

UNIVERSIDAD DE LA SABANA



Facultad de Ingeniería

Maestría en Gerencia de Ingeniería

INGENIERIA ANALITICA PARA LA PREDICCIÓN DE FRACASO DE LAS MICROEMPRESAS EN
COLOMBIA

Área de investigación

Chía, septiembre de 2020

Documento Final

INGENIERIA ANALITICA PARA LA PREDICCIÓN DE FRACASO DE LAS MICROEMPRESAS EN
COLOMBIA

Mónica Andrea Gómez Barrera

Director de Proyecto

Luis Carlos Rabelo, Ph.D.

Facultad de Ingeniería

Maestría en Gerencia de Ingeniería

Chía, septiembre de 2020

CONTENIDO

1. INTRODUCCIÓN	7
2. PROBLEMA	8
2.1 Contexto General	8
2.2 Contexto Particular	8
2.3 Pregunta.....	8
3. JUSTIFICACION	9
4. OBJETIVOS	12
4.1 General.....	12
4.2 Específicos.....	12
5. MARCO TEORICO	13
5.1 Indicadores Financieros y la Predicción de Fracaso Empresarial	13
5.2 Fracaso Empresarial en Colombia.....	16
5.3 Ingeniería Analítica Predictiva	19
5.4 Árboles de decisión como modelos de predicción	20
5.5 Random Forest como modelos de predicción	22
5.6 Redes Neuronales como Modelo de Predicción.....	22
5.6.1 Algoritmo de ajuste de decaimiento dinámico.....	25
5.6.2 Resilient back propagation	26
5.7 La Regresión como Modelo de Predicción	26
5.7.1 Regresión Lineal Simple	26
5.7.2 Regresión Múltiple.....	27
5.7.3 Modelo de probabilidad lineal.....	28
5.8 Herramientas para el análisis de desempeño de los modelos	28
5.8.1 Curvas ROC y AUC	28
5.8.2 Precisión.....	30
5.8.3 Tasa de verdaderos positivos:.....	31
6. METODOLOGIA	31
7. IDENTIFICACION DE LAS VARIABLES FINANCIERAS POR SECTOR	35
7.1 Variables financieras para el Sector Primario: Agricultura.....	35
7.2 Variables financieras para el Sector Secundario: Industrial	37
7.3 Variables financieras para el Sector Terciario: Servicios	39

8. RESULTADOS PARA TODA LA MUESTRA SIN CLASIFICAR POR SECTORES	43
8.1 Resultados del mejor modelo: Árboles Binarios.....	44
9. RESULTADOS SECTOR AGRICULTURA.....	47
9.1 Resultado mejor modelo: Random Forest.....	47
10. RESULTADOS SECTOR INDUSTRIAL.....	52
10.1 Resultados del mejor modelo: Random Forest	52
11. RESULTADOS SECTOR SERVICIOS.....	58
11.1 Resultados del mejor modelo: Random Forest	59
12. VALIDACIÓN DE LOS RESULTADOS CON EXPERTOS.....	63
13. CONCLUSIONES.....	64
14. CONTRIBUCIONES.....	65
15. TRABAJOS EN EL FUTURO	66
16. LIMITACIONES DEL TRABAJO	67
17. REFERENCIAS BIBLIOGRAFICAS.....	68
17. ANEXOS.....	77
17.1 Modelos Knime: Mejor modelo para toda la muestra: Árboles Binarios.....	77
17.2 Modelos Knime: Mejor modelo Sector Agricultura: Random Forest	78
17.3 Modelos Knime: Mejor modelo Sector Industrial: Random Forest.....	79
17.4 Modelos Knime: Mejor modelo Sector Servicios: Random Forest.....	80

INDICE DE TABLAS

Tabla 1. Evolución del salario mínimo en Colombia _____	32
Tabla 2. Clasificación por códigos CIU _____	33
Tabla 3. Indicadores financieros sector Agricultura _____	37
Tabla 4. Indicadores financieros sector Industrial _____	39
Tabla 5. Indicadores financieros sector Servicios _____	41
Tabla 6. Resumen de los indicadores financieros _____	42
Tabla 7. Desempeño de los diferentes modelos para toda la muestra sin clasificar por sectores _____	43
Tabla 8. Desempeño de los diferentes modelos Sector Agricultura _____	47
Tabla 9. Random Forest Sector Agricultura _____	49
Tabla 10. Porcentaje de elección Random Forest _____	50
Tabla 11. Desempeño de los diferentes modelos Sector Industrial _____	52
Tabla 12. Random Forest Sector Industrial _____	54
Tabla 13. Porcentaje de elección de cada indicador, Sector Industrial _____	55
Tabla 14. Resultados Sensibilidad DDA _____	58
Tabla 15. Desempeño de los diferentes modelos Sector Industrial _____	59
Tabla 16. Random Forest Sector Servicios _____	60
Tabla 17. Porcentaje elección de cada indicador, sector Servicios _____	61

INDICE DE ILUSTRACIONES

Ilustración 1. Búsqueda Avanzada Biblioteca Universidad de la Sabana	11
Ilustración 2. Evolución de los procesos de reorganización de las empresas en Colombia desde 2007 hasta 2018	18
Ilustración 3. Redes Neuronales Back Propagation	23
Ilustración 4. Set de variables utilizado por (Duarte, 2019)	40
Ilustración 5. Curva ROC Muestra sin clasificar por sectores, modelo Árboles Binarios	44
Ilustración 6. Matriz de confusión Árboles Binarios, toda la muestra	45
Ilustración 7. Árboles Binarios para toda la muestra	46
Ilustración 8. Curva ROC Sector Agricultura, Random Forest	48
Ilustración 9. Matriz de confusión Random Forest, Sector Agricultura	49
Ilustración 10. Promedio de los indicadores que predicen el fracaso sector Agricultura	51
Ilustración 11. Promedio otras cuentas sector Agricultura	51
Ilustración 12. Curva ROC, Sector Industrial, Random Forest	53
Ilustración 13. Matriz de confusión Random Forest, Sector Industrial	54
Ilustración 14. Promedio Indicadores Sector Industrial	56
Ilustración 15. Promedio cuentas - cálculo de los principales indicadores	57
Ilustración 16. Curva ROC, Sector Servicios, Random Forest	59
Ilustración 17. Matriz de Confusión Random Forest Sector Servicios	60
Ilustración 18. Promedio Otras Cuentas Sector Servicios	61
Ilustración 19. Promedio de los Indicadores que predicen el Fracaso en el Sector Servicios	62
Ilustración 20. Promedio cuentas - Indicadores Sector Servicios	62

1. INTRODUCCIÓN

Las microempresas son todas aquellas empresas cuyos activos se encuentran por debajo de los 500 SMMLV (es decir 414 millones de pesos colombianos para el año 2019) y un número inferior a los diez trabajadores, según definición dada por la Ley 590 de 2000 (Colombia, 2000). En América Latina y en Colombia, estas empresas son muchas veces una estrategia de autoempleo; algunas de las razones que afectan su competitividad son, la falta de financiamiento, dificultad de acceso a los mercados y limitada mano de obra calificada (CEPAL, 2015).

La ingeniería analítica es según el IISE, Institute of Industrial and Systems Engineers, por sus siglas en inglés, “El proceso de aprovechamiento de los datos en información procesable” (IISE, 2019) de manera que su adaptación para este caso en particular, nos permita llegar a conclusiones acerca de las variables financieras que predicen el fracaso empresarial de las microempresas en Colombia.

Muchas son las investigaciones realizadas que estudian la competitividad y supervivencia de las pequeñas y medianas empresas en Latinoamérica, sin embargo, pocos de estos estudios han utilizado modelos de predicción de fracaso, por esta razón, este trabajo pretende utilizar las herramientas de la ingeniería analítica que ayuden a identificar los indicadores financieros claves en el fracaso empresarial de las microempresas en Colombia, teniendo en cuenta el sector productivo al que pertenece, tomando como definición de fracaso empresarial aquellas empresas que se encuentren en Concordato, reorganización, reestructuración y liquidación, según definición de Pérez et al., (Perez G, Lorena, & Mauricio, 2013).

2. PROBLEMA

2.1 Contexto General

Según Confecamaras (Confecamaras, 2016), en un estudio realizado en 2016, el 70,9% de las microempresas fracasan en Colombia, algunas de las causas estudiadas para las pymes (no microempresas específicamente) están relacionadas con problemas de liquidez y endeudamiento (Romero Espinosa, 2013). De manera que la alta tasa de mortalidad de las microempresas constituye el problema de investigación de este trabajo.

2.2 Contexto Particular

“Cerca de 18,752 empresas se cierran cada año en Bogotá, lo cual disminuye la capacidad productiva de la ciudad, se reducen los ingresos tanto públicos como privados, se pierden esfuerzos y experiencia de los emprendedores y se limitan las fuentes de empleo en la ciudad” (Dirección de Gestión y Transformación del Conocimiento, 2019).

De acuerdo con un estudio realizado por la cámara de comercio de Bogotá a las empresas liquidadas en la misma ciudad el año anterior, se concluye que las principales razones financieras que llevan a la quiebra a las empresas son las bajas ventas y la falta de liquidez. De otro lado, el empresario decide lanzar su empresa basado en el conocimiento empírico del sector productivo, sin buscar asesoría especializada (Dirección de Gestión y Transformación del Conocimiento, 2019).

2.3 Pregunta

¿Qué modelo de Ingeniería analítica tiene mayor capacidad de predicción del fracaso de las microempresas en Colombia y cuáles son sus variables financieras?

3. JUSTIFICACION

Las mipymes (micro, pequeñas y medianas empresas) son quizás una de las mayores fuentes de generación de empleo (Gil Ospina, Armando A, Jimenez Sepulveda, 2015). Según datos del gobierno general, las micro y pequeñas empresas proveen el 63% del empleo, el 45% de la producción manufacturera y el 40% de los salarios de nuestro país (DANE, 2009). En las regiones el mayor porcentaje de las personas empleadas se da gracias a las microempresas, (Gil Ospina, Armando A, Jimenez Sepulveda, 2015), sin embargo la tasa de mortalidad en Colombia es del 70,9% (Confecamaras, 2016).

No son muchos los estudios relacionados con el fracaso empresarial en Colombia. En Bogotá, se realizó un análisis de empresas en donde se clasificó empresas sanas y empresas fracasadas y dicho análisis mostró que, durante el primer año sobreviven el 95% de las microempresas, sin embargo, solo el 50% de estas llegan hasta los ocho años. El mayor número de empleados y el mayor ROA inicial son unos de los factores que incidieron en el no cierre de las empresas (Santana, 2017).

En Colombia entre 2005 y 2011, el fracaso en las pequeñas y medianas empresas estuvo dado por su alto apalancamiento a largo plazo, problemas de liquidez y endeudamiento (Romero Espinosa, 2013). En 2015 Espinosa et al., (Espinosa, Molina, & Vera-Colina, 2015), identificaron las variables financieras que llevan a fracasar a las pymes, siendo una de ellas el costo de producción y los gastos de administración. En el caso del análisis a las empresas fracasadas, las ventas netas no alcanzaron a cubrir los costos de ventas ni gastos de administración, aun cuando estos habían sufrido una reducción.

En Latinoamérica también se han realizado estudios que intentan explicar los factores del fracaso de las microempresas. Por ejemplo, en República Dominicana, el fracaso de las microempresas está relacionado con características demográficas, educación, edad y genero del propietario. (Ortiz Medina, 2013), mientras que en Argentina, Chile y Perú las empresas sanas muestran unos

indicadores financieros de liquidez y de rentabilidad mayores que las empresas en crisis (Caro, 2016).

Dentro de las herramientas de ingeniería analítica que se van a utilizar en este trabajo, se encuentran, regresión lineal, Random Forest, redes neuronales y árboles binarios.

Los modelos de Redes Neuronales (RN) han sido utilizados en los últimos años para la predicción de precios de commodities en nuestro país, tales como petróleo (Ortiz Arango, 2017), oro (Villada, Muñoz, & García-Quintero, 2016), carbón (Franco Sepúlveda, Velilla, & Velez, 2014), energía eléctrica (Medina Hurtado, Santiago, 2011) incluso para determinar el índice de compromiso exportador de las pymes (Escandon, Diana, Hurtado, 2017). La efectividad de predicción de estos modelos resulta en la mayoría de los casos, ser mayor que la de los modelos tradicionales.

De otro lado, en Colombia, se han utilizado técnicas de machine learning como Random Forest en varios escenarios de clasificación de producción de cacao y de café. Por ejemplo, en Santander, se utilizaron varias herramientas para evaluar la clasificación de producción de cacao, ya que el cacao es soporte fundamental para cerca de 25.000 familias en la región y la producción se ve afectada por varios factores, entre ellos, clima, agua, tierra, etc. El Random Forest resulto ser la herramienta que mejor precisión mostró (Plazas, López, & Corrales, 2017).

La Regresión Lineal ha sido ampliamente utilizada en este campo en Colombia, para la predicción de fracaso de las Pymes, pequeñas y medianas, y grandes empresas, sin embargo, para la predicción de quiebra de las microempresas en Colombia, no se encontraron estudios. De manera que este trabajo es original e innovador, no solo en cuanto al sujeto de investigación, (las microempresas) sino también al uso de las herramientas aplicadas (Árboles Binarios, Random Forest y Redes Neuronales).

Teniendo en cuenta la importancia de las microempresas en la economía colombiana, esta investigación pretende realizar uso de las diferentes herramientas de la Ingeniería Analítica, para de esta manera, establecer los indicadores determinantes en el fracaso de las microempresas en Colombia, en donde las variables determinantes serán los indicadores financieros. Los resultados del modelo permitirán saber con anterioridad si una empresa va por el camino a la quiebra, de manera que se alcancen a tomar medidas por parte de los directivos o dueños. De igual manera será de utilidad para los bancos y demás agentes proveedores de liquidez, en cuanto a mejor asignación de recursos se refiere.

Este trabajo es innovador en Colombia, no solo por el sujeto de investigación (Microempresas en Colombia), sino también por las técnicas utilizadas en el estudio del fracaso empresarial, ver Ilustración 1.

Ilustración 1. Búsqueda Avanzada Biblioteca Universidad de la Sabana

The screenshot displays the Eureka! search interface for the Biblioteca Octavio Arizmendi Posada. The search criteria are set to 'Colombia' and 'fracaso'. The results page shows 1 to 9 of 9 results. A note indicates that exact repetitions have been removed. The results are as follows:

- El fracaso en los emprendedores colombianos: una aproximación desde la logística**
Ver registro en OpenAIRE | Exportar a Zotero/Mendeley
- Causas de la siniestralidad de las MiPymes en Colombia: factores internos**
Ver registro en OpenAIRE | Exportar a Zotero/Mendeley
- Alcances y limitaciones de los modelos de capacidad predictiva en el análisis del fracaso empresarial.**
Scope and limitations of predictive models in the analysis of business failure.
Publicación académica | Texto completo en PDF | Save PDF to Cloud (289KB) | Exportar a Zotero/Mendeley
- LOS MECANISMOS DE REDES Y LA DIMENSIÓN CULTURAL, INDIVIDUALISMO Y COLECTIVISMO: SUPERVIVENCIA DE ALGUNOS PRODES DE ANTIOQUIA, BOGOTÁ, TOLIMA Y CALDAS.**
MECHANISMS OF NETWORKS AND THE CULTURAL DIMENSION, INDIVIDUALISM AND COLLECTIVISM: SURVIVAL OF SOME OF PRODES OF ANTIOQUIA, BOGOTÁ, TOLIMA AND CALDAS. / OS MECANISMOS DE REDES E A DIMENSÃO CULTURAL, INDIVIDUALISMO E COLETIVISMO: SOBREVIVÊNCIA DE ALGUM PRODES DE ANTIOQUIA, BOGOTÁ, TOLIMA E CALDAS. / LES MECANISMES DE RÉSEAUX ET LA DIMENSION CULTURELLE, INDIVIDUALISME ET COLLECTIVISME: SURVIE DE QUELQUES PRODES D'ANTIOQUIA, BOGOTÁ, TOLIMA ET CALDAS.
Publicación académica | Texto completo en PDF | Save PDF to Cloud (1.2MB) | Exportar a Zotero/Mendeley
- INNOVACIÓN SOCIAL ABIERTA EN EL DISEÑO DE UNA POLÍTICA Y ESTRATEGIA DE FORMALIZACIÓN SOSTENIBLE: UN CASO COLOMBIANO DE GOBIERNO COLABORATIVO.**
OPEN SOCIAL INNOVATION IN POLICY AND STRATEGY DESIGN FOR SUSTAINABLE FORMALIZATION: COLOMBIAN CASE OF COLLABORATIVE GOVERNMENT. / INOVAÇÃO SOCIAL ABERTA NO DESENHO DE UMA POLÍTICA E DE UMA ESTRATÉGIA DE FORMALIZAÇÃO SUSTENTÁVEL: UM CASO COLOMBIANO DE GOVERNO COLABORATIVO.
Publicación académica | Texto completo en PDF | Save PDF to Cloud (1.4MB) | Exportar a Zotero/Mendeley
- Variables financieras determinantes del fracaso empresarial para la pequeña y mediana empresa en Colombia: análisis bajo modelo Logit.**
Determining financial variables in the business failure to small and medium enterprises in Colombia: analysis on Logit model.
Publicación académica | Texto completo en PDF | Save PDF to Cloud (224KB) | Exportar a Zotero/Mendeley

7. MODELO PARA LA ELABORACIÓN DE UN PLAN DE NEGOCIOS PARA LAS EMPRESAS PEQUEÑAS Y MEDIANAS



[Ver registro en OpenATRE](#) [Exportar a Zotero/Mendeley](#)



8. MODELO PARA LA ELABORACIÓN DE UN PLAN DE NEGOCIOS PARA LAS EMPRESAS PEQUEÑAS Y MEDIANAS



[Vea este registro desde ScienceDirect](#) [Exportar a Zotero/Mendeley](#)



9. MODELO PARA LA ELABORACIÓN DE UN PLAN DE NEGOCIOS PARA LAS EMPRESAS PEQUEÑAS Y MEDIANAS



[Vea este registro desde ScienceDirect](#) [Exportar a Zotero/Mendeley](#)



Publicación académica

Página: 1

4. OBJETIVOS

4.1 General

Mediante un marco de decisión, escoger el modelo con mayor capacidad predictiva del fracaso empresarial de las microempresas en Colombia para identificar los indicadores financieros determinantes de este.

4.2 Específicos

- Recopilar y clasificar la información de las microempresas por sector
- Identificar los indicadores financieros que harán parte del modelo basados en la teoría
- Calcular los indicadores financieros para las empresas según sector
- Entrenar el modelo Random Forest para las empresas fracasadas y activas
- Entrenar el modelo de árboles binarios para las empresas fracasadas y activas
- Entrenar el modelo de redes neuronales para las empresas fracasadas y activas
- Crear un modelo de probabilidad lineal para cada sector de la economía
- Probar el modelo Random Forest con datos que no hayan sido utilizados para el entrenamiento

- Probar el modelo de árboles binarios con datos que no hayan sido utilizados para el entrenamiento
- Probar el modelo de redes neuronales con datos que no hayan sido utilizados para el entrenamiento
- Crear un marco de decisión para escoger el modelo para cada sector de la economía
- Identificar los indicadores fundamentales en la predicción de fracaso para cada sector de la economía
- Validar los resultados con expertos en temas de gestión de las microempresas

5. MARCO TEORICO

5.1 Indicadores Financieros y la Predicción de Fracaso Empresarial

Los primeros estudios dedicados a la predicción del fracaso empresarial se dieron en 1966, por William Beaver y en 1968 por Edward Altman. Los resultados de Altman mostraron que la combinación de indicadores financieros en un modelo de análisis multivariado es más eficiente que la aplicación común de los ratios financieros (E. Altman, 1968), de otro lado, Beaver demostró que la diferencia de indicadores entre las empresas que fracasan y las que no fracasan son profundas y a medida que se acerca el fracaso se acentúan más (Beaver, 1966).

