

ESCUELA POLITÉCNICA NACIONAL

FACULTAD DE CIENCIAS

APLICACIÓN DE MODELOS NO PARAMÉTRICOS PARA
PREDECIR Y EXPLICAR EL RIESGO DE QUIEBRA EMPRESARIAL
DEL SECTOR INDUSTRIAL MANUFACTURERO ECUATORIANO.

TRABAJO DE TITULACIÓN PREVIO A LA OBTENCIÓN DEL TÍTULO DE
INGENIERO MATEMÁTICO

PROYECTO DE INVESTIGACIÓN

MAYRA ALEXANDRA MORENO VICENTE

mayra.moreno@epn.edu.ec

ROQUE IGNACIO MIÑO PAZOS

roque.mino@epn.edu.ec

Director: CARLOS ALBERTO ALMEIDA RODRIGUEZ, PH.D.

carlos.almeidar@epn.edu.ec

QUITO, DICIEMBRE 2022

DECLARACIÓN

“Nosotros, MAYRA ALEXANDRA MORENO VICENTE y ROQUE IGNACIO MIÑO PAZOS, declaramos bajo juramento que el trabajo aquí descrito es de nuestra autoría; que no ha sido previamente presentado para ningún grado o calificación profesional; y, que hemos consultado las referencias bibliográficas que se incluyen en este documento.

Sin perjuicio de los derechos reconocidos en el primer párrafo del artículo 114 del Código Orgánico de la Economía de los Conocimientos, Creatividad e Innovación -COESC-, cedo los derechos patrimoniales de este trabajo, a la Escuela Politécnica Nacional.”



MAYRA ALEXANDRA MORENO VICENTE



ROQUE IGNACIO MIÑO PAZOS

CERTIFICACIÓN

Certifico que el presente trabajo fue desarrollado por MAYRA ALEXANDRA MORENO VICENTE y ROQUE IGNACIO MIÑO PAZOS, bajo mi supervisión.

CARLOS ALBERTO ALMEIDA RODRIGUEZ, Ph.D.

Director del Proyecto

AGRADECIMIENTOS

Agradezco a Dios y a mi madre Virgen por la fortaleza y salud que me han dado para culminar mi carrera y brindarme fortaleza espiritual.

A mi madre Doris, hermanos y hermanas por el apoyo que me han brindado todos los días de mi vida estudiantil, por el esfuerzo y la dedicación que han puesto en mi para alcanzar una de mis metas.

A Roque, por apoyarme ante situaciones de desesperación, por ser un compañero leal, por su gran paciencia ante mis exigencias y por su esfuerzo dedicado en este trabajo de investigación.

A mi supervisor, Miguel Flores, por el tiempo empleado al desarrollo de esta investigación.

A mis amigos, Katherine Michay, Alexander Nenjer, Cristian Velastegui por hacer mi vida universitaria un ambiente más sociable y más armónico.

Mayra

AGRADECIMIENTOS

A Dios, por ser guía y fortaleza en todos los momentos que forjaron el destino de mi vida.

A mi madre, por haberme brindado la mejor educación de la vida para ser un hombre de bien.

A mi padre y mis hermanos, quienes son la motivación y fuente para alcanzar los grandes logros en mi vida.

A Alexandra, por haber sido la amiga y compañera en los momentos más complicados en mi vida.

Agradezco a mi director de tesis, Dr. Miguel Flores, por haberme brindado una adecuada orientación que permitió que esta tesis pueda ser finalizada a buen término.

A mis amigos de la Facultad a quienes agradezco de corazón cada momento compartido tanto buenos como malos y por quienes la Universidad se convirtió en mi segundo hogar.

Roque

DEDICATORIA

Dedico este trabajo de investigación a mi familia y amigos que siempre estuvieron a mi lado apoyándome.

Mayra

DEDICATORIA

Dedico esta tesis al más hermoso ángel que se encuentra en el cielo, a mi madre, a quien invito en espíritu el día de la defensa de este trabajo.

Roque

Índice general

| | |
|---|-----------|
| Resumen | 1 |
| Abstract | 2 |
| 1. Introducción | 3 |
| 1.1. Planteamiento del problema | 3 |
| 1.2. Objetivos | 5 |
| 1.2.1. Objetivo General | 5 |
| 1.2.2. Objetivos Específicos | 5 |
| 1.3. Justificación | 6 |
| 1.3.1. Justificación Teórica | 6 |
| 1.3.2. Justificación Metodológica | 7 |
| 1.3.3. Justificación Práctica | 9 |
| 1.4. Situación y desempeño del sector Industrial Manufacturero, durante el periodo 2016 – 2020 en Ecuador. | 10 |
| 1.4.1. Superintendencia de Compañías Valores y Seguros (SUPER-CÍAS) | 10 |
| 1.4.2. Sistema de Cuentas Nacionales (SCN) | 11 |
| 1.4.3. Sector "C" Industrias Manufactureras | 13 |
| 1.4.4. Incidencia del sector Industrial Manufacturero ecuatoriano en el Producto Interno Bruto (PIB), período 2016 – 2020 | 15 |
| 1.4.5. Evolución de las empresas en Ecuador, año 2016 – 2020. | 18 |
| 2. Marco Teórico | 23 |

| | | |
|--------|--|----|
| 2.1. | Definiciones y conceptos básicos | 24 |
| 2.1.1. | Conceptualización del riesgo | 24 |
| 2.1.2. | Tipos de riesgo | 26 |
| 2.1.3. | Medición del Riesgo | 27 |
| 2.2. | Riesgo de insolvencia | 29 |
| 2.3. | Modelos de predicción de quiebra | 31 |
| 2.3.1. | Etapa Descriptiva del desarrollo de modelos de predicción de quiebra a través de indicadores | 31 |
| 2.3.2. | Etapa Predictiva del desarrollo de modelos de predicción de quiebra | 32 |
| 2.4. | Modelos Paramétricos | 36 |
| 2.4.1. | Modelos Lineales Generalizados | 36 |
| 2.5. | Modelos No Paramétricos | 40 |
| 2.6. | Modelo Aditivo Generalizado | 43 |
| 2.6.1. | Introducción | 43 |
| 2.6.2. | Funciones Suaves - Caso Univariante | 45 |
| 2.6.3. | Regresión Polinómica | 45 |
| 2.6.4. | Splines Cúbicos de Regresión | 46 |
| 2.6.5. | Grado de suavizado: Regresión por Splines penalizadas | 51 |
| 2.6.6. | Estimación del Parámetro Suavizante λ | 52 |
| 2.6.7. | Modelos Aditivos | 53 |
| 2.6.8. | Modelos Aditivos Generalizados | 55 |
| 2.7. | Redes Neuronales | 55 |
| 2.7.1. | Redes neuronales como un modelo biológico | 56 |
| 2.7.2. | Redes neuronales artificiales | 57 |
| 2.7.3. | Algoritmo de aprendizaje <i>Backpropagation</i> | 67 |
| 2.7.4. | Formulación Matemática del algoritmo <i>Backpropagation</i> | 69 |
| 2.8. | Criterios Estadísticos para seleccionar variables y Medir el desempeño de los Modelos | 74 |
| 2.8.1. | Metodología para Selección de variables | 74 |

| | |
|---|------------|
| 2.8.2. Pruebas y Desempeño de los Modelos | 75 |
| 3. Metodología | 81 |
| 3.1. Datos y Variables | 81 |
| 3.1.1. Variable Dependiente | 91 |
| 3.2. Análisis exploratorio y tratamiento de datos | 92 |
| 3.2.1. Descripción Variables Cuantitativas | 92 |
| 3.2.2. Análisis de Datos Atípicos | 93 |
| 3.2.3. Tratamiento de Datos Atípicos | 97 |
| 3.2.4. Depuración de la base de datos | 100 |
| 3.2.5. Selección de la Muestra para Entrenamiento y Validación . . . | 104 |
| 3.2.6. Selección de Variables | 105 |
| 3.3. Estimación del Modelo Aditivo Generalizado | 107 |
| 3.3.1. Resultados y validación del modelo aditivo generalizado . . . | 114 |
| 3.4. Redes Neuronales Artificiales | 117 |
| 3.4.1. Construcción de una RNA | 117 |
| 4. Resultados e Interpretación | 125 |
| 5. Conclusiones y recomendaciones | 138 |
| 5.1. Conclusiones | 138 |
| 5.2. Recomendaciones | 140 |
| A. Estadísticos Descriptivos | 142 |
| B. Diagrama de Cajas y Bigotes Variables Numéricas | 144 |
| C. Promedio del riesgo de quiebra por Provincia y tamaño empresarial | 151 |
| Bibliografía | 166 |

Índice de figuras

| | |
|---|----|
| 1.1. PIB del sector industrial Manufacturero y PIB total (período 2016-2020) a millones de dólares del 2007 | 16 |
| 1.2. PIB a precios constantes (millones de dólares)/Tasa anual decrecimiento (período 2016-2020) | 17 |
| 1.3. Total empresas en Ecuador, período 2016-2020 | 18 |
| 2.1. Estados de evaluación | 25 |
| 2.2. Intervalos de corte para una base de $(m-1)$ polinomios. | 47 |
| 2.3. Función f suavizada con B -Spline | 48 |
| 2.4. Spline Cúbico. | 49 |
| 2.5. Partes de una neurona biológica | 56 |
| 2.6. Estructura de una neurona biológica | 57 |
| 2.7. Esquema de una red neuronal | 59 |
| 2.8. Neurona biológica vs neurona artificial | 59 |
| 2.9. Función de activación lineal | 61 |
| 2.10. Función de activación sigmoide | 62 |
| 2.11. Función de activación tangente hiperbólica | 63 |
| 2.12. Primer paso de una red neuronal | 65 |
| 2.13. Segundo paso de una red neuronal | 66 |
| 2.14. Tercer paso de una red neuronal | 67 |
| 2.15. Cálculo de los parámetros de la última capa | 70 |
| 3.1. Distribución y Diagrama de cajas de la variable <i>Liquidez Corriente</i> | 95 |
| 3.2. Distribución y Diagrama de cajas de la variable <i>Rotación de Cartera</i> | 96 |

| | |
|--|-----|
| 3.3. Distribución y Diagrama de cajas de la variable <i>liquidez corriente</i> tratada | 98 |
| 3.4. Distribución y Diagrama de cajas de la variable <i>Rotación de Cartera</i> tratada | 99 |
| 3.5. Diagrama de cajas de las variables <i>Total Activo</i> por tamaño Empresarial | 100 |
| 3.6. Diagrama de cajas de las variables <i>Total Pasivo</i> por tamaño Empresarial | 100 |
| 3.7. Diagrama de cajas de las variables <i>Total Patrimonio</i> por tamaño Empresarial | 101 |
| 3.8. Tasa de quiebra según el total de activos | 102 |
| 3.9. Tasa de quiebra según el total de pasivos | 102 |
| 3.10. Tasa de quiebra según el total de patrimonio | 103 |
| 3.11. Tasa de quiebra según el tamaño empresarial | 103 |
| 3.12. Predicción Parcial para la variable Endeudamiento Patrimonial . . . | 111 |
| 3.13. Predicción Parcial para la variable Rotación de ventas | 112 |
| 3.14. Predicción Parcial para la variable Margen Operacional | 113 |
| 3.15. Distribución de los residuos para el modelo (GAM - 3) | 113 |
| 3.16. Curva ROC para Modelamiento y Validación | 116 |
| 3.17. Red neuronal con una variable de entrada | 118 |
| 3.18. Red neuronal con una variable de entrada | 118 |
| 3.19. Red neuronal con dos variables de entrada | 119 |
| 3.20. Red neuronal con once variables de entrada y una capa oculta | 120 |
| 3.21. Red neuronal con once variables de entrada y una capa oculta con 2 nodos | 121 |
| 3.22. Red neuronal de 3 capas ocultas con 8 nodos total | 122 |
| 3.23. Curva ROC de la red neuronal para Modelamiento y Validación . . . | 123 |
| 4.1. Riesgo de Quiebra Estimado vs. Comportamiento Real en el sector Manufacturero. | 126 |
| 4.2. Riesgo de quiebra Estimado vs. comportamiento real para Empresas Grandes por año de estudio. | 128 |
| 4.3. Riesgo de quiebra Estimado vs. comportamiento real para Medianas Empresas por año de estudio. | 128 |

| | |
|--|-----|
| 4.4. Riesgo de quiebra Estimado vs. comportamiento real para Microempresas por año de estudio. | 129 |
| 4.5. Riesgo de quiebra Estimado vs. comportamiento real para Pequeñas Empresas por año de estudio. | 129 |
| 4.6. Riesgo de quiebra empresarial por segmento de riesgo en el año 2016. | 131 |
| 4.7. Riesgo de quiebra empresarial por segmento de riesgo en el año 2017. | 132 |
| 4.8. Riesgo de quiebra empresarial por segmento de riesgo en el año 2018. | 132 |
| 4.9. Riesgo de quiebra empresarial por segmento de riesgo en el año 2019. | 133 |
| 4.10. Riesgo de quiebra empresarial por segmento de riesgo en el año 2020. | 134 |
| 4.11. Riesgo de quiebra empresarial por segmento de riesgo en Empresas Grandes. | 135 |
| 4.12. Riesgo de quiebra empresarial por segmento de riesgo en Medianas Empresas. | 136 |
| 4.13. Riesgo de quiebra empresarial por segmento de riesgo en Pequeñas Empresas. | 137 |
| 4.14. Riesgo de quiebra empresarial por segmento de riesgo en Microempresas. | 137 |
| | |
| B.1. Distribución y Diagrama de cajas de la variable <i>Prueba Ácida</i> tratada | 145 |
| B.2. Distribución y Diagrama de cajas de la variable <i>Endeudamiento Patrimonial</i> tratada | 146 |
| B.3. Distribución y Diagrama de cajas de la variable <i>Endeudamiento Activo Fijo</i> tratada | 147 |
| B.4. Distribución y Diagrama de cajas de la variable <i>Rotación del Activo Fijo</i> tratada | 148 |
| B.5. Distribución y Diagrama de cajas de la variable <i>Rotación de Ventas</i> tratada | 149 |
| B.6. Distribución y Diagrama de cajas de la variable <i>Periodo medio de cobranza</i> tratada | 150 |

Índice de cuadros

| | |
|---|-----|
| 1.1. CIIU rev.4.0, estructura general | 13 |
| 1.2. Actividades del sector Industrial Manufacturero (2016 – 2020) | 14 |
| 1.3. PIB del sector Industrial Manufacturero y PIB total (período 2016-2020) en millones de dólares | 15 |
| 1.4. PIB a precios constantes (millones de dólares)/Tasa anual decrecimiento (período 2016-2020) | 17 |
| 1.5. Cantidad de empresas según sector económico en (valor numérico y porcentual) | 20 |
| 1.6. Empleo registrado por sector económico (número y porcentaje) | 21 |
| 2.1. Error de clasificación para predecir quiebras futuras | 34 |
| 2.2. Regla para el Valor de Información | 75 |
| 2.3. Matriz de Confusión | 79 |
| 2.4. Métricas de Desempeño | 80 |
| 3.1. Clasificación de las empresas según el tamaño | 88 |
| 3.2. Proporción de empresas, por tamaño | 89 |
| 3.3. Estadísticos Descriptivos de la Variable Endeudamiento del Activo | 92 |
| 3.4. Distribución número de empresas por año | 104 |
| 3.5. Distribución de la variable dependiente para entrenamiento y validación | 105 |
| 3.6. Test KS | 105 |
| 3.7. Valor de Información (VI) | 106 |
| 3.8. Modelo de estimación GAM-1 (Primer Modelo) | 108 |
| 3.9. Modelo de estimación GAM-2 (Segundo Modelo) | 109 |

| | |
|---|-----|
| 3.10. Modelo de estimación GAM-3 (Tercer Modelo) | 109 |
| 3.11. Modelo con mejor rendimiento estadístico | 110 |
| 3.12. Factor de Inflación generalizado (GVIF) GAM-3 | 114 |
| 3.13. Estadísticos de prueba | 115 |
| 3.14. Matriz de Confusión | 115 |
| 3.15. Métricas de evaluación | 116 |
| 3.16. Estadísticos de prueba | 123 |
| 3.17. Matriz de Confusión - Red Neuronal Artificial | 124 |
| 3.18. Métricas de evaluación - Red Neuronal Artificial | 124 |
| 4.1. Cantidad de empresas por periodo de tiempo | 125 |
| 4.2. Probabilidad de Riesgo por cada periodo de tiempo. | 126 |
| 4.3. Riesgo de quiebra empresarial por tamaño de empresa | 127 |
| 4.4. Análisis del riesgo de quiebra empersarial por provincia | 130 |
| A.1. Resumen Estadístico | 143 |
| C.1. Promedio del riesgo de quiebra por segmento de riesgo en el año 2016 | 152 |
| C.2. Promedio del riesgo de quiebra por segmento de riesgo en el año 2017 | 153 |
| C.3. Promedio del riesgo de quiebra por segmento de riesgo en el año 2018 | 154 |
| C.4. Promedio del riesgo de quiebra por segmento de riesgo en el año 2019 | 155 |
| C.5. Promedio del riesgo de quiebra por segmento de riesgo en el año 2020 | 156 |
| C.6. Riesgo de quiebra por segmento de riesgo en las grandes empresas . | 157 |
| C.7. Riesgo de quiebra por segmento de riesgo en las Medianas empresas | 158 |
| C.8. Riesgo de quiebra por segmento de riesgo en las Pequeñas empresas | 159 |
| C.9. Riesgo de quiebra por segmento de riesgo en las Microempresas . . . | 160 |

Resumen

El objetivo del presente trabajo es el aplicar métodos no paramétricos que ayuden a predecir el riesgo empresarial manufacturero ecuatoriano en el período de tiempo 2016-2020; analizando el comportamiento de características financieras que intervienen internamente en una empresa, así como también características sobre tamaño empresarial y región empresarial.

En primera instancia se realiza un análisis de estudio de la industria manufacturera; en este sentido se analiza su influencia en el PIB ecuatoriano dentro del período de estudio 2016-2020 y su influencia en sector laboral, dando a conocer cuán importante es para la economía del país y razón por la cual su estudio en particular para este sector económico.

Consecutivamente, se procede a estudiar métodos no paramétricos como el modelo aditivo generalizado y la técnica de redes neuronales artificiales, explicando la necesidad de poder acercarse más a la realidad de las variables que se involucran en cualquier problema del mundo real.

Debido a la complejidad de interpretación de los resultados que proporcionan las redes neuronales (RNA), ya que son una caja negra, es que se emplean los modelos aditivos generalizados (GAM), puesto que en ellos la interpretación de sus resultados es más flexible a la de las RNA, razón por la cual la RNA implementada en esta investigación recibe entradas o características de las empresas manufactureras, que son seleccionadas por el GAM que mejor estime. Con ello se logra aplicar ambos modelos, de tal manera que uno de ellos es complementario al otro en función de interpretación.

Palabras claves: Manufactura, riesgo empresarial, quiebra empresarial, GAM, RNA, aprendizaje de la red.

Abstract

The purpose of this study is to apply non-parametric methods that help predict the Ecuadorian manufacturing business risk in the 2016-2020 time period; analyzing the behavior of the financial characteristics that intervene internally in a company, as well as characteristics on business size and business region.

In the first instance, a study analysis of the manufacturing industry is carried out, in this sense its influence on the Ecuadorian GDP is analyzed within the study period 2016-2020 and its influence on the labor sector, revealing how important it is for the economy. of the country and reason why it was studied in particular for this economic sector.

Consecutively, non-parametric methods such as the generalized additive model and the artificial neural network technique are studied, explaining the need to be able to get closer to the reality of the variables involved in any real-world problem.

Due to the complexity of interpretation of the results guaranteed by neural networks (ANN), since they are a black box, generalized additive models (GAM) are used, since in them the interpretation of their results is more flexible to that of the ANNs, which is why the ANN implemented in this research receives inputs or characteristics from the manufacturing companies, which are selected by the GAM that it considers best. With this, it is possible to apply both models, in such a way that one of them is complementary to the other in terms of interpretation.

Keywords: Manufacturing, business risk, business bankruptcy, GAM, RNA, network learning.

Capítulo 1

Introducción

1.1. Planteamiento del problema

Todos los países que poseen un sistema financiero estable es gracias al desarrollo de varias actividades económicas; entre las principales se encuentran aquellas de carácter empresarial como son: industria petrolera, agricultura, transporte, construcción, enseñanza, industria manufacturera, entre otras. Todas estas actividades son importantes y con ellas las empresas, produciendo dinamismo y crecimiento económico para un país. [57]

Para Ecuador, las empresas constituyen un eje trascendental en el aporte de la economía local y regional, que según el Servicio de Rentas Internas (SRI) constituyen el 86 % de las recaudaciones efectuadas [59], teniendo una gran incidencia en el aporte a la economía ecuatoriana. Sin embargo, en los últimos años se ha experimentado un proceso de crisis o recesión debido a factores internos como externos que afectan a la economía del país, tales como: la reducción del precio del barril de petróleo que afecta el ingreso del estado, los bajos precios de las materias primas, la disminución de la demanda interna y el impacto directo de la emergencia sanitaria que se vive actualmente, entre otros.

Debido a lo expuesto con anterioridad es necesario plantear que en Ecuador se han complicado las problemáticas preexistentes, y con la presencia del Covid-19 todo esto conllevó a poner en riesgo la continuidad de algunas empresas afectando aún más al sector empresarial y en consecuencia a la economía del Ecuador[42], generando grandes cambios y afectaciones al proceso productivo nacional; esto se detecta en las estadísticas de la Cámara de Industrias y Producción donde detallan que en el año 2020 la industria manufacturera descendió en un 6,1 %, el comercio en

7,8 % y los servicios en 16 % [23]; presentando un sector manufacturero con pérdidas menores a diferencia de los sectores de comercio y servicios.

Con dichas estadísticas se observa que los sectores relacionados al comercio y servicio son los más afectados a diferencia del sector manufacturero; no obstante, el estudio de esta investigación no se centra en las empresas de los sectores económicos más afectados por la pandemia, sino en el riesgo de quiebra del sector empresarial que más movimiento económico genera dentro del país y que ha sido afectado y continúa siendo por la recesión económica; el cual corresponde al sector manufacturero. Dicho sector, durante el año 2019, fue el que más aportó a la economía con 9 mil millones de dólares, representando el 12,5 % de la producción total del país; de igual manera, fue el que más empleos generó para la población ecuatoriana; es así que para finales del año 2019 se contaba con 804.401 personas trabajando en el sector de transformación de materias primas [45], pero entre marzo y diciembre del año 2020 se tuvo una pérdida de 51.353 empleos [20].

Asimismo, según análisis y estudios de la Superintendencia de Compañías, Valores y Seguros (SUPERCIAS), las empresas han tenido una tendencia de comportamiento no esperado, ya que el número de empresas en el Ecuador desde el año 2016 hasta el año 2020 tuvo un crecimiento de 36.984 empresas (66.053 empresas en el año 2016 y 103.037 empresas para el 2020), representando un incremento del 56 %; no obstante, entre los años 2018 y 2020, en el sector manufacturero se observó una reducción del número de empresas, de 5.773 en el año 2018 a 5.395 empresas en el 2020 (lo que implica un decrecimiento del 6,55 %) [62].

Estos comportamientos, casos fortuitos y factores externos e internos son algunos problemas que repercuten en el sistema productivo, lo que implica un descenso en la producción, comercialización y como consecuencia una mayor exposición al riesgo de insolvencia financiera empresarial. Es por esta razón que los gerentes y la alta dirección en las empresas buscan constantemente herramientas de carácter cuantitativo, que les permita detectar a tiempo las condiciones financieras, para que de esta manera puedan anticiparse a las circunstancias y evitar inconvenientes que les conlleve a una insolvencia empresarial [38].

Muchos gerentes requieren del uso de diferentes modelos para realizar pronósticos y estimar la situación futura de su empresa, debido a la incertidumbre financiera que pueden afectar las actividades económicas de la misma. Por tal motivo, utilizan modelos de predicción de bancarrota (insolvencia financiera), los cuales permiten predecir si una empresa entra o no en quiebra y así medir el riesgo de insolvencia

financiera (bancarrota)[2].

Actualmente, existen métodos estadísticos de clasificación tales como: análisis múltiple discriminante (MDA), regresión Logística, modelo Probit, entre otros, que predicen el riesgo de quiebra empresarial; sin embargo, estos métodos consideran la linealidad de los datos dando paso a una predicción no muy certera.

Por ello, es necesario desarrollar modelos más robustos para el análisis de quiebra empresarial que consideren una posible relación no lineal que pueda existir entre variables predictoras y variable respuesta, permitiendo obtener un alto rendimiento de predicción y así evitar consecuencias grandes como el cierre de empresas y los altos costos que conllevan dichos procesos.

Es así que el presente trabajo propone el desarrollo de un modelo más robusto para ser aplicado en las empresas Industria Manufactureras ecuatorianas, midiendo así el riesgo de quiebra empresarial. El modelo del que se habla se denomina Redes Neuronales Artificiales (RNA), el cual tiene una estructura funcional no paramétrica que contemplan la posible no linealidad de las variables y un mayor poder de predictibilidad. El desarrollo del modelo propuesto contempla la creación de otros modelos no paramétricos que son los Modelos Aditivos Generalizados (GAM) para ayudarnos en la interpretabilidad de los resultados obtenidos por RNA, creando así un complemento entre ambos modelos.

1.2. Objetivos

1.2.1. Objetivo General

Aplicar un modelo no paramétrico, denominado red neuronal artificial, para determinar la probabilidad de quiebre empresarial en el sector de la Industria Manufacturera del Ecuador y, adicionalmente, complementar este modelo con la implementación de un modelo aditivo generalizado para facilitar la interpretabilidad de los resultados obtenidos por la red neuronal artificial.

1.2.2. Objetivos Específicos

- Describir y analizar las variables que intervienen en las empresas manufactureras correspondientes al tamaño de la empresa y los indicadores financieros, para preseleccionar aquellas que tienen mayor poder predictivo respecto a la

variable dependiente y sean incluidas como variables de entrada en el modelo GAM.

- Estimar un modelo GAM con las variables preseleccionadas, para obtener las variables que mejor expliquen el fracaso empresarial y con ellas desarrollar un modelo de Red Neuronal Artificial para la predicción de bancarrota.
- Validar la idoneidad de los modelos a través de criterios estadísticos y realizar un análisis retrospectivo (backtesting) de los modelos propuestos con base en los datos históricos.
- Pronosticar el quiebre empresarial de las empresas manufactureras a través de la RNA que muestre un menor error de predicción.

1.3. Justificación

1.3.1. Justificación Teórica

El problema de la crisis empresarial es un tema de actualidad; esta apreciación se ha derivado por el creciente interés que tienen los directivos de prevenir el fracaso empresarial, como también por la recesión económica del país, lo cual lleva a las empresas a enfrentar problemas de riesgos financieros[6].

Actualmente las organizaciones que son parte del sector manufacturero ecuatoriano cuentan con modelos estadísticos de predicción que permitan evaluar la probabilidad de bancarrota empresarial, de los cuales se pueden nombrar algunos de ellos: el análisis discriminante, la regresión logística, el método Altman, modelo de Ohlson; sin embargo, son de bajo poder predictivo y robustez, teniendo como consecuencia el reflejo de altos niveles de inestabilidad financiera por parte de las empresas del sector manufacturo [58].

Por tal motivo, el presente trabajo se realiza con la finalidad de aportar en el estudio del análisis de riesgo de insolvencia empresarial con el desarrollo de un nuevo modelo de quiebra que sea validado como instrumento de medición del riesgo de bancarrota para ser incorporado como una herramienta fundamental dentro del estudio de riesgo de quiebra en la Industria Manufacturera ecuatoriana, permitiendo segmentar las empresas de este sector como empresas en quiebra y no quiebra con un alto poder de acierto y un error de predicción mínimo. Dichos modelos son las

Redes Neuronales Artificiales (RNA), los cuales son algoritmos de inteligencia artificial que permiten cierto aprendizaje a través de la experiencia para discernir la relación entre las características de la situación financiera empresarial y la probabilidad de bancarrota [4].

A pesar que este algoritmo tiene un alto desempeño predictivo, tiene una gran falencia, la cual es que no son fácilmente explicables respecto a los resultados que proporciona; por ello, con el fin de solucionar este problema se propone también el desarrollo de un modelo aditivo generalizado (GAM), el cual considera, al igual que las RNA, la no linealidad de las variables y tiene una interpretabilidad más flexible que los algoritmos de inteligencia artificial; motivo por el cual se espera obtener de los GAM las variables que tienen mayor poder predictivo sobre la bancarrota empresarial para ser incluidas como variables principales en el desarrollo de las RNA y así proveer resultados no solamente con alto poder de predicción y un error de predicción pequeño, sino también una fácil interpretación de los mismos.

Se proporciona así, una alternativa a emplearse para minimizar el riesgo de bancarrota empresarial, complementándose con el desarrollo de un modelo estadístico no paramétrico (GAM) y de esta manera probar que el uso de estas metodologías pueden ayudar a mejorar el desempeño financiero de las empresas.

1.3.2. Justificación Metodológica

Los primeros estudios comenzaron con Fitzpatrick (1932) [32] utilizando el análisis financiero para tratar de entender el fenómeno de la quiebra empresarial; le siguieron los estudios de Beaver (1966) [7] y Altman (1968) [3] apoyados en las técnicas de análisis univariado y multivariado respectivamente, con el uso de indicadores financieros.

Con el paso del tiempo los estudios sobre la quiebra empresarial se han multiplicado, en parte por el desarrollo de la matemática y la estadística a través de la creación de nuevas metodologías, permitiendo el uso y la aplicación de técnicas avanzadas en la búsqueda del mejor modelo que permita predecir la quiebra empresarial; es así que, durante las últimas dos décadas se han dado numerosos estudios teóricos acerca de nuevos modelos para predicción de bancarrota fundamentados en métodos de machine learning (aprendizaje automático), los cuales están orientados hacia el rendimiento de la predicción con base en un conjunto de indicadores financieros. Entre los más populares de estos nuevos modelos de predicción se encuentran las redes neuronales [14].

Las Redes Neuronales Artificiales (RNA) surgieron originalmente como una simulación abstracta de los sistemas nerviosos biológicos, constituidos por un conjunto de unidades llamadas neuronas conectadas unas con otras y el primer modelo de red neuronal fue propuesto en 1943 por McCulloch y Pitts en términos de un modelo computacional de actividad nerviosa [44]; sin embargo, los estudios de investigación realizados con el uso de las redes neuronales para la predicción de quiebras empresariales comenzó en 1990, debido al desarrollo tecnológico efectuado en aquel entonces. Muchas investigaciones utilizaron y aún utilizan este método para estimar resultados empíricos, debido a que considera un enfoque no lineal proporcionando una relación más cercana a la realidad entre las variables[4].

En general, las redes neuronales artificiales tienen elevadas capacidades predictivas respecto a los modelos estadísticos paramétricos clásicos (regresión logística, el Probit, el análisis discriminante, entre otros), es decir, presentan menor sesgo en la predicción de bancarrota, lo que implica obtener resultados muy cercanos a la realidad; estos resultados se evalúan a través de un análisis ROC (característica operativa del receptor), el cual evalúa la precisión de las predicciones del modelo y se ayuda del cálculo del AUC (área bajo la curva), cuyo estadístico representa la probabilidad de que la predicción sea correcta.[64].

Otra forma de evaluar qué tan bien ha funcionado un modelo de red neuronal es a través de desarrollo de la matriz de confusión, de donde se obtiene la precisión o la exactitud (accuracy), el cual mide el porcentaje del número de empresas clasificadas correctamente en quiebra y no quiebra, señalando que mientras más alto es este porcentaje, mejor es el modelo [54].

Por otra parte, los modelos aditivos generalizados fueron introducidos a la literatura por Marx y Eilers para el caso unidimensional en 1998, por Durban en 2002 y Currie en 2004. Estos modelos surgen como una extensión de la regresión lineal múltiple, ya que combina la no linealidad y la regresión no paramétrica [27]. Berg (2006) usó modelos aditivos generalizados (GAM) como una alternativa no paramétrica para la predicción de quiebras, y las redes neuronales de capa única oculta, demostrando que el modelo GAM tiene un desempeño de interpretabilidad de los resultados significativamente más flexibles que las redes neuronales[8]. Asimismo, Castro et al. (2019) utilizó modelos aditivos generalizados para predecir el riesgo de insolvencia financiera, dando a conocer resultados donde el uso de este modelo se adapta de mejor manera a la estructura no lineal de los datos[18].

Por otro lado, respecto al poder de predicción de quiebre empresarial, Ginanni

et al. (2002) plantea que los modelos de RNA muestran mejor desempeño que los GAM; sin embargo, la interpretación de los resultados proporcionados por los modelos RNA son significativamente más complejos que los proporcionados por los GAM [34].

Por lo expuesto anteriormente, en este trabajo se propone desarrollar dos metodologías: Redes Neuronales Artificiales y Modelos Aditivos Generalizados. Las RNA para predecir el riesgo de bancarrota empresarial dentro del sector manufacturero y los Modelos Aditivos Generalizados como complemento del desarrollo de las Redes Neuronales, donde a través del GAM se proporcione las variables más significativas respecto a la quiebra empresarial, cuyas variables consideradas para este estudio corresponden a los indicadores financieros de las compañías, y puedan ser empleadas en las RNA para facilitar la interpretabilidad de los resultados que proporcione el modelo de red neuronal. Para validar los modelos se plantea calcular tanto el AUC como la exactitud o la precisión, los cuales permiten determinar que los modelos propuestos son viables y confiables para la evaluación de la probabilidad de riesgo de quiebre empresarial.

1.3.3. Justificación Práctica

El propósito de la investigación se fundamenta en los problemas financieros que viven día a día las empresas ecuatorianas del sector Industrial Manufacturero, los cuales se originan por una mala gestión administrativa financiera, el mal uso de recursos económicos, la falta de oportunidad para acceder al sector financiero formal y por los inconvenientes internos y externos que genera la recesión económica del país, además de la emergencia sanitaria que aún se vive. Esto deriva a una insolvencia económica y en consecuencia a una bancarrota. Por tal motivo, se proponen dos metodologías eficientes, complementando una a la otra (GAM complementa a RNA), que ayuden a las empresas a medir el riesgo de insolvencia financiera y de esta manera crear alertas que permitan evitar caer en una posible quiebra.

Los beneficiarios del desarrollo de esta investigación son las empresas manufactureras ecuatorianas, las cuales son tomadas para el estudio de esta investigación debido a que se enfrentan a cambios más rápidos en el tiempo y además es uno de los sectores que más aporta a los ingresos del país, así como a su desarrollo.

1.4. Situación y desempeño del sector Industrial Manufacturero, durante el periodo 2016 – 2020 en Ecuador.

La industria manufacturera es una rama de la manufactura y comercio basada en la fabricación, procesamiento y transformación de materias primas o bienes primarios en productos finales o insumos para otras industrias. "La industria manufacturera juega un rol preponderante en la economía de los países emergentes, pues es capaz de emplear a una porción importante de la población y producir materiales requeridos por sectores estratégicos".¹

Para Ecuador, el sector manufacturero es importante para la producción del país, puesto que juega un rol fundamental no solo en el sector productivo de la economía sino también como una fuente de empleo e ingresos al estado. La manufactura siempre ha sabido sobrevivir ante los factores de riesgo político, social o riesgos fortuitos (como pandemias o desastres naturales) que se han venido dando en el país y cuyos factores no le han permitido estabilizarse o crecer.[58]

Por ende, en esta sección lo que se pretende es describir la importancia de la industria manufacturera y su desempeño en la sociedad durante el período 2016 – 2020 en el Ecuador.

1.4.1. Superintendencia de Compañías Valores y Seguros (SUPER-CÍAS)

La Superintendencia de Compañías, Valores y Seguros (SUPER-CÍAS) "es el organismo técnico, con autonomía administrativa y económica, que controla la organización, actividades, funcionamiento, disolución y liquidación de las compañías y entidades en las circunstancias y condiciones establecidas por la ley".²

Este organismo proporciona con responsabilidad a la ciudadanía, valiosa información sobre las compañías registradas en Ecuador a través del *Portal de Información* de su página web www.supercias.gob.ec; siendo el Ranking empresarial e información de las compañías, algunas de ellas.

Respecto a la información de compañías, la información que ofrece la SUPER-

¹Encyclopedia Britannica, 2015

²www.supercias.gob.ec

CIAS corresponde a los estados financieros de las empresas activas que presentan sus balances. Adicionalmente, detalla el nombre legal de cada compañía, código y descripción de la actividad económica al sexto nivel, es decir, a la actividad económica específica que desarrolla la compañía; como también la región, provincia, ciudad, tamaño de la empresa e incluso la clasificación industrial internacional uniforme de todas las actividades económicas (CIIU). Este último es una parte importante dentro del Sistema de Cuentas Nacionales (SCN), del cual se hablará en la siguiente sección.

1.4.2. Sistema de Cuentas Nacionales (SCN)

El Sistema de Cuentas Nacionales (SCN) “consta de un conjunto coherente sistemático e integrado de cuentas macroeconómicas, balances y cuadros basados en un conjunto de conceptos, clasificaciones y reglas contables aceptadas internacionalmente. Ofrece un marco contable amplio dentro del cual pueden elaborarse y presentarse datos económicos en un formato destinado al análisis económico, a la toma de decisiones y a la formulación de la política económica; las cuentas en sí misma presentan, en forma condensada un gran volumen de información detallada, organizada de acuerdo con determinados principios y percepciones acerca del funcionamiento de la economía.

Constituyen un registro completo y pormenorizado de las complejas actividades económicas que tienen lugar dentro de una economía y de la interacción entre los diferentes agentes económicos, o grupos de los mismos, que tiene lugar en los mercados o en otros ámbitos”.³

Por ello, una parte importante dentro del SCN es la Clasificación Industrial Internacional Uniforme (CIIU), elaborado por las Naciones Unidas, la cual recoge información en determinados momentos del tiempo, tanto de las diversas actividades económicas como también de los niveles de los activos productivos de una economía.

En Ecuador, el CIIU es ajustado por el Instituto Nacional de Estadísticas y Censos (INEC), el cual es el ente coordinador y orientador del Sistema Estadístico Nacional (SEN) que tiene como obligación incursionar dentro del mundo de la ciencia y la tecnología, los avances en la materia económica política y social, de tal manera que pueda orientar sus estadísticas hacia la consecución de los estándares de innovación, calidad, excelencia y oportunidad.[51]

³NACIONES UNIDAS. Sistema de Cuentas Nacionales, 1993

Clasificación Industrial Internacional Uniforme (CIIU)

La clasificación Industrial Internacional Uniforme de todas las actividades económicas (CIIU), elaborada por la División de Estadísticas de las Naciones Unidas, “es la clasificación internacional de referencia de las actividades económicas productivas. Su principal finalidad es facilitar un conjunto de categorías de actividad que pueda utilizarse para la elaboración de actividades económicas que desarrollan las empresas”. [52]

La versión original de la CIIU fue aprobada en el año 1948, y desde aquel año la mayoría de los países de mundo han utilizado la CIIU. En tal virtud la CIIU ha aportado orientación a los países para elaborar las clasificaciones de las actividades, convirtiéndose en una herramienta importante para la comparación de datos estadísticos sobre actividades económicas a escala internacional.

En 1956 la Comisión de Estadística analizó y revisó la CIIU y en 1958 publicó la primera revisión como Informes estadísticos, Serie M, No. 4.0, Rev. 1.

La segunda revisión fue hecha en 1968 como Informes Estadísticos, Serie M, No.4, Rev. 2; la tercera fue aprobada en febrero de 1989, esta revisión, Rev. 3 se la realizó con la finalidad de que se armonice con otras clasificaciones de actividades con las clasificaciones de bienes y servicios. Para el año 2002, se aprobó la revisión 3.1 de la CIIU, la cual tenía la misma estructura de la CIIU Revisión 3, a excepción de tres grupos de actividad económica. [53]

Con la llegada del siglo XXI, se trajo consigo un mundo dinámico y globalizado, basado en la comunicación y en la accesibilidad de la información al instante, contando con la concientización para la preservación del medio ambiente, se han generado nuevas actividades en los diferentes sectores económicos, con el fin de fortalecer y optimizar la utilización de los recursos y el tiempo en el desarrollo de las actividades.

Con este panorama la Comisión de Estadística de Estadística de Naciones Unidas vio la pertinencia de generar y presentar a la comunidad estadística una nueva revisión de la CIIU, llegando a ser la cuarta revisión, donde estuvieran todas las actividades que no se encontraban en versiones anteriores y que dieran importancia en la actualidad, en especial en actividades relacionadas a la tecnología de la información y las comunicaciones, las actividades financieras, profesionales y de apoyo a las empresas, además de la reorganización de las actividades en todos los sectores económicos. [50]

Es así que la versión 4, Rev. 4, de la CIIU es la versión a ser utilizada en el desarrollo del presente estudio, ya que esta versión corresponde al período de estudio, 2016–2020. Esta versión de la CIIU contiene desagregadas las actividades económicas en 21 secciones, las cuales se detallan a continuación en el cuadro 1.1.

Cuadro 1.1: CIIU rev.4.0, estructura general

| SECCIÓN | DESCRIPCIÓN |
|---------|---|
| A | Agricultura, ganadería, silvicultura y pesca |
| B | Explotación de minas y canteras |
| C | Industria manufacturera |
| D | Suministro de electricidad, gas, vapor y aire acondicionado |
| E | Distribución de agua; alcantarillado, gestión de riesgo y actividades de saneamiento. |
| F | Construcción |
| G | Comercio al por mayor y menor; reparación de vehículos automotores y motocicletas. |
| H | Transporte y almacenamiento. |
| I | Actividades de alojamiento y servicios de comidas |
| J | Información y comunicación |
| K | Actividades financieras y de seguros |
| L | Actividades Inmobiliarias |
| M | Actividades profesionales, científicas y técnicas |
| N | Actividades de servicios administrativos y de apoyo |
| O | Administración pública y defensa; planes de seguridad social de afiliación obligatoria |
| P | Enseñanza |
| Q | Actividades de atención de la salud humana y de asistencia social |
| R | Artes, entretenimiento recreación |
| S | Otras actividades de servicios |
| T | Actividades de los hogares como empleadores; actividades no diferenciadas de los hogares como productos de bienes servicios para uso propio |
| U | Actividades de organizaciones y órganos extraterritoriales |

Fuente: Naciones Unidas, CIIU, Revisión 4.0

Elaboración: Los Autores

1.4.3. Sector "C" Industrias Manufactureras

La CIIU identifica a las industrias manufactureras como las actividades dedicadas a la transformación química o física, sustancias o componentes en productos nuevos. A la transformación de los materiales, sustancias o componentes se los co-

noce como materias primas procedentes de la agricultura, ganadería, silvicultura, pesca y explotación de minas, así como productos de otras actividades manufactureras. Este sector incluye 23 actividades que se muestran en el cuadro 1.2.

Cuadro 1.2: Actividades del sector Industrial Manufacturero (2016 – 2020)

| CÓDIGO | ACTIVIDAD |
|---------------|--|
| C10 | Elaboración de productos alimenticios |
| C11 | Elaboración de bebidas |
| C12 | Elaboración de productos de tabaco |
| C13 | Fabricación de productos textiles |
| C14 | Fabricación de prendas de vestir |
| C15 | Fabricación de cueros y productos conexos |
| C16 | Producción de madera y fabricación de productos de madera y corcho, excepto muebles |
| C17 | Fabricación de papel y de productos de papel |
| C18 | Impresión y reproducción de grabaciones |
| C19 | Fabricación de Coque y de productos de la refinación del petróleo |
| C20 | Fabricación de sustancias y de químicos |
| C21 | Fabricación de farmacéuticos, sustancias químicas medicinales y productos botánicos de uso farmacéutico. |
| C22 | Fabricación de productos de caucho y de plástico |
| C23 | Fabricación de otros productos mineraes no metálicos |
| C24 | Fabricación de metales comunes |
| C25 | Fabricación de productos elaborados de metal, excepto maquinaria y equipo |
| C26 | Fabricación de productos de informática, electrónica y óptica |
| C28 | Fabricación de maquinaria y equipo n.c.p. ³ |
| C29 | Fabricación de vehículos, automotores, remolques y semiremolques |
| C30 | Fabricación de otros tipos de equipos de transporte |
| C31 | Fabricación de muebles |
| C32 | Otras industrias manufactureras |
| C33 | Reparación e instalación de maquinaria y equipo |

Fuente: Instituto Nacional de Estadísticas y Censos (INEC)

Elaboración: Los Autores

4

⁴No clasificado en otra parte (n.c.p)

1.4.4. Incidencia del sector Industrial Manufacturero ecuatoriano en el Producto Interno Bruto (PIB), período 2016 – 2020

En esta sección se mostrará la influencia de las industrias manufactureras sobre el PIB, analizando los valores del PIB anual, como también la tasa anual del PIB correspondientes a dicho sector, durante los años 2016, 2017, 2018, 2019 y 2020.

Cabe señalar que la información para este análisis se la obtiene del Banco Central del Ecuador (BCE), cuya institución proporciona sus bases de datos, en donde discriminan la producción manufacturera de la refinación de petróleo, la cual según la CIIU es parte del sector “C” Industria Manufacturera. Sin embargo, esto no es problema alguno, pues se agrega dicha información de acuerdo a la CIIU.

Cuadro 1.3: PIB del sector Industrial Manufacturero y PIB total (período 2016-2020) en millones de dólares

| ACTIVIDAD ECONÓMICA/AÑO | 2016 | 2017 | 2018 | 2019 | 2020 |
|--------------------------------------|---------------|---------------|---------------|---------------|---------------|
| Manufacturera (*) | 8.016 | 8.265 | 8.364 | 8.414 | 7.835 |
| Refinación de petróleo | 639 | 699 | 652 | 530 | 428 |
| Manufactura y Refinación de petróleo | 8,655 | 8.964 | 9.016 | 8.944 | 8.263 |
| PIB TOTAL | 69.314 | 70.956 | 71.870 | 71.879 | 66.308 |

(*) Excluye la actividad de refinación y petróleo

Fuente: Banco Central del Ecuador (BCE)

Elaboración: Los Autores

En la tabla 1.3 se puede evidenciar el crecimiento del PIB ecuatoriano (ver el PIB total) durante el periodo 2016-2019, registrándose el mayor crecimiento del período al año 2017, con 1.642 millones de dólares. Sin embargo, el 2019 no fue un gran crecimiento debido a que, en este año la producción y el consumo de combustibles disminuyeron; por una parte, la producción de combustible cayó un 6 % respecto al año 2018 y, por otra, el consumo bajó 2,3 % al pasar de 255.289 barriles diarios en 2018 a 249.315 barriles diarios en 2019.⁵

Para el 2020, se tiene una gran caída de 5.571 millones de dólares respecto al año 2019, y esto debido a la pandemia del COVID-19 que se sufrió mundialmente.

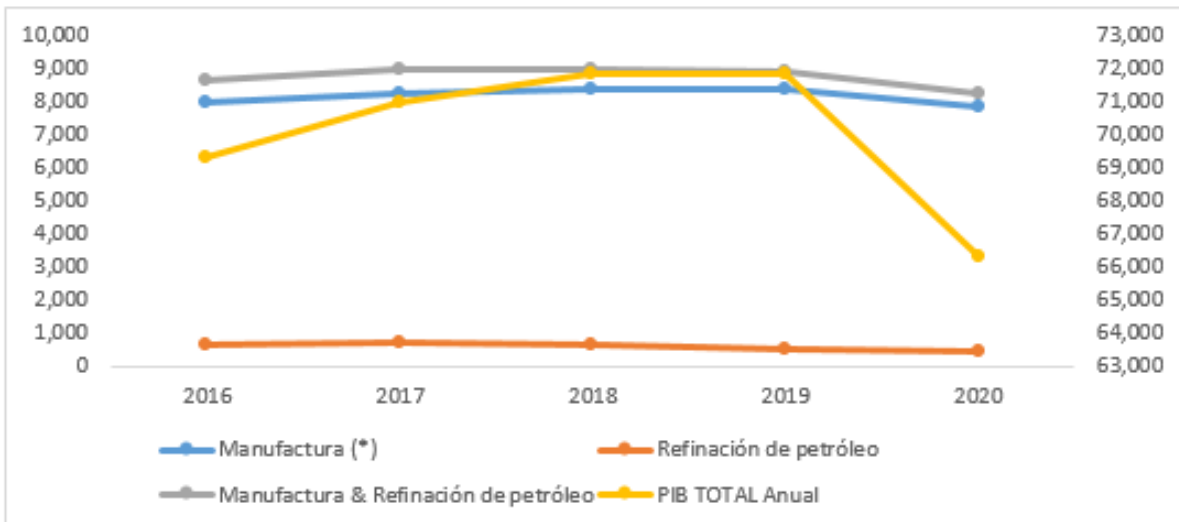
Además, se evidencia en la figura 1.1 la información de la tabla 3, donde el PIB generado por la Industria Manufacturera tiene un crecimiento en los años 2016, 2017 y 2018; pero en los dos últimos años del período en estudio, 2019 y 2020, empieza a

⁵Primicias. Refinerías pararon 23 veces en Ecuador, año 2019.

decaer. En el 2019 decae por la refinación de petróleo que como se dijo anteriormente se debe a la baja producción de combustible y decreciente consumo del mismo para este año. En el 2020, no sólo cae la refinación de petróleo, sino también la manufactura no petrolera, debido a la pandemia que afectó negativamente a la economía ecuatoriana, tomando en cuenta que ya venía con una recesión económica desde el año 2019, provocando un decrecimiento en los sectores económicos y en particular en la industria manufacturera de un 5,2%.⁶

Cabe destacar que la industria manufacturera no petrolera, se mantuvo en crecimiento hasta el año 2019, como se aprecia en la figura 1.1 y la tendencia del PIB generada por el sector industrial manufacturero es similar a la tendencia del PIB total, lo que muestra el impacto que tiene este sector en el PIB total para la economía ecuatoriana, en donde excluyendo o no la refinación petrolera, no deja de ser un sector estratégico.

Figura 1.1: PIB del sector industrial Manufacturero y PIB total (período 2016-2020) a millones de dólares del 2007



(*) Excluye la actividad de refinación de petróleo

Fuente: Banco Central del Ecuador (BCE)

Elaboración: Autores

Asimismo, se presenta la tasa anual de crecimiento del PIB, desagregado en manufactura y refinación de petróleo, como se observa en la tabla 1.4.

