

SEMINARIO DE INVESTIGACIÓN

DESARROLLO DE UN MODELO PREDICTIVO DE RENTABILIDAD EMPRESARIAL EN EMPRESAS DE LA CADENA ALIMENTARIA DEL HUILA MEDIANTE EL ANÁLISIS DE VARIABLES FINANCIERAS Y TENDENCIAS DEL MERCADO REGIONAL

Presentado por:

CARLOS ALBERTO AMADOR ROSALES

DANIEL ALEJANDRO RODRÍGUEZ DÍAZ

JUAN CARLOS ORTÍZ

LEANDRO REYES JORDÁN

UNIDAD DE ESTUDIO:

SEMINARIO INVESTIGACIÓN - ESPECIALIZACIÓN - GRUPO 2

SEGUNDO SEMESTRE – 2024

FACULTAD DE INGENIERÍA

BOGOTÁ, 25 DE NOVIEMBRE DE 2024

UNIVERSIDAD EAN

RESUMEN

El departamento del Huila enfrenta desafíos económicos que amenazan la estabilidad de su agroindustria, especialmente en la producción de café, debido a la volatilidad en los mercados internacionales. Esta situación ha impactado negativamente las exportaciones y la supervivencia empresarial, con solo una cuarta parte de las empresas fundadas en 2017 manteniéndose operativas cinco años después. Ante este panorama, la investigación se enfoca en desarrollar un modelo predictivo basado en técnicas avanzadas de machine learning para analizar variables financieras clave y tendencias del mercado que afectan la rentabilidad y supervivencia de las microempresas. Este enfoque busca ofrecer soluciones estratégicas que promuevan la sostenibilidad y competitividad económica en el Huila, alineando los esfuerzos regionales con los objetivos de desarrollo a largo plazo, y fomentando la innovación en la agroindustria y otros sectores clave.

Palabras clave: Agroindustria, supervivencia empresarial, machine learning, rentabilidad, volatilidad económica, Huila.

TABLA DE CONTENIDO

PROBLEMA DE INVESTIGACIÓN	7
OBJETIVOS.....	9
Objetivo General:	9
Objetivos Específicos:	9
Conveniencia de la investigación.....	10
JUSTIFICACIÓN.....	11
MARCO TEÓRICO	13
Sector de la cadena alimentaria en el Huila.....	13
Factores de Éxito y Fracaso Empresarial en el Sector Alimentario.....	14
Impacto Económico y Social de la Longevidad Empresarial	15
Modelos Predictivos en la Gestión Empresarial.....	16
Tecnología y Transformación Digital en el Sector Alimentario	17
Fundamentos de Machine Learning	18
Aplicación de Machine Learning al Análisis de Supervivencia Empresarial	19
Desafíos y Oportunidades de la Implementación de Machine learning en Empresas del Huila	19
Impacto de Machine Learning en la Predicción Financiera y la Rentabilidad Empresarial	20
Marco Legal y Ético del Uso de Machine Learning en Colombia	21
Marco Institucional.....	22

METODOLOGÍA DE LA INVESTIGACIÓN	24
Primer nivel	24
Enfoque, alcance y diseño de la investigación.	24
Definición de Variables.	25
Población y Muestra	30
Segundo Nivel	30
Selección de Métodos o Instrumentos para la Recolección de Información	30
Características del Instrumento Utilizado: Base de Datos de la Cámara de Comercio del Huila.....	31
Técnicas de Análisis de Datos	32
Análisis Cuantitativo.....	32
Estadística Descriptiva	32
Regresión de Cox para el Análisis de Supervivencia.....	32
Modelos Predictivos: Random Forest y Árboles de Decisión	32
Validación de Modelos	33
Justificación de las Técnicas.....	33
ANÁLISIS Y DISCUSIÓN DE LOS RESULTADOS	34
Análisis Exploratorio de Datos (EDA)	34
Variable 1: Estado de Matrícula	34
Variable 2: Período	35
Variable 3: Organización Jurídica	36
Variable 4: Categoría	37
Variable 5: Zona	38
Variable 6: Barrio	39
Variable 7: Total Valor Activos	39
Variable 8: Total Valor Pasivos	40

Variable 9: Total Patrimonio	41
Variable 10: Total Ventas Netas.....	42
Variable 11: Tamaño.....	43
Variable 12: Sector agregado	44
Variable 13: Sector Actividad Económica.....	45
Variable 14: Actividad Económica Principal	46
Variable 15: Personal ocupado	47
Variable 16: Municipio.....	47
Modelado Predictivo	48
Análisis de la gráfica de correlación	51
Comparación de modelos.....	54
DISCUSIÓN	57
CONCLUSIONES	60
REFERENCIAS	62

TABLA DE FIGURAS

Figura 1. Estado Matrícula.....	35
Figura 2. Organización Jurídica	36
Figura 3. Categoría.....	37
Figura 4. Zona	38
Figura 5. Barrio.....	39
Figura 6. Total Valor Activos.....	40
Figura 7. Total Valor Pasivos.....	41
Figura 8. Total Patrimonio.....	42
Figura 9. Total Ventas Netas	43
Figura 10. Tamaño	43
Figura 11. Sector agregado	44
Figura 12. Sector Actividad Económica	45
Figura 13. Categorías actividad económica principal de las empresas.	46
Figura 14. Cantidad de personal ocupado en las empresas.	47
Figura 15. Municipio	48
Figura 16. Matriz de Correlación.....	51
Figura 17. Top 10 Importancia Variables	53
Figura 18. Análisis de Residuos.....	55

PROBLEMA DE INVESTIGACIÓN

La economía del Huila ha sido tradicionalmente impulsada por la agroindustria, con un énfasis particular en la producción de café, que, junto a otros alimentos, constituye una parte considerable de la actividad industrial del departamento. Sin embargo, algunos sectores específicos han mostrado una volatilidad notable, en gran parte debido a las fluctuaciones en los mercados internacionales, lo que se tradujo en una reducción del 29,4% en las exportaciones durante el año 2023 (Departamento Administrativo Nacional de Estadística, 2023).

A nivel nacional, durante el 2022, el 62,8% de las empresas que nacieron en el 2021 supervivieron, mientras que solo el 33,9% de las que nacieron en 2018 sobrevivieron para el 2022 (Departamento Administrativo Nacional de Estadística, 2023). Lo cual también podría ser una cuestión de género, pues, la tasa de supervivencia de las empresas activas económicas da cuenta de que para 2022, por cada 100 empresas creadas en 2021, 71,1% empresas con representante legal mujer y 72,5% empresas con representante legal hombre seguían activas; mientras que, por cada 100 empresas creadas en 2018, 49,7% empresas con representante legal mujer y 53,3% empresas con representante legal hombre seguían activas (Departamento Administrativo Nacional de Estadística, 2023); lo que puede implicar algunas desigualdades o inequidades de género que requieren de un profundo análisis para determinar las causas de esta disparidad. En este mismo sentido, el departamento del Huila también enfrenta una alta tasa de fracaso empresarial; de las 8.427 compañías que se fundaron en 2017, solo el 60,21% continuaba en operación al cabo de un año; el cual se redujo al 34,54% después de tres años, y únicamente el 28,54% de estas empresas logró sobrevivir cinco años (Cámara de Comercio del Huila, 2023). Estos datos revelan no solo los riesgos para la estabilidad económica en el país y en la región, sino también la preocupación por la sostenibilidad a largo plazo de la cadena alimentaria.

En cuanto a la ciudad de Neiva, tanto factores internos como externos influyen significativamente en la supervivencia de las empresas. Según un estudio, el 60% de los empresarios señala la falta de mercado como la principal causa de la iliquidez que enfrentan (Navia, 2018). En respuesta a estos desafíos, el departamento del Huila ha formulado diversas estrategias productivas dentro del marco de actualización de la agenda interna de productividad y competitividad (García, Vargas, & Pérez Gómez, 2015). Entre estas estrategias se encuentran la agroindustria basada en tecnología, con productos como café, cacao y diferentes variedades de frutas de la familia Passiflora, como la granadilla, cholupa, maracuyá, badea, curuba y gulupa. También se ha potenciado el sector turístico, que ofrece una variedad de actividades al aire libre, como navegación de aventura, espeleología, cabalgatas, ciclismo de montaña, senderismo y turismo rural. Otros sectores destacados incluyen la piscicultura, que ha sido reconocida como un sector de clase mundial por el programa de transformación productiva, así como la minería y la energía, considerados fundamentales para el desarrollo económico regional (García G. P., 2015).

La predicción de los ingresos de estas actividades resulta compleja y multifacética, debido a la variabilidad de la demanda del mercado, la competencia, eventos inesperados y factores naturales (Hernández, Julio, & Pantoja Wilches, 2024). Cada uno de estos elementos puede tener un impacto significativo y difícil de prever en la economía regional. Estudios recientes han destacado cómo las técnicas de machine learning pueden ser empleadas para prever la supervivencia de nuevas empresas en el mercado colombiano. Estos estudios utilizan modelos de clasificación y la regresión de Cox para analizar las características tanto del emprendedor como del negocio, que podrían influir en su estabilidad en el mercado (Román Ramírez D. , Supervivencia de las nuevas empresas: una aproximación desde el Machine Learning, 2021).

A pesar del potencial que estas técnicas ofrecen, su adopción en regiones como el Huila sigue siendo limitada. Sin embargo, la integración de estas tecnologías en el tejido empresarial podría

no solo mejorar la competitividad regional, sino también alinearse con los objetivos estratégicos de largo plazo del departamento. Esto permitiría promover la innovación, agregar valor y fomentar un desarrollo sostenible que posicione al Huila a la vanguardia del desarrollo tecnológico y empresarial.

Pregunta de Investigación: ¿En qué medida las tendencias del mercado regional del Huila afectan la rentabilidad de las microempresas y cómo puede preverse la rentabilidad futura teniendo en cuenta estas tendencias?

OBJETIVOS

Objetivo General:

Desarrollar un modelo predictivo para estimar la rentabilidad de empresas en la cadena alimentaria del Huila usando indicadores financieros y características regionales

Objetivos Específicos:

1. Identificar y categorizar las variables financieras críticas que influyen significativamente en la rentabilidad de las empresas de la cadena alimentaria del Huila, empleando métodos avanzados de análisis estadístico y econométrico.
2. Desarrollar un modelo de análisis multivariante que integre las tendencias del mercado regional y las variables financieras, evaluando su interacción y el impacto combinado en la rentabilidad de las empresas de la cadena alimentaria.
3. Implementar y validar un modelo predictivo basado en técnicas de machine learning, que permita anticipar la rentabilidad futura de las empresas de la cadena alimentaria, ajustando las predicciones en función de las dinámicas del mercado regional.

4. Proponer un marco teórico-práctico para la toma de decisiones empresariales, que utilice los resultados del modelo predictivo como herramienta estratégica para mejorar la sostenibilidad y el crecimiento económico de las empresas de la cadena alimentaria en entornos volátiles como el del Huila.

Conveniencia de la investigación

La investigación sobre el desarrollo de un modelo predictivo de la rentabilidad empresarial en las empresas de la cadena alimentaria del Huila presenta varias conveniencias. En primer lugar, identificar y categorizar las variables financieras críticas permite a las empresas comprender de forma anticipada cuáles son sus factores de riesgo, facilitando la toma de decisiones oportunas para proteger las utilidades presupuestadas. Por otra parte, el desarrollo de un modelo de análisis multivariante que integra tendencias del mercado regional con variables financieras proporciona un enfoque integral del entorno económico, lo que facilita la toma de decisiones basadas en modelos predictivos que consideran los cambios del sector. Además, el uso de técnicas de machine learning permitirá a las empresas mejorar su eficiencia operativa y obtener una ventaja competitiva frente a sus competidores. Finalmente, la propuesta de un marco teórico-práctico ofrece una guía para mejorar la sostenibilidad y el crecimiento económico de las empresas, especialmente en un entorno tan volátil como el del Huila. En resumen, esta investigación no solo proporciona herramientas avanzadas y conocimiento específico para las empresas locales, sino que también contribuye al desarrollo económico de la región.

JUSTIFICACIÓN

La creciente complejidad y dinamismo de los mercados globales y regionales exige a las empresas, particularmente a las pequeñas y medianas empresas (PYMES) en regiones como el Huila, adoptar estrategias avanzadas para asegurar su rentabilidad y sostenibilidad. En este contexto, la predicción de la rentabilidad empresarial mediante el análisis de variables financieras y las tendencias del mercado regional se presenta no solo como una necesidad, sino como una oportunidad crítica para mejorar la toma de decisiones estratégicas y operativas (Nguyen, 2022).

El sector agroindustrial en el Huila, que ha sido tradicionalmente impulsado por la producción de café y otros productos agrícolas, enfrenta retos significativos debido a la volatilidad del mercado y a las fluctuaciones en los precios internacionales (Departamento Administrativo Nacional de Estadística (DANE, 2024). Esta volatilidad ha llevado a una alta tasa de fracaso empresarial en la región, con solo el 28,54% de las empresas fundadas en 2017 sobreviviendo después de cinco años (Cámara de Comercio del Huila, 2023). Este escenario subraya la urgente necesidad de herramientas predictivas que permitan a las empresas anticiparse a los cambios en el mercado y ajustar sus estrategias para garantizar su rentabilidad y supervivencia.

El uso de técnicas de machine learning en la predicción de la rentabilidad empresarial representa un avance significativo en la gestión financiera y estratégica. Estas técnicas permiten analizar grandes volúmenes de datos y detectar patrones complejos que serían invisibles mediante métodos tradicionales (Domingos, 2015). Además, la capacidad de estos modelos para adaptarse y mejorar con el tiempo ofrece a las empresas una ventaja competitiva sustancial en un entorno económico volátil y altamente competitivo (Jordan & Mitchell, 2015).

La investigación que se propone no solo contribuirá al cuerpo de conocimiento académico sobre la aplicación de machine learning en la gestión empresarial, sino que también ofrecerá un marco práctico para la toma de decisiones en empresas del sector agroindustrial del Huila. Este marco permitirá a los empresarios locales identificar y mitigar riesgos financieros, mejorar la eficiencia operativa, y optimizar su desempeño en el mercado regional. En última instancia, el éxito de esta investigación puede servir como un modelo replicable para otras regiones con dinámicas económicas similares, promoviendo así el desarrollo económico sostenible a nivel local y nacional (Nguyen, 2022).

El desarrollo de un modelo predictivo de rentabilidad empresarial mediante el análisis de variables financieras y tendencias del mercado regional es no solo relevante, sino imprescindible para enfrentar los desafíos actuales del sector agroindustrial en el Huila. Este proyecto tiene el potencial de transformar la gestión empresarial en la región, promoviendo la innovación, la sostenibilidad y la competitividad en un mercado cada vez más globalizado.

MARCO TEÓRICO

Sector de la cadena alimentaria en el Huila

El departamento del Huila en Colombia es un pilar fundamental en la producción agrícola y la cadena alimentaria del país, destacándose en sectores como la piscicultura, la producción de café, y otras agroindustrias. Según diversos estudios, el modelo de asociatividad ha sido crucial para fortalecer a las pequeñas y medianas empresas (pymes) agroindustriales del Huila (Bada Carbajal, L. M., Rivas Tovar, L. A., & Littlewood Zimmerman, H. F., 2017). Este modelo permite a las pymes mejorar su competitividad a través de alianzas estratégicas, acceso a recursos y mercados, reducción de costos, e impulso a la innovación.

El sector piscícola, en particular, ha posicionado al Huila como uno de los principales productores y exportadores de tilapia en Colombia. La piscicultura es prioritaria en la Agenda Interna de Competitividad del Huila, y la exportación de tilapia a mercados como Estados Unidos, Canadá, y Europa ha sido un motor de crecimiento económico regional. Este éxito refleja la capacidad del Huila para aprovechar sus ventajas comparativas y especialización en el mercado acuícola, consolidándose como un actor relevante a nivel nacional e internacional (Cerquera Losada, Gómez Segura, & Arias Barrera, 2020).