El análisis de los indicadores financieros no solo es importante para la empresa sino indispensable a la hora de evaluar la situación de las compañías y su manejo (Fedorova, Gilenko, & Dovzhenko, 2013), sin embargo, un análisis de estos ratios es de mayor utilidad cuando se realiza de manera dinámica y por sector, es decir se evalúan varios periodos de tiempo teniendo en cuenta el sector competitivo de cada empresa (Creus, 2000). Esta clasificación es de vital importancia ya que, los sectores productivos difieren en cuanto ciclos de vida de los productos, estructura competitiva y modelos de distribución (Platt, Harlan, Platt, 1990). Por ejemplo, para el sector construcción la

“rentabilidad económica” estuvo relacionada con el ROA y el ROAll, para el sector servicios, el ROA y la rotación del activo total RAC definieron el “margen de beneficios” (Mures Quintana, Ma Jesus, Garcia Gallego, Ana, Vallejo Pascual, 2012).

Turetsky et al., (Turetsky, Howard, McEwen, 2001) identificó factores potenciales de riesgo definidos como ROA (Rentabilidad de los activos) y ROS (Rentabilidad de las ventas). Del Llano et al., (De Llano Monelos, Piñeiro Sanchez, & Rodriguez Lopez, 2016) seleccionaron 59 indicadores financieros de los cuales los indicadores de rotación, la liquidez del ciclo de operaciones, la rentabilidad del negocio y la estructura de la deuda son los indicadores financieros relacionados con el riesgo en los plazos cortos.

Altman, el primero en proponer un modelo formal de fracaso empresarial evaluó cinco ratios contables, dentro de los cuales se encuentran Capital de trabajo/activos totales, ganancias retenidas/activos totales, EBIT (earnings before interest and taxes)/activos totales, capitalización/deuda y ventas/activos totales, a esta combinación de indicadores llamo el Z-Score (E. Altman, 1968).

La Función discriminante del Z-Score es la siguiente:

$$Z = 0.012X_1 + 0.014X_2 + 0.033X_3 + 0.006X_4 + 0.999X_5$$

Donde, X1 es Capital de trabajo/activos totales, X2 es ganancias retenidas/activos totales, X3 es EBIT/activos totales, X4 es capitalización/deuda y X5 es ventas/activos totales.

El Z-Score es un modelo reconocido en las finanzas para la predicción del fracaso empresarial. La información obtenida para elaborar este modelo se tomó de Moody’s Industrial Manual, por lo que este modelo es aplicable para empresas industriales (Giner, 2013).

Otro modelo comúnmente utilizado para predecir el fracaso empresarial es el “Z Credit Risk Model” que es un modelo más sofisticado estadísticamente y que permite la predicción del fracaso con cinco años de anterioridad, mientras que el modelo Z-Score lo hace con dos años de anticipación (E. I. Altman, 2000).

Los modelos desarrollados para predecir el fracaso de las empresas tienen como insumo principal los estados financieros de las compañías a analizar, para grandes compañías con estados financieros de un año es suficiente (Beaver, 1966), para las empresas pequeñas es necesario tener al menos tres años consecutivos de información (Edmister, 1972), de otro lado, realizar un estudio que implique analizar empresas grandes y pequeñas sería un error. William Beaver (Beaver, 1966) encontró evidencia probabilística que sugiere que el tamaño de los activos (esto es lo que diferencia las microempresas de las pymes y de las grandes empresas) altera la relación entre indicadores financieros y fracaso empresarial.

Los indicadores financieros han sido ampliamente utilizados para predecir el fracaso de las empresas, sin embargo, especial atención se ha presentado al tema desde 1968, cuando Altman introdujo el análisis de fracaso empresarial con el uso de indicadores financieros (E. Altman, 1968), siendo lo más utilizados en los modelos, los indicadores de rentabilidad y solvencia (E. Altman, Iwanicz-Drozdowska, Laitinen, & Suvas, 2017), sin embargo, indicadores de liquidez y de flujo de caja deben ser considerados también (Lukason & Laitinen, 2018).

Un estudio realizado por Lukason et al, (Lukason, Laitinen, & Suvas, 2016) para micro empresas jóvenes en Francia, demuestra que los indicadores más poderosos para predecir fracaso son liquidez y solvencia, siendo la liquidez, la habilidad para pagar las deudas de corto plazo, mientras la solvencia la capacidad de pagar las deudas de largo plazo, de otro lado también encontró que la edad de la empresa es inversamente proporcional a la probabilidad de fracaso.

Laitinen et al, (Lukason, Oliver, Laitinen, 2017) utilizaron indicadores de apalancamiento, rentabilidad y liquidez, mientras que (Lukason et al., 2016) añadió indicadores de actividad y de eficiencia

5.2 Fracaso Empresarial en Colombia

Al estudiar el fracaso empresarial en Colombia encontramos que no son muchos los trabajos de investigación realizados en esta área, sobre todo si se reduce la búsqueda al fracaso de las microempresas o pymes en el país; de hecho, Cesar León escribe para la revista de la facultad de Contaduría Pública de la Universidad del Externado, “ Aunque esta clase de análisis –como se mencionó al iniciar éste artículo– es prácticamente desconocida en el país, creemos que vale la pena iniciar su divulgación en los medios empresariales del país como una herramienta adicional para la toma de decisiones oportunas y que conduzcan a una mejor y más eficiente administración financiera de las empresas colombianas” (Leon Valdes, 2002) casi veinte años más tarde, nos encontramos con que la situación en Colombia no ha cambiado significativamente.

La variación del modelo Z-Score de Altman se ha aplicado un par de veces a empresas colombianas. Byington Colombia hace una modificación de la formula original, aplicando la siguiente (Leon Valdes, 2002):

$$(1.2 * (\text{Activo Corriente} - \text{Pasivo Corriente}) / \text{Activo Total}) + (1.4 * (\text{Utilidades Retenidas} / \text{Activo Total})) + (3.3 * (\text{Utilidades antes de Impuestos y Corrección} + \text{Intereses Pagados}) / \text{Activo Total}) + (0.6 * (\text{Patrimonio Líquido} / \text{Activo Total})) + (\text{Ingresos Netos} / \text{Activo Total}).$$

La modificación se realizó en las utilidades retenidas cambiándolas en este modelo a utilidades en periodos anteriores y reservas.

Los rangos establecidos por Altman para el análisis de resultados son los siguientes:

1. Resultados entre 1.8 y 3.0 se encuentran en una franja gris y el resultado es no determinado
2. Resultados por encima de 3.0 les da a las empresas una seguridad de seguir operando por los siguientes dos años

3. Resultados por debajo de 1.8 corresponden a empresas cuya quiebra es casi inevitable por insolvencia

De acuerdo con lo anterior, el estudio de Byinton en Colombia en 2002, el 77% de las empresas obtuvieron resultados por debajo de 2.7, de estos, los sectores con empresas que iban camino a la insolvencia se encontraban pesca, hoteles, restaurante y transporte. Mientras que los mejores resultados se obtuvieron de empresas de construcción y actividad inmobiliaria (Leon Valdes, 2002).

Cruz et al., (Cruz, Espinosa, & Aristizabal, 2014) encontró que las empresas pyme que reportan pérdidas consecutivas tienen ocho veces más probabilidad de caer en estado de insolvencia, mientras que (Perez G et al., 2013) observó que los sectores de agricultura e industria manufacturera son los más propensos a caer en riesgo de quiebra. El menos propenso es el sector de actividades inmobiliarias.

Fredy Espinosa (Romero Espinosa, 2013) destacó que los principales problemas de las pymes en Colombia son de liquidez y de endeudamiento, además rechazó la hipótesis que la empresas jóvenes son las más propensas al fracaso, de hecho las empresas con mayor probabilidad de fracaso son las que se encuentran entre 22 y 30 años de edad. Lo anterior confirma el estudio realizado por Confecamaras en 2017 (Confecamaras, 2017), donde identificó 612 pymes que crecieron sostenidamente entre el 2011 y 2015, este crecimiento iba de la mano con factores de innovación y edad de la empresa, las empresas jóvenes tienen mayor crecimiento sostenido. Las características que se identificaron como aceleradores del crecimiento son direccionamiento estratégico, mejora continua de productividad y eficiencia, compañías jóvenes, incentivos para retener el capital humano, procesos de innovación, trabajadores con mayor nivel académico, conocimiento del cliente y uso de nuevas tecnologías. (Confecamaras, 2017).

Respecto a la finalidad, principios y alcance de la Insolvencia de las empresas en Colombia, el gobierno decreto en la ley 1116 de 2006 que “tiene por objeto la protección del crédito y la

recuperación y conservación de la empresa como unidad de explotación económica y fuente generadora de empleo, a través de los procesos de reorganización y de liquidación judicial, siempre bajo el criterio de agregación de valor” («Leyes desde 1992 - Vigencia expresa y control de constitucionalidad [LEY_1116_2006]», 2019)

En cuanto a las estadísticas de empresas que se acogen a esta ley, las micro y pequeñas empresas son las que más se declaran en estado de insolvencia. Con corte a marzo de 2018, 537 microempresas y 504 pequeñas empresas se acogieron a esta ley, 295 empresas medianas y 195 empresas grandes. Los sectores más perjudicados son manufactura, comercio y servicios, de las cuales, 389 de las empresas que se acogen a esta ley pertenecen al sector comercio, mientras que 351 son del sector servicios y 349 del sector manufactura (La Republica, 2018).

Vale la pena anotar que muchas de las empresas pequeñas de Colombia ni siquiera están registradas en cámara de comercio, cuando dejan de funcionar simplemente cierran sus puertas y no operan más, esto se debe a la gran cantidad de informalidad en estos tamaños de empresas, de manera que el número de microempresas quebradas en realidad es mayor.

En la Ilustración 2 se puede observar la evolución desde 2007 de los procesos en reorganización de las empresas en Colombia, hasta 2018 (La Republica, 2018)

Ilustración 2. Evolución de los procesos de reorganización de las empresas en Colombia desde 2007 hasta 2018



Fuente: (La Republica, 2018)

En cuanto al acceso a créditos bancarios de las pequeñas empresas en Colombia, según Mckinsey, “los dueños de pequeñas empresas pasan más tiempo luchando con la burocracia que haciendo negocios. Más del 70% de lo que hacen está relacionado con el trabajo administrativo. ¿No es eso decepcionante?” (Revista Semana, 2020) y es que el trámite para acceder a los créditos bancarios va desde presentación de cifras con diferentes cortes firmadas por el revisor fiscal, reclasificación de estas, composición accionaria, y en muchos casos, flujos de caja del proyecto a financiar. La obtención de cifras para las empresas grandes que tienen departamentos de finanzas y de contabilidad estructurados es cuestión de un par de días, sin embargo, para una microempresa, en donde el dueño es el mismo financiero y en muchas ocasiones no tiene el conocimiento financiero requerido, la preparación de los requerimientos exigidos por las entidades bancarias se convierten en impedimentos.

5.3 Ingeniería Analítica Predictiva

De acuerdo con el Institute of Industrial and Systems Engineers, IISE, la ingeniería analítica es el “aprovechamiento de los datos para la obtención de ideas procesables” y su uso tiene aplicaciones entre otras, segmentación de clientes para una mejor utilización de recursos, seguimiento del rendimiento para el diagnóstico y mantenimiento, estrategias de clasificación para una mejor gestión del inventario, así como un mejor uso de los datos para incrementar ventas y así obtener mayores ingresos y mayor satisfacción de clientes (IISE, 2019). Este proceso va desde la recolección de los datos, la interpretación de estos y el preguntarse las preguntas adecuadas. En cuanto al procesamiento de datos, este va desde la recolección hasta la identificación de tendencias, clasificación y reconocimiento de patrones, esto se hace gracias al uso de machine learning y minería de datos (Politechnic, 2020).

En cuanto al desarrollo de productos, Siemens utiliza las capacidades predictivas de la ingeniería analítica, y lo describe de la siguiente manera: “El análisis predictivo de la ingeniería analítica es la aplicación de simulación y prueba de ingeniería multidisciplinaria con informes inteligentes y

análisis de datos, para desarrollar gemelos digitales que puedan predecir el comportamiento de los productos en el mundo real durante todo el ciclo de vida del producto. La ingeniería analítica predictiva incluye las tácticas y herramientas que los fabricantes pueden aprovechar para expandir la verificación y validación del diseño tradicional en un papel predictivo en apoyo del desarrollo de productos impulsado por sistemas. El objetivo final de implementar una estrategia de análisis de ingeniería predictiva es ofrecer innovación para productos complejos más rápido y con mayor confianza” (Siemens, 2020)

Dentro de las ventajas de la ingeniería analítica predictiva que Siemens menciona se encuentran (Siemens, 2020):

1. el ofrecimiento de modelos de fidelidad múltiple, que se encuentran lo más apegado a la realidad, durante cada etapa de desarrollo de los productos.
2. Ayuda a escoger los mejores conceptos y modelos de sistemas para llevar a cabo el proyecto.
3. Los procesos de simulación que influyen el diseño son más rápidos
4. Equilibrio de los requisitos multidisciplinarios
5. Mantiene los modelos sincronizados con el producto real en uso, de manera que las simulaciones son más realistas.
6. Aprovechamiento de los enfoques de simulación y prueba en conjunto para obtener una mayor confianza en el cumplimiento de los requisitos
7. El uso del análisis de los datos con las simulaciones en ambiente real que permiten la captura de conocimientos que llevan a la innovación.

5.4 Árboles de decisión como modelos de predicción

Una herramienta de la ingeniería analítica para el estudio del fracaso empresarial es el análisis mediante árboles de decisión. Los árboles de decisiones son un tipo de inteligencia artificial supervisada, en donde cada nodo evalúa el atributo de la muestra, mientras que las ramas muestran el resultado de dicha evaluación. Las reglas de clasificación provienen de la evaluación

de cada nodo o atributo y lo que pretenden los árboles de decisión, es evaluar la probabilidad de que un registro pertenezca o no a un grupo o clasificación predeterminado (Au, 2018), que para el caso de este estudio será el fracaso empresarial.

El tipo de árboles de decisión que se utilizaran en este trabajo son los CART, Classification and regression tree, los cuales muestran un puntaje en cada rama una vez realizada la partición, de manera que se facilita aún más la interpretación. Existen varios criterios de partición o división de los árboles de decisión, tales como Índice de Gini, entropía, test de chi cuadrado y proporción de ganancia de información. El criterio de división utilizado en la partición de los árboles CART de este trabajo es el índice de Gini. Se utiliza este criterio porque es apropiado en los problemas de selección, como el que se está tratando en este trabajo (Wang & Liu, 2020).

El Índice de Gini pretende reducir la impureza o incrementar la homogeneidad, o que los siguientes dos nodos tengan la mayor pureza posible, esto se hace buscando la probabilidad de no sacar dos registros t de la misma clase de nodo. Así, (Hurtado, 2012),

$$Gini\ t = 1 - \sum_{c \in C} P_{t,c}^2$$

C es el conjunto de clases y $P_{t,c}$ es la probabilidad de ocurrencia de c en el nodo t (Hurtado, 2012).

Una ventaja de esta herramienta es que no requiere la definición de supuestos sobre la distribución de la muestra, además que los parámetros no se seleccionan ni se optimizan. Otra gran ventaja es que la interpretación se hace fácil, sobre todo si se compara con las redes neuronales, otra de las herramientas utilizadas en los análisis de predicción (Marek, Durika, Jaroslav, Frnda, Svabova, 2019).

5.5 Random Forest como modelos de predicción

Random Forest o bosques aleatorios en español es una combinación de árboles de decisión individuales que funcionan como un conjunto, en donde cada árbol de decisión que pertenece al bosque aleatorio realiza una predicción, de manera que la predicción con más “votos” se constituye en la predicción para el bosque aleatorio (Breiman, 2001).

El fundamento detrás de los árboles de decisión es lo que comúnmente se llama la “sabiduría de la multitud” (Yiu, 2019).

El algoritmo de árboles de decisión funciona así:

1. Se toman n muestras aleatorias de tamaño N
2. Se construye un árbol de decisión para cada muestra aleatoria
3. La predicción de los datos se realizará a través de los atributos con mayor cantidad de votos para cada clase o nivel.

Para que el Random Forest tenga un mayor poder de predicción se requiere que la muestra contenga atributos con capacidad de predicción, así como datos no correlacionados.

El error estimado de la predicción en el Random Forest se puede calcular a través de la medición de los datos que no cayeron dentro de la muestra o como se conoce comúnmente como “out of bag predictions”, de manera que la tasa de error de los datos “out of bag” se constituye una medida de error estimado (Liaw & Wiener, 2001).

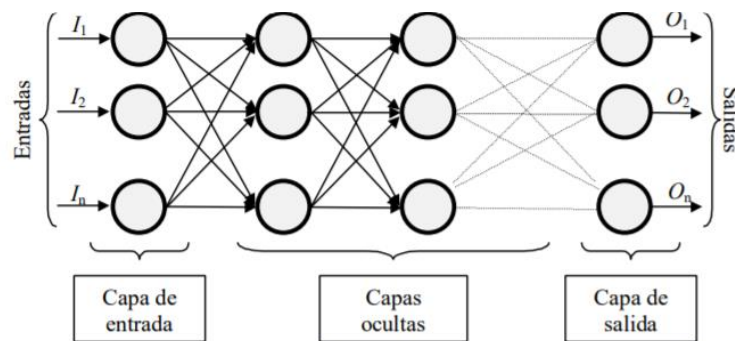
5.6 Redes Neuronales como Modelo de Predicción

El estudio de fracaso empresarial y los modelos de predicción de quiebra se pueden dividir en dos corrientes, modelos estadísticos y modelos de inteligencia artificial (Lee Sangjae, Sung Choi, 2013). Algunos de los autores que utilizaron modelos estadísticos y econométricos son William Beaver (Beaver, 1966), Edward Altman (Altman, 1968), James Ohlson (Ohlson, 1980), Fredy

Romero (Romero Espinosa, 2013), Norma Patricia Caro (Caro, 2016). Los modelos de redes neuronales han tenido gran acogida en los últimos años dentro del grupo de investigadores, algunos de ellos son Lee et al., (Lee Sangjae, Sung Choi, 2013), Diana Escandón et al., (Escandon, Diana, Hurtado, 2017), (Chauhan, Ravi, & Karthik Chandra, 2009), Fredy Pérez (Castaño, Horacio, Perez, 2007).

Las redes neuronales son “un nuevo sistema para el tratamiento de la información, cuya unidad básica de procesamiento está inspirada en la célula fundamental del sistema nervioso humano, la neurona” (Ruiz & Basualdo, 2001). Dentro de las diferentes clases de redes neuronales, la más utilizada es “Back Propagation” que dibuja la función de mapeo entre la entrada y la salida de un conjunto de datos proporcionado. La red contiene una capa de entrada, una o dos capas ocultas y una capa de salida. Las unidades de salida de una capa son las unidades de entrada de la siguiente capa. Los modelos de redes neuronales ayudan a solucionar problemas no lineales y el objetivo de la retro-propagación es minimizar la suma del error al cuadrado entre el valor real y el valor predicho (Lee Sangjae, Sung Choi, 2013), o minimizar la diferencia entre el vector de salida de la red y el vector deseado (Rumelhart, Hinton, & Williams, 1986).

Ilustración 3. Redes Neuronales Back Propagation



Fuente: (Ruiz & Basualdo, 2001)

Las Redes Neuronales como modelos de capacidad predictiva han demostrado ser más acertadas que los modelos convencionales, por ejemplo, De Llano et al., (De Llano Monelos et al., 2016) realizaron un estudio de comparación de las técnicas de predicción de los diferentes modelos utilizados en la literatura para predecir el fracaso empresarial, llegando a la conclusión que las redes neuronales son “las candidatas idóneas para el desarrollo de instrumentos de análisis y pronósticos” con una tasa de acierto del modelo del 94% en la fase de entrenamiento y del 80,5% en la fase de validación. Los resultados de la investigación de García Salgado et al., (García Salgado & Morales Castro, 2016) acerca del desempeño financiero de las empresas, mostraron que las redes neuronales tienen un mejor pronóstico de evaluación y clasificación que los resultados obtenidos por los modelos Logit.