⁶Ekos.ec, ecuador productivo, año 2021

Cuadro 1.4: PIB a precios constantes (millones de dólares)/Tasa anual decrecimiento (período 2016-2020)

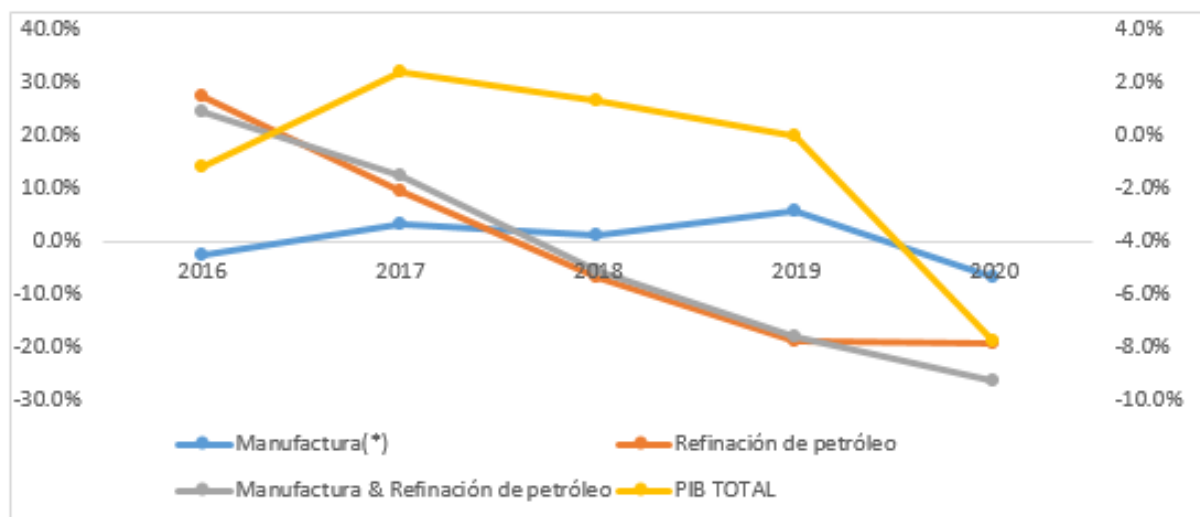
| ACTIVIDAD ECONÓMICA/AÑO | 2016 | 2017 | 2018 | 2019 | 2020 |
|--------------------------------------|---------------|--------------|--------------|---------------|---------------|
| Manufacturera(*) | -2,6 % | 3,1 % | 1,2 % | 6,0 % | -6,9 % |
| Refinación de petróleo | 27,4 % | 9,5 % | -6,7 % | -18,7 % | -19,3 % |
| Manufactura y Refinación de petróleo | 24,8 % | 12,6 % | -5,5 % | -18,1 % | -26,2 % |
| PIB TOTAL | -1,2 % | 2,4 % | 1,3 % | 0,01 % | -7,8 % |

(*) Excluye la actividad de refinación y petróleo

Fuente: Banco Central del Ecuador (BCE)

Elaboración: Los Autores

Figura 1.2: PIB a precios constantes (millones de dólares)/Tasa anual decrecimiento (período 2016-2020)



(*) Excluye la refinación de petróleo

Fuente: Banco Central del Ecuador (BCE)

Elaboración: Autores

Para el periodo de estudio 2016-2020, el sector manufacturero no petrolero tiene una tasa de crecimiento hasta el año 2019; el 2020 muestra un duro golpe en este sector económico, llegando a tener una tasa de crecimiento de $-6,9\%$ anual, efecto de la emergencia sanitaria por el COVID-19. La incidencia de la producción de petróleo se evidencia en la tasa anual de crecimiento del PIB, la baja producción, el bajo consumo de combustible y la pandemia, generó un alto decrecimiento en más del 26 % en la Manufactura y Refinación de Petróleo. Situación diferente se tuvo durante el 2016 y 2017, donde su crecimiento tuvo una tasa anual de más de 24 % y 12 % respectivamente.

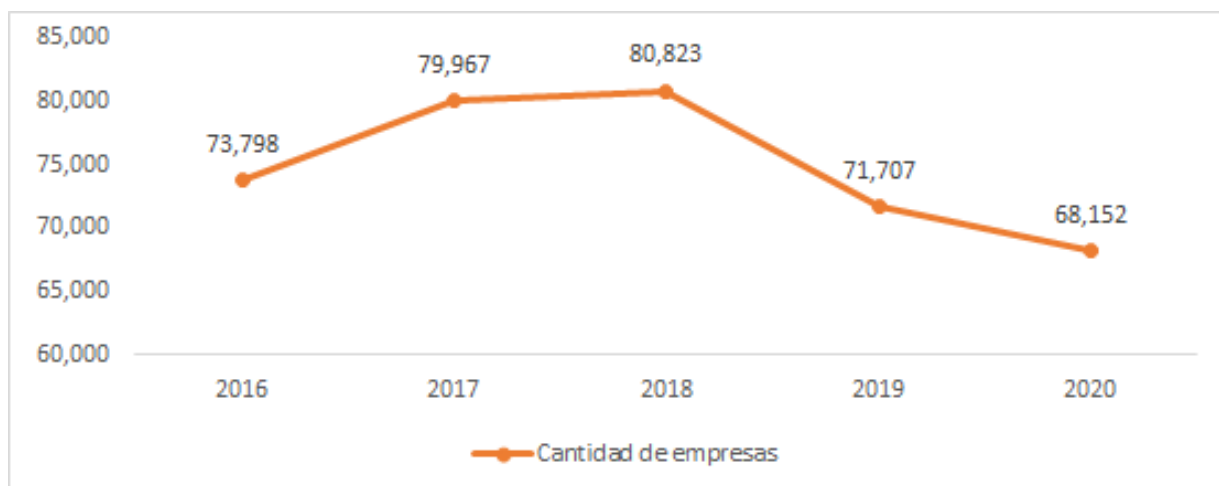
El PIB total muestra que la economía ecuatoriana decreció en un 7,8 %, lo que equivale a que, en términos constantes, totalizó USD 63,308 millones en el año 2020.⁷

En la figura 1.2, se observa como el sector de la manufactura y refinación de petróleo empieza a decrecer desde el año 2018 en adelante, mostrando una tendencia similar la tasa de crecimiento del PIB total, resaltando una vez más la importancia que del sector manufacturero dentro de la economía ecuatoriana.

1.4.5. Evolución de las empresas en Ecuador, año 2016 – 2020.

En la gráfica 1.3 se puede observar de manera general la cantidad de empresas que se han ido generando a lo largo de los años.

Figura 1.3: Total empresas en Ecuador, período 2016-2020



Elaboración: Los Autores

Se visualiza que la cantidad de empresas ha ido en crecimiento durante los años 2016, 2017, 2018, pero para los dos últimos dos años, 2019 y 2020 la disminución de empresas respecto al año 2018 es aproximadamente del 11 % y 15,7 %, respectivamente.

La caída de empresas en el año 2019 se debe al escaso número de constituciones (creaciones de empresas) en este año y la causa de ello según Santiago García, economista y docente de la Universidad Central del Ecuador, es por el pesado trámite que se debe realizar para la constitución de una compañía, en donde el empresario se demora al menos el 50 % de horas como mínimo para inicios de la tramitología. Otra de las causas es porque se tiene poco acceso al crédito. Según la Cámara de la

⁷Banco Central del Ecuador, boletines de prensa. la economía ecuatoriana decrece. Año 2020

Pequeña y Mediana Industria (Capeipi), solo para el funcionamiento de una Pyme existen 60 instituciones públicas que piden permisos, con lo que se reafirma que en Ecuador no hay mayores oportunidades y sea un país difícil para la constitución de nuevas empresas.⁸

Para el año 2020, debido a la emergencia sanitaria que vivió el país, el número de empresa decreció aún más, y a pesar de que hubo emprendedores que decidieron formalizarse como personas jurídicas, a través del nuevo tipo de sociedad mercantil⁹ que surgió en el año 2020, el cual corresponde a la sociedad por acciones simplificadas (SAS)¹⁰, no se pudo evitar el cierre de muchas empresas, sobretodo de las pequeñas compañías que estaban constituidas entre 10 y 49 trabajadores, debido a la pandemia que afectó no solo a nivel nacional sino a nivel mundial.

Ahora, es necesario también considerar la evolución del número de empresas a nivel sectorial y con ello observar, en particular, el sector de las industrias manufactureras, el cual es el centro de estudio en esta investigación.

En el cuadro 1.5, al analizar la estructura del número de empresas de los distintos sectores económicos en el período 2016 – 2020, el sector de *Servicios* sobresale con mayor número de empresas,¹¹ con una participación promedio de aproximadamente el 53 % a lo largo del período; y, ha pasado de representar el 53 % el 2016 a ser el 52 % en 2020. La sigue de cerca el Comercio, que muestra un crecimiento en su participación de 4 puntos porcentuales, del 23 % en el 2016 al 27 % en el 2020.

Respecto a la Industria Manufacturera, en el período 2016-2020 se tiene un crecimiento del 0,67 %, y para el año 2019 se presenta en valores absolutos un decrecimiento de 916 empresas respecto al año 2018, que corresponde en términos porcentuales al 0,15 %.

En el año 2020 los sectores Industrial Manufacturero y Comercio, a pesar de haber vivido la emergencia sanitaria, la tendencia aumenta en comparación al año 2019, debido a que dichos sectores trataron de adaptarse rápidamente al cambio di-

⁸Revista Primicias, empresas en Ecuador decrecen en comparación a los países vecinos.

⁹Según el código de Comercio, la sociedad mercantil es la personalidad jurídica que se crea para iniciar una actividad económica con fines de lucro.

¹⁰Según señala la revista Primicias, la figura mercantil de sociedades por acciones simplificadas entró en vigencia en Ecuador el 18 de mayo del 2020, y este tipo de figura mercantil pueden ser constituidas por una sola persona y no tienen un capital mínimo para constituirse.

¹¹A nivel de grandes sectores económicos, los Servicios incluyen las actividades de: Suministros de electricidad, gas, vapor; agua alcantarillado, desechos y saneamiento; transporte y almacenamiento; alojamiento y servicios de comidas; información y comunicación; actividades financieras y de seguros; actividades inmobiliarias; actividades profesionales, científicas y técnicas; servicios administrativos y de apoyo; administración pública y defensa; enseñanza; atención a la salud humana y de asistencia social; artes; entretenimiento, recreación y otras actividades de servicios.

Cuadro 1.5: Cantidad de empresas según sector económico en (valor numérico y porcentual)

| SECTOR ECONÓMICO/AÑO | 2016 | 2017 | 2018 | 2019 | 2020 |
|---|-------------------|-------------------|-------------------|-------------------|-------------------|
| Servicios | 39.080 52,96 % | 42.638 53,32 % | 43.485 53,80 % | 38.872 54,21 % | 35.167 51,60 % |
| Comercio | 16.773 22,73 % | 17.939 22,43 % | 18.190 22,51 % | 16.422 22,90 % | 18.527 27,18 % |
| Agricultura, ganadería, silvicultura y pesca | 5.016 6,80 % | 5.531 6,92 % | 5.549 6,87 % | 4.749 6,62 % | 5.262 7,72 % |
| Industria Manufacturera | 5.679 7,70 % | 6.054 7,57 % | 6.119 7,57 % | 5.203 7,26 % | 5.707 8,37 % |
| Construcción | 6.403 8,68 % | 6.807 8,51 % | 6.488 8,03 % | 5.634 7,86 % | 2.557 3,75 % |
| Explotación Minas y Canteras | 847 1,15 % | 998 1,25 % | 992 1,23 % | 827 1,15 % | 932 1,37 % |
| TOTAL | 73.798 | 79.967 | 80.823 | 71.707 | 68.152 |

Fuente: Superintendencia de Compañías, Valores y Seguros (SUPERCIAS)

Elaboración: Los Autores

gital. Es así que para el mes de octubre fue cuando más empresas se crearon, según la SUPERCIAS, y esto fue justamente por el nuevo tipo de sociedad mercantil, SAS; tomando en cuenta que la constitución de este tipo de sociedad mercantil no señala que se hayan creado nuevas empresas, sino más bien que hubo emprendedores que decidieron formalizarse como personas jurídicas.¹²

Además, el cuadro 1.5 muestra que, entre los tres primeros sectores económicos que están conformados por una sola actividad económica (excluye al sector de Servicios) y son más representativas en cuanto al número de empresas dentro del período de estudio, se encuentra la Industria Manufacturera, dándose así a conocer una vez más la importancia que tiene dentro del sector empresarial en el Ecuador; por tanto, generar más empresas en este sector conlleva a generar más empleo e innovación para el país.

A continuación, se analizará cómo afecta la variación del número de empresas en el período de estudio 2016 – 2020 al empleo en el país y en particular la influencia del sector manufacturero en el ámbito laboral.

¹²Revista Primicias, Economía mas empresas, constitución 2020, Ecuador.

Empresas manufactureras en el Ecuador y su influencia en el sector laboral

Ahora, se procede a realizar un análisis del empleo, señalando antes que este estudio se lo efectúa debido a que en una economía siempre es relevante conocer cuáles serán los sectores que en forma estratégica influyan en la producción y generación de ingresos.

En el cuadro 1.6 se puede observar el número de personal ocupado por sector económico y su representación porcentual respecto al total por cada año.

Cuadro 1.6: Empleo registrado por sector económico (número y porcentaje)

| SECTOR ECONÓMICO/AÑO | 2016 | 2017 | 2018 | 2019 | 2020 |
|---|----------------------|----------------------|----------------------|----------------------|----------------------|
| Servicios | 1,612,961 54.94 % | 1,632,613 55.51 % | 1,681,273 55.81 % | 1,698,838 56.04 % | 1,606,809 56.48 % |
| Comercio | 542,840 18.49 % | 536,753 18.25 % | 552,793 18.35 % | 555,670 18.33 % | 527,356 18.54 % |
| Industria Manufacturera | 396,636 13.51 % | 387,639 13.18 % | 391,925 13.01 % | 387,726 12.79 % | 355,268 12.49 % |
| Agricultura, ganadería, silvicultura y pesca | 214,318 7.30 % | 217,643 7.40 % | 225,636 7.49 % | 232,817 7.68 % | 223,358 7.85 % |
| Construcción | 132,407 4.51 % | 131,762 4.48 % | 125,018 4.15 % | 118,834 3.92 % | 98,398 3.46 % |
| Explotación Minas y Canteras | 36,698 1.25 % | 34,705 1.18 % | 35,849 1.19 % | 37,590 1.24 % | 33,829 1.19 % |
| TOTAL | 2,935,860 | 2,941,115 | 3,012,494 | 3,031,475 | 2,845,018 |

Fuente: Directorio de empresas y establecimientos 2016, 2017, 2018, 2019, 2020

Elaboración: Los Autores

Al estudiar el empleo se observa de manera general que en los años 2016 al 2019 el empleo iba en crecimiento para los sectores de Servicios, Comercio, Manufactura y Agricultura; pero para el año 2020, en todos los sectores el empleo disminuye, siendo los más afectados los sectores de Servicios con 92,029; Comercio con 28,314; Manufactura con 32,458 y Construcción con 20,436 empleados despedidos, respectivamente. Esto como consecuencia del cierre de varias empresas como se vio en la tabla 1.5.

El sector manufacturero para el 2019 ya tuvo una disminución de plazas de empleo de aproximadamente 4.000 que se venían dando por la baja del precio del petróleo; sin embargo, el declive más grande fue en el año 2020, donde el impacto de la emergencia sanitaria provocó menor actividad productiva económica de la que

ya venía ocurriendo. Esto, pese a que desde el sector productivo se procuró la menor reducción posible de empleos a través de un acoplamiento a un método digital, evitando así tener más pérdidas de las que ya se habían generado. Para septiembre del 2020 se tuvo una gradual reactivación productiva y económica, pero, aunque se dio dicha reactivación, fuentes oficiales declararon que solo 32 de cada 100 personas lograron encontrar un empleo adecuado, mientras que otras 60 de cada 100 permanecen en situación de empleo inadecuado y otras 7 de cada 100 se encuentran en situación de desempleo.¹³

¹³Revista Primicias. Industrias, noviembre 2020 suplemento.

Capítulo 2

Marco Teórico

En el presente capítulo se exponen las definiciones y conceptos teóricos de las metodologías propuestas acerca de los modelos aditivos generalizados y las redes neuronales artificiales, con el propósito de conocer la estructura matemática que existe detrás de cada modelo y así realizar un adecuado proceso de modelización. Cabe mencionar que también se describen métodos de selección de variables que permiten obtener un mejor rendimiento de cada modelo; además, de estadísticos e índices usados para evaluar su desempeño y así complementar un adecuado proceso de modelización que permita encontrar resultados acorde a la realidad y a través de los cuales se logren cumplir los objetivos planteados.

Los países que poseen un sistema financiero estable, es gracias al desarrollo de varias actividades económicas, las mismas que son importantes para el dinamismo y crecimiento económico de una región. Sin embargo, al experimentar procesos de crisis o recesión económica, las principales actividades que se ven afectadas son las empresariales, lo que implica un impacto en los sectores sociales que impulsan la economía de un país; ya que según Estallo y Giner (2010) el principal motor que promueve la economía de un país son las empresas, donde se considera que la riqueza de las naciones radica en la producción de bienes y servicios que ellas generan [30].

No obstante, los comportamientos antes mencionados son algunos de los problemas que repercuten en el sistema productivo, lo que implica un decrecimiento en la producción, comercialización y como consecuencia una mayor exposición al riesgo de insolvencia financiera. Por tal motivo, los gerentes y la alta dirección en las empresas buscan herramientas de carácter cuantitativo que les permita detectar a tiempo las condiciones financieras y de esta manera anticiparse a eventos que

conlleven a una insolvencia empresarial y posteriormente a una posible bancarrota [38].

Los modelos de predicción de bancarrota permiten predecir y clasificar si una empresa entra o no en quiebra y de esta manera medir el riesgo de insolvencia financiera [2].

Existe una amplia variedad de metodologías que permiten predecir la bancarrota empresarial, entre los cuales se consideran a los modelos paramétricos clásicos como la regresión logística o el análisis discriminante múltiple. Con el desarrollo tecnológico se han creado nuevos métodos estadísticos como los algoritmos de Machine Learning. En particular, se hará énfasis a este último método ya que investigaciones realizadas por Valencia et al. (2019) demostraron que los métodos de Machine Learning, en particular el modelo de redes neuronales artificiales, poseen un mejor rendimiento de clasificación y predicción de bancarrota respecto a los modelos estadísticos paramétricos clásicos como la regresión logística. Por otro lado, Berg (2006) usó modelos aditivos generalizados como una alternativa semiparamétrica para la predicción de quiebras y demostró que su desempeño es significativamente mejor que las técnicas estadísticas tradicionales como el análisis discriminante, la regresión logística, los modelos lineales, entre otros [64] [8].

En el presente trabajo de investigación se busca complementar el uso de los modelos aditivos generalizados junto con las redes neuronales artificiales y así obtener un rendimiento robusto para la predecir y explicar el riesgo de quiebra empresarial en el sector industrial manufacturero ecuatoriano.

2.1. Definiciones y conceptos básicos

2.1.1. Conceptualización del riesgo

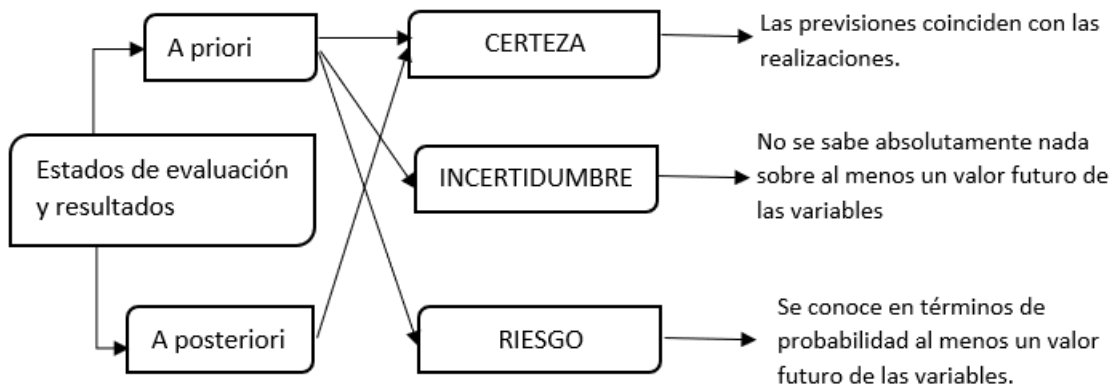
Antes de dar una definición concreta al riesgo es fundamental entender que toda decisión está sujeta a la incertidumbre ya que no se puede conocer o anticipar la certeza de los resultados que provengan de ella. Como lo afirma Durbán (2008), toda decisión se desarrolla en tres estados de evaluación[28]:

- **Certeza:** se conocen los resultados de cada una de las variables.
- **Riesgo:** se conocen cuáles podrían ser las probabilidades de ocurrencia de los resultados de cada una de las variables.

- **Incertidumbre:** se desconocen absolutamente los resultados de cada una de las variables.

Cada uno de estos estados se los puede apreciar de mejor manera en la siguiente Figura 2.1:

Figura 2.1: Estados de evaluación



Fuente: Durbán (2008)

Es por ello que toda decisión considera un riesgo bajo una incertidumbre cuyo resultado es desconocido. Por ejemplo, las inversiones para un activo se desarrollan bajo un estado de riesgo; no generar ganancia para una empresa ocasionando una posible bancarrota; los inversionistas que están dispuestos a asumir un nivel de riesgo con base en un capital invertido, entre otros. Por eso, el análisis del riesgo es una disciplina que ha surgido como consecuencia de eventos pasados que han mostrado que la inestabilidad económica y las crisis financieras se hubieran evitado con una adecuada administración y gestión de riesgos.

A continuación, se presentan algunos conceptos que permiten formalizar la definición del riesgo desde diferentes puntos de vista. En particular, desde la perspectiva financiera se puede tener las siguientes definiciones:

- La probabilidad de que una empresa no pueda enfrentar alguna situación inherente a su actividad (Celaya y López, 2004)[19].
- Gitman y Joehnk (2009) definen el riesgo como una condición, en la cual existe la posibilidad de desviarse del resultado esperado o deseado en donde el margen del error debe ser lo más pequeño posible, Por lo tanto, si se habla de rendimientos, el riesgo será el grado de variación de los rendimientos relacionados con un activo específico. Con esta definición, el riesgo indica el daño o

la posibilidad de beneficios, en donde daño o un beneficio implica pérdida o aumento del valor de los activos [35].

- Cruz (2016) define el riesgo como todo aquello que puede generar un evento no deseado y traer como consecuencia pérdidas y/o daños [22].
- A partir de los riesgos financieros para empresas de todos los sectores económicos, concretamente manufacturero, se definirá al riesgo como la posibilidad de ocurrencia de un hecho generador de pérdidas que afecte el valor económico de las empresas.

A partir de las definiciones previamente mencionadas se pueden generar los siguientes tipos de riesgo.

2.1.2. Tipos de riesgo

Al considerar una decisión de alto impacto que puede afectar procesos establecidos dentro de un entorno social permite dar inicio a una incertidumbre debido a que son desconocidos los resultados que puede proporcionar. Las decisiones con base en el riesgo son aplicadas constantemente en diferentes áreas de trabajo; por ejemplo: otorgamiento de créditos, inversiones sobre un portafolio de activos, inversiones en los mercados de valores, transacciones bancarias, entre otros. Por lo tanto, los riesgos según su naturaleza pueden ser clasificados en financieros y no financieros.

Riesgos Financieros

Dentro de los riesgos financieros podemos encontrar los de mercado, crédito y liquidez.

- **Riesgo de Mercado:** Es la pérdida que puede sufrir una inversionista por variaciones en los precios de las acciones que se registran en los mercados bursátiles.
- **Riesgo de Crédito:** Es la posibilidad del incumplimiento del prestatario o contraparte de las obligaciones financieras pactadas que conlleva a el no pago de las mismas.
- **Riesgo de Liquidez:** Es la probabilidad que una empresa no pueda obtener en el mercado flujos de caja esperados y no esperados por el crecimiento de sus

activos o es la contingencia de pérdida que se manifiesta por la incapacidad de la empresa para enfrentar la escasez de fondos y cumplir sus obligaciones y que determina la necesidad de conseguir recursos alternativos o vender activos en condiciones desfavorables.

En el contexto del desarrollo de este trabajo de investigación, se pondrá más énfasis en el riesgo de insolvencia financiera empresarial, el cual se lo definirá más ampliamente en la siguiente sección.

Riesgos No Financieros

Por otro lado, existen otros tipos de riesgos no financieros como el operativo, legal y reputacional.

- **Riesgo Operativo:** Es la posibilidad de que se produzcan pérdidas por fallos en los procesos, personas, tecnologías de información y eventos externos.
- **Riesgo Legal:** Es la posibilidad de que una empresa sufra pérdidas directas o indirectas sobre sus activos debido a que están expuestos a mayor situación de vulnerabilidad o que sus pasivos y contingentes se vean incrementados más allá de los niveles esperados debido a error, negligencia, impericia o de que deriven de la incorrecta aplicación de disposiciones legales o normativas.

2.1.3. Medición del Riesgo

Para la cuantificación del riesgo se utilizan múltiples técnicas que permiten asignar un valor numérico al cálculo del riesgo; en muchos de los casos el resultado es un valor de probabilidad. En esta sección se resumen algunas de las medidas más comunes de riesgo y se mencionan sus potenciales beneficios y posibles dificultades. A continuación, se muestran algunas medidas:

Probabilidad de ocurrencia

Este enfoque es simple y efectivo. Se considera el siguiente ejemplo: se tiene el 10 % de probabilidad de que un proyecto no alcance su punto de equilibrio o a cubrir los gastos generados (devolverá un valor presente neto negativo o cero, indicando pérdidas y no rentabilidad) en los próximos 5 años. Además, se supone que dos

proyectos similares tienen los mismos costos de implementación y de ganancias esperadas. Se lleva a cabo un análisis de riesgo para todos los casos en donde el primer proyecto podría revelar un 70 % de probabilidad de pérdidas en comparación con sólo el 5 % en el segundo proyecto; por lo que es evidente que el segundo proyecto es más rentable cuando se analizan los riesgos.

Desviación Estándar y Varianza

La desviación estándar es una medida que indica el promedio de la variación de cada caso con respecto a la media de los datos, en el cual una desviación estándar más grande implica una mayor amplitud de la distribución y, por ende, implica un mayor riesgo. Algunos analistas consideran al riesgo como las posibles pérdidas o caídas; por lo tanto, la desviación estándar y la varianza penalizan los movimientos al alza así como a la baja.

Valor en Riesgo (VaR)

Las empresas están expuestas a riesgos financieros, los cuales pueden ser riesgos operacionales, de liquidez, de crédito y de mercado. Estos riesgos están relacionados con las pérdidas potenciales asociadas al comportamiento del mercado. Visto de manera informal, el VaR es una medida de la variación potencial máxima del valor de un activo (o una cartera de activos), sobre un período pre-fijado, con una probabilidad dada; es decir, cuánto se puede perder, con una probabilidad p , sobre un horizonte de tiempo h fijo. Por lo tanto, el VaR estima el valor por el cual el riesgo podría disminuir debido a los movimientos generales del mercado financiero durante un período de tiempo determinado.

Ahora bien, para dar una definición formal al VaR, se considera el siguiente supuesto: Para un tiempo t se busca medir el riesgo de un activo o una cartera de activos para un horizonte de tiempo $h > 0$. Sea $\Delta P(h) = P_{t+h} - P_t$ la variación del valor de los activos desde el tiempo t hasta $t + h$, donde esta cantidad se mide en dólares y es una variable aleatoria en el tiempo t . Se considera como $F_h(\cdot)$ la función de distribución acumulada de $\Delta P(h)$. Entonces se define al VaR de un activo o cartera de activos en un horizonte h con probabilidad p , como la cantidad VaR que satisface lo siguiente:

$$p = Pr[\Delta P(h) \leq VaR] = F_h(VaR)$$

En la siguiente sección se presentará la manera de medir el riesgo de insolvencia. Esto se la hará a través del cálculo de la probabilidad de ocurrencia de que una empresa pueda o no caer en bancarrota (insolvencia financiera).

2.2. Riesgo de insolvencia

El riesgo de insolvencia es uno de los más críticos dentro del estudio de los riesgos financieros, ya que esta relacionado con la quiebra de empresas debido al fracaso empresarial. Dentro de este contexto, según Vargas (2015), existen cuatro términos que permiten determinar el riesgo de insolvencia en el sector empresarial [65].

Fracaso : Cuando la tasa de rendimiento¹ del capital invertido, considerando el riesgo, es más baja que la tasa de retorno de inversiones similares. También, se considera el escenario en el cual los ingresos son insuficientes para cubrir los costos y cuando el promedio del rendimiento sobre la inversión es continuamente inferior al costo de capital de la empresa.

Insolvencia : Sucede cuando una compañía no puede cumplir con su deuda a corto plazo debido a la ausencia de liquidez.

Incumplimiento : El incumplimiento se da cuando una compañía falla en el cumplimiento de pago de un contrato, donde el acreedor puede ejercer una acción legal para su cobro.

Bancarrota : La bancarrota es el resultado de las decisiones financieras equivocadas de una empresa ya que no puede cumplir con sus obligaciones económicas; los signos se dan antes de que la empresa pueda continuar con sus operaciones o que esta sea intervenida por un ente regulatorio.

Las causas del fracaso empresarial pueden surgir como consecuencia de un proceso de crisis o recesión debido a inconvenientes internos como externos, tales como la negligencia gerencial, fraude, desastres naturales, crisis económicas o financieras, entre otros. López et. al (2015) señalan que una empresa es financieramente sana cuando los recursos generados son suficientes, como mínimo, para mantener la capacidad de pago y, la empresa fracasa cuando no logra para sus accionistas un aumento satisfactorio de riqueza que les compense por el riesgo asumido y por la pérdida de la disponibilidad del capital invertido [46].

¹El rendimiento hace referencia al cambio de valor de un activo o portafolio en un tiempo determinado con respecto a su valor inicial, dado que el riesgo de mercado está ligado con la volatilidad de los precios.

Para ejemplificar, la crisis financiera de las empresas es el resultado que ocasiona el impacto de la aplicación de medidas económicas y sociales pensadas para el bien común. No obstante, la situación futura de una empresa se encuentra estrechamente relacionada con la información que esta maneje, como por ejemplo los informes financieros que debe presentar constantemente. Esta información se la puede resumir a través de los indicadores financieros, los cuales permiten conocer y estimar la situación futura de una empresa. Esta relación, justifica la creación de modelos de predicción de bancarrota, en donde estos proveen de información eficaz que proporciona información para generar alertas que identifiquen a tiempo un diagnóstico financiero que posibilita pronosticar si una empresa puede llegar a tener insolvencia financiera, lo que implica una posible bancarrota.

El fracaso empresarial ha sido definido de incontables maneras por individuos encargados del estudio del riesgo financiero, los cuales han aportado con varios modelos estadísticos que tienen como objetivo predecir la quiebra empresarial, entre los cuales podemos mencionar a los siguientes:

Los primeros estudios comienzan con Fitzpatrick (1932) que utilizó el análisis financiero reportado por las empresas para entender el fenómeno de la quiebra empresarial [32], en donde indicó que la falta de financiamiento por parte de los inversionistas, resultó en el deterioro en mayor grado sobre los indicadores financieros; lo que implica que, requieren mayores retornos esperados lo que incrementa el riesgo de quiebra. Por otro lado, le siguieron después los estudios de Beaver (1966) y Altman (1968), que definen el fracaso empresarial como la incapacidad de una empresa en pagar sus obligaciones financieras a medida que maduran. Apoyados en las técnicas del análisis univariado y multivariado respectivamente y con el uso de indicadores financieros, sus trabajos se caracterizaron por realizar modelos de predicción de quiebra empresarial [7] [3]. Las empresas consideradas por Beaver fueron aquellas que incumplieron el pago de bonos con cuentas bancarias sobregiradas; mientras que Altman consideró aquellas empresas que presentaron una petición de bancarrota bajo la *Ley Nacional de Bancarrota*, durante el período 1946-1965.

Deakin (1972) consideró en su estudio como empresas quebradas las que no cumplieron con sus obligaciones financieras y atravesaron por problemas financieros tales como: sobre endeudamiento, falta de liquidez, utilidad negativa, entre otros. Cada empresa se comparó con una no afectada en función de la calificación de la industria, el año de la información financiera entregada y el tamaño de los activos [25].

El modelo de Ohlson (1980) utilizó empresas declaradas legalmente en bancarrota. Según la definición de fracaso empresarial que el autor consideró fue puramente legalista y las empresas fallidas deben haberse declarado en bancarrota notificando el procedimiento de quiebra [56].

Se puede apreciar que las definiciones y las empresas declaradas en quiebra varían según los criterios de varios autores. En las investigaciones realizadas, el proceso que lleva a una empresa a la bancarrota o quiebra se da por el incumplimiento de sus obligaciones financieras y el deterioro de su patrimonio. De acuerdo a Calderón (2016), si los términos como insolvencia, quiebra y fragilidad se encuentran específicos en aspectos legales y de cumplimiento de deudas, pueden ser abarcados dentro de la terminología del fracaso empresarial [15].

Para entender el riesgo de insolvencia y evitar una bancarrota muchos gerentes optan por diferentes modelos matemáticos para realizar pronósticos y estimar la situación futura de su empresa, debido a la incertidumbre financiera que pueden afectar las actividades económicas de la misma. Por tal motivo, usan modelos de predicción de quiebra (bancarrota), los cuales permiten predecir y clasificar si una empresa entra o no en quiebra y así medir el riesgo de insolvencia financiera [2].

2.3. Modelos de predicción de quiebra

Los primeros estudios de predicción de bancarrota empresarial se basaron en estudios que se enfocaron, básicamente, en el análisis de los indicadores (ratios) financieros y en compararlos entre empresas fallidas y exitosas. A lo largo de la historia se han aplicado diferentes metodologías estadísticas que permitan predecir la quiebra empresarial. No obstante, de acuerdo a Caro et al. (2013) se diferencian dos etapas claves para su desarrollo [17].

2.3.1. Etapa Descriptiva del desarrollo de modelos de predicción de quiebra a través de indicadores

Dentro de la etapa descriptiva se encuentran algunos estudios relacionados al fracaso empresarial realizados por investigadores que analizaron su relación con los estados financieros. FitzPatrick (1932) analizó los indicadores financieros de veinte empresas industriales durante el período 1920 – 1929, con el fin de determinar el impacto en el momento del fracaso empresarial. Durante la investigación se ana-

lizó la variación de los indicadores hasta uno y dos años antes de la quiebra [32]. Posteriormente, los indicadores financieros fueron considerados como las variables que muestran un diagnóstico de la situación financiera real en el que se encuentran las empresas. Ya que según Merwin (1942) quien publicó su estudio centrado en pequeños fabricantes, concluyó que, al comparar empresas exitosas con empresas en quiebras, las empresas en quiebra mostraban indicadores con signos de debilidad, incluso, cuatro o cinco años antes del fracaso. Sus principales conclusiones fueron que las pequeñas empresas tienen un riesgo de quiebra cuando incurren en muy poca deuda a largo plazo y dependen para su respectivo financiamiento, principalmente, del capital social de fuentes internas y de préstamos a corto plazo; la bancarrota de las pequeñas empresas se refleja en la información histórica, en donde se puede identificar cuál sería el pronóstico a futuro. La principal debilidad por la cual las empresas fracasaron fue por lo descensos en las proporciones de capital de trabajo neto a activos totales, activos corrientes a pasivos corrientes y patrimonio neto a total deuda [48] .

Jackendoff (1962) realizó un estudio de los problemas críticos referentes a la clasificación, selección y uso de los indicadores, en donde el autor realizó el cálculo de los mismos a partir de información financiera recopilada por el Servicio de Impuestos de los Estados Unidos para los años fiscales 1947 – 1948 a 1957 – 1958; esto, con el propósito de adquirir antecedentes acerca de las características de varias industrias, empresas de diversos tamaños y empresas rentables versus no rentables. El autor encontró que entre los principales problemas que son evidentes entre los ratios publicados por la industria, está la falta de estandarización en el uso adecuado de los términos y, con frecuencia, el uso indiscriminado de la relación. Dentro de su estudio el autor seleccionó los siguientes indicadores básicos: liquidez, rentabilidad, apalancamiento y rentabilidad [39].

2.3.2. Etapa Predictiva del desarrollo de modelos de predicción de quiebra

En esta sección se muestran las diferentes metodologías estadísticas utilizadas por diferentes autores que permitieron estimar el riesgo de quiebra, en donde la "predicción" del fracaso empresarial dio paso a crear advertencias que indican cuál será la situación futura de una empresa y si la misma puede o no llegar a una bancarrota.

A continuación, se presentan las metodologías paramétricas clásicas que dieron origen a la predicción del fracaso empresarial, entre los cuales podemos mencionar el Análisis discriminante, los modelos de regresión lineal, los modelos de probabilidad lineal y los modelos Logit. Para las metodologías mencionadas su principal insumo matemático fueron los indicadores financieros.

Análisis Discriminante

El análisis discriminante aparece como una opción en situaciones en las que se necesita clasificar elementos con información incompleta; en general, con variables medibles que en la actualidad ayudan a determinar un conocimiento futuro (Peña, 2002). Esta es una técnica de clasificación supervisada dado que se conoce una muestra de elementos clasificados de manera correcta y que sirve de pauta para la clasificación del resto de observaciones de la población en estudio. Desde el punto de vista estadístico, se debe mencionar que el análisis parte del supuesto de normalidad multivariante de las variables consideradas y que es óptimo bajo esta hipótesis.

Bajo el contexto anterior descrito, Beaver (1966) propone que para estimar la predicción de la quiebra empresarial se utilice una metodología que permita evaluar datos contables para diferentes fines y no solo para determinar la insolvencia financiera. Para ello, el autor realiza un análisis multivariado de una serie de indicadores de las empresas que habían quebrado en *Moody's Industrial Manual* en el periodo 1954 - 1964, en donde seleccionó a treinta indicadores a partir de los siguientes criterios:

- Aparición frecuente en literatura.
- Indicadores que tuvieron buenos resultados en estudios anteriores.
- Los indicadores se encuentran definidos a través de un concepto llamado "flujo de caja".

El autor señala que la información recopilada en los estados financieros o flujos de caja son de mucha utilidad y la misma se puede definir en términos de capacidad predictiva; en otras palabras, facilita la predicción de la insolvencia empresarial; además, la distribución de los indicadores de las empresas fallidas presentan un evidente deterioro a medida que se acerca la bancarrota. Finalmente, el autor llegó a la conclusión que la relación entre el flujo de efectivo y la deuda total, tienen la capacidad de clasificar correctamente entre las empresas quebradas y las no quebradas [7].

Altman (1968), por otro lado, eligió el análisis múltiple (AMD) como metodología para su investigación y afirma que los indicadores de empresas quebradas son significativamente diferentes de las empresas que continúan con sus respectivas operaciones; esto, porque se evidenció que la información recabada durante los cinco últimos años previo a la bancarrota puede ser de suma importancia en la predicción del quiebre empresarial. En síntesis, Altman desarrolló los modelos *Z-score* en base a su técnica de investigación [3]. Es así que Beaver (1966) y Altman (1968) sentaron las bases para el desarrollo de la etapa predictiva de la bancarrota empresarial a través de modelos univariantes y multivariantes, respectivamente.

No obstante, Deakin (1972) realizó un estudio posterior, en el cual utilizó los mismos indicadores que Beaver pero los aplicó en una serie de modelos discriminantes multivariados, en donde el autor asevera que la quiebra de una empresa genera pérdidas a acreedores y accionistas; por tal motivo, fue necesario la creación de un modelo que pueda predecir posibles fallas empresariales y de esa manera crear alertas que sirvan para reducir tales pérdidas al proporcionar una amplia advertencia a las partes interesadas.

Se concluye que la capacidad predictiva del modelo de Altman disminuye a medida que se aumenta el número de años anteriores a la bancarrota. A continuación, se muestra el rendimiento de las técnicas descritas anteriormente [25].

Cuadro 2.1: Error de clasificación para predecir quiebras futuras

| Año antes de Quiebra | Beaver Flujo de caja/total de deuda | Función discriminante de Altman |
|----------------------|---|---------------------------------------|
| 1 | 13 | 5 |
| 2 | 21 | 28 |
| 3 | 23 | 52 |
| 4 | 24 | 71 |
| 5 | 22 | 64 |

Fuente: Deakin (1972)

En el cuadro 2.1 se puede ver que la capacidad predictiva de la técnica de Altman disminuye conforme aumenta el número de años previo a la quiebra empresarial; mientras que, el de Beaver concentra una distribución uniforme en cada año antes de la bancarrota.

Blum (1974) utilizó un análisis discriminante para probar que el modelo de predicción de bancarrota puede distinguir entre empresas en quebradas y no quebradas. El modelo distingue a las empresas fallidas de las no fallidas con una precisión

de aproximadamente 94 % cuando la bancarrota se dio dentro de una año a partir de la fecha de predicción, 80 % para la falla dentro de dos años y el 70 % para las fallas a tres, cuatro y cinco años. El modelo planteado por el autor fue elaborado con base en tres indicadores sobre el marco de flujo de efectivo: liquidez, rentabilidad y variabilidad. Para su estudio consideró a 115 empresas que quebraron durante el periodo 1954 y 1968 [12].

Edmister (1972) demostró como el uso de indicadores financieros permitieron predecir la bancarrota en las pequeñas empresas; para ello, utilizó como insumo los resultados empíricos realizados por Altamn (1968), Beaver (1966) y Blum (1974) y posteriormente el autor realizó un análisis multivariante para un consolidado de pequeñas empresas correspondientes a tres años previos a la quiebra empresarial durante el periodo 1954 – 1969. Para la metodología se empleó 19 indicadores financieros como variables independientes, de los cuales consideró a los más significativos aquellos que habían sido considerados en estudios previos. De hecho, solo 7 variables fueron seleccionadas las cuales discriminan 39 de 42 casos, es decir, el 93 % del total de empresas se logró predecir si podría entrar en un proceso de bancarrota [29].

Finalmente, Altman, Haldeman y Narayanan (1977) analizan un nuevo modelo de clasificación de quiebra que incorpora mejoras actuales en la utilización de técnicas estadísticas discriminantes con una muestra amplia de empresas en bancarrota que cubre el periodo 1969 – 1975. El nuevo modelo es denominado “Zeta”, para la clasificación de bancarrota, el cual posee un alto grado de precisión hasta 5 años antes del fracaso, con una clasificación exitosa de más del 90 % de la muestra un año antes de la quiebra y 70 % de precisión hasta 5 años. Para el modelo Z se consideraron las siguientes variables (indicadores financieros):

- Retorno sobre activos.
- Estabilidad de ganancias.
- Rentabilidad acumulada.
- Liquidez.
- Capitalización.
- Tamaño empresarial.

A partir del modelo Z de Altman, un modelo de predicción de bancarrota empresarial, hubo un amplio desarrollo en la investigación de nuevas metodologías

estadísticas que tenían como fin predecir la bancarrota empresarial, a través de las cuales se podría determinar cuando una empresa podía o no entrar en un proceso de quiebra. Estos modelos empezaron a utilizar la medida de probabilidad el cual es un resultado que permite determinar qué tan probable es que una empresa pueda llegar a una quiebra o no en un determinado periodo de tiempo. A continuación, se mostrarán algunos de los modelos más utilizados que permiten calcular la probabilidad de insolvencia financiera y una posible bancarrota empresarial. No obstante, primero se muestra la estructura estadística detrás de cada metodología, muchas de las cuales inician a partir de los modelos estadísticos paramétricos tradicionales.

2.4. Modelos Paramétricos

2.4.1. Modelos Lineales Generalizados

Modelos de Regresión lineal

El análisis de una regresión lineal es una colección de técnicas estadísticas para modelar e investigar la relación entre una variable respuesta de interés "y" y un conjunto de variables predictoras o explicativas (x_1, x_2, \dots, x_p) . Las aplicaciones de una regresión son numerosas y ocurren en casi todos los campos aplicados, incluyendo la ingeniería, las ciencias físicas y químicas, ciencias biológicas y de la vida, las ciencias sociales, la gestión y la economía. El modelo de regresión lineal tiene como objetivo modelar la dependencia de la media de la variable "y" en función de las variables (x_1, \dots, x_p) ; estas últimas también se llaman covariables.

Se desea describir dicha dependencia de forma que se pueda evaluar la contribución relativa de cada una de las variables predictoras en la explicación de "y", y de esa manera predecir esta variable para algún conjunto de variables explicativas. Para ello se supone $n > p$ observaciones sobre la variable respuesta, por ejemplo, y_1, y_2, \dots, y_n . Luego, para cada variable respuesta y_i se tiene una observación para cada variable predictora x_j para todo $j = 1, \dots, p$. Entonces, un modelo de regresión múltiple viene dado por:

$$y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \dots + \beta_p x_{ip} + \epsilon_i$$

$$y_i = \beta_0 + \sum_{j=1}^p \beta_j x_{ij} + \epsilon_i, \quad \forall i = 1, 2, \dots, n \quad (2.1)$$

Donde; ϵ_i es el vector de errores normales que se encuentran idénticamente distribuidos ($\epsilon_i \sim N(0; \sigma^2)$), β_j son parámetros o coeficientes del modelo desconocidos y x_{ij} es el valor de la j -ésima variable explicativa para la i -ésima observación e y_i el valor de la variable respuesta para la i -ésima observación, siendo $i = 1, 2, \dots, n$ y $j = 1, 2, \dots, p$. Como se puede ver se esta suponiendo que la dependencia de $E[y|x_1, \dots, x_p]$ es lineal. A menudo la ecuación 2.1 es llamada modelo de regresión lineal. En este tipo de modelos lo que se busca es la estimación para β_j con $j = 1, \dots, p$ por mínimos cuadrados de tal manera que se minimice la suma de los cuadrados de los errores.

Estos modelos son útiles para:

- Proporcionar una descripción simple de los datos.
- Describir la contribución de cada una de las variables explicativas mediante un coeficiente.
- Proporcionar un método simple para predecir nuevas observaciones.

Estos modelos no siempre son adecuados, ya que en muchos casos suponer que la variable respuesta se relaciona linealmente con las variables explicativas es demasiado restrictivo si la relación no es lineal.

Modelos de Regresión no lineal

De lo mencionado en la sección anterior, se tiene que un modelo de regresión lineal tiene la siguiente expresión:

$$y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \dots + \beta_p x_{ip} + \epsilon_i$$

Sin embargo, un modelo lineal no solo incluye relaciones de primer orden, sino que también considera de segundo orden, tercer orden, que son llamados modelos polinomiales y que son considerados de orden complejo. De hecho, un modelo de regresión lineal también se puede escribir de la siguiente manera:

$$y = \beta_0 + \beta_1 z_1 + \beta_2 z_2 + \dots + \beta_p z_p + \epsilon.$$

Donde z_j , con $j = 1, \dots, p$, representan cualquier función que se encuentra en base a las variables explicativas o regresoras ($x_{i1}, x_{i2}, \dots, x_{ip}$); por ejemplo: $z_1 = \exp x_{i1}$, $z_2 = \sqrt{x_{i2}}$ para todo $i = 1, \dots, n$. Por otra parte, los parámetros β_j del modelo lineal también son desconocidos pero, los modelos lineales se caracterizan por

ser lineales en el parámetro desconocido. Por tal motivo, la ecuación 2.1 se puede escribir de forma general como sigue:

$$y = f(\mathbf{x}, \beta) + \epsilon \quad (2.2)$$

Donde, $f(\mathbf{x}, \beta) = \beta_0 + \beta_1 x_{i1} + \dots + \beta_p x_{ip}$ y $\epsilon_i \sim N(0, \sigma^2)$. Por lo tanto, el valor esperado de la variable respuesta es:

$$\begin{aligned} E(y) &= E[f(\mathbf{x}, \beta) + \epsilon] \\ &= f(\mathbf{x}, \beta). \end{aligned}$$

Donde, $f(x, \beta)$ es una función estimada lineal de parámetros desconocidos.

Con lo anterior mencionado se tiene que no todos los modelos de regresión son lineales; para este caso, la variable predictora o respuesta puede estar definida en base a funciones que permiten evaluar qué función se puede considerar para definir la variable respuesta según los parámetros de entrada de la función.

Hay muchas situaciones en las cuales hacer uso de un modelo de regresión lineal no es adecuado. Por ejemplo, el análisis teórico de una amplia variedad de fenómenos físicos o químicos, en donde la relación que existe entre la variable respuesta y las predictoras se asume que la relación es lineal; este supuesto parte de la teoría que existe debajo de cada fenómeno. La verdadera relación entre la variable respuesta y predictora es una ecuación diferencial o la solución a una ecuación diferencial. Así que, el modelo debe ser de la forma no lineal.

Cabe mencionar que, cualquier modelo que no sea lineal en los parámetros desconocidos es un modelo de regresión no lineal. Para ejemplificar, se considera el siguiente modelo:

$$y = \beta_1 e^{\beta_2 x} + \epsilon, \quad (2.3)$$

es no lineal en los parámetros β_1 y β_2 . Generalizando la ecuación 2.3, la misma puede reescribirse como un modelo de regresión no lineal para un vector β_j de j parámetros

$$y = f(x, \beta_j) + \epsilon \quad \forall j = 1, 2, \dots, p.$$

Donde, β_j es el vector de parámetros y ϵ sigue una distribución normal con $E(\epsilon) = 0$ y $Var(\epsilon) = \sigma^2$. Se asume que los errores están normalmente distribuidos,

como en una regresión lineal.

$$\begin{aligned} E(y) &= E[f(x, \beta) + \epsilon] \\ &= f(x, \beta) \end{aligned}$$

A la expresión $f(x, \beta)$ se la llamará la función estimada para un modelo de regresión no lineal. Esto es muy similar al caso de la regresión lineal, a excepción de que ahora la función estimada es una función no lineal de parámetros desconocidos.

Para diferenciar entre un modelo de regresión lineal y no lineal con base en la función estimada, se analiza la dependencia de las derivadas parciales de f respecto a los parámetros; en la regresión lineal, estas derivadas no están en función de los parámetros desconocidos; mientras que, para una regresión no lineal, se tiene el caso contrario. Para entender la diferencia, se considera la siguiente función estimada que se encuentra expresada como en la ecuación 2.1

$$f(x, \beta) = \beta_0 + \sum_{j=1}^p \beta_j x_{ij} \quad \forall i = 1, \dots, n.$$

Las derivadas parciales de la función estimada son:

$$\frac{\partial f(x, \beta)}{\partial \beta_j} = x_{ij}, j = 1, \dots, p; i = 1, \dots, n$$

Se puede apreciar que las derivadas parciales no están en función de los parámetros β_j para todo $j = 1, 2, \dots, p$. Ahora, si se considera el modelo de regresión no lineal definido en la ecuación 2.3, las derivadas parciales de la función estimada con respecto a β_1 y β_2 son:

$$\begin{aligned} \frac{\partial f(x, \beta)}{\partial \beta_1} &= e^{\beta_2 x} \\ \frac{\partial f(x, \beta)}{\partial \beta_2} &= \beta_1 x e^{\beta_2 x} \end{aligned}$$

Con este resultado, se tiene que las derivadas parciales están en función de los parámetros β_1 y β_2 , por lo tanto se verifica que el modelo es no lineal.

