(model, param_grid, cv=5, scoring='f1_weighted', verbose=2, n_jobs=-1)
grid_search.fit(X_train, y_train)

# Imprimir los mejores hiperparámetros
print('Mejores hiperparámetros:', grid_search.best_params_)
```

Figura 12. Validación cruzada para el modelo Isolation Forest

```
param_grid = {
    'n_estimators': [50, 100, 200],
    'max_samples': [0.5, 0.7, 1.0],
    'contamination': [0.01, 0.05, 0.1]
}

# Crear el modelo Isolation Forest
model = IsolationForest(random_state=42)
grid_search = GridSearchCV(model, param_grid, cv=5, scoring='f1_weighted', verbose=2, n_jobs=-1)
grid_search.fit(X_train, y_train)

# Imprimir los mejores hiperparámetros
print('Mejores hiperparámetros:', grid_search.best_params_)

# Entrenar el modelo con los mejores hiperparámetros
best_model = grid_search.best_estimator_
best_model.fit(X_train, y_train)

# Evaluar el modelo con los mejores hiperparámetros en el conjunto de prueba
y_pred = grid_search.best_estimator_.predict(X_test)
f1 = f1_score(y_test, y_pred, average='weighted')

# Ajustar umbral para mapear la etiqueta
threshold = np.percentile(best_model.decision_function(X_train), 100 * 0.01)

# Mapear etiquetas
labels = np.where(best_model.decision_function(X_train) < threshold, "anomalía", "normal")
```

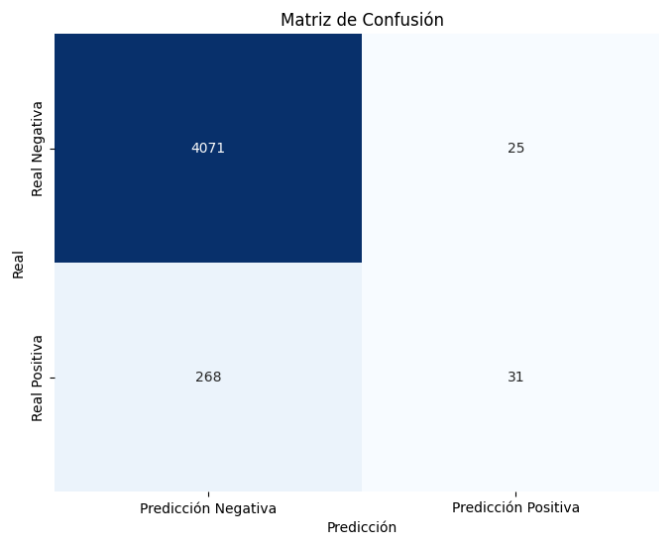
Como se mencionó anteriormente, los tres modelos utilizados en el trabajo se entrenaron con los mismos criterios para la validación cruzada. Lo que cambia son los parámetros que se van a sintonizar, ya que cada modelo tiene parámetros diferentes, los cuales están almacenados en el diccionario Param_Grid. Adicionalmente, para el modelo Isolation Forest se realizó el mapeo de etiquetas. Para el primer modelo se entrenó un modelo *Catboost*, los hiperparámetros escogidos para el modelo y sus resultados los podemos visualizar en la figura 13 y la figura 14.

Figura 13. Hiperparametros que mejor se ajustan al modelo CatBoost

```
# Imprimir los mejores hiperparámetros
print('Mejores hiperparámetros:', grid_search.best_params_)
✓ 0.0s

Mejores hiperparámetros: {'depth': 8, 'iterations': 200, 'l2_leaf_reg': 10, 'learning_rate': 0.1, 'random_strength': 1}
```

Figura 14. Matriz de confusión modelo CatBoost



A partir de la matriz de confusión anterior, podemos calcular la precisión y el Recall para posteriormente calcular el F1 Score, calculándolos de la siguiente manera:

$$\text{Precisión} = \text{TP} / (\text{TP} + \text{FP}) = 0.1037$$

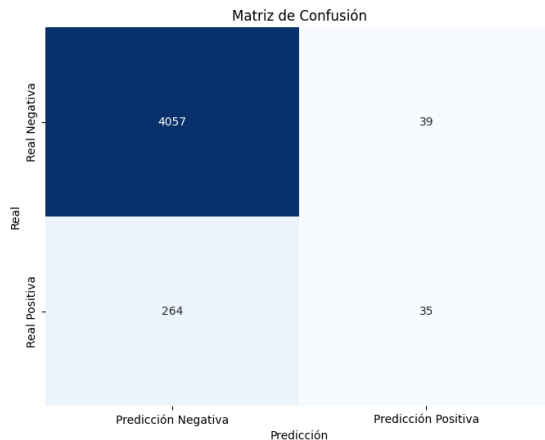
$$\text{Recall} = \text{TP} / (\text{TP} + \text{FN}) = 0.5536$$

$$\text{F1 Score} = 2 * \text{Precisión} * \text{Recall} / (\text{Precisión} + \text{Recall}) = 0.1747$$

Figura 15. Mejores Hiperparámetros del modelo Random Forest