Existen cinco tipos de redes neuronales artificiales, son las siguientes (Calvo, 2017):

1. Red Monocapa: consiste en una capa de neuronas de entrada, conectada a una capa de neuronas de salida.
2. Red Multicapas o MLP: Contienen capa de entrada, capa de salida y además capas ocultas.
3. Red neuronal convolucional: La principal diferencia es que las neuronas de cada capa no se conectan con todas las neuronas de las capas siguientes, sino con un subgrupo de estas, esto se realiza para disminuir la complejidad computacional utilizada.
4. Red Neuronal Recurrente: No existen unas capas de entrada, salida y ocultas definidas como en las redes multicapas, aquí las neuronas se conectan arbitrariamente.
5. Redes de base Radial: En estas redes, las capas ocultas se conectan a un mismo centro (pesos de la función). La diferencia con las MLP es que la distancia calculada es la distancia euclidiana entre el vector de pesos y la capa de entrada, mientras que en las redes MLP esta distancia es una función sigmoideal.

5.6.1 Algoritmo de ajuste de decaimiento dinámico

Variante de las redes de base radial RBF, Este algoritmo genera reglas basadas en datos numéricos. Cada regla se define como una función gaussiana de alta dimensión que se ajusta mediante dos umbrales, theta menos y theta más, para evitar conflictos con las reglas de diferentes clases (KNIME, 2020)

“El algoritmo de ajuste de decaimiento dinámico es un método de entrenamiento constructivo rápido para la red de funciones básicas radiales (RBF en inglés) cuando se usa para la clasificación” (Ke-Lin Du, 2019).

La gran ventaja de esta técnica es que usualmente se obtienen clasificaciones más precisas que para aquellas técnicas basadas en MLP (Multi Layer Perceptrons), además el entrenamiento se hace más rápido, comparado con back propagation.

Las ventajas de este algoritmo son (Niels, 1995):

1. Se agregan nodos RBF (redes de función básicas radiales) cuando es requerido.
2. La red se construye desde cero, el número de capas ocultas se define durante el entrenamiento de la red.
3. Entrenamiento rápido
4. Siempre hay un final en el entrenamiento y este se identifica claramente
5. Las clasificaciones incorrectas se encuentran por debajo del umbral theta (-) y las correctas por encima del umbral theta (+)
6. Los umbrales theta (-) y theta (+) se definen manualmente

5.6.2 Resilient back propagation

Variante del tipo MLP, Resilient back propagation o RPROP, es un algoritmo que se utiliza para entrenar redes neuronales y es similar al método comúnmente utilizado que es back propagation, explicado arriba. Tiene dos ventajas, es más rápido, y no requiere que especifique ningún valor de parámetro (Igel & Hüsken, 2002).

Esta variante del modelo original back propagation, ha sido ampliamente utilizada ya que, no requiere una tasa inicial de aprendizaje. El modelo es sensible a dicha tasa inicial. La principal diferencia con el modelo back propagation radica en que las derivadas parciales de la función del error se utilizan únicamente para corregir el sentido de los pesos de la red, lo que lo hace más rápido que los modelos tradicionales.

La técnica RPROP ha probado ser tolerante a las bases de datos que contienen ruido, por esta razón es que su uso se ha ampliado al análisis de situaciones con datos empíricos. “El algoritmo de optimización RPROP es considerado como uno de los algoritmos más poderosos para la estimación de los parámetros (o pesos) de una red neuronal” (Ortíz, Villa, & Velásquez, 2007).

5.7 La Regresión como Modelo de Predicción

5.7.1 Regresión Lineal Simple

En la regresión lineal simple, la variable dependiente es explicada por una variable independiente, y busca explicar la relación entre las dos variables, la fórmula que la representa es la siguiente:

$$y = B_0 + B_1X + u$$

Siendo y la variable explicada o dependiente y X la variable independiente o que explica la variable dependiente, aunque es ampliamente utilizado en la teoría, no es el tipo de regresión

que estaremos utilizando en este trabajo, ya que en nuestro modelo se tiene más de una posible variable explicativa.

5.7.2 Regresión Múltiple

El análisis de regresión múltiple es adecuado cuando se tiene más de una variable que explica a la variable dependiente y su uso se hace más importante cuando se están interpretando problemas empíricos. La siguiente es la ecuación de la regresión múltiple:

$$y = B_0 + B_1X_1 + B_2X_2 + \dots + B_nX_n + u$$

Siendo B_0 el intercepto, B_1 el coeficiente asociado a la variable independiente X_1 , B_2 el coeficiente asociado a la variable independiente X_2 y u el error de la ecuación.

En este caso se debe tener cuidado con la correlación en las variables independientes que invaliden el modelo.

El valor de R^2 es un indicador del funcionamiento del modelo y representa el porcentaje de ajuste al mismo. Es una medida que va entre cero y uno y entre más cercana a uno significa que la explicación de la variable dependiente a través de las variables independientes es mejor.

En cuanto a la interpretación de los coeficientes, estos representan los cambios en la variable dependiente ante una unidad de cambio en los coeficientes, asumiendo que todas las demás variables permanecen constantes.

Es importante evaluar los p-value de cada coeficiente para revisar si son o no significativos, de manera que si el p-value > 0.05 (nivel de alfa común) el coeficiente no es estadísticamente significativo.

5.7.3 Modelo de probabilidad lineal

En los métodos revisados anteriormente, tanto la variable dependiente como la variable independiente tenían un valor cuantitativo. Sin embargo, las variables binarias o dummy son igualmente importantes para describir variables de carácter cualitativo.

Dado que las variables independientes del modelo de este trabajo son los indicadores financieros, y estos no son bivariados, pero la variable dependiente es binaria, (fracasada o no), se explicará el caso en el que la variable dependiente es dummy.

La ecuación que explica el modelo es la siguiente, muy parecida a la de regresión múltiple, sin embargo, la variable y es dummy.

$$y = B_0 + B_1X_1 + B_2X_2 + \dots + B_nX_n + u$$

La interpretación de los coeficientes en este caso se hace un poco diferente, como y es binaria, lo que se va a medir es la probabilidad de éxito o la probabilidad de que $y = 1$, y el coeficiente B_j mide la variación de la probabilidad de éxito al variar X_j , permaneciendo los demás factores constantes (Wooldridge, 2010)

5.8 Herramientas para el análisis de desempeño de los modelos

Kalra et al (Kalra, Kumar, & Das, 2020) utilizan para evaluar el desempeño de los diferentes modelos, la precisión, la tasa de verdaderos positivos, AUC (área bajo la curva) ROC. Se utilizarán las mismas medidas en este trabajo. A continuación, una breve explicación de cada una de ellas.

5.8.1 Curvas ROC y AUC

Las curvas ROC (Receiver operating characteristic curve) son una herramienta estadística que permite ver la relación de la sensibilidad del modelo en función del complemento de la especificidad, para distintos puntos de corte. Los propósitos de la curva ROC son varios, primero

determinar la precisión/exactitud del modelo, segundo, encontrar el punto en la curva con mayor sensibilidad y especificidad, y tercero, la comparación entre varios modelos, mediante el análisis del área bajo la curva ROC.

Teniendo en cuenta el problema estudiado en este trabajo, y para explicar los términos, tenemos lo siguiente:

1. Las empresas clasificadas como fracasadas por el modelo, y que en realidad se encuentran fracasadas, son los verdaderos positivos.
2. Las empresas clasificadas como fracasadas por el modelo, pero que en realidad son empresas activas, se consideran falsos positivos.
3. Las empresas clasificadas como activas (No fracasadas) por el modelo, y que en realidad se encuentran activas, son los verdaderos negativos.
4. Las empresas clasificadas como activas por el modelo, pero que en realidad se encuentran fracasadas, son los falsos negativos.

El área bajo la curva ROC (AUC) que se analizara más adelante en los modelos, lo que nos va a decir es, la capacidad del modelo de clasificar una empresa activa seleccionada aleatoriamente como activa y una fracasada también seleccionada aleatoriamente como fracasada.

De acuerdo con las definiciones anteriores tenemos que, la sensibilidad y especificidad se calculan como se muestra en las siguientes formulas:

$$\text{Sensibilidad} = \frac{\textit{Verdaderos positivos}}{\textit{Verdaderos Positivos} + \textit{Falsos Negativos}}$$

$$\text{Especificidad} = \frac{\textit{Verdaderos Negativos}}{\textit{Verdaderos Negativos} + \textit{Falsos Positivos}}$$

5.8.2 Precisión

La precisión del modelo, que es otra medida para tener en cuenta en la selección del mejor modelo, se calcula así:

$$\text{Precisión/Exactitud} = \frac{\textit{Verdaderos positivos} + \textit{Verdaderos Negativos}}{\textit{Total de datos}}$$

La precisión o exactitud, evalúa la capacidad del modelo de clasificar correctamente los datos.

Teniendo en cuenta la restricción de datos que se puede observar en algunos sectores de la economía, como por ejemplo agricultura; en este trabajo, se evaluará la precisión del modelo con la validación cruzada dejando uno por fuera, o Leave One Out Cross Validation LOOCV.

Este método de validación pretende estimar el error del modelo de una manera mucho más certera, ya que, evalúa toda la muestra N dejando un solo dato para la prueba del modelo. De manera que, para el entrenamiento de este, se van a utilizar N-1 datos. Adicionalmente, se van a realizar N modelos dejando en cada caso para la validación de este, un dato diferente no utilizado anteriormente. Se pretende con este método, tener una mejor evaluación de la precisión de los modelos

El procedimiento es el siguiente (Zeng, Sun, & Farnham, 2020):

1. Dejar un dato de lado para la evaluación del modelo
2. Utilizar para el entrenamiento del modelo los datos restantes, (n-1)
3. Evaluar la precisión del modelo
4. Se repiten los modelos n veces, utilizando en cada caso para la evaluación un dato diferente.
5. El error cuadrático del modelo será el promedio del cuadrado de los errores de los n modelos, así:

$$\text{Promedio de error al cuadrado o MSE} = \frac{1}{n} \sum_{n=1}^n (O_n - P_n)^2$$

Siendo O_n el dato observado para el modelo n y P_n el dato predicho para el modelo n .

5.8.3 Tasa de verdaderos positivos:

Mide la cantidad de datos positivos clasificados como tal, dentro del universo de datos positivos, la fórmula es:

$$\text{Tasa de verdaderos positivos} = \frac{\textit{Verdaderos positivos}}{\textit{Total de datos positivos}}$$

6. METODOLOGIA

La Metodología de recolección y procesamiento de información de este trabajo de investigación seguirá la del trabajo de William Beaver (Beaver, 1966). Se seleccionará una empresa activa por cada empresa fracasada que pertenezcan al mismo sector productivo.

La información que se utiliza en este proyecto son los balances y estados de resultados de las microempresas reportada a la Super Intendencia de Sociedades y que se encuentra en el SIREM, Sistema de información y reporte empresarial.

Como la información no se encuentra clasificada para micro, pymes o grandes empresas, la primera actividad a realizar será la identificación de las microempresas, para esto se organizarán los activos totales de las empresas de menor a mayor, (información de la hoja de balance) de manera que todas aquellas empresas con activos inferiores a 500 SMMLV para cada año, desde el 2013 hasta el 2019, serán las microempresas en Colombia.

La Tabla 1 muestra la evolución de salarios SMMLV por año, que servirá de guía para clasificación de información (BANREP, 2018)

Tabla 1. Evolución del salario mínimo en Colombia

AÑO	SALARIO MÍNIMO	MONTO MAXIMO PARA SER CLASIFICADA COMO MICROEMPRESA
2013	\$ 589,500	\$ 294,750,000
2014	\$ 616,000	\$ 308,000,000
2015	\$ 644,350	\$ 322,175,000
2016	\$ 689,455	\$ 344,727,500
2017	\$ 737,717	\$ 368,858,500
2018	\$ 781,242	\$ 390,621,000
2019	\$ 828,116	\$ 414,058,000
2020	\$ 878,000	\$ 439,000,000

Fuente: Elaboración propia con información de (BANREP, 2018)

De otro lado, Colombia, adopto el código industrial internacional uniforme, que clasifica las empresas según su actividad económica (DANE, 2012), y dichas empresas se pueden además clasificar en tres grandes sectores de la economía colombiana, dichos sectores son el primario, secundario y terciario o, agricultura, industrial y servicios (Universidad Católica de Oriente, 2013).

El sector primario o de Agricultura, incluye a todas aquellas actividades relacionadas con la explotación de los recursos naturales. Aquí se incluyen, agricultura, caza, pesca y silvicultura.

El sector secundario o industrial, incluye las actividades que transforman los bienes extraídos del sector primario y sector terciario o de servicios, se refiere a la producción de bienes intangibles (Universidad Católica de Oriente, 2013).

Lo anterior permite clasificar los códigos CIU de las empresas colombianas en los tres grandes sectores de la economía (DANE, 2012), así:

Tabla 2. Clasificación por códigos CIU

SECCION	DIVISION	DESCRIPCION	SECTOR ECONOMICO
A	01_03	Agricultura, ganadería, caza, silvicultura y pesca	Agricultura
B	05_09	Explotación de minas y canteras	Industrial
C	10_33	Industrias manufactureras	Industrial
D	35	Suministro de electricidad, gas, vapor y aire acondicionado	Industrial
E	36_39	Distribución de agua; evacuación y tratamiento de aguas residuales, gestión	Industrial
F	41_43	Construcción	Industrial
G	45_47	Comercio al por mayor y al por menor; reparación de vehículos automotores y motocicletas	Servicios
H	49_53	Transporte y almacenamiento	Servicios
I	51_56	Alojamiento y servicios de comida	Servicios
J	58_63	Información y comunicaciones	Servicios
K	64_66	Actividades financieras y de seguros	Servicios
L	68	Actividades inmobiliarias	Servicios
M	69_75	Actividades profesionales, científicas y técnicas	Servicios
N	77_82	Actividades de servicios administrativos y de apoyo	Servicios
O	84	Administración pública y defensa; planes de seguridad social de afiliación obligatoria	Servicios
P	85	Educación	Servicios
Q	86_88	Actividades de atención de la salud humana y de asistencia social	Servicios
R	90_93	Actividades artísticas, de entretenimiento y recreación	Servicios
S	94_96	Otras actividades de servicios	Servicios
T	97_98	Actividades de los hogares en calidad de empleadores; actividades no diferenciadas de los hogares individuales como productores de bienes y servicios para uso propio	Servicios
U	99	Actividades de organizaciones y entidades extraterritoriales	Servicios

Fuente: (DANE, 2012)

La segunda actividad que se realizará es la clasificación de las empresas por sectores: Los balances y estados de resultados de SIREM tienen una columna con el código CIU, de manera que la agrupación se basará por la clasificación de la DIAN por sectores económicos, que resulta de la agrupación de códigos de actividad CIU en sectores económicos, tal como lo muestra la tabla 2.

Los indicadores financieros que van a ser parte del modelo serán todos aquellos que son de relevancia, de acuerdo con el marco teórico de este trabajo. Dichos indicadores serán de liquidez, indicadores de eficiencia, indicadores de desempeño, indicadores de productividad, de endeudamiento, solvencia y de diagnóstico financiero.

Los cálculos de estos indicadores se realizarán en Excel, con la información obtenida de SIREM siguiendo las fórmulas de los indicadores financieros más relevantes para las empresas dependiendo del sector productivo al cual pertenece.

La información que se utilizará serán los estados financieros de las microempresas liquidadas y que hayan durado en el mercado más de cuatro años. Esto con el fin de controlar y evitar el efecto de la quiebra de la microempresa recién creada, ya que, la tasa de mortalidad aquí es mayor (De Llano Monelos et al., 2016).

Como métodos de validación de los diferentes modelos aquí estudiados se utilizará información financiera adicional que no haya sido usada como input para entrenar el modelo, así como validación con expertos en la gestión técnica y financiera de las microempresas.

Es importante tener en cuenta el año en que la empresa se liquidó, canceló o fue el último año de renovación de la matrícula mercantil. Si la empresa se encuentra activa en el RUES (Registro Único Empresarial) pero su último año de renovación de matrícula mercantil fue en 2016, también será objeto de estudio. Por lo anterior, serán consideradas como microempresas fallidas o quebradas todas aquellas que de una muestra aleatoria entre 2011 y 2016 aparezcan ante la RUES como en liquidación, canceladas o activas, estas últimas deben cumplir que su último año de renovación de matrícula es inferior a 2016 inclusive. y que además cumplan con el tamaño de activos que por ley definen las microempresas en Colombia.

De otro lado, todas aquellas empresas que se encuentren activas ante la RUES, pero con último año de renovación de matrícula mercantil entre 2017 y 2018, no serán tenidas en cuenta dentro

de la base de información para el estudio, mientras que las microempresas que se encuentren activas y con último año de renovación de 2019, serán consideradas como empresas sanas o no fallidas.

El periodo t es el año en el que la empresa dejó de funcionar o se liquidó. Conocido ese año, se obtendrá la información del periodo $t-1$, de acuerdo con Altman et al., (E. Altman et al., 2017) de la página SIREM, que es donde publican la información de los estados financieros de todas las empresas que reportan información ante la Superintendencia de sociedades.

Antes de realizar la búsqueda de los estados financieros del periodo $t-1$, se debió verificar en la página de la RUES, Registro Único Empresarial, que la matrícula mercantil de la empresa estuviera cancelada o que la empresa estuviera en Liquidación, o que cumpliera los requisitos arriba descritos para la clasificación de microempresas fallidas.

7. IDENTIFICACION DE LAS VARIABLES FINANCIERAS POR SECTOR

A la hora de escoger que variables financieras utilizar, se realizó una revisión de literatura por sector industrial para las empresas en diferentes países, sobre todo en Colombia y en países en vía de desarrollo, de manera que la problemática que se quiere estudiar se llevase a cabo en condiciones similares.

7.1 Variables financieras para el Sector Primario: Agricultura

El sector agrícola en Colombia representó el 9% del PIB para 2016 y genera cerca de 2.1 millones de empleos directos (Villanueva Mejía, 2018). En cuanto al tamaño de las empresas del sector, el 48,2% son microempresas mientras que el 41,8% son empresas unipersonales (Ministerio de Agricultura, 2016). Este sector se caracteriza por su alta informalidad, de hecho según la encuesta Integrada de Hogares de 2016, la cifra de informalidad para el mismo año en este sector fue del 84.9% (DANE, 2016). Según el ministerio de Agricultura, el sector enfrenta tres clases de riesgos

que afectan la estabilidad de sus empresas, el primero tiene que ver con los riesgos de producción, que agrupa los cambios climáticos, la aparición de las enfermedades en las plantas o animales y demás temas referentes a la mano de obra y producción. El segundo tipo de riesgos son los asociados al mercado, el cambio en los precios, oferta o demanda de los productos y el tercer tipo de riesgos son los de contexto. Estos últimos se refieren a los riesgos inherentes a la comercialización de los productos, como entorno legal, paros camioneros, entre otros (Ministerio de Agricultura, 2018). Estos riesgos afectan los indicadores financieros de las empresas de este sector, de una manera diferente a las empresas que pertenecen al sector industrial o de servicios, por lo que las variables financieras escogidas para analizar las microempresas de agricultura en el país deben tener en cuenta lo anteriormente descrito.

En un análisis del consejo privado de competitividad para el sector agropecuario, se llega a la conclusión de que el bajo crecimiento del sector se debe al bajo retorno esperado del proyecto y/o el costoso financiamiento de la inversión, estos dos factores son explicados a su vez por el nivel educativo de los trabajadores de las empresas de este sector, la mala distribución o asignación de tierras, el no aprovechamiento de las economías de escala, estructura deficiente, la falta de acceso a mercados externos y problemas de financiamiento, (Consejo privado de Competitividad, 2017).

De otro lado, Uribe et al. (Uribe, Fonseca, Bernal, Contreras, & Castellanos, 2010), indican que es necesario incluir indicadores cuantitativos para medir el impacto del direccionamiento estratégico y su aporte al sector agrícola, dichos indicadores son de valoración y tecnología, de valoración económica e indicadores de conocimiento (Uribe et al., 2010).

Un estudio comparativo entre el sector agro de Colombia y Brasil para evaluar el desempeño financiero del sector en cada país, incluye dentro de los ratios analizados, indicadores de rentabilidad, EBITDA, Indicadores de endeudamiento, Indicadores de eficiencia e indicadores de liquidez (Gil, Cruz, & Yaelt, 2018).