Modelo lineal generalizado

El planteamiento de métodos estadísticos en los que se trata de explicar el comportamiento de una o varias variables respuesta, a través de un conjunto de variables explicativas o covariables, requiere la elección de un modelo que describa la

estructura de la relación entre variables.

Generalmente, el modelo más utilizado es del tipo lineal en el que se modeliza la variable respuesta, a través de una combinación lineal de las variables explicativas (ver ecuación 2.1).

Por tanto, el modelo lineal clásico consiste en expresar la esperanza condicionada de la variable respuesta como combinación lineal de las variables explicativas bajo hipótesis de normalidad y homocedasticidad. Esta modelización lineal clásica se puede extender a una familia más general, propuesta por Nelder y Wedderburn (1972) y ampliada por McCullagh y Nelder (1989), conocida como **Modelo Lineal Generalizado**.

Un modelo lineal generalizado (GLM) es un algoritmo de aprendizaje supervisado que generaliza el modelo lineal clásico, de manera que la variable dependiente “Y” está relacionada linealmente con las covariables mediante una determinada función de enlace. Además, el modelo permite que la variable Y se distribuya de manera diferente a una distribución normal (binomial, poisson, gamma, entre otras). Por lo general se utilizan cuando las variables bajo estudio incluyen datos categóricos.

La función de enlace es una componente estructural dentro de un GLM, la que permite relacionar el valor esperado de la variable respuesta con las variables explicativas y se representa mediante $g(f(x, \beta)) = z\beta^t$. Donde; $z = (1, x_1, x_2, \dots, x_p)$ y $\beta = (\beta_0, \beta_1, \dots, \beta_p)$ es el vector de parámetros o coeficientes de la regresión.

No obstante, los modelos indicados tienen una debilidad importante, pues no incorporan la posible no linealidad de las variables explicativas y la variable respuesta. Para sobrellevar esta debilidad, dentro de este proyecto de investigación se consideran los siguientes modelos:

2.5. Modelos No Paramétricos

Al considerar un comportamiento lineal entre la variable respuesta y las predictoras se genera un sesgo en la estimación y pronóstico de posibles eventos futuros, ya que este supuesto crea una mala representación de la relación real que existe entre variables, que probablemente no sea lineal y muy compleja de ser expresada analíticamente. Por otro lado, al trabajar con una relación no lineal se obtiene una mejor aproximación del comportamiento real que tienen los datos, lo que implica una predicción más acertada en los eventos futuros. Ahora bien, para obtener un alto rendimiento de predicción e interpretabilidad es necesario permitir una relación

no paramétrica entre variables predictoras y variable respuesta [64].

A diferencia de los modelos paramétricos, la estimación para sus coeficientes se la realiza a través de funciones, las cuales cumplen características que permiten encontrar el comportamiento real entre variables, aunque el problema se centra en la estimación de tales funciones.

Elegir un modelo que se adapte al comportamiento real de los datos puede imponer condiciones muy diferentes a la forma funcional real de f (ver ecuación 2.2), que es la desventaja principal de cualquier enfoque paramétrico, en donde el modelo resultante no se ajustará bien a los datos, lo que implica una estimación errónea.

El problema descrito se lo puede resolver aplicando modelos no paramétricos, los cuales permiten una mayor flexibilidad al no establecer hipótesis sobre la estructura subyacente de los datos, y esto se traduce en soluciones más flexibles.

Los modelo no paramétricos, según James, Witten, Hastie y Tibshirani (2014), a diferencia de los modelos paramétricos, antes de ajustar la función f no realiza ninguna suposición sobre la forma funcional de la misma. Este método ofrece más flexibilidad que los modelo paramétricos, al permitir que la forma funcional de f se pueda asumir dentro de un rango más amplio de funciones posibles. Este resultado es la principal ventaja, pues al hacer uso de un modelo paramétrico se tendría que realizar un análisis descriptivo previo para saber qué función describe mejor cada una de las variables explicativas. Por ejemplo, puede ser x^2 , $\log(x)$, etc.

Dentro de la amplia literatura de los modelos no paramétricos existen métodos que permiten relajar la condición de linealidad mediante el uso de funciones suaves, y son usadas en la búsqueda de f las cuales pueden ser funciones polinómicas, spline cúbicos, spline suavizado (*smooth spline*), regresión polinómica local (Loess), entre otros.

Previamente se describió que el uso de los modelos paramétricos puede implicar en una equivocada elección de la forma funcional de f , lo que conlleva a que el modelo resultante realice predicciones erróneas y sesgadas. Este problema se resuelve con el uso de modelos más flexibles que permitan formas funcionales más robustas para f . Sin embargo, ajustar modelos más flexibles requiere estimar más parámetros, en donde el riesgo que se asume con estos modelos es el fenómeno conocido como sobreajuste (James et. al 2014). Esto significa que el ajuste de la forma funcional de f aprende muy rápido; en otras palabras, la función queda muy ajustada a características específicas de los datos de entrenamiento que son usados para estimar los parámetros, a pesar de omitir información de suma importancia.

Este escenario no es el que se busca, ya que no se conseguirían predicciones precisas para nuevas observaciones que no se encontraron dentro del conjunto de datos de entrenamiento.

Además del problema del sobreajuste. Studenmund, (2016) menciona que las variables independientes correlacionadas pueden influir en la interpretación de los coeficientes, porque no es posible aislar completamente el efecto individual de cada variable independiente. De hecho, las variables independientes correlacionadas afectan la varianza de las estimaciones de los coeficientes, lo que lleva a pruebas de significancia distorsionadas.

En la práctica, se debe aplicar un procedimiento por pasos que permita incluir solo variables explicativas que agreguen un poder predictivo significativo al modelo. La inclusión de variables explicativas altamente correlacionadas puede causar errores en la interpretación de los efectos individuales de las variables explicativas sobre el modelo. En síntesis, al incluir variables altamente correlacionadas, tales interpretaciones deben evitarse, debido a los fenómenos de multicolinealidad. No obstante, si un modelo se construye únicamente con el propósito de predicción, entonces la multicolinealidad no es motivo de preocupación.

Los métodos no paramétricos al no reducir el problema de la estimación de f a un número pequeño de parámetros $\beta_0, \beta_1, \dots, \beta_n$ como sucede en los métodos paramétricos, es necesario un número mayor de observaciones que permitan obtener una estimación óptima de f . Estos métodos sí admiten variables correlacionadas, las cuales dependen de la necesidad del modelo que se esté realizando.

Los modelos paramétricos presentan valiosas ventajas por encima de los no paramétricos cuando su principal interés es la inferencia estadística, pues a través de estas técnicas se obtienen mejores interpretaciones sobre la relación entre la variable respuesta y las variables explicativas. Por ejemplo, el modelo lineal (ver ecuación 2.1) no presentará mayor dificultad al momento de entender la relación entre Y y las variables explicativas ($f(X_1, X_2, \dots, X_p)$).

Por otro lado, el hacer uso de modelos no paramétricos tiene como desventaja realizar estimaciones complejas en f , lo que implica un sobreajuste que conlleva un serio problema al momento de interpretar los resultados, y como consecuencia se vuelve difícil comprender la relación entre las variables predictoras o explicativas y la variable respuesta Y .

El **Modelo Aditivo Generalizado** ofrece un punto medio, ya que puede adaptarse a relaciones complejas y no lineales entre la variable respuesta Y y las variables

explicativas $f(X_1, X_2, \dots, X_p)$ siempre que mantenga la característica de aditividad, con el uso de elementos paramétricos como no paramétricos, lo que implica realizar buenas predicciones acorde a la realidad. Es más, permite hacer estadística inferencial, explicar y entender la estructura subyacente de los modelos y el porqué de las estimaciones que realiza. A continuación, se presenta la estructura interna de este modelo.

2.6. Modelo Aditivo Generalizado

2.6.1. Introducción

Los modelos aditivos generalizados fueron desarrollados por Hastie y Tibshirani (1986, 1990). Estos modelos son una extensión de los modelos lineales generalizados (GLM) al considerar funciones no lineales en cada una de las variables explicativas o predictoras.

El modelo lineal generalizado supone que la influencia de las variables explicativas sobre la variable respuesta es de forma lineal; es decir, una función de enlace g , considera el valor esperado de un modelo de regresión con n observaciones de una variable respuesta y_i con r variables predictoras

$$g(E[y_i]) = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \dots + \beta_p x_{ip} + \epsilon_i, \quad \forall i = 1, \dots, n$$

Sin embargo, puede suceder que la relación entre las variables predictoras y la variable respuesta tenga forma desconocida y, en ese caso, la estructura del modelo toma la siguiente expresión:

$$g(E[y_i]) = X_i^* \theta + f_1(x_{i1}) + f_2(x_{i2}) + \dots + f_p(x_{ip}) + \epsilon_i, \quad \forall i = 1, \dots, n \quad (2.4)$$

donde, y_i sigue una distribución de la familia exponencial, X_i^* es la i -ésima fila de la matriz del modelo que corresponde a las variables explicativas que definen las componentes paramétricas del modelo, θ es el vector de coeficientes de la regresión y, f_j para todo $j = 1, \dots, p$ son las funciones suavizadoras de variables.

En el modelo aditivo generalizado, a diferencia del modelo lineal generalizado, el supuesto de linealidad es relajado y proporciona suficiente flexibilidad permitiendo que no existan suposiciones sobre la forma funcional de las funciones f_j , pero

definiendo el modelo en términos de las funciones suaves, en lugar de las relaciones paramétricas. Sin embargo, esta flexibilidad produce dos problemas teóricos:

- Cómo presentar las funciones suaves.
- Qué tan suaves deben ser estas funciones.

Para el primer problema planteado, se tiene que las técnicas de estimación de las funciones f_j se basan en técnicas denominadas gráficos de dispersión suaves; por ejemplo: Hastie y Tibshirani (1990) describió una variedad de enfoques para estimar las funciones suaves. Algunos de los métodos de suavizamiento son los *spline* de suavizado y *spline* de regresión, que se diferencian de otros métodos entre los cuales se encuentra el suavizamiento por núcleo (Kernel Smooth) y regresión polinomial local (LOESS), entre otros [1].

La diferencia entre estos dos grupos radica en la forma en la que se realiza la estimación de los modelos aditivos generalizados cuando las funciones suaves están completamente parametrizadas. Los *splines* de suavizado (Smooth Splines), como los *splines* de regresión cúbica, se pueden expresar utilizando expansiones de base, y por lo tanto, PIRLS (mínimos cuadrados reponderados iterativamente penalizados) se puede aplicar directamente a los modelos aditivos generalizados con *splines* de regresión.

No obstante, el suavizamiento por núcleo (Kernel Smooth), como la regresión polinomial local, no se pueden expresar mediante la expansión de la base, por lo que el algoritmo PIRLS no se aplica directamente en este caso. Para realizar la estimación, se puede utilizar en su lugar el algoritmo *backfitting* [14]. La idea del algoritmo *backfitting* es ajustar los residuos parciales de forma iterativa en cada componente aditivo del modelo medio hasta la convergencia. El detalle de la descripción del algoritmo se encuentra fuera del alcance de esta investigación.

De los enfoques actualmente usados han destacado el uso de *splines* de regresión para la estimación de las funciones de suavizado. Wood (2017) afirma que la suavización de la función puede representarse mejor como *splines* de regresión [66].

Dentro de las siguientes secciones se indicará la representación de la función suave desconocida mediante *splines* de regresión penalizados, en particular los *splines* cúbicos de regresión y cómo seleccionar el parámetro de suavizado para f_j mediante REML o "Máxima probabilidad restringida", método recomendando por Ross(2019). Para poder implementar el modelo se hace uso de la librería '*mgcv*' del software R,

que permite estimar explícitamente los coeficientes para cada término suave mediante PIRLS (estimación de suavidad integrada).

2.6.2. Funciones Suaves - Caso Univariante

Con el propósito de representar y estimar las funciones suaves, se considera la ecuación (2.4) simplificada a una función suave como regresor, es decir, para n observaciones de una variable y_i con una sola variable explicativa x_{i1} , se busca un función f que cumpla:

$$y_i = f(x_{i1}) + \epsilon_i, \quad \forall i = 1, 2, \dots, n \quad (2.5)$$

donde,

- y_i : Variable dependiente.
- x_{i1} : Variable explicativa (regresor), por facilidad se supondrá que $x_{i1} \in [0, 1]$.
- f : función suave.
- ϵ_i : Variables aleatorias independientes idénticamente distribuidas (i.i.d) que siguen una distribución $N \sim (0, \sigma^2)$.

Con ayuda de las técnicas estadísticas anteriormente mencionados, se estima f , aunque es necesario representar f de tal manera que la ecuación (2.5) se convierta en modelo lineal. Esto se consigue eligiendo una base, y así definir el espacio de funciones base b_j de dimensión $m + 1$ en donde f se encuentre presente. Wood (2017) define esta base mediante una combinación lineal entre algunas funciones básicas conocidas y un vector de parámetros desconocidos α , con lo cual f se expresa como sigue:

$$f(x_{i1}) = \sum_{s=0}^m b_s(x_{i1})\alpha_s, \quad i = 1, 2, \dots, n \quad (2.6)$$

Sustituyendo (2.6) en (2.5) se obtiene el siguiente modelo lineal que puede ser estimado fácilmente:

$$y_i = \sum_{s=0}^m b_s(x_{i1})\alpha_s + \epsilon_i, \quad i = 1, 2, \dots, n$$

2.6.3. Regresión Polinómica

La idea intuitiva para la suavización es realizar un ajuste polinomial, es decir, utilizar la idea de representación de una función suave en términos de los polinomios

básicos. Es por ello que la construcción de f se puede realizar mediante la técnica conocida como regresión polinómica, la cual hace uso de una base polinómica. Esto permite ajustar la función f mediante una función polinómica de grado m , con esto se obtiene el siguiente modelo:

$$y_i = \alpha_0 + \alpha_1 x_{ij} + \alpha_2 x_{ij}^2 + \alpha_3 x_{ij}^3 + \dots + \alpha_m x_{ij}^m + \epsilon_i \quad \forall i = 1, \dots, n \quad (2.7)$$

Donde, $1, x_{ij}, x_{ij}^2, x_{ij}^3, \dots, x_{ij}^m$ para algún $j = 1, \dots, p$, son los polinomios básicos asociados a las funciones básicas y , según James et al. (2014), los parámetros α pueden ser estimados fácilmente con la técnica conocida como mínimos cuadrados [40].

Supongamos que f es un polinomio de orden m , de modo que una base para el espacio de funciones básicas de orden menor o igual a m , es la siguiente:

$$b_0(x_{ij}) = 1, \quad b_1(x_{ij}) = x, \quad b_2(x_{ij}) = x^2, \quad b_3(x_{ij}) = x^3 \quad \dots \quad b_m(x_{ij}) = x^m,$$

con $m \in \mathbb{N}$, de tal forma que si $b_s(x_{ij})$ es una función base, con $s \in [1, m]$. Así, la ecuación 2.6 se transforma a

$$f(x_{ij}) = \alpha_0 + \alpha_1 x_{ij} + \alpha_2 x_{ij}^2 + \dots + \alpha_m x_{ij}^m \quad \forall i = 1, \dots, n;$$

para algún j . Sin embargo, la regresión polinómica genera una curva extremadamente no lineal, demasiado flexible adoptando comportamientos atípicos. Por tal motivo, la alternativa más adecuada es la estimación por *splines*.

2.6.4. Splines Cúbicos de Regresión

Los *splines* de regresión a diferencia de la regresión polinómica permiten estimar la función f dividiendo la función original en secciones y ajustando cada sección con un polinomio individual de menor grado. Cada punto que une las distintas secciones se conoce como "nodo". Esto nos lleva a la siguiente clase de bases de funciones, donde una función f con la siguiente representación matemática

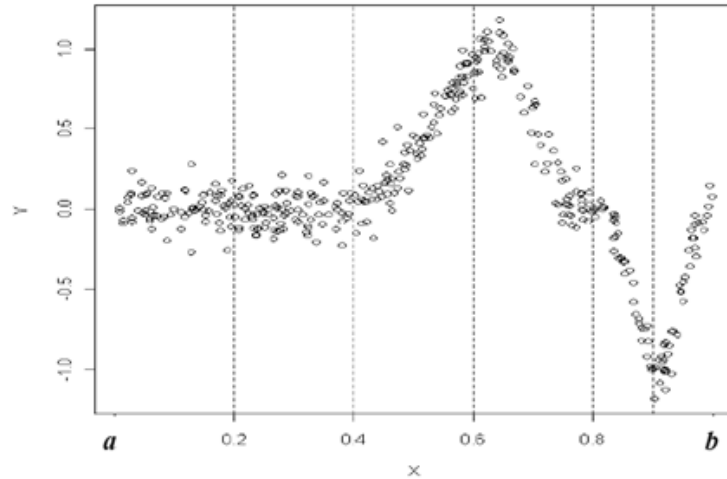
$$f: [a, b] \longrightarrow \mathbb{R},$$

es un *spline* polinómico de grado m , si cumple las siguientes condiciones:

1. $f(x)$ es $(m - 1)$ veces continuamente diferenciable.
2. $f(x)$ es un polinomio de grado m para todo $x \in [k_s, k_{s+1})$ con $s = 1, \dots, m - 1$.

Se puede considerar una base de $(m - 1)$ polinomios a trozos de grado m , ajustados por separado en cada uno de los intervalos definidos por una sucesión de nodos de corte (*Knots*).

Figura 2.2: Intervalos de corte para una base de $(m-1)$ polinomios.



Fuente: Hastie, T. J. (1990)

Donde la partición para el intervalo $[a, b]$ de la figura 2.2 es:

$$\mathbf{a} = k_1 < k_2 < \dots < k_{m-1} < k_m = \mathbf{b}.$$

Por tanto, cada *spline* polinómico puede ser representado por una base

$$d = (m + m - 1)$$

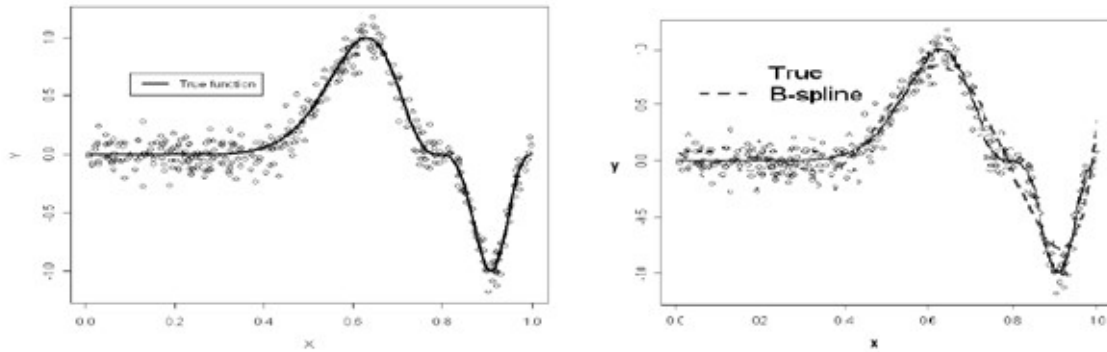
funciones. Así, la ecuación (2.6) se puede escribir también como sigue:

$$f(x_{i1}) = \sum_{s=1}^d b_s(x_{i1})\alpha_s$$

Una de las bases más simples es aquella formada por polinomios de orden que se considera para f ; además, es una de las más utilizadas en varias situaciones; sin embargo, cuando f está definida en el dominio $[0, 1]$ se tienen algunos problemas con esta base. En este punto, aparecen las bases formadas por *splines* debido a las propiedades teóricas que poseen. Es así que una posible base local es la formada por *Basic - Splines* o simplemente ***B-splines***. Estas bases de grado m se obtienen fusionando $(m + 1)$ polinomios de grado m en los $(m - 1)$ nodos (*Knots*) del *spline*. Para visualizar de manera objetiva lo descrito se considera el siguiente ejemplo "simulado": Se

considera la función $f(x_i) = \sin(2\pi x_i^3) + \epsilon_i$ donde $x_i \sim U[0, 1]$ y $\epsilon_i \sim N(0; 0, 1)$, con $i = 1, \dots, 400$.

Figura 2.3: Función f suavizada con B -Spline



(a) función *Spline*

(b) función suavizada con B -Spline

Fuente: Hastie, T. J. (1990)

La figura 2.3 (a) representa la función *spline* y la figura 2.3 (b) indica la función suavizada a través de la base B -Spline. Los círculos en ambas gráficas representan los nodos del *Spline*.

Ahora, si se toma $d = 3$ para la ecuación 2.7, se obtiene el siguiente modelo de regresión cúbico:

$$y_i = \alpha_0 + \alpha_1 x_i + \alpha_2 x_i^2 + \alpha_3 x_i^3 + \epsilon_i \quad (2.8)$$

Donde, los coeficientes $(\alpha_0, \alpha_1, \alpha_2, \alpha_3)$ son diferentes en las distintas secciones y el ajuste de y_i se realiza mediante un polinomio cúbico definido a trozos.

Por ejemplo, un polinomio cúbico definido a trozos con un solo nodo en un punto x^* se define de la siguiente forma:

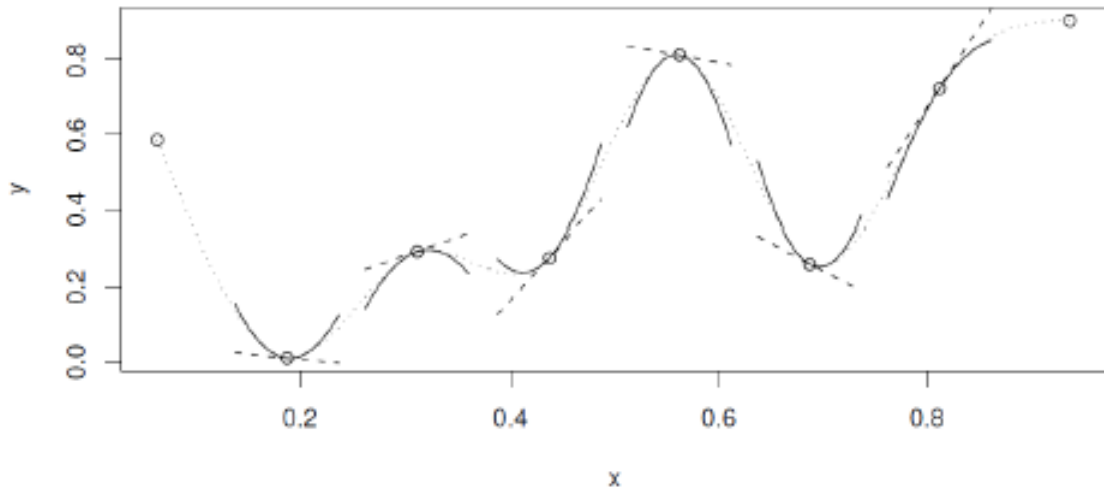
$$y_i = \begin{cases} \alpha_{01} + \alpha_{11}x_i + \alpha_{21}x_i^2 + \alpha_{31}x_i^3 + \epsilon_i & \text{si } x_i < x^* \\ \alpha_{02} + \alpha_{12}x_i + \alpha_{22}x_i^2 + \alpha_{32}x_i^3 + \epsilon_i & \text{si } x_i \geq x^* \end{cases} \quad (2.9)$$

De modo que (2.9) realiza el ajuste de los datos mediante dos funciones polinómicas diferentes, una para aquellas observaciones en donde $x_i < x^*$ y otra para el resto tales que $x_i \geq x^*$.

El primer polinomio tiene coeficientes $\alpha_{01}, \alpha_{11}, \alpha_{21}, \alpha_{31}$ y el segundo $\alpha_{02}, \alpha_{12}, \alpha_{22}, \alpha_{32}$. Para ambos casos, los coeficientes se pueden estimar mediante mínimos cuadrados.

De forma gráfica un spline cúbico según Wood (2017), es una curva compuesta por secciones de polinomios cúbicos unidos de tal forma que la curva sea continua hasta la segunda derivada en cada nodo[66].

Figura 2.4: Spline Cúbico.



Fuente: Wood (2017)

En la figura 2.4 se puede apreciar que un *spline* cúbico (curva punteada) queda representado por una curva construida por distintos polinomios cúbicos unidos de manera que la curva sea continua hasta la segunda derivada. Los puntos de unión (o) son los nodos del *spline*, los cuales dividen a la curva en siete secciones. Cada polinomio tiene distintos coeficientes, pero en el valor de los nodos se igualarán sus valores y las primeras dos derivadas con respecto a las zonas vecinas.

Si el rango de x (Figura 2.4) es dividido en suficientes regiones, se genera un ajuste extremadamente suave. En general, cuantos más nodos se usen, más flexible será el ajuste realizado sobre f . Si se ajustan k nodos a lo largo del rango x , se ajustan $k + 1$ polinomios cúbicos.

En conclusión, un *spline* de grado d es aquel definido mediante un polinomio de grado d definido a trozos cuyas primeras $d - 1$ derivadas son continuas en cada nodo. En la práctica, los *splines* más usados son los de grado 3, conocidos como ***splines* cúbicos**.

Bases de Splines Cúbicos

Para determinar un base para *splines* cúbicos, Wood (2017) propone una de las bases más usadas, la cual está basada en $q - 2$ nodos x_i^* con $i = 1, \dots, q - 2$ y está generada por las siguientes funciones básicas:

$$b_1(x) = 1, \quad b_2(x) = x, \quad , b_{q+2} = R(x, x_i^*),$$

tal que

$$R(x, z) = \frac{\left[\left(z - \frac{1}{2} \right)^2 \frac{1}{12} \right] \left[\left(x - \frac{1}{2} \right)^2 \frac{1}{12} \right]}{4} - \frac{\left[\left(|x - z| - \frac{1}{2} \right)^4 - \frac{1}{2} \left(|x - z| - \frac{1}{2} \right)^2 + \frac{7}{240} \right]}{24}$$

Con el uso de esta base para ajustar f se consigue que (2.5) se convierta en un modelo lineal de la forma $y = X\beta + \epsilon$, en donde la i -ésima fila de la matriz X es:

$$X_i = \left[1, x_i, R(x_i, x_1^*), R(x_i, x_2^*), \dots, R(x_i, x_{q-2}^*) \right]$$

y por tanto los parámetros desconocidos pueden ser estimados por mínimos cuadrados.

Otra de las bases de funciones muy usada para representar *splines* cúbicos, propuesta por James et. al (2014) está basada en k nodos x_i^k con $i = 1, 2, \dots, k$ y está generada por:

$$b_1(x) = x, \quad b_2(x) = x^2, \quad , b_3(x) = x^3$$

$$b_4(x) = h(x, x_1^*), \quad , b_5(x) = h(x, x_2^*), \dots, b_{k+3}(x) = h(x, x_k^*),$$

donde,

$$h(x, x_i^*) = (x - x_i^*)_+^3 = \begin{cases} (x - x_i^*)^3 & \text{si } x > x_i^* \\ 0 & \text{si } x \leq x_i^* \end{cases}$$

Entonces, (2.5) quedaría como

$$y_i = \alpha_0 + \alpha_1 x_i + \alpha_2 x_i^2 + \alpha_3 x_i^3 + \alpha_4 h(x, x_1^*) + \alpha_5 h(x, x_2^*) + \dots + \alpha_{k+2} h(x, x_k^*) + \epsilon_i$$

Por tanto, el vector de los parámetros desconocidos α puede ser estimado por mínimos cuadrados.

2.6.5. Grado de suavizado: Regresión por Splines penalizadas

Al encontrar un *spline* que permita suavizar la función estimada para una variable explicativa, esta debe ser lo más cercanamente posible al comportamiento real entre la variable respuesta y la explicativa. Esto se lo puede determinar a través del orden de la base. No obstante, Wood (2017) afirma que el orden de la base no es un criterio suficiente para controlar la suavidad del *spline* sobre la función. Por lo tanto, una alternativa para controlar la suavidad de la función (segundo problema teórico) es fijar la dimensión de la base en un tamaño superior, pero para controlar la suavidad del modelo se debe añadir una penalización por posibles "ondulaciones" que permitan controlar la suavidad de la función. Por ejemplo, en lugar de ajustar el modelo (ver ecuación 2.10) por mínimos cuadrados

$$\|y - X\alpha\|^2 \quad (2.10)$$

Se lo puede ajustar por mínimos cuadrados de:

$$\|y - X\alpha\|^2 + \lambda \int_0^1 [f''(x)]^2 dx,$$

donde la integral al cuadrado de la segunda derivada penaliza al modelo por ondulaciones. La compensación entre el modelo ajustado y el modelo suavizado, está controlada por el parámetro *suavizante* λ . Si $\lambda \rightarrow \infty$ conduce a una estimación de f que daría como resultado una recta; en caso contrario, si $\lambda = 0$, entonces se tendría la estimación de la regresión por *splines* sin penalización.

Dado que f es lineal con respecto a α_i , la penalización se puede escribir como:

$$\int_0^1 [f''(x)]^2 dx = \alpha' S \alpha$$

Donde, S es una matriz de coeficientes conocidos. Esto hace que la forma de la base de *splines* sea complicada, ya que se tiene:

$$S_{i+2, j+2} = R(x_i^*, x_j^*), \quad i, j = 1, \dots, q-2$$

Mientras que, las dos primeras columnas de S son 0 [36].

Por lo tanto, el problema de ajuste por regresión de *splines* penalizados consiste en minimizar

$$\|y - X\alpha\|^2 + \lambda \alpha' S \alpha$$

El problema para la estimación del grado de suavidad del modelo se resume en la estimación del parámetro de suavizado λ . No obstante, es importante considerar la estimación de α dado λ . El estimador de mínimos cuadrados penalizado de α dado λ es:

$$\hat{\alpha} = (X'X + \lambda S)^{-1} X'y$$

2.6.6. Estimación del Parámetro Suavizante λ

Idealmente, si se escoge bien el valor de λ esto hace que \hat{f} sea lo más cercana posible a la f real. Un criterio adecuado puede ser la minimización del siguiente estimador:

$$M = \frac{1}{n} \sum_{i=1}^n (\hat{f}_i - f_i)^2$$

Sin embargo, f no es conocida, y no se puede obtener M directamente, pero se puede obtener una estimación de $E(M) + \sigma^2$, que es el error cuadrático de la nueva variable. Entonces, se define lo siguiente.

Sea $\hat{f}^{[-i]}$ el modelo que se ajusta todos los datos excepto a y_i y se define el valor de validación cruzada ordinario como:

$$\vartheta_0 = \frac{1}{n} \sum_{i=1}^n \left(\hat{f}^{[-i]} - y_i \right)^2$$

Este valor resulta de omitir cada dato por vez, ajustando el modelo a los demás datos y calculando el promedio de la diferencia al cuadrado entre el valor perdido y su predicción. Reemplazando $y_i = f_i + \epsilon_i$:

$$\vartheta_0 = \frac{1}{n} \sum_{i=1}^n \left(\hat{f}^{[-i]} - f_i - \epsilon_i \right)^2 \quad (2.11)$$

$$= \frac{1}{n} \sum_{i=1}^n \left(\hat{f}^{[-i]} - f_i \right)^2 - \left(\hat{f}^{[-i]} - f_i \right) \epsilon_i + \epsilon_i^2 \quad (2.12)$$

donde $E(\epsilon_i) = 0$, ϵ_i y $\hat{f}^{[-i]}$ son independientes.

Por otro lado, al tomar la esperanza a cada lado de la ecuación (2.12), se obtiene el siguiente resultado:

$$E(\vartheta_0) = \frac{1}{n} E \left(\sum_{i=1}^n \left(\hat{f}^{[-i]} - f_i - \epsilon_i \right)^2 \right) + \sigma^2 \quad (2.13)$$

Ahora, en 2.13, $\hat{f}^{[-i]} \approx \hat{f}$ para n suficientemente grande, entonces $E(\vartheta_0) \approx E(M) + \sigma^2$ si se considera el mismo n . Por lo tanto, elegir λ que minimice ϑ_0 es un enfoque adecuado si la idea es minimizar M .

2.6.7. Modelos Aditivos

Para simplificar la exposición de estos modelos, sin pérdida de generalidad, se consideran dos covariables y una variable dependiente; por tanto, una estructura simple de un modelo aditivo es:

$$y_i = f_1(x_i) + f_2(z_i) + \epsilon_i \quad (2.14)$$

Donde;

- $x_i, z_i \in [0, 1]$ son covariables.
- f_i son las funciones suaves.
- $\epsilon_i \sim N(0, \sigma^2)$ son independientes e idénticamente distribuidas.

Cabe notar dos puntos importantes en el modelo 2.14. En primer lugar, el supuesto de aditividad es muy fuerte, ya que $f_1(x) + f_2(x)$ es un caso restrictivo especial de la función general suave de dos variables $f(x, z)$. En segundo lugar, el hecho de que el modelo contiene más de una función, introduce un problema de identificabilidad ya que las funciones solo se pueden estimar en una constante aditiva; es decir, se puede sumar un valor a f_1 y restarlo de f_2 sin afectar las predicciones del modelo. Por tanto, es importante tener en cuenta esto al plantear un modelo de este tipo.

El problema de identificabilidad se lo puede tratar utilizando la representación por *splines*, estimando los coeficientes por mínimos cuadrados penalizados y determinando el parámetro de suavidad por validación cruzada, de manera similar a la presentada para el caso univariante.

Representación de la Regresión de *Splines* Penalizada de un Modelo Aditivo

Las funciones suaves de los modelos aditivos pueden ser representadas con el uso de una base de regresión por *splines* penalizada, tal como sigue:

$$f_1(x) = \delta_1 + x\delta_2 + \sum_{i=1}^{r_1-2} R(x, x_i^2)\delta_{i+2}$$

$$f_2(x) = \gamma_1 + x\gamma_2 + \sum_{i=1}^{r_2-2} R(z, z_i^*)\gamma_{i+2}$$

Donde;

- δ_i y γ_i son parámetros desconocidos de f_1 y f_2 , respectivamente.
- r_1 y r_2 representan el número de parámetros desconocidos de f_1 y f_2 , respectivamente.
- x_i^* y z_i^* son los nodos para f_1 y f_2 .

Entonces, el problema de la identificabilidad significa que se puede considerar a δ_i como γ_i , lo que produce confusión al momento de trabajar con los parámetros. Una manera sencilla de eliminar esta confusión es asignando el valor de cero a cualquiera de los dos parámetros ($\gamma_1 = 0$). Es así que el modelo puede también ser escrito de manera lineal como: $y = X\alpha + \epsilon$, donde la i -ésima de la matriz de diseño se escribe de la siguiente manera:

$$X_i = [1, x_i, R(x_i, x_1^*), \dots, R(x_i, x_{r_1-2}^*), z_i, R(z_i, z_1^*), \dots, R(x_i, x_{r_2-2}^*)]$$

$$\alpha = [\delta_1, \delta_{r_1}, \gamma_2, \dots, \gamma_{r_2}]'$$

Luego, las ondulaciones de las funciones debido a las suavizaciones aplicadas a través de una base, se pueden medir de manera similar al caso univariante:

$$\int_0^1 [f_1''(x)]^2 dx = \alpha' S_1 \alpha \quad \text{y} \quad \int_0^1 [f_2''(x)]^2 dx = \alpha' S_2 \alpha,$$

donde, S_1 y S_2 son matrices con valor de cero en cada entrada, a excepción de:

$$S_{1i+2, j+2} = R(x_i^*, x_j^*) \quad \text{para} \quad i, j = 1, \dots, r_1 - 2$$

$$S_{2i+r_1+1, j+r_1+1} = R(z_i^*, z_j^*) \quad \text{para} \quad i, j = 1, \dots, r_2 - 2$$

Ajuste de mínimos cuadrados Penalizados para Modelos Aditivos

Los parámetros α del modelo se obtiene minimizando la siguiente ecuación:

$$\|y - X\alpha\| + \lambda_1 \alpha' S_1 \alpha + \lambda_2 \alpha' S_2 \alpha,$$

donde, λ_1 y λ_2 son los parámetros desconocidos suavizados que se pueden estimar por la validación cruzada ordinaria aplicada de manera independiente a cada

ecuación.

2.6.8. Modelos Aditivos Generalizados

Como ya se había mencionado, los modelos aditivos generalizados vienen a ser una generalización de los modelos lineales, ampliando los conceptos vistos hacia los modelos aditivos; en otras palabras, el predictor lineal predice una función monótona suave conocida de los valores esperados de la variable dependiente y , además, la misma puede seguir cualquier distribución de la familia exponencial, permitiendo así aprovechar las técnicas de estimación por máxima verosimilitud.

Para resumir, todo lo desarrollado para los modelos aditivos se extiende al caso general. Por ejemplo, si se considera un modelo *logit* para una variable respuesta binaria, el modelo se especifica de la siguiente manera:

$$\ln \left(\frac{p(X)}{1 - p(X)} \right) = \alpha_0 + \alpha_1 x_{i1} + \dots + \alpha_n x_{in} + \epsilon_i \quad \text{con} \quad p(X) = P(y_i = 1 | x_i) \quad (2.15)$$

La ecuación (2.15) puede ser generalizada de la siguiente manera:

$$\ln \left(\frac{p(X)}{1 - p(X)} \right) = \alpha_0 + \sum_{j=1}^r f_j(x_{ij}) + \epsilon_i \quad \forall i = 1, 2, \dots, n, \quad (2.16)$$

donde, f_j , en la ecuación (2.16), es una función suave desconocida. La forma no paramétrica de estas funciones permite que el modelo se vuelva más flexible. No obstante, la aditividad se conserva y esta característica permite interpretar el modelo de la misma manera que un modelo de regresión logística.

2.7. Redes Neuronales

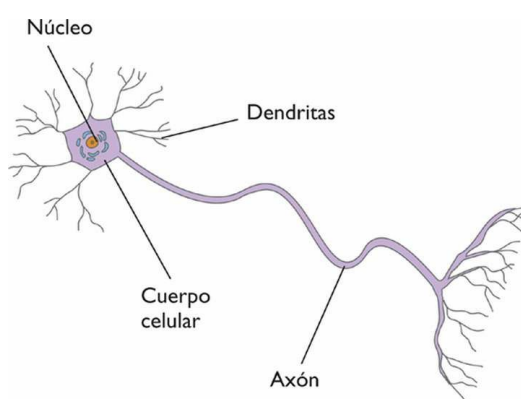
En la siguiente sección se muestra el origen de las redes neuronales partiendo del modelo biológico y mostrando las similitudes de modelo artificial con este. Asimismo, se expone las matemáticas del proceso de las redes neuronales, mostrando su estructura básica e indicando las operaciones que se realizan en cada parte de la red (sumatorios, funciones de transferencia, entre otros).

2.7.1. Redes neuronales como un modelo biológico

Las redes neuronales son un sistema que reciben entradas numéricas y generan salidas de uno o más valores numéricos. También se la puede conocer como una herramienta que imita la forma en que el cerebro humano funciona.

Partiendo de un concepto biológico, se conoce que el cerebro humano está compuesto por varias neuronas, los cuales son elementos básicos para el funcionamiento del cerebro. Estas, a su vez, están formadas por un cuerpo central, llamado núcleo, y un mecanismo de conexión con otras neuronas (axón y dendritas). Ver Fig. 2.5.

Figura 2.5: Partes de una neurona biológica



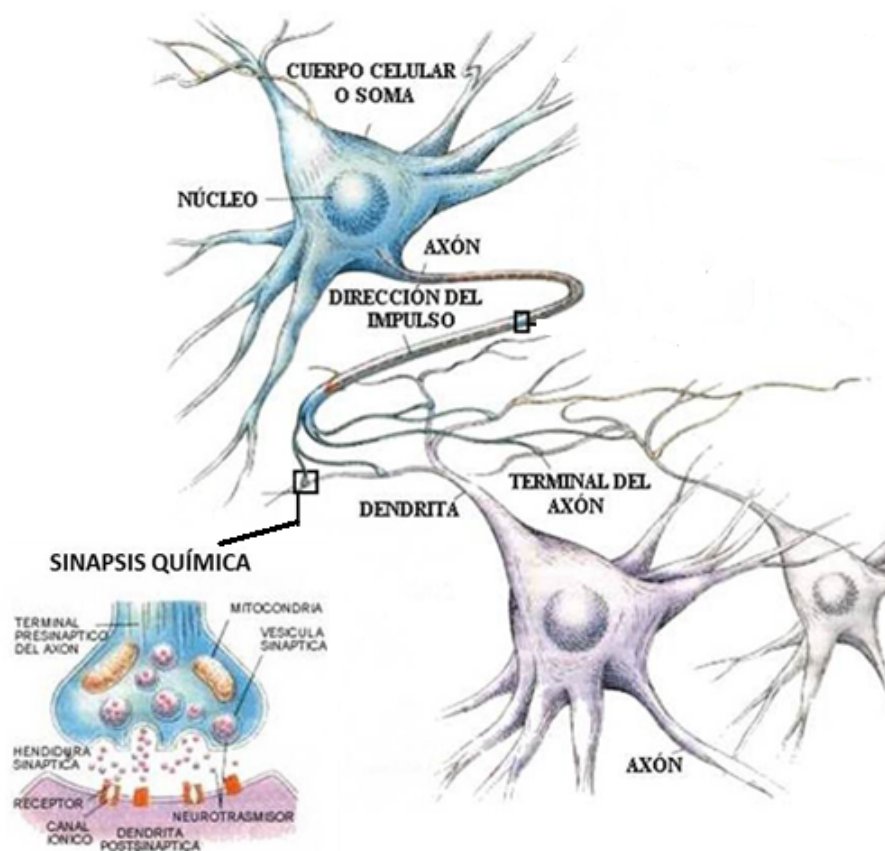
Los estímulos recibidos en el cerebro son transmitidos entre las neuronas mediante las conexiones sinápticas. Cuando una neurona es estimulada libera una pequeña cantidad de un componente químico (neurotransmisor); este viaja a través del axón hasta llegar a las dendritas de otras neuronas en las cuales el proceso se repite. Este proceso sirve para incrementar o disminuir la relación entre las neuronas involucradas en él. Así, ante un determinado estímulo ciertas neuronas se activan y otras se inhiben. Este proceso se lo puede observar en la figura 2.6.

El proceso de crear nuevas conexiones entre neuronas, es conocido como *aprendizaje*, y cuando este proceso se completa se tiene conexiones fijas que se conocen como memoria. El aprendizaje se consigue a través de la fluctuación de la efectividad de la sinapsis, cambiando la presión que unas neuronas realizan sobre otras. Es así que el conocimiento que se adquiere está en la relación entre neuronas logradas en el proceso de aprendizaje.[49]

Con lo dicho en el párrafo anterior, se puede decir que el cerebro es entrenado a través de la repetición de estímulos, por lo que mediante un proceso de aprendizaje se logra obtener conexiones adecuadas entre las neuronas.

Por tanto, una red neuronal que, como se dijo anteriormente, funciona de manera

Figura 2.6: Estructura de una neurona biológica



muy similar al cerebro humano, puede enviar o recibir una señal hacia o desde otras neuronas y representar un modelo simplificado del cerebro como tal.

2.7.2. Redes neuronales artificiales

Las redes neuronales artificiales son modelos computacionales que aparecieron a mediados del siglo XX tras experimentos realizados con el fin de obtener formalizaciones matemáticas sobre la estructura del cerebro. Esto impulsó el nacimiento de la cibernética y más tarde nació el término Inteligencia Artificial. Con ello, se replica la estructura del sistema nervioso conforme al funcionamiento del cerebro humano, basándose en el aprendizaje a partir de la experiencia, con la posterior extracción de conocimiento a partir de la misma, con el objetivo de imitar al cerebro humano gracias a procesadores artificiales, que además, puedan subsanar algunos errores humanos.[33]

Por lo tanto, son modelos artificiales que intentan reproducir una de las características naturales de los seres humanos, la cual es memorizar y enlazar hechos.

También esta definida como una técnica no paramétrica, capaz de modelar patrones complejos que no pueden ser modelados por los métodos estadísticos de uso común, cuya característica le brinda ventajas como el poder resolver problemas altamente no lineales o el no hacer suposiciones sobre la distribución de los datos.

Definiciones

1. **Capas.-** Son un conjunto de neuronas cuyas entradas provienen de una capa anterior o de los datos de entrada en el caso de una primera capa.
2. **Nodos.-** Representan a las neuronas artificiales, las cuales reciben y envían mensajes dentro de la red a través de una conexión artificial (en el cerebro es conocido como el axón).
3. **Pesos.-** Son coeficientes que indican la proporción de importancia que tiene la información de entrada que recibe cada neurona dentro de la red.
Matemáticamente, se lo expresa como $w_{i,j}$, donde i representa la neurona de la capa anterior y j representa la neurona de la siguiente capa.

Elementos de la red neuronal artificial

Debido a que la red neuronal es un sistema que recibe entradas numéricas y genera salidas del mismo tipo, dependen de elementos esenciales los cuales son los nodos, cuyos elementos están distribuidos en varias capas que trabajan conjuntamente para transformar los datos de entrada en entidades de salida.

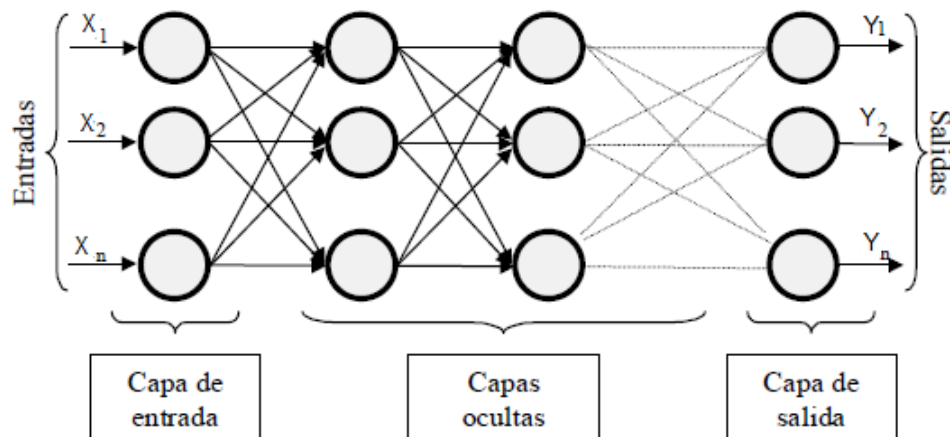
A partir de la situación en que se encuentren los nodos (neuronas) se puede distinguir tres tipos de capas.

- *Capa de entrada:* Es conocida como la capa cero que recibe directamente la información de las fuentes externas de la red, es decir, está conformada por las neuronas que introducen la información en la red y cuyo objetivo es redistribuir los valores de entrada hacia la siguiente capa.
- *Capa Oculta:* Conocida como capa intermedia, constituida por nodos (neuronas) donde sus entradas vienen de capas anteriores y sus salidas serán las entradas de la siguientes capas. Estas capas se caracterizan por no tener contacto con el exterior y pueden estar formadas por varias capas.

- *Capa de Salida* Esta conformado por neuronas cuyo objetivo es transferir la información de la red hacia el exterior; por ello, representan la salida de toda la red.

En la figura 2.7 se puede observar los elementos básicos de la red neuronal, nombrados anteriormente, donde la gráfica muestra cómo las neuronas se interconectan y forman las capas. La capa oculta puede constituirse por más de una capa.

Figura 2.7: Esquema de una red neuronal

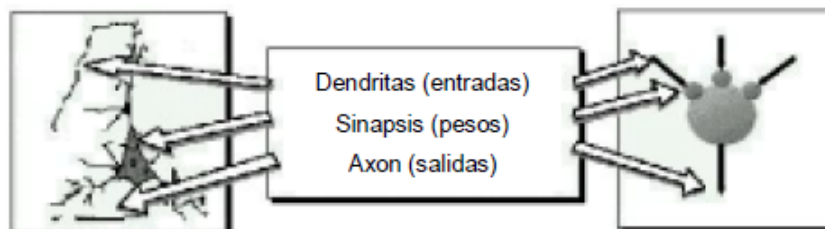


Fuente: Sancho, F. (2019)

Las variables X_1, X_2, \dots, X_n , con $n \in \mathbb{N}$, son las variables de entrada en la red y las variables Y_1, Y_2, \dots, Y_n , con $n \in \mathbb{N}$, representan las variables de salida. Cabe aclarar que las variables de salida pueden ser en cantidad menor a n .

Adicionalmente, en la figura 2.8 se puede observar cómo la neurona artificial se asemeja a una neurona biológica, como se había dicho en párrafos anteriores.

Figura 2.8: Neurona biológica vs neurona artificial



Fuente: Sancho, F. (2019)

La neurona artificial tiene entradas, pesos y salidas, las cuales, son conocidas en

una neurona biológica como dendritas, sinapsis y axón, respectivamente.

Proceso de la red neuronal artificial

- **Función de entrada**

Dado que la neurona recibe muchos valores de entrada como si fuera uno solo, se trata de combinar a todas estas entradas a través de la *función de entrada*, la cual se expresa como una *suma ponderada* como se indica en la ecuación (2.17)

$$z_i = (X_{i1} * w_{i1}) + (X_{i2} * w_{i2}) + \dots + (X_{ij} * w_{ij}) + b_i \quad (2.17)$$

Donde:

X_{ij} : Representa la entrada j , con $j = 1, 2, \dots, n$, $n \in \mathbb{N}$ (n entradas), a la neurona i , con $i = 1, 2, \dots, N$ (N neuronas).

w_{ij} : Corresponde al peso de la entrada j para la neurona i .

b_i : Es sesgo (*bias*) de la neurona i . Este parámetro es una constante que incrementa o disminuye en la neurona la suma ponderada; es decir, varía al valor de la función de entrada del nodo i . A menudo es considerado como un peso que proviene de una entrada de valor unitario.

En esta función los valores de entrada se multiplican por los pesos ingresados con anterioridad a la neurona; estos pesos son aleatorios entre -1 y 1.

- **Función de activación**

Dado que la neurona artificial se asemeja a una neurona biológica y esta tiene estados de activación que pueden ser activa o inactiva, entonces la neurona artificial también tiene un estado de activación; por tanto, el objetivo de la *función de activación*, conocida también como *función de transferencia*, es calcular el estado de actividad de una neurona, para lo cual transforma el valor de entrada global de la neurona (el valor de la suma ponderada menos el umbral o el sesgo (*bias*) b_i) a un estado de activación, cuyo rango va normalmente de (-1 a 1) o de (0 a 1); indicando así, que una neurona puede estar totalmente inactiva (0 o -1) o totalmente activa (1).