El Huila es uno de los mayores productores de café, aguacate, cacao, y plátano en Colombia. Las agroindustrias del café y cacao, en particular, tienen un impacto significativo en la economía local (Gobernación del Huila, 2023). El café del Huila, reconocido por su calidad, es prioritario en la Agenda Interna de Competitividad, impactando considerablemente en el empleo y en el Producto Interno Bruto (PIB) del departamento (Cerquera Losada, Gómez Segura, & Arias Barrera, 2020).

Durante 2012-2016 el café ocupó la mayor área cosechada con un promedio de 103.741 hectáreas, seguido del arroz riego (31.722 ha), fríjol tecnificado (16.250 ha), maíz tecnificado

amarillo (11.424 ha), y maíz tradicional amarillo (11.326 ha). Desde el comportamiento del índice de participación, los cultivos más representativos del departamento del Huila, en términos de oferta agropecuaria, fueron el café con 41,25%, y el arroz riego que alcanzó el 12,61% (Muñoz-Nieto, 2018). El café ha mostrado un crecimiento constante en los mercados internacionales, lo que evidencia la competitividad del café huilense y su importancia en las exportaciones del departamento (Losada, Gómez, & Chavarro, 2020).

La cadena alimentaria del Huila es un motor clave en la economía regional, representando aproximadamente el 20% del PIB departamental, un porcentaje significativamente alto en comparación con el promedio nacional (DANE, 2024). Este sector también genera un 35% del empleo, abarcando desde la producción hasta la comercialización (DANE, 2024). La comparación con otros sectores como el minero y manufacturero resalta aún más la relevancia del sector alimentario en la economía del Huila.

La innovación y la adopción de tecnologías, como la agricultura inteligente y los Sistemas de Información Geográfica (SIG), están transformando el sector agrícola del Huila. Estas tecnologías mejoran la productividad y competitividad, promueven prácticas más sostenibles, alineadas con las demandas globales de sostenibilidad y protección ambiental (Gobernación del Huila, 2020).

No obstante, el sector enfrenta retos significativos, como la mejora de infraestructuras de transporte y almacenamiento para reducir pérdidas postcosecha, que pueden afectar hasta el 40% de la producción en algunos subsectores (Rodríguez Borray, Cruz Castiblanco, Tauta Muñoz, Huertas Carranza, & Polo Murcia, 2022).

Factores de Éxito y Fracaso Empresarial en el Sector Alimentario

El éxito empresarial en los sectores agrícola, pecuario y agroindustrial se fundamenta en tres teorías clave, según (Islas Moren, Rodríguez Ortega, & Aguirre Álvarez, 2023). En el ámbito

agrícola, la eficiencia productiva es esencial, lograda mediante la gestión óptima de recursos como tierra y agua para maximizar la rentabilidad. En el sector pecuario, la integración vertical permite a las empresas agregar valor y mejorar su competitividad al controlar toda la cadena de valor. En el sector agroindustrial, la diferenciación se destaca como crucial, enfocándose en la calidad, innovación y sostenibilidad para crear una ventaja competitiva sostenible en el mercado. En resumen, todo lo referente al ámbito agrícola y la eficiencia productiva se interpreta como una columna esencial, que, mediante la gestión administrada de recursos naturales, estas pueden extender la utilidad y la sostenibilidad a corto y largo plazo de los cultivos. La implementación de tecnologías emergentes y de aprendizaje como la agricultura de precisión y la automatización de procesos, permite a las pequeñas empresas en general mejorar sus modelos empresariales, disminuir costos y el impacto ambiental. Estas tecnologías no solo mejoran la eficiencia, sino que también son coherentes con las progresivas demandas de los consumidores por productos más sostenibles y responsables creando una ventaja competitiva, que contribuye con el desarrollo económico y social en la región.

Impacto Económico y Social de la Longevidad Empresarial

El parque empresarial formal en Colombia está compuesto por 1,5 millones de empresas, donde el 91,5% son microempresas, el 8% pymes y el 0,5% grandes empresas (Asobancaria, 2021). Para apoyar la supervivencia de las pymes, se proponen estrategias como el acceso a instrumentos crediticios (factoring, leasing), ocupaciones exportadoras y programas de educación financiera. Según (Benavides Pupiales, 2023), la adaptación estratégica en entornos desafiantes, como la crisis del COVID-19, y acciones como la innovación, mejora de calidad y posicionamiento de marca son esenciales para que las pymes se mantengan competitivas y satisfagan las demandas del mercado. También, la unificación de conjunto de técnicas digitales y la ejecución de análisis predictivo han demostrado ser necesarios para mejorar los resultados empresariales, permitiendo a las pequeñas empresas, de manera práctica y sencilla predecir

alteraciones en la demanda y modificar sus indicadores y planes de negocio en el momento oportuno. Estas técnicas de aprendizaje automatizado no solo optimizan la eficiencia operativa, sino que también proporcionan la individualización de nuevas oportunidades en el mercado, lo que resulta muy favorable por un lado para la sostenibilidad económica de las empresas del huila si no como en el desarrollo social de la región objeto del estudio.

Modelos Predictivos en la Gestión Empresarial

El artículo de (Palencia Fajardo, Cortés Lemus, & Benavides Parra, 2022), destaca la importancia de la responsabilidad social empresarial (RSE) para la supervivencia de las pymes en La Plata, Huila. El compromiso con la RSE, especialmente en áreas como la protección ambiental y la adopción de prácticas socialmente responsables, es clave para el crecimiento económico sostenible, fortaleciendo la reputación, relaciones con stakeholders y competitividad a largo plazo.

Por otro lado, (Ccopa Mamani, M., & Chavez Viza, 2015) subrayan en su estudio sobre la mejora en la toma de decisiones en el sector hospitalario la importancia de la minería de datos, logrando un 91.97% de mejora en la toma de decisiones. En cuanto a las mipymes en Colombia, (Blandón, Zuluaga, Vásquez, López, & Moreno, 2024), presentan un modelo de analítica de flujo de caja para mitigar su desaparición. Este modelo, compuesto por módulos descriptivos, diagnósticos y predictivos basados en inteligencia artificial, facilita la comprensión y toma de decisiones sobre el flujo de caja, ayudando a prever la probabilidad de insolvencia. Esto contribuye significativamente al fortalecimiento del tejido empresarial y a la creación de empleo en el país.

La inteligencia artificial está modificando extremadamente la forma en el sector empresarial ejecuta planes financieros. Esta permite realizar pronósticos más exactos y mejorar los métodos de la gestión de riesgos. Un estudio reciente de (DYNA, 2023) resalta que las

organizaciones que acogen IA en sus operaciones financieras logran fortalezas competitivas como factor diferenciador entre sus competidores. Estas organizaciones logran predecir mejor sus balances financieros, estas también logran ventajas competitivas si automatizan revisiones críticas que ayudan con la detección de fraudes. En la región objeto de estudio donde la volatilidad en el sector es relevante como lo es la agroindustria del Huila, estas herramientas del IA son fundamentales para el crecimiento y sostenibilidad del sector. (DYNA, 2023)

Tecnología y Transformación Digital en el Sector Alimentario

La transformación digital en el sector alimentario implica la adopción de Tecnologías de la Información y Comunicación (TIC) para optimizar procesos, mejorar la interacción con los clientes y abrir nuevas oportunidades de negocio. Este cambio va más allá de simplemente implementar herramientas digitales; requiere una reconfiguración profunda de modelos de negocio, procesos productivos y la cultura organizacional en empresas y cooperativas agroalimentarias.

La digitalización busca responder a demandas actuales como la trazabilidad de productos, la personalización de la oferta y la eficiencia en la cadena de suministro, mediante tecnologías como Big Data, Internet de las Cosas (IoT) y comercio electrónico. Esto no solo mejora la eficiencia operativa, sino que también promueve la sostenibilidad y mejora la experiencia del cliente (Vázquez, Cebolla, & Ramos, 2019).

Un estudio sobre cooperativas agroalimentarias en España destaca la digitalización como clave para enfrentar los retos del mercado y asegurar la competitividad. Sin embargo, el nivel de transformación digital varía según el tamaño y sector de actividad, con sectores como el lácteo y vitivinícola mostrando mayor integración de tecnologías (Vázquez, Cebolla, & Ramos, 2019).

En Colombia, la transformación digital es vista como fundamental para mejorar la competitividad en el sector alimentario. Ejemplos como una empresa dedicada a la acuaponía e

hidroponía, fundada en 2016, muestran cómo la integración de tecnologías de la industria 4.0 puede promover el cuidado del medio ambiente y la alimentación saludable.

El avance hacia la digitalización en Colombia refleja tanto el potencial de las TIC para impulsar el desarrollo económico como los desafíos que aún existen para lograr una transformación integral en el sector alimentario. La adopción de tecnologías digitales se convierte en un pilar esencial para la competitividad y sostenibilidad en el contexto actual (Parejo, Nuñez, & Nuñez, 2021).

Fundamentos de Machine Learning

El Machine Learning es una rama clave de la inteligencia artificial enfocada en desarrollar algoritmos que aprenden de los datos para hacer predicciones o decisiones, sin necesidad de programación explícita. Este aprendizaje se realiza a través del análisis de patrones y estructuras en los datos. Existen diferentes metodologías dentro del ML: el aprendizaje supervisado, no supervisado y por refuerzo (Bishop, 2013).

El aprendizaje supervisado utiliza datos etiquetados para entrenar algoritmos, permitiéndoles predecir resultados en nuevos datos, siendo útil en tareas como la clasificación y regresión (Murphy, 2012). El aprendizaje no supervisado, en cambio, identifica patrones en datos no etiquetados, útil en análisis exploratorios y agrupamientos (Hastie, Tibshirani, Friedman, Tibshirani, & Friedman, 2009). Por otro lado, el aprendizaje por refuerzo se centra en que agentes tomen decisiones basadas en recompensas o penalizaciones, aplicándose en sistemas de recomendación y juegos (Sutton & Barto, 2018).

El ML tiene aplicaciones vastas, ofreciendo insights valiosos en áreas como medicina, finanzas y manufactura, gracias a su capacidad para aprender de grandes volúmenes de datos (Jordan & Mitchell, 2015). Sin embargo, enfrenta desafíos como la necesidad de grandes conjuntos de datos, el riesgo de sesgos y la dificultad en la interpretación de modelos,

especialmente en sectores donde la replicabilidad es crucial (Doshi-Velez & Kim, 2017). La investigación continúa para superar estos retos y maximizar los beneficios del ML de manera ética y efectiva.

Aplicación de Machine Learning al Análisis de Supervivencia Empresarial

El uso de Machine Learning en el análisis de supervivencia empresarial busca identificar los factores clave que permiten a las empresas, especialmente startups, mantenerse y prosperar en el mercado. A través del análisis de grandes bases de datos, se pueden detectar patrones y correlaciones que podrían pasar desapercibidos con métodos tradicionales. Aunque el ML se ha aplicado en varios sectores, su uso específico en la gestión de productos perecederos aún es limitado (Garcés Camacho, 2020). Sin embargo, su capacidad para manejar variables predictoras y optimizar la toma de decisiones lo convierte en una herramienta valiosa para mejorar la supervivencia empresarial.

El estudio de (Román Ramírez D. , Supervivencia de las nuevas empresas: una aproximación desde el Machine Learning, 2021) en Colombia, utilizando algoritmos como Random Forest, demuestra que el ML puede predecir la supervivencia empresarial con mayor precisión que los métodos tradicionales, proporcionando insights sobre las estrategias efectivas para el desarrollo empresarial. Otro estudio en Bogotá (Romero Acosta, 2019) destaca la necesidad de adaptar estos modelos a las particularidades del sector y la región.

Desafíos y Oportunidades de la Implementación de Machine learning en Empresas del Huila

La implementación del ML en empresas del Huila presenta desafíos y oportunidades que, con una estrategia adecuada, pueden transformar significativamente la forma en que estas empresas operan y compiten globalmente. Según (Barrera Molina, 2022), aunque la adopción de tecnologías digitales en el sector financiero del Huila es prometedora, aún persisten desafíos

importantes. El informe del (Departamento Administrativo Nacional de Estadística , 2019) destaca una brecha en la adopción de tecnologías avanzadas en regiones como el Huila, debido a la insuficiente infraestructura tecnológica y la falta de capital humano especializado en TIC y ML. Además, la (Cámara Colombiana de Informática y Telecomunicaciones , 2020) señala que la resistencia al cambio organizacional y la falta de una cultura de innovación son obstáculos clave.

Sin embargo, la implementación de ML ofrece oportunidades para mejorar procesos, personalizar servicios y desarrollar nuevos modelos de negocio. La (Comisión Económica para América Latina y el Caribe, 2019) indica que la digitalización y el uso de ML pueden mejorar la eficiencia operativa y el acceso a mercados internacionales. El (Banco Interamericano de Desarrollo, 2018) también sugiere que el ML puede impulsar la creación de empleo de calidad y el desarrollo de habilidades en la región. Para aprovechar estas oportunidades, es necesario un esfuerzo conjunto entre el sector privado, el gobierno y las instituciones educativas, lo que justifica la inversión en transformación digital.

Impacto de Machine Learning en la Predicción Financiera y la Rentabilidad

Empresarial

Los adelantos en las técnicas de machine learning están transformando la forma en que se predicen las variables financieras y la rentabilidad de las empresas. Un estudio reciente de (HANSEN, 2020) muestra cómo el uso de algoritmos de machine learning ha mejorado significativamente la precisión en las predicciones financieras en diversos sectores, lo que facilita una toma de decisiones más acertada en entornos competitivos. Este estudio también destaca que la capacidad de machine learning para analizar grandes cantidades de datos ayuda a identificar patrones financieros complejos que influyen directamente en la rentabilidad, proporcionando así una herramienta sólida para la planificación financiera estratégica (HANSEN, 2020), El uso de machine learning llego para aportar significativamente en los pronóstico financieros, estos ha sido explorado en diferentes estudios que fortalecen las actuales

investigaciones . Un estudio publicado en la Revista científica y académica (Salazar, 2024) este autor fortalece los estudios y destaca cómo las técnicas avanzadas de aprendizaje automático pueden ser aplicadas eficazmente en la predicción de la rentabilidad empresarial, refuerza que estas pueden ser adaptadas y funcionales especialmente en contextos volátiles. El autor robustece la relevancia de utilizar estas tecnologías para mejorar la precisión en las proyecciones financieras y optimizar la toma de decisiones estratégicas en las empresas (Salazar, 2024) destaca cómo la capacidad de machine learning para procesar grandes volúmenes de datos mejora sustancialmente la precisión de las predicciones financieras. Estos modelos de predicción son de gran relevancia principalmente en las micro y pequeñas empresas del sector agroindustrial del Huila, estas por múltiples factores como político, económico, ambiental afrontan un entorno volátil debido a la incertidumbre de los mercados nacionales e internacionales.

Marco Legal y Ético del Uso de Machine Learning en Colombia

El marco legal y ético del uso del ML en Colombia se basa en leyes y regulaciones que buscan equilibrar los avances tecnológicos con la protección de los derechos individuales y fomentar un desarrollo responsable. La Ley de Protección de Datos Personales de 2012 (Congreso de Colombia, 2012) y el Decreto 1377 de 2013 (Presidencia de la República de Colombia, 2013), son fundamentales, estableciendo condiciones para el tratamiento de información personal y garantizando la privacidad en proyectos de ML. Estas normas aseguran que los sistemas de ML manejen datos de manera ética, protegiendo los derechos de los individuos y exigiendo consentimiento informado y medidas de seguridad.

La Estrategia de Gobierno del (Ministerio de Tecnologías de la Información y las Comunicaciones, 2018) promueve el uso de tecnologías digitales, incluido el ML, para mejorar los servicios gubernamentales, estableciendo directrices sobre accesibilidad, seguridad y usabilidad aplicables más allá del sector público.

A nivel global, la recomendación de la (UNESCO, 2021) sobre la ética de la inteligencia artificial, ofrece un marco que Colombia podría adoptar, enfatizando la transparencia, la equidad y la responsabilidad. Además, en el contexto de la industria alimentaria y la salud, el uso de IA para optimizar procesos y personalizar servicios debe considerar riesgos como el sesgo algorítmico y la pérdida de empleo, así como el impacto ambiental.