Teniendo en cuenta la revisión de la literatura anterior, los indicadores financieros que se utilizarán para realizar el análisis de las microempresas del sector agrícola en Colombia son:

Tabla 3. Indicadores financieros sector Agricultura

TIPO	VARIABLE	CALCULO
Agricultura	Coeficiente de deuda corto plazo CP	Deuda corriente/Pasivo total
	Coeficiente de deuda largo plazo LP	Deuda no corriente/Pasivo total
	EBITDA	Utilidad Operativa + depreciaciones + amortizaciones
	Indicador Apalancamiento	Deuda Total/Patrimonio Total
	Ratio de Costo financiero	intereses pagados/Pasivo total
	Ratio de deuda	deuda/pasivo total
	Ratio de eficiencia	costo de ventas/ingresos
	ROA	Utilidad/Activos totales
	Equity Strength	Patrimonio/Total pasivo
	Indicador de capital de trabajo	Activos corrientes - Pasivos corrientes / activos totales
	ROE	Utilidad/Patrimonio

Fuente: Elaboración Propia

7.2 Variables financieras para el Sector Secundario: Industrial

Dentro del sector industrial se encuentran varios subsectores de producción, entre ellos están, manufactura, construcción, suministro de electricidad y evacuación de agua. Sin embargo, las microempresas en Colombia se ubican en su mayoría en dos subsectores: manufactura y construcción. La principal razón del porque no observamos más microempresas en subsectores como suministro de electricidad o evacuación de agua, es por el tamaño de los activos, es decir, para realizar este tipo de actividades, la inversión en infraestructura y adecuaciones de planta sobrepasa el tamaño de los activos de las microempresas.

En cuanto a los factores que explican el comportamiento de la quiebra de las microempresas en el sector manufacturero, un estudio realizado para empresas de este sector en Indonesia, muestra que el flujo de caja operacional tiene un impacto positivo en el fracaso empresarial, mientras que el retorno sobre los activos o ROA, tiene un impacto negativo, es decir, a mayor ROA menor probabilidad de fracaso (Finishtya, 2019). Lo anterior es debido a que, a mayor flujo

de caja operativo, la compañía puede sortear más fácilmente una crisis financiera y a mayor ROA, mayores utilidades por lo que es menor la probabilidad de una crisis.

Otro estudio para empresas del sector manufacturero en el mismo país demostró en 2018 que la liquidez y las ganancias retenidas no tienen impacto cuando se trata de predecir fracaso empresarial, sin embargo, el EBIT (Earnings Before Interest) y ROE (Return on Equity) si tienen impacto a la hora de predecir dificultades financieras (Tya, Restianti, Agustina, 2018).

Además de los indicadores tradicionales utilizados en los análisis de predicción de fracaso empresarial, las tendencias de las ventas, de los ingresos o de los activos también han sido tenidas en cuenta (Bae, 2012).

De otro lado, para analizar el comportamiento del sector de la construcción, (subsector de Industrial) es necesario tener en cuenta que, los indicadores de endeudamiento son muy altos y los retornos de la inversión se ven hasta las fases finales de los proyectos además son empresas sensibles a las crisis económicas (Heo & Yang, 2014). En China, otro país emergente, para diferenciar las empresas solventes de las insolventes, del sector de la construcción, se realizó un modelo que tenía en cuenta indicadores de liquidez, rentabilidad, apalancamiento y de actividad, los indicadores utilizados por Thomas et al son en total 22, sin embargo, los indicadores con mayor poder de predicción son, para medir actividad, ventas netas/activo corriente y capital de trabajo/total activos; para medir rentabilidad, ventas netas/utilidad y ROA, y para medir solvencia, razón corriente y EBIT/gastos financieros (Thomas Ng, Wong, & Zhang, 2011).

Aunque los subsectores de manufactura y construcción tienen comportamientos diferentes dentro del ciclo económico colombiano, (siendo estos los sectores impulsados por el gobierno a la hora de aumentar el gasto publico) los indicadores que, según la literatura revisada, han servido para predecir dificultades financieras son los mismos, con una o dos variaciones puntuales. Teniendo en cuenta lo anterior, los indicadores financieros que se tendrán en cuenta para realizar el cálculo de las microempresas que pertenecen al sector industrial de la economía colombiana son:

Tabla 4. Indicadores financieros sector Industrial

TIPO	VARIABLE	CALCULO
Industrial	EBIT	Utilidad + Impuestos + Intereses
	Indicador Liquidez	Disponible/Activo Total
	Flujo de caja operativo FCO	EBIT + Amortización + Impuestos
	Indicador de capital de trabajo	Activos corrientes - Pasivos corrientes / activos totales
	Indicador de actividad	(ventas netas) /activos corrientes
	Indicador de solvencia	EIBT/Gastos financieros
	ROS	Utilidad/ventas netas
	Razón corriente	activos corrientes - inventarios / pasivo corriente
	ROA	Utilidad/Activos totales
	ROE	Utilidad/Patrimonio
	EBIT	Utilidad + Impuestos + Intereses
	Indicador Apalancamiento	Deuda Total/Patrimonio Total

Fuente: Elaboración Propia

7.3 Variables financieras para el Sector Terciario: Servicios

El sector terciario de la economía es el sector que más número de microempresas contiene, la explicación a este hecho también tiene que ver con el tamaño de los activos. Las empresas de servicios se caracterizan en su mayoría, por no poseer grandes cantidades de inventario ni hacer uso de grandes maquinarias. Este sector compone más del 50% del PIB nacional y es quizá la mayor apuesta de los inversionistas en el país. De acuerdo con la gran encuesta Pyme de 2018, del universo de empresas encuestadas, el 90% de estas pertenecía al sector servicios (61% a comercio y 29% servicios).

En el análisis a la gran encuesta de microempresarios de 2018, se puede observar un sector microempresarial con acceso limitado a créditos, baja capacidad de ahorro, baja capacidad de diversificación y alta informalidad(LA REPUBLICA, 2018).

En cuanto a las variables que explican el comportamiento de las microempresas en Bogotá para el sector servicios específicamente, Duarte (Duarte, 2019) utiliza el siguiente set de variables del siguiente cuadro:

Ilustración 4. Set de variables utilizado por (Duarte, 2019)

Variable	Descripción
X1	Activo corriente / Pasivo corriente
X2	Pasivo total / Activo total
X3	Ventas / Activo total
X4	Disponible / Activo total
X5	Disponible / Pasivo corriente
X6	(Activo corriente - Inventario) / Pasivo corriente
X7	Capital de trabajo / Activo total
X8	EBIT / Activo total
X9	Pasivo total / Patrimonio
X10	Utilidades acumuladas / Activo total
X11	Utilidad neta / Ventas
X12	Activo corriente / Activo total
X13	Capital de trabajo / Ventas
X14	Pasivo corriente / Patrimonio
X15	Patrimonio / Activo total
X16	Utilidad neta / Patrimonio
X17	Utilidad operacional / Activo total
X18	(Cuentas por cobrar / Ventas)*360
X19	(Inventario / Costo de ventas)*360
X20	Activo corriente / Ventas
X21	Disponible / Ventas
X22	Efectivo / Ventas
X23	Flujo de caja / Pasivo total
X24	Flujo de efectivo generado en la operación / Activo Total
X25	Obligaciones financieras / Activo total
X26	Pasivo no corriente / Activo total
X27	Patrimonio / Pasivo total
X28	Utilidad neta / Activo total
X29	Utilidad antes de impuestos / Activo total
X30	(Activo corriente - Pasivo corriente) / Activo total

Fuente: (Duarte, 2019)

De las cuales encuentra que el EBIT (indicador de rentabilidad), Disponible/Activo Total (indicador de liquidez), utilidad antes de impuestos /activo total (indicador de rentabilidad) y utilidad

neta/activo total (indicador de rentabilidad) son las variables que mejor explican las diferencias entre empresas fracasadas y no fracasadas (Duarte, 2019).

De lo anterior es importante resaltar que indicadores de Liquidez y de rentabilidad son los que mejor explican la situación económica complicada de las microempresas de servicios en Bogotá, esto coincide con el estudio realizado para pymes en Letonia del sector servicios, en el que indicadores de rentabilidad y de deuda fueron los principales predictores del fracaso empresarial (Kotane & Kuzmina-Merlino, 2017).

En Malasia, otro país emergente, para las empresas comerciales y de servicios, se encontró que dentro de los principales indicadores financieros que impactan en la predicción de fracaso empresarial se encuentran el indicador de deuda, indicador de capital de trabajo, índice de rotación de activos totales, ventas netas/activos totales e indicador de deuda (Alifiah, 2014).

Teniendo en cuenta la revisión de literatura para empresas del sector servicios en países emergentes y en Colombia, las variables a utilizar para el estudio de las microempresas de este sector son las que el siguiente cuadro muestra:

Tabla 5. Indicadores financieros sector Servicios

TIPO	VARIABLE	CALCULO
Servicios	Indicador de actividad	ventas/activos corrientes
	Indicador Liquidez	Disponible/Activo Total
	Indicador Rentabilidad_ activos	Utilidad Neta/Activo Total
	ROS	Utilidad/ventas netas
	Ratio de deuda	deuda/Pasivo Total
	EBITDA	Utilidad Operativa + depreciaciones + amortizaciones
	Indicador Rentabilidad_ activos	Utilidad Neta/Activo Total
	ROA	Utilidad/Activos totales
	ROE	Utilidad/Patrimonio

Fuente: Elaboración Propia

Esta es la lista completa de indicadores que se utilizaran en los diferentes modelos:

Tabla 6. Resumen de los indicadores financieros

TIPO	VARIABLE	CALCULO
Agricultura	Coeficiente de deuda corto plazo CP	Deuda corriente/Pasivo total
	Coeficiente de deuda largo plazo LP	Deuda no corriente/Pasivo total
	EBITDA	Utilidad Operativa + depreciaciones + amortizaciones
	Indicador Apalancamiento	Deuda Total/Patrimonio Total
	Ratio de Costo financiero	intereses pagados/Pasivo total
	Ratio de deuda	deuda/pasivo total
	Ratio de eficiencia	costo de ventas/ingresos
	ROA	Utilidad/Activos totales
	Equity Strength	Patrimonio/Total pasivo
	Indicador de capital de trabajo	Activos corrientes - Pasivos corrientes / activos totales
	ROE	Utilidad/Patrimonio
Industrial	EBIT	Utilidad + Impuestos + Intereses
	Indicador Liquidez	Disponible/Activo Total
	Flujo de caja operativo FCO	EBIT + Amortización + Impuestos
	Indicador de capital de trabajo	Activos corrientes - Pasivos corrientes / activos totales
	Indicador de actividad	(ventas netas) /activos corrientes
	Indicador de solvencia	EBIT/Gastos financieros
	ROS	Utilidad/ventas netas
	Razón corriente	activos corrientes - inventarios / pasivo corriente
	ROA	Utilidad/Activos totales
	ROE	Utilidad/Patrimonio
	EBIT	Utilidad + Impuestos + Intereses
	Indicador Apalancamiento	Deuda Total/Patrimonio Total
Servicios	Indicador de actividad	ventas/activos corrientes
	Indicador Liquidez	Disponible/Activo Total
	Indicador Rentabilidad_ activos	Utilidad Neta/Activo Total
	ROS	Utilidad/ventas netas
	Ratio de deuda	deuda/Pasivo Total
	EBITDA	Utilidad Operativa + depreciaciones + amortizaciones
	Indicador Rentabilidad_ activos	Utilidad Neta/Activo Total
	ROA	Utilidad/Activos totales
	ROE	Utilidad/Patrimonio

Fuente: Elaboración Propia

8. RESULTADOS PARA TODA LA MUESTRA SIN CLASIFICAR POR SECTORES

En esta sección, se aplicaron todas las herramientas a la muestra completa sin clasificar por sectores. Se pretende observar que tan diferentes son los resultados por sectores de los que arrojen los modelos para las empresas sin clasificar por sectores.

Antes de evaluar la capacidad de predicción de los modelos y su desempeño, se realizó la prueba de correlación de Pearson¹. En todos los casos, donde la correlación era superior a 0.20 se eliminaron las variables causantes del problema. Al aplicar los criterios del filtro de correlación² del programa Knime³, las variables que se eliminaron de todos los modelos (Árboles Binario, Random Forest, DDA, RPROP) para la muestra sin clasificar por sectores fueron: ROE y EBITDA.

El desempeño de los modelos se evaluará teniendo en cuenta: AUC-ROC, tasa de verdaderos positivos, y precisión o exactitud. Los resultados son los siguientes:

Tabla 7. Desempeño de los diferentes modelos para toda la muestra sin clasificar por sectores

	Coeficiente de correlación Pearson R=0.20			
	Árboles Binarios	Random Forest	RN DDA	RN RPROP
AUC	0.66	0.719	0.694	0.605
Tasa de verdaderos positivos	77.37%	69.47%	72.87%	67.44%
Precisión	66.58%	66.58%	66.27%	57.94%
Indicadores	ROA, INDICADOR DE ACTIVIDAD, APALANCAMIENTO			
Variables eliminadas	ROE, EBITDA			

¹ Evalúa la correlación lineal entre dos variables. El coeficiente de correlación resultado de la prueba puede tomar valores entre -1 y 1.

² Función que evalúa iterativamente los indicadores que tienen más correlación con las demás variables. Una vez identificados, los elimina.

³ Software de libre descarga, utilizado en este trabajo.

8.1 Resultados del mejor modelo: Árboles Binarios

Para el caso del análisis de la muestra sin clasificar por sectores, el modelo con el mejor desempeño son los Árboles Binarios, ya que, dos de las tres variables de desempeño resultan ser las más altas. La ilustración 5 muestra la curva ROC, mientras que la ilustración 6 muestra la matriz de confusión de los Árboles Binarios. La precisión se calculó teniendo en cuenta el método de validación cruzada Leave One Out Cross Validation, explicado previamente en el marco teórico.

De los 380 datos utilizados para entrenar el modelo, 253 se clasificaron correctamente, mientras que 127 fueron clasificados de manera incorrecta. De estos 127, 84 corresponden a falsos positivos (empresas predichas como fracasadas que en realidad se encuentran activas) y 43 a falsos negativos (empresas predichas como activas que en realidad son fracasadas).

La tasa de verdaderos positivos es **77.37%**, y la precisión es **66.58%**.

Ilustración 5. Curva ROC Muestra sin clasificar por sectores, modelo Árboles Binarios

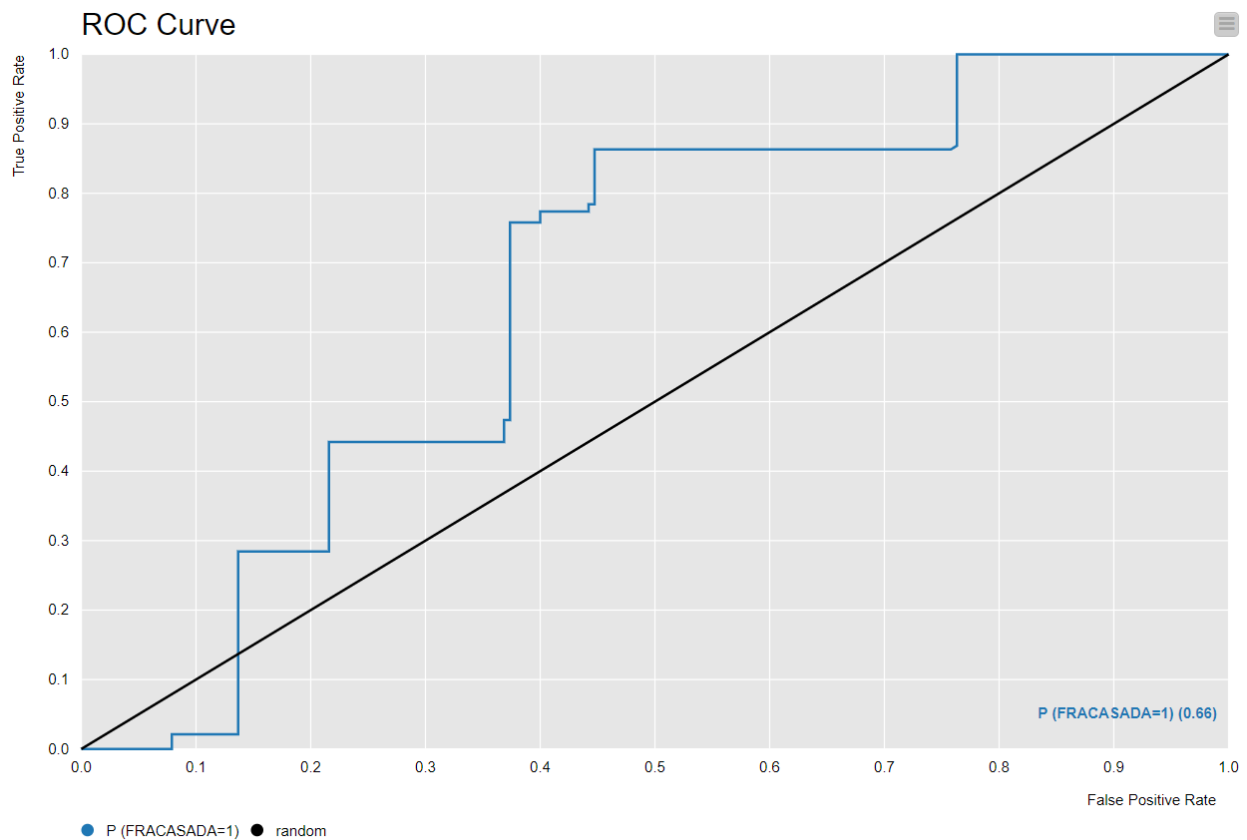


Ilustración 6. Matriz de confusión Árboles Binarios, toda la muestra

	0 (Predicted)	1 (Predicted)	
0 (Actual)	106	84	55.79%
1 (Actual)	43	147	77.37%
	71.14%	63.64%	

Overall Statistics

Overall Accuracy	Overall Error	Cohen's kappa (κ)	Correctly Classified	Incorrectly Classified
66.58%	33.42%	0.332	253	127

En cuanto a las particiones realizadas por los Árboles Binarios, en primer lugar, tenemos el ROA, luego el indicador de actividad y por último el indicador de apalancamiento.

Si el ROA es mayor a -0.2596, pero el indicador de actividad es inferior a 1.4038, el 50% de las microempresas fracasan, pero si además el indicador de apalancamiento es inferior a 0.002, el porcentaje de fracasos aumenta al 65.91%. Ver ilustración 7.

Se podrá observar más adelante, que, así como el ROA es el indicador que realiza la primera partición de los árboles binarios para toda la muestra sin clasificar por sectores, al analizar el sector servicios específicamente, este indicador también resulta ser el primero en realizar la primera división de la muestra; esto no es coincidencia, la base de datos en su mayoría tiene microempresas del sector servicios.

Ilustración 7. Árboles Binarios para toda la muestra



9. RESULTADOS SECTOR AGRICULTURA

Existe una particularidad con las empresas de este sector en Colombia, y es que la informalidad supera el 86% en el campo (El país, 2019), además si tenemos en cuenta que este porcentaje es aún más alto para las microempresas, encontramos la razón de la falta de información financiera para las microempresas en este sector. De manera que, las herramientas se construyeron bajo una restricción de información, esto es 16 empresas en total, de las cuales, 8 están activas y 8 fracasadas.

Antes de entrenar los diferentes modelos, se realizó la prueba de correlación de Pearson y se eliminaron las variables que tenían correlación de más del 20% con otras variables. Los indicadores eliminados en este caso fueron: ROA, Asset turnover, ROE, Ebitda, Razón Corriente.