Las funciones de activación más comunes son:

1. Función lineal

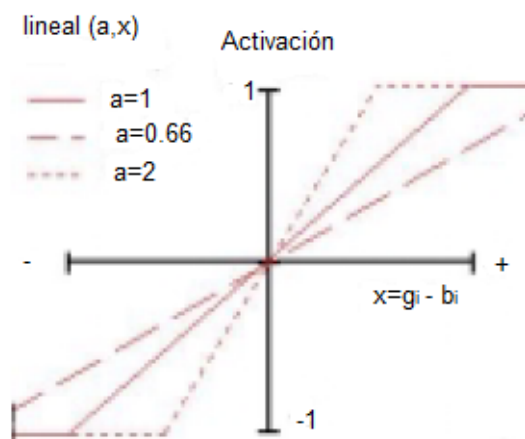
La función lineal viene dada por la ecuación (2.18):

$$f(x) = \begin{cases} -1 & \text{si } x \leq -1/a \\ ax & \text{si } -1/a < x < 1/a \\ 1 & \text{si } x \geq 1/a \end{cases} \quad (2.18)$$

donde: $x = z_i - b_i$, con $i = 1, \dots, N$, $N \in \mathbb{N}$ y $a > 0$.

Es así que los valores de salida de esta función serán: ax en caso de x este dentro del intervalo $(-1/a, 1/a)$; y, cuando x esté por debajo o sobre esta zona, el valor de salida se fijará en -1 o 1 , respectivamente, como se muestra en la figura 2.9

Figura 2.9: Función de activación lineal



2. Función sigmoide

La función sigmoide es la función más antigua y usada para activar las neuronas dentro de una red neuronal, debido a que es continua diferenciable, lo cual ayudará para el cálculo del entrenamiento de la red (esto lo veremos más adelante). Esta representada por la ecuación (2.19)

$$f(z) = \frac{1}{1 + e^{-z}} \quad \text{con } z = b_i + \sum_j^n w_{ij}x_{ij} \quad (2.19)$$

donde:

n es el número de entradas a la neurona i .

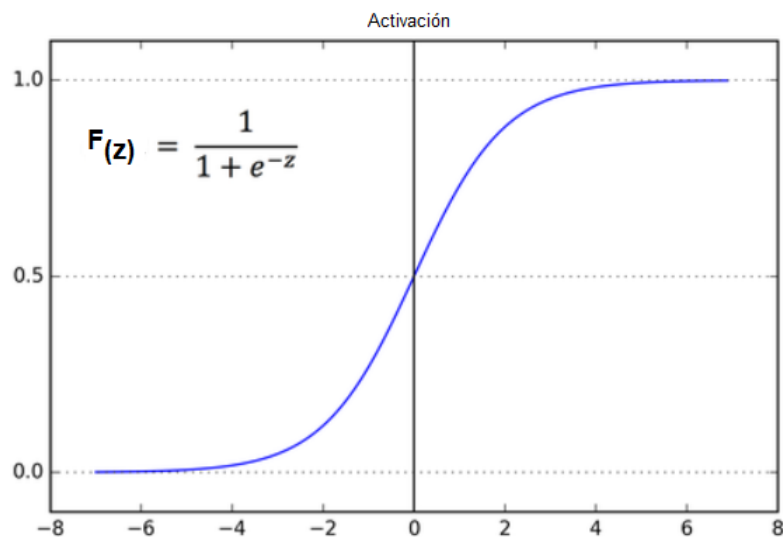
X_{ij} : Es la entrada j , con $j = 1, 2, \dots, n$, a la neurona i , con $i = 1, 2, \dots, N$.

w_{ij} : Es el peso de la entrada j para la neurona i .

b_i : Es el bias o sesgo de la neurona i .

En la gráfica 2.10 se puede observar que los valores de salida proporcionados por esta función están en el rango de (0,1). Además, para valores de z mayores a 0, e^{-z} converge a 0, provocando que $f(z)$ se aproxime a 1. De similar manera, para valores de z menores a 0, e^{-z} crece exponencialmente, lo que provoca que $f(z)$ se aproxime a 0.

Figura 2.10: Función de activación sigmoide



Esta función de activación es preferible usarla para cuando se desee valores de salidas probabilísticos.

3. Función Tangente Hiperbólica

La función tanh tiene similar forma a la de la función sigmoidea, excepto que se reescala horizontalmente y verticalmente a $[-1, 1]$ (se estira).

Matemáticamente esta representada por la ecuación (2.20)

$$f(z) = \frac{2}{1 + e^{-2z}} - 1 \quad \text{con} \quad z = b_i + \sum_j^n w_{ij}x_{ij} \quad (2.20)$$

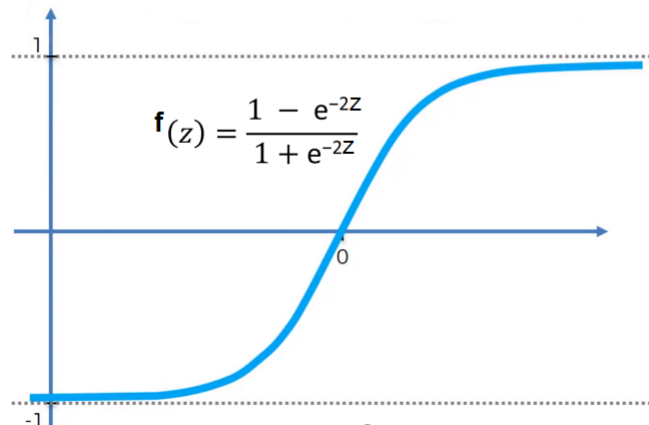
Para todo $i = 1, \dots, N$, con N cantidad total de neuronas en la red.

En la figura 2.11 se ve que los valores de salida de la función va de $[-1, 1]$.

Esta función de activación es recomendable usar cuando se desee que los valores de salida sean tanto positivas, como negativas.

Una especificación más general, al aplicar la función de transferencia a la red para obtener el valor final de la salida (*output*) se puede representar como el

Figura 2.11: Función de activación tangente hiperbólica



apartado (2.21)

$$\hat{y} = \beta_0 + \sum_{i=1}^q \beta_i f_q \left(\sum_{j=1}^p w_{ij} x_{ij} + b_i \right) \quad (2.21)$$

Donde:

f_q : Función de transferencia correspondiente a la última capa oculta (penúltima capa de todas las capas).

w_{ij} : Peso de la conexión entre la j -ésima entrada (*input*) con la i -ésima neurona oculta.

b_i : Término independiente o "bias" que le corresponde la i -ésima neurona oculta.

β_i : Peso correspondiente a la conexión entre la i -ésima neurona oculta y la neurona de salida (output).

β_0 : término independiente de la neurona de salida n (output).

- **Aprendizaje de la red neuronal**

Una red neuronal al recibir información tiene la obligación de aprender, de tal manera que entienda las entradas que recibe, las procese y los clasifique o etiquete.

Para que la red aprenda, se hace uso de algoritmos de aprendizaje automático, de los cuales dos son los más conocidos: supervisado, y no supervisado.[24]

1. *Supervisado*: En el aprendizaje supervisado la red genera resultados a partir de entradas determinadas. Está controlado por un agente externo que controla la salida de la red y en caso de que esta no coincida con la deseada, se modifican los pesos de las conexiones, con el fin de conseguir que la salida obtenida se aproxime a la deseada.

Este tipo de aprendizaje puede ser llevado a cabo a través de *aprendizaje por corrección de error*, el cual consiste en ajustar los pesos de las conexiones de la red en función de la diferencia entre los valores deseados y los obtenidos a la salida de la red, es decir, en función del error cometido en la salida.

2. *No-supervisado*: En este tipo de aprendizaje no se entrega o proporciona datos de entrada más bien, se da la libertad para que el algoritmo encuentre por si solo variantes de patrones y correlaciones. Suele ser usado para cuando se desea descubrir qué hay detrás de los datos que no vemos; por ejemplo, para perfilado o “*clustering*” de clientes.

Los algoritmos de aprendizaje automático tienen como objetivo minimizar o maximizar una función, la cual llamamos **función objetivo** o **función de costo**. El grupo de funciones que son minimizadas se llaman las funciones de costo; estas son una medida de cuán bueno es un modelo. Para ir avanzando por la minimización de estas funciones se utiliza comúnmente el método “descenso de gradiente”. Este método se explicará más adelante.

La función de costo más utilizada, por ser considerada como la mejor, es la del error cuadrático medio (MSE en inglés), cuya función está representada matemáticamente en la ecuación (2.22)

$$MSE = \frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{2} \quad (2.22)$$

donde:

n : Es el número de observaciones de la base.

y_i : Representa la variable objetivo para cada observación $i = 1, \dots, n$

\hat{y}_i : Representa los valores predichos.

Sin embargo, no hay una sola función de costo que funcione con todo tipo de datos, pues depende de muchos factores como la presencia de puntos anormales, el algoritmo de aprendizaje automático escogido, la capacidad computacional que se puede permitir o la facilidad de encontrar derivadas.[24]

Las funciones de costo se pueden categorizar en dos grandes tipos, o bien para problemas de clasificación o bien para problemas de regresión. En este trabajo nos concentraremos más en funciones de costo para regresión puesto que están más ligadas a la predicción de quiebra de las empresas manufactureras ecuatorianas.

En resumen, para las redes neuronales, aprender significa que al recibir un con-

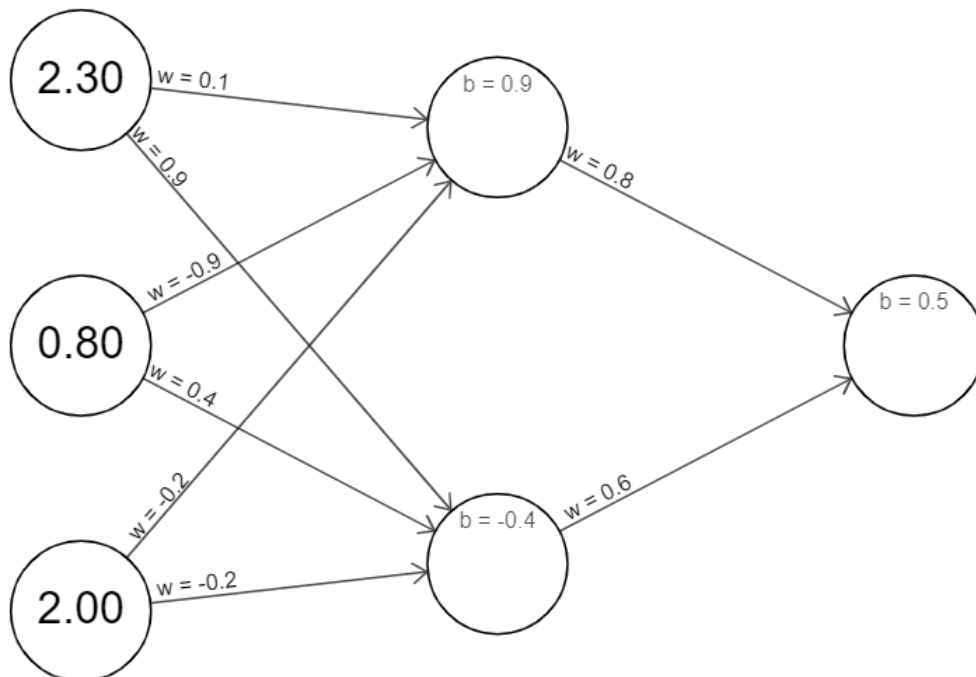
junto de características u observaciones y una estructura de función, utilice un algoritmo para encontrar un proceso que optimice la función de costo relacionada con las observaciones. Generalmente las funciones de costo dependen de la tarea de aprendizaje necesario.

Para entender mejor el proceso hasta este punto, se procede a explicar con un pequeño ejemplo de red neuronal y lo que se hace en principio con los pasos señalados anteriormente.

Considerar que se tiene una red neuronal con dos capas, una capa oculta y una capa de salida. Antes de continuar, hay que tomar en cuenta que cuando se cuenta el número de capas en una red, solo se cuenta las capas que reciben entradas, es decir, la capa de entrada será omitida.

Entonces el proceso de análisis empieza en la capa de entrada, la cual consta de tres variables de entrada; se otorgan valores aleatorios al conjunto de pesos y sesgos, para ponderar a las neuronas de la primera capa oculta y, de igual manera, se prueba un conjunto de pesos aleatorios para la capa de salida, como también el sesgo de la neurona de salida final. Como se puede observar en la figura 2.12, donde nuestros valores de entrada (variables de entrada) corresponde a $X = [X_1, X_2, X_3] = [2,3; 0,8; 2,0]$

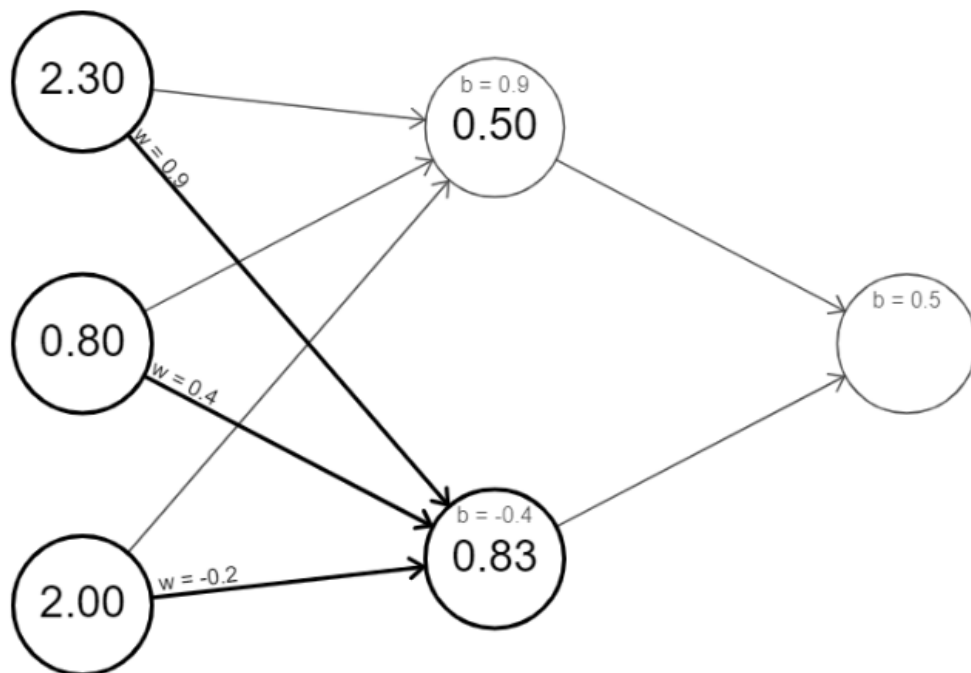
Figura 2.12: Primer paso de una red neuronal



Como segundo paso se procede a aplicar la función de entrada para las unidades de la capa oculta, cuya función se dijo en párrafos anteriores que es la suma ponde-

rada entre el producto de la variable X y el peso asignando a la misma, adicionando el sesgo b . Así, se obtiene las entradas para la capa oculta y simultáneamente se aplica la función de activación al valor que se obtenga de la suma ponderada. Este resultado se puede observar en la figura (2.13)

Figura 2.13: Segundo paso de una red neuronal



Es decir, los valores de las neuronas de la capa oculta han sido calculados tal como se muestra en las ecuaciones (2.23), (2.24) y (2.25), (2.26), donde las ecuaciones (2.23) y (2.24) se aplicó la función de entrada (suma ponderada) y las ecuaciones (2.25) y (2.26) se aplica la función de activación a la suma ponderada de cada neurona respectivamente:

$$z_1 = (0,1 * 2,3) + (-0,9 * 0,8) + (-0,2 * 2) + (0,9) = 0,0 \quad (2.23)$$

$$z_2 = (0,9 * 2,3) + (0,4 * 0,8) + (-0,2 * 2) + (0,4) = 1,6 \quad (2.24)$$

$$f(z_1) = f(0,0) = 0,5 \quad (2.25)$$

$$f(z_2) = f(1,6) = 0,83 \quad (2.26)$$

Siguiente paso, es muy similar al anterior, pero ahora será aplicado para la(s) neuronas de la capa de salida.

Se toma estos dos últimos valores calculados, para alimentar en el proceso de entrada de la neurona siguiente. Se observan las ecuaciones (2.27) y (2.28), donde

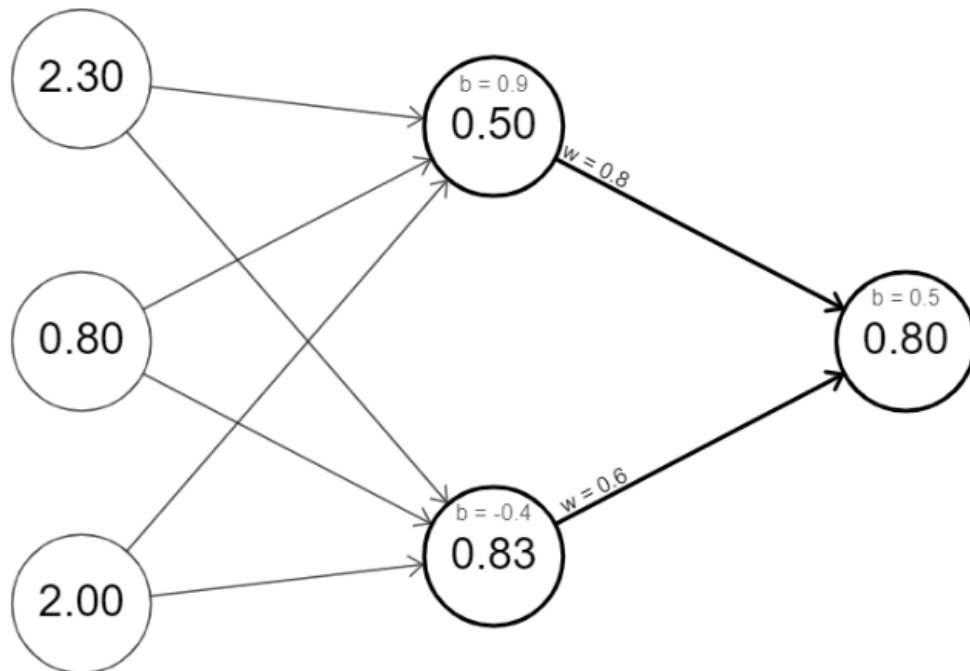
indican la función de entrada y función de activación:

$$z_3 = (0.8 * 0.5) + (-0.6 * 0.83) + (0.5) = 1.4 \quad (2.27)$$

$$f(z_3) = f(1.4) = 0.80 \quad (2.28)$$

En la figura 2.14 se puede ver el resultado obtenido:

Figura 2.14: Tercer paso de una red neuronal



Es así que se ha obtenido un primer valor; sin embargo, el error de este resultado será medido por la función de coste, que como se dijo, será el error cuadrático medio (MSE), que es la más común para cuantificar el error en problemas de regresión.

Suponiendo que el el valor de salida deseado era 0,45, entonces se procede a medir el error, tal que al aplicar la función MSE, se tiene como resultado que el error del aprendizaje de la red es de 0,1225, por lo que se procede a continuar entrenando la red y esto se lo hará a través del aprendizaje *Backpropagation*, el cual se verá a continuación.

2.7.3. Algoritmo de aprendizaje *Backpropagation*

Para entrenar la red, por muchos años se buscó algoritmos de entrenamiento de redes multicapa (que tienen más de una capa), sin conseguir éxito alguno; por lo que al limitarse en obtener resultados solo con una capa y no más de dos, el mundo

de la red neuronal entró en un abandono durante décadas. Sin embargo, por el año 1985 se da la invención del algoritmo de *backpropagation*, el cual, desde entonces ha desempeñado un papel fundamental en el proceso de aprendizaje de las redes neuronales artificiales por utilizar un aprendizaje sencillo y por ser potente en su capacidad de entrenar capas ocultas.[60]

Backpropagation es un método de entrenamiento supervisado. Por cada entrenamiento los pesos son ajustados de forma que disminuya el error entre la salida deseada y la respuesta de la red.

Este algoritmo de aprendizaje conlleva realizar una fase de propagación hacia adelante y otra fase de propagación hacia atrás por cada patrón presentado en el entrenamiento.

- **Propagación hacia delante:**

Inicia cuando se presenta un vector patrón (variables externas) en la capa de entrada de la red y cada neurona en la capa de entrada se conecta a estos patrones. Consecutivamente, el resto de capas empieza a realizar el proceso del que se habló en la sección anterior sobre el proceso y aprendizaje de la red neuronal artificial, determinando el nivel de activación de cada capa. Las operaciones se van haciendo de izquierda a derecha, es decir, de la capa de entrada, van a la primera capa oculta, después van a la segunda capa oculta y así consecutivamente hasta llegar a la última capa encargada de generar el resultado (capa de salida). Esto es la propagación hacia adelante.

- **Propagación hacia atrás**

Consecuentemente después de completarse la fase de propagación hacia adelante se compara el resultado de salida obtenido por el que se deseaba obtener y calcular un valor de error para cada neurona de tal manera que a este se lo optimice. Para ello se inicia con la fase de corrección o fase de propagación hacia atrás en combinación con el descenso del gradiente. Es, decir, se modifica todos los pesos de las conexiones empezando por la capa de salida y continuando hacia atrás a través de todas las capas de la red hasta la capa de entrada.

Por ende, este algoritmo se lo utiliza como un método para calcular las derivadas parciales de cada uno de los parámetros de la red con respecto a la función de costo. Las derivadas parciales se necesitan calcular porque son parte fundamental del vector del gradiente, que ayuda a que el algoritmo del descenso del gradiente funcione y se pueda optimizar la función de costo.

En este proceso de retropropagación se mide cuantitativamente el aprendizaje mediante el error cuadrático medio de la red, de tal manera que refleje el modo en el que la red está logrando respuestas correctas y así, a medida que la red aprende su error cuadrático medio decrece. Esto se logra visualizar cuando la red converge; pero la convergencia no siempre es fácil de conseguir porque a veces el proceso puede requerir un tiempo excesivo o bien porque la red alcanza un mínimo local y deja de aprender.

2.7.4. Formulación Matemática del algoritmo *Backpropagation*

Una red neuronal inicialmente tiene sus parámetros de entrada inicializados de forma aleatoria, significando que el resultado que se obtendrá en un inicio será aleatorio, y cuando se lo compare con la predicción real, probablemente la predicción sea completamente ineficiente y la función de costo le asignará a la red un error muy elevado.

Ahora usemos dicho error para entrenar a la red.

Se quiere calcular para cada parámetro dentro de la red neuronal la derivada parcial del coste respecto al peso w y al término bias b . Se muestran en la expresión (2.29)

$$\frac{\partial C}{\partial w} \quad y \quad \frac{\partial C}{\partial b} \quad (2.29)$$

donde:

C : representará la función de Coste que se utilice, en este trabajo será el error cuadrático medio;

w : representa el vector de pesos de la red y,

b : representa el vector de sesgos de cada neurona en las capas (excepto de la capa de entrada).

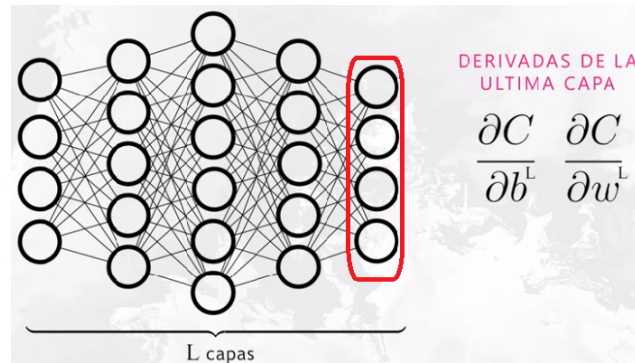
Se empieza a trabajar hacia atrás y para ello se calcula los parámetros de la última capa, como se visualiza en la figura 2.15, donde L indica el número de capa en la que se está trabajando, suponiendo que se tiene L capas.

Recordando que el objetivo es optimizar el error de la red, el cual, se podría decir está representado como una composición de funciones, como se ve en la expresión (2.30).

$$ERROR = C(f(Z^L)) \quad (2.30)$$

donde:

Figura 2.15: Cálculo de los parámetros de la última capa



Z^L : Es el resultado de la suma ponderada en la capa L, la última capa.

$f(\cdot)$: Representa la función de activación utilizada.

C: Función de coste que evaluará el resultado de activaciones en la última capa. En esta investigación la función de coste será el error cuadrático medio.

Se procede entonces a calcular la derivada parcial de una función compuesta, a través de la regla de la cadena.

Tomando en cuenta la expresiones (2.29) y (2.30), lo que se necesita calcular son todas las derivadas internas que se presentan en las ecuaciones (2.31) y (2.32).

$$\frac{\partial C}{\partial w^L} = \frac{\partial C}{\partial f^L} * \frac{\partial f^L}{\partial z^L} * \frac{z^L}{w^L} \quad (2.31)$$

$$\frac{\partial C}{\partial b^L} = \frac{\partial C}{\partial f^L} * \frac{\partial f^L}{\partial z^L} * \frac{z^L}{b^L} \quad (2.32)$$

Señalando que $Z^L = W^L f^{L-1} + b^L$, donde f^{L-1} es el valor de salida de la capa anterior, es decir, la función de activación de la capa anterior a la capa L aplicada a la suma pondera de las neuronas de la capa anterior. W^L representa el vector de pesos de la capa L y b^L el *bias* de las neuronas de la capa L.

Debido a que la función de costo en este trabajo de investigación será el error cuadrático medio (ecuación 2.33), entonces la derivada parcial del costo respecto a la función de activación se señala en la ecuación (2.34).

$$C(f_j^L) = \frac{1}{2} \sum_j (y_j - f_j^L)^2 \quad (2.33)$$

$$\frac{\partial C}{\partial f_j^L} = (f_j^L - y_j) \quad (2.34)$$

Donde:

f_j^L : Representa la función de activación de la neurona j dentro de la capa L .

y_j : Es el valor observado de la neurona j .

Luego, se procede a calcular la variación de la salida de la neurona f^L , cuando varía la suma ponderada de la neurona z^L .

Si la función de activación es la sigmoide (ver expresión 2.35), y justamente será la empleada en este trabajo de investigación, entonces se sigue el cálculo como se muestra en la ecuación 2.36:

$$f^L(z^L) = \frac{1}{1 + e^{-z^L}} \quad (2.35)$$

$$\frac{\partial f^L}{\partial z^L} = f^L(z^L) * (1 - f^L(z^L)) \quad (2.36)$$

Ahora, calculando la variación de la suma ponderada cuando varían los parámetros w y b (ver ecuaciones 2.38 y 2.39), tomando en cuenta que la suma ponderada en la capa L es como se muestra en la expresión 2.37, se tiene que:

$$z^L = \sum_j f_j^{L-1} w_j^L + b_j^L \quad (2.37)$$

$$\frac{\partial z^L}{\partial w^L} = f_j^{L-1} \quad (2.38)$$

$$\frac{\partial z^L}{\partial b^L} = 1 \quad (2.39)$$

Una vez calculadas todas las derivadas internas, se reemplaza en las ecuaciones 2.29 y 2.30 para obtener las derivadas de los parámetros de la red.

Se va a interpretar a $\frac{\partial C}{\partial z^L}$ como el error imputado a la neurona, debido a que esta derivada va a decir qué responsabilidad tiene la neurona en el resultado final y por tanto, en el error. Se la va a representar con el símbolo δ^L . Así, con esto último dicho y por las ecuaciones (2.38) y (2.39), las derivadas de los parámetros de la red en la capa de salida serán como se indica en las expresiones (2.40) y (2.41).

$$\frac{\partial C}{\partial w^L} = \delta^L f_j^{L-1} \quad (2.40)$$

$$\frac{\partial C}{\partial b^L} = \delta^L(1) = \delta^L \quad (2.41)$$

donde:

$$\delta^L = \frac{\partial C}{\partial f^L} \frac{\partial f^L}{\partial z^L} \quad (2.42)$$

será el error imputado de la neurona.

Con todo esto se ha determinado las derivadas de la capa de salida, pero hace falta una más para determinar las derivadas del resto de la red. Se analiza cómo se podría obtenerla.

Se considera ahora que se desea calcular los parámetros de la capa anterior, (L-1), a la capa de salida (L). Entonces, usando el mismo razonamiento anterior se tendría que:

En este caso cuando se calcula las derivadas de los parámetros, al aplicar la regla de la cadena, la expresión se prolonga hasta convertirse en lo que se señala en las ecuaciones (2.43) y (2.44)

$$\frac{\partial C}{\partial w^{L-1}} = \frac{\partial C}{\partial f^L} * \frac{\partial f^L}{\partial z^L} * \frac{\partial z^L}{\partial f^{L-1}} * \frac{\partial f^{L-1}}{\partial z^{L-1}} * \frac{z^{L-1}}{w^{L-1}} \quad (2.43)$$

$$\frac{\partial C}{\partial b^{L-1}} = \frac{\partial C}{\partial f^L} * \frac{\partial f^L}{\partial z^L} * \frac{\partial z^L}{\partial f^{L-1}} * \frac{\partial f^{L-1}}{\partial z^{L-1}} * \frac{z^{L-1}}{b^{L-1}} \quad (2.44)$$

Porque ahora el error sería como se indica en la expresión (2.45).

$$ERROR = C(f^L(W^L f^{L-1}(W^{L-1} f^{L-2} + b^{L-1}) + b^L)) \quad (2.45)$$

Pero se vio que:

$$\delta^L = \frac{\partial C}{\partial f^L} \frac{\partial f^L}{\partial z^L}$$

y que el cálculo de las derivadas del peso y del *bias* es igual a como se calculó anteriormente, solo que ahora representará la activación de la capa $L - 2$, como se ve en las ecuaciones (2.46) y (2.47)

$$\frac{\partial z^{L-1}}{\partial w^{L-1}} = f_j^{L-2} \quad (2.46)$$

$$\frac{\partial z^{L-1}}{\partial b^{L-1}} = 1 \quad (2.47)$$

Además, $\frac{\partial f^{L-1}}{\partial z^{L-1}}$ representa la derivada de la función de activación y $\frac{\partial z^{L-1}}{\partial f^{L-1}}$ corresponde a la matriz de pesos W^L , que conecta ambas capas, la cual señala cómo varía

la suma ponderada de una capa cuando varía la salida de una neurona en la capa previa.

Con esto, se tiene al igual que en la capa de salida que el error imputado de las neuronas en la capa $L - 1$ será como se muestra en la expresión (2.48)

$$\frac{\partial C}{\partial z^{L-1}} = \delta^{L-1} \quad (2.48)$$

Por tanto, con el análisis hecho en esta capa, ya se es posible extender al resto de las capas de la red.

En resumen lo que hace al algoritmo es:

1. Computar del error de la última capa

$$\delta^L = \frac{\partial C}{\partial f^L} \frac{\partial f^L}{\partial z^L} \quad (2.49)$$

2. Retropropagar el error de la capa anterior (ecuación 2.50)

$$\delta^{l-1} = W^l \delta^l \frac{\partial f^{L-1}}{\partial z^{l-1}} \frac{\partial f^L}{\partial z^L} \quad (2.50)$$

3. Calcular las derivada de las capas usando el error

$$\frac{\partial C}{\partial b^{l-1}} = \delta^{l-1} \quad (2.51)$$

$$\frac{\partial C}{\partial w^{l-1}} = \delta^{l-1} f^{l-2} \quad (2.52)$$

y así, sucesivamente recorriendo todas las capas de la red hasta el final. Con esto se habrá calculado todos los errores y las derivadas parciales de la red haciendo uso sólo de las cuatro expresiones anteriores.

2.8. Criterios Estadísticos para seleccionar variables y Medir el desempeño de los Modelos

2.8.1. Metodología para Selección de variables

Valor de Información (VI)

En los modelos de clasificación binaria (Bueno/Malo), Fair Isaac (FI)² elaboraron una medida que se denominó valor de información (*Information Value*) que se utiliza para medir el poder predictivo de una categoría de una variable cualitativa sobre un modelo estadístico. Esta medida se la conoce técnicamente como la medida de Kullback.

En el proceso de análisis preliminar de los datos, es de suma importancia elegir las variables predictoras o covariables adecuadas para el modelo, a través de las cuales se pueda distinguir entre empresas en quiebra y no quiebra.

Este método permite una adecuada selección de variables categóricas y luego hacer su inclusión al modelo. EL valor de información (VI) se lo calcula de la siguiente manera:

$$VI = \sum_{i=1}^n \left[\left(\frac{b_i}{B} - \frac{m_i}{M} \right) \ln \left(\frac{\frac{b_i}{B}}{\frac{m_i}{M}} \right) \right]$$

Donde:

- n : Número de categorías en que se ha clasificado la variable categórica.
- m_i : Número de elementos etiquetados como Bueno dentro de la categoría i .
- B : Total de elementos etiquetados como Bueno.
- b_i : Número de elementos etiquetados como Malo dentro de la categoría i .
- M : Total de elementos etiquetados como Malo.

El valor de información se encuentra en el rango de cero a infinito. Aunque los valores más comunes son los que se encuentran entre 0 y 1, y mientras más grande sea su valor, más predictiva será la variable categórica, influyendo significativamente el rendimiento del modelo. Para determinar la capacidad predictiva de cada variable categórica, se hará uso de la siguiente tabla de clasificación por rangos:

²Los pioneros más conocidos de la clasificación crediticia son el Ingeniero Bill Fair y el Matemático Earl Isaac, quienes fundaron su consultora, Fair Isaac (FI), en San Francisco en 1956 [5].

Cuadro 2.2: Regla para el Valor de Información

| VI | Poder de Predicción |
|------------|----------------------|
| <0,02 | Predictor Ineficaz |
| 0,02 a 0,1 | Predictor Débil |
| 0,1 a 0,3 | Predictor Medio |
| 0,3 a 0,5 | Predictor Fuerte |
| >0.5 | Predictor Sospechoso |

Fuente: (SAS, 2009)

Lo ideal sería elegir a las variables que tengan un poder predictivo medio y fuerte, por lo que bastaría considerar al valor de información superior al 0,1. Se debe tomar en cuenta que las variables con valores superiores a 0,5 serán revisadas ya que puede existir una sobreestimación o sobreajuste del modelo.

2.8.2. Pruebas y Desempeño de los Modelos

Multicolinealidad

Para analizar la relación que existe entre la variable respuesta Y y las covariables X_1, X_2, \dots, X_n se utiliza la multicolinealidad; es decir, se comprueba que no exista correlación entre las variables explicativas dentro del modelo de regresión. Esta correlación entre variables es un problema ya que genera un error estándar, lo que implica que los coeficientes estimados del modelo sean inestables y por lo tanto no sean precisos. Además, la multicolinealidad entre variables no permite distinguir el impacto individual de cada variable sobre la variable dependiente; lo que conlleva a una mala interpretación del modelo y también crea problemas de sobreajuste.

Para calcular el nivel de multicolinealidad entre las covariables se aplica la medida del *Factor de inflación de la varianza (VIF)*. Esta medida indica el grado en el que la varianza del coeficiente estimado de una variable explicativa aumenta debido a la correlación de la misma respecto al resto de variables explicativas del modelo.

El *Factor de inflación de la varianza (VIF)*, se define como:

$$VIF_i = \frac{1}{1 - R_i^2}$$

donde, R_i^2 , el coeficiente de determinación: es una proporción de la varianza que

resulta de la regresión entre la variable explicativa i y el resto de variables explicativas. Con este resultado se puede conocer qué tan correlacionada se encuentra la covariable i con el resto de covariables. Por último, un valor alto para R_i^2 indica una alta correlación.

Esta correlación es capturada por el VIF , cuanto más cerca se encuentre el valor de R_i^2 de 1, mayor será el valor de VIF y mayor será la multicolinealidad con la variable explicativa i .

Kolmogorov - Smirnov (KS)

Uno de los estadísticos comúnmente utilizados, es la estadística KS, la cual se basa de un análisis de la función de distribución empírica acumulada (ECDF).

El test de Kolmogorov-Smirnov es una prueba de bondad de ajuste, donde se contrasta la hipótesis de si las muestras aleatorias e independientes provienen de distribuciones continuas idénticas.

A continuación se describe esta prueba para dos muestras aleatorias donde se contrastan las hipótesis:

$$\begin{cases} H_0 : F_1(x) = F_2(x) \forall x \\ H_1 : F_1(x) \neq F_2(x) \end{cases}$$

donde F_1 es la función de acumulación de una muestra X_1 de tamaño n_1 y F_2 es la función de acumulación de una muestra X_2 de tamaño n_2 . El valor del estadístico KS usado para contrastar la hipótesis nula H_0 resulta de la siguiente expresión:

$$KS = \max_x |\tilde{F}_1(x) - \tilde{F}_2(x)|$$

La hipótesis nula se rechaza si el estadístico KS es mayor que el valor crítico KS_α para un nivel de significancia α dado.

En el presente trabajo se comparan las distribuciones de los individuos obteniendo las variables que presenten un mayor estadístico KS que se encuentre por encima del 4 %, es decir, las que mayor poder predictivo tengan para realizar el cálculo de la predicción de la probabilidad de impago.

Área bajo la curva ROC

Para evaluar el poder discriminatorio de un sistema de clasificación, Basilea propone varias metodologías estadísticas para ser usadas dentro de la industria financiera. En este trabajo se hará uso de la Curva ROC. Para validar un proceso adecuado discriminatorio la curva ROC debe probarse en un conjunto de entrenamiento y otro de validación.

La curva ROC (Receiver Operating Characteristic) es un gráfico de propabilidad que indica el rendimiento de una metodología de clasificación evaluando así la capacidad que tiene para discriminar entre buenos y malos, en este caso particular entre empresas quebradas y no quebradas. La curva ROC esta representada mediante la relación:

- **Sensibilidad:** Razón de positivos verdaderos, es decir, es la proporción de empresas no quebradas que fueron calificadas como no quebradas respecto al total de empresas clasificadas.
- **1-especificidad:** Razón de falsos positivos, es decir, es la proporción de empresas quebradas que fueron calificadas como no quebradas respecto al total de empresas clasificadas.

Estos resultados se consiguen recorriendo el umbral de clasificación permitiendo así maximizar la sensibilidad, al mismo tiempo que minimiza el complemento de la especificidad. Este punto de corte toma valores en el intervalo [0,1]

Con el uso de la curva ROC se procede a calcular uno de los índices que miden el rendimiento de un modelo de clasificación (capacidad de clasificar correctamente a los sujetos) conocido como AUC (Area Under the Curve), que no es más que el área bajo la curva ROC. Esta medida, es una transformación lineal, donde en el eje horizontal se encuentra el porcentaje de falsos positivos, es decir, total de sujetos o empresas que el modelo clasificó como buenos, en otras palabras empresas con una mínima probabilidad de bancarrota; en el eje vertical se encuentra el porcentaje de verdaderos positivos que son empresas buenas y que fueron clasificados como buenos en el modelo.

El área bajo la curva AUC puede tomar valores entre 0,5 y 1. Un porcentaje del AUC del 50 % implica que el modelo no es el adecuado para medir el rendimiento de clasificación; y un valor de 100 % indica la improbable aparición de predicciones perfectamente correctas. No obstante, para el caso de sistemas de clasificación lo

mejor sería tomar valores del AUC que sean mayores a 0,7 y, entonces se dirá que esos modelos tienen una buena capacidad de clasificación [5].

Análisis de la *Deviance*

La deviance para un modelo Y está definida como:

$$D(Y, \mu) = 2[\iota(\mu_{max}; y) - \iota(\mu; y)]$$

Donde μ_{max} es el valor del parámetro que maximiza $\iota(\mu; y)$ en el modelo, siendo $\iota(\cdot)$ el logaritmo de la verosimilitud. La *deviance* desempeña el papel de la suma de cuadrados del error y se utiliza para evaluar la bondad de ajuste y para comparar modelos. Para la comparación de modelos, supongamos \hat{M}_1 y \hat{M}_2 dos modelos ajustados con \hat{M}_1 anidado dentro de \hat{M}_2 , bajo algunas condiciones de regularidad

$$D(\hat{M}_2, \hat{M}_1) = D(y; \hat{M}_1) - D(y; \hat{M}_2)$$

Para los modelos no paramétricos y aditivos, la deviance tiene sentido para evaluar los modelos y ver sus diferencias.

Matriz de confusión

La matriz de confusión Kohavi y Provost (1998) es una herramienta que evalúa la precisión y exactitud de un modelo de clasificación. Se encarga de comparar entre los valores reales y las pronosticadas por el modelo para la variable respuesta.

Para evaluar que tan bien ha funcionado un modelo predictivo, con resultados binarios consiste en calcular el porcentaje de cuantos han sido calificados erróneamente. Es así que, los casos clasificados correctamente son los verdaderos positivos (malos) y negativos (buenos). Si ellos no corresponden a las categorías mencionadas, se etiquetan falso positivo (Error Tipo I) esperando que el malo sea bueno y negativo (Error tipo II) esperando que el bueno sea malo.

Cuadro 2.3: Matriz de Confusión

| | | Observación | |
|------------|-----------|-------------------------|-------------------------|
| | | Positivos | Negativos |
| Predicción | Positivos | Verdadero Positivo (VP) | Falso Positivo (FP) |
| | Negativos | Falso Negativo (FN) | Verdadero Negativo (VN) |

Elaboración: Los autores

En donde (ver cuadro 2.3) cada fila corresponde a los valores pronosticados para la variable objetivo; cada columna representa los valores reales para la variable objetivo. Cada entrada de la matriz de confusión tiene las siguientes significados:

- **Verdadero Positivo (VP):** Si una entidad o institución Buena es clasificada como buena.
- **Falso Negativo (FN):** Si una entidad o institución Buena es clasificada como Mala.
- **Falso Positivo (FP):** Si una entidad o institución Mala es clasificada como Buena.
- **Verdadero Negativo (VN):** Si una entidad o institución Mala es clasificada como Mala.

Con base en el del cuadro 2.3 la matriz tiene las siguientes métricas asociadas más importantes:

Cuadro 2.4: Métricas de Desempeño

| Métrica | Definición | Fórmula |
|---------------------|--|-----------------------|
| Precisión | Porcentaje de predicciones correctas del total de sujetos | $\frac{VP+VN}{TP+TN}$ |
| Sensibilidad | Porcentaje de Buenos clasificados correctamente del total de Buenos. | $\frac{VP}{VP+FN}$ |
| Especificidad | Porcentaje de Malos clientes correctamente frente al total de Malos | $\frac{VN}{FP+VN}$ |
| Predicción positivo | Porcentaje de Buenos clasificados correctamente frente al total de buenos predichos | $\frac{VP}{VP+FP}$ |
| Predicción Negativa | Porcentaje de Malos clasificados correctamente respecto al total de Malos predichos. | $\frac{VN}{FN+VN}$ |

Elaboración: Los autores

Capítulo 3

Metodología

En el presente capítulo comienza el desarrollo de la estructura metodológica aplicada en esta investigación para cumplir con los objetivos planteados. A partir de ello, se expone la metodología a realizarse la cual consiste en primera instancia analizar las variables que intervienen en el estudio, las cuales corresponden a una base empresarial manufacturera; Es necesario, estimar y validar un modelo aditivo generalizado que proporcione las variables que predigan el quiebre empresarial para ser usadas en la construcción de varios modelos de redes neuronales artificiales, luego comparar todos los modelos RNA realizados y seleccionar el de menor error, para posteriormente validarlo y con ello proceder a realizar el pronóstico de fracaso empresarial en el sector manufacturero. En el siguiente capítulo se expondrán los resultados obtenidos.

3.1. Datos y Variables

Para la estimación de los modelos estadísticos mencionados en el capítulo anterior, se dispone de una base de datos a partir de la cual se obtuvo información financiera de la lista total de empresas que reportaron sus finanzas desde el año 2016 hasta el 2020 en la SUPERCIAS, que corresponde a los estados y reportes financieros; conjuntamente se adicionó la información de clasificación de las empresas según su tamaño y variables geográficas tales como la provincia y la región a la que pertenecen, la que será extraída desde el ente regulador.

Como ya se había mencionado, la información que servirá de insumo para cumplir con los objetivos de esta investigación serán todas las empresas que reportaron sus estados financieros. No obstante, al considerar solo las empresas manufacture-

ras, las mismas fueron filtradas de acuerdo a la sección C del código CIIU (Clasificación Industrial Internacional Uniforme), la cual hace referencia al sector industrial manufacturero.

Con este análisis previo se obtuvo un total de 5.679 registros en el año 2016; 6.054 en el 2017; 6.119 para el 2018; 5.203 en el 2019 y, 5.707 registros para el año 2020. Posteriormente, se calcularon los indicadores financieros a partir de todas las cuentas registradas en cada año de estudio. Cabe mencionar que cada empresa informó acerca de los estados y cuentas de resultados en 799 variables, las que permitieron los cálculos de los indicadores financieros; entre los principales se puede mencionar a los indicadores de rentabilidad, liquidez, solvencia y endeudamiento. Incluso, para mostrar el comportamiento financiero de todas las empresas se consideró las variables de total ingresos, total saldos, costos, total pasivos, patrimonio neto, entre otros de todos los estados de resultados.

Toda la información descrita en los anteriores párrafos corresponde a la información disponible de las variables explicativas que provienen de tres distintas fuentes:

Información Interna

Son todas las variables que permiten describir un diagnóstico de la situación real de una empresa. Para este caso se hará uso de los siguientes indicadores financieros.

1. Indicadores de Liquidez.

Los indicadores de liquidez surgen por la necesidad de medir la capacidad que tienen las empresas para cancelar sus obligaciones de corto plazo. Además, permiten establecer la facilidad o dificultad que presenta una compañía para cubrir sus pasivos corrientes al convertir en efectivo sus activos corrientes.

- **Liquidez Corriente:** Este índice relaciona los activos corrientes frente a los pasivos de la misma naturaleza. Cuanto más alto será el coeficiente, entonces la empresa tendrá mayores posibilidades de efectuar sus pagos en el corto plazo.

$$Liquidez\ Corriente = \frac{\text{Activo Corriente}}{\text{Pasivo Corriente}}$$

Se entiende que la liquidez corriente tiene un valor óptimo cuando se encuentra entre 1,50 y 2. Esto quiere decir que: si el resultado es inferior 1,50 entonces una entidad no cuenta con la solvencia adecuada para hacer frente a sus obligaciones financieras al corto plazo, y si es mayor a 2 se

tiene pérdida de rentabilidad por ausencia de inversión de los activos ociosos.

- **Prueba Ácida:** Se conoce también con el nombre de prueba del ácido o liquidez seca. Es un indicador más riguroso, el cual pretende verificar la capacidad de la empresa para cancelar sus obligaciones corrientes, pero sin depender de la venta de sus existencias.

$$\text{Prueba Ácida} = \frac{\text{Activo Corriente} - \text{Inventarios}}{\text{Pasivo Corriente}}$$

El valor óptimo para la prueba ácida es cuando se encuentra cerca de 1, esto indica que una empresa o entidad se encuentra en buena posición financiera para responder a sus obligaciones financieras. Si es menor a 1, existe peligro de suspensión de pagos a terceros por activos corrientes insuficientes; mientras que si es mayor a 1, entonces se tiene exceso de liquidez.

2. Indicadores de Solvencia.

Los indicadores de endeudamiento o solvencia tienen por objetivo medir en qué grado y de qué forma participan los acreedores dentro del financiamiento de la empresa. Se trata de establecer también el riesgo que corran tales acreedores, los dueños de la compañía y los intereses del endeudamiento.

- **Endeudamiento del Activo:** Este índice permite determinar el nivel de independencia financiera. Cuando el índice es elevado (mayor a 1 o 100 %) indica que la empresa depende mucho de sus acreedores, inversionistas, patrocinadores y que dispone de una limitada capacidad de endeudamiento y funciona con una estructura financiera más arriesgada, lo que conlleva a que una compañía no pueda hacer frente a sus obligaciones financieras. Por el contrario, un índice bajo representa un grado elevado de independencia frente a sus acreedores.

$$\text{Endeudamiento del Activo} = \frac{\text{Pasivo Total}}{\text{Activo Total}}$$

- **Endeudamiento Patrimonial:** Este indicador mide el grado de compromiso del patrimonio para con los acreedores de la empresa. No debe entenderse que los pasivos se puedan pagar con patrimonio, puesto que, en

el fondo, ambos constituyen un compromiso para la empresa.

$$\text{Endeudamiento Patrimonial} = \frac{\text{Pasivo Total}}{\text{Patrimonio}}$$

Para este indicador, se estima que el valor óptimo se encuentra entre el 0,4 y 0,6 indicando que el porcentaje total de los activos está siendo financiado por los acreedores del corto y largo plazo; si es mayor a 0,6 entonces se está perdiendo autonomía financiera frente a los terceros; si es menor a 0,4 se tiene exceso de capitales propios, por lo que se recomienda tener una proporción de deudas. Por ejemplo: si el valor del indicador del endeudamiento es del 0,55 indica que los recursos ajenos de la entidad supone un 55 % de los propios, que es lo mismo decir que: por cada 55 unidades monetarias de financiación ajena la empresa cuenta con 100 unidades monetarias de financiación propia.

- **Endeudamiento del Activo fijo:** El coeficiente resultante de esta relación indica la cantidad de unidades monetarias que se tiene de patrimonio por cada unidad invertida en activos fijos.

$$\text{Endeudamiento del Activo Fijo} = \frac{\text{Patrimonio}}{\text{Activo Fijo Neto Tangible}}$$

En otras palabras, cuantos dólares de patrimonio se tiene por cada dólar invertido en activos fijos, cuando el indicador es superior a 1 se dice que la empresa pudo adquirir todos sus equipos a partir del patrimonio, con necesidad de financiamiento externo.

- **Apalancamiento:** Se interpreta como el número de unidades monetarias de activos que se han conseguido por cada unidad monetaria de patrimonio. En otras palabras, determina el grado de apoyo de los recursos internos de la empresa sobre los recursos de terceros.

$$\text{Apalancamiento} = \frac{\text{Activo Total}}{\text{Patrimonio}}$$

El indicador del apalancamiento es superior a 1 indica que la rentabilidad es superior a la que se obtendría si no existiera endeudamiento.

- **Apalancamiento Financiero:** El apalancamiento financiero indica las ventajas o desventajas del endeudamiento con terceros y como este contribuye a la rentabilidad del negocio. Su análisis es fundamental para comprender los efectos de los gastos financieros sobre las utilidades. En consecuencia, a medida que las tasas de interés de una deuda sean más ele-

vadas, será más difícil para las empresas apalancarse financieramente.

$$\text{Apalancamiento Financiero} = \frac{\frac{\text{Utilidad Antes de Impuestos}}{\text{Patrimonio}}}{\frac{\text{Utilidad antes de Impuestos e Intereses}}{\text{Activo Total}}}$$

Cuando el índice es mayor a 1 indica que los fondos propios son superiores a lo que si la empresa no se endeudaría. Cuando el índice es menor o igual a 1, indica que la utilización de fondos ajenos es indiferente desde el punto de vista económico.