Marco Institucional

Marco Institucional del Sector Agroindustrial en el Huila

El sector agroindustrial en el departamento del Huila ha sido históricamente un motor fundamental de la economía regional desde la época colonial, cuando la región comenzó a destacarse por su producción de café. Este cultivo rápidamente se convirtió en el principal impulsor económico no solo del Huila, sino de todo el país. Durante el siglo XIX, gran parte de la geografía del departamento se transformó en extensas plantaciones cafetaleras, acompañadas por el desarrollo de infraestructuras básicas. La construcción de caminos y puentes no solo facilitó el transporte y la comercialización del café, sino que también conectó regiones, impulsando el crecimiento de una agroindustria que permitió llevar este producto a los mercados nacionales e internacionales.

El sector agroindustrial en el Huila es una parte integral de la economía regional, este sector aporta al sostenimiento agrícola y económico del país, esta se destaca por su diversificación y capacidad de agregar valor con una amplia gama de productos agrícolas. Este sector agroindustrial incluye alimentos esenciales para la alimentación como la producción de café, cacao, frutas, también incluye la transformación de estos productos en alimentos procesados, combinando no solo la producción agrícola si no el procesamiento y la cadena de abastecimiento alimentaria como por ejemplo aceites esenciales, y biocombustibles.

El CIIU (Clasificación Industrial Internacional Uniforme) es un sistema estándar utilizado para clasificar las actividades económicas en diferentes sectores. (Camara de comercio Bogota, 2024) La agroindustria en el Huila se encuentra bajo el código CIIU 10-12 (Industria de alimentos y bebidas) y el CIIU 01 (Agricultura, ganadería, caza y actividades de servicios conexas). El Huila se está consolidando como líder nacional en productos agrícolas. El café, la granadilla, el arroz y la pitahaya son los renglones productivos bandera. Según el DANE, el Huila tiene 79.392 unidades productivas y 1.368.042 de hectáreas con uso de suelo en actividades agrícolas. (FINAGRO, 2024)

El sector agroindustrial del departamento del Huila presenta una estructura organizacional que se puede considerar bastante diversa, esta diversificación incluye pequeños productores agrícolas organizados en cooperativas que fomentan el crecimiento de los agricultores, también incluye grandes empresas agroindustriales que operan a nivel nacional e internacional con estructura organizacional de la alta competitividad. Se puede considerar que pequeñas cooperativas juegan un rol fundamental en la recolección, procesamiento y comercialización de productos, mientras que las grandes empresas lideran la transformación de materias primas y la exportación de productos con valor agregado. La región se destaca por una fuerte cultura de asociatividad de trabajo colaborativo, lo que permite que pequeños productores puedan acceder a mercados más grandes y beneficiarse de economías de escala que tiene actualmente un mercado globalizado.

El departamento del huila con su fuerte agroindustria no solo se centra en la producción primaria, esta industria también trabaja día a día en la transformación de los alimentos valor agregado a su potencial económico. Las agroindustrias locales han invertido grandes sumas de dinero en tecnologías y procesos modernos para desarrollar técnicas que logren una mejor la calidad de los productos, para de esta forma reducir costos y aumentar la eficiencia operativa.

METODOLOGÍA DE LA INVESTIGACIÓN

Primer nivel

Enfoque, alcance y diseño de la investigación.

1.1. Enfoque de la investigación.

El enfoque de esta investigación es mixto, puesto que integra elementos cualitativos y cuantitativos entre las variables que se integrarán al modelo de machine learning con el que se investigarán los determinantes del éxito y la longevidad de las empresas de la cadena alimentaria en el departamento del Huila.

1.2. Alcance de la investigación.

La investigación se centrará en analizar cuáles son las variables críticas que tienen influencia sobre la supervivencia de empresas de la cadena alimentaria del departamento del Huila, con los datos disponibles hasta finales del año 2023.

1.3. Diseño de investigación.

En cuanto al tipo de diseño de la investigación, se podría establecer que será un tipo de estudio descriptivo, correlacional o aplicado, puesto que:

Descriptivo: Se describirán las características de las empresas y las variables relevantes en la cadena alimentaria, se identificarán patrones y tendencias sin buscar establecer relaciones causales directas.

Correlacional: Se identificarán las variables que podrían influir en la supervivencia o no de las empresas, se analizarán relaciones entre estas variables.

Aplicado: Dado que los hallazgos se podrían aplicar a intervenciones específicas en organizaciones de la cadena alimentaria. Los resultados se podrían utilizar para ofrecer recomendaciones prácticas a las empresas sobre cómo mejorar su supervivencia.

Definición de Variables.

La selección de variables tiene como base diferentes investigaciones que se han desarrollado previamente, y a través de las cuales se ha demostrado un impacto significativo en la sobrevivencia de las empresas (Román Ramírez J. , 2021) entre las que destacan: tamaño de la empresa, actividad empresarial, ingresos operacionales, utilidad operacional y nacionalidad empresarial. Entre las variables que harán parte del estudio se pueden mencionar las que se encuentran en la siguiente tabla:

Variable	Definición Conceptual	Definición Operacional	Dimensiones
Tamaño de la empresa	Clasificación de las empresas según su capacidad económica y operativa categoriza a las empresas en micro, pequeñas, medianas y grandes. Esta clasificación se basa en la generación de ingresos, el número de empleados y el valor total de sus activos, lo que permite identificar su rol y alcance en la economía (Ministerio de Comercio, Industria y Turismo, 2019)	Micro, pequeña, mediana o grande, determinada por rangos específicos en el número de empleados o el total de ingresos anuales.	Número de empleados, ingresos anuales, activos totales.
Actividad empresarial	Abarca las acciones que una empresa realiza para ofrecer bienes o servicios, determinadas por sus operaciones,	Clasifica según el sector económico de operación (como manufactura, servicios, o comercio), los tipos	Sector económico, Tipo de producto o servicio, Estrategias de mercado.

Variable	Definición Conceptual	Definición Operacional	Dimensiones
	sector y estrategias. Esta actividad varía según la industria y el tamaño de la empresa, influyendo en su competitividad. La clasificación de la actividad empresarial por sector económico, tipo de producto o servicio y estrategias de mercado ayuda a entender su modelo de negocio (Deskings, 2012)	de productos o servicios ofrecidos, y las principales estrategias de mercado empleadas.	
Ingresos operacionales	Representan el total de ingresos generados por una empresa a partir de sus actividades comerciales y operativas principales durante un periodo contable. Se cuantifican como la suma total de los ingresos netos recibidos por la venta de bienes y servicios en el curso normal de las operaciones de la empresa durante el último año fiscal (Almagtome, 2020).	Se cuantifica como la suma total de los ingresos netos recibidos por la venta de bienes y servicios en el curso normal de las operaciones de la empresa durante el último año fiscal.	Flujo de ingresos.
Utilidad operacional	La utilidad operacional refleja la eficiencia de una empresa para generar ganancias a partir de sus operaciones habituales antes de deducir intereses e impuestos. Esta eficiencia se mide	Se determina por el beneficio neto obtenido de las actividades operativas antes de intereses, impuestos, depreciación y amortización.	Rentabilidad operativa.

Variable	Definición Conceptual	Definición Operacional	Dimensiones
	por el beneficio neto de las actividades operativas antes de intereses, impuestos, depreciación y amortización (Aspilcueta Arias, 2020)		
Nacionalidad empresarial	Se determina por el país bajo cuya legislación está constituida y registrada, lo cual es clave para entender su identidad legal, incluyendo la jurisdicción aplicable y obligaciones fiscales. Esta clasificación se basa en el país de su incorporación legal y registro para operar, según se documenta en su certificado de incorporación o registro mercantil (Cafer Şafak Eysel, 2019)	La clasificación de la empresa basada en el país en el que está legalmente incorporada y registrada para hacer negocios, de acuerdo con su certificado de incorporación o registro mercantil.	País de constitución, Jurisdicción legal.
Fecha de Constitución	Indica el inicio formal de las operaciones de una empresa (Grennan, 2021)	Se registra como la fecha en la que la empresa se inscribe oficialmente ante la autoridad competente.	Fecha en la que la empresa se constituyó.
Fecha de Liquidación	Es el momento en que se lleva a cabo el proceso formal de disolución y liquidación de sus activos y pasivos. Esta fecha es crucial porque marca el inicio de la fase en la que la empresa	Se registra como la fecha en la que se realizan los trámites legales necesarios para cerrar la empresa oficialmente, lo que incluye la presentación de informes a las	Fecha en la que la empresa se liquidó.

Variable	Definición Conceptual	Definición Operacional	Dimensiones
	cesa sus operaciones comerciales y comienza a pagar sus deudas, distribuyendo cualquier activo remanente a los accionistas o propietarios.	autoridades fiscales y comerciales.	
Antigüedad de la empresa	Se refiere al tiempo que ha transcurrido desde su constitución formal hasta la fecha actual.	Se resta la fecha de constitución de la empresa de la fecha actual.	Años que la empresa lleva activa en el mercado.
Ingresos Totales	Monto total de dinero que una empresa recibe por la venta de bienes o servicios durante un período específico, sin deducir ningún costo o gasto. Es una medida clave de la actividad económica de una empresa y es fundamental para evaluar su desempeño financiero.	Los ingresos totales representan la principal fuente de financiamiento para las operaciones de una empresa.	Monto total de dinero que una empresa recibió.
Utilidad Neta	Representa el beneficio total de una empresa después de deducir todos los costos y gastos, incluidos impuestos e intereses. Es un indicador clave de la rentabilidad de la empresa y se utiliza para evaluar su desempeño financiero.	La utilidad neta muestra cuánto dinero gana una empresa después de cubrir todos sus gastos, lo que es crucial para evaluar su eficiencia operativa.	Cantidad de dinero que tiene la empresa después de deducir todos los costos y gastos.
Pasivo Total	Representa todas las obligaciones	Se calcula sumando todas las	Endeudamiento a Corto Plazo,

Variable	Definición Conceptual	Definición Operacional	Dimensiones
	<p>financieras que una empresa debe a terceros. Esta cifra es un indicador crítico del nivel de endeudamiento y se utiliza para evaluar la estructura de capital y la solidez financiera de una empresa. El Pasivo Total incluye tanto las obligaciones a corto plazo como el endeudamiento a largo plazo y es relevante para los inversores y analistas que buscan comprender la capacidad de una empresa para generar valor y su sostenibilidad financiera a largo plazo (Barreto Granda, 2020)</p>	<p>obligaciones actuales y futuras reflejadas en el balance de la empresa, incluyendo deudas a corto y largo plazo, cuentas por pagar y otras obligaciones financieras.</p>	<p>Endeudamiento a Largo Plazo.</p>
<p>Margen Operativo</p>	<p>Indicador financiero que mide la eficiencia de una empresa en la generación de ganancias a partir de sus operaciones principales.</p>	<p>Se calcula como el porcentaje de los ingresos totales que queda después de deducir todos los costos y gastos operativos, excluyendo impuestos e intereses. Este margen proporciona una visión clara de la rentabilidad de las operaciones básicas de una empresa sin considerar los efectos de la estructura de capital y las políticas fiscales.</p>	<p>Porcentaje de ingresos luego de deducción de costos y gastos operativos.</p>

Variable	Definición Conceptual	Definición Operacional	Dimensiones
EBITDA	Indicador financiero que mide el rendimiento operativo de una empresa, excluyendo los efectos de decisiones de financiamiento, impuestos y costos no monetarios, como la depreciación y la amortización.		Cantidad de dinero luego de decir costos y gastos operacionales.

Población y Muestra

La población objetivo de esta investigación está constituida por las empresas de la cadena alimentaria en el departamento del Huila, que comprende un total de 2064 empresarios. La muestra se elegirá de manera aleatoria, considerando la variedad de empresas en cuanto a tamaño, tipo de actividad y ubicación geográfica, utilizando como base la información proporcionada por la Cámara de Comercio del Huila. Para calcular el tamaño de la muestra, se empleó la herramienta Netquest, teniendo en cuenta las dimensiones de las empresas, con un nivel de confianza del 95% y un margen de error del 5%. Como resultado, se obtuvieron 325 empresarios en la muestra, garantizando así la representatividad de los datos.

Segundo Nivel

Selección de Métodos o Instrumentos para la Recolección de Información

En este estudio descriptivo-correlacional, la fuente de información principal será la base de datos proporcionada por la Cámara de Comercio del Huila, que contiene más de 20,000 registros empresariales del sector alimentario en el departamento. Esta base de datos incluye variables

clave que han sido recopiladas de manera sistemática por la entidad, garantizando altos estándares de fiabilidad y validez. Estas características hacen que esta fuente sea un instrumento adecuado para el análisis de la supervivencia empresarial, asegurando la consistencia en la medición de las variables objeto de estudio.

Características del Instrumento Utilizado: Base de Datos de la Cámara de Comercio del Huila

La base de datos proporcionada por la Cámara de Comercio del Huila actúa como una fuente secundaria de gran valor para el análisis de variables fundamentales relacionadas con la supervivencia empresarial. Entre los datos que ofrece se encuentran registros sobre el tamaño de las empresas, ingresos operativos, antigüedad, sector económico, utilidad operativa, y otras variables económicas y financieras relevantes.

Debido a su naturaleza longitudinal, esta base de datos ha sido empleada en estudios previos sobre análisis empresarial, destacando por su capacidad de rastrear la evolución de las empresas a lo largo del tiempo, lo que facilita un enfoque integral para la comprensión de su comportamiento (Cámara de Comercio del Huila., 2023). Este tipo de información es especialmente útil en estudios descriptivos y correlacionales, ya que proporciona una visión detallada y cronológica del desarrollo de las empresas en el sector alimentario.

La riqueza y extensión de estos datos permite un análisis exhaustivo del estado y evolución de las empresas del sector alimentario en el Huila, utilizando técnicas avanzadas como el análisis de supervivencia empresarial y modelos de predicción. Estas herramientas no solo permiten estudiar los factores asociados a la continuidad o desaparición de las empresas, sino que también contribuyen a generar predicciones sobre su sostenibilidad futura, ofreciendo una base

sólida para la toma de decisiones estratégicas y políticas públicas orientadas al fortalecimiento del sector.

Técnicas de Análisis de Datos

Para el análisis de los datos extraídos de la base de datos, se utilizarán técnicas cuantitativas orientadas a identificar las relaciones entre las variables y a predecir la supervivencia de las empresas.

Análisis Cuantitativo

Estadística Descriptiva

La primera fase del análisis consistirá en describir las características de las empresas del sector alimentario mediante técnicas de estadística descriptiva. Se calcularán medidas de tendencia central (media, mediana) y dispersión (desviación estándar) para las variables como ingresos operativos, antigüedad y tamaño de la empresa. Estos análisis ayudarán a identificar patrones y a obtener una visión general de los datos proporcionados (Hernández Julio, 2024).

Regresión de Cox para el Análisis de Supervivencia

Para determinar los factores que influyen en la supervivencia empresarial, se aplicará un modelo de Regresión de Cox, una técnica comúnmente utilizada en estudios de análisis de supervivencia. Este modelo permitirá identificar qué variables (como tamaño de la empresa, ingresos operativos o sector económico) tienen un mayor impacto en la probabilidad de que una empresa continúe operando a lo largo del tiempo (Román Ramírez J. , 2021).

Modelos Predictivos: Random Forest y Árboles de Decisión

Como parte del análisis predictivo, se implementarán algoritmos de Random Forest y Árboles de Decisión, técnicas de machine learning que son útiles para identificar relaciones complejas

entre las variables. Estos modelos permitirán predecir la probabilidad de supervivencia de las empresas en función de las características identificadas en la base de datos de la Cámara de Comercio (Afolabi, 2019). Estos algoritmos han demostrado ser altamente eficaces en la predicción de la longevidad empresarial, particularmente en estudios con grandes volúmenes de datos.

Validación de Modelos

Con base a la validación de modelos predictivos como paso fundamental para evaluar su rendimiento y asegurarse de que proporcionan predicciones confiables, específicamente para este estudio, los modelos predictivos serán validados utilizando técnicas como la validación cruzada y el bootstrapping, lo que garantiza que los resultados obtenidos sean robustos y generalizables. Estas técnicas permiten evaluar la precisión de los modelos y asegurar que no haya sobreajuste, asegurando que las predicciones se puedan aplicar a diferentes contextos (Hastie T. T., 2021).