De la misma manera que se realizó con la muestra sin clasificar por sectores, se va a evaluar los indicadores AUC-ROC, Tasa de verdaderos positivos y precisión. Los resultados son los siguientes:

Tabla 8. Desempeño de los diferentes modelos Sector Agricultura

	Coeficiente de correlación Pearson R=0.20			
Marco de Decisión	Árboles Binarios	Random Forest	RN DDA	RN RPROP
AUC	-	0.578	0.43	0.00%
Tasa de verdaderos positivos	25.00%	75.00%	50.00%	0.00%
Precisión	12.00%	56.25%	37.50%	16.67%
Indicadores		INDICADOR DE APALANCAMIENTO		
Variables eliminadas	ROA, ASSET TURNOVER, ROE, EBITDA, RAZON CORRIENTE			

9.1 Resultado mejor modelo: Random Forest

A continuación, se muestran el área bajo la curva y la matriz de confusión del mejor modelo, Random Forest. La ilustración 8 muestra la curva ROC, mientras que la ilustración 9 muestra la matriz de confusión. La precisión se calculó teniendo en cuenta el método de validación cruzada Leave One Out Cross Validation.

De los 16 datos utilizados para entrenar el modelo, 9 se clasificaron correctamente, mientras que 7 fueron clasificados de manera incorrecta. De estos 7, 5 corresponden a falsos positivos (empresas predichas como fracasadas que en realidad se encuentran activas) y 2 a falsos negativos (empresas predichas como activas que en realidad son fracasadas).

La tasa de verdaderos positivos es **75%**, y la precisión es **56.25%**.

Ilustración 8. Curva ROC Sector Agricultura, Random Forest

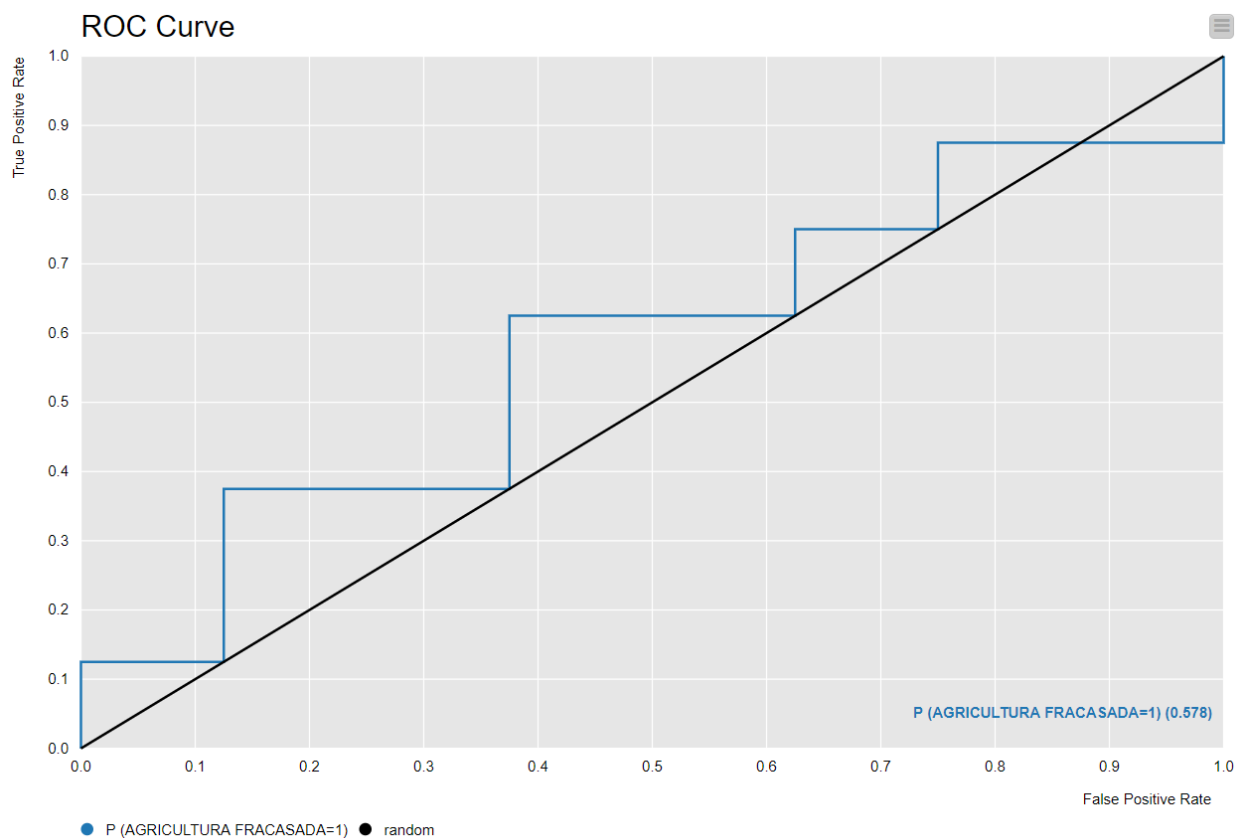


Ilustración 9. Matriz de confusión Random Forest, Sector Agricultura

Confusion Matrix		0 (Predicted)	1 (Predicted)	
0 (Actual)		3	5	37.50%
1 (Actual)		2	6	75.00%
		60.00%	54.55%	
Overall Statistics				
Overall Accuracy	Overall Error	Cohen's kappa (κ)	Correctly Classified	Incorrectly Classified
56.25%	43.75%	0.125	9	7

El principal indicador de predicción es el indicador de apalancamiento, en segundo lugar, el ratio de deuda. La tabla 9 muestra los resultado del modelo, mientras que la tabla 10 nos dice cuántas veces fueron escogidos efectivamente los diferentes indicadores, de las veces que fueron candidatos⁴, para realizar el Split⁵ de la muestra:

Tabla 9. Random Forest Sector Agricultura

Indicador Financiero	Splits nivel 1	Splits nivel 2	Splits nivel 3	Candidatos nivel 1	Candidatos nivel 2	Candidatos nivel 3
Indicador de Apalancamiento	20	19	7	32	66	64
Indicador de rentabilidad_Utilidad	9	7	3	30	48	65
Coeficiente de deuda de largo Plazo LP	6	8	5	30	57	66
Coeficiente de deuda de largo Plazo CP	5	2	1	26	47	61
Ratio de deuda	4	12	4	23	55	55
Ratio de eficiencia	4	6	5	24	50	63

⁴ Los candidatos del modelo son los indicadores financieros que fueron tenidos en cuenta para realizar el Split de la muestra.

⁵ Split: División de la muestra.

Tabla 10. Porcentaje de elección Random Forest

Indicador Financiero	Splits	Candidato	% elección
Indicador de Apalancamiento	46	162	28%
Ratio de deuda	20	133	15%
Coficiente de deuda de largo Plazo LP	19	153	12%
Indicador de rentabilidad_Utilidad	19	143	13%
Ratio de eficiencia	15	137	11%
Coficiente de deuda de largo Plazo CP	8	134	6%

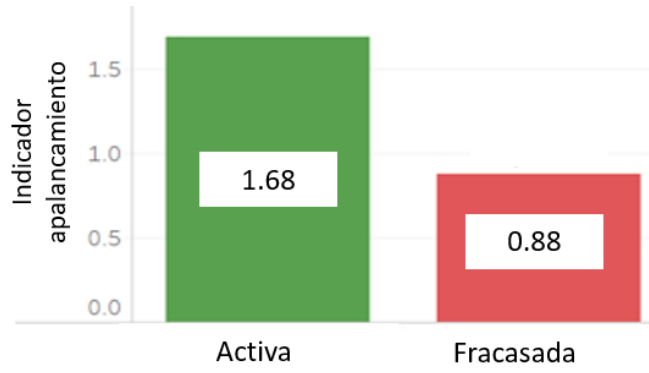
Fuente: Elaboración Propia

Analizando el indicador de apalancamiento, este nos muestra unos resultados sorprendentes, pues es mayor en empresa activas, siendo 1.6840 versus 0.8789 de las empresas fracasadas. Ver ilustración 10. la explicación a estos resultados se encuentra en el rubro de deuda total, perteneciente a los pasivos corrientes, siendo mayor para empresas activas. Lo que nos muestra que las empresas activas tienen un mayor porcentaje de deuda, quizá es gracias a la deuda que estas empresas han tenido acceso a capital para su producción. Ver Ilustración 11.

Estos resultados coinciden con los estudios realizados por Anif y Confecamaras para empresas MiPymes en donde el acceso a los créditos bancarios le permite al empresario desarrollar y crecer su empresa o cubrir gastos cuando las ventas no son suficientes (ANIF, 2019).

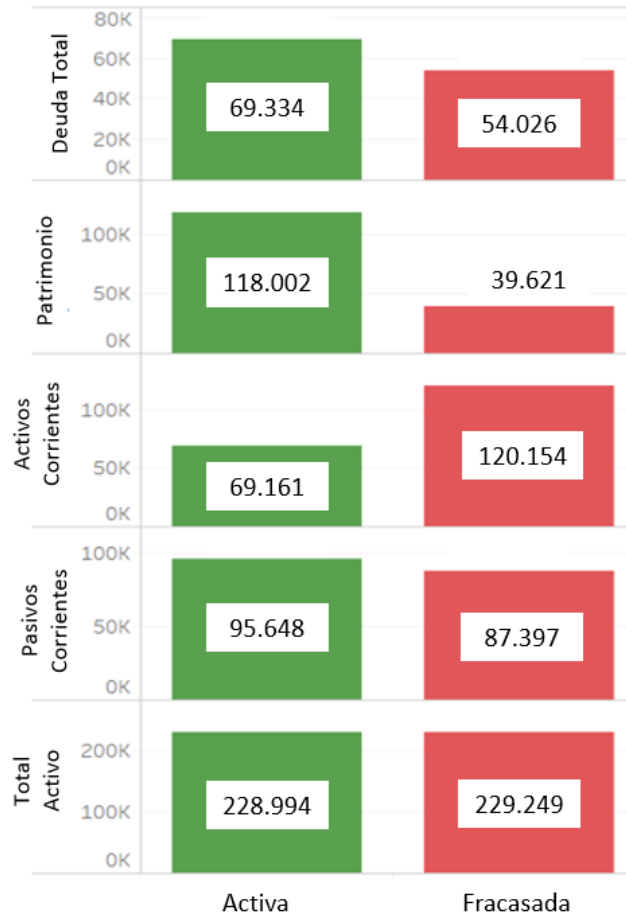
Sin embargo, dada la cantidad de datos disponible para este sector, es muy difícil realizar conclusiones acerca de los indicadores financieros que predicen el fracaso en las microempresas agrícolas.

Ilustración 10. Promedio de los indicadores que predicen el fracaso sector Agricultura



Fuente: Elaboración Propia

Ilustración 11. Promedio otras cuentas sector Agricultura



Fuente: Elaboración Propia

10. RESULTADOS SECTOR INDUSTRIAL

Para este sector contamos con más información de microempresas que en el sector agrícola. Durante la recolección y construcción de la base de datos, se pudo observar que la mayoría de las empresas del sector industrial son grandes empresas, esto tiene sentido ya que, generalmente las empresas industriales tienen una mayor inversión en maquina y equipo para la operación.

A continuación, se muestran los indicadores que evalúan el desempeño de los modelos para este sector: AUC-ROC, tasa de verdaderos positivos y precisión; teniendo en cuenta las pruebas de correlación de Pearson, de manera que esta no fuese superior al 0.20. Toda correlación superior se corrige eliminando la variable causante del problema.

Tabla 11. Desempeño de los diferentes modelos Sector Industrial

	Coeficiente de correlación Pearson R=0.20			
Marco de Decisión	Árboles Binarios	Random Forest	RN DDA	RN RPROP
AUC	0	0.804	0.659	0.363
Tasa de verdaderos positivos	70.45%	70.45%	70.45%	56.00%
Precisión	59.09%	71.59%	61.36%	61.22%
Indicadores		INDICADOR DE CAPITAL DE TRABAJO, RAZON CORRIENTE, ROA		
Variables eliminadas	ROE, INDICADOR DE RENTABILIDAD, LIQUIDEZ, EBITDA			

10.1 Resultados del mejor modelo: Random Forest

A continuación, se muestran en la Ilustración 12 y en la Ilustración 13 la curva ROC y la matriz de confusión, respectivamente.

La matriz de confusión del Random Forest muestra que, de las 88 empresas que hacen parte de la muestra, 63 se clasificaron correctamente, mientras que, las 25 restantes quedaron mal clasificadas por el modelo. De estas 25, 12 corresponden a falsos positivos y 13 a falsos negativos.

La tasa de verdaderos positivos es **70.45%**, la precisión es de **71.59%**.

Ilustración 12. Curva ROC, Sector Industrial, Random Forest

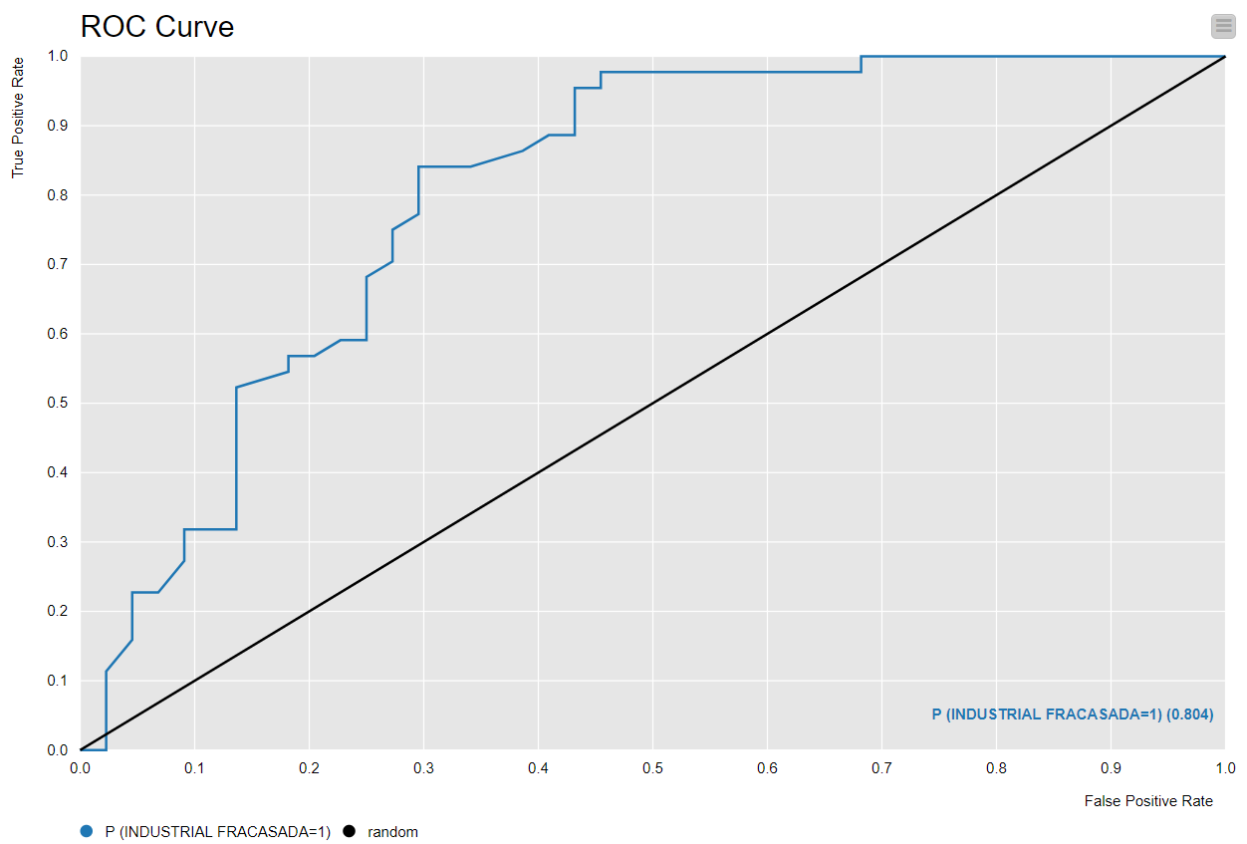


Ilustración 13. Matriz de confusión Random Forest, Sector Industrial

Confusion Matrix		0 (Predicted)	1 (Predicted)	
0 (Actual)		32	12	72.73%
1 (Actual)		13	31	70.45%
		71.11%	72.09%	
Overall Statistics				
Overall Accuracy	Overall Error	Cohen's kappa (κ)	Correctly Classified	Incorrectly Classified
71.59%	28.41%	0.432	63	25

En cuanto a los indicadores principales según el modelo Random Forest, para la predicción del fracaso empresarial en el sector industrial, se tienen al Indicador de capital de trabajo, ROA y Razón corriente. Ver tabla 12.

La tabla 13 nos muestra el porcentaje de selección de cada indicador.

Tabla 12. Random Forest Sector Industrial

Indicador Financiero	Splits nivel 1	Splits nivel 2	Splits nivel 3	Candidatos nivel 1	Candidatos nivel 2	Candidatos nivel 3
Indicador de capital de trabajo	27	30	47	37	71	113
ROA	26	33	49	34	71	113
Razón corriente	26	31	54	34	60	110
Ratio de eficiencia	9	24	28	31	59	116
Indicador Apalancamiento	7	24	37	36	65	106
Indicador de actividad	5	24	25	28	74	106

Fuente: Elaboración Propia

Este sector se caracteriza por tener unos indicadores financieros de predicción particulares, diferentes de los indicadores de agricultura y servicios. Las ratios que explican el fracaso empresarial son la razón Corriente, ROA e Indicador de capital de trabajo, para los niveles 1,2 y 3 respectivamente.

Tabla 13. Porcentaje de elección de cada indicador, Sector Industrial

Indicador Financiero	Splits	Candidato	% escogencia
Razón corriente	111	204	54%
ROA	108	218	50%
Indicador de capital de trabajo	104	221	47%
Indicador Apalancamiento	68	207	33%
Ratio de eficiencia	61	206	30%
Indicador de actividad	54	208	26%

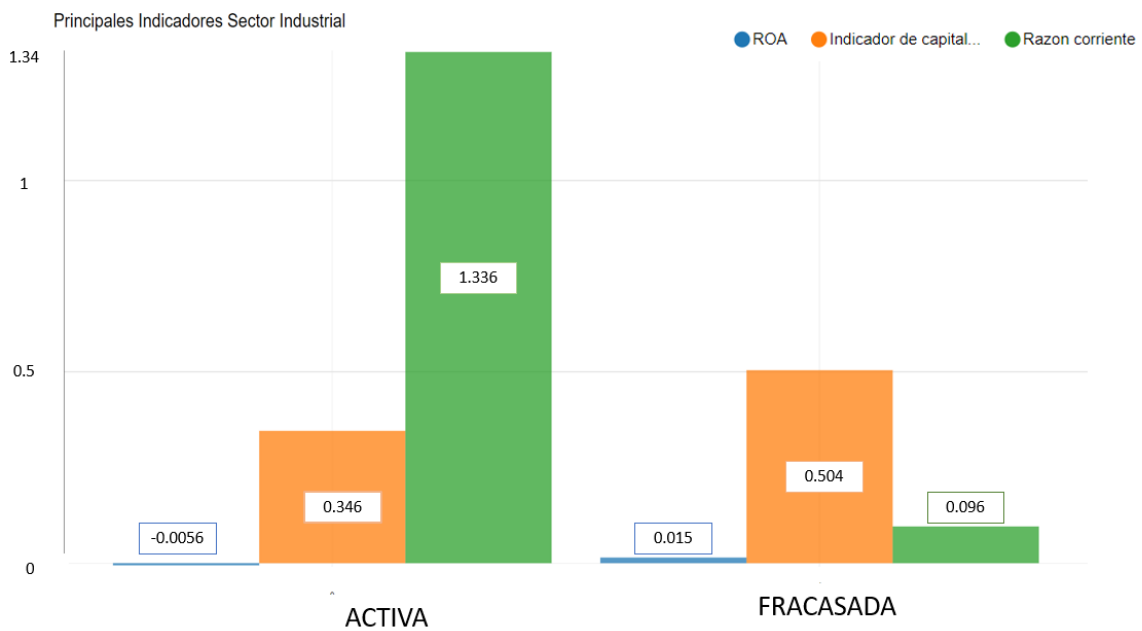
Fuente: Elaboración Propia

En resumen, los indicadores que predicen el fracaso empresarial en el sector industrial para las microempresas en Colombia son el indicador de capital de trabajo, razón corriente y ROA. Al graficar los promedios de los indicadores (Ilustración 14) para ver las diferencias de estos entre las empresas activas y fracasadas. En primer lugar obtuvimos que la razón corriente es mayor para las empresas activas, en promedio 1.336 contra 0.09 de las empresas fracasadas, y la razón es que los inventarios en las empresas activas son menores. En segundo lugar, el indicador de capital de trabajo es mayor para las empresas fracasadas, lo cual resulta sorprendente. La explicación la encontramos en los pasivos corrientes y en los activos totales, en ambos casos, estos son más grandes en las empresas activas. Los activos no corrientes también son mayores en este grupo de empresas, lo que puede ser entendido como una mayor inversión en maquinaria y equipo. En tercer lugar, el ROA en las empresas activas es menor. En este punto la razón principal es la utilidad negativa presentada por las empresas activas. Sorprendentemente, las empresas fracasadas reportaron en promedio una mayor utilidad que las activas, quizá por un recorte de costos, propio de las situaciones de estrés financiero que caracteriza los periodos pre quiebra.