3. Indicadores de Gestión

Estos indicadores tienen por objetivo medir la eficiencia con la cual las empresas utilizan sus recursos. De esta forma, medir el nivel de rotación de los componentes del activo; medir también, el grado de recuperación de los créditos y del pago de las obligaciones; además de, medir la eficiencia con la cual una empresa utiliza sus activos según la velocidad de recuperación de los valores aplicados en ellos y medir el impacto de diversos gastos de una empresa en relación con los ingresos generados por ventas.

- **Rotación de Cartera:** Este indicador muestra el número de veces que las cuentas por cobrar giran, en promedio, en un periodo determinado de tiempo, generalmente un año.

$$\text{Rotación de Cartera} = \frac{\text{Ventas}}{\text{Cuentas por Cobrar}}$$

- **Rotación Activo fijo:** El índice de rotación del activo fijo indica la cantidad de unidades monetarias vendidas por cada unidad monetaria invertida en activos inmovilizados. Señalan también una eventual insuficiencia en ventas; por ello, las ventas deben estar en proporción de lo invertido en los equipos de la empresa. Caso contrario, las utilidades se reducirán pues se verían afectadas por la depreciación de los equipos o los activos tangibles.

$$\text{Rotación del Activo fijo} = \frac{\text{Ventas}}{\text{Activo Fijo Neto Tangible}}$$

- **Rotación de Ventas:** La eficiencia en el uso del activo total se mide a través de esta relación que indica el número de veces que, en un determinado nivel de ventas, se utilizan los activos. Este indicador también es

conocido como " coeficiente de eficiencia directiva" puesto que mide la efectividad sobre la administración de una empresa.

$$\text{Rotación Ventas} = \frac{\text{Ventas}}{\text{Activo Total}}$$

Mientras mayor sea el volumen de ventas que se pueda realizar con determinada inversión, más eficiente será la dirección del negocio.

- **Período Medio de Cobranza:** Permite apreciar el grado de liquidez (días) de las cuentas y documentos por cobrar, lo cual se refleja en la gestión y administración de la empresa. Debido a que el índice pretende medir la habilidad de la empresa para recuperar el dinero de sus ventas, para elaborarlo se utilizan las cuentas por cobrar al corto plazo.

$$\text{Período medio de cobranza} = \frac{\text{Cuentas y Documentos por Cobrar} * 365}{\text{Ventas}}$$

4. Indicadores de Rentabilidad.

Los indicadores de rendimiento, llamados también de rentabilidad, sirven para medir la efectividad de la administración de la empresa para controlar los costos y gastos. De esta manera, convertir las ventas en utilidades. Desde el punto de vista del inversionista, lo más importante es hacer uso de estos indicadores para analizar la manera en como se produce el retorno de capital invertido en la empresa.

- **Rentabilidad Neta del Activo:** Esta razón muestra la capacidad del activo para producir utilidades, independientemente de la forma como haya sido financiado, ya sea con deuda o patrimonio.

$$\text{Rentabilidad Neta del activo} = \frac{\text{Utilidad Neta}}{\text{Ventas}} * \frac{\text{Ventas}}{\text{Activo Total}}$$

Para interpretar este indicador se considera el siguiente criterio, por cada unidad monetaria invertida en activos, entonces la empresa obtiene de utilidades netas X unidades monetarias.

- **Margen Bruto:** Este índice permite conocer la rentabilidad de las ventas frente al costo de ventas y la capacidad de la empresa para cubrir los gastos operativos y generar utilidades antes de deducciones e impuestos. En el caso de las empresas industriales, el costo de ventas corresponde al

costo de producción más el de los inventarios de productos terminados.

$$\text{Margen Bruto} = \frac{\text{Ventas} - \text{Costo de Ventas}}{\text{Ventas}}$$

El valor de este indicador puede ser negativo en caso de que el costo de ventas sea mayor a las ventas totales.

- **Margen Operacional:** La utilidad operacional está influenciada no sólo por el costo de las ventas, sino también por los gastos operacionales de administración y ventas. Los gastos financieros, no deben considerarse como gastos operacionales, puesto que teóricamente no son absolutamente necesarios para que la empresa pueda operar. El margen operacional tiene gran importancia dentro del estudio de la rentabilidad de una empresa, puesto que indica si el negocio es o no lucrativo, en sí mismo, independientemente de la forma como ha sido financiado.

$$\text{Margen Operacional} = \frac{\text{Utilidad Operacional}}{\text{Ventas}}$$

- **Margen Neto:** Los índices de rentabilidad de ventas muestran la utilidad de la empresa por cada unidad de venta. Se debe tener especial cuidado al estudiar este indicador, pues comparándolo con el margen operacional, se debe establecer si la utilidad procede principalmente de la operación propia de la empresa, o de ingresos externos.

$$\text{Margen Neto} = \frac{\text{Utilidad Neta}}{\text{Ventas}}$$

- **Rentabilidad Operacional del Patrimonio:** La rentabilidad operacional del patrimonio permite identificar la rentabilidad que le ofrece a los socios o accionistas el capital que han invertido en la empresa, sin tomar en cuenta los gastos financieros ni de impuestos y participación de los trabajadores. Por lo tanto, para su análisis es importante tomar en cuenta la diferencia que existe entre este indicador y el de rentabilidad financiera, para conocer cuál es el impacto de los gastos financieros e impuestos en la rentabilidad de los accionistas.

$$\text{R. Operacional del Patrimonio} = \frac{\text{Utilidad Operacional}}{\text{Patrimonio}}$$

Se podría interpretar que, por cada unidad monetaria de patrimonio aportado por los propietarios, se generan X unidades monetarias de utilidad

operacional.

- **Rentabilidad Financiera:** Cuando un accionista o socio decide mantener la inversión en la empresa, es porque la misma le responde con un rendimiento mayor a las tasas de mercado o indirectamente recibe otro tipo de beneficios que compensan su frágil o menor rentabilidad patrimonial. De esta forma, la rentabilidad financiera se constituye como un indicador sumamente importante, pues mide el beneficio neto generado en relación a la inversión de los propietarios de la empresa. Refleja además, las expectativas de los accionistas o socios, que suelen estar representadas por el denominado costo de oportunidad, el cual indica la rentabilidad que dejan de percibir en lugar de optar por otras alternativas de inversiones de riesgo.

$$\begin{aligned} \text{Rentabilidad Financiera} &= \frac{\text{Venta}}{\text{Activo}} * \frac{\text{UAII}}{\text{Ventas}} * \frac{\text{Activo}}{\text{Patrimonio}} * \frac{\text{UAI}}{\text{UAII}} * \frac{\text{UN}}{\text{UAI}} \\ &= \frac{\text{Utilidad Neta}}{\text{Patrimonio}} \end{aligned}$$

Donde, UAI es la utilidad antes de Impuestos; UAII es la utilidad antes de Impuestos e Intereses. UN es la utilidad neta después del 15% de trabajadores e impuesto a la renta. Este indicador puede interpretarse como: por cada unidad de patrimonio aportado por los socios se generan X unidades monetarias de utilidad neta.

Información Externa

Son las variables que permiten segmentar a todas las empresas según su volumen de ventas, capital social, cantidad de trabajadores, y su nivel de producción o activos presentan. Es decir, se clasifica a las empresas según su tamaño. Existen cuatro tipos de tamaño de empresas.

Cuadro 3.1: Clasificación de las empresas según el tamaño

| Tamaño Empresarial | Volumen de Ingresos | Trabajadores |
|--------------------|-----------------------------|--------------|
| Microempresa | menor a \$100000 | 1 a 9 |
| Pequeña Empresa | entre \$100001 y \$1000000 | 10 a 49 |
| Mediana Empresa | entre \$1000001 y \$5000000 | 50 a 199 |
| Empresa Grande | mayor a \$5000001 | mayor a 200 |

Fuente: Superintendencia de Compañías, Valores y Seguros (SUPERCIAS)

Elaboración: Los autores

Para esta investigación, considerar el tamaño de la empresa para predecir el riesgo de bancarrota es de suma importancia debido a que se está identificando qué segmento es el más riesgoso. No obstante, existe una variedad de investigaciones que utilizan como insumo este parámetro, con el fin de descifrar cuál es la verdadera razón de la quiebra prematura empresarial de muchas compañías ecuatorianas, ya sea esta propia de una empresa o del sistema de la economía en la que se desarrolla.

Como ya se había mencionado, la adecuada categorización de las empresas para definir su tamaño se lo realiza según el volumen de ventas. Sin embargo, hay empresas que durante los primeros años de creación y desarrollo reportan sus estados financieros con utilidades menores a los \$100.000 o en el peor de los casos por debajo de los \$1.000. Estas empresas pertenecen al segmento de las Microempresas, en donde muchas de ellas nacen como respuesta a procesos de crisis financieras o necesidades económicas.

Según datos presentados por la SUPERCIAS para el año 2022, el Ecuador comprende el 96,2 % entre micro, pequeñas y medianas empresas (MiPyme); mientras que el 3,8 % son empresas grandes (Ver cuadro 3.2). Estas últimas tienen facilidades para acceder a operaciones de crédito por parte del sistema financiero nacional o privado, situación que es muy alejada de las MiPyme, ya que estas han tenido dificultades para acceder a créditos. Esto se justifica porque los períodos de vida de una micro o mediana empresa son muy cortos. De hecho, investigaciones realizadas indican que un poco más del 50 % de estas empresas no llegan a cumplir los cinco años de vida empresarial, mostrando incluso elevados riesgos de quiebra antes de sus primeros años [43].

Cuadro 3.2: Proporción de empresas, por tamaño

| País | Microempresa | Pequeña Empresa | Mediana Empresa | Empresa Grande |
|---------|--------------|-----------------|-----------------|----------------|
| Ecuador | 63,5 % | 24,6 % | 8,1 % | 3,8 % |

Fuente: Super Intendencia de Compañías, Valores y Seguros (SUPERCIAS - 2022)

Elaboración: Los autores

Con respecto al servicio financiero que recibe el sector de las MiPyme, los sistemas financieros de la región carecen de alcance. Primero la brecha financiera es significativa. Segundo, a nivel micro, el sistema no sirve para operaciones crediticias para las MiPyme; en el mejor de los casos, más de la mitad de las MiPyme no tienen acceso al sector financiero formal.

Bernanke (1983), mostró como el crédito se vuelve más costoso y difícil de obtener para el sector real en crisis, donde los deudores tienen necesidades financieras o quiebras. Como señaló el autor, los hogares, los agricultores y las pequeñas empresas son las más afectadas [10].

Bernanke & Gertler (1995) mencionan la situación de las pequeñas empresas cuando el canal de crédito se ve afectado. Primero, muestran que las grandes empresas aumentan sus préstamos a corto plazo en momentos en que adquirir el crédito es escaso, porque tienen acceso a los mercados de papeles comerciales¹ y otros financiamientos de corto plazo. Las pequeñas empresas no tienen esta ventaja y responden recortando costos e inventarios, primero para responder a los impactos cuando los préstamos a corto plazo se cierran [11].

Dietsch y Kremp (1998) argumentan que las pequeñas empresas tienen mayor probabilidad de quebrar por insolvencia, ya que según la teoría económica y financiera, la probabilidad de quiebra es alta según el tamaño empresarial, lo que implica que los bancos al otorgales un crédito a pequeñas empresas representa un mayor riesgo para la institución financiera [26].

Berger y Udell (1988) ratifican que el flujo de efectivo de las pequeñas empresas es más volátil y cambiante para afrontar deducos. Es por ello, que los bancos otorgan menos crédito a las pequeñas empresas por ser más riesgosas [9].

Crutzen y Caille (2008) afirman que las pequeñas empresas durante crisis económicas (financieras), tienden a quebrar en mayor medida. Esto porque dependen del financiamiento otorgado por bancos [21].

Para el Ecuador, el estudio de la quiebra empresarial resulta atractivo y conveniente para las micro y pequeñas empresas debido a su escasa visión de negocio para estos segmentos. Por tal motivo, el riesgo de una bancarrota es alto, aunque se hayan identificado alertas que indicaban ya un riesgo empresarial. Otra de las causas de quiebre es el sobreendeudamiento injustificado que adquieren las micro y pequeñas empresas, generando una incapacidad de pago al corto y mediano plazo.

Por lo tanto, uno de los temas sustanciales a ser considerados es precisamente la situación financiera y el riesgo de quiebra de una compañía, razón por la cual existen varios métodos de gestión financiera. Según Delgado (2009) los estados e indicadores financieros, son los mejores instrumentos para medir la gestión financiera y administrativa de las empresas.

¹El papel Comercial es un tipo de obligación, pero de corto plazo. Es un valor emitido como un mecanismo que permite captar recursos del público y financiar sus actividades productivas

Información Geográfica

Son las variables que indican el lugar en donde se encuentra la empresa en estudio, ya sean que se pueden encontrar en la Costa, Sierra o también en la región Amazónica. Con esto se puede determinar si el sector social puede influir en la estimación de bancarrota de las empresas del sector industrial manufacturero.

3.1.1. Variable Dependiente

Para definir la variable dependiente se consideró toda la información proporcionada por la Superintendencia de Compañías que indica cuando una empresa se encuentra activa, cancelada, disuelta o en proceso de liquidación. No obstante, esta información no permite definir un criterio que permita identificar cuales son las razones por las cuáles las empresas quiebran. Además, la información que emite la SUPERCIAS, en su mayoría tiene problemas de temporalidad debido a procesos operacionales, burocráticos y administrativos propios de la entidad. Por ejemplo, cuando una empresa quebró en un año puntual, su disolución y posterior liquidación conlleva un determinado tiempo posterior al evento de quiebra que después es registrado por la entidad competente. Este proceso, en particular, no permite fijar un criterio que permita reconocer cuando una empresa ha entrado en un proceso de insolvencia financiera llevándola a una eventual bancarrota.

Por tal motivo, al no contar con un indicador proporcionado por la Superintendencia que identifique a las empresas quebradas por insolvencia financiera, se propone un indicador que refleje cuando una empresa se encuentra insolvente. Considerando la revisión de la literatura que precede a este trabajo de investigación y por los criterios de bancarrota empresarial definidos por algunos autores, se determinó como principal indicador de riesgo de bancarrota empresarial al indicador del endeudamiento del activo, definiendo el siguiente criterio para la construcción de la variable dependiente:

$$y = \begin{cases} 1 & \text{Empresa quebrada.} \\ 0 & \text{Empresas no quebrada} \end{cases}$$

Se considera a las empresas quebradas cuando el índice del indicador del endeudamiento del activo es mayor a 1 o 100 %. Este argumento se justifica según Brealey, Myers y Marcus (1996), quienes afirman que la insolvencia es mayor cuando mayor es el nivel de endeudamiento de la empresa, porque esta situación disminuye la co-

tización de sus títulos en el mercado; por ende, el valor de la empresa se reduce, lo que implica un mayor riesgo de quiebra empresarial. En conclusión, se define a las empresas quebradas cuando los activos no pueden hacer frente a sus obligaciones y su indicador de endeudamiento es mayor a 1 [13].

3.2. Análisis exploratorio y tratamiento de datos

La información que se hará uso para la realización de este trabajo está conformada por las diferentes fuentes de información que se mencionaron en el apartado anterior, la que comprende información respecto a indicadores financieros, variables sociodemográficas, tamaño empresarial, entre otras, información importante que permitirá una correcta estimación de los modelos propuestos.

A partir del contexto anterior, se procederá a realizar un análisis exploratorio de los datos y así depurar la base de datos para trabajar con la información adecuada y concisa.

3.2.1. Descripción Variables Cuantitativas

En esta sección se muestra la descripción de las variables numéricas; de esta manera se podrá entender el comportamiento estadístico de todas las empresas que se consideran en este estudio. Por ejemplo, la variable Endeudamiento del Activo tiene la siguiente descripción. Ver cuadro (3.3)

Cuadro 3.3: Estadísticos Descriptivos de la Variable Endeudamiento del Activo

| Estadístico | Valor |
|-------------|---------|
| Media | 0.95 |
| Desv_Est | 19.60 |
| Mínimo | 0.00 |
| Perc 25 % | 0.43 |
| Mediana | 0.66 |
| Perc 75 % | 0.86 |
| Máximo | 2347.13 |
| %NA | 0.00 |

Fuente: Superintendencia de compañías, valores y seguros (SUPERCIAS)

Elaboración: Los Autores

Los estadísticos descriptivos para el resto de variables continuas se presentan en el Anexo A.

3.2.2. Análisis de Datos Atípicos

A partir de la base de datos consolidada se realiza un estudio con todas las variables numéricas, analizando toda la información inconsistente o que presente anomalías; de esta manera se pueden identificar casos que contengan información irrelevante para el estudio. Estos casos, pueden tener un efecto desproporcionado en los estudios estadísticos, como la media, lo que implica realizar estimaciones con interpretaciones engañosas.

Un valor atípico es un punto de datos que es significativamente diferente de los datos restantes. Hawkins (1980), define el concepto de un valor atípico de la siguiente manera: "Un valor atípico es una observación que se desvía tanto de las otras observaciones como para despertar sospechas de que fue generado por un mecanismo diferente". Los valores atípicos también se denominan anomalías, variaciones o anomalías en la literatura sobre minería de datos y estadística. Por tal motivo, es importante identificar los datos atípicos antes de modelar y realizar un estudio estadístico.

Considerar el tratamiento de los valores atípicos puede proporcionar información útil sobre los datos. Frecuentemente se los identifica a través de los diagramas de cajas y bigotes que son una representación gráfica de la distribución de los datos numéricos a partir de los *cuartiles* de información.

El método que se utiliza para la detección de los datos atípicos es la representación gráfica del diagrama de caja y bigotes.

Para su realización se presentan los tres cuartiles y los valores mínimo, máximo de los datos sobre un rectángulo vertical u horizontal. Consecuentemente se determina los valores superiores e inferiores a los umbrales máximo y mínimo respectivamente. En donde los umbrales máximo y mínimo están representados de tal manera que toma como base la diferencia entre el primer cuartil (Q_1) y el tercer cuartil (Q_3), llamado rango intercuartílico (IQR). Es así que, en un diagrama de caja se considera un valor atípico aquel que se encuentre 1.5 veces el rango intercuartílico de uno de los cuartiles (Q_1, Q_3), llamado atípico leve; o 3 veces el IQR denominado atípico extremo. Por tanto, se considera un valor atípico leve, q , aquel que verifique cualquiera de las siguientes desigualdades:

$$q < Q_1 - 1.5 * IQR$$

$$q > Q_3 + 1.5 * IQR$$

Donde;

- Q_1 : Primer Cuartil.
- Q_3 : Tercer Cuartil.
- IQR : Rango Intercuartil.
- $Q_1 - 1.5 * IQR$: Valla interna inferior.
- $Q_3 + 1.5 * IQR$: Valla interna superior.

Por otra parte, se considera un valor atípico extremo q , a aquel que verifique las siguientes desigualdades:

$$q < Q_1 - 3 * IQR$$

$$q > Q_3 + 3 * IQR$$

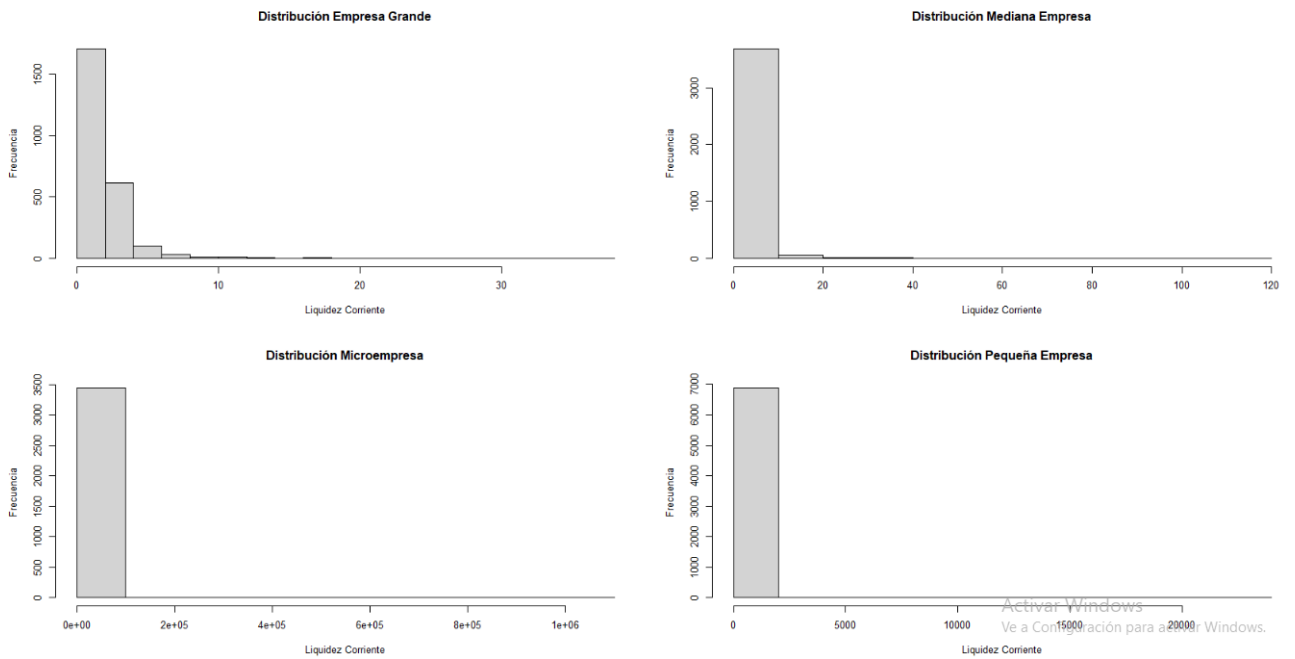
Donde;

- Q_1 : Primer Cuartil.
- Q_3 : Tercer Cuartil.
- IQR : Rango Intercuartil.
- $Q_1 - 3 * IQR$: Valla interna inferior.
- $Q_3 + 3 * IQR$: Valla interna superior.

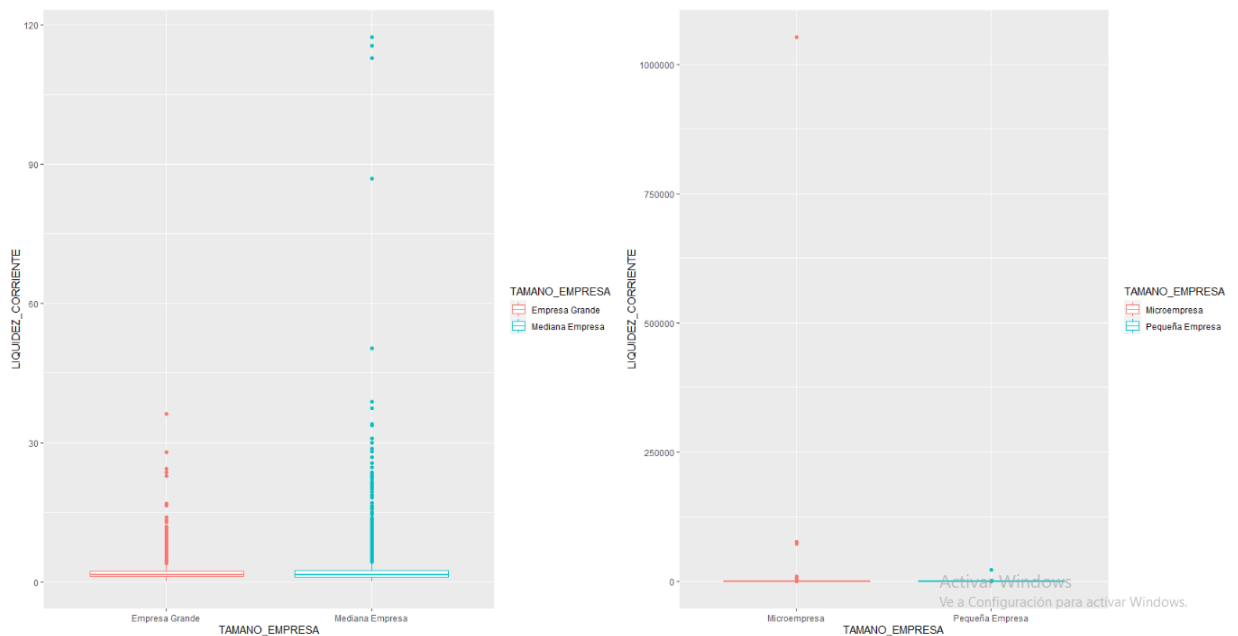
A continuación, se presentan las variables con el mayor número de datos atípicos identificados:

En el indicador de la Liquidez corriente (ver figura 3.1(a)) se puede notar que existe un valor por encima del 100 en sectores de microempresa y pequeña empresa, mientras que en empresas grandes y medianas a pesar de que no existe valores superiores a 100, se encuentran atípicos que sobrepasan el umbral máximo pues un valor óptimo sería entre 1,5 y 2; sin embargo, al tener valores que sobrepasan sus umbrales correspondientes a cada sector de empresas se producen anomalías, lo que conlleva a que el indicador sea tratado por atípicos en relación al tamaño empresarial.

Figura 3.1: Distribución y Diagrama de cajas de la variable *Liquidez Corriente*



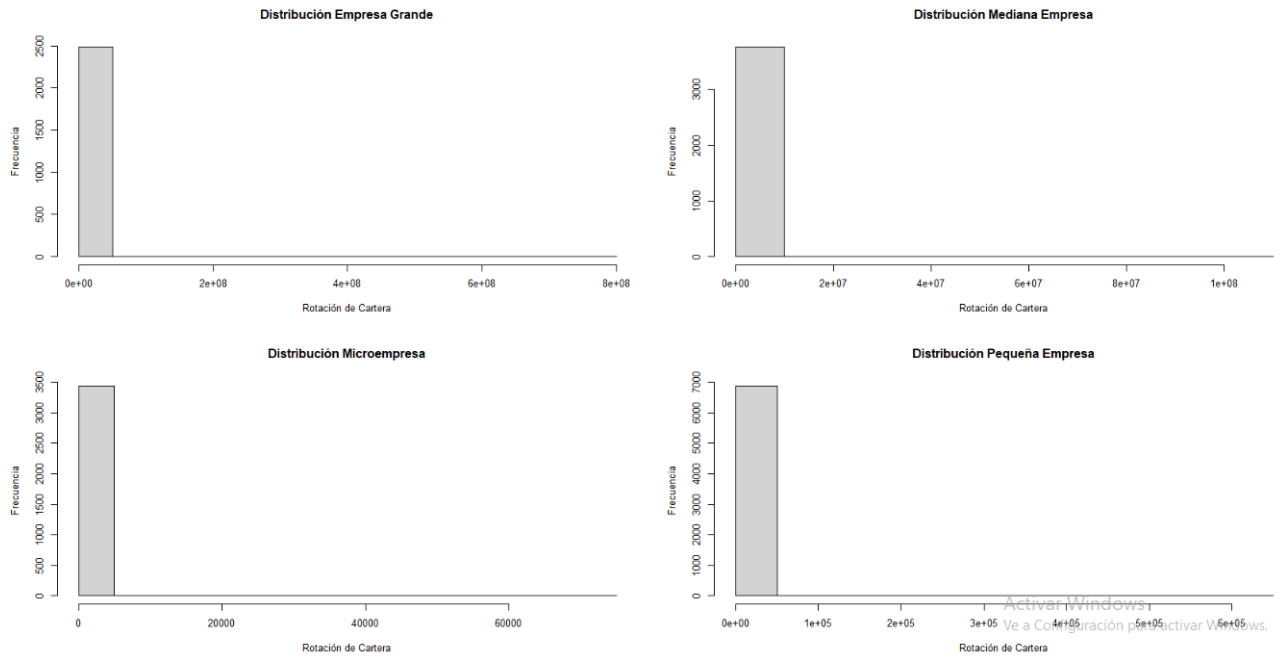
(a) Distribución de la Liquidez Corriente por tamaño empresarial considerando atípicos



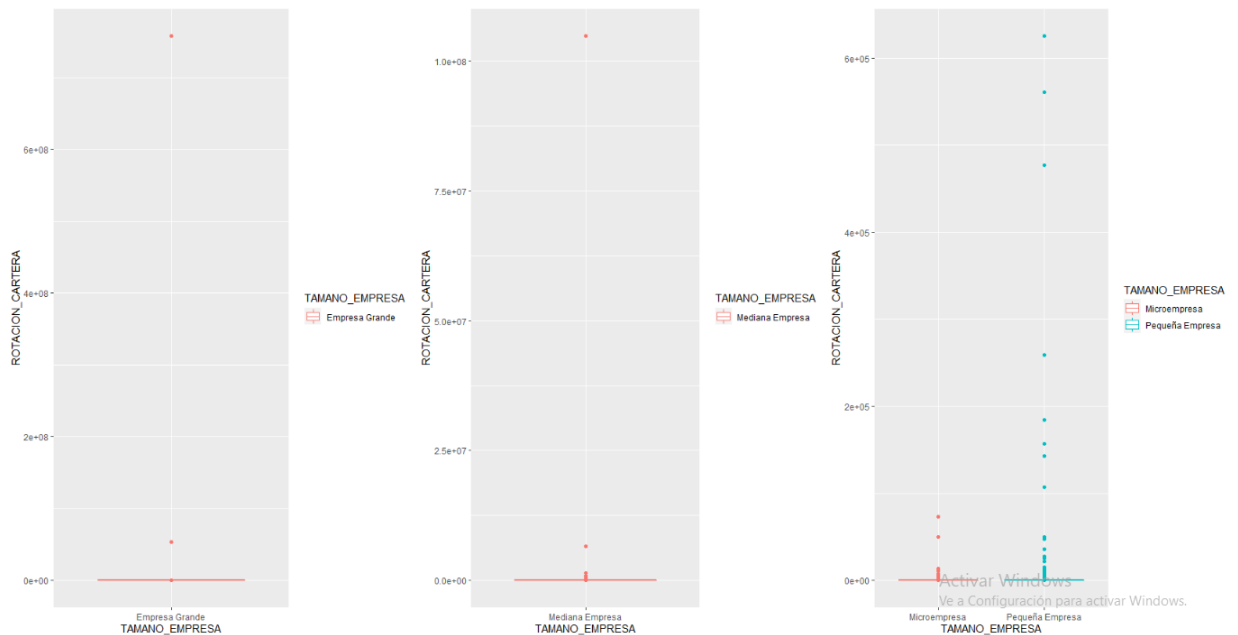
(b) Detección de atípicos en la liquidez corriente por tamaño empresarial

Elaboración: Los Autores

Figura 3.2: Distribución y Diagrama de cajas de la variable *Rotación de Cartera*



(a) Distribución de la Rotación de Cartera por tamaño empresarial considerando atípicos



(b) Detección de atípicos en la Rotación de cartera por tamaño empresarial

Elaboración: Los Autores

Además, se puede ver en la figura 3.1(b) que la distribución para micro y pequeña empresa no pueden ser definidas de manera correcta debido a los valores atípicos que la alteran.

Otra de las variables con mayores valores atípicos corresponde al indicador de *Rotación de Cartera*. Similar al indicador de liquidez corriente, en la figura 3.2(a) se puede ver que tiene valores superiores al umbral máximo correspondiente a 86 (hallado a través del cálculo del *IQR*). Su distribución es alterada justamente por la detección de estos datos atípicos (ver la figura 3.2(b)).

En la siguiente sección se determina el tratamiento de las variables que tengan valores anómalos.

3.2.3. Tratamiento de Datos Atípicos

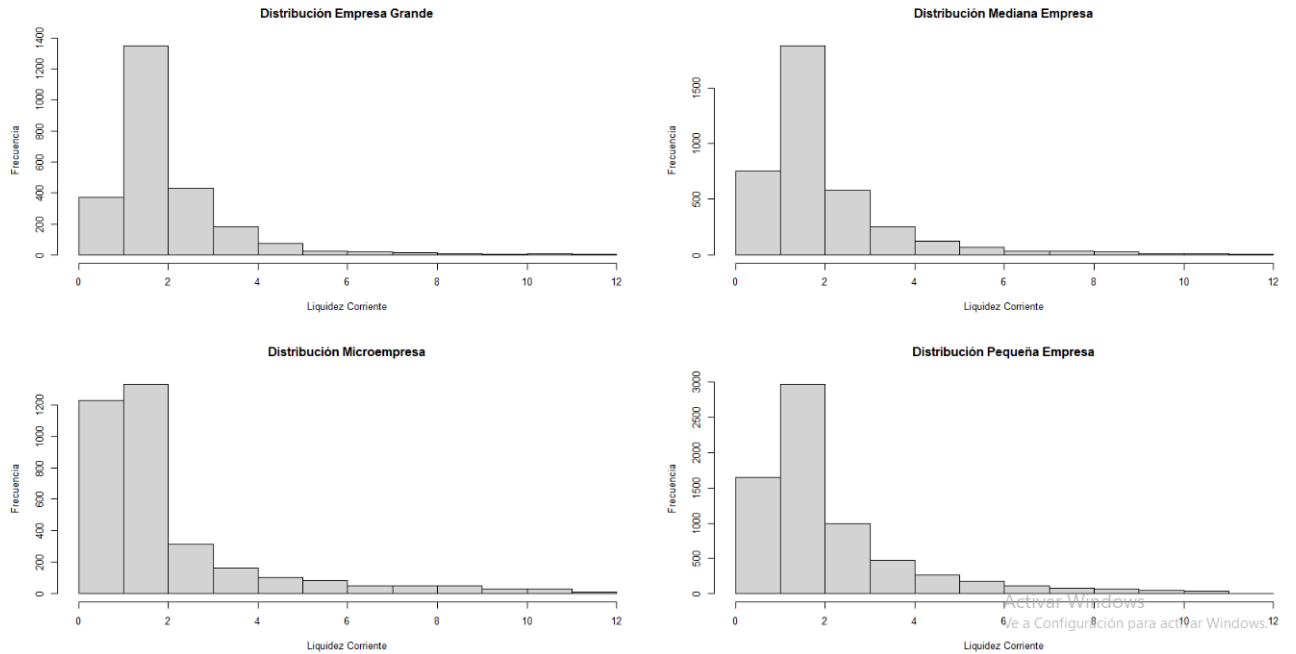
Una vez identificados los datos atípicos de las variables numéricas, se procede a tratar los datos de tal manera que no se imputen de la base a menos que sea necesario o con justificación razonable. Para esto, se ha hecho uso de la estadística clásica que usa la media, la varianza o la mediana; particularmente se procede a tomar los valores atípicos como la mediana, debido a que esta medida es menos susceptible a las otras dos.

En la imagen 3.3 se muestra la variable tratada por tamaño de empresa y conjuntamente la distribución, en ella se puede observar más claramente cómo están distribuidos los datos del indicador de liquidez para cada sector de empresa (grande, mediana, micro y pequeña). A pesar de haber tratado a los valores atípicos, aún quedan valores por encima del umbral máximo; sin embargo, dichos valores pueden ser considerados reales, es decir, existen empresas que tienen valores entre 3 y 12 como indicador de liquidez indicando que tienen sus activos corrientes ociosos lo que produciría una pérdida de rentabilidad para ellas.

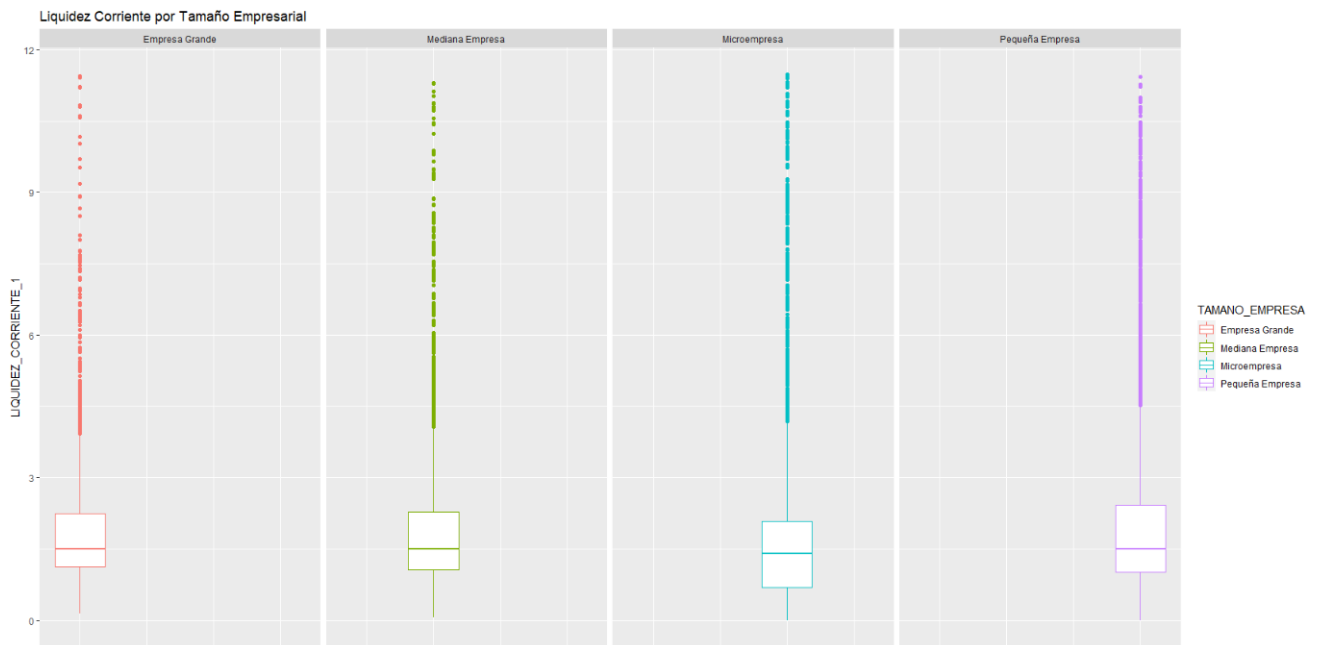
Así mismo, se ha tratado a la variable *Rotación de Cartera* (ver imagen 3.4). Se observa de manera más clara la distribución de los datos, mostrándose como una distribución asimétrica positiva. Además, cabe señalar que aún existen datos por sobre el valor del umbral máximo, pero no se puede asegurar con certeza si dichos valores son erróneos o no; sin embargo, no existe un número de días exacto como tal que sea ideal, básicamente entre menos días la empresa tarde en recuperar la cartera de clientes es mejor, pero para un análisis más riguroso este indicador debería analizarse con la rotación de cuentas por pagar para de esa manera poder validar si

la empresa tiene o no liquidez.

Figura 3.3: Distribución y Diagrama de cajas de la variable *liquidez corriente* tratada



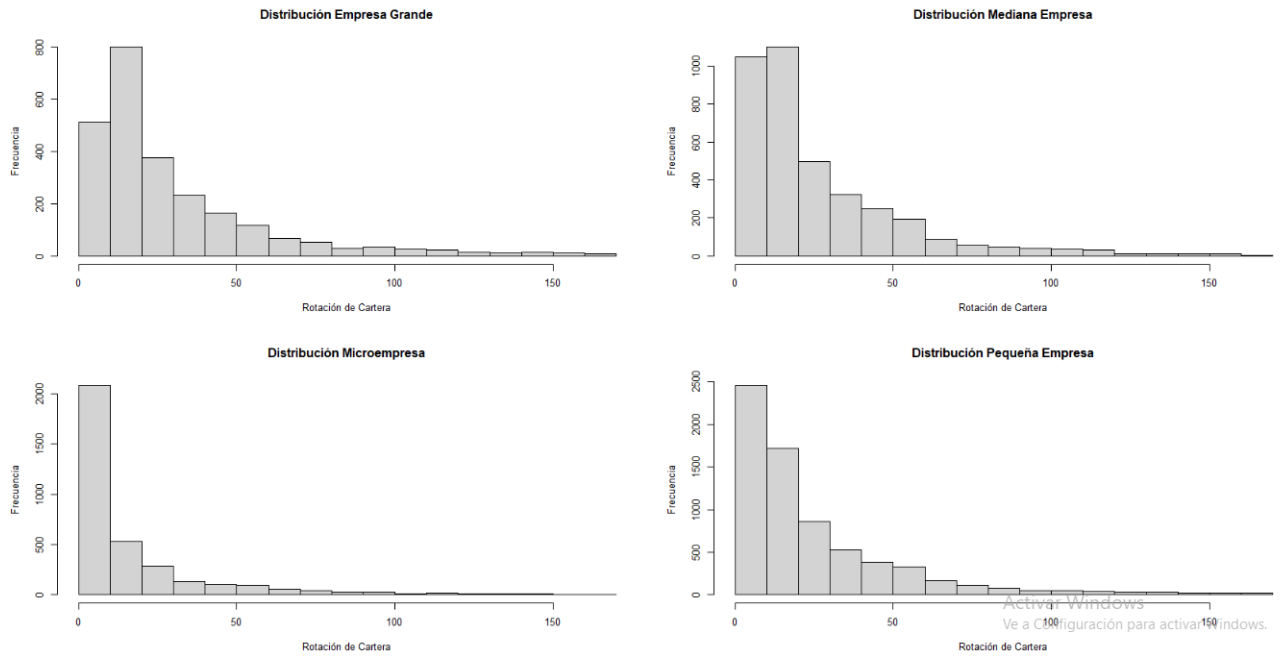
(a) Distribución de la liquidez corriente por tamaño empresarial con tratamiento sobre atípicos



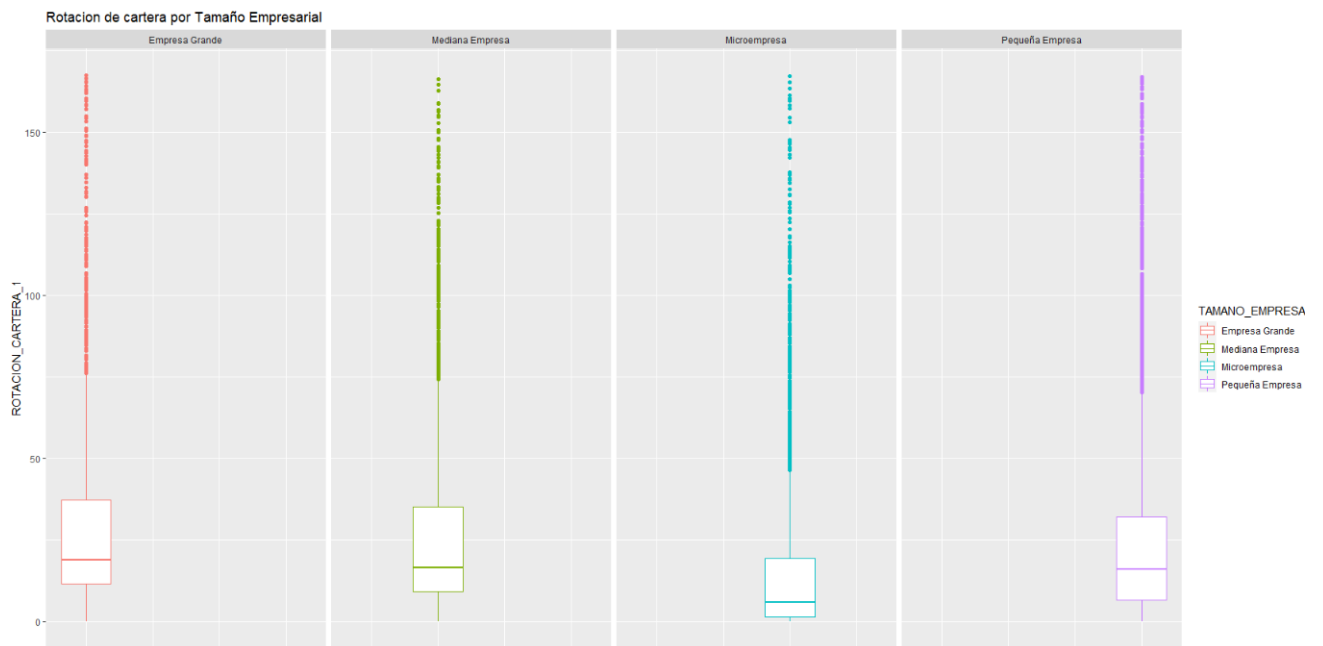
(b) Detección de atípicos de la liquidez corriente por tamaño empresarial con tratamiento de atípicos

Elaboración: Los Autores

Figura 3.4: Distribución y Diagrama de cajas de la variable *Rotación de Cartera* tratada



(a) Distribución de la Rotación de Cartera por tamaño empresarial con tratamiento sobre atípicos



(b) Detección de atípicos de la Rotación de Cartera por tamaño empresarial con tratamiento de atípicos

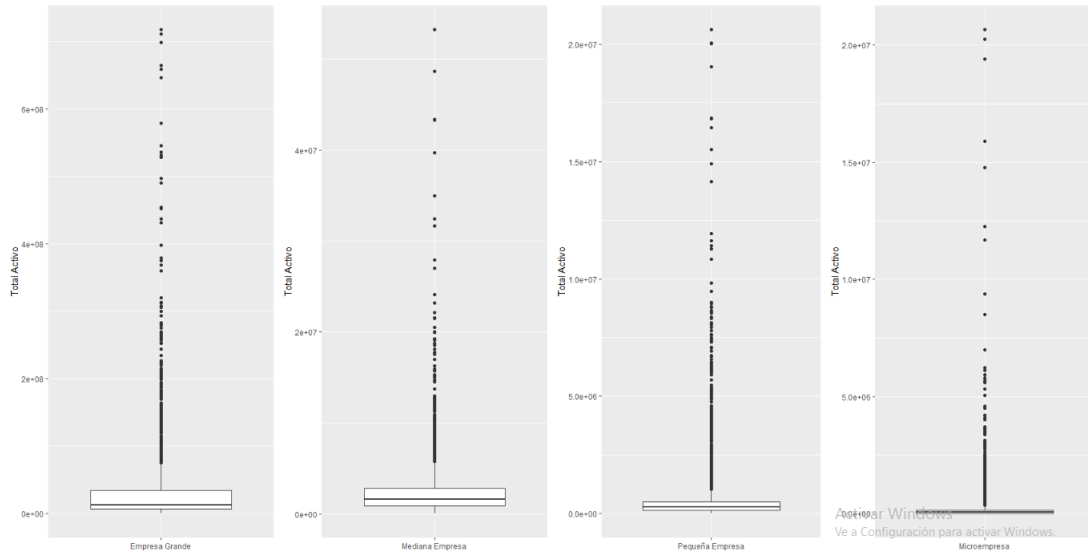
Elaboración: Los Autores

El resto de las variables que presentan atípicos son tratadas y se pueden visualizar en el Anexo B.

3.2.4. Depuración de la base de datos

A partir del primer filtro aplicado sobre los datos, se decidió realizar un nuevo estudio adicional para trabajar con información que beneficie en el cálculo de la probabilidad de bancarrota, pues se identificaron valores inusuales para algunas variables.

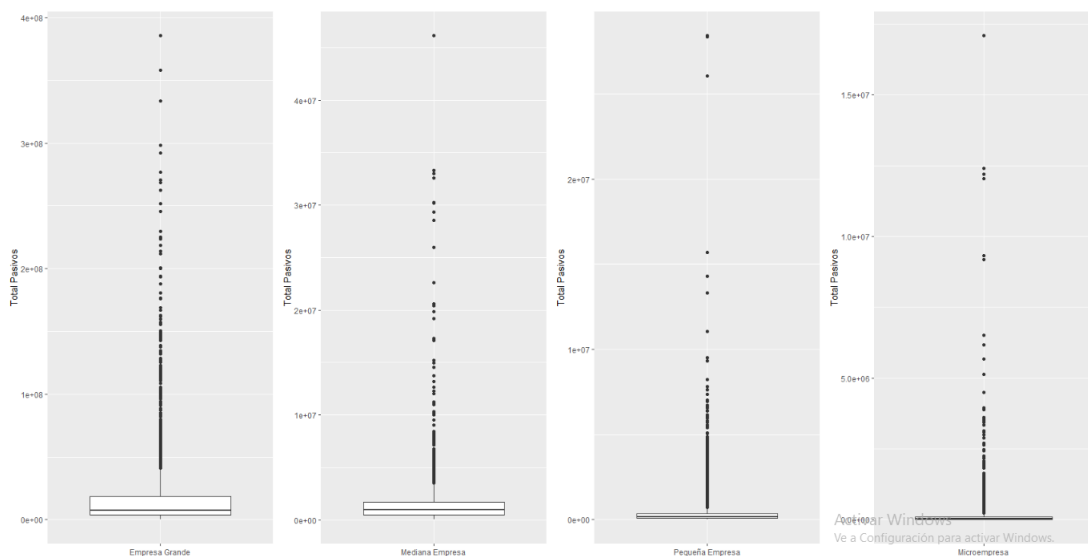
Figura 3.5: Diagrama de cajas de las variables *Total Activo* por tamaño Empresarial



Fuente: Superintendencia de Compañías, Valores y Seguros (SUPERCIAS)

Elaboración: Los Autores

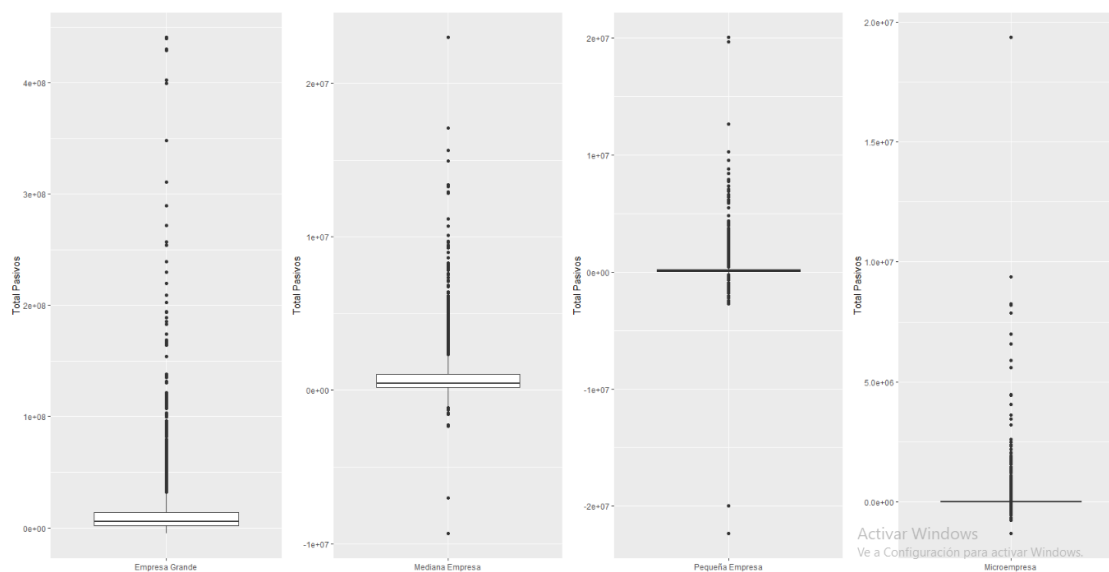
Figura 3.6: Diagrama de cajas de las variables *Total Pasivo* por tamaño Empresarial



Fuente: Superintendencia de Compañías, Valores y Seguros (SUPERCIAS)

Elaboración: Los Autores

Figura 3.7: Diagrama de cajas de las variables *Total Patrimonio* por tamaño Empresarial



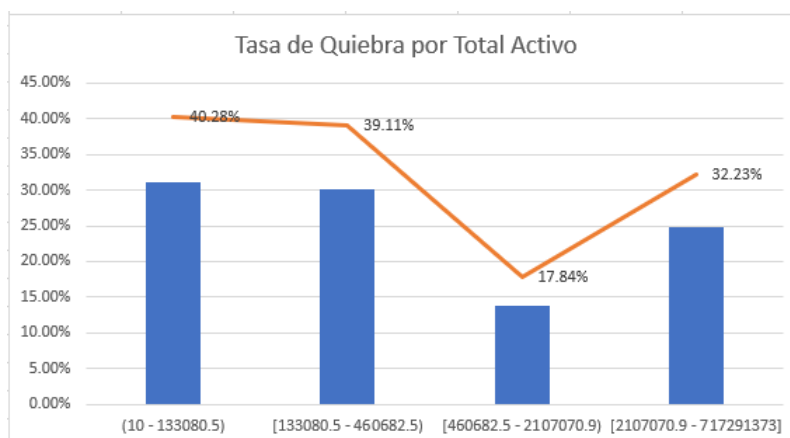
Fuente: Superintendencia de Compañías, Valores y Seguros (SUPERCIAS)

Elaboración: Los Autores

En las figuras 3.5 y 3.6 se puede ver que el activo y pasivo total exceden los 400 millones, lo que indica que las empresas con estos activos y pasivos poseen un patrimonio superior a los 25 millones; valores de capital que corresponden a las grandes empresas. Estos resultados muestran la situación real del rendimiento financiero que poseen, sin embargo no se puede obtener un diagnóstico certero para predecir si una empresa podría entrar en un proceso de insolvencia financiera. Para medir este comportamiento se procede a encontrar la relación que tienen las empresas entre sus activos, pasivos, patrimonio y la probabilidad de entrar o no en un proceso de quiebra empresarial.