Justificación de las Técnicas

El uso de técnicas como la Regresión de Cox y modelos de machine learning como Random Forest y Árboles de Decisión está plenamente justificado por su capacidad para manejar grandes volúmenes de datos y su efectividad en la identificación de factores críticos de supervivencia empresarial. Estas técnicas han sido utilizadas en estudios previos con datos similares para predecir la longevidad de las empresas (Román Ramírez J. , 2021), (García G. P., 2022).

ANÁLISIS Y DISCUSIÓN DE LOS RESULTADOS

En este capítulo, se presentan los resultados del análisis exploratorio de datos (EDA) y el desarrollo de los modelos predictivos para estimar la rentabilidad de empresas en la cadena alimentaria del Huila. A partir de estos análisis, se extraen conclusiones basadas en los factores que más influyen en la rentabilidad empresarial en esta región.

Análisis Exploratorio de Datos (EDA)

Al realizar el Análisis Exploratorio de Datos de desarrollo podemos comprender la distribución de los datos, identificar patrones, detectar valores atípicos y evaluar las relaciones entre variables de la base de datos escogida para este estudio. Con base a los resultados obtenidos, se observa lo siguiente:

Al analizar la base de datos, que tiene un total de 17 columnas y 26,129 entradas, podemos hacer algunas conclusiones sobre cada variable, considerando su tipo de dato, su propósito y su relevancia potencial en el análisis. A continuación, se presenta una interpretación de cada una de las variables con base al análisis exploratorio de datos:

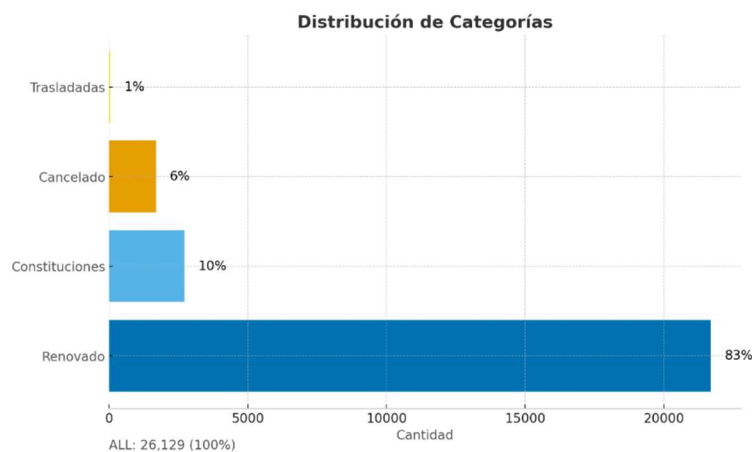
Variable 1: Estado de Matrícula

La variable Estado de Matrícula es categórica y refleja el estado de registro de las empresas en el sistema. La distribución de esta variable muestra una clara predominancia de la categoría "Renovado", que representa el 83% del total de registros (21,688 empresas), seguida de "Constituciones" con un 10% (2,711 registros). Las categorías "Cancelado" y "Trasladadas" tienen una representación mucho menor, con el 6% (1,685 registros) y menos del 1% (45 registros), respectivamente. Esta distribución indica que la mayoría de las empresas mantienen su registro activo, mientras que una minoría ha cancelado o trasladado su registro. En términos de asociaciones, "Estado de Matrícula" presenta las relaciones más altas con las variables

"Período" y "Año", ambas con un coeficiente de incertidumbre de 1, lo que indica que no hay variabilidad temporal en estos datos. Las asociaciones con otras variables, como "Tamaño" (0.05) y "Organización Jurídica" (0.02), son bajas, al igual que las correlaciones con variables numéricas como "Total Valor Patrimonio" y "Total Valor Activos" (ambas 0.01), lo cual sugiere una relación limitada entre el estado de matrícula y estas características.

Figura 1.

Estado Matrícula



Variable 2: Período

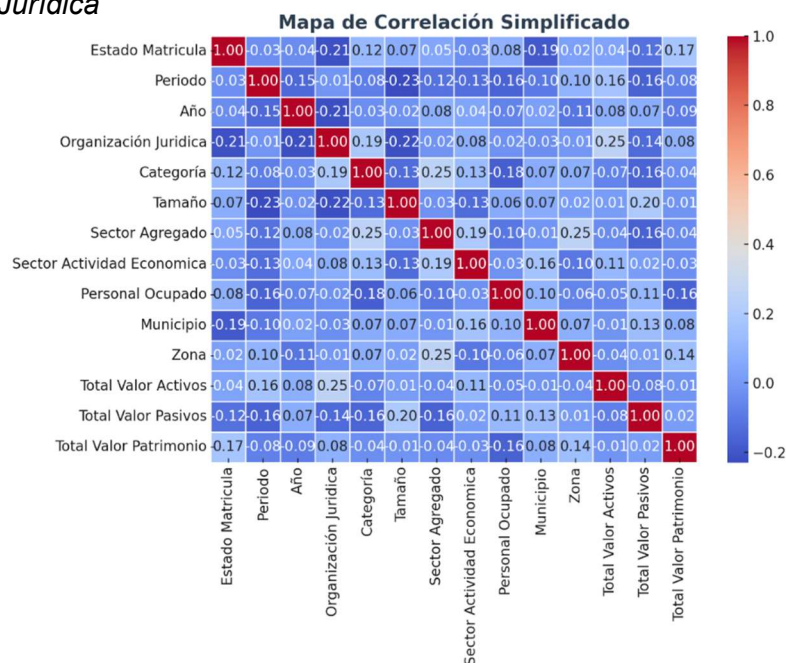
La variable Período es categórica y presenta un único valor, "1er Trimestre", que abarca el 100% de los 26,129 registros en el dataset. Esto significa que no existe variabilidad en los datos para esta variable, ya que todos los registros pertenecen al mismo período temporal. Este comportamiento convierte a "Período" en una variable constante en esta muestra, sin valores faltantes ni diferencias entre registros. Las métricas de longitud muestran una longitud constante de 13 caracteres para cada valor, y no se presentan valores únicos adicionales. Las asociaciones de "Período" con otras variables, como "Estado de Matrícula", "Año" y "Organización Jurídica", indican una correlación perfecta (1.0), lo que es esperable para una variable constante, pero estas asociaciones no aportan información diferencial.

Variable 3: Organización Jurídica

La variable Organización Jurídica es categórica y presenta 8 categorías distintas. La categoría predominante es "Persona Natural", que abarca el 84% de los registros (21,968 empresas), seguida por "Sociedades por Acciones Simplificadas" con un 12% (3,174 empresas). Las demás categorías, como "Sociedad Limitada" (2%), "Sociedad Anónima" (1%), y otras estructuras jurídicas, representan menos del 1% cada una. No hay valores faltantes en esta variable. En términos de asociaciones, "Organización Jurídica" presenta una alta correlación (0.81) con la variable "Categoría", lo que sugiere que ambas variables podrían estar relacionadas en la caracterización de las empresas. Además, muestra asociaciones moderadas con "Sector Actividad Económica" (0.23) y "Tamaño" (0.13), lo cual indica que la estructura jurídica podría estar parcialmente vinculada con la actividad económica y el tamaño de las empresas. Entre las correlaciones con variables numéricas, "Organización Jurídica" tiene una relación baja con "Total Valor Patrimonio" (0.13), "Total Valor Activos" (0.12), y "Total Valor Pasivos" (0.11), sugiriendo que el tipo de estructura jurídica tiene una ligera influencia sobre el valor financiero de las empresas.

Figura 2.

Organización Jurídica

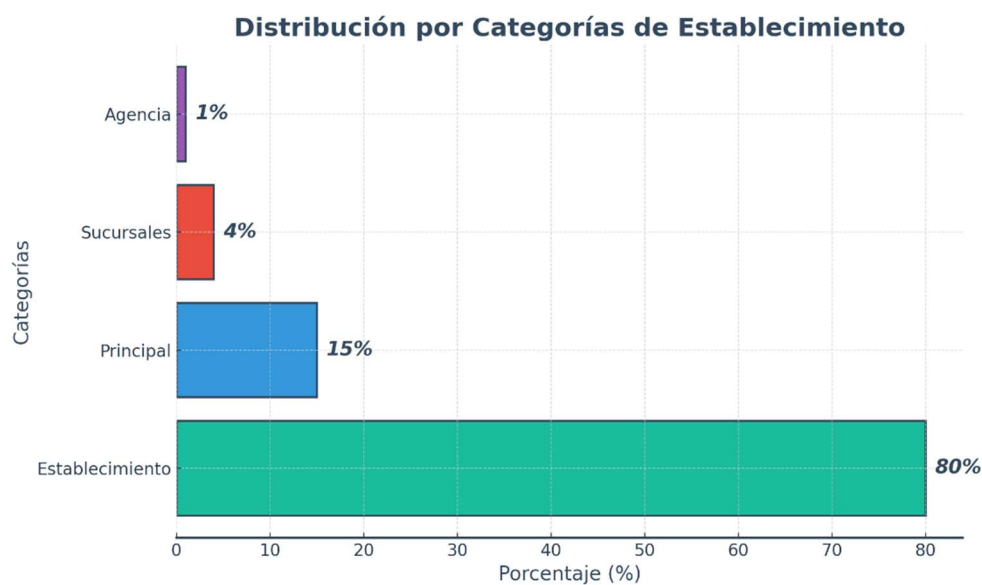


Variable 4: Categoría

La variable Categoría es categórica y cuenta con cuatro valores distintos: "Establecimiento" (84% de los registros), "Principal" (14%), "Sucursales" (2%) y "Agencia" (menos del 1%). La predominancia de "Establecimiento" refleja que la mayoría de los registros pertenecen a empresas que funcionan en un único lugar o sede principal. No presenta valores faltantes. En términos de asociaciones, "Categoría" muestra una correlación alta con "Organización Jurídica" (0.90) y una relación moderada con "Tamaño" (0.36), lo que sugiere que la categoría de la empresa podría estar vinculada con su estructura jurídica y su tamaño. También se observa una asociación más baja con variables financieras, como "Total Valor Patrimonio" (0.08) y "Total Valor Activos" (0.07), lo cual indica que la categoría de la empresa podría tener una leve influencia en sus características financieras.

Figura 3.

Categoría



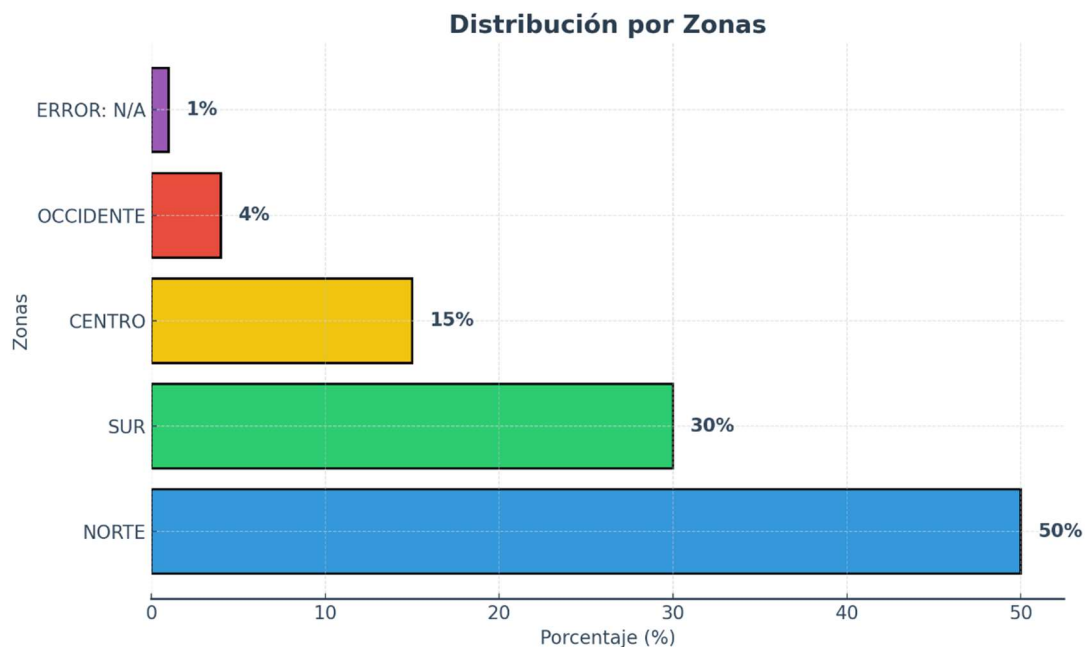
Variable 5: Zona

Este análisis sugiere que la mayoría de los datos se encuentran en la zona Norte, lo que podría indicar que es la región más representativa o con mayor actividad en el contexto laboral del Huila.

Las diferencias en la representación entre zonas pueden ser relevantes para entender la distribución geográfica del trabajo y otros factores económicos o sociales que pueden influir en cada región. Es importante considerar cómo estas diferencias entre zonas pueden impactar los resultados del análisis de la cadena alimentaria en el Huila.

Figura 4.

Zona

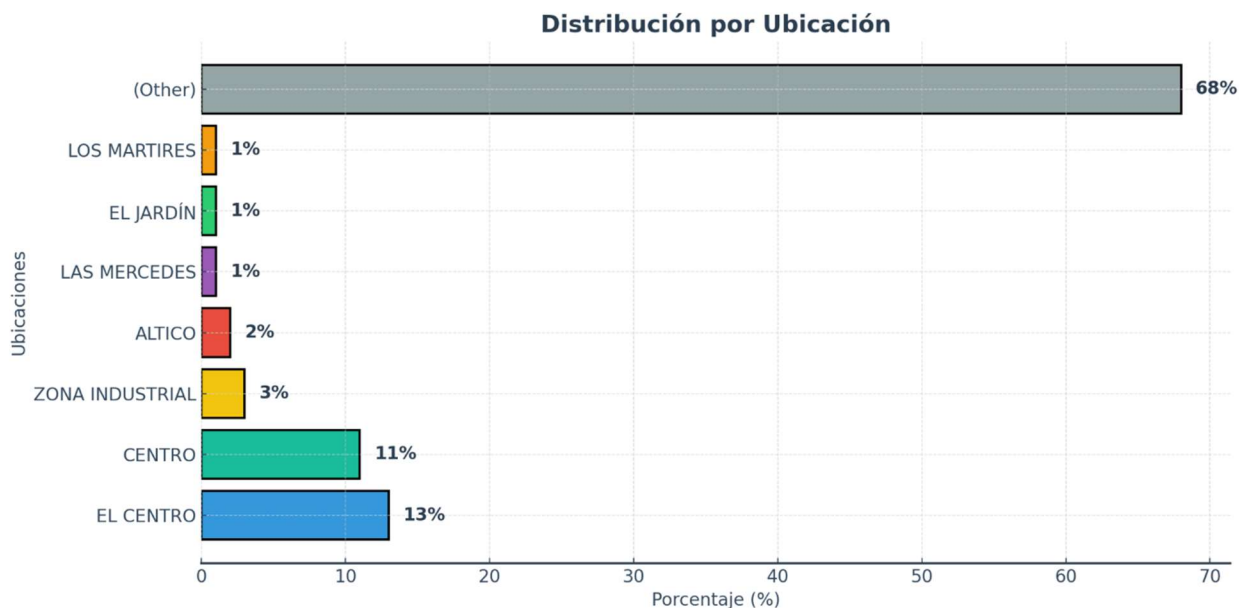


Variable 6: Barrio

Este Análisis muestra una categorización de los datos, donde los barrios "El Centro" y "Centro" representan los porcentajes más altos, con un 13% y un 11% respectivamente. Este desglose refleja una concentración de datos significativa en estas áreas, mientras que otros barrios, como "Zona Industrial" y "Altico", cuentan con menores representaciones (3% y 2% respectivamente)

Figura 5.

Barrio



Variable 7: Total Valor Activos

La gráfica correspondiente al "Total Valor Activos" revela una amplia distribución en los valores. El rango entre los valores mínimo y máximo es de 915.1 mil millones, con una media y una desviación estándar significativas, lo que sugiere una alta variabilidad en los activos. Esto es indicativo de una distribución sesgada con una asimetría alta (87.3) y curtosis elevada (8,560), que podrían implicar la presencia de datos extremos o outliers

Figura 6.