Los resultados para el sector industrial coinciden con la teoría, en cuanto a que los indicadores de liquidez y capital de trabajo son dos indicadores que predicen fracaso en empresas de este sector (Thomas Ng et al., 2011), sin embargo, la teoría también nos muestra que un indicador de

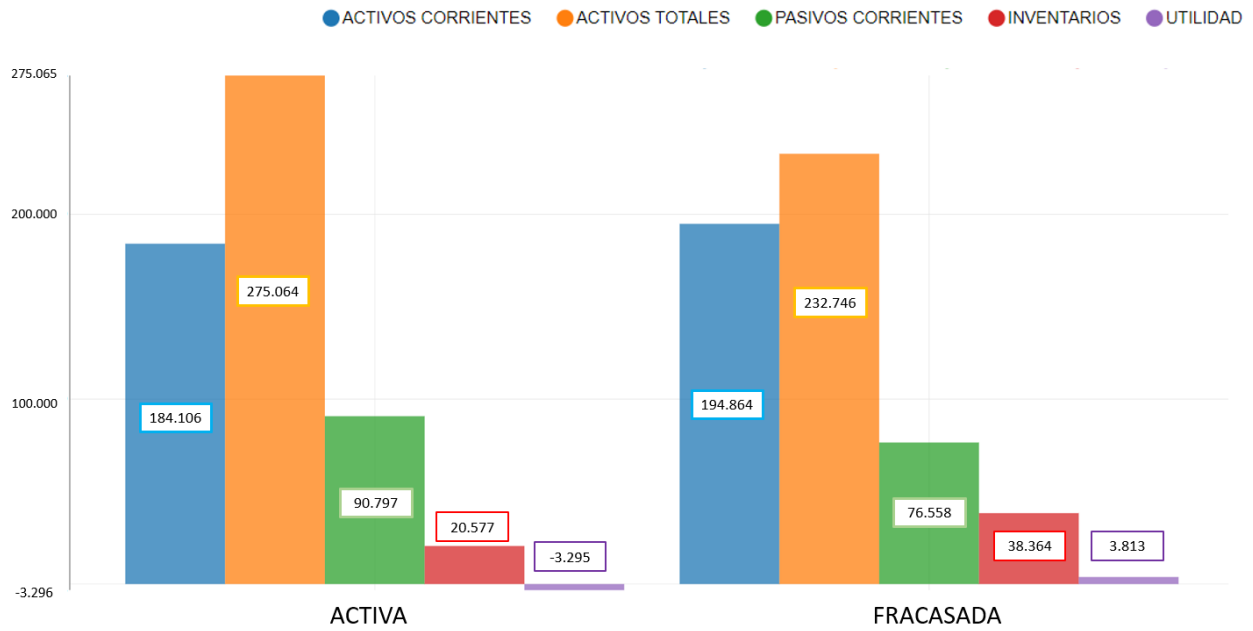
capital de trabajo le ayuda a las empresas a sortear las situaciones complicadas (Finishtya, 2019), y al estudiar las microempresas colombianas del sector industrial, encontramos que las fracasadas tienen un indicador de capital de trabajo mayor que las activas.

Ilustración 14. Promedio Indicadores Sector Industrial



Fuente: Elaboración Propia

Ilustración 15. Promedio cuentas - cálculo de los principales indicadores



Fuente: Elaboración Propia

Teniendo en cuenta que las diferencias entre los dos mejores modelos: Random Forest y DDA, pueden no ser estadísticamente significativas, se procede a aplicar la sensibilidad de Moodys (Moody & Utans, 1995) para extraer información de las redes neuronales y de esta manera conocer los indicadores que son fundamentales para el modelo DDA. La sensibilidad S_{β} tiene la siguiente ecuación:

$$S_{\beta} = \frac{1}{N} \sum_{j=1}^N ASE(\bar{x}_{\beta}) - ASE(x_{\beta})$$

$$\text{Donde, } \bar{x}_{\beta} = \frac{1}{N} \sum_{j=1}^N X\beta_j$$

“Aquí, $X\beta_j$ es la β del ejemplar j . La sensibilidad S_{β} mide el efecto en el error ASE (Average Square Error – error elevado al cuadrado) de reemplazar la variable de entrada β_j por su promedio \bar{x}_{β} .”

El reemplazo de la variable de entrada por su promedio remueve la influencia en el output del modelo” (Moody & Utans, 1995).

Para resumir, se quiere evaluar el cambio en el error del modelo, al reemplazar el valor de cada dato específico de cada clase, por el valor promedio de dicha clase. Entre más aumente el error al incluir el promedio, más importancia toma dicha variable.

Tabla 14. Resultados Sensibilidad DDA

RPROP	Error (\bar{x}_β)	Error x_β	ABS[(\bar{x}_β)-(x_β)]
Indicador de actividad	0.49300	0.39	0.1072
Indicador de Capital de trabajo	0.46480	0.39	0.079
ROA	0.42250	0.39	0.0367
Indicador de Liquidez	0.40850	0.39	0.0227
Razón corriente	0.36620	0.39	0.0196

Fuente: Elaboración Propia

Para las redes neuronales DDA las variables más importantes son indicador de actividad, Indicador de capital de trabajo y ROA, estos dos últimos indicadores coinciden con el modelo Random Forest.

11. RESULTADOS SECTOR SERVICIOS

Este es el sector productivo con mayor proporción de microempresas, lo cual tiene sentido ya que, la inversión o capital inicial para tener una empresa de servicios es mínima. De hecho, el 72.6% de las empresas que hacen parte de la base de datos de este trabajo son del sector servicios.

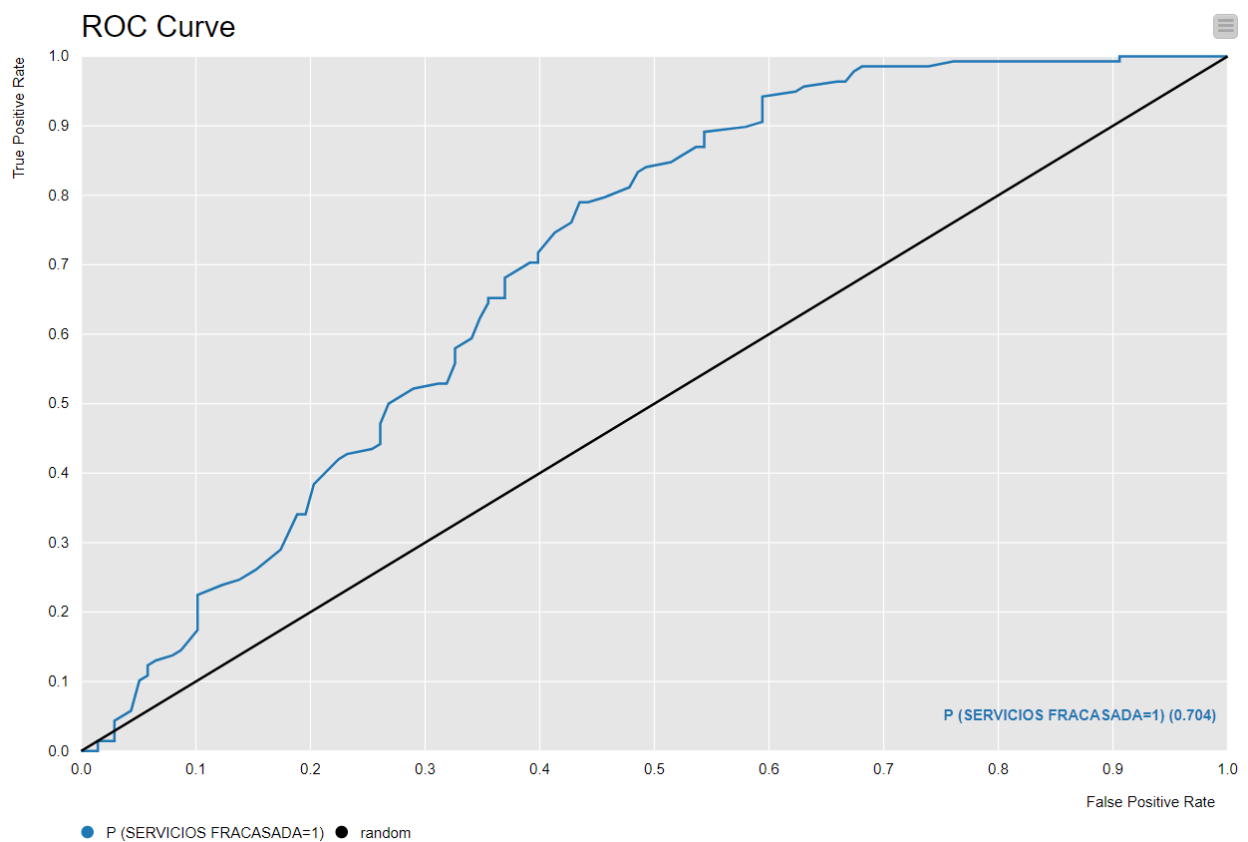
Antes de realizar los respectivos modelos, se evaluó que la correlación de las variables entre si fuese inferior a 0.20, cuando la correlación era superior a este nivel se eliminó la variable causante del problema. Ver tabla 15. A continuación, se muestran los indicadores que evalúan el desempeño de las diferentes herramientas aplicadas para este sector.

Tabla 15. Desempeño de los diferentes modelos Sector Industrial

Marco de Decisión	Coeficiente de correlación Pearson R=0.20			
	Árboles Binarios	Random Forest	RN DDA	RN RPROP
AUC	0.696	0.704	0.606	0.678
Tasa de verdaderos positivos	62.32%	68.12%	54.35%	69.31%
Precisión	61.59%	65.58%	58.33%	60.91%
Indicadores		ROA, ROE		
Variables eliminadas	ASSET TURNOVER, INDICADOR APALANCAMIENTO, INDICADOR ACTIVIDAD, INDICADOR LIQUIDEZ, EBITDA			

11.1 Resultados del mejor modelo: Random Forest

Ilustración 16. Curva ROC, Sector Servicios, Random Forest



La matriz de confusión del Random Forest nos muestra que, de las 95 empresas mal clasificadas, 51 son falsos positivos y 44 son falsos negativos. De otro lado, 181 microempresas se clasificaron correctamente por el modelo.

La tasa de verdaderos positivos es **68.12%** y la precisión es **65.58%**.

Ilustración 17. Matriz de Confusión Random Forest Sector Servicios

Confusion Matrix		0 (Predicted)	1 (Predicted)	
0 (Actual)	87	51		63.04%
1 (Actual)	44	94		68.12%
	66.41%	64.83%		

Overall Statistics				
Overall Accuracy	Overall Error	Cohen's kappa (κ)	Correctly Classified	Incorrectly Classified
65.58%	34.42%	0.312	181	95

El resultado del modelo muestra que el indicador que realiza la clasificación de las microempresas fracasadas, en el primer nivel es el ROA o la rentabilidad del activo. En segundo nivel, es el ROE o rentabilidad del patrimonio. Ver tabla 16.

Tabla 16. Random Forest Sector Servicios

Indicador Financiero	Splits nivel 1	Splits nivel 2	Splits nivel 3	Candidatos nivel 1	Candidatos nivel 2	Candidatos nivel 3
ROA	28	42	48	28	57	95
ROE	13	16	26	27	55	91
Coeficiente de deuda de largo Plazo CP	13	8	22	31	71	106
Coeficiente de deuda de largo Plazo LP	12	4	8	32	60	94
Ratio de deuda	8	20	18	30	57	89
Ratio de eficiencia	8	19	18	26	64	97
Indicador de capital de trabajo	5	11	29	32	53	107
Cash ratio	3	12	37	28	56	96
razón corriente	3	14	30	30	66	93

En la tabla 17, se puede ver el porcentaje de elección de los indicadores en el modelo. Por ejemplo, el ROA fue candidato 180 veces en diferentes divisiones, pero fue escogido para la primera división del modelo 118 veces, de aquí que el porcentaje de elección es del 66%.

Tabla 17. Porcentaje elección de cada indicador, sector Servicios

Indicador Financiero	Splits	Candidato	% escogencia
ROA	118	180	66%
ROE	55	173	32%
Cash ratio	52	180	29%
Ratio de deuda	46	176	26%
razón corriente	47	189	25%
Ratio de eficiencia	45	187	24%
Indicador de capital de trabajo	45	192	23%
Coficiente de deuda de largo Plazo CP	43	208	21%
Coficiente de deuda de largo Plazo LP	24	186	13%

El indicador fundamental en la predicción de fracaso de las microempresas del sector servicios en Colombia es el ROA. La ilustración 19, muestra el resumen del promedio de dos indicadores de rentabilidad, ROE y ROA. Los resultados para este sector son los esperados, ya que, al no tener inventarios ni producción, es la rentabilidad quien define la quiebra. Para las empresas activas, el promedio de ROA es -0.575 versus -2.175 para empresas fracasadas. El ROE es -0.179 versus -2.380, activas y fracasadas respectivamente.

La principal explicación al resultado de los ROA Y ROE se refleja en la utilidad del ejercicio, si bien es negativa para ambos grupos, es menos negativa para empresas activas. Las barras verdes representan el promedio de dichas cuentas para las empresas activas, mientras las barras rojas representan el promedio de las mismas cuentas para empresas fracasadas, Ilustración 20.

(Duarte, 2019) encuentra que, en Bogotá para empresas del sector servicios, tres de los cuatro indicadores que diferencian las empresas activas de las fracasadas en el sector servicios, son indicadores de rentabilidad.

Ilustración 19. Promedio de los Indicadores que predicen el Fracaso en el Sector Servicios

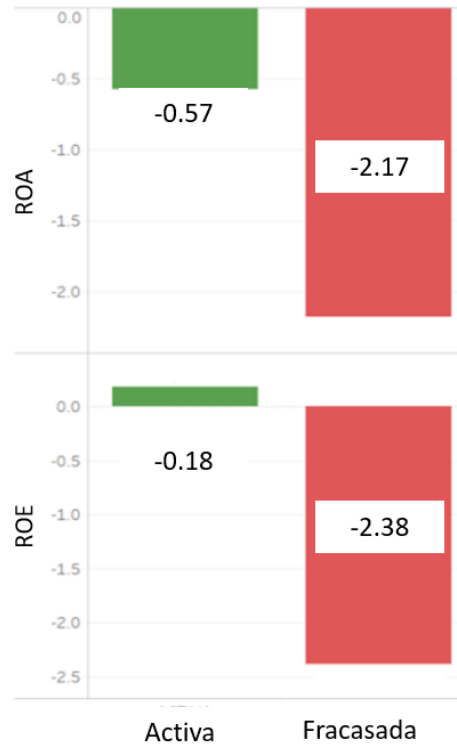
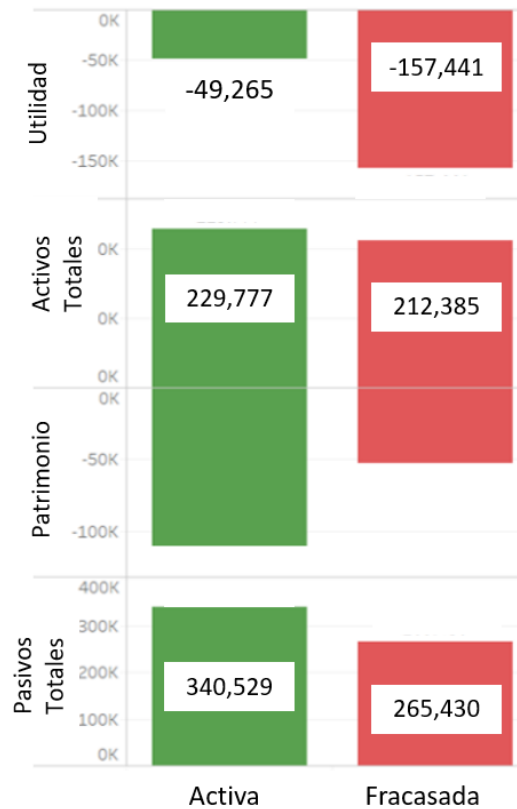


Ilustración 20. Promedio cuentas - Indicadores Sector Servicios



12. VALIDACIÓN DE LOS RESULTADOS CON EXPERTOS

Para la validación con expertos se consultó con el departamento de crédito de uno de los Bancos más importantes de Colombia, en donde se realiza una revisión exhaustiva a los balances de las diferentes empresas para la asignación créditos. En esta área, no solo se determina si la empresa debe tener cupo de crédito abierta con el banco, también se determina la cantidad de crédito aprobado, periodicidad del pago de intereses y capital, y plazo de los créditos. Todo esto, después de estudiar el flujo de caja del cliente, su actividad y cifras, de acuerdo con el sector al que pertenece.

Los indicadores con mayor impacto a la hora de revisar las cifras son, para el sector Agro, el endeudamiento, como primera medida, el ciclo operacional, los inventarios, la relación deuda/Ebitda y la cartera del cliente.

Para el sector Industrial, en primer lugar, el Capital de trabajo, además del endeudamiento con el sistema financiero. Se revisa el ciclo de vida del producto, así como la rotación de inventarios.

Para el sector servicios, se estudia a fondo el pasivo del cliente, revisando la acumulación de obligaciones laborales. Es de vital importancia estudiar el ROA para empresas de servicios que tienen activos, mientras que si la empresa no tiene activos significativos se revisa el EBITDA, pero en general para este sector se estudian las rentabilidades obtenidas.

En la validación con expertos académicos, la profesora María Teresa Macias, jefe del departamento de Finanzas de la Universidad de la Sabana, sugirió además de los indicadores utilizados y descritos en la sección de selección de variables, un indicador que “permita saber qué tan apalancada está la operación en los proveedores o si la empresa requiere financiación adicional para su operación, pues estas necesidades de efectivo para el día a día podrían restarle capacidad a la empresa o fortalecerla” (Macias, 2020). Dicho indicador es el ciclo de conversión de efectivo. Sin embargo, no se pudo incluir dentro de la lista de indicadores para tener en cuenta,

ya que la base de datos no permite calcular los diferentes días (inventario, cobros y pagos) de operación que se requieren para calcular el ciclo de conversión de efectivo.

13. CONCLUSIONES

Después de aplicar algunas de las herramientas de la ingeniería analítica al problema de este trabajo, se puede concluir que los modelos individuales tienen como variables de predicción indicadores diferentes a los resultantes para el estudio de la muestra sin clasificar por sectores, esto se da porque la muestra en su mayoría está compuesta por empresas del sector servicios, de manera que los resultados de las empresas agrícolas e industriales se verían sesgados al realizar un análisis para toda la muestra, sin clasificar por sectores.

El análisis por sectores es de vital importancia, sobre todo para empresas del sector industrial y agrícola, ya que, si bien es cierto, los indicadores de rentabilidad ROA y ROE son fundamentales a la hora de realizar un análisis de cifras en las empresas, para empresas del sector industrial, las variables definitivas son capital de trabajo, ROA y razón corriente, mientras que, para el sector agrícola el acceso al capital de trabajo mediante apalancamiento es fundamental.