Con base en lo anterior mencionado, se procedió a discretizar las variables en función del riesgo de quiebra que presentan según los activos, pasivos totales y patrimonio neto que registran las empresas según su tamaño. Para ello, se procedió a seguir el siguiente análisis:

Figura 3.8: Tasa de quiebra según el total de activos

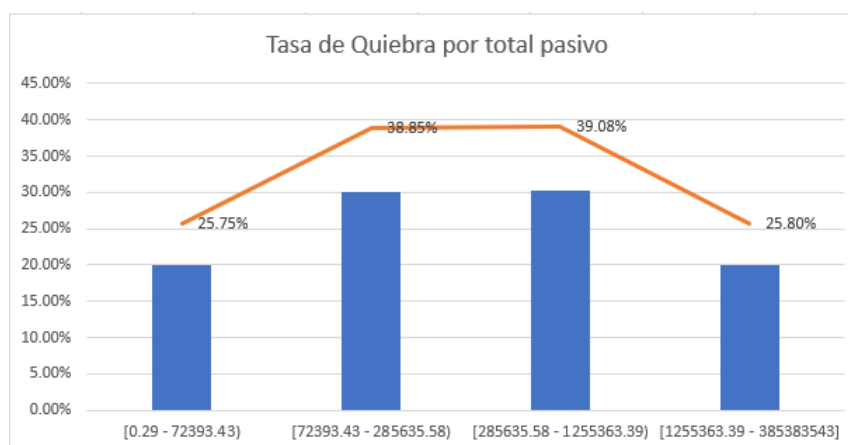


Fuente: Superintendencia de Compañías, Valores y Seguros (SUPERCIAS)

Elaboración: Los Autores

En la figura 3.8 se puede ver que las empresas que tienen activos entre 10 y 133k poseen una tasa de caer en bancarrota del 40 %; mientras que las empresas que tienen activos valorados entre 450k y 2 M tienen una menor probabilidad de caer en quiebra. Para resumir, las empresas que poseen menor valor en todos sus activos tienen una mayor probabilidad de quiebra empresarial.

Figura 3.9: Tasa de quiebra según el total de pasivos

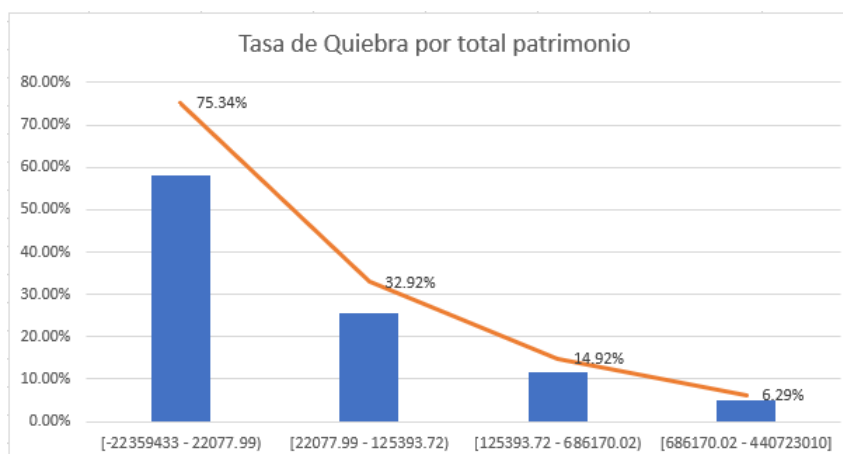


Fuente: Superintendencia de Compañías, Valores y Seguros (SUPERCIAS)

Elaboración: Los Autores

Por otro lado, en la figura 3.9 se puede ver que las empresas con un total de pasivos entre 72k y 1M tienen una mayor probabilidad de tener una quiebra empresarial a comparación con los que tienen un total de pasivos por debajo de 72K y por encima de 1M.

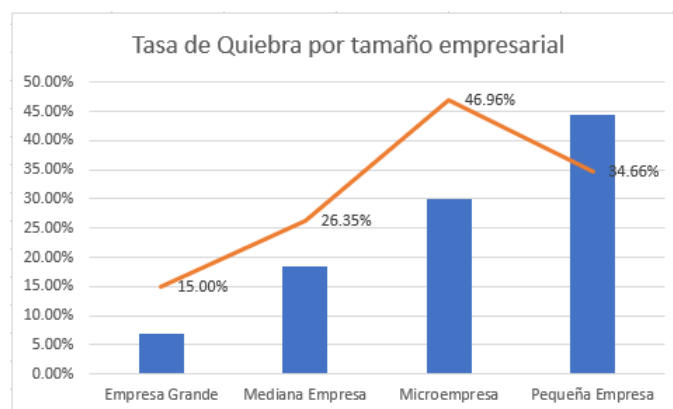
Figura 3.10: Tasa de quiebra según el total de patrimonio



Fuente: Superintendencia de Compañías, Valores y Seguros (SUPERCIAS)
Elaboración: Los Autores

Respecto al total del patrimonio, es notorio que las empresas entre más patrimonio poseen su tasa de quiebra es menor respecto a las empresas que tienen un patrimonio por debajo de los 22K.

Figura 3.11: Tasa de quiebra según el tamaño empresarial



Fuente: Superintendencia de Compañías, Valores y Seguros (SUPERCIAS)
Elaboración: Los Autores

En la figura 3.11 se puede evidenciar lo descrito en la sección anterior, en donde se evidencia que las micro y pequeñas empresas poseen una mayor tasa de caer en quiebra que las grandes y medianas empresas. Esto se debe al poco acceso a fuentes de financiación para sus respectivas operaciones internas.

La tasa de quiebra empresarial respecto al resto de variables se encuentra en el Apéndice C. A partir de estos resultados se discretizó a las variables y así aportar un mejor *performace* al rendimiento de los modelos propuestos.

3.2.5. Selección de la Muestra para Entrenamiento y Validación

Después de completar un detallado y exhaustivo proceso en el tratamiento de los datos, el conjunto de datos a considerar para el cálculo de la probabilidad de quiebra quedó consolidado por 16.591 casos, los cuales tienen la siguiente distribución temporal:

Cuadro 3.4: Distribución número de empresas por año

| Año | N° Empresas |
|------|-------------|
| 2016 | 4142 |
| 2017 | 3361 |
| 2018 | 3552 |
| 2019 | 2850 |
| 2020 | 3016 |

Fuente: Superintendencia de Compañías, Valores y Seguros (SUPERCIAS)

Elaboración: Los Autores

Para la aplicación de las metodologías que se proponen dentro de este estudio se debe considerar dos muestras de datos, entrenamiento y validación. Según Siddiqi (2006) un conjunto de datos debe dividirse en dos submuestras: Muestra de entrenamiento (Permite entrenar los modelos) y una Muestra de validación (Permite validar el comportamiento y el funcionamiento de los modelos) [61].

Existen diferentes proporciones para definir un conjunto de entrenamiento y otro para validación. Para este trabajo en particular, se consideró el 70 % para la muestra de entrenamiento y el 30 % restante para validación. De esta manera se obtiene que ambas muestras sean representativas. A continuación, se indica la distribución de empresas identificadas en quiebra y no quebradas, tanto en la muestra de entrenamiento y validación.

Cuadro 3.5: Distribución de la variable dependiente para entrenamiento y validación

| Y | Descripción | Entrenamiento | | Validación | |
|---|-------------|---------------|------------|------------|------------|
| | | Empresas | Porcentaje | Empresas | Porcentaje |
| 0 | Bueno | 7854 | 67.63 % | 3367 | 67,63 % |
| 1 | Malo | 3759 | 32.36 % | 1611 | 32.36 % |
| | Total | 11613 | 100 % | 4978 | 100 % |

Elaboración: Los Autores

3.2.6. Selección de Variables

Como se había mencionado en secciones anteriores, se aplicarán técnicas que permitan identificar a las variables potenciales con mayor poder predictivo que influyan fuertemente sobre la variable dependiente Y . Por tal motivo, a continuación se presenta la prueba de Kolmogorov-Smirnov (KS) aplicada sobre las variables cuantitativas ordenadas de mayor a menor.

Cuadro 3.6: Test KS

| Variable | KS |
|-------------------------------------|--------|
| Apalancamiento | 0,8659 |
| Apalancamiento Financiero | 0,8554 |
| Endeudamiento Patrimonial | 0,6203 |
| Liquidez Corriente | 0,4277 |
| Prueba Ácida | 0,3341 |
| Rentabilidad Operacional Patrimonio | 0,3201 |
| Margen Operacional | 0,2786 |
| Margen Neto | 0,2708 |
| Rentabilidad Neta del Activo | 0,2707 |
| Rentabilidad Financiera | 0,2336 |
| Margen Bruto | 0,1107 |
| Rotación Cartera | 0,1028 |
| Rotación Ventas | 0,0919 |
| Periodo Medio del Cobranza | 0,0845 |
| Permanencia | 0,0801 |
| Rotación Activo del Fijo | 0,0703 |

Elaboración: Los Autores

Del cuadro 3.6 se puede observar la variable explicativa que más influye sobre el comportamiento de la variable dependiente es el *Apalancamiento*, el cual indica si una empresa puede o no endeudarse para aumentar los beneficios de una inversión; señalando que una empresa puede recurrir a una financiación externa cuando este indicador registre un valor de mayor a 1. Otra variable que también influye sobre la variable dependiente es el *Endeudamiento Patrimonial*, que indica cuanto del patrimonio de la entidad destina para con los acreedores de la empresa, en donde este valor debe estar comprometido con los acreedores desde el 40 % hasta el 60 %. Seguido se tiene que la *liquidez corriente* influye sobre la variable dependiente cuando alcanza valores entre 1,5 a 2. Este rango muestra que una empresa puede hacer frente a sus obligaciones financieras mediante la rentabilidad generada a través de los activos. Al igual que la liquidez corriente, la *prueba ácida* también es una variable que influye fuertemente en el comportamiento de la variable dependiente cuando la misma tiene valores cercanos a uno, esto indica que una empresa puede responder a sus obligaciones financieras a partir de sus ingresos generados. Por último, el índice del *Apalancamiento Financiero* también influye sobre la variable objetivo cuando considera valores superiores a uno, esto implica que los fondos propios son superiores a lo que si una empresa busca financiamiento externo.

En el cuadro 3.6 se pueden ver las variables potenciales que permitirán estimar un modelo con un mayor predictivo y con un error mínimo de precisión. Por otro lado, se aplica el criterio del valor de información (VI) para determinar cuáles variables cualitativas tienen mayor efecto sobre la variable dependiente.

Cuadro 3.7: Valor de Información (VI)

| Variable | VI |
|----------------------------|-----------|
| Categoría Total Patrimonio | 1.662 |
| Tamaño Empresarial | 0.2225 |
| Categoría Total Ingresos | 0.2161 |
| Categoría Total Activo | 0.1766 |
| Categoría Total Costos | 0.1555 |
| Categoría Total Pasivos | 0.0813 |
| PROVINCIA | 0.0278 |
| REGION | 0.0081 |

Elaboración: Los Autores

Una vez identificadas las variables que podrán explicar y predecir la bancarrota empresarial, se procede con la elaboración de los modelos propuestos para el desarrollo de esta investigación.

3.3. Estimación del Modelo Aditivo Generalizado

Después de haber identificado y seleccionado las variables más representativas que influyen en el comportamiento de la variable dependiente se procede a construir el modelo aditivo generalizado que toma como insumo las variables seleccionadas.

Para esta metodología estadística se consideran las variables cuantitativas a las cuales se les aplicará las funciones de suavizamiento tipo *splines cúbicos* de regresión, permitiendo así capturar la relación no lineal entre la variable dependiente y las variables explicativas.

Al tratar de encontrar un modelo adecuado que permita estimar la probabilidad de bancarrota y así predecir si una empresa entra o no en un quiebra empresarial, se tiene una restricción particular la cual consiste en determinar el número de grados de suavizamiento (nodos de la función) que se aplica a una variable.

Bajo este contexto, se elaboró una lista de todas las variables que obtuvieron un valor de KS significativo. Esta lista esta conformada de la siguiente manera: se considera cada variable con su estructura original y bajo una función suavizadora, en este caso los *splines cúbicos* y, por último, considera un valor aleatorio para el parámetro de suavizado.

Después se aplica la función *step.gam* de la librería *mgcv*² la cual pertenece al software estadístico R. Esta función utiliza el algoritmo *backward* el cual permite encontrar el modelo con mejor desempeño en n iteraciones. Este último podrá ser medido a través del criterio de Akaike, en donde las variables consideradas como menos influyentes para el modelo serán eliminadas en cada iteración hasta que no existan variables por eliminar.

En la tabla (3.8) se presenta el primer modelo resultante luego de haber aplicado la metodología explicada en el párrafo anterior.

²Es una librería creada para modelos GAM en la que se corrige la incertidumbre del parámetro de suavizado.

Cuadro 3.8: Modelo de estimación GAM-1 (Primer Modelo)

| Parametric coefficients: | | | | |
|--------------------------|-----------|------------|---------|----------|
| | Estimate | Std. Error | z value | Pr(> z) |
| (Intercept) | 4.596e+04 | 1.181e+05 | 0.389 | 0.697 |
| ROTACION_VENTAS | 5.551e+0 | 1.689e+01 | 0.329 | 0.742 |

| Approximate significance of smooth terms: | | | | |
|---|-------|--------|--------|---------|
| | edf | Ref.df | Chi.sq | p-value |
| s(LIQUIDEZ_CORRIENTE) | 1.046 | 1.068 | 0.132 | 0.770 |
| s(ENDEUDAMIENTO_PATRIMONIAL) | 2.381 | 2.739 | 1.753 | 0.485 |
| s(APALANCAMIENTO) | 2.932 | 2.974 | 3.024 | 0.382 |
| s(PRUEBA_ACIDA) | 1.027 | 1.039 | 0.121 | 0.730 |
| s(ROTACION_CARTERA) | 1.000 | 1.000 | 0.009 | 0.923 |
| s(MARGEN_OPERACIONAL) | 1.000 | 1.000 | 0.006 | 0.937 |
| s(RENTABILIDAD_OPERACIONAL_PATRIMONIO) | 1.000 | 1.000 | 0.132 | 0.716 |
| s(RENTABILIDAD_FINANCIERA) | 1.000 | 1.001 | 0.009 | 0.924 |
| s(RENTABILIDAD_NETA_ACTIVO) | 1.000 | 1.000 | 0.133 | 0.716 |

| | | | | |
|--------------|---------|----------------------|-------|-----------|
| --- | | | | |
| R-sq.(adj) = | 0,22 | Deviance explained = | 24,6% | |
| -REML = | -75.838 | Scale est. = | 1 | n = 11613 |

Elaboración: Los Autores

En la tabla 3.8 se puede ver que ninguna variable del modelo aditivo generalizado es significativa, incluso cuando las variables han sido suavizadas. Por tal motivo se vuelve a ejecutar la función *step.gam* pero, en esta ocasión se utiliza las variables cualitativas, para de esta manera poder apreciar el comportamiento de las empresas según su total de ingresos, total pasivos y total patrimonio según el tamaño empresarial. A continuación, se presenta un nuevo modelo con la anterior observación:

Como se puede ver en la tabla 3.9 todas las variables son estadísticamente significativas. A partir de esto, se procede con la elaboración de un nuevo modelo, presentando así al modelo GAM-3 (cuadro 3.10).

De similar manera, se puede ver en la cuadro (3.10) que todas las variables son estadísticamente significativas, por lo que corresponden a un modelo adecuado; además, los datos explicados corresponden a más del 74.2 %

Cuadro 3.9: Modelo de estimación GAM-2 (Segundo Modelo)

| Parametric coefficients: | | | | | |
|---|----------|------------|---------|------------|-----|
| | Estimate | Std. Error | z value | Pr(> z) | |
| (Intercept) | 24.549 | 1.380 | 17.789 | < 2e-16 | *** |
| LIQUIDEZ_CORRIENTE | -0.428 | 0.029 | -14.467 | < 2e-16 | *** |
| RENTABILIDAD | 2.906 | 0.161 | 18.006 | < 2e-16 | *** |
| _OPERACIONAL_PATRIMONIO | | | | | |
| RENTABILIDAD_FINANCIERA | 3.832 | 0.294 | 13.031 | < 2e-16 | *** |
| RENTABILIDAD_NETA_ACTIVIVO | -17.553 | 1.348 | -13.018 | < 2e-16 | *** |
| TAMANOEMPRESA | 0.856 | 0.185 | 4.612 | < 3.99e-06 | *** |
| MedianaEmpresa | | | | | |
| TAMANOEMPRESAMicroempresa | 1.571 | 0.180 | 8.693 | < 2e-16 | *** |
| TAMANOEMPRESA | 1.151 | 0.173 | 6.657 | < 2.79e-11 | *** |
| PequenaEmpresa | | | | | |
| Approximate significance of smooth terms: | | | | | |
| | edf | Ref.df | Chi.sq | p-value | |
| s(ENDEUDAMIENTO_PATRIMONIAL) | 2.994 | 3.000 | 897.35 | <2e-16 | *** |
| s(MARGEN_OPERACIONAL) | 3.786 | 4.319 | 128.96 | <2e-16 | *** |
| s(ROTACION_VENTAS) | 1.002 | 1.004 | 0.477 | 0.491 | |
| R-sq.(adj) = 0.751 Deviance explained = 69.7% | | | | | |
| -REML = 2240.5 Scale est. = 1 n = 11613 | | | | | |

Elaboración: Los Autores

Cuadro 3.10: Modelo de estimación GAM-3 (Tercer Modelo)

| Parametric coefficients: | | | | | |
|---|----------|------------|---------|----------|-----|
| | Estimate | Std. Error | z value | Pr(> z) | |
| (Intercept) | 24.949 | 1.391 | 17.926 | < 2e-16 | *** |
| LIQUIDEZ_CORRIENTE | -0.429 | 0.029 | -14.432 | < 2e-16 | *** |
| RENTABILIDAD | 2.885 | 0.161 | 17.827 | < 2e-16 | *** |
| _OPERACIONAL_PATRIMONIO | | | | | |
| RENTABILIDAD_FINANCIERA | 3.800 | 0.293 | 12.935 | < 2e-16 | *** |
| RENTABILIDAD_NETA_ACTIVIVO | -17.601 | 1.344 | -13.093 | 6.74e-10 | *** |
| PERMANENCIA | -0.115 | 0.027 | -4.187 | 2.82e-05 | *** |
| TAMANOEMPRESA | 0.842 | 0.186 | 4.518 | 6.23e-06 | *** |
| MedianaEmpresa | | | | | |
| TAMANOEMPRESA | 1.513 | 0.181 | 8.322 | < 2e-16 | *** |
| Microempresa | | | | | |
| TAMANOEMPRESA | 1.128 | 0.173 | 6.496 | 8.24e-11 | *** |
| PequenaEmpresa | | | | | |
| Approximate significance of smooth terms: | | | | | |
| | edf | Ref.df | Chi.sq | p-value | |
| s(ENDEUDAMIENTO_PATRIMONIAL) | 2.994 | 3.000 | 897.23 | <2e-16 | *** |
| s(MARGEN_OPERACIONAL) | 3.789 | 4.322 | 129.28 | <2e-16 | *** |
| s(ROTACION_VENTAS) | 1.002 | 1.004 | 0.69 | 0.0407 | . |
| R-sq.(adj) = 0.798 Deviance explained = 74.2% | | | | | |
| -REML = 2234 Scale est. = 1 n = 11613 | | | | | |

A partir de los dos último modelos (GAM-2) y (GAM-3) se obtienen los siguientes resultados comparativos. (Ver cuadro 3.11)

Cuadro 3.11: Modelo con mejor rendimiento estadístico

| Modelo | Deviance Explained | AIC |
|---------------|---------------------------|------------|
| GAM-2 | 69,7 % | 4466,21 |
| GAM-3 | 74,2 % | 4449,86 |

Elaboración: Los Autores

Como se puede ver el modelo que mejor se adapta al comportamiento de los datos es el GAM-3; por lo tanto será considerado para el estudio de esta investigación y se procede a dar su interpretación dentro del estudio.

Es así, que se puede observar que todos los coeficientes son significativos al 95 % y la interpretación de los coeficientes se la puede realizar de manera similar a la regresión lineal; sin embargo, hay que considerar que este modelo trabaja con la función de logit y depende de una probabilidad de que ocurra un suceso. Así se puede obtener un adecuado valor de probabilidad de riesgo de quiebra.

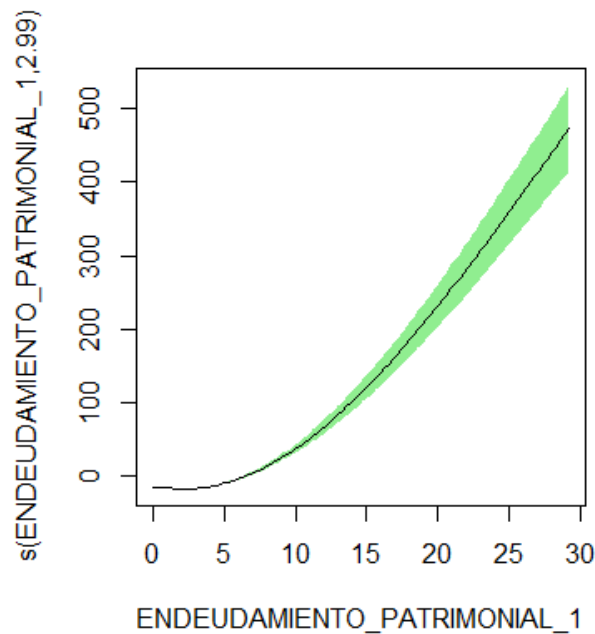
Se presenta la explicación y el comportamiento de las variables seleccionadas en el modelo.

- En primer lugar se puede ver que el signo asociado al coeficiente de la variable liquidez corriente es negativo, lo que indica que las empresas con un mayor valor en pasivo son menos propensas a la quiebra empresarial pues indican buenos niveles de liquidez. Con respecto a la rentabilidad del patrimonio, la posibilidad de bancarrota es menor debido a que a mayor patrimonio mejor rentabilidad.
- El signo del coeficiente de la variable rentabilidad financiera indica que el riesgo de bancarrota para empresas con un menor valor de patrimonio son propensas a la insolvencia financiera. Mientras que la rentabilidad neta del activo indica que el riesgo de quiebra tienden a disminuir cuando existe una mayor utilidad generada por el total de activos.
- Por otro lado, a medida que las empresas son recurrentes en sus actividades de desarrollo económico a lo largo del período de estudio son menos propensas al riesgo de quiebra debido a su estabilidad en los años. De manera similar se puede ver que, si las empresas corresponden al segmento de medianas empre-

sas, su probabilidad de caer en riesgo es baja a comparación con las micro y pequeñas empresas cuya probabilidad de riesgo es mayor.

- Para visualizar el impacto que tiene un modelo GAM se presentan a continuación las variables cuantitativas suavizadas con la técnica de splines cúbico de regresión.

Figura 3.12: Predicción Parcial para la variable Endeudamiento Patrimonial

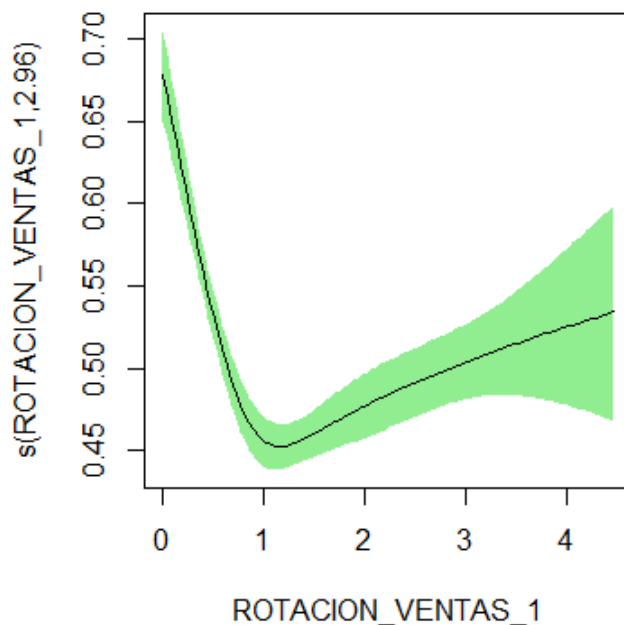


Elaboración: Los Autores

En la figura 3.12 se puede ver que la aplicación de la función de suavizamiento ajusta de manera correcta el comportamiento no lineal de la relación entre la variable riesgo de quiebra y el endeudamiento patrimonial. Por otro lado, la variable toma un comportamiento cúbico, dado que los datos se abren ligeramente para tomar todos los casos, en otras palabras, las posibilidades de que una empresa llegue a quebrar en el rango dentro del endeudamiento patrimonial de 5 a 30, aumentan debido que se está perdiendo autonomía financiera frente a terceros. Además, entre mayor sea el endeudamiento patrimonial, la posibilidad de categorizar a una empresa como riesgosa o no riesgosa es más incierta, debido a que ingresa en una zona de incertidumbre como se lo puede apreciar en la figura.

Se puede observar en la figura 3.13 que las posibilidades de que una empresa caiga en quiebra son altas cuando la rotación de ventas es menor a 1, esto se debe

Figura 3.13: Predicción Parcial para la variable Rotación de ventas



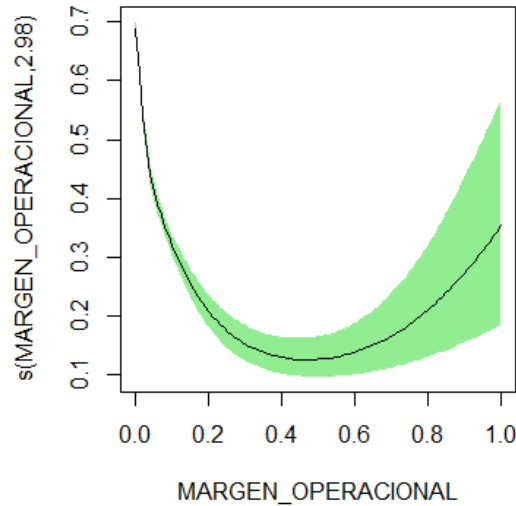
Elaboración: Los Autores

a que una entidad o empresa podría tener mucho capital inutilizado en sus activos. Así como también la posibilidad de riesgo de las empresas cuando este indicador es mayor a 1 y eso podría darse cuando la empresa registre pocos activos para hacer frente a un repunte de ventas o bien que los activos necesitan ser reemplazados por nuevos activos.

A partir de la figura 3.14 se puede observar que las posibilidades de determinar si una empresa puede llegar a tener una bancarrota es cuando su margen operacional es cercano a cero; mientras que, si el valor es menor a uno, la posibilidad es aún más incierta debido al área de incertidumbre que cubre este rango.

Finalmente, para el suavizamiento del margen operacional se puede ver en la figura (3.14) que las posibilidades de que una empresa se vuelva de alto riesgo ocurre cuando este indicador es menor a 0,2 o mayor a 0,6; además que, a partir de 0,6 el rango de incertidumbre de clasificar correctamente a una empresa aumenta. Por otra parte, entre los valores de 0,2 y 0,6 se muestra que las posibilidades de caer en quiebra es menor. Esto se debe a que la utilidad operaciones antes de impuestos y participación de los trabajadores es mayor a los ingresos por ventas.

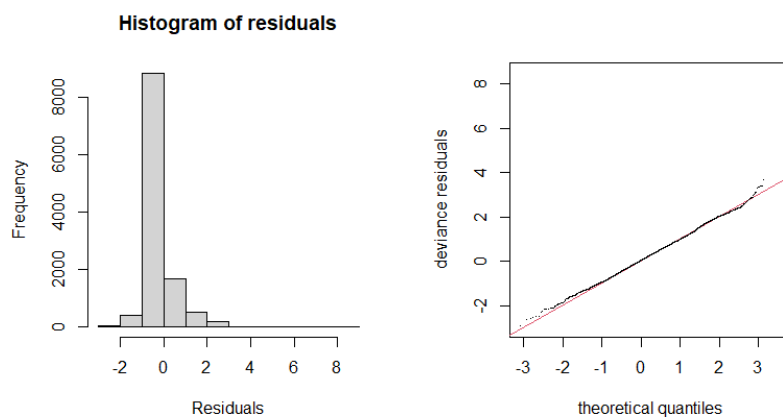
Figura 3.14: Predicción Parcial para la variable Margen Operacional



Elaboración: Los Autores

A continuación, se presentan los gráficos de rendimiento para verificar que el modelo encontrado (GAM-3) se ajusta de manera correcta a los datos sin presentar sobreajuste.

Figura 3.15: Distribución de los residuos para el modelo (GAM - 3)



(a) Histograma de los Residuos

(b) Residuos de la *deviance*

Elaboración: Los Autores

En la figura 3.15 se puede ver que los residuos siguen una distribución normal

con media cero y varianza mayor a cero. Esto quiere decir que el modelo es adecuado debido a la significancia del vector de los errores que muestra el modelo al momento de la estimación y realizar un adecuado proceso predictivo. El histograma se asemeja a una forma de campana que es la característica de la distribución normal. El gráfico de la probabilidad normal *QQ-Plot* muestra que la mayoría de los datos están sobre la línea diagonal roja, patrón que permite concluir que el modelo normal es razonable o adecuado para modelar los datos.

3.3.1. Resultados y validación del modelo aditivo generalizado

Dentro de esta nueva sección se procederá con la evaluación estadística del modelo previamente elaborado. Se lo hará a través del conjunto de datos de validación y los criterios estadísticos definidos en la sección 2.8; con esto se mide la capacidad predictiva que posee el modelo aditivo generalizado.

Para determinar si las variables seleccionadas para el modelo están correlacionadas o no, se aplica la prueba del cálculo del factor de inflación generalizado (GVIF) al modelo *GAM* – 3.

Cuadro 3.12: Factor de Inflación generalizado (GVIF) *GAM*-3

| Variable | GVIF |
|---|--------|
| Liquidez Corriente | 1.075 |
| Rentabilidad Operacional del Patrimonio | 2.4367 |
| Rentabilidad Financiera | 1.0235 |
| Rentabilidad Neta del Activo | 1.124 |
| Permanencia | 1.254 |
| Tamaño Empresa | 3.425 |
| Endeudamiento Patrimonial | 1.025 |
| Margen Operacional | 0.0524 |
| Rotación Ventas | 1.687 |

Elaboración: Los Autores

De la tabla 3.12 se puede ver que las variables no están correlacionadas; es decir, no existe multicolinealidad entre variables, pues los valores de GVIF para cada variable es menor a 10. Ahora, para medir que tan bien discrimina el modelo se calculan los estadísticos *AUC*, *KS* y *GINI*, tanto para el conjunto de modelamiento

como para el de validación. A continuación, se indican los valores calculados para cada estadístico:

Cuadro 3.13: Estadísticos de prueba

| Estadístico | Modelamiento | Validación |
|-------------|--------------|------------|
| KS | 0.541 | 0.541 |
| GINI | 0.941 | 0.919 |
| AUC | 0.970 | 0.959 |

Elaboración: Los Autores

Se puede ver en la tabla (3.13) que el estadístico *AUC* es un valor representativo superior al 80 %. Esto quiere decir que el modelo propuesto presenta una adecuada discriminación entre empresas en quiebra y no quebradas. El estadístico *KS* supera el 50 %, lo que indica que el modelo tiene un buen rendimiento. Por último, el coeficiente del estadístico *GINI* muestra un rendimiento adecuado del modelo para explicar y predecir el riesgo de quiebra empresarial.

Para una idea gráfica, se presenta a continuación la curva ROC para cada conjunto de datos (Modelamiento/Validación) y así mostrar la capacidad de clasificación que posee el modelo.

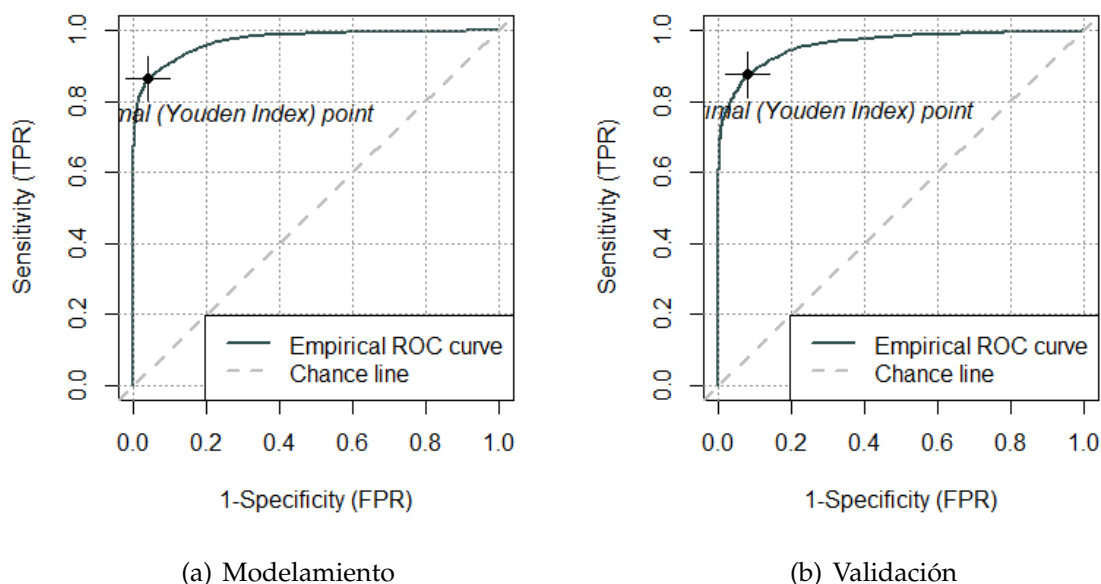
Posteriormente, se construyeron las matrices de confusión para las dos muestras, junto con las métricas descritas en la tabla 2.4.

Cuadro 3.14: Matriz de Confusión

| | | Modelamiento | | Validación | |
|------------|---|--------------|------|-------------|------|
| | | Observación | | Observación | |
| | | 0 | 1 | 0 | 1 |
| Predicción | 0 | 7597 | 257 | 3228 | 139 |
| | 1 | 565 | 3194 | 303 | 1308 |

Elaboración: Los Autores

Figura 3.16: Curva ROC para Modelamiento y Validación



Elaboración: Los Autores

Cuadro 3.15: Métricas de evaluación

| Modelamiento | | Validación | |
|---------------|--------|---------------|--------|
| Precisión | 0.9292 | Precisión | 0.9112 |
| Sensibilidad | 0.8496 | Sensibilidad | 0.8119 |
| Especificidad | 0.9672 | Especificidad | 0.9587 |

Elaboración: Los Autores

Del cuadro (3.15) se puede ver que la precisión del modelo propuesto tiene una precisión del 92,92 % para la muestra de modelamiento, mientras que en la muestra de validación se tiene una precisión del 91,12 %. Estos resultados muestran que el modelo está clasificando correctamente alrededor del 90 % de todas las empresas. La sensibilidad es un porcentaje superior al 80 %, lo que indica que más del 80 % de empresas están clasificadas correctamente y, la especificidad tiene un valor cerca del 95 %; esto indica que, los porcentajes de clasificación son altos y muy similares en ambas muestras.

Con estos resultados se puede ver que el modelo aditivo generalizado es una herramienta fundamental para encontrar el comportamiento real entre la variable

dependiente y las variables predictoras y así realizar un buen ajuste para predecir el riesgo de bancarrota. En la siguiente sección se procede a mostrar la estimación del riesgo de bancarrota empresarial, partiendo DE las variables seleccionadas por el modelo *GAM*, mediante el uso de las redes neuronales artificiales.

3.4. Redes Neuronales Artificiales

En esta nueva sección se procede aplicar el algoritmo supervisado de las redes neuronales artificiales (RNA), donde su principal insumo son las variables seleccionadas para estimar el riesgo de quiebra empresarial con la metodología *GAM*. Se realiza este proceso con el propósito de que los resultados puedan ser interpretables, pues los métodos de Machine Learning son considerados algoritmos de caja negra. Además, se busca que el rendimiento predictivo para el cálculo de la probabilidad de quiebra sea lo más robusto posible.

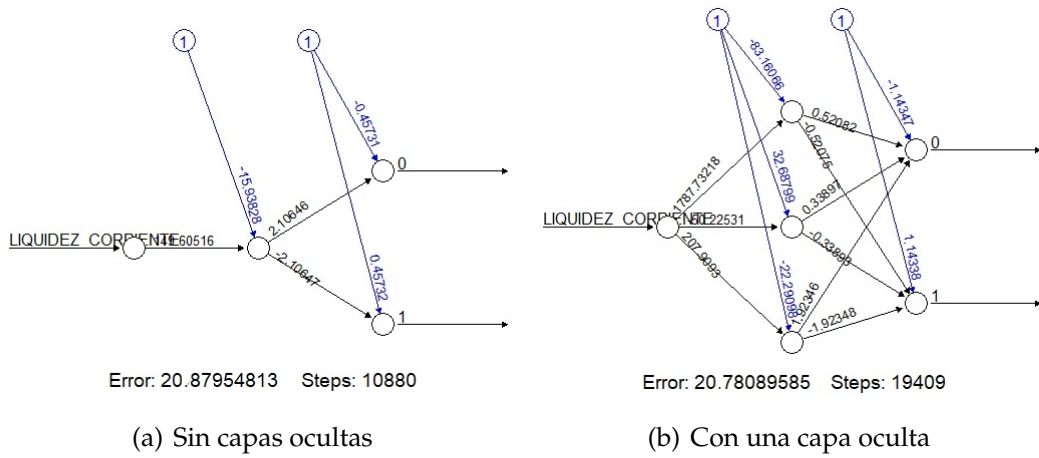
Para la metodología RNA se utilizan los mismo conjuntos de modelamiento y validación, lo que permite que los resultados sean válidos respecto al modelo aditivo generalizado elaborado en la sección precedente. Bajo estas consideraciones, se precede con la aplicación de la red neuronal artificial.

3.4.1. Construcción de una RNA

En una primera instancia se pretende usar las mismas variables que fueron seleccionadas para el cálculo de la quiebra con el modelo aditivo generalizado. No obstante, se considera ver el comportamiento y desarrollo de una red neuronal seleccionando las primeras variables que tuvieron impacto sobre la variable dependiente, es decir, las variables con mayor influencia para discriminar entre empresas en riesgo de quiebra y no quiebra. Para ello se consideró la variable *Liquidez Corriente*, a partir de la cual se puede observar la construcción de la primera red neuronal artificial.

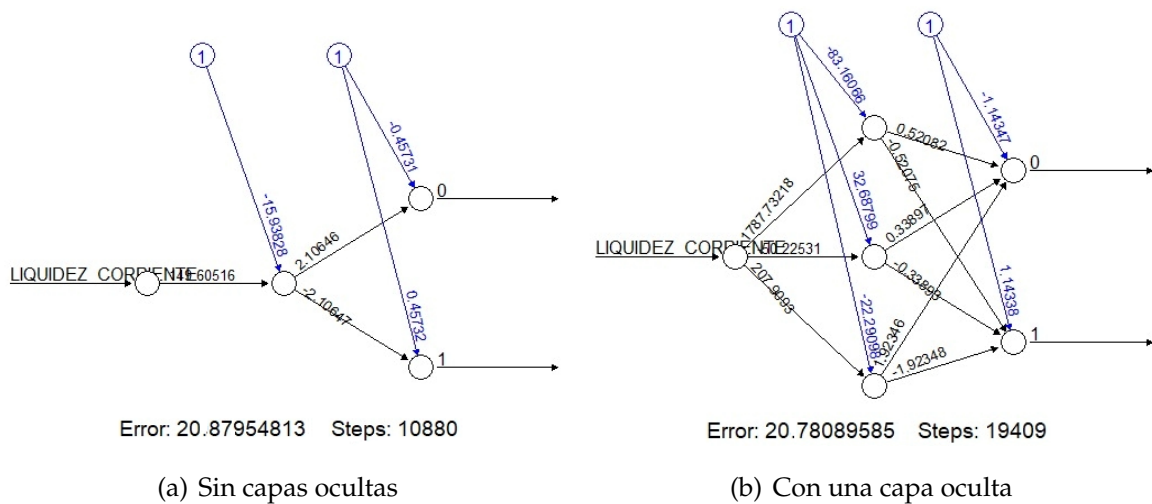
Se puede ver en la figura 3.17(a) que la red esta conformada por una sola capa oculta de una neurona, dando como resultado un error de aprendizaje de 20,87, el cual es un error alto puesto que la red puede mejorar. En cambio en la figura 3.17(b) la red esta conformada de igual manera por una capa oculta pero ahora con tres neuronas, permitiendo que el error de aprendizaje se reduzca a diferencia de la anterior; sin embargo, aún carece de un aprendizaje eficiente.

Figura 3.17: Red neuronal con una variable de entrada



Elaboración: Los Autores

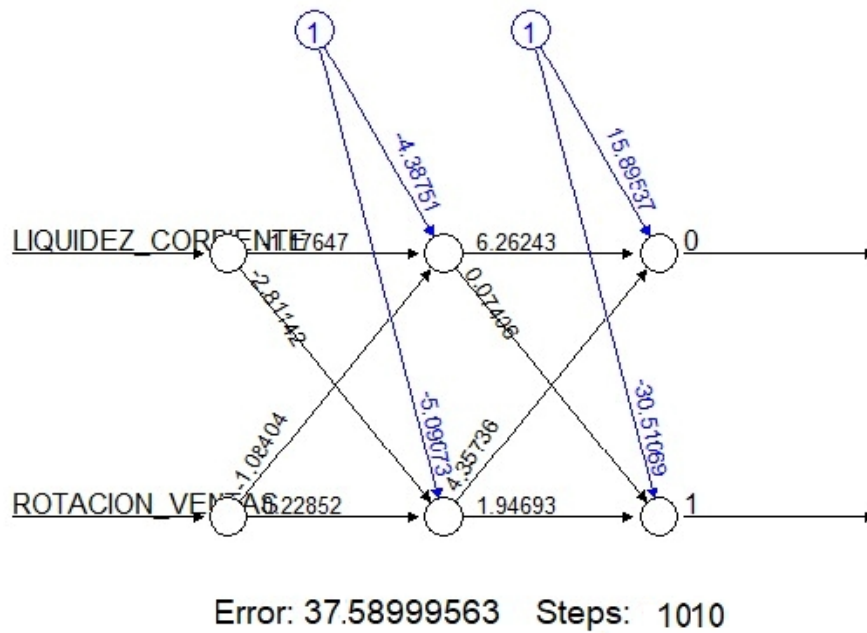
Figura 3.18: Red neuronal con una variable de entrada



Elaboración: Los Autores

En la figura (3.19) se muestra la red con dos variables de entrada y una capa oculta que contiene dos neuronas, donde a diferencia de las dos anteriores, al tener más de una variable de entrada permite interactuar estas características con la variables resultante y esta interacción la logra a través de estos dos nodos ocultos. Pero, la red sigue sin mejorar es su entrenamiento porque su error de aprendizaje aumentó.

Figura 3.19: Red neuronal con dos variables de entrada

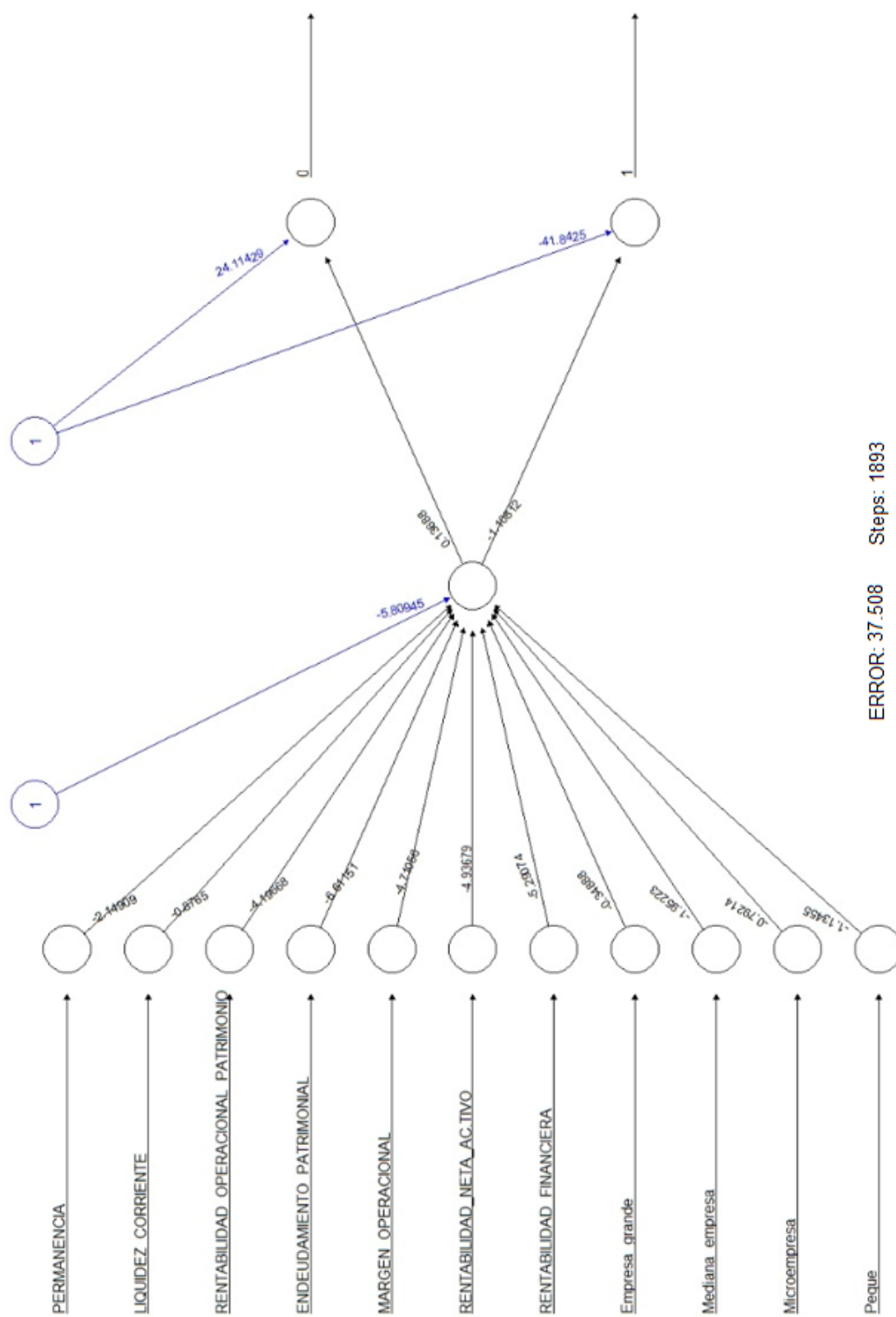


Elaboración: Los Autores

Ahora se ha creado una red en donde intervienen todas las variables identificadas por el modelo GAM, que en total son once. Esta red está conformada por una sola capa oculta de un solo nodo y un error de 37,508, indicando un aprendizaje no eficiente por parte de la red, lo cual es razonable pues al tener únicamente un nodo las variables de entrada no interactúan.