```
Fitting 5 folds for each of 288 candidates, totalling 1440 fits
Mejores hiperparámetros: {'bootstrap': False, 'max_depth': None, 'min_samples_leaf': 2, 'min_samples_split': 5, 'n_estimators': 10}
```

Figura 16. Matriz de confusión modelo Random Forest



En las figuras 15 y 16, además de identificar el mejor ajuste para parámetros como '*Bootstrap*', '*max_depth*', '*min_samples_leaf*', '*min_samples_split*' y '*n_estimators*', realizando el mismo cálculo del apartado del modelo CatBoost (para el cálculo de Precisión, Recall y F1) observamos que este modelo registra un *F1-Score* de 0,1875. Con esto se concluye que, el mejor *F1-Score* de los tres modelos entrenados para este estudio lo tiene el modelo *Random Forest*. Por otro lado, es importante señalar que el modelo *Isolation Forest* resulta ser el menos adecuado para los propósitos de este estudio.

Figura 17. Mejores Hiperparámetros y resultados del modelo Isolation Forest

```
Mejores hiperparámetros: {'contamination': 0.01, 'max_samples': 1.0, 'n_estimators': 200}
C:\Users\espit\AppData\Local\Packages\PythonSoftwareFoundation.Python.3.10_qbz5n2kfra8p0\
warnings.warn(
F1-score: 0.008503981797497156
```

Como se mencionaba anteriormente, el modelo *Isolation Forest* fue el menos adecuado para el presente estudio. Para este modelo nos arroja un valor muy bajo (0.0085). Por lo tanto, nos inclinaríamos por los modelos anteriores antes que por este.

Plan de gestión de datos

Se utilizarán datos obtenidos de la Superintendencia de Sociedades, que incluyen los estados financieros de las pymes para el año 2022, así como otra base de datos que contiene información sobre las empresas que se declararon en insolvencia económica para el año siguiente. Estos datos se consolidarán en una sola base que incluirá la información financiera de la empresa y si esta se declaró en insolvencia económica en 2023. Es importante destacar que estos datos son de acceso público y están disponibles en el portal web de la Superintendencia de Sociedades.

Finalmente, se calcularán los índices financieros seleccionados previamente utilizando los datos financieros disponibles de la empresa. Estos índices incluirán medidas como la prueba de acidez, el nivel de endeudamiento, entre otros, y se calcularán a partir de los datos proporcionados.

Aspectos éticos

Los datos obtenidos de la Superintendencia se utilizarán exclusivamente para las actividades y objetivos establecidos en el trabajo. De igual manera, los resultados obtenidos del proyecto se emplearán para los fines previamente definidos.

Por otro lado, se espera que, con el desarrollo del modelo, las empresas puedan prever, basándose en sus índices financieros, si caerán o no en una situación de insolvencia económica.

CONCLUSIONES

En el presente trabajo, se abordó el problema de la insolvencia económica desde la perspectiva financiera de las pequeñas y medianas empresas del país, a través de la construcción de indicadores financieros para estas empresas en el año 2021. Se entrenaron diferentes modelos de machine learning para determinar, a través de la métrica de evaluación *F1-Score*, cuál sería el modelo más adecuado y con mejores métricas para ser utilizado como modelo de predicción de la insolvencia económica.

Por otro lado, mediante el análisis exploratorio de los datos se obtuvo una idea general de cómo está compuesto el dataset, observando la distribución de las empresas por departamento, por tipo de sociedad, etc. De esta forma, se tiene una idea más clara de cómo se componen los datos con los que vamos a trabajar. Asimismo, usando técnicas como la correlación entre las variables, fue posible eliminar aquellas que no serían relevantes para el estudio, dejando el dataset más limpio.