Un mejor indicador de capital de trabajo, de liquidez o de apalancamiento no significa siempre que la empresa va por buen camino, hay empresas que pueden tener mejores resultados en cualquier de estos indicadores, por ejemplo, las empresas que entran en ley 1116 pueden tener indicadores de apalancamiento bajos que se da por la falta de acceso a créditos bancarios, pero además porque se encuentran en acuerdos de reestructuración de deuda, de manera que gran parte de su ingreso es destinado al pago de las deudas bancarias, limitando así, el crecimiento, producción o ventas de las empresas.

Es preocupante que las empresas activas del sector servicios muestran un indicador ROE en promedio negativo, que es mejor que el de las empresas fracasadas, sin embargo, la rentabilidad

del patrimonio no es lo suficientemente alta como para cubrir costos y gastos, situando a las empresas en una zona de peligro en la que es necesario el cambio de estrategia bien sea, la diferenciación de productos que aporte más a la rentabilidad o mayor inversión en productividad.

En cuanto a la precisión de las herramientas utilizadas, el Random Forest y los árboles binarios muestran ser las herramientas no solo con mayor capacidad predictiva, también con mayor capacidad explicativa de manera, que los resultados que brindan pueden ser utilizados por parte de directivos y hacedores de políticas a la hora de tomar decisiones.

14. CONTRIBUCIONES

Como se mencionó al principio, en Colombia los estudios acerca del fracaso de las pymes no son muchos, la mayoría de estos utilizan herramientas de análisis de regresión lineal. En cuanto al análisis del fracaso de las microempresas si hay dos trabajos al respecto en Colombia son muchos. De manera que este trabajo es innovador en cuanto a las herramientas implementadas y al tamaño de las empresas estudiadas. En parte, el acceso a la información de las microempresas creo que ha sido limitante a la hora de estudiarlas a fondo.

En cuanto a los aportes del trabajo a la ingeniería analítica en Colombia, hasta ahora se habían establecido cifras de cierre de las microempresas, algunos factores e indicadores, sin embargo, el estudio por sectores les brinda a los directivos de las empresas, bancos y a quienes realizan las políticas públicas, herramientas para la toma de decisión, en cuanto a asignación de fondos, ayudas e implementación de planes que apoyan a las micro y pequeñas empresas en Colombia.

15. TRABAJOS EN EL FUTURO

En Colombia no tenemos fácil acceso a los informes financieros de las pequeñas empresas, si bien esta situación ha mejorado de manera considerable con el paso de los años, la base de datos requerida para realizar este trabajo se tomó meses en conseguir y armar. Para trabajos futuros se podría realizar una visita a la cámara de comercio de Bogotá y establecer convenios con las diferentes cámaras de comercio del país, de manera que se obtuviese una base de datos más completa y con mayor cobertura de las diferentes ciudades de Colombia.

En este trabajo se realizó el modelo de probabilidad lineal para toda la muestra y para los diferentes sectores, sin embargo, ni los coeficientes ni los ajustes de bondad fueron significativos. Como una variación a este trabajo, el modelo Logit-probit puede mostrar mejores resultados, y ser una mejor aproximación que el modelo de probabilidad lineal utilizado en este punto.

Para trabajos futuros, se pueden evaluar otras herramientas de ingeniería analítica, tales como K-Means y realizar agrupaciones o clústeres a través de esta herramienta para empresas del sector servicios, porque, aunque son empresas que pertenecen al mismo sector, todavía se puede observar bastante heterogeneidad.

Debido a que el objetivo principal de este trabajo es identificar los indicadores financieros que predicen el fracaso empresarial para los diferentes sectores de la economía colombiana, no se realizó un análisis de componentes principales por la falta de interpretabilidad de los resultados, sin embargo, para un trabajo en el que se quiera realizar un modelo de predicción sin necesidad de conocer los resultados de los modelos, esta podría ser una herramienta útil para reducir la correlación entre las diferentes variables.

Otro aporte interesante para trabajos futuros es el impacto de los ciclos económicos en el fracaso de las empresas, en particular si se tiene en cuenta los sectores productivos. Por ejemplo, los fenómenos climáticos en Colombia que provocan inflaciones altas y que pueden perjudicar a las empresas agrícolas. La devaluación del peso colombiano y la tasa de cambio en la importación de

bienes y maquinaria que seguramente puede influir en el bienestar de una empresa industrial. Las altas tasas de intervención del Banco de la Republica que pueden encarecer los créditos de las empresas o la imposición de aranceles a ciertos productos de la industria. El considerar los ciclos económicos y como afectan las variables macro podría enriquecer los resultados y los planes de acción.

De otro lado, se podría plantear la realización de este tipo de estudios según el lector principal; es diferente un estudio de fracaso empresarial cuando el principal interesado es un banco o proveedor de liquidez o cuando es una entidad gubernamental. Si bien es cierto, ambos lectores quisieran saber si una empresa fracasa o no, el primero querrá ser más acido en la implementación de las herramientas y se enfocara más en los falsos negativos; porque el banco no va a querer prestarle dinero a una empresa fracasada que luce como sana en el modelo, mientras que el segundo quisiera disminuir los falsos positivos, ya que no se quiere gastar dinero público en ayudas a empresas que realmente no lo necesiten.

Creo que, en trabajos futuros, una contribución adicional a la ingeniería analítica en Colombia podría ser la creación de planes de ayuda para las microempresas dependiendo de los resultados obtenidos en el estudio. Se puede adicionalmente, establecer convenios con organizaciones como Anif o las cámaras de comercio de cada ciudad para de esta manera hacer públicos los resultados y las herramientas a cada dueño de empresa, así, cada uno puede realizar el análisis de su empresa en particular.

16. LIMITACIONES DEL TRABAJO

Como se ha mencionado en los capítulos anteriores, quizá la gran limitante de este trabajo fue el acceso a información de microempresas que fuera representativa para los tres sectores, pero en especial para el sector de agricultura. La información utilizada en este trabajo fue construida con

base en datos obtenidos de varios proveedores de información gubernamentales (Sirem, RUES, Cámaras de comercio) sin embargo, comprar una base de datos mediante convenios con las diferentes cámaras de comercio del país hubiese permitido tener una muestra representativa para el sector agrícola.

Otra limitación del trabajo tiene que ver con la correlación entre variables, si bien se aplicaron las pruebas de correlación de Pearson, y se solucionaron hasta que R^2 fuese 0.20, eliminar mas variables para reducir la correlación puede sesgar los modelos. De manera que la mejor alternativa para reducir la correlación entre variables, es aumentar el número de la muestra (Wooldridge, 2010).

Una segunda limitante de este trabajo fue el asumir que una empresa con cámara de comercio cuyo último año de renovación es 2016, pertenece al grupo de las empresas fracasadas. Si bien es cierto, la cámara de comercio y su renovación es indispensable para acceder a los diferentes servicios bancarios y para la interacción con los demás agentes de la economía, como proveedores y clientes empresariales; poder acceder a una lista de empresas que se hayan declarado en ley 1116 o ley de insolvencia, nos hubiese permitido tener plena certeza de que esta empresa pertenece realmente al grupo de fracasadas.

17. REFERENCIAS BIBLIOGRAFICAS

- Alifiah, M. N. (2014). Prediction of Financial Distress Companies in the Trading and Services Sector in Malaysia Using Macroeconomic Variables. *Procedia - Social and Behavioral Sciences*, 129, 90-98. <https://doi.org/https://doi.org/10.1016/j.sbspro.2014.03.652>
- Altman, E. (1968). Financial Ratios, Discriminan Analysis and the prediction of Corporate Bankruptcy. *The Journal of Finance*, 23(4), 568-609. Recuperado de <http://www.jstor.org/stable/pdf/2978933.pdf>
- Altman, E. I. (2000). PREDICTING FINANCIAL DISTRESS OF COMPANIES: REVISITING THE Z-SCORE AND ZETA

- ® MODELS Predicting Financial Distress of Companies: Revisiting the Z-Score and ZETA ® Models Background. Recuperado de <http://pages.stern.nyu.edu/~ealtman/Zscores.pdf>
- Altman, E., Iwanicz-Drozdowska, M., Laitinen, E. K., & Suvas, A. (2017). Financial Distress Prediction in an International Context: A Review and Empirical Analysis of Altman's Z-Score Model. *Journal of International Financial Management and Accounting*, 28(2), 131-171. <https://doi.org/10.1111/jifm.12053>
- ANIF. (2019). La inclusión financiera de las MIPYME en Colombia | Anif. Recuperado 29 de febrero de 2020, de <http://www.anif.co/Biblioteca/politica-fiscal/la-inclusion-financiera-de-las-mipyme-en-colombia>
- Au, T. C. (2018). *Random Forests, Decision Trees, and Categorical Predictors: The «Absent Levels» Problem*. *Journal of Machine Learning Research* (Vol. 19). Recuperado de <https://www.stat.berkeley.edu/>
- Bae, J. K. (2012). Predicting financial distress of the South Korean manufacturing industries. *Expert Systems With Applications*, 39(10), 9159-9165. <https://doi.org/10.1016/j.eswa.2012.02.058>
- BANREP. (2018). Salarios | Banco de la República (banco central de Colombia). Recuperado 20 de abril de 2018, de <http://www.banrep.gov.co/es/mercado-laboral/salarios>
- Beaver, W. H. (1966). Financial Ratios As Predictors of Failure. *Source Journal of Accounting Research Empirical Research in Accounting: Selected Studies*, 4, 71-111. Recuperado de <http://www.jstor.org/stable/2490171>
- Breiman, L. (2001). RANDOM FOREST. *Scimago Journal & Country Rank*. Recuperado de <https://www.stat.berkeley.edu/~breiman/randomforest2001.pdf>
- Calvo, D. (2017). Clasificación de redes neuronales artificiales. Recuperado 31 de mayo de 2020, de <http://www.diegocalvo.es/clasificacion-de-redes-neuronales-artificiales/>
- Caro, N. P. (2016). Predicción del Fracaso Empresarial en empresas de Argentina, Chile y Perú a través de indicadores contables., 130-147.
- Castaño, Horacio, Perez, F. (2007). Las Redes Neuronales y la Evaluación del Riesgo de Crédito. *Revista Ingenierias Universidad de Medellin*, 6(10).
- CEPAL. (2015). Reflexiones acerca del desarrollo de las Mipyme.
- Chauhan, N., Ravi, V., & Karthik Chandra, D. (2009). Differential evolution trained wavelet neural networks: Application to bankruptcy prediction in banks. *Expert Systems With Applications*, 36, 7659-7665. Recuperado de <http://10.0.3.248/j.eswa.2008.09.019>
- Colombia, C. de la R. de. (2000). Ley 590 de 2000.
- Confecamaras. (2016). Nacimiento y supervivencia de las empresas en Colombia. Recuperado de

http://www.confecamaras.org.co/phocadownload/Cuadernos_de_analisis_economico/Cuaderno_de_Analisis_Economico_N_11.pdf

Confecamaras. (2017). Determinantes del crecimiento acelerado de las empresas. *Red de cámaras de comercio*, 5-21. Recuperado de

http://www.confecamaras.org.co/phocadownload/Cuadernos_de_analisis_economico/Cuaderno_de_Analisis_Economico_N_13.pdf

Consejo privado de Competitividad. (2017). Competitividad del sector agropecuario colombiano. Recuperado de <https://compite.com.co/wp-content/uploads/2017/05/208Agro.pdf>

Creus, T. (2000). CENTRE METAL·LÚRGIC EL ANÁLISIS DE LA EMPRESA A TRAVÉS DE LOS RATIOS. Recuperado de <http://www.centrem.cat/ecomu/upfiles/publicacions/analisi.pdf>

Cruz, E. A., Espinosa, J., & Aristizabal, S. (2014). Modelo para la medición del riesgo de insolvencia empresarial: PYME de Colombia, un caso de estudio. *Entre Ciencia e Ingeniería*, 16. Recuperado de <https://login.ez.unisabana.edu.co/login?url=https://search.ebscohost.com/login.aspx?direct=true&db=edsgii&AN=edsgcl.455782852&lang=es&scope=site>

DANE. (2012). *Clasificación Industrial Internacional Uniforme de todas las actividades Económicas*. Recuperado de https://www.dane.gov.co/files/nomenclaturas/CIIU_Rev4ac.pdf

DANE. (2016). *Gran Encuesta Integrada de Hogares 2016*.

DANE, D. N. de E. (2009). Encuesta Anual Manufacturera 2009.

De Llano Monelos, P., Piñeiro Sanchez, C., & Rodriguez Lopez, M. (2016). Predicción del fracaso empresarial . Una contribución a la síntesis de una teoría mediante el análisis comparativo de distintas técnicas de predicción * *Resumen*, 43, 163-199.

Dirección de Gestión y Transformación del Conocimiento. (2019). *Causa de liquidación de las empresas 2018*. Recuperado de https://bibliotecadigital.ccb.org.co/bitstream/handle/11520/22862/Causas_de_liquidacion_de_las_empresas_en_Bogota_2018.pdf?sequence=1&isAllowed=y

Duarte, M. (2019). Indicadores Financieros que denotan el Fracaso Empresarial en las Mipymes Bogotanas del sector Servicios. Recuperado de <http://bdigital.unal.edu.co/73631/1/1032433125.2019.pdf>

Edmister, R. O. (1972). An Empirical Test of Financial Ratio Analysis for Small Business Failure Prediction. *Source: The Journal of Financial and Quantitative Analysis*, 7(2), 1477-1493. Recuperado de <http://www.jstor.org/stable/2329929>

El país. (2019). «Informalidad supera el 86% en el campo»: Presidente de la SAC. Recuperado 1 de febrero de 2020, de <https://www.elpais.com.co/economia/informalidad-supera-el-86-en-el-campo-presidente-de-la-sac.html>

- Escandon, Diana, Hurtado, A. (2017). El compromiso exportador en Colombia : un análisis de redes neuronales Export Commitment in Colombia : An Analysis of Neural Networks, (24), 362-388.
- Espinosa, F. R., Molina, Z. A. M., & Vera-Colina, M. A. (2015). Fracaso empresarial de las pequeñas y medianas empresas (pymes) en Colombia. *Suma de Negocios*, 6(13), 29-41. <https://doi.org/https://doi.org/10.1016/j.sumneg.2015.08.003>
- Fedorova, E., Gilenko, E., & Dovzhenko, S. (2013). Bankruptcy prediction for Russian companies: Application of combined classifiers. *Expert Systems with Applications*, 40, 7285-7293. <https://doi.org/10.1016/j.eswa.2013.07.032>
- Finishtya, F. C. (2019). THE ROLE OF CASH FLOW OF OPERATIONAL, PROFITABILITY, AND FINANCIAL LEVERAGE IN PREDICTING FINANCIAL DISTRESS ON MANUFACTURING COMPANY IN INDONESIA. *JURNAL APLIKASI MANAJEMEN*, 17(1), 110-117. <https://doi.org/10.21776/ub.jam.2019.017.01.12>
- Franco Sepúlveda, G., Velilla, D. A., & Velez, I. E. (2014). ANÁLISIS DEL PRECIO DEL CARBÓN MEDIANTE REDES NEURONALES ARTIFICIALES (RNA). *Boletín de Ciencias de la Tierra*, (35), 31-35. <https://doi.org/10.15446/rbct.n35.39682>
- García Salgado, O., & Morales Castro, A. (2016). DESEMPEÑO FINANCIERO DE LAS EMPRESAS: UNA PROPUESTA DE CLASIFICACIÓN POR RNA. (Spanish). *DESEMPENHO FINANCEIRO DAS EMPRESAS: UMA PROPOSTA DE CLASSIFICAÇÃO POR RNA. (Portuguese)*, 14(2), 11. Recuperado de <https://login.ez.unisabana.edu.co/login?url=https://search.ebscohost.com/login.aspx?direct=true&db=edb&AN=118359063&lang=es&scope=site>
- Gil, J., Cruz, J., & Yaelt, A. (2018). Desempeño financiero empresarial del sector agropecuario: un analisis comparativo entre Colombia y Brasil 2011 - 2015. *Revista EAN*. Recuperado de <http://www.scielo.org.co/pdf/ean/n84/0120-8160-ean-84-00109.pdf>
- Gil Ospina, Armando A, Jimenez Sepulveda, J. J. (2015). El contexto económico global de la Pyme. *Revista Académia e Institucional*, 95, 155-179.
- Giner, B. (2013). Predicción del fracaso empresarial en los sectores de construcción e inmobiliario: Modelos generales versus específicos* Corporate failure prediction in the construction and real estate industries: General versus industry-focused models, 1698-5117. Recuperado de <http://repositori.uji.es/xmlui/bitstream/handle/10234/84531/61113.pdf?sequence=1>
- Heo, J., & Yang, J. Y. (2014). AdaBoost based bankruptcy forecasting of Korean construction companies. *Applied Soft Computing*, 24, 494-499. <https://doi.org/10.1016/J.ASOC.2014.08.009>
- Hurtado, C. (2012). Arboles de Decisión, (I), 51.
- Igel, C., & Hüsken, M. (2002). Empirical evaluation of the improved Rprop learning. *Neurocomputing*, 50,

105-123. [https://doi.org/10.1016/S0925-2312\(01\)00700-7](https://doi.org/10.1016/S0925-2312(01)00700-7)

IISE. (2019). Introduction to Engineering Analytics. Recuperado 17 de marzo de 2020, de <https://www.iise.org/TrainingCenter/CourseDetail.aspx?EventCode=EAS>

Kalra, M., Kumar, S., & Das, B. (2020). Seismic Signal Analysis Using Empirical Wavelet Transform for Moving Ground Target Detection and Classification. *IEEE sensors journal*. Recuperado de http://explore.bl.uk/primo_library/libweb/action/display.do?tabs=detailsTab&gathStatTab=true&ct=display&fn=search&doc=ETOCvdc_100102805435.0x000001&indx=1&reclds=ETOCvdc_100102805435.0x000001

Ke-Lin Du, M. N. S. S. (2019). *Neural Networks and Statistical Learning*. Recuperado de <https://books.google.com.co/books?id=IUmvDwAAQBAJ&printsec=frontcover#v=onepage&q&f=false>

KNIME. (2020). DDA Node Description.