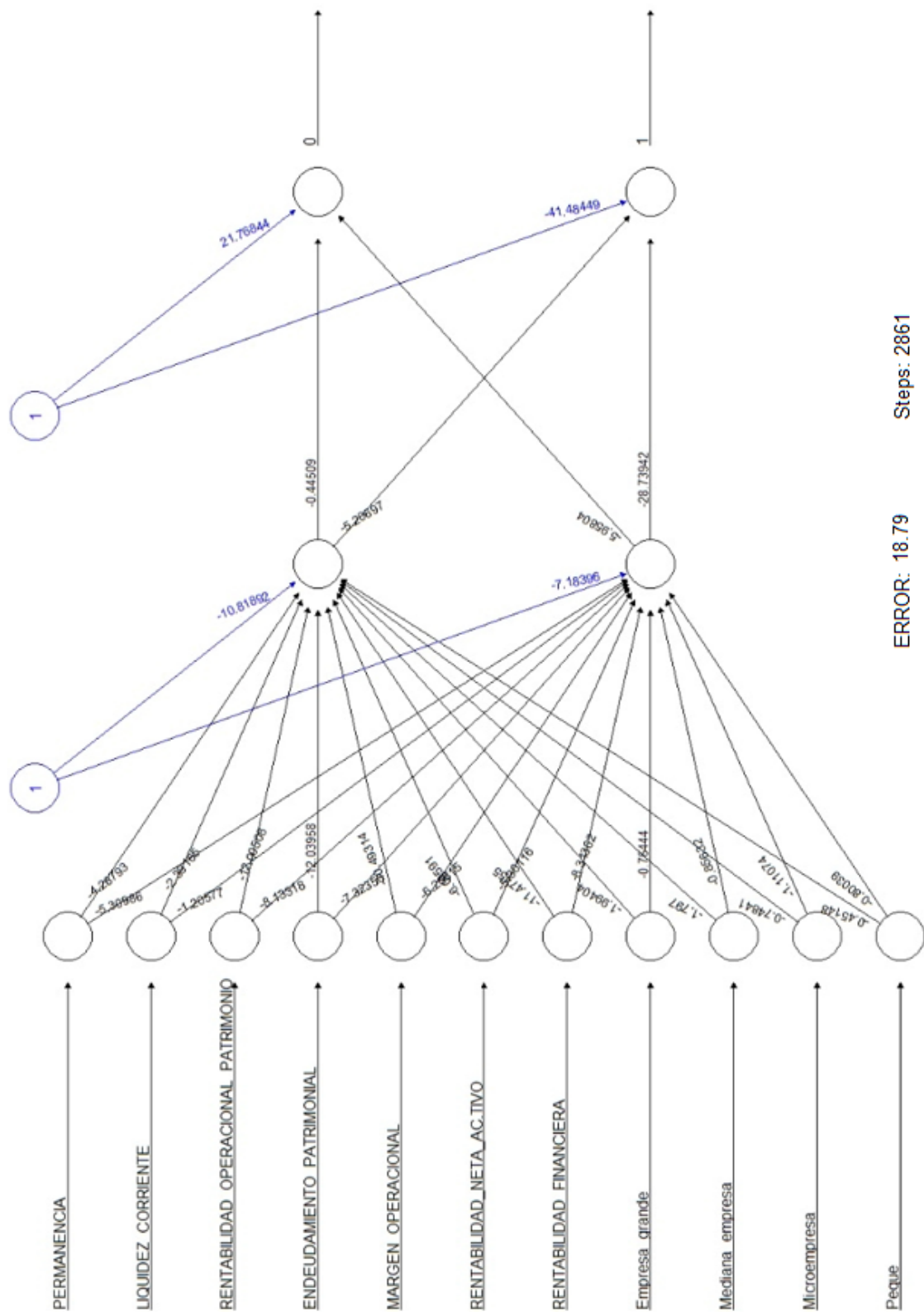
En la figura 3.21 se encuentra la red formada por dos nodos para la capa oculta, por lo que logran interactuar las variables y la red empieza a aprender, reduciendo su error a 18,75, lo cual es bajo a comparación de la red que se observa en la figura 3.20.

Figura 3.20: Red neuronal con once variables de entrada y una capa oculta



Elaboración: Los Autores

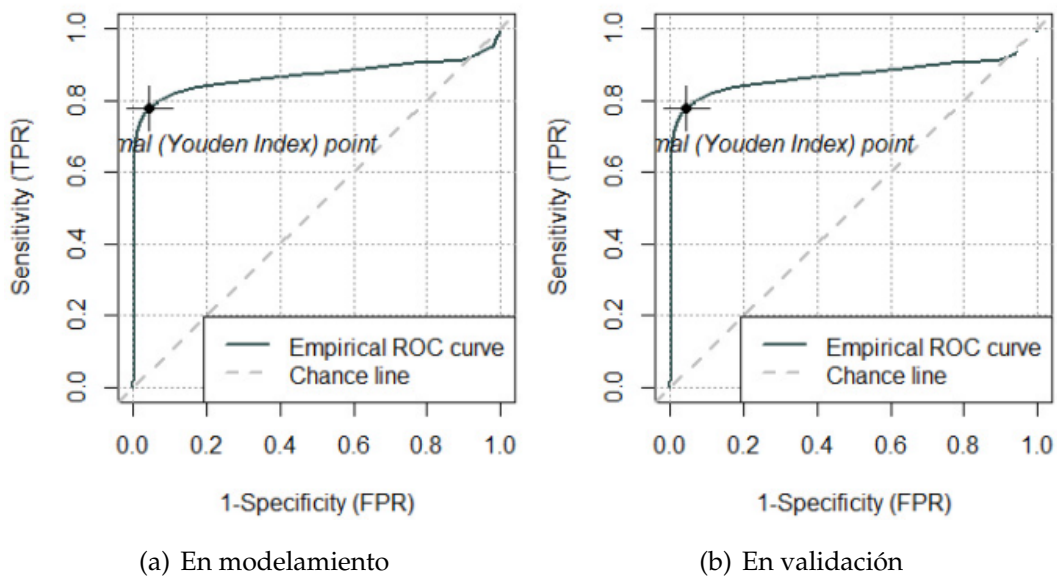
Figura 3.21: Red neuronal con once variables de entrada y una capa oculta con 2 nodos



Elaboración: Los Autores

En la última red compilada (figura 3.22), se observa que el error bajó considerablemente y para lograr esto la red ha tenido que aprender, pero gracias a la introducción de más de dos capas ocultas, donde la primera esta conformada por 4 neuronas, la segunda por 3 neuronas y la tercera por 1 neurona, total 8 nodos (neuronas). La red ha alcanzado aprender lo suficiente interactuando las entradas por varias capas y mejorando la predicción, de tal manera que se ha optimizado la función de coste y así se obtienen los resultados lo más cercanos a la realidad.

Figura 3.23: Curva ROC de la red neuronal para Modelamiento y Validación



Elaboración: Los Autores

Cuadro 3.16: Estadísticos de prueba

| Estadístico | Modelamiento | Validación |
|-------------|--------------|------------|
| GINI | 0.7331 | 0.6954 |
| AUC | 0.8665 | 0.8783 |

Elaboración: Los Autores

Se puede ver en el cuadro 3.16 que el estadístico AUC es más que representativo pues su valor es superior al 86 %, lo que implica que el modelo de la red neuronal tiene un comportamiento robusto para la realizar la predicción del riesgo de quiebra empresarial. Además, el estadístico GINI el cual es superior al 70 % implica que las empresas son clasificadas correctamente y así se puede determinar las que tienen mayor riesgo de quiebra.

Cuadro 3.17: Matriz de Confusión - Red Neuronal Artificial

| | | Modelamiento | | Validación | |
|------------|---|--------------|------|-------------|------|
| | | Observación | | Observación | |
| | | 0 | 1 | 0 | 1 |
| Predicción | 0 | 7276 | 578 | 3059 | 308 |
| | 1 | 961 | 2798 | 442 | 1169 |

Elaboración: Los Autores

Cuadro 3.18: Métricas de evaluación - Red Neuronal Artificial

| Modelamiento | | Validación | |
|---------------|--------|---------------|--------|
| Precisión | 0.8675 | Precisión | 0.8493 |
| Sensibilidad | 0.8288 | Sensibilidad | 0.7915 |
| Especificidad | 0.8833 | Especificidad | 0.8737 |

Elaboración: Los Autores

Se puede observar en la tabla 3.18 que el modelo clasifica correctamente al 85 % del total de la población, lo que quiere decir que el modelo de la red neuronal discrimina de manera adecuada a las empresas con mayor riesgo de quiebra. Además, respecto al entorno de entrenamiento el ajuste disminuye aproximadamente en un 1.82 % en la clasificación general; sin embargo, la diferencia es mínima por lo que es un margen de error aceptable. Se concluye de esta manera que la metodología propuesta tiene un buen rendimiento para predecir el riesgo de quiebra empresarial.

Capítulo 4

Resultados e Interpretación

Después de seleccionar las variables que mejor discriminan a las empresas, se procede a detallar el comportamiento que influye en la probabilidad de riesgo, lo que se hace a través de la predicción de la incertidumbre del quiebre empresarial. Para ello, se considera el número de empresas por cada año de estudio que fueron clasificadas con posible riesgo de bancarrota. En el cuadro (4.1) el 70,75 % del total de la población estudiada no son consideradas en riesgo de acuerdo al modelo de Redes Neuronales Artificiales.

Cuadro 4.1: Cantidad de empresas por periodo de tiempo

| QUIEBRA | Período de Tiempo | | | | | Total | % Participacion |
|---------|-------------------|------|------|------|------|-------|-----------------|
| | 2016 | 2017 | 2018 | 2019 | 2020 | | |
| 0 | 2874 | 2398 | 2276 | 2046 | 2144 | 11738 | 70,75 % |
| 1 | 1238 | 963 | 976 | 804 | 872 | 4853 | 29,25 % |
| Total | 4112 | 3361 | 3252 | 2850 | 3016 | 16591 | 100 % |

Elaboración: Los Autores

Además, se encontró como resultado que las empresas consideradas en quiebra tienen una mayor posibilidad de bancarrota que las empresas que no se encuentran clasificadas como quebradas o insolventes. A continuación, se presenta el resultado de la probabilidad de quiebra por año.

Cuadro 4.2: Probabilidad de Riesgo por cada periodo de tiempo.

| QUIEBRA | 2016 | 2017 | 2018 | 2019 | 2020 | General |
|---------|---------|---------|---------|---------|---------|---------|
| 0 | 7,15 % | 6,34 % | 6,26 % | 5,96 % | 5,53 % | 6,31 % |
| 1 | 94,69 % | 95,12 % | 95,87 % | 95,92 % | 95,59 % | 95,38 % |

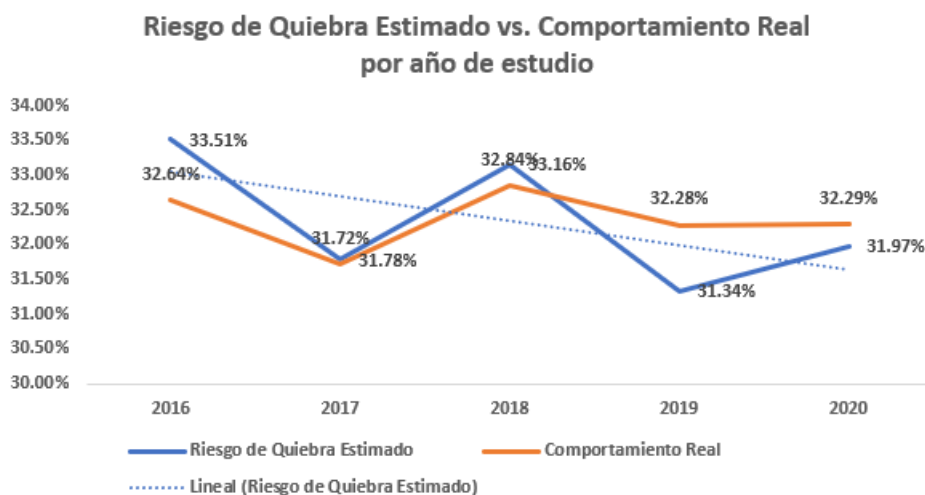
Elaboración: Los Autores

En la tabla 4.2 se puede ver que, en efecto las empresas calificadas en quiebra son las que tienen una mayor posibilidad de ingresar a un proceso de insolvencia, pues se identifica que el riesgo de bancarrota promedio para las mismas es del 95,38 %.

Resultados del riesgo de quiebra empresarial anual

En la figura 4.1 se puede observar el comportamiento del riesgo de quiebra temporal, en el que se puede apreciar una tendencia decreciente en el riesgo con su valor más alto en el año 2016 y un ligero incremento del 0,63 % para el año 2020.

Figura 4.1: Riesgo de Quiebra Estimado vs. Comportamiento Real en el sector Manufacturero.



Elaboración: Los Autores

Para complementar, en la figura 4.1 se muestra una comparativa entre la tasa de riesgo de la bancarrota y el valor estimado del riesgo de quiebra calculado a partir de la Red neuronal artificial. Por ejemplo: se puede ver que la tasa de riesgo de quiebra para el año 2016 es del 32,64 %; mientras que, el valor estimado de la

probabilidad de riesgo es del 33,51 %. Este resultado indica que la estimación del riesgo de quiebra no está tan alejado de la realidad, pues el valor de error entre el valor real y el estimado es del 0,87 %. Este resultado indica que las estimaciones realizadas por la RNA son cercanas al valor del comportamiento real en la tasa de riesgo de quiebra empresarial.

Resultados del riesgo de quiebra empresarial por tamaño de empresa en el período de estudio

El modelo fue también aplicado a las empresas según su tamaño empresarial, pues de esta manera se puede observar cuales empresas son propensas a una insolvencia financiera y cuales no, para de esta manera crear alertas tempranas que permitan conocer la situación real de una empresa. Por lo tanto, en la siguiente tabla se observa que las micro y pequeñas empresas tienen una mayor probabilidad de insolvencia. Este resultado está acorde a la realidad de las Pymes, pues muchas de estas carecen de fuentes de financiamiento para un desarrollo y crecimiento económico interno, a diferencia de las medianas y grandes empresas donde estas presentan una menor probabilidad de quiebra, pues muchas de estas, aparte de tener fuentes de financiamiento, son las que poseen grandes valores en sus patrimonios.

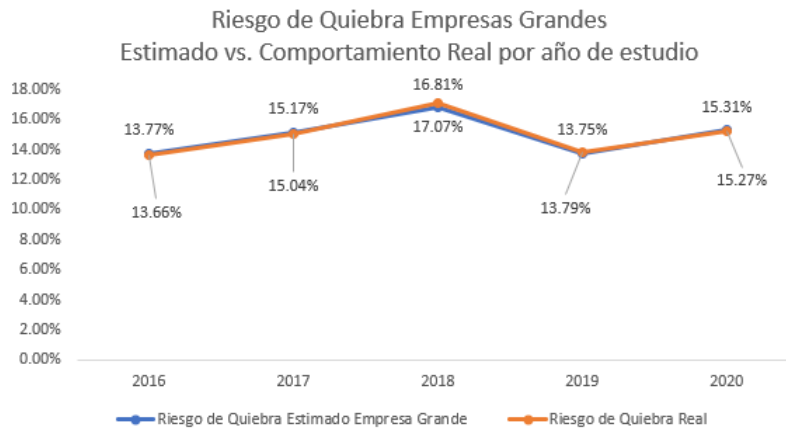
Cuadro 4.3: Riesgo de quiebra empresarial por tamaño de empresa

| AÑO | Empresa Grande | Mediana Empresa | Pequeña Empresa | Microempresa |
|------------|-----------------------|------------------------|------------------------|---------------------|
| 2016 | 13,77 % | 27,32 % | 35,48 % | 45,77 % |
| 2017 | 15,17 % | 25,84 % | 33,79 % | 47,89 % |
| 2018 | 16,81 % | 27,82 % | 36,05 % | 48,50 % |
| 2019 | 13,75 % | 25,59 % | 34,80 % | 45,70 % |
| 2020 | 15,31 % | 24,69 % | 32,89 % | 47,58 % |

Fuente: Superintendencia de Compañías, Valores y Seguros (SUPERCIAS)

Elaboración: Los Autores

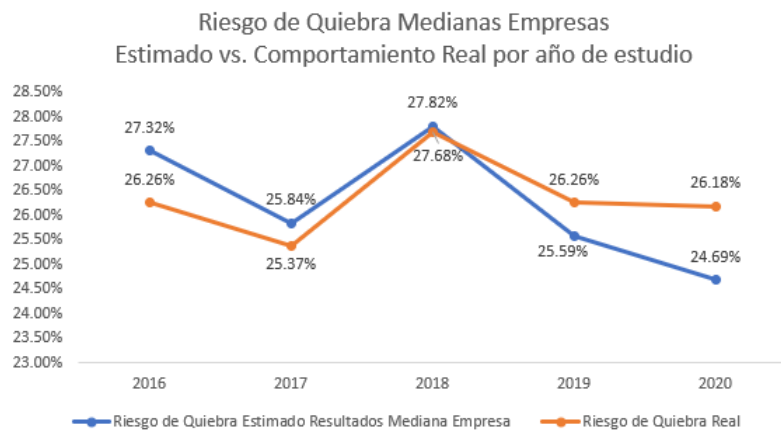
Figura 4.2: Riesgo de quiebra Estimado vs. comportamiento real para Empresas Grandes por año de estudio.



Elaboración: Los Autores

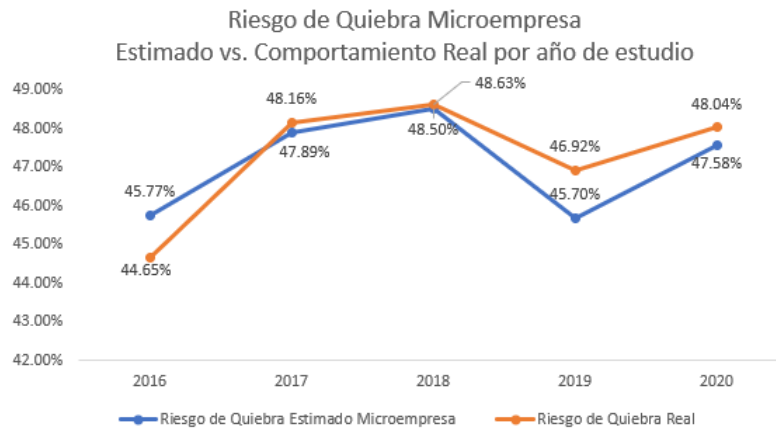
En la figura 4.2 se puede observar que las grandes empresas poseen un menor riesgo de quiebra en comparación de las micro y pequeñas empresas manufactureras ecuatorianas. Se puede ver también que la estimación del riesgo de quiebra es muy cercano al valor calculado del comportamiento real de las empresas grandes. Esto quiere decir que, la estimación realizada por el modelo indica un adecuado desempeño al momento de calcular la probabilidad de riesgo de bancarrota. A continuación, se presenta el riesgo de quiebra estimado contra el comportamiento real para medianas, pequeñas y microempresas.

Figura 4.3: Riesgo de quiebra Estimado vs. comportamiento real para Medianas Empresas por año de estudio.



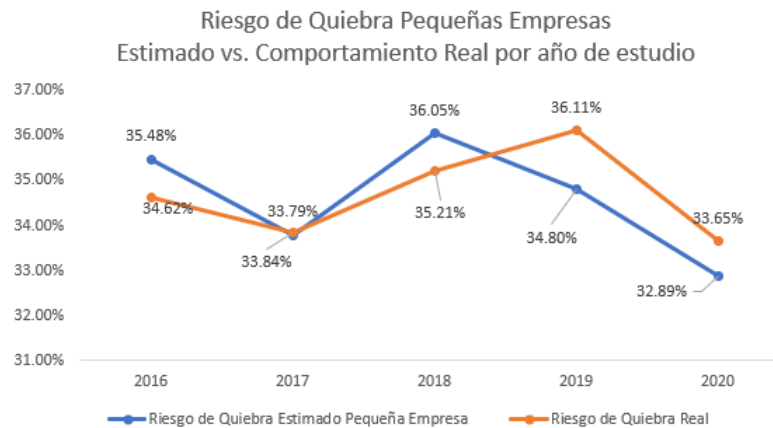
Elaboración: Los Autores

Figura 4.4: Riesgo de quiebra Estimado vs. comportamiento real para Microempresas por año de estudio.



Elaboración: Los Autores

Figura 4.5: Riesgo de quiebra Estimado vs. comportamiento real para Pequeñas Empresas por año de estudio.



Elaboración: Los Autores

Para finalizar, se puede que para medianas, pequeñas y microempresas el desempeño del modelo es similar al comportamiento real, por lo que el modelo propuesto identifica y calcula de manera adecuada el riesgo de bancarrota empresarial.

Riesgo de quiebra empresarial por Provincia

En los siguientes resultados (Ver cuadro 4.4) se observa que la provincia del Napo posee el mayor riesgo de quiebra promedio anual con un 70,83 %, seguido por la

provincia de Santa Elena con un riesgo de quiebra promedio del 52,51 %, Loja con un riesgo del 37,68 % y Azuay con un riesgo del 35,02 %.

Cuadro 4.4: Análisis del riesgo de quiebra empersarial por provincia

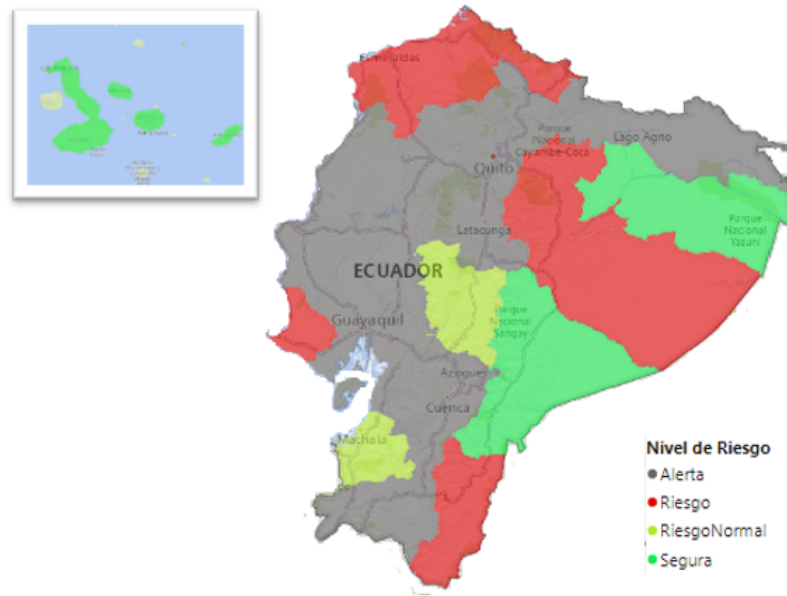
| PROVINCIA | 2016 | 2017 | 2018 | 2019 | 2020 |
|-----------------------------------|---------|---------|---------|---------|---------|
| AZUAY | 35,14 % | 32,54 % | 38,11 % | 34,54 % | 34,75 % |
| BOLIVAR | 13,44 % | 11,72 % | 27,05 % | 0,00 % | 0,00 % |
| CAÑAR | 7,22 % | 23,43 % | 14,97 % | 15,36 % | 15,40 % |
| CARCHI | 38,12 % | 36,38 % | 7,48 % | 3,98 % | 7,34 % |
| COTOPAXI | 29,50 % | 32,66 % | 19,18 % | 29,71 % | 26,36 % |
| CHIMBORAZO | 30,64 % | 11,38 % | 35,78 % | 21,67 % | 23,97 % |
| EL ORO | 21,26 % | 8,93 % | 20,84 % | 26,30 % | 23,21 % |
| ESMERALDAS | 20,77 % | 50,46 % | 45,01 % | 18,04 % | 21,24 % |
| GUAYAS | 31,92 % | 30,74 % | 31,97 % | 28,92 % | 29,17 % |
| IMBABURA | 27,37 % | 30,96 % | 33,04 % | 30,71 % | 20,36 % |
| LOJA | 44,44 % | 33,44 % | 35,46 % | 32,19 % | 42,87 % |
| LOS RIOS | 24,52 % | 29,05 % | 21,97 % | 6,35 % | 23,36 % |
| MANABI | 32,36 % | 30,97 % | 33,54 % | 39,65 % | 35,60 % |
| MORONA SANTIAGO | 8,47 % | 4,30 % | 32,05 % | 99,99 % | 0,00 % |
| NAPO | 99,99 % | 99,99 % | 29,41 % | 24,75 % | 99,00 % |
| PASTAZA | 59,89 % | 57,45 % | 1,87 % | 16,17 % | 18,57 % |
| PICHINCHA | 35,98 % | 33,62 % | 33,99 % | 33,69 % | 33,62 % |
| TUNGURAHUA | 27,57 % | 30,42 % | 37,55 % | 27,59 % | 36,99 % |
| ZAMORA CHINCHIPE | 0,00 % | 66,12 % | 1,21 % | 0,00 % | 0,00 % |
| GALAPAGOS | 21,47 % | 0,07 % | 1,18 % | 0,18 % | 10,51 % |
| SUCUMBIOS | 25,23 % | 19,92 % | 23,46 % | 23,62 % | 22,85 % |
| ORELLANA | 31,94 % | 1,69 % | 26,12 % | 34,40 % | 6,81 % |
| SANTO DOMINGO DE LOS TSACHILAS | 36,49 % | 23,95 % | 36,06 % | 29,34 % | 21,31 % |
| SANTA ELENA | 44,11 % | 99,99 % | 53,27 % | 63,68 % | 1,52 % |

Elaboración: Los Autores

Para una mejor visualización del comportamiento de la incertidumbre del quiebra empresarial se procedió a elaborar un mapa de riesgos en donde se puede ver por sectores y tamaño empresarial qué empresas tienen una menor y mayor posi-

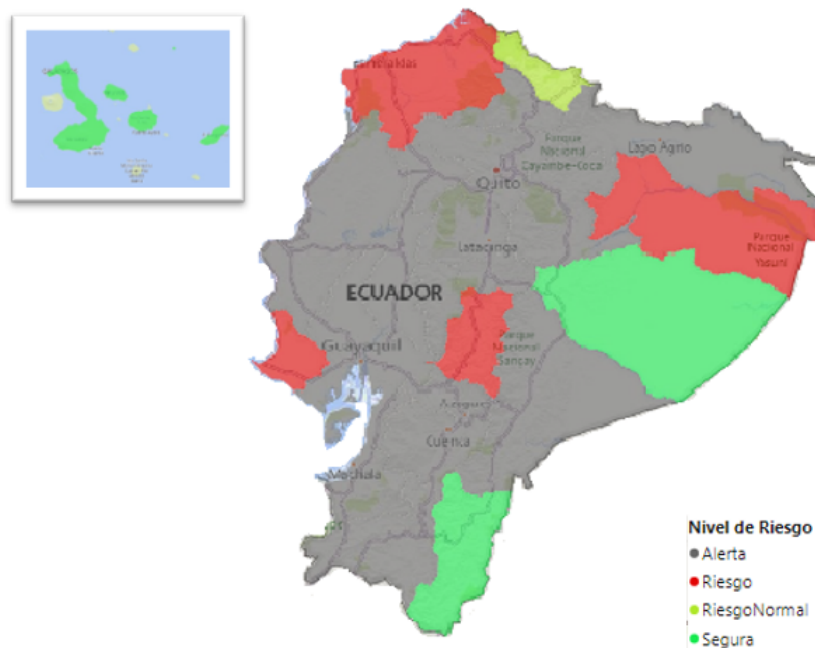
del 44,44 % disminuyendo para el año 2017 nueve puntos porcentuales (33,34 %). Esto se manifiesta debido a que empresas con riesgo alto lograron tener una mejor administración reflejando un mejor desempeño financiero.

Figura 4.7: Riesgo de quiebra empresarial por segmento de riesgo en el año 2017.



Elaboración: Los Autores

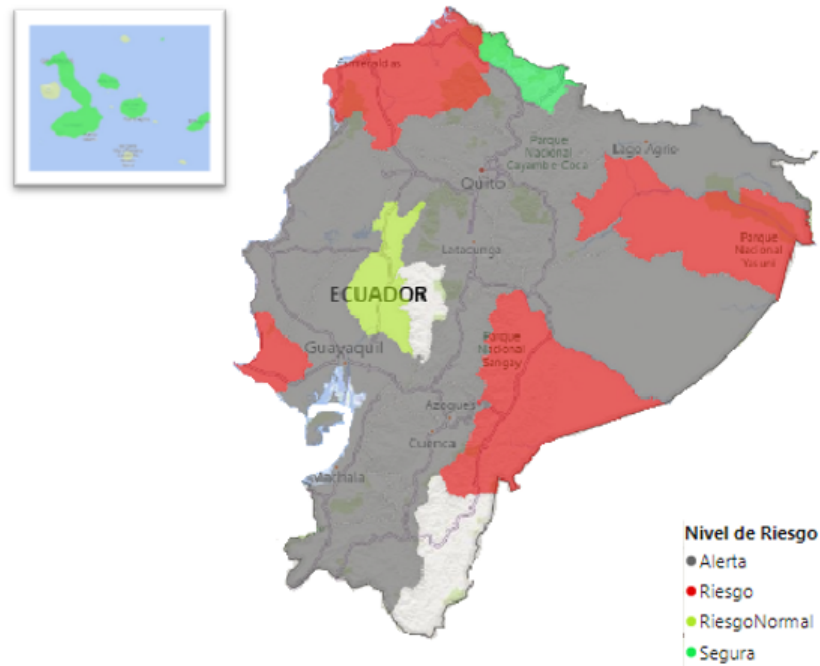
Figura 4.8: Riesgo de quiebra empresarial por segmento de riesgo en el año 2018.



Elaboración: Los Autores

Las probabilidades de riesgo de quiebra entre las años 2017 y 2018 para las provincias de Pastaza y Zamora Chinchipe, bajaron del 99 % al 2 % de riesgo y del 66,2 % al 1 %, respectivamente. Esta variación se debió a que las empresas dentro de estas provincias enviaron sus informes financieros al ente regular y por lo tanto existió mayor control para este período.

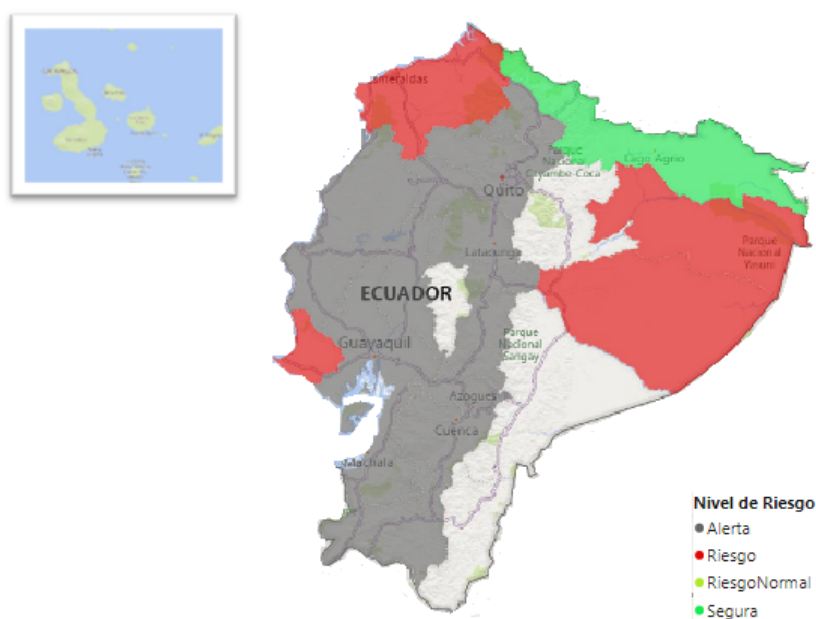
Figura 4.9: Riesgo de quiebra empresarial por segmento de riesgo en el año 2019.



Elaboración: Los Autores

A pesar de la emergencia sanitaria, se obtuvo como resultado que no hay provincias con empresas que tengan un mayor riesgo de quiebra pues durante el periodo de Pandemia muchas empresas optaron por el comercio digital (Ver figura 4.10), y realizar operaciones de negocio de manera virtual, ya que se puede observar que todas las empresas para el año 2020 entraron en zona de alerta. La innovación digital ayudó a que muchas empresas no incurran en una inminente bancarrota, pues a diferencia del año 2019 las empresas con mayor riesgo se mantienen en las provincias de Santa Elena, Esmeraldas, Zamora Chinchipe y Orellana

Figura 4.12: Riesgo de quiebra empresarial por segmento de riesgo en Medianas Empresas.



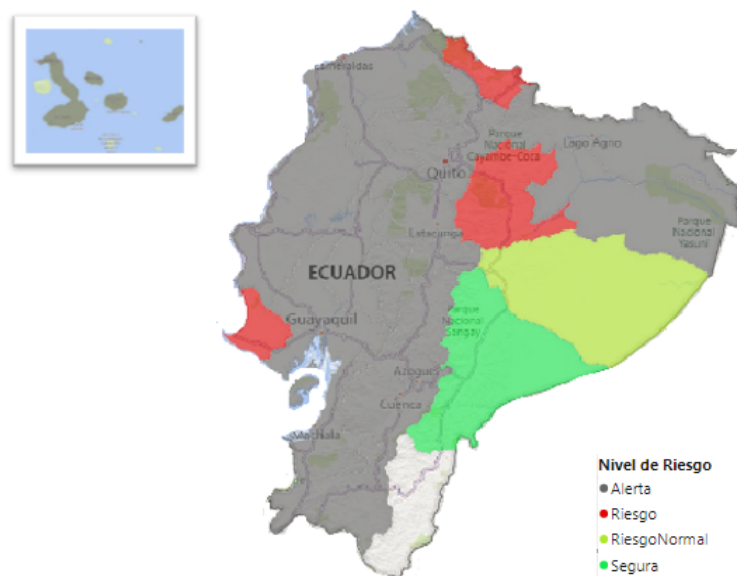
Elaboración: Los Autores

En la figura 4.13 se observa que las pequeñas empresas del sector manufacturero presentan actividad en 23 de las 24 provincias; con respecto a la zona segura, Morona Santiago (4,4%) es la provincias que presenta menor riesgo de quiebra. En el segmento de Riesgo normal, las empresas clasificadas en este segmento son Galápagos (11,6%), Los Ríos (12,7%) y Esmeraldas (13,4%), pues las mismas presentan un menor riesgo de quiebra en este segmento. Finalmente en la zona de Riesgo se tiene a las provincias de Santa Elena (76,5%) y Napo (99%) donde presentan la mayor probabilidad de riesgo del quiebra. La diferencia de información se debe a la carencia de reporte de información al ente regulador.

Finalmente, en la figura 4.14 las provincias que abarcan las empresas con mayor riesgo de quiebra empresarial en el sector manufacturero son Pichincha (54,6%), Orellana (84,7%), Zamora Chinchipe (66,3%) y Azuay (54,8%). Pues, las mismas indican un mayor riesgo de quiebra por su falta de desarrollo y crecimiento económico debido alto riesgo de insolvencia financiera.

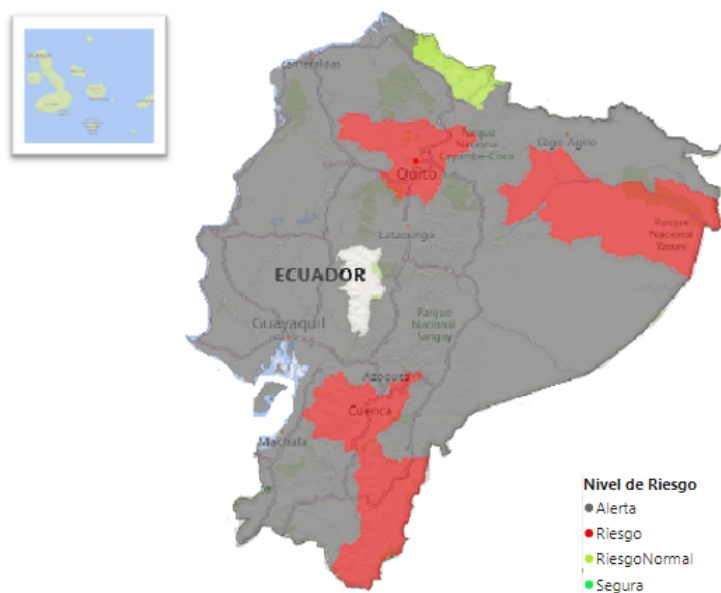
Todos los resultados descritos se encuentran a detalle en el Anexo C.

Figura 4.13: Riesgo de quiebra empresarial por segmento de riesgo en Pequeñas Empresas.



Elaboración: Los Autores

Figura 4.14: Riesgo de quiebra empresarial por segmento de riesgo en Microempresas.



Elaboración: Los Autores

Capítulo 5

Conclusiones y recomendaciones

5.1. Conclusiones

- Las industrias manufactureras son una fuente importante dentro del Ecuador porque generan productividad económica y social. Es uno de los sectores económicos que más aporta en el producto interno bruto ecuatoriano.
- Con el análisis exploratorio de todas las variables consideradas para la explicación y predicción del riesgo de quiebra empresarial, se identificó el comportamiento de la tasa de quiebra empresarial según su tamaño, en donde se determinó que entre más grande es la empresa su riesgo de quiebra es menor, a comparación con las pequeñas y micro empresas, por lo que este resultado permite explicar el impacto al medir el riesgo de quiebra por tamaño de empresa.
- Las variables de mayor impacto que reflejaron el comportamiento financiero de las empresas fueron: los indicadores de liquidez corriente, rentabilidad operacional del patrimonio, rentabilidad financiera, tamaño de empresa, entre otros. Estas variables fueron seleccionadas para el modelo debido a la significancia asociada al coeficiente de cada variable, y también porque tuvieron un valor de KS y VI superior.
- Los datos atípicos en las variables financieras que se estudiaron en esta investigación, fueron tratados acorde a métodos estadísticos como la mediana, media o desviación estándar, dependiendo de la distribución y simetría de las variables. Cabe señalar que no fue eliminado ningún registro sino que se reemplazaron por la medida de tendencia central *mediana* debido a la robustez

que tiene respecto a las otras 2 medidas.

- En este trabajo de investigación se estimó un modelo aditivo generalizado con las variables seleccionadas a partir de estadísticos de prueba, las cuales explican el riesgo del fracaso empresarial validando los resultados del modelo a través de las pruebas de rendimiento. Por ejemplo, una empresa es más riesgosa si posee un valor en patrimonio menor al de los activos, pues indica que la empresa no puede hacer frente a sus obligaciones financieras.
- Los modelos GAM utilizan funciones de suavizamiento para determinar con mayor exactitud que tipo de tendencia poseen las variables, debido a que no todas tienen tendencia lineal; la mayoría de ellas tienen tendencia no lineal y al no aplicarles un suavizamiento entonces el modelo no las selecciona, ya que no tienen significancia para con la variable dependiente; sin embargo, no es correcto no tomarlas en cuenta, pues lo más seguro es que si tenga relación con la variable dependiente pero no de forma lineal.
- Para validar un rendimiento adecuado al modelo se procedió a calcular los estadísticos, GINI, ROC y Ks, los que dan a conocer la robustez, la explicación y el poder predictivo que tiene este modelo para estimar eventos futuros. De la misma manera, se utilizó la matriz de confusión, cuyo principal resultado es el discriminante del *Accuracy*, que indica cuántas empresas fueron correctamente clasificadas como quebradas y no quebradas. En el caso del GAM, el *Accuracy* sobrepasa el 90 %.
- Con base en los resultados que muestra el rendimiento del modelo GAM, se seleccionaron como variables de entrada para la red neuronal a las variables que mejor explican el quiebre empresarial en el GAM, lo cual hace más eficiente el resultado de la RNA debido al alto poder predictivo que posee este modelo, generando una mayor intrerpretabilidad de los resultados que se obtuvieron con la ejecución de la red neuronal.
- Al igual que en modelo GAM, se procedió a medir el desempeño del modelo de la red neuronal artificial mediante los estadísticos de prueba, con los cuales se logró constatar que el valor del área bajo la curva (ROC) es superior al 86 %. Esto quiere decir que el rendimiento del modelo para predecir el riesgo de quiebra es real, con un *accuracy* del 86,75 %.
- Es notorio ver que los estadísticos de prueba midieron un buen desempeño

del modelo de las redes neuronales artificiales, tanto en el conjunto de entrenamiento como en el de validación.

- Con los resultados de la red neuronal artificial se busca automatizar patrones para la estimación del riesgo de quiebra, lo que no se puede realizar únicamente con cualquier otro tipo de metodología, en este caso con el modelo GAM.
- Entre los resultados que se obtuvieron dentro de la investigación fue que en efecto el pronóstico del riesgo de quiebra empresarial por año indica que las empresas grandes tienen un promedio de riesgo menor en el tiempo de estudio, alcanzando su valor máximo en el año 2018; para las medianas empresas la tasa promedio de riesgo alcanzó el 27.82 %, su valor más alto en el año 2018. A su vez, el riesgo de quiebra promedio para las microempresas alcanza valores altos en cada año de estudio, llegando en el año 2020 a una tasa del 47.58 %, este comportamiento fue propio de la emergencia sanitaria que afectó a gran parte de la economía ecuatoriana perjudicando al microempresario.
- Se pudo evidenciar en los resultados que al predecir la incertidumbre de quiebra, las pequeñas y micro empresas son en efecto las más riesgosas pues su riesgo promedio supera el 44,5 %, lo que justifica el porqué estos sectores al no tener fuentes de financiamiento interno como externo tienden a la insolvencia financiera y una eventual bancarrota en el corto plazo.

5.2. Recomendaciones

- Cuando se hace uso de las redes neuronales artificiales, es recomendable siempre usar capas ocultas, ya que si no se las tuviese, entonces se tendría que trazar una conexión directa entre las entradas y la salida; la contribución de cada entrada hacia el valor de salida sería independiente de las otras entradas. Pero en la mayoría de los problemas del mundo real, las variables de entrada tienden a ser altamente interdependientes y afectan la salida de forma combinatoria, por lo que las neuronas de las capas ocultas permiten capturar justamente estas interacciones sutiles entre las entradas.
- Al realizar una red neuronal artificial es aconsejable hacer uso de al menos dos capas ocultas, ya que esto permite tener mejor interacción entre las variables y determinar más aproximado el valor de salida predicho en comparación al valor real.

- Se debe considerar para trabajos futuros en el ámbito del riesgo empresarial metodologías robustas que permitan identificar el comportamiento real entre variables, pues en la mayoría de los casos no siempre la relación es de carácter lineal, por lo que buscar técnicas que permitan encontrar el comportamiento real será de mucha utilidad pues las mismas minimizan el riesgo de sesgo en los resultados deseados.
- Se sugiere establecer políticas de desarrollo y dinamismo económico dentro del sector manufacturero de mayor riesgo, para de esa manera identificar a empresas potenciales para la bancarrota y crear alertas tempranas que permitan a las empresas predecir eventos futuros en base a un diagnóstico real.

Apéndice A

Estadísticos Descriptivos

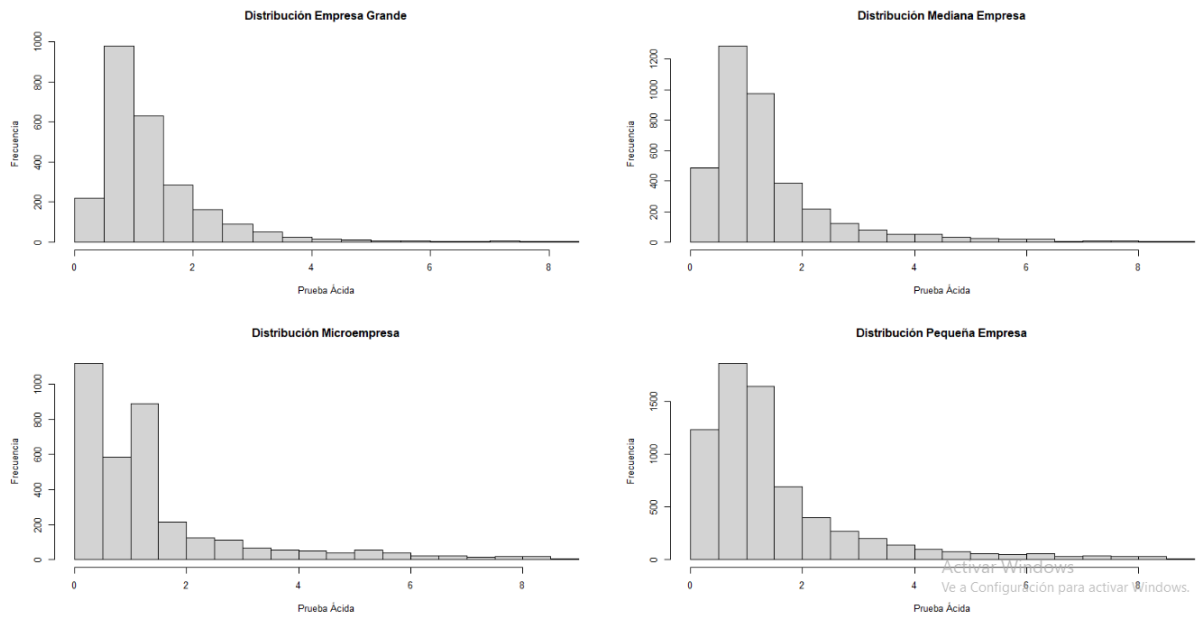
Cuadro A.1: Resumen Estadístico

| Variables | Media | Desv_Est | Mínimo | 1er Cuartil | Mediana | 3er Cuartil | Máximo |
|-------------------------------------|--------------|-----------------|---------------|--------------------|----------------|--------------------|---------------|
| Endeudamiento Activo | 0.95 | 19.60 | 9.14E-06 | 0.43 | 0.66 | 0.86 | 2.35E+03 |
| Liquidez Corriente | 84.94 | 8,209.94 | 0.00 | 1.01 | 1.50 | 2.68 | 1.05E+06 |
| Prueba Acida | 19.32 | 802.13 | 0.00 | 0.62 | 1.06 | 1.92 | 7.75E+04 |
| Endeudamiento Patrimonial | 10.44 | 129.43 | 0.00 | 0.52 | 1.50 | 4.00 | 1.22E+04 |
| Eneudamiento_Activo Fijo | 1.71E+14 | 1.01E+16 | 0.00 | 0.52 | 1.18 | 3.06 | 6.54E+17 |
| Apalancamiento | 12.99 | 132.69 | 4.26E-04 | 1.68 | 2.74 | 5.68 | 1.22E+04 |
| Apalancamiento Financiero | 12.01 | 131.14 | 0.00 | 1.48 | 2.43 | 5.07 | 1.22E+04 |
| Rotacion Cartera | 5.61E+04 | 5.95E+06 | 0.00 | 6.10 | 15.87 | 37.80 | 7.58E+08 |
| Rotacion_Activo Fijo | 2.06E+14 | 1.76E+16 | 0.00 | 1.46 | 4.12 | 13.19 | 1.99E+18 |
| Rotacion Ventas | 1.75 | 7.64 | 0.00 | 0.64 | 1.16 | 1.92 | 7.52E+02 |
| Periodo_Medio Cobranza | 8,134.94 | 4.53E+05 | 0.04 | 4.11 | 10.05 | 25.00 | 5.19E+07 |
| Rentabilidad_Neta Activo | 0.07 | 0.30 | 0.00 | 0.00 | 0.02 | 0.08 | 3.47E+01 |
| Margen Bruto | 0.42 | 0.28 | 0.00 | 0.21 | 0.35 | 0.57 | 1.00 |
| Margen Operacional | 0.05 | 0.10 | 0.00 | 0.00 | 0.02 | 0.07 | 1.00 |
| Margen Neto | 0.04 | 0.08 | 0.00 | 0.00 | 0.01 | 0.04 | 1.00 |
| Rentabilidad Operacional Patrimonio | 0.92 | 14.42 | 0.00 | 0.06 | 0.19 | 0.49 | 1,605.25 |
| Rentabilidad Financiera | 0.26 | 1.34 | 0.00 | 0.00 | 0.08 | 0.26 | 73.40 |

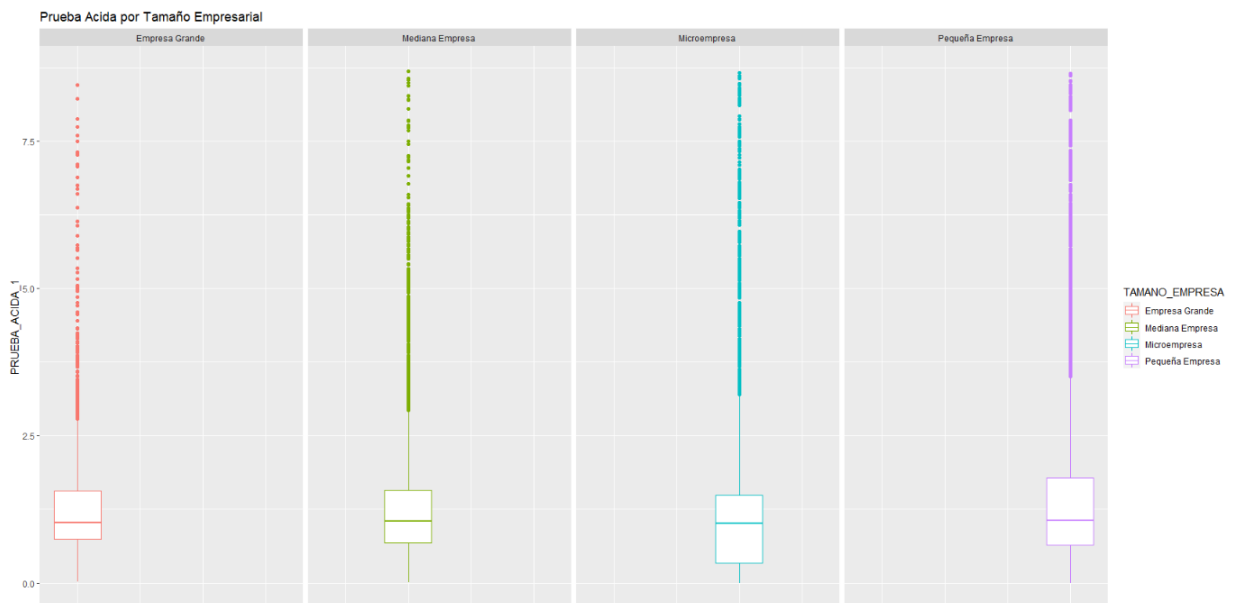
Apéndice B

Diagrama de Cajas y Bigotes Variables Numéricas

Figura B.1: Distribución y Diagrama de cajas de la variable *Prueba Ácida* tratada

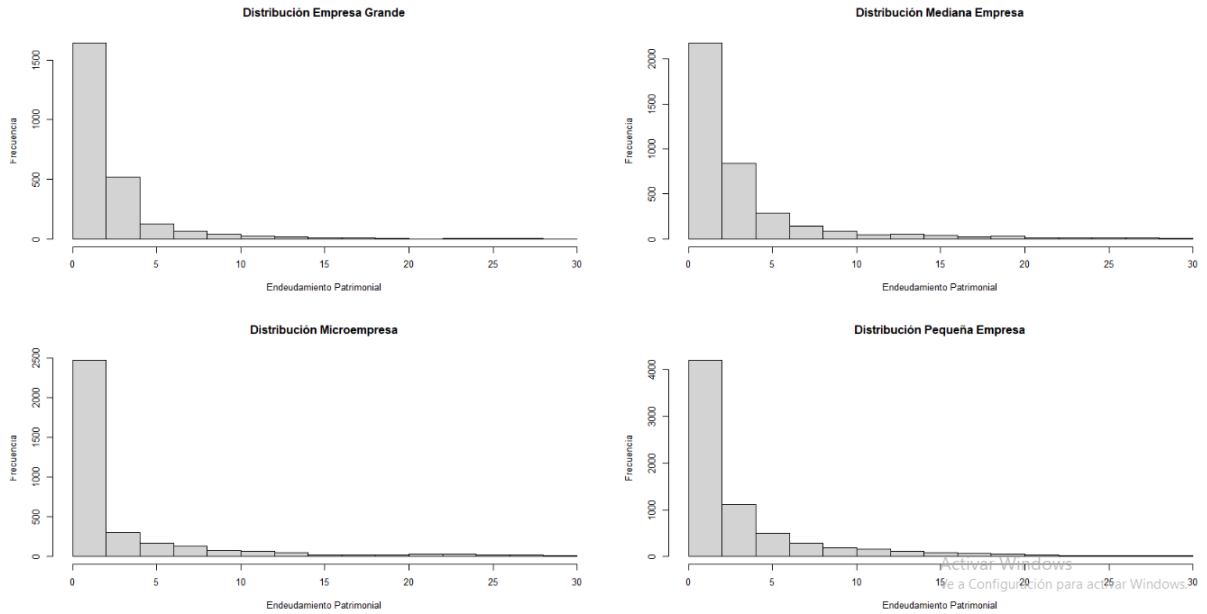


(a) Distribución de la Prueba Ácida por tamaño empresarial con tratamiento sobre atípicos