Total Valor Activos



Variable 8: Total Valor Pasivos

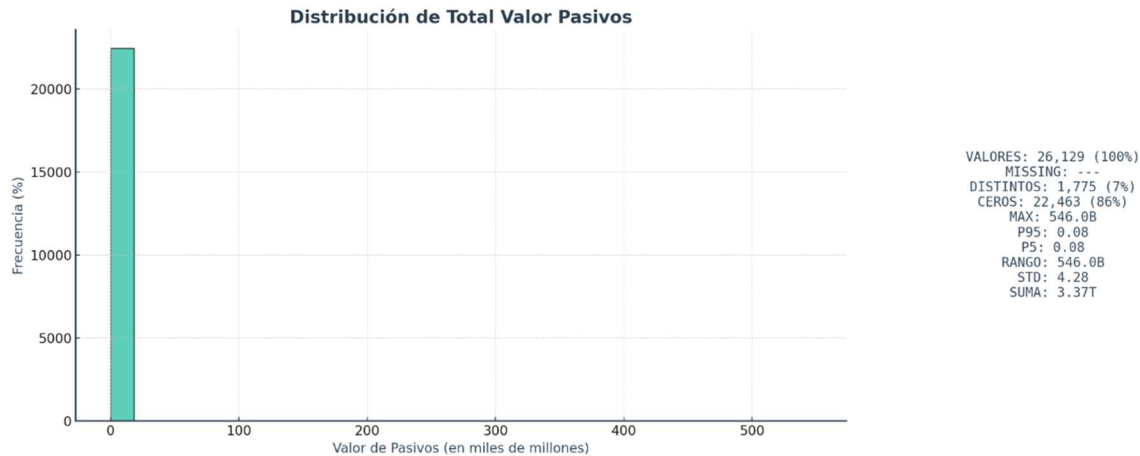
La mayoría de las empresas tienen pasivos bajos o inexistentes, mientras que unas pocas concentran grandes cantidades de deuda o pasivos. Esto sugiere una estructura en la que solo ciertas empresas tienen obligaciones financieras significativas.

La alta dispersión y asimetría reflejan la desigualdad en el nivel de endeudamiento, lo cual podría implicar que sólo unas pocas empresas están asumiendo mayores riesgos financieros o cuentan con un apalancamiento considerable.

Esta distribución indica una concentración significativa en los niveles de endeudamiento dentro del sector empresarial del Huila, con solo algunas empresas que tienen pasivos elevados en comparación con la mayoría, que tienden a operar con pocos o nulos pasivos.

Figura 7.

Total Valor Pasivos



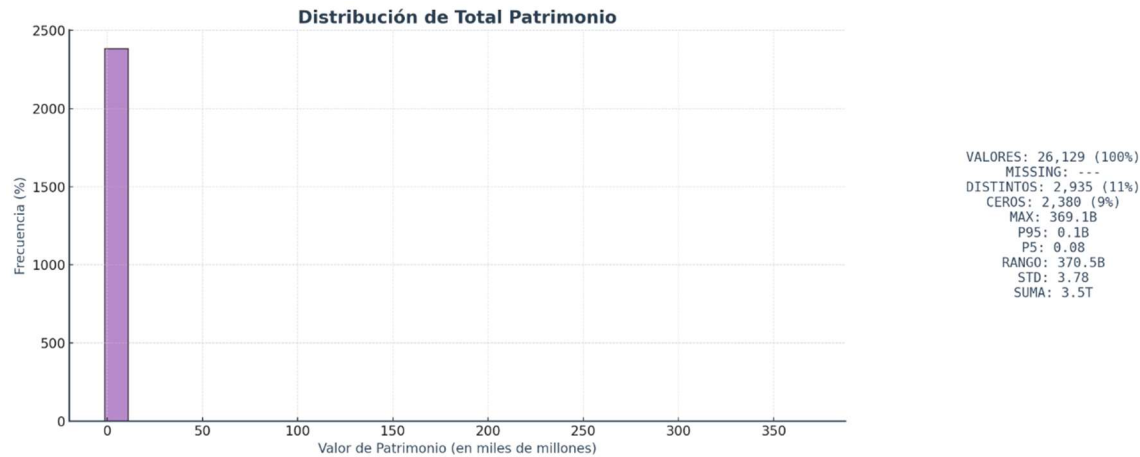
Variable 9: Total Patrimonio

La gráfica de "Total Valor Patrimonio" La mayoría de las observaciones tienen valores de patrimonio bajos o nulos, con una minoría de empresas que concentran valores de patrimonio muy elevados.

La alta asimetría y curtosis en la distribución de la variable reflejan una gran desigualdad en el patrimonio de las empresas, lo que sugiere que solo algunas empresas tienen una acumulación significativa de recursos propios. Este análisis apunta a una estructura empresarial en la que solo un pequeño porcentaje de empresas tiene altos niveles de patrimonio, mientras que la mayoría tiene recursos propios limitados o nulos. Esto podría ser indicativo de una concentración económica en ciertas empresas dentro del Huila.

Figura 8.

Total Patrimonio



Variable 10: Total Ventas Netas

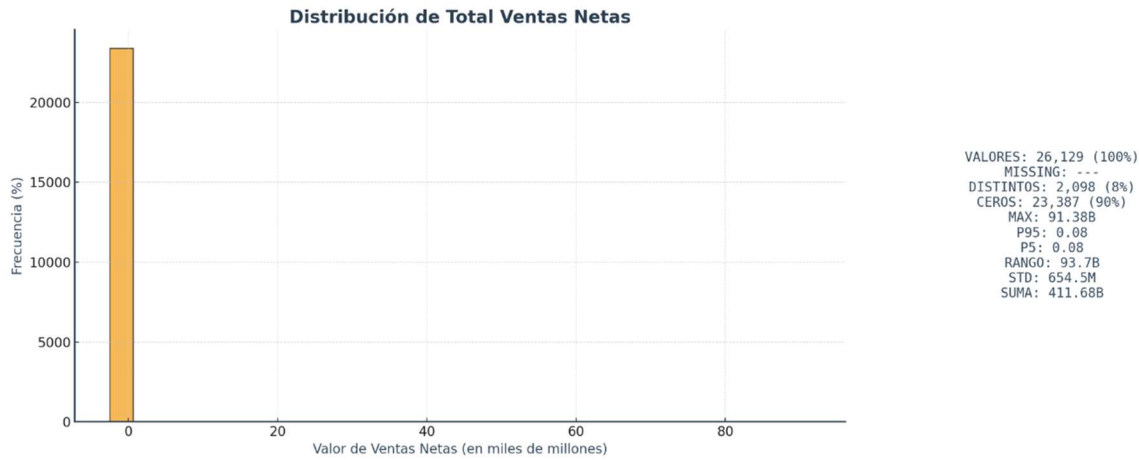
La mayoría de las observaciones tienen ventas netas en cero, lo que podría ser el resultado de empresas inactivas en ventas o falta de reporte.

Existen valores extremadamente altos en la distribución, que elevan considerablemente el promedio y la dispersión, sugiriendo que sólo una minoría de las empresas o sectores tienen ventas netas significativas.

La distribución altamente sesgada indica una gran desigualdad en el volumen de ventas netas entre las observaciones, con algunos pocos casos de ventas muy altas en comparación con el resto.

Figura 9.

Total Ventas Netas

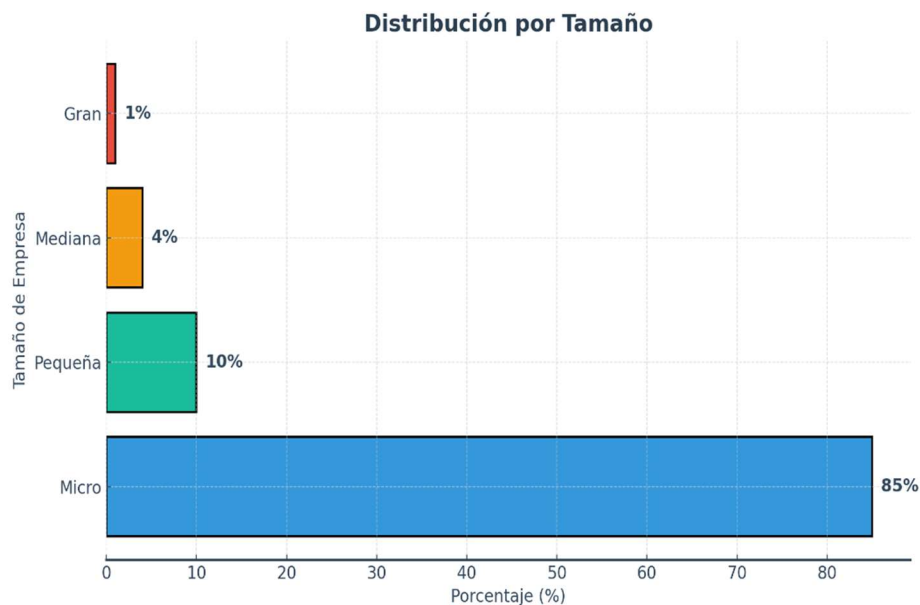


Variable 11: Tamaño

Hace referencia al tamaño de las empresas; tal y como se observa en la figura, el 97% de las empresas están categorizadas como microempresas, seguidas de las pequeñas empresas (2%), mediana y grandes empresas, éstas dos últimas representan menos del 1%, cada una, de la información en la base de datos.

Figura 10.

Tamaño

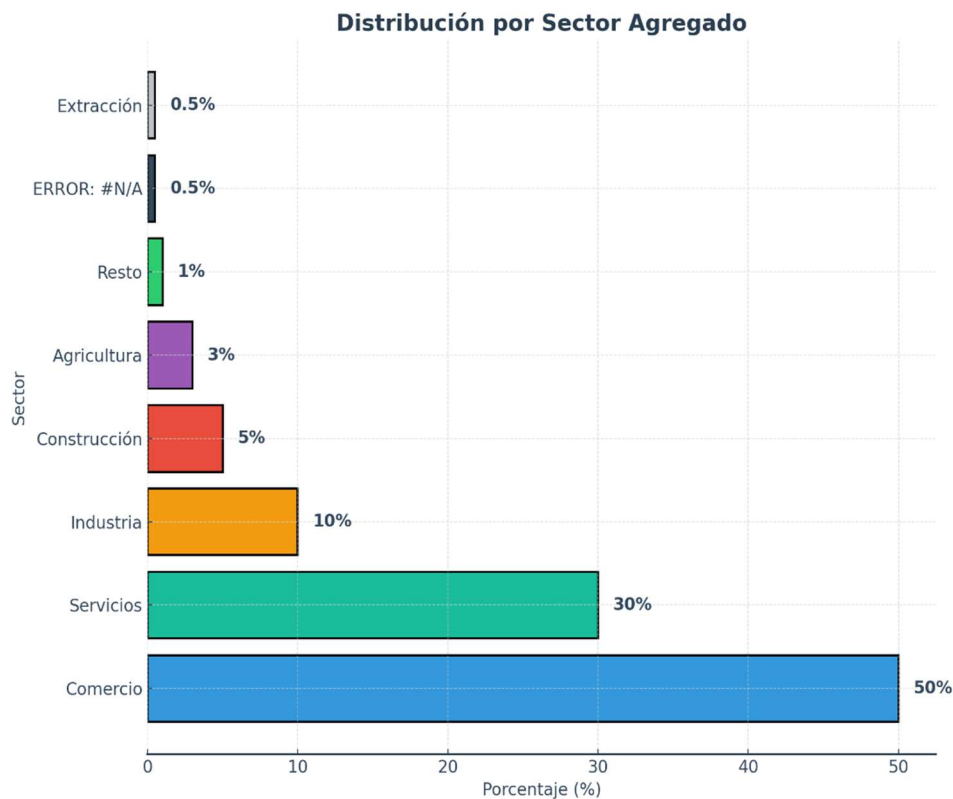


Variable 12: Sector agregado

Hace referencia al sector en el que las empresas actúan. Como se observa en la figura, el sector en el que más empresas, dentro de la base de datos, tienen presencia es el comercio, con el 51% de la información, seguido de servicios (33%), industria (8%), construcción (3%), agricultura (2%), resto de sectores (2%), y extracción y ERROR: #N/A (empresas que no tienen ningún valor registrado para esta columna, es decir, esta vacía la casilla), cada una, con menos del 1% de la información.

Figura 11.

Sector agregado

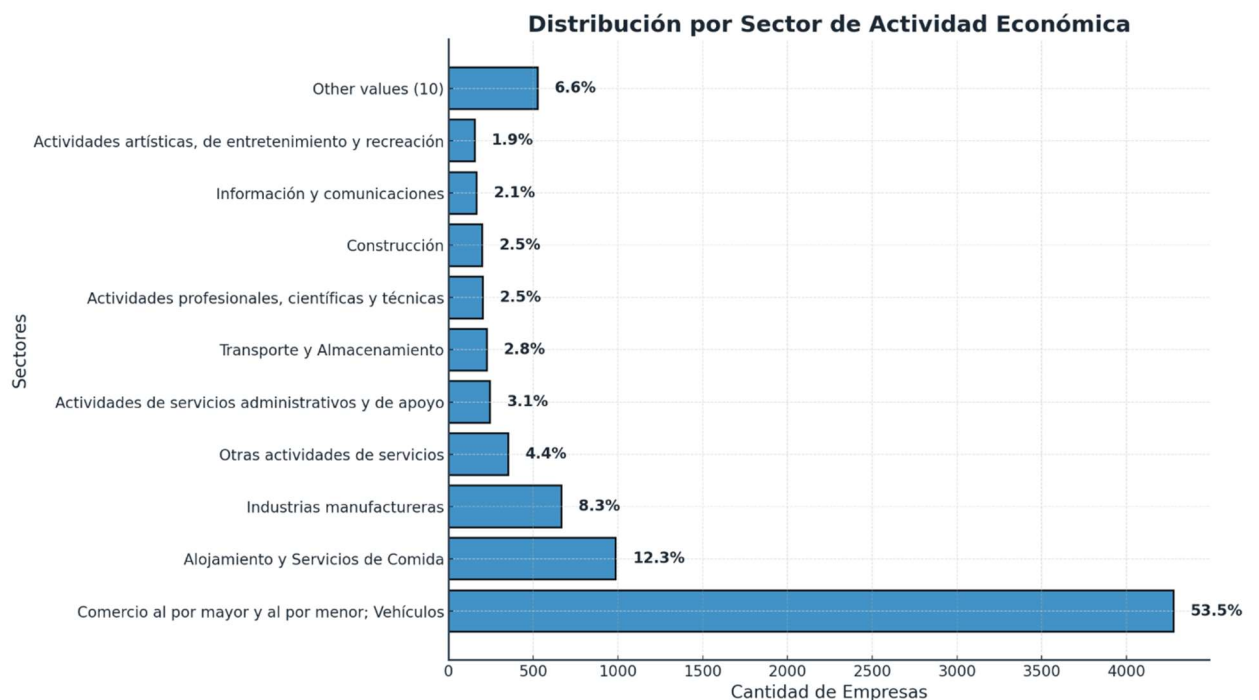


Variable 13: Sector Actividad Económica

Hace referencia al sector en el que actúa la empresa de forma desagregada. Como se observa en la figura, el 51% de las empresas se dedican al Comercio al por mayor y al por menor de vehículos; seguido de Alojamiento y Servicios de Comida (13%), Industrias manufactureras (8%) y Otras actividades de servicios (4%), entre otras.

Figura 12.

Sector Actividad Económica



Variable 14: Actividad Económica Principal

Hace referencia a la actividad económica de la empresa. La actividad económica en la que se registra más información es “Comercio al por menor en establecimientos no especializados con surtido compuesto principalmente por alimentos bebidas” con el 10%, seguido de “Otros” (21%), “Expendio a la mesa de comidas preparadas” (4%) y “Comercio al por menor de productos farmacéuticos y medicinales cosméticos y artículos de tocador en establecimientos especiales” (4%), tal y como se observa en la figura.