Kotane, I., & Kuzmina-Merlino, I. (2017). Analysis of Small and Medium Sized Enterprises' Business Performance Evaluation Practice at Transportation and Storage Services Sector in Latvia. En *Procedia Engineering* (Vol. 178, pp. 182-191). Elsevier Ltd. <https://doi.org/10.1016/j.proeng.2017.01.093>

La Republica. (2018). La Ley 1116 ha acogido a más de 3.600 firmas con problemas financieros. Recuperado 21 de marzo de 2020, de <https://www.larepublica.co/especiales/proteccion-patrimonial/cual-es-el-balance-de-la-ley-de-insolvencia-en-colombia-2719926>

LA REPUBLICA. (2018). Resultados de la Gran Encuesta a las Microempresas 2018. Recuperado 11 de octubre de 2019, de <https://www.larepublica.co/analisis/sergio-clavijo-500041/resultados-de-la-gran-encuesta-a-las-microempresas-2018-2718177>

Lee Sangjae, Sung Choi, W. (2013). A multi-industry bankruptcy prediction model using back-propagation neuronal network and multivariate discriminant analysis. *Expert Systems with Applications*, 40, 2941-2946. Recuperado de https://ac.els-cdn.com/S095741741201250X/1-s2.0-S095741741201250X-main.pdf?_tid=cdca43df-7b81-4126-821f-3b79ace12d56&acdnat=1524435711_5d6c9f3bc777e467a0d94eff3e0551fa

Leon Valdes, C. (2002). El análisis financiero como herramienta en la predicción de quiebra e insolvencia financiera. Recuperado de <https://revistas.uexternado.edu.co/index.php/contad/article/view/1295/1232>

Leyes desde 1992 - Vigencia expresa y control de constitucionalidad [LEY_1116_2006]. (2019). Recuperado 21 de marzo de 2020, de http://www.secretariasenado.gov.co/senado/basedoc/ley_1116_2006.html

- Liaw, A., & Wiener, M. (2001). *Classification and Regression by RandomForest*. Recuperado de <https://www.researchgate.net/publication/228451484>
- Lukason, Oliver, Laitinen, E. (2017). Firm failure processes and components of failure risk: An analysis of European bankrupt firms. *Journal of Business Research*. Recuperado de <https://pdf.sciencedirectassets.com/271680/1-s2.0-S0148296319X00028/1-s2.0-S0148296318303126/main.pdf?X-Amz-Security-Token=AgoJb3JpZ2luX2VjED8aCXVzLWVhc3QtMSJHMEUCIDJ4f7n1ml%2BR7KPrV5%2FsKHzQuQqdRMfNjOWZHkPi%2Fa88AiEAwfa1JKXzW5pItUxVPv94IxeYlOm2fRrBRclxbV>
- Lukason, O., & Laitinen, E. K. (2018). Failure of exporting and non-exporting firms: do the financial predictors vary? *Review of International Business and Strategy*, 28(3/4), 317-330. <https://doi.org/10.1108/RIBS-02-2018-0015>
- Lukason, O., Laitinen, E. K., & Suvas, A. (2016). Failure processes of young manufacturing micro firms in Europe. *Management Decision*, 54(8), 1966-1985. <https://doi.org/10.1108/MD-07-2015-0294>
- Macias, M. T. (2020). Comentario Tesis.
- Marek, Durika, Jaroslav, Frnda, Svabova, L. (2019). View of Decision tree based model of business failure prediction for Polish companies. Recuperado 21 de diciembre de 2019, de <http://economic-research.pl/Journals/index.php/oc/article/view/1718/1607>
- Medina Hurtado, Santiago, M. J. (2011). Pronóstico de la demanda de energía eléctrica horaria en Colombia. *Revista facultad de Ingeniería U. Antioquia*, 59, 98-107.
- Ministerio de Agricultura. (2016). Marco Nacional Cualificaciones Sector Agropecuario. Recuperado de https://www.mineducacion.gov.co/1759/articles-362822_recurso.pdf
- Ministerio de Agricultura. (2018). *Estrategia de política pública para la gestión integral de riesgos agropecuarios en Colombia*. Recuperado de <https://www.minagricultura.gov.co/Documents/LIBRO ESTRATEGIA VERSION FINAL.pdf>
- Moody, J., & Utans, J. (1995). Principled Architecture Selection for Neural Networks: Application to Corporate Bond Rating Prediction. *International Journal of Forecasting*, 11(4), 602-603. [https://doi.org/10.1016/s0169-2070\(95\)90007-1](https://doi.org/10.1016/s0169-2070(95)90007-1)
- Mures Quintana, Ma Jesus, Garcia Gallego, Ana, Vallejo Pascual, M. E. (2012). Análisis del fracaso empresarial por sectores: Factores diferenciadores.
- Niels, M. (1995). The Dynamic Decay Adjustment Algorithm. Recuperado 31 de mayo de 2020, de <http://www.ra.cs.uni-tuebingen.de/SNNS/UserManual/node193.html>
- Ohlson, J. A. (1980). Financial Ratios and the Probabilistic Prediction of Bankruptcy. *Journal of Accounting*

Research, 18(1), 109. <https://doi.org/10.2307/2490395>

Ortiz Arango, F. (2017). Pronóstico de precios de petróleo: una comparación entre modelos garch y redes neuronales diferenciales. *Investigación Económica*, 76(300), 105-126. <https://doi.org/10.1016/J.INVECO.2017.06.002>

Ortíz, D., Villa, F., & Velásquez, J. (2007). *Una Comparación entre Estrategias Evolutivas y RPROP para la Estimación de Redes Neuronales*. *Revista Avances en Sistemas e Informática* (Vol. 4).

Ortiz Medina, M. (2013). El fracaso de la microempresa relacionado con las características individuales del propietario: un estudio empírico en Republica Dominicana. *FAEDPYME INTERNATIONAL REVIEW*, 2(3). <https://doi.org/10.15558/fir.v2i3.34>

Perez G, J. I., Lorena, G. C. K., & Mauricio, L. C. (2013). 8. Modelos De Prediccion De La Fragilidad Empresarial: Aplicacion Al Caso Colombiano Para El Año 2011. *Red de Revistas Cientificas de America Latina, el Caribe, España y Portugal*, (22), 205-228. Recuperado de <http://www.redalyc.org/pdf/861/86131758010.pdf>

Platt, Harlan, Platt, M. (1990). DEVELOPMENT OF A CLASS OF STABLE PREDICTIVE VARIABLES: THE CASE OF BANKRUPT...: Biblioteca Octavio Arizmendi Posada. *Journal of Business Finance and Accounting*, 17, 31-50. Recuperado de <https://eds-a-ebSCOhost-com.ez.unisabana.edu.co/eds/pdfviewer/pdfviewer?vid=4&sid=c6754dc8-0fdb-41c5-803b-6d0501524b83%40sessionmgr4010>

Plazas, J. E., López, I. D., & Corrales, J. C. (2017). A Tool for Classification of Cacao Production in Colombia Based on Multiple Classifier Systems. *Lecture notes in computer science*. Recuperado de http://explore.bl.uk/primo_library/libweb/action/display.do?tabs=detailsTab&gathStatTab=true&ct=display&fn=search&doc=ETOCCN602558834&indx=1&recIds=ETOCCN602558834

Politechnic, T. (2020). Engineering Analytics and Machine Learning | Temasek Polytechnic. Recuperado 17 de marzo de 2020, de <https://www.tp.edu.sg/courses/part-time-courses/skillsfuture/stackable-modular-courses/engineering-analytics-and-machine-learning>

Revista Semana. (2020). ¿Por qué los bancos no saben satisfacer las necesidades de las PYMEs? Recuperado 23 de marzo de 2020, de <https://www.semana.com/hablan-las-marcas/articulo/porque-los-bancos-no-saben-satisfacer-las-necesidades-de-las-pymes/654795>

Romero Espinosa, F. (2013). Variables financieras determinantes del fracaso empresarial para la pequeña...: Biblioteca Octavio Arizmendi Posada. *Pensamiento & Gestión*, 34, 235-277. Recuperado de <https://eds-b-ebSCOhost-com.ez.unisabana.edu.co/eds/pdfviewer/pdfviewer?vid=18&sid=9969ccc5-9248-4878-b675->

25e3e0f45449%40sessionmgr4009

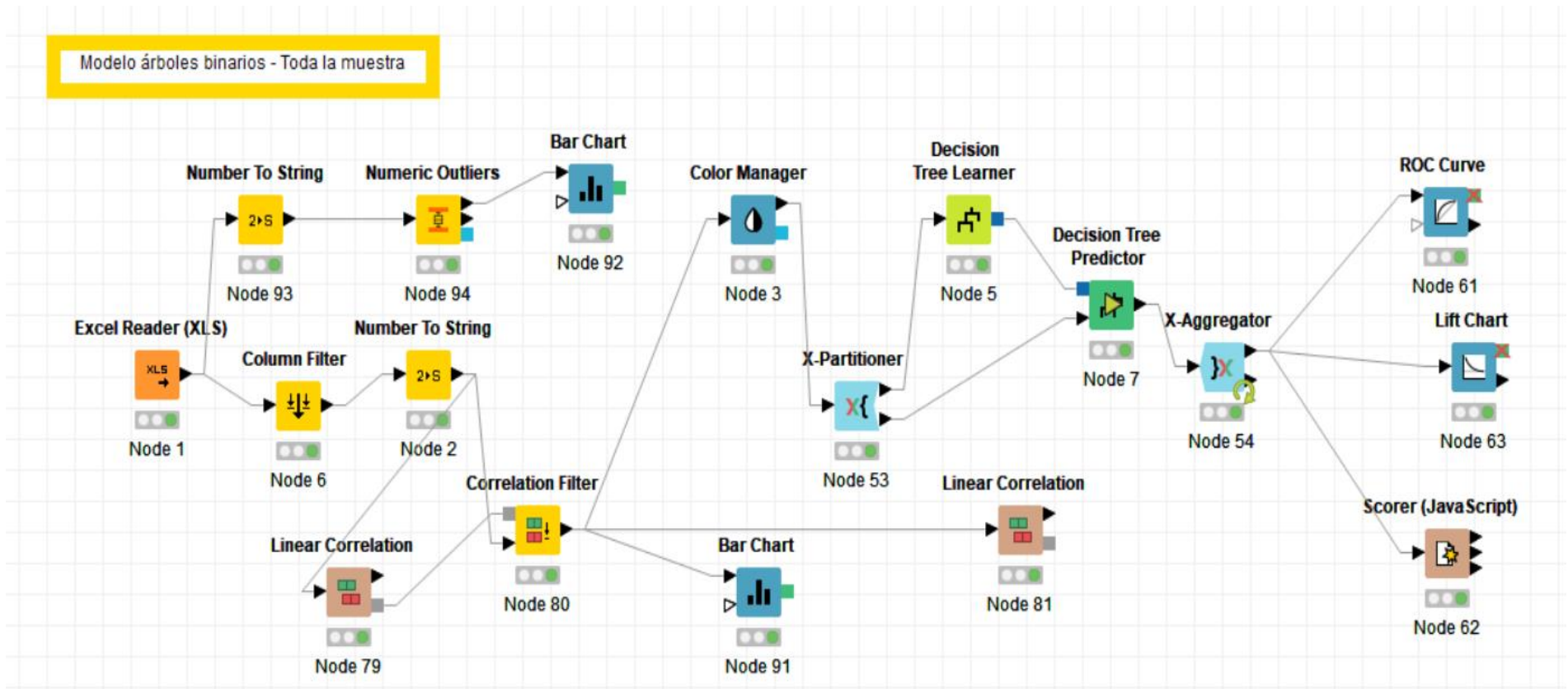
- Ruiz, C., & Basualdo, M. S. (2001). Redes Neuronales: Conceptos Básicos y Aplicaciones. Recuperado de <ftp://decsai.ugr.es/pub/usuarios/castro/Material-Redes-Neuronales/Libros/matich-redesneuronales.pdf>
- Rumelhart, D. E., Hinton, G. E., & Williams, R. J. (1986). Learning representations by back-propagating errors. *Nature*, 323(6088), 533-536. <https://doi.org/10.1038/323533a0>
- Santana, L. (2017). Determinantes de la supervivencia de microempresas en Bogotá: un análisis con modelos de duración. *Revista INNOVAR Journal*, 27(64), 50-61. <https://doi.org/10.15446/innovar.v27n64.62368>
- Siemens. (2020). Predictive Engineering Analytics. Recuperado 17 de marzo de 2020, de <https://www.plm.automation.siemens.com/global/en/our-story/glossary/predictive-engineering-analytics/13222>
- Thomas Ng, S., Wong, J. M. W., & Zhang, J. (2011). Applying Z-score model to distinguish insolvent construction companies in China. *Habitat International*, 35(4), 599-607. <https://doi.org/https://doi.org/10.1016/j.habitatint.2011.03.008>
- Turetsky, Howard, McEwen, R. (2001). An Empirical Investigation of Firm Longevity: A Model of the Ex Ante Predic...: Biblioteca Octavio Arizmendi Posada. *Review of Quantitive and Accounting*, 16, 323-342. Recuperado de <https://eds-a-ebcohost-com.ez.unisabana.edu.co/eds/pdfviewer/pdfviewer?vid=3&sid=eaeaf7ae-d4b5-4d6a-b906-774369b24d40@sessionmgr4008>
- Tya, Restianti, Agustina, L. (2018). The Effect of Financial Ratios on Financial Distress Conditions in Sub Industrial Sector Company. *Accounting Analysis Journal*. Recuperado de <https://journal.unnes.ac.id/sju/index.php/aaj/article/view/18996/11200>
- Universidad Católica de Oriente. (2013). Sectores productivos en Colombia. Recuperado de http://www.uco.edu.co/ova/OVA_Economia_Colombiana/Objetos_informativos/Unidad_1/2.SECTORES_PRODUCTIVOS_EN_COLOMBIA.pdf
- Uribe, C., Fonseca, S., Bernal, G., Contreras, C., & Castellanos, O. (2010). Sembrando Innovación para la competitividad del sector agropecuario Colombiano. Recuperado de http://www.bdigital.unal.edu.co/3567/1/Diagramacion_Libro_MADR_V2.pdf
- Villada, F., Muñoz, N., & García-Quintero, E. (2016). Redes Neuronales Artificiales aplicadas a la Predicción del Precio del Oro. *Información tecnológica*, 27(5), 143-150. <https://doi.org/10.4067/S0718-07642016000500016>

- Villanueva Mejia, D. (2018). Estudio sobre Bioeconomía como fuente de nuevas industrias basadas en el capital natural de Colombia Fase II. *Estudios sobre Bioeconomía*. Recuperado de [https://www.dnp.gov.co/Crecimiento-Verde/Documents/ejes-tematicos/Bioeconomia/Informe 2/ANEXO 1_Análisis sector agrícola.pdf](https://www.dnp.gov.co/Crecimiento-Verde/Documents/ejes-tematicos/Bioeconomia/Informe%202/ANEXO%201_Análisis%20sector%20agrícola.pdf)
- Wang, X., & Liu, F. (2020). Data-driven relay selection for physical-layer security: A decision tree approach. *IEEE Access*, 8, 12105-12116. <https://doi.org/10.1109/ACCESS.2020.2965963>
- Wooldridge, J. (2010). *Introducción a la Econometría*.
- Yiu, T. (2019). Understanding Random Forest - Towards Data Science. Recuperado 22 de febrero de 2020, de <https://towardsdatascience.com/understanding-random-forest-58381e0602d2>
- Zeng, P., Sun, X., & Farnham, D. J. (2020). Skillful statistical models to predict seasonal wind speed and solar radiation in a Yangtze River estuary case study. *Scientific Reports*, 10(1). <https://doi.org/10.1038/s41598-020-65281-w>

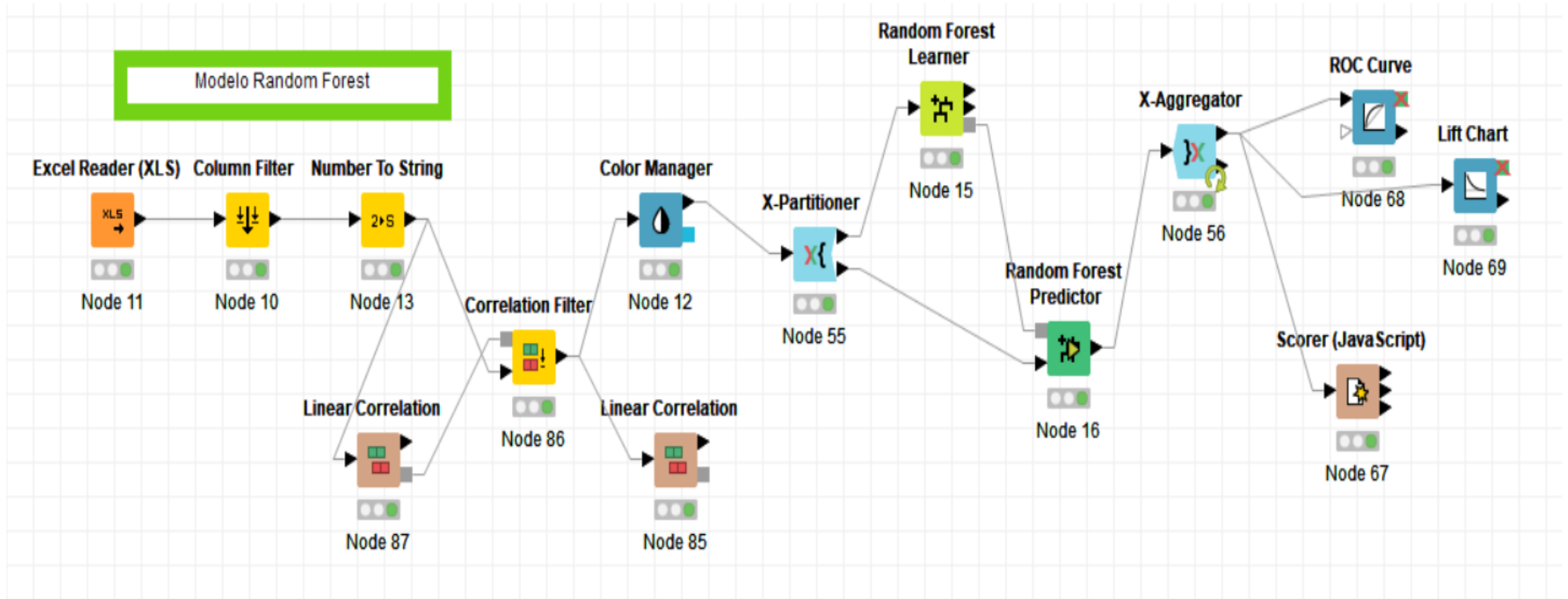
17. ANEXOS

17.1 Modelos Knime: Mejor modelo para toda la muestra: Árboles Binarios

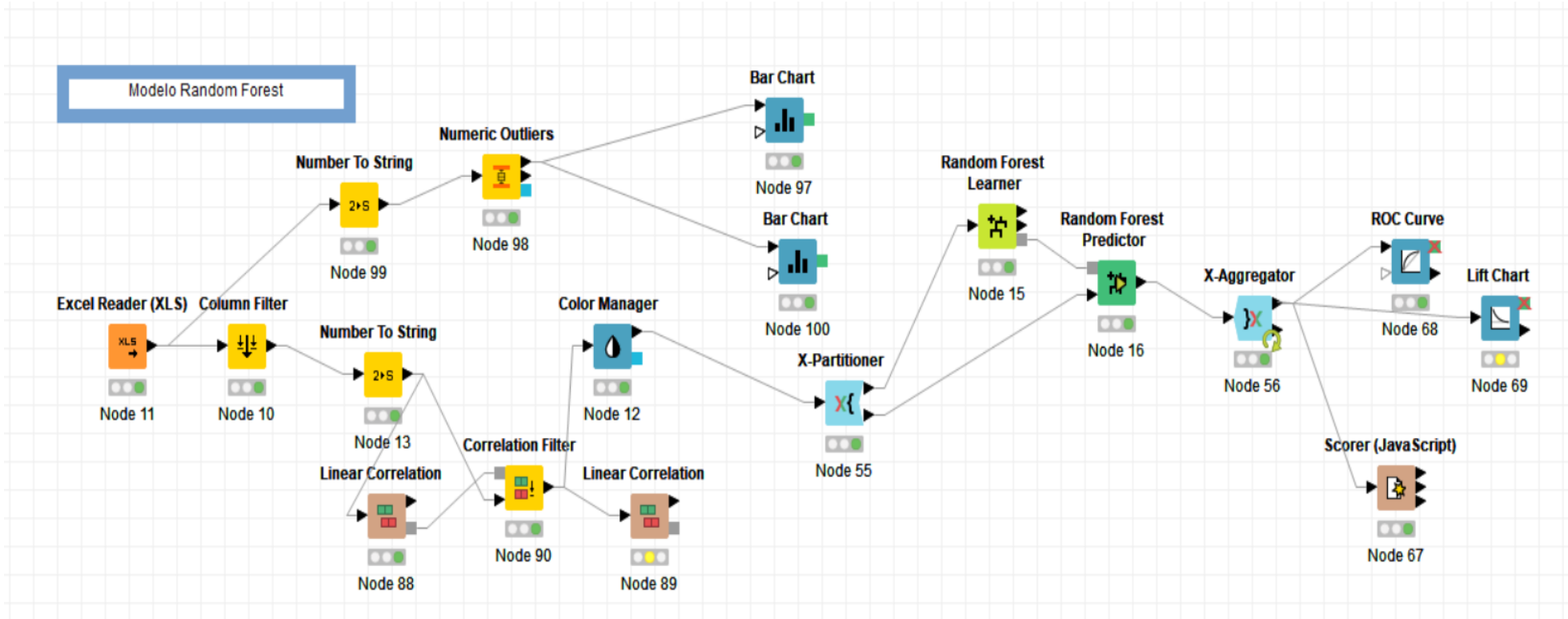
Estos modelos se construyeron en el programa KNIME y los diagramas muestran desde la lectura de la base de datos de Excel, la depuración, validación LOOCV, la construcción de los diferentes modelos, para terminar por último en las estadísticas de desempeño.



17.2 Modelos Knime: Mejor modelo Sector Agricultura: Random Forest



17.3 Modelos Knime: Mejor modelo Sector Industrial: Random Forest



17.4 Modelos Knime: Mejor modelo Sector Servicios: Random Forest