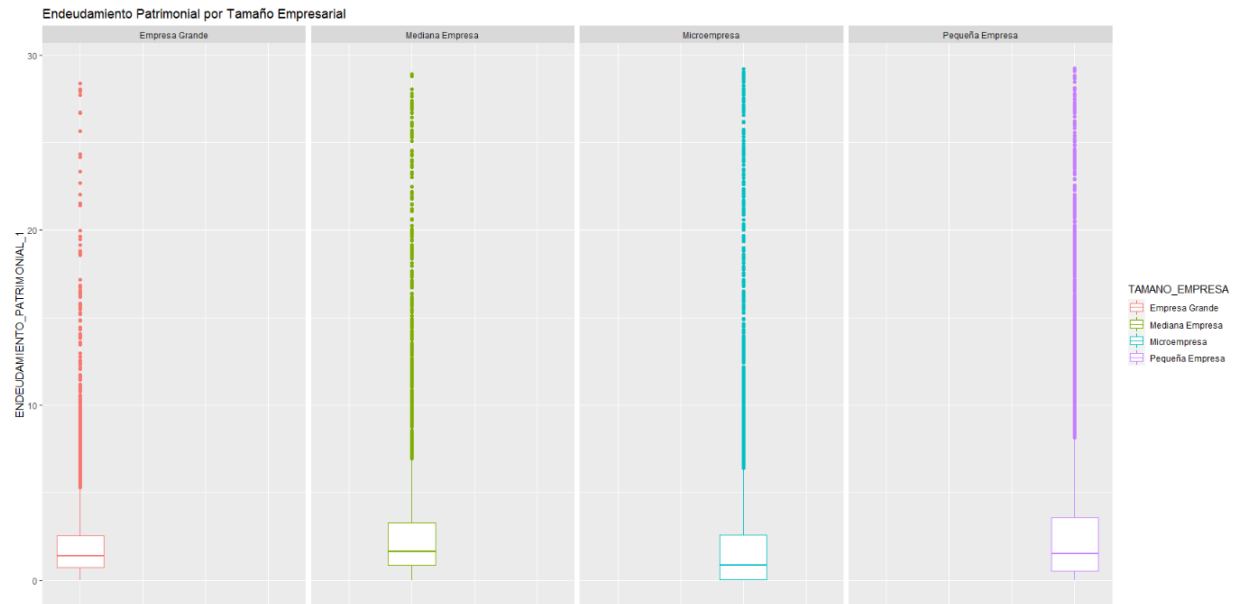


(b) Detección de atípicos de la Prueba Ácida por tamaño empresarial con tratamiento de atípicos

Figura B.2: Distribución y Diagrama de cajas de la variable *Endeudamiento Patrimonial* tratada

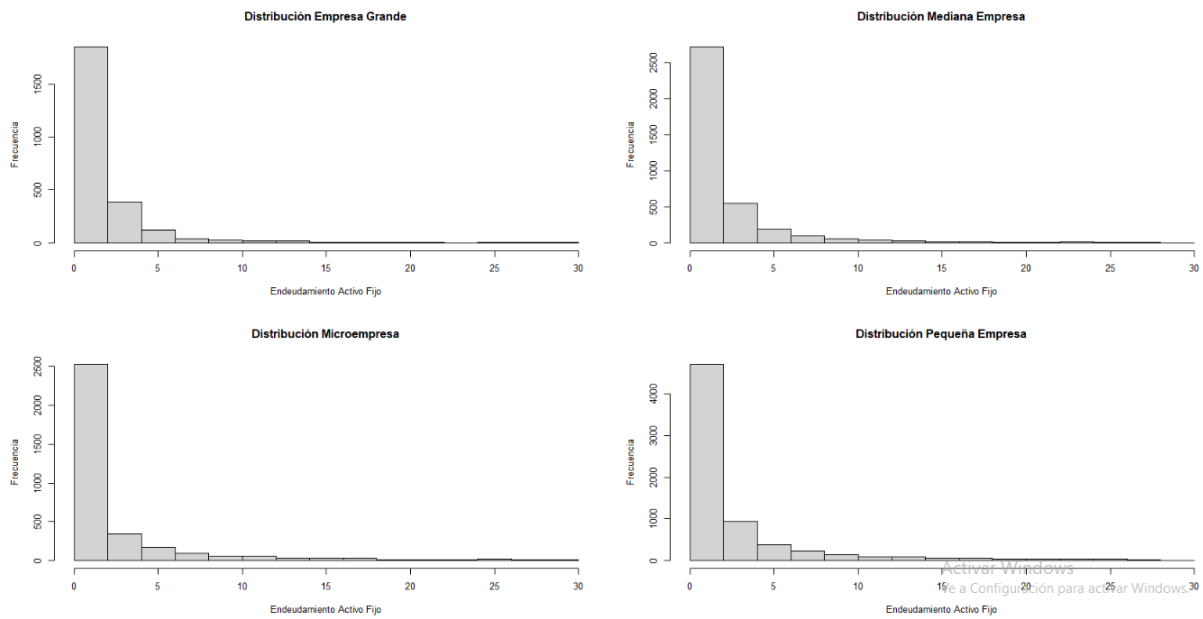


(a) Distribución del Endeudamiento Patrimonial por tamaño empresarial con tratamiento sobre atípicos

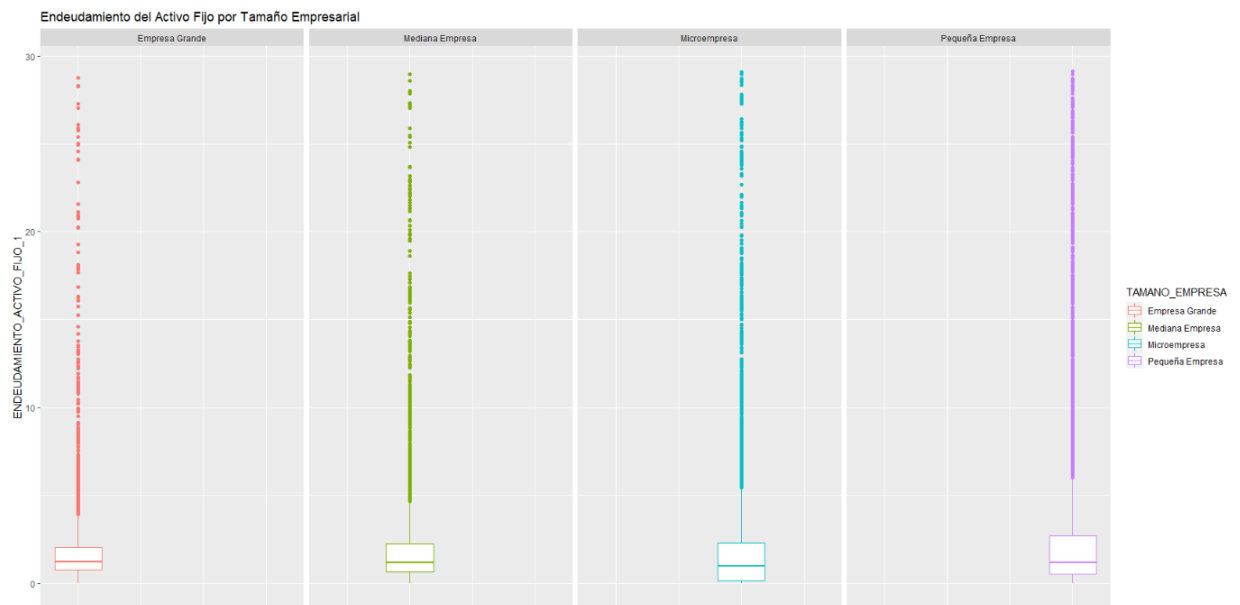


(b) Detección de atípicos del Endeudamiento Patrimonial por tamaño empresarial con tratamiento de atípicos

Figura B.3: Distribución y Diagrama de cajas de la variable *Endeudamiento Activo Fijo* tratada

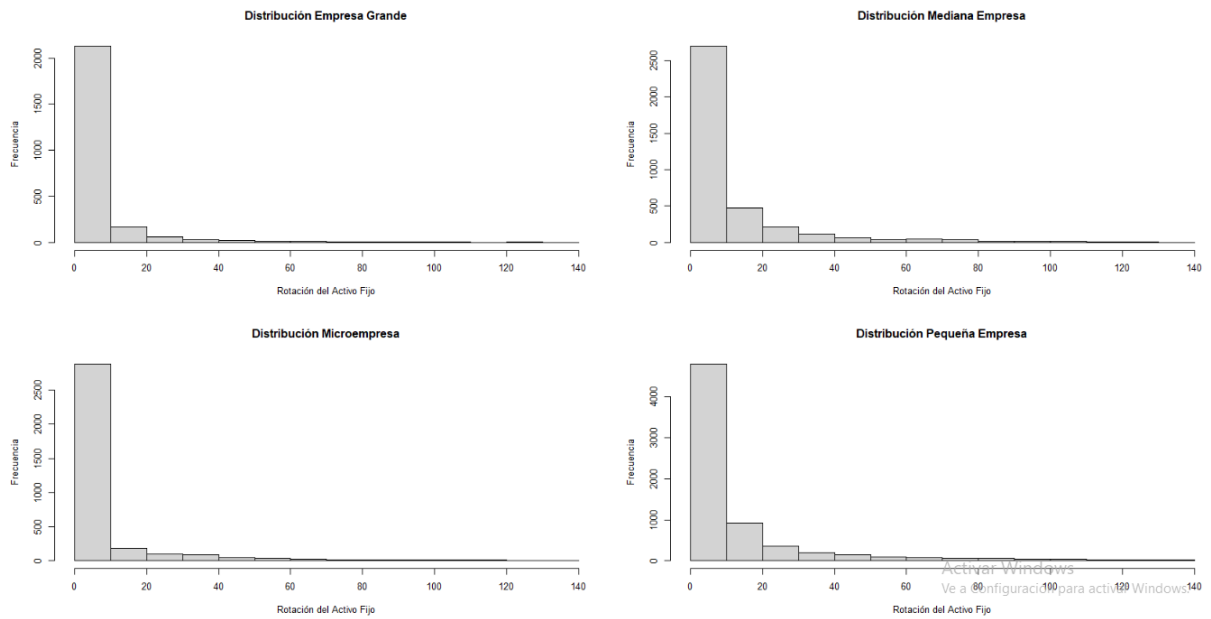


(a) Distribución del Endeudamiento Activo Fijo por tamaño empresarial con tratamiento sobre atípicos

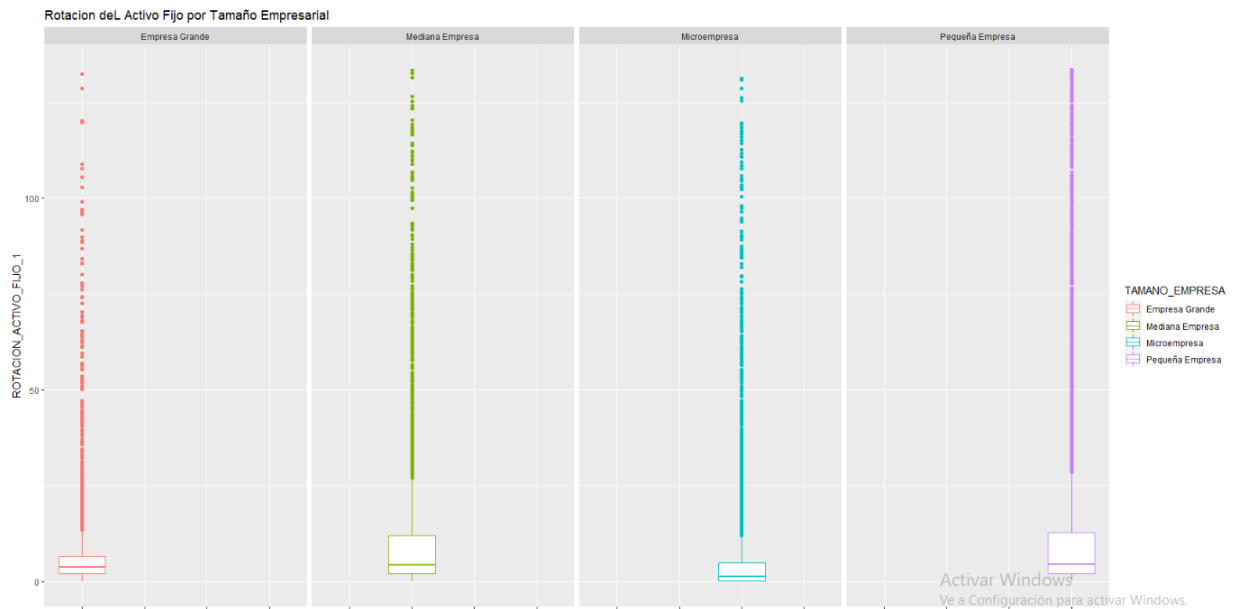


(b) Detección de atípicos del Endeudamiento Activo Fijo por tamaño empresarial con tratamiento de atípicos

Figura B.4: Distribución y Diagrama de cajas de la variable *Rotación del Activo Fijo* tratada

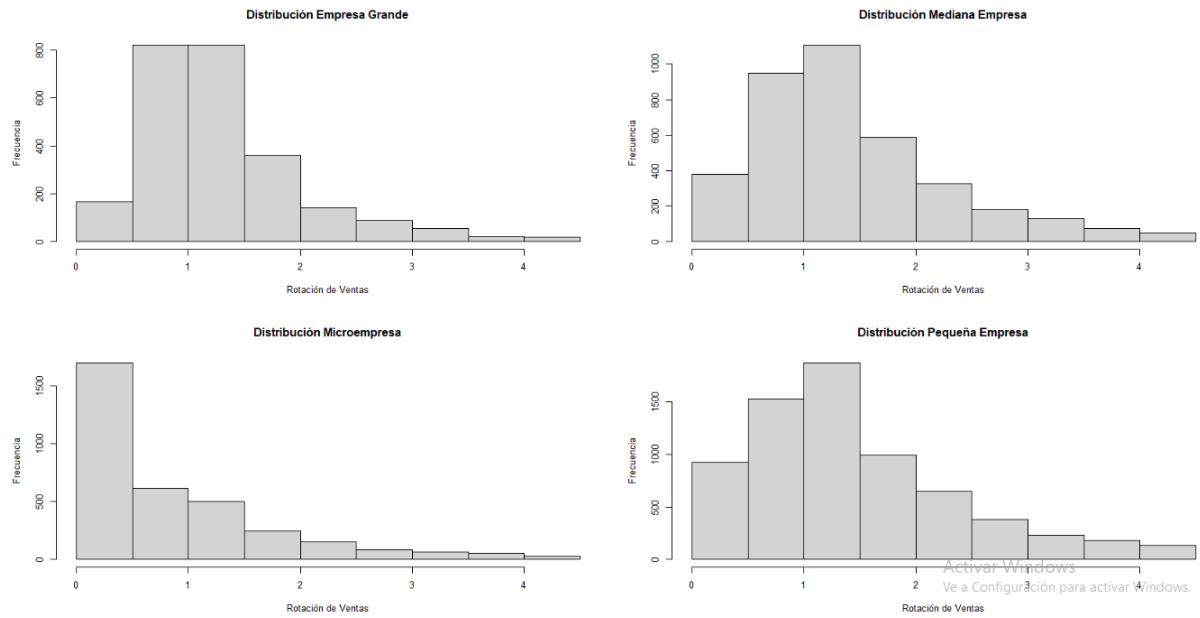


(a) Distribución de la Rotación del Activo Fijo por tamaño empresarial con tratamiento sobre atípicos

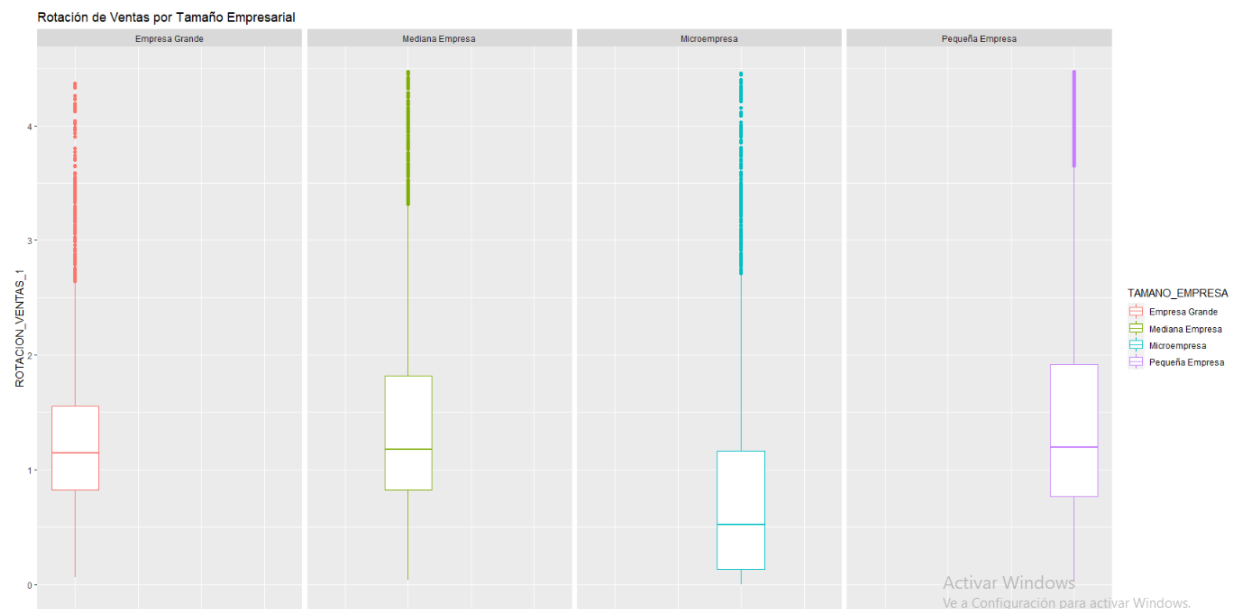


(b) Detección de atípicos de la Rotación del Activo Fijo por tamaño empresarial con tratamiento de atípicos

Figura B.5: Distribución y Diagrama de cajas de la variable *Rotación de Ventas* tratada

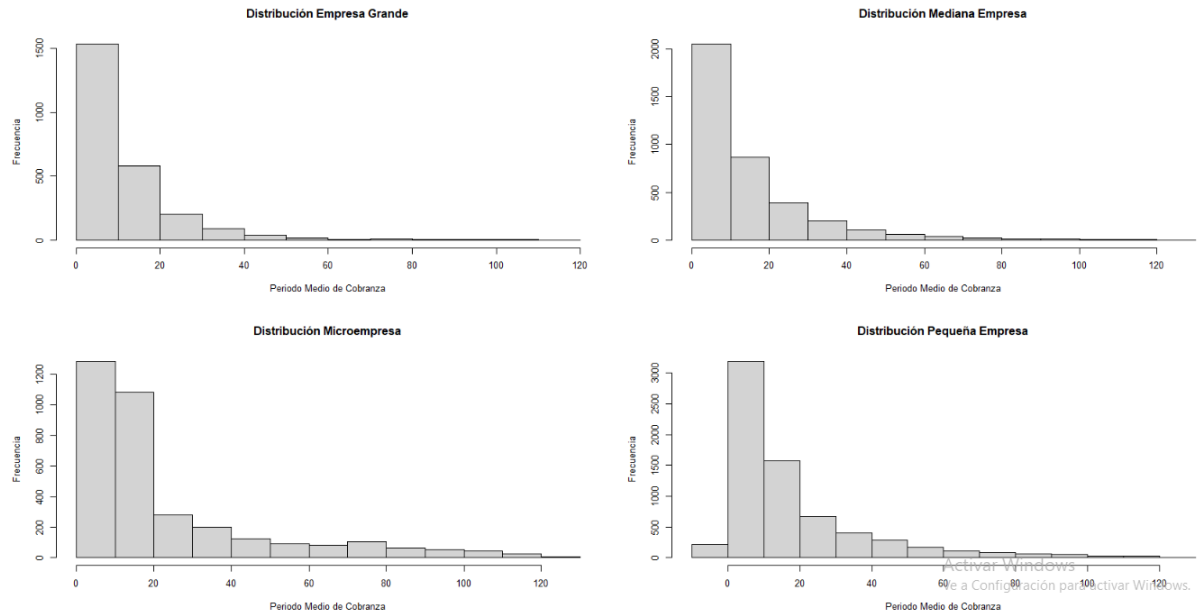


(a) Distribución de la Rotación de Ventas por tamaño empresarial con tratamiento sobre atípicos

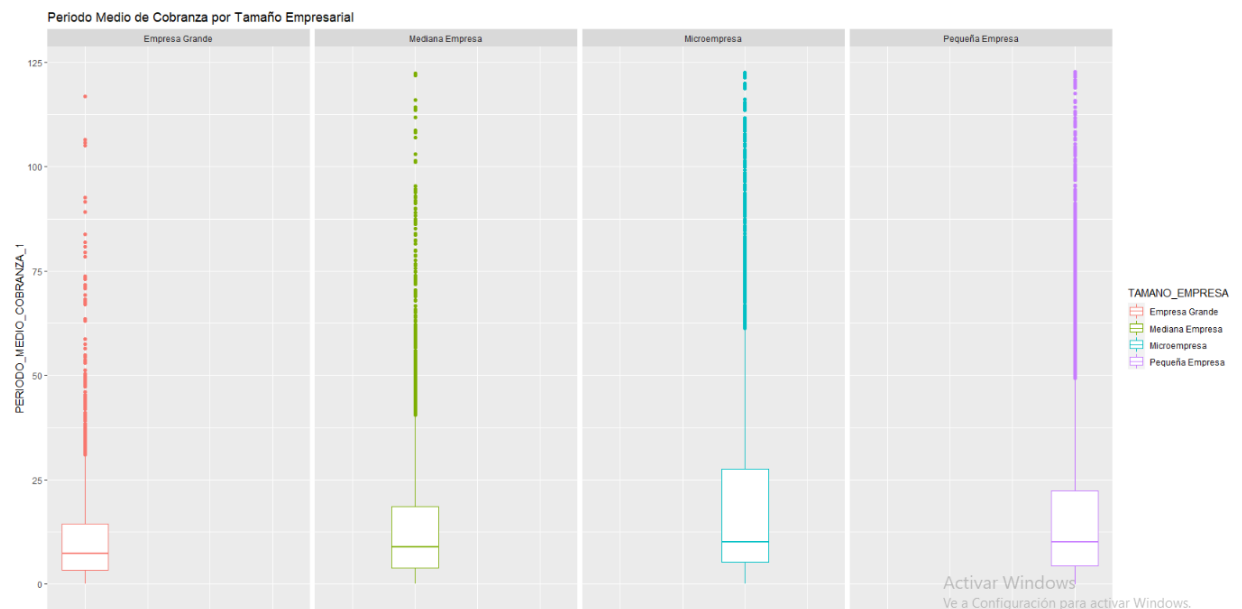


(b) Detección de atípicos de la Rotación de Ventas por tamaño empresarial con tratamiento de atípicos

Figura B.6: Distribución y Diagrama de cajas de la variable *Periodo medio de cobranza* tratada



(a) Distribución del Periodo medio de cobranza por tamaño empresarial con tratamiento sobre atípicos



(b) Detección de atípicos del Periodo medio de cobranza por tamaño empresarial con tratamiento de atípicos

Apéndice C

Promedio del riesgo de quiebra por Provincia y tamaño empresarial

Cuadro C.1: Promedio del riesgo de quiebra por segmento de riesgo en el año 2016

| Provincia | Segura | Riesgo Normal | Alerta | Riesgo |
|---------------------------------------|---------------|----------------------|---------------|---------------|
| AZUAY | 6,0 % | | 32,0 % | 54,0 % |
| BOLIVAR | | 13,0 % | | |
| CAÑAR | 4,0 % | 15,0 % | | |
| CARCHI | 2,0 % | 10,0 % | | 75,0 % |
| CHIMBORAZO | 4,0 % | | 32,0 % | |
| COTOPAXI | 3,0 % | | 37,0 % | |
| EL ORO | 1,0 % | 18,0 % | 38,0 % | |
| ESMERALDAS | 1,0 % | 13,0 % | 26,0 % | 64,0 % |
| GALAPAGOS | | | 21,0 % | |
| GUAYAS | | 15,0 % | 34,0 % | |
| IMBABURA | 1,0 % | 16,0 % | 32,0 % | 44,0 % |
| LOJA | 3,0 % | | | 47,0 % |
| LOS RIOS | 3,0 % | | 30,0 % | |
| MANABI | | 10,0 % | 35,0 % | |
| MORONASANTIAGO | 8,0 % | | | |
| NAPO | | | | 99,0 % |
| ORELLANA | 5,0 % | | | 68,0 % |
| PASTAZA | 6,0 % | | | 83,0 % |
| PICHINCHA | | 16,0 % | 34,0 % | 51,0 % |
| SANTA ELENA | | | 36,0 % | 99,0 % |
| STO. DOMINGO DE LOS TSACHI- LAS | | | 36,0 % | |
| SUCUMBIOS | | 18,0 % | 31,0 % | |
| TUNGURAHUA | 8,0 % | | 32,0 % | |
| ZAMORA CHINCHIPE | | | | |

Cuadro C.2: Promedio del riesgo de quiebra por segmento de riesgo en el año 2017

| Provincia | Segura | Riesgo Normal | Alerta | Riesgo |
|---------------------------------------|--------|---------------|--------|--------|
| AZUAY | 9,0 % | | 34,0 % | 44,0 % |
| BOLIVAR | | 12,0 % | | |
| CAÑAR | 1,0 % | | 35,0 % | 99,0 % |
| CARCHI | | 17,0 % | | 51,0 % |
| CHIMBORAZO | 7,0 % | 13,0 % | | |
| COTOPAXI | 8,0 % | | 37,0 % | 58,0 % |
| EL ORO | 6,0 % | 12,0 % | | |
| ESMERALDAS | 7,0 % | | | 63,0 % |
| GALAPAGOS | 0,0 % | | | |
| GUAYAS | | 21,0 % | 28,0 % | 44,0 % |
| IMBABURA | 2,0 % | | 29,0 % | 61,0 % |
| LOJA | 3,0 % | 16,0 % | 35,0 % | 60,0 % |
| LOS RIOS | 3,0 % | | 34,0 % | 53,0 % |
| MANABI | | 16,0 % | 26,0 % | 50,0 % |
| MORONASANTIAGO | 4,0 % | | | |
| NAPO | | | | 99,0 % |
| ORELLANA | 2,0 % | | | |
| PASTAZA | | 15,0 % | | 99,0 % |
| PICHINCHA | | 12,0 % | 31,0 % | 55,0 % |
| SANTA ELENA | | | | 99,0 % |
| STO. DOMINGO DE LOS TSACHI- LAS | 1,0 % | | 34,0 % | |
| SUCUMBIOS | 1,0 % | | 35,0 % | |
| TUNGURAHUA | | 11,0 % | 30,0 % | 50,0 % |
| ZAMORA CHINCHIPE | | | | 66,0 % |

Cuadro C.3: Promedio del riesgo de quiebra por segmento de riesgo en el año 2018

| Provincia | Segura | Riesgo Normal | Alerta | Riesgo |
|------------------|--------|---------------|--------|--------|
| AZUAY | | 14,0 % | 38,0 % | 60,0 % |
| BOLIVAR | | | 27,0 % | |
| CAÑAR | 2,0 % | | 35,0 % | |
| CARCHI | 5,0 % | 12,0 % | | |
| CHIMBORAZO | 1,0 % | 20,0 % | | 66,0 % |
| COTOPAXI | 3,0 % | 16,0 % | 35,0 % | |
| EL ORO | 1,0 % | 15,0 % | 34,0 % | |
| ESMERALDAS | 1,0 % | | | 68,0 % |
| GALAPAGOS | 1,0 % | | | |
| GUAYAS | | 21,0 % | 35,0 % | |
| IMBABURA | 1,0 % | | 39,0 % | 44,0 % |
| LOJA | 3,0 % | | 35,0 % | 44,0 % |
| LOS RIOS | | 12,0 % | 32,0 % | |
| MANABI | | 17,0 % | 31,0 % | 48,0 % |
| MORONASANTIAGO | 1,0 % | | 42,0 % | |
| NAPO | | | 29,0 % | |
| ORELLANA | 2,0 % | | | 99,0 % |
| PASTAZA | 2,0 % | | | |
| PICHINCHA | | 14,0 % | 32,0 % | 55,0 % |
| SANTA ELENA | 7,0 % | | | 99,0 % |
| STO. DOMINGO DE | 1,0 % | | 41,0 % | 58,0 % |
| LOS TSACHILAS | | | | |
| SUCUMBIOS | 4,0 % | | 29,0 % | |
| TUNGURAHUA | | | 28,0 % | 47,0 % |
| ZAMORA CHINCHIPE | 1,0 % | | | |

Cuadro C.4: Promedio del riesgo de quiebra por segmento de riesgo en el año 2019

| Provincia | Segura | Riesgo Normal | Alerta | Riesgo |
|--|--------|---------------|--------|--------|
| AZUAY | 6,0 % | | 24,0 % | 47,0 % |
| BOLIVAR | | | | |
| CAÑAR | 3,0 % | | 43,0 % | |
| CARCHI | 3,0 % | | | |
| CHIMBORAZO | | 10,0 % | 24,0 % | |
| COTOPAXI | 2,0 % | | 25,0 % | 58,0 % |
| EL ORO | 1,0 % | 10,0 % | 36,0 % | |
| ESMERALDAS | 4,0 % | | | 99,0 % |
| GALAPAGOS | 0,0 % | | | |
| GUAYAS | | 17,0 % | 32,0 % | |
| IMBABURA | 2,0 % | | 37,0 % | |
| LOJA | | | 31,0 % | |
| LOS RIOS | 3,0 % | 19,0 % | | |
| MANABI | | | 28,0 % | 48,0 % |
| MORONASANTIAGO | | | | 99,0 % |
| NAPO | | | 25,0 % | |
| ORELLANA | 2,0 % | | | 84,0 % |
| PASTAZA | | 11,0 % | 22,0 % | |
| PICHINCHA | | 13,0 % | 31,0 % | 53,0 % |
| SANTA ELENA | | 15,0 % | | 88,0 % |
| STO. DO- MINGO DE LOS TSA- CHILAS | 1,0 % | | 29,0 % | 46,0 % |
| SUCUMBIOS | 0,0 % | 10,0 % | 28,0 % | |
| TUNGURAHUA | | 10,0 % | 31,0 % | |
| ZAMORA CHINCHIPE | | | | |

Cuadro C.5: Promedio del riesgo de quiebra por segmento de riesgo en el año 2020

| Provincia | Segura | Riesgo Normal | Alerta | Riesgo |
|--|--------|---------------|--------|--------|
| AZUAY | 8,0 % | | 30,0 % | 62,0 % |
| BOLIVAR | | | | |
| CAÑAR | 4,0 % | | | 99,0 % |
| CARCHI | 4,0 % | 12,0 % | | |
| CHIMBORAZO | 1,0 % | | 29,0 % | |
| COTOPAXI | 1,0 % | | 35,0 % | |
| EL ORO | 1,0 % | 17,0 % | 33,0 % | |
| ESMERALDAS | 3,0 % | | 40,0 % | |
| GALAPAGOS | | 11,0 % | | |
| GUAYAS | | 19,0 % | 31,0 % | |
| IMBABURA | 1,0 % | 16,0 % | 24,0 % | |
| LOJA | | 14,0 % | | 60,0 % |
| LOS RIOS | 3,0 % | | 24,0 % | 63,0 % |
| MANABI | | | 31,0 % | 55,0 % |
| MORONASANTIAGO | | | | |
| NAPO | | | | 99,0 % |
| ORELLANA | 3,0 % | | 29,0 % | |
| PASTAZA | | 19,0 % | | |
| PICHINCHA | | 13,0 % | 30,0 % | 54,0 % |
| SANTA ELENA | 2,0 % | | | |
| STO. DO- MINGO DE LOS TSA- CHILAS | 3,0 % | | 36,0 % | |
| SUCUMBIOS | 2,0 % | | 42,0 % | |
| TUNGURAHUA | | 15,0 % | 33,0 % | 70,0 % |
| ZAMORA CHINCHIPE | | | | |

Cuadro C.6: Riesgo de quiebra por segmento de riesgo en las grandes empresas

| PROVINCIA | Empresas Grandes | | | |
|------------------|------------------|---------------|---------|--------|
| | Segura | Riesgo Normal | Alerta | Riesgo |
| AZUAY | | 14,20 % | | |
| BOLIVAR | | | | |
| CAÑAR | 2,20 % | | | |
| CARCHI | | | | 99 % |
| COTOPAXI | 4,10 % | | | |
| CHIMBORAZO | 9,40 % | | | |
| EL ORO | 8,25 % | | | |
| ESMERALDAS | 3,20 % | | | |
| GUAYAS | | 15,30 % | | |
| IMBABURA | 2,30 % | | | |
| LOJA | | | 27,40 % | |
| LOS RIOS | | | 43,20 % | |
| MANABI | | 10,30 % | | |
| MORONA SANTIAGO | | | | |
| NAPO | | | | |
| PASTAZA | | | | |
| PICHINCHA | | 16,40 % | | |
| TUNGURAHUA | | | 25,70 % | |
| ZAMORA CHINCHIPE | | | | |
| GALAPAGOS | | | | |
| SUCUMBIOS | | | | |
| ORELLANA | | | | |
| SANTO DOMINGO | 1 % | | | |
| DE LOS TSACHI- | | | | |
| LAS | | | | |
| SANTA ELENA | | | | |

Cuadro C.7: Riesgo de quiebra por segmento de riesgo en las Medianas empresas

| PROVINCIA | Mediana Empresa | | | |
|------------------|-----------------|---------------|---------|---------|
| | Segura | Riesgo Normal | Alerta | Riesgo |
| AZUAY | | | 26,20 % | |
| BOLIVAR | | | | |
| CAÑAR | | | 35,30 % | |
| CARCHI | 1,10 % | | | |
| COTOPAXI | | | 33,30 % | |
| CHIMBORAZO | | 15,40 % | | |
| EL ORO | | | 35,20 % | |
| ESMERALDAS | | | | 60 % |
| GUAYAS | | | 26,70 % | |
| IMBABURA | | | 32,60 % | |
| LOJA | | | 38,40 % | |
| LOS RIOS | | | 32,40 % | |
| MANABI | | | 26,70 % | |
| MORONA SANTIAGO | | | | |
| NAPO | | | | |
| PASTAZA | | 11,40 % | | |
| PICHINCHA | | | 26,40 % | |
| TUNGURAHUA | | | 27,20 % | |
| ZAMORA CHINCHIPE | | | | |
| GALAPAGOS | | | | |
| SUCUMBIOS | 1.40 % | | | |
| ORELLANA | | | | 84,30 % |
| SANTO DOMINGO | | | | |
| DE LOS TSACHILAS | | | | |
| SANTA ELENA | | | | 99,0 % |

Cuadro C.8: Riesgo de quiebra por segmento de riesgo en las Pequeñas empresas

| PROVINCIA | Pequeña Empresa | | | |
|------------------|-----------------|---------------|---------|---------|
| | Segura | Riesgo Normal | Alerta | Riesgo |
| AZUAY | | | 40,00 % | |
| BOLIVAR | | 13,40 % | | |
| CAÑAR | | | 35,00 % | |
| CARCHI | | | 76 % | |
| COTOPAXI | | | 30,2 % | |
| CHIMBORAZO | | 16,70 % | | |
| EL ORO | | 16,40 % | | |
| ESMERALDAS | | | 13,40 % | |
| GUAYAS | | | 31,20 % | |
| IMBABURA | | | 34,40 % | |
| LOJA | | | 30 % | |
| LOS RIOS | | | 12,70 % | |
| MANABI | | | 35,70 % | |
| MORONA SANTIAGO | 4,40 % | | | |
| NAPO | | | | 99 % |
| PASTAZA | | 15.10 % | | |
| PICHINCHA | | | 37,80 % | |
| TUNGURAHUA | | | | 46,20 % |
| ZAMORA CHINCHIPE | | | | |
| GALAPAGOS | | | 11,60 % | |
| SUCUMBIOS | | | | |
| ORELLANA | | | 29,60 % | |
| SANTO DOMINGO | | | 36,40 % | |
| DE LOS TSACHILAS | | | | |
| SANTA ELENA | | | | 76,5 % |

Cuadro C.9: Riesgo de quiebra por segmento de riesgo en las Microempresas

| PROVINCIA | MicroEmpresa | | | |
|------------------|--------------|---------------|---------|---------|
| | Segura | Riesgo Normal | Alerta | Riesgo |
| AZUAY | | | | 54,80 % |
| BOLIVAR | | | | |
| CAÑAR | | | 43,50 % | |
| CARCHI | | 13,40 % | | |
| COTOPAXI | | | 40,00 % | |
| CHIMBORAZO | | | 43,80 % | |
| EL ORO | | | 36,30 % | |
| ESMERALDAS | | | 33,25 % | |
| GUAYAS | | | 42,74 % | |
| IMBABURA | | | 32,41 % | |
| LOJA | | | 30 % | |
| LOS RIOS | | | 42,60 % | |
| MANABI | | | 40,00 % | |
| MORONA SANTIAGO | | | 42,23 % | |
| NAPO | | | 27,12 % | |
| PASTAZA | | | 22,50 % | |
| PICHINCHA | | | | 54,60 % |
| TUNGURAHUA | | | | |
| ZAMORA CHINCHIPE | | | | 66,30 % |
| GALAPAGOS | | | | |
| SUCUMBIOS | | | 33,40 % | |
| ORELLANA | | | | 84,70 % |
| SANTO DOMINGO | | | 39,60 % | |
| DE LOS TSACHILAS | | | | |
| SANTA ELENA | | | 36,9 % | |

Bibliografía

- [1] *Generalized Additive Models*, Chapman and Hall, Great Britain, 1990.
- [2] K. AHMAD AHMADPOUR, S. SEYYED HASAN, AND E. FATEMEH, *La predicción de bancarrota mediante Redes Neuronales y Regresión logística*, Universidad de Mazandaran, Iran, 2013.
- [3] E. ALTMAN, *Financial ratios, discriminant analysis and the prediction of corporate bankruptcy.*, *The Journal of Finance*, 23 (1968), pp. 589–609.
- [4] F. A. AMIR, *Bankruptcy Prediction for Credit Risk Using Neuronal Networks: A Survey and New Results*, *IEEE TRANSACTIONS ON NEURAL NETWORKS*, 12.
- [5] R. ANDERSON, *The Credit scoring toolkit: Theory and practice for retail credit risk management and decision automation*, Oxford University Press, 2007.
- [6] M. BAQUE, B. CEDEÑO, J. CHELE, AND V. GAONA, *Fracaso de las pymes: Factores desencadenantes, Ecuador 2020*, FIPCAEC (Edición 22), 5 (4), pp. 3–25.
- [7] W. BEAVER, *Financial ratios as predictors of failures.*, *Journal of Accounting Research*, 4 (1966), pp. 71–111.
- [8] D. BERG, *Bankruptcy prediction by generalized additive models*, *Applied Stochastic Models in Business and Industry*, 23 (2007), pp. 129–143.
- [9] A. BERGER AND G. UDELL, *The economics of Small Business finance: the roles equity and debt markets in the financial growth cycle*, *Journal at Banking and Finance*, 2 (1998), pp. 73–613.
- [10] B. BERNANKE, *Non-Monetary Effects or the Financial Crisis in the Propagation of the Great Depression*, *The American Economic Review*, 73 (3) (1983), pp. 257 – 276.

- [11] B. BERNANKE AND M. GERTLER, *Inside the black box: The credit channel of monetary policy transmission*, Journal of Economic Perspectives, 9 (4) (1995), pp. 27 – 48.
- [12] M. BLUM, *Failing Company Discriminant Analysis*, Journal of Accounting Research, (1974), pp. 1 – 25.
- [13] M. BREALEY, S. MYERS, AND MARCUS, *Principios de dirección financiera*, Madrid, España: McGraw Hill, 1996.
- [14] L. BREIMAN, J. H. FRIEDMAN, R. A. OLSHEN, AND C. STONE, *Classification and Regression Trees*, Wadsworth, Belmont, CA, 1984.
- [15] E. CALDERÓN, *Evaluación de los modelos de predicción de fracaso empresarial en el sector manufacturero colombiano en los años 2010-2014*, Master's thesis, Universidad Nacional de Colombia, Bogotá, 2016.
- [16] H. CAPA, *Probabilidades y Estadística: Para una gestión científica de la información*, Escuela Politécnica Nacional, (2016).
- [17] N. CARO, M. DÍAZ, AND M. PORPORATO, *Revista de métodos cuantitativos para la economía y la empresa*, Revista de Métodos Cuantitativos para la Economía y la Empresa, 16 (2013), pp. 200–2015.
- [18] Y. CASTRO, A. HUERTAS, E. OBANDO, AND F. VALENCIA, *Análisis de Supervivencia para predicción de bancarrota: Caso de las industrias minoristas de Colombia*, Revista Espacios, 40 (2019), p. 18.
- [19] R. CELAYA AND M. LÓPEZ, *¿Cómo determinar su riesgo empresarial?*, Escuela de Administración de Negocios, 52 (2004), pp. 69–75.
- [20] B. CENTRAL DEL ECUADOR, *Evaluación impacto macroeconómico del covid-19 en la economía ecuatoriana*, https://contenido.bce.fin.ec//documentos/publicacionesnotas//impmaccovid_122020.pdf, (2020), pp. 16–17.
- [21] N. CRUTZEN AND D. V. CAILLE, *The Business Failure Process. An Integrative Model of the Literature*, Review of Business and Economic Literature, (2008), pp. 287–316.
- [22] E. D. CRUZ, *Teoría de Riesgo (4. ed.)*, Bogotá: Ecoe Ediciones, 2016.

- [23] I. CÁMARA DE INDUSTRIAS Y PRODUCCIÓN, *Proyecciones económicas 2021*, <https://www.cip.org.ec/2021/01/14/proyecciones-economicas-2021/>, (2021).
- [24] J. DAMIÁN, *Redes Neuronales: Conceptos básicos y aplicaciones*, PhD thesis, Universidad Tecnológica Nacional – Facultad Regional Rosario Departamento de Ingeniería Química Grupo de Investigación Aplicada a la Ingeniería Química (GIAIQ), 2001.
- [25] E. DEAKIN, *Research Reports A Discriminant Analysis of Predictors of Business Failure.*, *Journal of Accountin Research*, 10(1) (1972), pp. 167–179.
- [26] M. DIETSCH AND E. KREMP, *El crédito entre empresas beneficia más a las grandes empresas que a las Pymes*, *Economía y Estadística*, 314 (1988), pp. 25–37.
- [27] M. DURBÁN, *Modelos Aditivos Generalizados con P-splines*, Universidad Carlos III de Madrid, Madrid, España, 2015.
- [28] S. DURBÁN, *Dirección Financiera*, McGraw-Hill Companies, 2008.
- [29] R. EDMISTER, *An Empirical Test of Financial Ratio Analysis for Small Bussines Failure Prediction*, *Financial and Quantitative Analysis*, (1972), pp. 1477 – 1493.
- [30] M. A. ESTALLO AND F. GINER, *Cómo crear y hacer funcionar una empresa*, Madrid, España: ESIC, 2010.
- [31] S. FERNANDO, *Redes neuronales: una visión superficial*, (2019).
- [32] P. FITZPATRICK, *A comparison of ratios of successful industrial enterprises with those of falied companies*, *The certified Public Accountant*, (1932), pp. 598–731.
- [33] R. FLÓREZ AND J. FERNÁNDEZ, *Las Redes Neuronales Artificiales. Fundamentos teóricos y aplicaciones prácticas*, Ed. Netbiblo, Coruña, España, 2008.
- [34] A. GINANNI, C. ROMANI, P. GONZALES AROCA, N. AGUIRRE AGUIERRE, P. VEGA LEITON, AND J. CARRAZANA MUÑOZ, *Modelos de clasificación y predicción de quiebras de empresas; una aplicación a empresas chilenas*, *Forum empresarial*, 7(1), San Juan, Puerto Rico.
- [35] L. GITMAN AND M. JOEHNK, *Fundamentos de inversiones (10. ed.)*, Pearson educación, 2009.
- [36] C. GU, *Smoothing Spline ANOVA Models*, Springer, New York, 2002.

- [37] N. GUJARATI DAMODAR AND C. PORTER DAWN, *Econometría*, 5ta ed. México: McGraw-Hill, 2010.
- [38] M. HERNÁNDEZ RAMÍREZ, *Modelo Financiero para la detección de Quiebras con el uso de análisis discriminante múltiple*, *InterSedes*, 15(32) (2014), pp. 4–19.
- [39] N. JACKENDOFF, *A study of Published Industry Financial and Operating Ratios*, Temple University, Bureau of Economic and Business Research, Philadelphia, University of Michigan, 1962.
- [40] G. JAMES, D. WITTEN, T. HASTIE, AND R. TIBSHIRANI, *An Introduction to Statistical Learning with Applications in R*, Springer Publishing Company, 2014.
- [41] S. JONES AND D. A. HENSHER, *Predicting firm financial distress: A mixed logit model.*, *Accounting Review*, 29(4) (2004), pp. 1011–1038.
- [42] D. JUMBO, J. CAMPUZANO, F. VEGA, AND A. LUNA, *Crisis económicas y covid-19 en Ecuador: Impacto en las exportaciones*, *Revista Universidad y Sociedad*, 12 (6) (2020).
- [43] A. JUNG AND I. PEÑA, *Factores asociados al fracaso de nuevas empresas*, III Conferencia internacional de Investigadores en Emprendedurismo de América Latina, (2004), pp. 56 – 58.
- [44] M. LASTRE, A. LASTRE, AND G. GARCÍA, *Redes Neuronales en la predicción de insolvencia. Un cambio de paradigma ante recetas tradicionales de prácticas empresariales*, Universidad Tecnológica Equinoccial, Quito, Ecuador.
- [45] K. LUCERO, *La manufactura, una tabla de salvación en medio de la crisis*, *Revista Gestión*, (2020).
- [46] M. LÓPEZ, C. SÁNCHEZ, AND P. MONELOS, *Predicción de insolvencia y fracaso financiero: medio siglo después de Beaver (1966). Avneces y nuevos resultados*, University of A Coruña., 2015.
- [47] E. MANSFIELD AND H. B, *Detecting multicollinearity*, *The American Statistician*, 36 (3a) (1982), pp. 158 – 160.
- [48] C. MERWIN, *Financing Small Corporations in Five Manufacturing Industries*, *Journal of the American Statistical Association*, 39 (225) (1942), pp. 129–130.
- [49] B. MÜLLER, J. REINHARDT, AND M. STRICKLAND, *Neural Networks: An Introduction*, Springer Science Business Media, 2012.

- [50] I. NACIONAL DE ESTADÍSTICAS Y CENSOS, *Manual de Usuario CIIU*, 2010.
- [51] ———, *Clasificación Nacional de actividades económicas CIIU Rev 4.0*, Unidad de análisis, (2012).
- [52] O. NACIONES UNIDAS, *Clasificación Industrial Uniforme de todas las actividades económicas CIIU Rev. 3.1*, (2005).
- [53] ———, *Clasificación Industrial Internacional Uniforme de todas las actividades económicas CIIU Rev. 4*, (2009).
- [54] E. NORDBY AND T. VIKEN, *Machine Learning in Bankruptcy Prediction*, Master's thesis, Norwegian School of Economics, Bergen, 2019.
- [55] A. NOVALES, *Econometría*, McGraw-Hill, 1993.
- [56] J. OHLSON, *Financial Ratios and the Probabilistic Prediction of Bankruptcy*, *Journal of Accounting Research*, 18(1) (1980), p. 109.
- [57] F. P. ORDÓÑEZ AND L. X. REYNA, *Riesgo de Insolvencia empresarial en el Ecuador durante el período 2012-2016 (Tesis de pregrado)*, Cuenca, Ecuador: Universidad de Cuenca, 2018.
- [58] I. ORELLANA, L. TONON, M. REYES, L. PINOS, AND E. CEVALLOS, *Riesgos financieros en el sector manufacturero del Ecuador*, Universidad del Azuay, Azuay, Ecuador, 2020.
- [59] I. SERVICIO DE RENTAS INTERNAS, *Estadísticas generales de recaudación sri*, <https://www.sri.gob.ec/estadisticas-generales-de-recaudacion-sri>, (2020).
- [60] S. SEUNG, *Multilayer Perceptron and Backpropagation learnign*, 9641 Lecture 4, MIT, 2002.
- [61] N. SIDDIQI, *Credit Risk scorecards: developing and implementing intelligent credit scoring*, Hoboken, N.J:Wiley, 2006.
- [62] I. SUPERINTENDENCIA DE COMPAÑÍAS VALORES Y SEGUROS, *Ranking empresarial: Empresas sujetas al control de la superintendencia de compañías, valores y seguros*, <https://appscvs.supercias.gob.ec/rankingcias/>, (2020).
- [63] R. TAFFLER, *Empirical model for the monitoring of UK corporations*, *Journal of Banking and Finance*, 8 (1984), pp. 1–29.

- [64] C. VALENCIA, S. CABRALES, L. GARCÍA, J. RAMIREZ, AND D. CALDERONA, *Generalized additive model with embedded variable selection for bankruptcy prediction: Prediction versus Interpretation*, *Cogent Economics & Finance*, (2019). 7:1, 1597956.
- [65] J. VARGAS, *Modelos de Beaver, Ohlson y Altman ¿Son realmente capaces de predecir la bancarrota en el sector empresarial costarricense?*, *TEC Empresarial*, 8 (2015), pp. 29–40.
- [66] S. N. WOOD, *Generalized Additive Models: An Introduction with R. 2nd ed.*, London: Chapman & Hall/CRC Press, 2017.
- [67] M. ZMIJWESKI, *Methodological Issues Related to the Estimation of Financial Distress Prediction Models.*, *Journal of Accounting Research*, 22 (1984), pp. 59–82.