Figura 13.

Categorías actividad económica principal de las empresas.

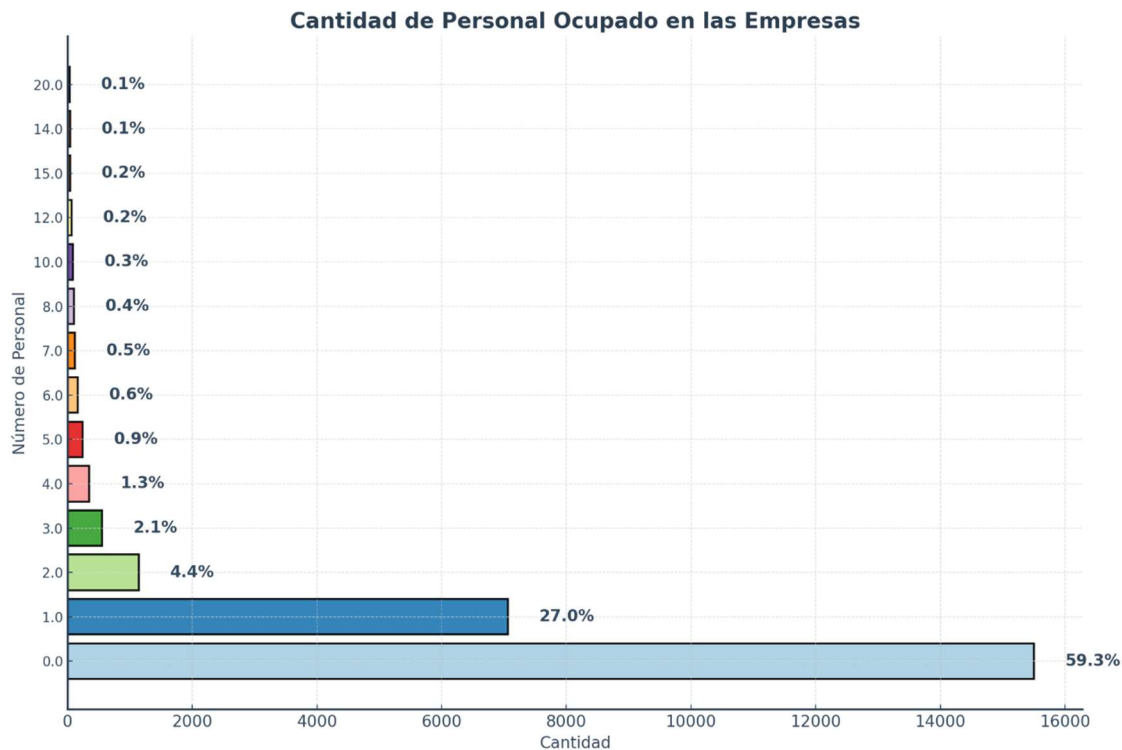
2,585	10%	G4711 ** Comercio al por menor en establecimientos no especializados con surtido compuesto principalmente por alimentos bebidas (alco...
1,095	4%	I5611 ** Expendio a la mesa de comidas preparadas
1,052	4%	G4773 ** Comercio al por menor de productos farmaceuticos y medicinales cosméticos y artículos de tocador en establecimientos especial...
955	4%	G4771 ** Comercio al por menor de prendas de vestir y sus accesorios (incluye artículos de piel) en establecimientos especializados
952	4%	G4719 ** Comercio al por menor en establecimientos no especializados con surtido compuesto principalmente por productos diferentes de ...
748	3%	I5630 ** Expendio de bebidas alcoholicas para el consumo dentro del establecimiento
704	3%	G4752 ** Comercio al por menor de artículos de ferreteria pinturas y productos de vidrio en establecimientos especializados
679	3%	S9602 ** Peluqueria y otros tratamientos de belleza
494	2%	G4520 ** Mantenimiento y reparacion de vehiculos automotores
478	2%	I5619 ** Otros tipos de expendio de comidas preparadas n.c.p.
449	2%	C1081 ** Elaboracion de productos de panaderia
436	2%	G4723 ** Comercio al por menor de carnes (incluye aves de corral) productos carnicos pescados y productos de mar en establecimientos e...
433	2%	G4761 ** Comercio al por menor de libros periodicos materiales y artículos de papeleria y escritorio en establecimientos especializados
409	2%	G4759 ** Comercio al por menor de otros artículos domesticos en establecimientos especializados
389	1%	G4724 ** Comercio al por menor de bebidas y productos del tabaco en establecimientos especializados
387	1%	G4541 ** Comercio de motocicletas y de sus partes piezas y accesorios
366	1%	G4620 ** Comercio al por mayor de materias primas agropecuarias animales vivos
364	1%	R9329 ** Otras actividades recreativas y de esparcimiento n.c.p.
354	1%	I5613 ** Expendio de comidas preparadas en cafeterias
350	1%	I5511 ** Alojamiento en hoteles
323	1%	G4530 ** Comercio de partes piezas (autopartes) y accesorios (lujos) para vehiculos automotores
321	1%	G4721 ** Comercio al por menor de productos agricolas para el consumo en establecimientos especializados
314	1%	G4741 ** Comercio al por menor de computadores equipos perifericos programas de informatica y equipos de telecomunicaciones en estab...
303	1%	F4290 ** Construccion de otras obras de ingenieria civil
290	1%	N8299 ** Otras actividades de servicio de apoyo a las empresas n.c.p.
282	1%	C1410 ** Confeccion de prendas de vestir excepto prendas de piel
275	1%	J6190 ** Otras actividades de telecomunicaciones
268	1%	G4772 ** Comercio al por menor de todo tipo de calzado y artículos de cuero y sucedaneos del cuero en establecimientos especializados
264	1%	G4729 ** Comercio al por menor de otros productos alimenticios n.c.p. en establecimientos especializados
246	<1%	H4923 ** Transporte de carga por carretera
231	<1%	G4799 ** Otros tipos de comercio al por menor no realizado en establecimientos puestos de venta o mercados
230	<1%	H5221 ** Actividades de estaciones vias y servicios complementarios para el transporte terrestre
206	<1%	G4754 ** Comercio al por menor de electrodomesticos y gasodomicos muebles y equipos de iluminacion en establecimientos especi...
198	<1%	M7310 ** Publicidad
189	<1%	G4774 ** Comercio al por menor de otros productos nuevos en establecimientos especializados
176	<1%	C3312 ** Mantenimiento y reparacion especializado de maquinaria y equipo
171	<1%	G4542 ** Mantenimiento y reparacion de motocicletas y de sus partes y piezas
170	<1%	G4755 ** Comercio al por menor de artículos y utensillos de uso domesticos en establecimientos especializados
166	<1%	K6621 ** Actividades de agentes y corredores de seguros
157	<1%	C2511 ** Fabricacion de productos metalicos para uso estructural
147	<1%	G4631 ** Comercio al por mayor de productos alimenticios
144	<1%	L6810 ** Actividades Inmobiliarias realizadas con bienes propios o arrendados
134	<1%	A0123 ** Cultivo de cafe
131	<1%	M7020 ** Actividades de consultoria de gestion
130	<1%	A0322 ** Acuicultura de agua dulce
127	<1%	Q8621 ** Actividades de la practica medica sin internacion
123	<1%	G4663 ** Comercio al por mayor de materiales de construccion artículos de ferreteria pinturas productos de vidrio equipo y materiales de fo...
122	<1%	J6120 ** Actividades de telecomunicaciones inalambricas
122	<1%	I5514 ** Alojamiento rural
114	<1%	K6412 ** Bancos comerciales
110	<1%	F4111 ** Construccion de edificios residenciales
104	<1%	I5519 ** Otros tipos de alojamiento para visitantes
103	<1%	M7112 ** Actividades de ingenieria y otras actividades conexas de consultoria técnica

Variable 15: Personal ocupado

Hace referencia a la cantidad de personal que cada empresa tiene trabajando. Como se observa en la figura, el 59,3% de las empresas no tiene personal ocupado, seguido de 1 trabajador (27,0%), 2 trabajadores (4,4%) y 3 trabajadores (2,1%).

Figura 14.

Cantidad de personal ocupado en las empresas.

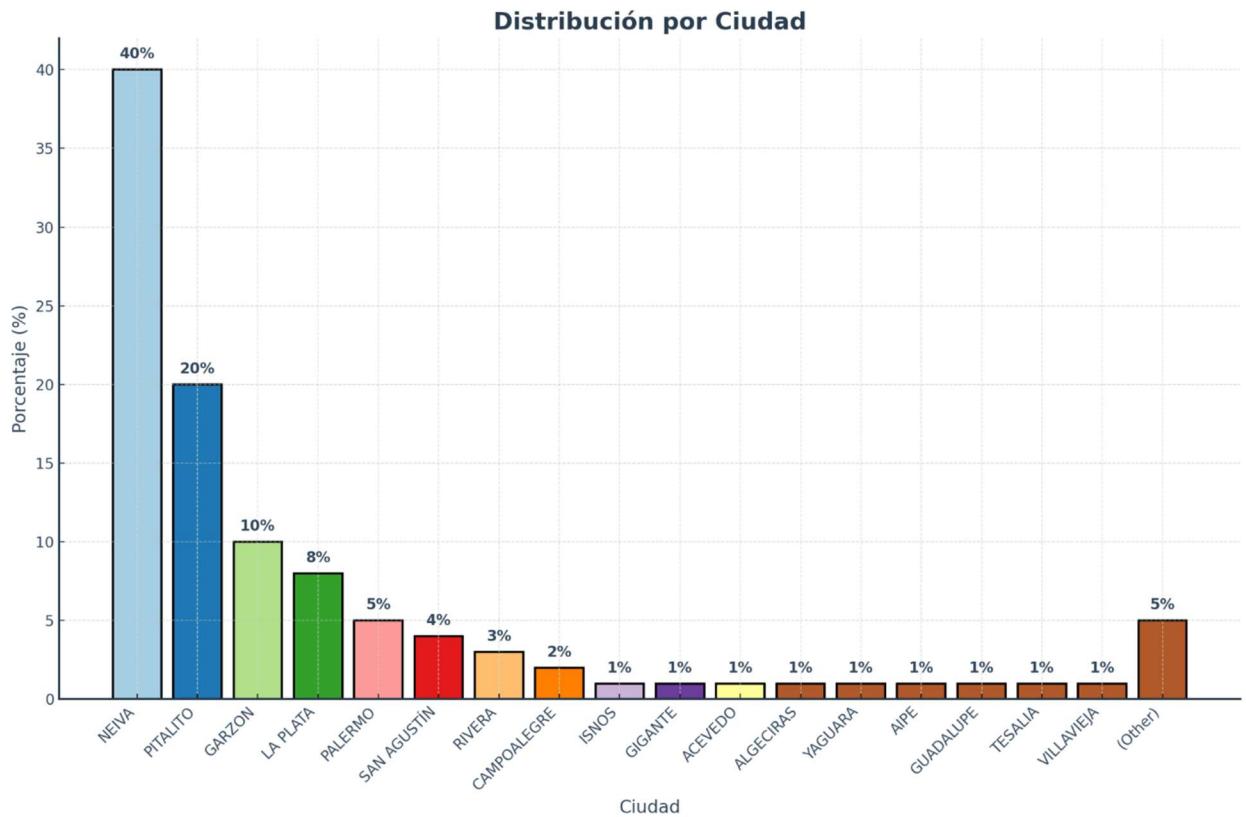


Variable 16: Municipio

Hace referencia al municipio en el cual se encuentra radicada la empresa. Como se observa en la figura, el 41% de las empresas se encuentra en Neiva, seguida de Pitalito (16%), Otros (11%) y Garzón (7%).

Figura 15.

Municipio



La base de datos incluye una variedad de variables muestran las organizaciones registradas y cuyas variables financieras (activos, pasivos, patrimonio y ventas) son especialmente importantes para evaluar la salud financiera de las organizaciones, mientras que las variables categóricas (como el tipo de organización, sector y ubicación) permiten segmentar el análisis y proporcionan un contexto para los resultados financieros correlacionados con la variable objetivo.

Modelado Predictivo

Para la selección de los modelos se escogieron los modelos Random Forest, Gradient Boosting y Regresión Lineal debido a sus altas capacidades de predicción de variables

financieras y debido a que los tres son complementarios entre sí. A continuación, se presentan sus características principales:

La regresión lineal es un modelo simple y ampliamente utilizado en análisis predictivos, especialmente cuando se busca comprender la relación lineal entre variables. Al ser un modelo básico, sirve como un punto de referencia inicial para evaluar el desempeño y la capacidad predictiva frente a otros modelos más complejos. En este caso, se utilizó para ver si la relación entre las variables financieras y la rentabilidad podía modelarse de manera efectiva mediante una fórmula lineal.

El modelo Random Forest E es un ensamble de árboles de decisión que permite capturar relaciones no lineales y complejas entre las variables. Además, este modelo es robusto frente a overfitting, maneja bien las interacciones entre las variables y ofrece un alto rendimiento sin necesidad de una gran cantidad de ajuste de hiperparámetros. Su habilidad para manejar grandes volúmenes de datos y extraer patrones relevantes lo hace adecuado para este tipo de análisis, donde las interacciones entre las variables financieras no son estrictamente lineales.

Gradient Boosting es otro modelo basado en árboles de decisión, pero con un enfoque de aprendizaje secuencial, donde cada árbol corrige los errores del anterior. Este enfoque le permite mejorar el rendimiento predictivo, especialmente cuando se trabaja con datos complejos y no lineales. Gradient Boosting es conocido por su alta precisión y ha demostrado ser efectivo en problemas de predicción, aunque es más sensible al overfitting si no se ajustan bien los parámetros. Se incluyó para comparar su rendimiento con el de Random Forest, evaluando si la corrección secuencial de errores mejora la capacidad predictiva en comparación con el enfoque de ensamble más general de Random Forest.

Modelos Correlación

Teniendo en cuenta que la variable objetivo no se encuentra definida , se decidió definir una variable que permita relacionar las variables financieras. En primera instancia, se hizo una selección y preprocesamiento de las variables críticas y se definió una variable llamada “Rentabilidad” considerando las variables financieras obtenidas de la base de datos “Total Valor Ventas Netas”, “Total Valor Pasivos” y “Total Valor Activos”, a través de la siguiente relación matemática:

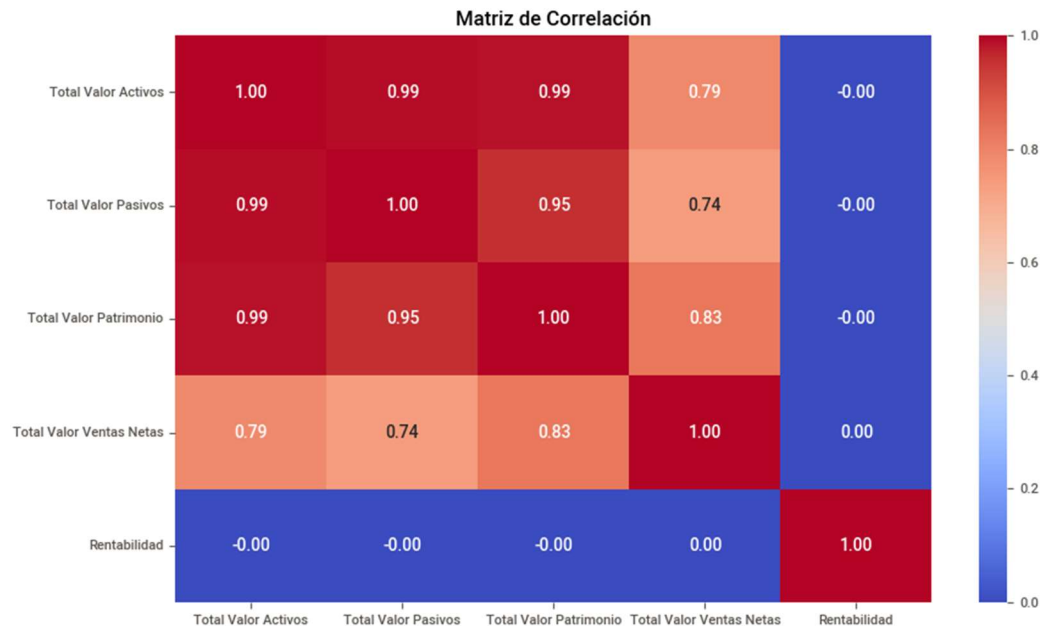
$$Rentabilidad = \frac{Total\ Valor\ Ventas\ Netas - Total\ Valor\ Pasivos}{Total\ Valor\ Activos}$$

Una vez generada la variable rentabilidad, se hizo un análisis de correlación con el propósito de conocer la fuerza y dirección de las relaciones entre la variables e identificar la multicolinealidad para poder identificar , combinar o eliminar variables redundantes y simplificar el análisis del modelo para finalmente definir la variable objetivo.