Se entrenaron modelos de CatBoost, Random Forest e Isolation Forest para determinar cuál de estos sería el modelo final, empleando el *F1-Score* como métrica de evaluación. Finalmente, en base a las métricas obtenidas, se escogió el modelo *Random Forest* como el mejor para el presente estudio, con un *F1-Score* de 0.1875. Cabe mencionar que el modelo *CatBoost* registra un puntaje similar al modelo Random Forest registrando un F1 Score de 0.1747, el cual está muy cerca del valor registrado en el entrenamiento del *Random Forest*. Por otro lado, el F1-Score obtenido con el modelo Isolation Forest (0.0085) es muy bajo, por lo que este modelo no sería una buena opción para predecir la insolvencia económica con los datos obtenidos para este estudio.

Finalmente, se recomienda el desarrollo de nuevas investigaciones derivadas del presente estudio o desde una perspectiva diferente al *machine learning* que puedan mitigar el impacto de la insolvencia económica, aplicando los modelos utilizados anteriormente o profundizando en el uso de otros métodos diferentes que también puedan resultar útiles para estudios en este ámbito. Asimismo, se recomienda realizar estudios con periodos de análisis más recientes para observar cómo se comporta este fenómeno en años posteriores a los escogidos en la presente investigación y poder compararlos.

REFERENCIAS

- Castro, C. (2024). Trabajo de Grado. GitHub.
<https://github.com/camilocastro12/Trabajo-de-Grado.git>
- Cano, J. (2023). Optimización del proceso de captación de clientes mediante machine learning: un estudio aplicado a Doofinder. Universidad Politécnica de Madrid, 1-58.
https://oa.upm.es/74930/1/TFM_JAVIER_CANO_DE_LA_TORRE.pdf
- Cánovas, F. Alonso, F. Gomariz, F. & Oñate, F. (2017). Modification of the random forest algorithm to avoid statistical dependence problems when classifying remote sensing imagery. Computers & Geosciences, 103, 1-11.
<https://doi.org/10.1016/j.cageo.2017.02.012>.
- Colombia, Presidencia de la República (1940). Decreto 750. “Se halla en estado de quiebra todo comerciante que sobresee en el pago corriente de sus obligaciones”. Bogotá, Diario Oficial, 24589.
- Colombia, Congreso de la República (2006). Ley 1116. “Por la cual se establece el Régimen de Insolvencia Empresarial en la República de Colombia y se dictan otras disposiciones”. Bogotá, Diario Oficial 46, 494.
https://www.sic.gov.co/recursos_user/documentos/normatividad/Ley_1116_2006.pdf
- Contreras, J. (2015). Análisis de quiebra empresarial: modelo de ecuaciones de estimación generalizadas sobre datos panel. Universidad Computense Madrid, 1-175. <https://docta.ucm.es/rest/api/core/bitstreams/a5ce59db-47ff-4afa-b493-2c973b1fb335/content>
- Correa, D. Castaño, M. (2019). Pronóstico de insolvencia empresarial en Colombia a través de indicadores financieros. Panorama económico, 27 (2), 510-526.

<https://revistas.unicartagena.edu.co/index.php/panoramaeconomico/article/view/2639/2220>

Cruz, A. & Pairetty, M. (2021). Analisis de insolvencia de las pymes del sector textil en Bogota a partir de sus indicadores financieros. Universidad de la Salle, 1-37. https://ciencia.lasalle.edu.co/finanzas_comercio/629

Espinosa, J. (2020). Aplicación de algoritmos Random Forest y XGBoost en una base de solicitudes de tarjetas de crédito. Ingeniería, Investigación y Tecnología. 21 (3), 1-16.

<https://www.revistaingenieria.unam.mx/numeros/2020/v21n3-02.pdf>

García, F. (2023). Estudio de modelos predictivos de insolvencia en el sector de la construcción en Colombia en los períodos pre pandemia y pos pandemia del Covid-19. Universidad EAFIT, 1-53.

https://repository.eafit.edu.co/bitstream/handle/10784/32644/FernandoArturo_GarciaVillegas_2023.pdf?sequence=2&isAllowed=y

García, G. (2018). Aplicación de la metodología CRISP-DM a la recolección y análisis de datos georreferenciados desde Twitter. Universidad Militar Nueva Granada, 1-23.

<https://repository.unimilitar.edu.co/bitstream/handle/10654/20099/GarciaVelezGustavoAdolfo2018.pdf?sequence=3>