En la imagen de la matriz de correlación se pueden observar las variables financieras "Total Valor Activos", "Total Valor Pasivos", "Total Valor Patrimonio", "Total Valor Ventas Netas" y "Rentabilidad", las cuales se relacionan mediante una escala de colores que va del azul (correlación negativa) al rojo (correlación positiva), con valores numéricos que indican la fuerza y dirección de la correlación entre las variables.

Figura 16.

Matriz de Correlación



Análisis de la gráfica de correlación

1. Correlaciones Altas y Positivas:

- Existe una correlación muy alta y positiva entre "Total Valor Activos" y "Total Valor Pasivos" (0.99), "Total Valor Patrimonio" (0.99) y "Total Valor Ventas Netas" (0.79).
- "Total Valor Pasivos" también muestra una correlación alta y positiva con "Total Valor Patrimonio" (0.95) y "Total Valor Ventas Netas" (0.74).
- "Total Valor Patrimonio" tiene una correlación alta y positiva con "Total Valor Ventas Netas" (0.83).

2. Correlaciones con Rentabilidad:

- La "Rentabilidad" muestra una correlación muy baja y negativa con "Total Valor Activos" (-0.00), "Total Valor Pasivos" (-0.00) y "Total Valor Patrimonio" (-0.00).
- La "Rentabilidad" tiene una correlación muy baja y positiva con "Total Valor Ventas Netas" (0.00).

3. Interpretación general de las correlaciones:

- Las altas correlaciones positivas entre "Total Valor Activos", "Total Valor Pasivos", "Total Valor Patrimonio" y "Total Valor Ventas Netas" sugieren que estas variables tienden a aumentar o disminuir juntas. Esto puede indicar que las empresas con mayores activos también tienen mayores pasivos y patrimonio, así como mayores ventas netas.
- La baja correlación entre "Rentabilidad" y las demás variables indica que la rentabilidad no está fuertemente relacionada con los valores totales de activos, pasivos, patrimonio o ventas netas en esta muestra de empresas.

4. Regresión R2 Score:

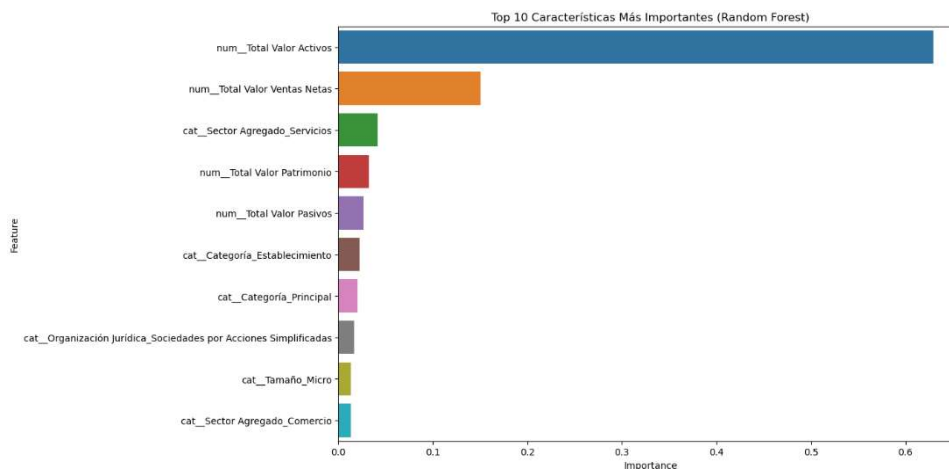
- El R2 Score de la regresión lineal es -0.17093195159975427, lo que indica que el modelo de regresión lineal no es adecuado para predecir la rentabilidad basada en las variables incluidas en la matriz de correlación. Un R2 negativo indica que el modelo está peor que una línea horizontal que pasa por la media de los datos.

Para evaluar la importancia de las variables se hizo un análisis con el modelo Random Forest, obteniendo los siguientes resultados:

R2 Score en entrenamiento (Random Forest): 0.8254589620584297
R2 Score en prueba (Random Forest): 0.9504716629402064
RMSE en prueba (Random Forest): 0.30356327627261004

Figura 17.

Top 10 Importancia Variables



Al analizar la gráfica de importancia, se puede observar que las variables mas importantes del modelo son la variable "num_Total Valor Activos" y la variable num_Total Ventas Netas lo que significa que estas 2 variables son las más influyentes en el modelo Random Forest para la toma de decisiones.

Respecto a los valores obtenidos durante las pruebas, el primer indicador R2 Score en Entrenamiento (0.8254589620584297) muestra que el modelo de Random Forest explica aproximadamente el 82.55% de la variabilidad de los datos de entrenamiento, lo cual es un buen indicador de que el modelo se ajusta bien a los datos con los que fue entrenado. El

segundo indicador R2 Score en Prueba (0.9504716629402064) es aún más alto que el R2 en entrenamiento, lo que sugiere que el modelo tiene un excelente desempeño en los datos de prueba, explicando el 95.05% de la variabilidad, es decir que es un signo muy positivo e indica que el modelo generaliza bien a datos nuevos y no vistos y es capaz de hacer predicciones muy precisas en datos nuevos. Finalmente, el Root Mean Squared Error RMSE en Prueba (0.30356327627261004) mide la diferencia promedio entre los valores predichos por el modelo y los valores reales, por lo tanto un valor de 0.3035 es relativamente bajo, lo que indica que las predicciones del modelo son bastante precisas y cercanas a los valores reales.

De estos resultados, se puede concluir que el modelo de Random Forest tiene un desempeño excelente tanto en los datos de entrenamiento como en los de prueba, lo que indica que es un modelo robusto, bien ajustado y se puede considerar muy efectivo para predecir las variables de interés en la base de datos.

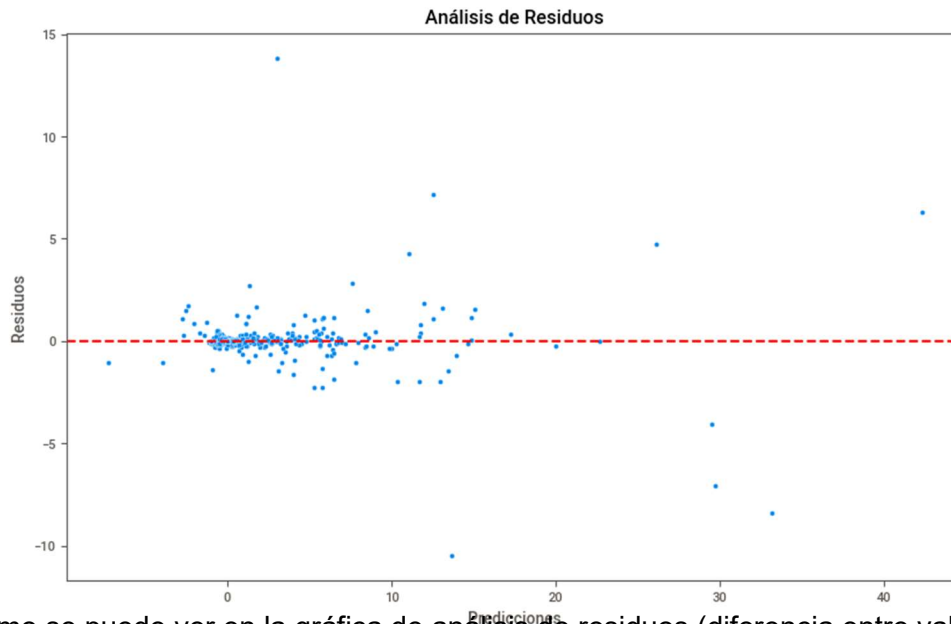
Con base a estos resultados y teniendo en cuenta que la variable Rendimiento no guarda correlación fuerte con las variables financieras, se tomó la decisión de asignar como la variable objetivo a la variable “Total Valor Ventas Netas”, toda vez que esta variable guarda una fuerte correlación con todas las variables financieras y es fuertemente sensible e influyente en la predicción de datos nuevos.

Comparación de modelos

A continuación se muestra el resultado del análisis de residuos, los valores obtenidos son los siguientes:

Figura 18.

Análisis de Residuos



Como se puede ver en la gráfica de análisis de residuos (diferencia entre valores predichos y valores reales) en el eje y, y las predicciones en el eje x. La línea roja discontinua en $y=0$ representa la línea de residuo cero. La mayoría de los puntos están cerca de esta línea, lo que es un buen indicador, pero hay algunos puntos dispersos que indican posibles valores atípicos o variabilidad en los residuos.

Los valores obtenidos de los tres modelos escogidos son los siguientes:

Random Forest R2 Score en prueba: 0.9504716629402064
Gradient Boosting R2 Score en prueba: 0.6202429330289292
Regresión Lineal R2 Score en prueba: -0.5405702623335871

Al comparar los resultados de los modelos, Random Forest con R2 Score en prueba de 0.9504716629402064 muestra un excelente desempeño, explicando el 95.05% de la variabilidad en los datos de prueba y los residuos están bien distribuidos alrededor de la línea de residuo cero, lo que indica un buen ajuste del modelo. La Regresión Lineal con

R2 Score en prueba de -0.540570262333587 tiene un desempeño muy pobre, con un R2 Score negativo, lo que indica que las predicciones son peores que simplemente usar la media de los valores observados. Los residuos muestran una gran dispersión, lo que indica un mal ajuste del modelo y el tercer modelo Gradient Boosting con R2 Score en prueba de 0.6202429330289292 tiene un desempeño aceptable, explicando el 62.02% de la variabilidad en los datos de prueba. Sin embargo, es menos preciso que el modelo de Random Forest. Por lo tanto el mejor modelo para hacer predicciones es el modelo Random Forest.

Para realizar pruebas de predicciones, se generó una variable con el incremento de la variable objetivo “num__Total Valor Ventas Netas”, considerando un incremento en ventas del 20% y el resultado obtenido es el siguiente:

Predicciones con incremento en Ventas Netas: [2.58124292 0.77968691 0.30259284 ... 0.38648909 1.56748095 0.75615507]

Un valor de 2.58 puede indicar una predicción de un caso con un rendimiento significativamente positivo y los valores como 0.78 o 0.30 reflejan predicciones más moderadas, sugiriendo que no todos los casos se benefician igualmente del incremento en las ventas netas.

Por lo tanto, el modelo tiene la capacidad de diferenciar entre los casos de acuerdo con otros factores subyacentes en los datos y el ajuste realizado en la variable num__Total Valor Ventas Netas, al incrementar las mismas en un 20%, el cual ha generado un aumento en los valores predichos por el modelo. Este cambio refleja que el modelo responde a la variación de las ventas netas, lo cual es consistente con la expectativa de que un aumento en las ventas debería tener un impacto positivo en las métricas relacionadas, como rentabilidad, ingresos o rendimiento.

DISCUSIÓN

Vinculación de los Resultados del Análisis con Decisiones Empresariales

Con base a los resultados obtenidos del análisis, particularmente la identificación de "Total Valor Activos" y "Total Valor Ventas Netas" como las variables más influyentes, son las que ofrecen un marco práctico para guiar decisiones estratégicas en las empresas de la cadena alimentaria del Huila. Estas decisiones pueden incluir:

Estrategias de Inversión:

Empresas con un bajo "Total Valor Activos" pueden priorizar inversiones en infraestructura o tecnología para mejorar su capacidad productiva. Por ejemplo, aquellas con una alta proporción de activos improductivos podrían enfocarse en la adquisición de activos más rentables, como maquinaria moderna o software de gestión.

Optimización de Recursos Financieros:

Las empresas con niveles desproporcionadamente altos de pasivos pueden utilizar los resultados del modelo para evaluar la viabilidad de renegociar deudas o buscar fuentes de financiamiento con mejores condiciones. Por ejemplo, un análisis detallado de las tendencias regionales podría recomendar la búsqueda de incentivos locales para empresas en sectores clave.

Estrategias de Venta Personalizadas:

Con base en la fuerte correlación entre "Total Valor Ventas Netas" y los otros indicadores, las empresas podrían desarrollar estrategias específicas para maximizar ventas en mercados locales. Por ejemplo, una empresa pequeña puede beneficiarse más al enfocarse en la diversificación de productos o nichos de mercado específicos.

Ejemplos Prácticos Basados en Simulaciones

Las simulaciones realizadas con un incremento del 20% en las ventas netas también pueden traducirse en decisiones específicas y a través de ejemplos Prácticos Basados en Simulaciones se pueden ver los siguientes escenarios.

Escenario 1: Empresa con baja capacidad de respuesta a un aumento en ventas.

Si el modelo predice que una empresa no se beneficiará significativamente de un incremento en ventas, se podría sugerir reorientar su estrategia hacia la optimización de costos o la mejora de procesos internos antes de intentar expandir ventas.

Escenario 2: Empresa con alta capacidad de respuesta a un aumento en ventas.

En este caso, las empresas podrían aprovechar las conclusiones para implementar estrategias agresivas de expansión, como campañas de marketing dirigidas o alianzas estratégicas para escalar su operación.

Estrategias de Sostenibilidad Empresarial

A partir del marco teórico-práctico propuesto, las siguientes estrategias pueden fomentar la sostenibilidad empresarial en un entorno volátil.

Digitalización y Automatización:

Basándose en el impacto de los activos sobre la rentabilidad, se podría priorizar la adopción de herramientas tecnológicas como software de gestión agrícola o sistemas IoT para optimizar la productividad en empresas del sector agroindustrial.

Ejemplo: Implementar sensores en cultivos para medir indicadores en tiempo real y tomar decisiones informadas.

Desarrollo de Capacidades Gerenciales:

Capacitar a los tomadores de decisiones en el uso del modelo predictivo para interpretar resultados y desarrollar planes de acción. Esto incluye talleres prácticos o la integración del modelo en sistemas de gestión empresarial.

Ejemplo: Crear dashboards interactivos que resuman las métricas críticas identificadas por el modelo.

Políticas de Sostenibilidad Financiera:

Fomentar el uso de los resultados del modelo para identificar empresas con altos riesgos financieros y diseñar estrategias de rescate o reestructuración de pasivos.

Ejemplo: Proponer esquemas de microcréditos ajustados al perfil de riesgo de cada empresa, basado en su desempeño histórico y proyectado.

Uso del Modelo para la Competitividad Regional

Los hallazgos no solo benefician a empresas individuales, sino que también pueden informar políticas públicas y programas de desarrollo empresarial en el Huila. Por ejemplo:

Clusters Regionales:

Crear clusters empresariales que agrupen a empresas con características similares para compartir recursos y conocimiento.

Ejemplo: Empresas con bajos "Total Valor Activos" podrían acceder a plataformas colaborativas para compartir maquinaria o equipos especializados.

Programas de Subsidios y Apoyos:

Utilizar las predicciones del modelo para identificar empresas clave que, con apoyo gubernamental, puedan generar un mayor impacto económico en la región.

Ejemplo: Proveer incentivos fiscales para empresas con altos niveles de ventas pero bajo capital para expandirse.

CONCLUSIONES

La economía del Huila enfrenta desafíos críticos en productividad y eficiencia, exacerbados por la reciente crisis económica, que ha resultado en el cierre de numerosas empresas. Actualmente, solo una cuarta parte de las empresas fundadas en 2017 sigue operativa. Ante este panorama, el análisis de datos y las técnicas de machine learning se posicionan como herramientas clave para mejorar la sostenibilidad y competitividad empresarial, facilitando la toma de decisiones estratégicas basadas en información.

Este estudio abordó el desarrollo de un modelo predictivo para estimar la rentabilidad de empresas en la cadena alimentaria del Huila, empleando indicadores financieros y características regionales. En la etapa de identificación y categorización de variables financieras críticas, se evidenció que la variable "Rentabilidad" (relación entre ventas netas, pasivos y activos) mostró una baja correlación con otras variables. Por ello, se optó por utilizar "Total Valor Ventas Netas" como variable objetivo, la cual presentó una fuerte correlación con "Total Valor Activos", "Total Valor Pasivos" y "Total Valor Patrimonio". Estas relaciones reflejan patrones típicos de empresas con mayor volumen de activos y ventas, lo que sugiere que estas variables son determinantes en la dinámica financiera regional.

En el desarrollo del modelo de análisis multivariante, se incorporaron tanto tendencias del mercado regional como variables financieras. Los resultados indicaron que el modelo más robusto fue el basado en Random Forest, alcanzando un R^2 de 0.95 en los datos de prueba,

demostrando alta precisión y capacidad para generalizar. En comparación, la Regresión Lineal mostró un desempeño deficiente (R^2 de -0.17), y Gradient Boosting presentó un desempeño moderado (R^2 de 0.62). El análisis de importancia de variables destacó que "Total Valor Activos" y "Total Valor Ventas Netas" son los principales factores que influyen en la rentabilidad empresarial, alineándose con las dinámicas observadas en empresas de mayor tamaño y alcance.

En cuanto a la validación del modelo predictivo, se llevaron a cabo simulaciones que exploraron el impacto de un aumento del 20% en las ventas netas. Estas simulaciones demostraron que el modelo puede diferenciar entre distintos escenarios y que no todas las empresas se benefician igual de un incremento en las ventas. Este hallazgo es crucial para diseñar estrategias personalizadas que aborden las necesidades específicas de cada empresa, aumentando así la relevancia práctica del modelo.

Finalmente, con base en estos resultados, se propone un marco teórico-práctico para la toma de decisiones empresariales orientado a la sostenibilidad y el crecimiento económico. Este marco utiliza los resultados del modelo predictivo como herramienta estratégica, permitiendo a los tomadores de decisiones priorizar inversiones, optimizar recursos y anticipar el impacto de las dinámicas de mercado en un entorno volátil como el del Huila. Este enfoque no solo fomenta la competitividad empresarial, sino que también contribuye a la sostenibilidad del sector alimentario regional al integrar técnicas avanzadas de análisis con la realidad económica local.

REFERENCIAS

- Afolabi, I. I. (2019). A Model for Business Success Prediction using Machine Learning Algorithms. *Journal of Physics: Conference Series*, 1299. <https://doi.org/10.1088/1742-6596/1299/1/012050>.
- Almagtome, A. &. (2020). Value relevance of financial performance measures: An empirical study. *International Journal of Psychological Rehabilitation.*, 24(7), 6777-6791.
- Asobancaria. (Junio de 2021). *Informe de tipificación la banca colombiana 2020*. Recuperado el 20 de Abril de 2024, de asobancaria.com: https://www.asobancaria.com/wp-content/uploads/2021/06/Informe-Tipificacion-2020_VF_web.pdf
- Aspilcueta Arias, A. N. (2020). *Gestión de Aprovisionamiento adaptado a MyPEs Comercializadoras de Productos de Consumo Masivo para incrementar las Utilidades Operacionales.* . Lima, Perú: <https://doi.org/10.1080/0013788X.2020.1811117>, Newman Business Review, 6(1), 27–53. . Obtenido de <https://doi.org/10.1080/0013788X.2020.1811117>
- Bada Carbajal, L. M., Rivas Tovar, L. A., & Littlewood Zimmerman, H. F. (2017). Model of associativity in the production chain in Agroindustrial SMEs. *Contaduría y administración*, 62(4), 1100–1117. doi:<https://doi.org/10.1016/j.cya.2017.06.006>
- Banco Interamericano de Desarrollo. (2018). *El futuro del trabajo en América Latina y el Caribe*.
- Barrera Molina, D. (2022). *Estrategias tecnológicas para mejorar la atención al cliente en las entidades financieras del Huila*. Universidad Militar Nueva Granada.

Barreto Granda, N. (2020). *Análisis financiero: factor sustancial para la toma de decisiones en una empresa del sector comercial*. . Revista Universidad y sociedad, 12(3), 129-134.

Benavides Pupiales, L. E. (2023). Estrategias de supervivencia empresarial de las pymes de la Ciudad de Pasto. *Revista Venezolana De Gerencia*, 28(No. Especial 9), 85-96. .
doi:<https://doi.org/10.52080/rvgluz.28.e9.6>

Bishop, C. (2013). *Pattern recognition and machine learning*. Springer (India) Private Limited.

Blandón, G., Zuluaga, B., Vásquez, J., López, R., & Moreno, E. (2024). Propuesta de modelo de analítica para flujo de caja en mipymes en Colombia. *Revista CEA*, 10(22), e2607-e2607.

Cafer Şafak Eysel, İ. B. (2019). *The Effect of Emotional Capital on Individual Innovativeness: A Research on Bahcesehir University Business Administration Undergraduate Students*. . Procedia Computer Science, 680-687. Obtenido de <https://www.sciencedirect.com>.

Cámara Colombiana de Informática y Telecomunicaciones . (2020). *Transformación Digital en Colombia*. Bogotá: CCIT.

Camara de comercio Bogota. (2024). *Camara de comercio Bogota*. Obtenido de Camara de comercio Bogota: https://recursos.ccb.org.co/ccb/instructivos/acerca_CIIU/01.html

Cámara de Comercio del Huila. (2023). *Informe Económico Empresarial*.

Cámara de Comercio del Huila. (2023). *Base de Datos Empresariales del Sector Alimentario*. [Proporcionado para análisis de investigación].

Ccopa Mamani, M., & Chavez Viza. (2015). Modelo Predictivo Basado en Minería de Datos Para la Mejora en la Toma de decisiones del departamento de Cirugía del Hospital Regional Manuel Núñez Butrón. Recuperado el 10 de Abril de 2024, de <https://renati.sunedu.gob.pe/handle/sunedu/3436947>

Cerquera Losada, O., Gómez Segura, C., & Arias Barrera, C. (2020). Competitividad de las exportaciones de tilapia en el Huila. *Revista Venezolana de Gerencia*, 26, 19.

doi:<https://doi.org/10.52080/rvgluz.26.e5.38>

Comisión Económica para América Latina y el Caribe. (2019). *Informe de Economía Digital 2019*.

Congreso de Colombia. (2012). Ley 1581 de 2012. Bogotá, D.C.

DANE. (2024). *Mercado laboral por departamentos*. Bogotá. Obtenido de

<https://www.dane.gov.co/files/operaciones/GEIH/bol-GEIHDepartamentos-2023.pdf>

Departamento Administrativo Nacional de Estadística . (2019). *Uso de tecnologías de la información y las comunicaciones (TIC) en empresas 2019*. Bogotá, Colombia: DANE.

Departamento Administrativo Nacional de Estadística. (2023). *Boletín Técnico Exportaciones (EXPO)*. Bogotá D.C.

Deskins, J. G.-C. (2012). *Understanding Small Business Activity at the State Level*. *Small Business Administration Office of Advocacy*. Obtenido de <https://advocacy.sba.gov/wp-content/uploads/2019/05/Understanding-Small-Business-Activity-at-the-State-Level-Full-Report.pdf>

Domingos, P. (2015). *The master algorithm: How the quest for the ultimate learning machine will remake our world*. Basic Books.

Doshi-Velez, F., & Kim, B. (2017). Towards a rigorous science of interpretable machine learning. *arXiv preprint arXiv:1702.08608*.

DYNA. (Agosto de 2023). Impacto de la inteligencia artificial en la toma de decisiones financieras: oportunidades y desafíos para los líderes empresariales. Recuperado el 30

de Septiembre de 2024, de

<https://revistas.unal.edu.co/index.php/dyna/article/view/114660/93182>

FINAGRO. (2024). *FINAGRO*. Obtenido de FINAGRO:

[https://www.finagro.com.co/noticias/sector-agropecuario-del-huila-enfrenta-grandes-retos#:~:text=Seg%C3%BAn%20el%20estudio%2C%20el%20Huila,infraestructura%20agropecuaria%20\(2926%20ha\)](https://www.finagro.com.co/noticias/sector-agropecuario-del-huila-enfrenta-grandes-retos#:~:text=Seg%C3%BAn%20el%20estudio%2C%20el%20Huila,infraestructura%20agropecuaria%20(2926%20ha)).

Garcés Camacho, H. (27 de Febrero de 2020). *Aplicación de Machine Learning en la demanda de productos permisibles*. Recuperado el 22 de Abril de 2024, de

<http://repositorio.ucsg.edu.ec/bitstream/3317/14264/1/T-UCSG-PRE-ECO-GES-621.pdf>

García, G. P. (Octubre de 2015). *Las apuestas productivas en el plan regional de Huila*.

Obtenido de Gobernación del Huila:

<https://www.huila.gov.co/loader.php?IServicio=Tools2&ITipo=descargas&IFuncion=descargar&idFile=7440>

García, G. P. (2022). *Factores claves en la predicción de la supervivencia empresarial en sectores agroindustriales*. . Revista Venezolana de Gerencia, 28(Especial 9), 85-96.

García, P., Vargas, F., & Pérez Gómez, V. (2015). *Actualización de la Agenda Interna de Productividad y Competitividad del Huila*. Obtenido de Gobernación del Huila: <chrome-extension://efaidnbmnnnibpcajpcglclefindmkaj/https://corhuila.edu.co/wp-content/uploads/2022/08/Anexo-7-Agenda-Interna-%E2%80%93-Plan-Regional-de-Competitividad-del-Huila-1.pdf>

Gobernación del Huila. (2020). *PLAN DE REACTIVACIÓN ECONÓMICA DEL DEPARTAMENTO*. NEIVA. Obtenido de <https://www.cchuila.org/wp-content/uploads/Plan-de-Reactivacion-Economica-del-Huila-CRECI.pdf>

Gobernación del Huila. (19 de 11 de 2023). *Caficultores del Huila más competitivos gracias a la*

Extensión Agropecuaria. Obtenido de

<https://www.huila.gov.co/publicaciones/14516/caficultores-del-huila-mas-competitivos-gracias-a-la-extension-agropecuaria/>

Grennan, J. &. (2021). *Fintechs and the market for financial analysis*. . Journal of Financial and Quantitative Analysis, 56(6), 1877-1907.

HANSEN, S. (2020). (U. d. Oxford, Ed.) *APLICACIÓN DEL APRENDIZAJE AUTOMÁTICO AL ANÁLISIS ECONÓMICO Y LA FORMULACIÓN DE POLÍTICAS*. Obtenido de

https://www.funcas.es/wp-content/uploads/Migracion/Articulos/FUNCAS_PEE/157art12.pdf

HANSEN, S. (2020). *APLICACIÓN DEL APRENDIZAJE AUTOMÁTICO AL ANÁLISIS ECONÓMICO Y LA FORMULACIÓN DE POLÍTICAS*. Obtenido de

https://www.funcas.es/wp-content/uploads/Migracion/Articulos/FUNCAS_PEE/157art12.pdf

Hastie, T. T. (2021). *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*. Springer.

Hastie, T., Tibshirani, R., Friedman, J., Tibshirani, R., & Friedman, J. (2009). *Boosting and additive trees. The elements of statistical learning: data mining, inference, and prediction*.

Hernández Julio, E. &. (2024). *Modelo de machine learning para predecir el rendimiento académico en la asignatura producción de artefactos tecnológicos del programa de licenciatura en informática de la Universidad de Córdoba*. . Universidad de Córdoba.

Hernández, E., Julio, J., & Pantoja Wilches, J. D. (2024). *Modelo de machine learning para predecir el rendimiento académico en la asignatura producción de artefactos tecnológicos del programa de licenciatura en informática de la Universidad de Córdoba.*

Obtenido de Universidad de Córdoba:

<https://repositorio.unicordoba.edu.co/entities/publication/051eba68-1df1-46b5-abb6-7e2fbebbae008>

Islas Moren, A., Rodríguez Ortega, A., & Aguirre Álvarez, G. (Septiembre de 2023). Éxito empresarial en los sectores agrícola, pecuario y agroindustrial: definiciones emitidas por expertos. *10 n° 30, 269-281*, 13. doi:<https://doi.org/10.35588/rivar.v10i30.5943>

Jordan, M., & Mitchell, T. (2015). Machine learning: Trends, perspectives, and prospects. *Science*, *349*, 255-260.

Losada, O., Gómez, V., & Chavarro, J. (2020). Análisis de la competitividad de las exportaciones del café del Huila. *Tendencias*, *21(2)*, 19-44.

Ministerio de Comercio, Industria y Turismo. (2019). *Por el cual se adiciona el capítulo 13 al Título 1 de la Parte 2 del Libro 2 del Decreto 1074.*

Ministerio de Tecnologías de la Información y las Comunicaciones. (2018). Decreto número 1008 de 2018. Bogotá, D.C.

Muñoz-Nieto, J. (2018). *Priorización de alternativas productivas agropecuarias y diagnóstico del mercado de productos agropecuarios del departamento del Huila.* Neiva, Huila: Convenio UPRA-Departamento del Huila.

Murphy, K. (2012). *Machine learning: a probabilistic perspective.* MIT press.

Navia, J. M. (2018). Factores determinantes de fracasos empresariales en Neiva (Colombia) durante el periodo 2000-2014. *Revista Espacios*, *39(9)*.

Nguyen, D. C. (2022). *Applications of Machine Learning in Finance and Economics*. Springer.

Palencia Fajardo, O., Cortés Lemus, M. Y., & Benavides Parra, Ó. A. (30 de Septiembre de 2022). La responsabilidad social, crecimiento de las empresas y medio ambiente en las pymes de La Plata, Huila (Colombia). *Perspectiva Empresarial*. Recuperado el 27 de Abril de 2024, de <https://revistas.ceipa.edu.co/index.php/perspectiva-empresarial/article/view/797/1107>

Parejo, I., Nuñez, L., & Nuñez, W. (2021). Análisis de la transformación digital de las empresas en Colombia: dinámicas globales y desafíos actuales. *Aglala*, 12(1), 129-141.

Presidencia de la República de Colombia. (2013). Decreto número 1377 de 2013. Bogotá, D.C., Colombia.

Rodríguez Borray, G. A., Cruz Castiblanco, G. N., Tauta Muñoz, J. L., Huertas Carranza, B., & Polo Murcia, S. M. (2022). Diversidad de empresas agroindustriales rurales: tipologías de producción de panela en Huila, Colombia. *Agronomía Mesoamericana*, 33(2), 23. doi:10.15517/am.v33i2.47969

Román Ramírez, D. (2021). *Supervivencia de las nuevas empresas: una aproximación desde el Machine Learning*. Universidad EAFIT.

Román Ramírez, D. (2021). *Supervivencia de las nuevas empresas: una aproximación desde el Machine Learning*. Universidad EAFIT.

Román Ramírez, J. (2021). Aplicación del machine learning en la predicción de supervivencia empresarial en Colombia. . *Revista Científica de Gestión Empresarial*, 34(1), 19-35.

Romero Acosta, D. (2019). *Diseño y aplicación de una metodología de análisis de las decisiones para caracterizar los factores de supervivencia empresarial en pequeñas*

empresas del sector manufactura de la ciudad de Bogotá DC. Uniandes. Obtenido de
<http://hdl.handle.net/1992/43923>

Salazar, M. (2024). Algoritmos Evolutivos en la Economía y las Finanzas . 289,441.

Recuperado el 2024, de <https://revistavitalia.org/index.php/vitalia/article/view/289/441>

Sutton, R., & Barto, A. (2018). *Reinforcement learning: An introduction*. MIT press.

UNESCO. (2021). Recomendación sobre la ética de la inteligencia artificial. París.

Vázquez, J., Cebolla, M., & Ramos, F. (2019). La transformación digital en el sector cooperativo agroalimentario español: situación y perspectivas. *Revista de Economía pública, social y cooperativa*, 95, 39-70.