García, J. Sánchez, J. Solano, E. (2016). Fracaso empresarial y efectos contagio. Un análisis espacial para España. El trimestre económico, 83 (330).

https://www.scielo.org.mx/scielo.php?script=sci_arttext&pid=S2448-718X2016000200429

García, M. (2018). Análisis de sensibilidad mediante Random Forest. Escuela Técnica Superior de Ingenieros Industriales de la Universidad Politécnica de

Madrid, 1-104.

https://oa.upm.es/53368/1/TFG_MARTA_GARCIA_RUIZ_DE_LEON.pdf

González, L. Prieto, F. (2018). Causas de insolvencia económica y estrategias para su solución en pequeñas y medianas empresas del calzado en Bogotá: barrio Restrepo. Repositorio Institucional Universidad EAFIT, 1-26.

<https://core.ac.uk/download/pdf/160244945.pdf>

Hernández, L. (2022). Crisis empresarial en Colombia, Probabilidad de entrar en proceso de insolvencia: 2016-2019. Universidad de los Andes, 1-28.

<https://repositorio.uniandes.edu.co/server/api/core/bitstreams/927f6f07-645e-4aea-83db-7587cecb49b2/content>

Infobae. (21 de septiembre de 2022). Ley de Insolvencia: este es el preocupante número de colombianos que se declararon en quiebra en los últimos 10 años. <https://www.infobae.com/america/colombia/2022/09/21/ley-de-insolvencia-este-es-el-preocupante-numero-de-colombianos-que-se-declararon-en-quiebra-en-los-ultimos-10-anos/>

La República. (16 de noviembre de 2023). En el primer trimestre del año crecieron 13,26% las empresas en insolvencia económica.

<https://www.larepublica.co/empresas/aumento-de-las-empresas-en-estado-de-insolvencia-economica-en-colombia-3749546>

León, P. Vargas, J. (2023). Factores que Afectan el Riesgo de Insolvencia Financiera de las Empresas Manufactureras de Colombia Según su Tamaño, Mediante los Modelos Z-Score De Altman, CA Score, Springate y Fulmer Durante el Periodo Pre-Pandemia y Pandemia (2017-2021). 1-153. https://repository.unab.edu.co/bitstream/handle/20.500.12749/21244/Tesis_Paula_Leon.pdf?sequence=1&isAllowed=y

- Ministerio de Comercio, Industria y Turismo. (21 de abril de 2023). Las microempresas fortalecen el tejido empresarial colombiano.
<https://www.mincit.gov.co/prensa/foto-noticias/microempresas-fortalecen-el-tejido-empresarial>
- Ministerio de Comercio, Industria y Turismo. (8 de marzo de 2024). Colombia cerró 2023 con número histórico de empresas activas.
<https://www.mincit.gov.co/prensa/noticias/industria/colombia-cerro-2023-con-historico-empresasactivas>
- Penagos, J. Muñoz, O. (2015). Modelo de predicción de insolvencia financiera aplicado al sector farmacéutico colombiano. 1-18.
<https://repository.eafit.edu.co/server/api/core/bitstreams/3d426760-3e64-492f-b048-8ba78e94971e/content>
- Naciones Unidas. (s.f). Fomentar las MIPYME mediante el apoyo a las mujeres y la juventud emprendedora y las cadenas de suministro resilientes.
<https://www.un.org/es/observances/micro-small-medium-businesses-day>
- Pérez, J. Lopera, M. Vásquez, F. (2019). Una breve aplicación a la predicción de la fragilidad de empresas colombianas, mediante el uso de modelos estadísticos. Borradores departamento de economía, 84, 1-26.
https://bibliotecadigital.udea.edu.co/bitstream/10495/12027/1/PerezJorge_2019_AplicacionPrediccionFragilidad.pdf
- Portafolio. (22 de julio de 2022). Cuántas empresas se han declarado en insolvencia en 2022.
<https://www.portafolio.co/economia/finanzas/disminuyo-la-insolvencia-y-la-confianza-sectorial-se-recupero-a-mayo-568526>
- Romero, J. (2023). Detección de anomalías transaccionales aplicando técnicas de machine learning con grafos. Universidad del Rosario, 1-46.

<https://repository.urosario.edu.co/server/api/core/bitstreams/148e3ef8-0f0c-4563-9e9e-71480ef7b41e/content>

Salomon, D. (2022). Exactitud y modelamiento de la insolvencia empresarial en los sectores colombianos durante la pandemia Covid-19. Universidad ICESI, 1-44.

https://repository.icesi.edu.co/biblioteca_digital/bitstream/10906/96477/1/TG03614.pdf