


A dark blue vertical bar on the left side of the page. A blue arrow points to the right from this bar, containing the date.

5/1/2024

# Modelo de Credit Scoring con beneficios para posibles "unicornios".

Several thin, curved lines in shades of blue and grey that sweep upwards from the bottom left towards the center of the page.

Galermes, Joaquin  
Tutor: Chichel, Cristian Ezequiel  
Universidad del Salvador  
Carrera: Actuario

Primer cuatrimestre 2024

<b>Índice</b>	
<b>Resumen</b>	4
<b>Introducción</b>	5
<b>Capítulo 1: Contexto y Marco Teórico</b>	5
<b>Capítulo 2: Desarrollo y Aplicación de Modelos de Credit Scoring</b>	5
<b>Capítulo 3: Análisis del Modelo de Credit Scoring y Resultados</b>	6
<b>Capítulo 1: Contexto y Marco Teórico</b>	7
<b>Introducción al Credit Scoring</b>	7
<b>Contexto Histórico</b>	7
<b>Aplicaciones del Credit Scoring</b>	7
<b>Importancia del Credit Scoring</b>	7
<b>Evolución del Credit Scoring en Argentina</b>	8
<b>Limitaciones del Enfoque Convencional</b>	9
<b>Necesidad de un Enfoque Holístico</b>	10
<b>Teorías y Herramientas Relevantes</b>	11
<b>Teoría de Credit Scoring</b>	11
<b>Teoría de la Inteligencia Artificial</b>	11
<b>Herramientas Matemáticas Avanzadas</b>	12
<b>Modelos de Machine Learning</b>	12
<b>Análisis de Regresión</b>	12
<b>Modelos Bayesianos</b>	12
<b>Herramientas de Validación y Evaluación</b>	13
<b>Empresas “Unicornio”</b>	13
1. Innovación Tecnológica	13
2. Crecimiento Rápido	13
3. Financiamiento de Capital de Riesgo	14
4. Disrupción de Industrias Tradicionales	14
5. Equipo Fundador Talentoso	14
6. Enfoque en la Experiencia del Usuario	14
7. Internacionalización	14
8. Ecosistema y Red de Apoyo	14
<b>Z”-Score</b>	14
<b>Modelo de Flujo de Caja Descontado (DCF)</b>	15
<b>Modelo de Gordon-Shapiro</b>	16
Tasa de Crecimiento de Ventas	16
Tasa de Descuento (r)	16
Supuestos y Limitaciones	16
Aplicación Práctica	16
<b>Modelo de Credit Default Probabilities (CDP)</b>	17

<b>Capítulo 2: Desarrollo y Aplicación de Modelos de Credit Scoring</b>	18
<b>Metodología Utilizada</b>	18
<b>Inteligencia Artificial</b>	18
<b>Modelos de Machine Learning</b>	18
1. Random Forest	19
2. Gradient Boosting	19
3. Redes Neuronales Artificiales	19
4. Support Vector Machines (SVM)	19
<b>Análisis de Regresión</b>	19
1. Regresión Lineal	20
2. Regresión Logística	20
3. Regresión de Árboles de Decisión	20
4. Regresión Ridge y Lasso	20
<b>Análisis Tornado de Variables</b>	21
Proceso de Análisis Tornado	21
1. Identificación de Variables Importantes	21
2. Visualización de Resultados	21
3. Interpretación de Resultados	21
4. Refinamiento del Modelo	22
<b>Evaluación de los Modelos</b>	22
<b>Coeficiente de Gini</b>	22
1. Concepto del Coeficiente de Gini	22
2. Cálculo del Coeficiente de Gini	22
3. Interpretación del Coeficiente de Gini	23
4. Importancia del Coeficiente de Gini en el Credit Scoring	23
5. Uso del Coeficiente de Gini en la Evaluación de Modelos	23
<b>Curva ROC</b>	23
1. Concepto de la Curva ROC	24
2. Interpretación de la Curva ROC	24
3. Área bajo la Curva ROC (AUC)	24
4. Importancia de la Curva ROC en el Credit Scoring	24
5. Uso de la Curva ROC en la Evaluación de Modelos	24
<b>Error Cuadrático Medio (MSE)</b>	24
1. Concepto del Error Cuadrático Medio (MSE)	25
2. Interpretación del MSE	25
3. Importancia del MSE en el Credit Scoring	25
4. Uso del MSE en la Evaluación de Modelos	25
<b>Validación Cruzada</b>	25

1. Importancia de la Validación Cruzada .....	26
2. Proceso de la Validación Cruzada.....	26
3. Beneficios de la Validación Cruzada.....	26
<b>Capítulo 3: Análisis del Modelo de Credit Scoring y Resultados.....</b>	<b>28</b>
<b>Armado de la base de datos.....</b>	<b>28</b>
<b>Análisis y elección del modelo más adecuado .....</b>	<b>32</b>
Modelo y validación del modelo de Árbol de Decisión .....	32
Modelo y validación del Modelo Gradient Boosting .....	33
Modelo y validación del Modelo Redes Neuronales .....	33
Modelo y validación de Support Vector Machines.....	34
Modelo y validación de Ridge y Lasso.....	34
Modelo y validación de Regresión Logística.....	35
<b>Desarrollo del modelo.....</b>	<b>36</b>
<b>Resultados y muestras:.....</b>	<b>38</b>
<b>Conclusiones.....</b>	<b>40</b>
<b>Bibliografía.....</b>	<b>41</b>
<b>Agradecimientos .....</b>	<b>42</b>
<b>Anexo: .....</b>	<b>43</b>

**Resumen**

El presente trabajo final de grado de mi autoría se enfoca en desarrollar un modelo de Credit Scoring, es decir, una metodología de evaluación crediticia que consiste en asignar un puntaje al potencial deudor midiendo la solvencia del cliente pero con un foco innovador sin historial pasado, en donde también se puedan identificar posibles empresas “unicornios” en el contexto de la República Argentina para poder otorgarle mayores beneficios económicos a estas. Utilizando técnicas avanzadas de Machine Learning e Inteligencia Artificial, como Random Forest, Gradient Boosting y Redes Neuronales, el estudio busca superar las limitaciones de los enfoques tradicionales de Credit Scoring que dependen excesivamente de datos históricos.

La metodología incluye la recopilación y preprocesamiento de datos relevantes, la exploración y análisis de estos datos para identificar patrones, y la selección de variables clave para la construcción del modelo. Luego, se desarrollan y validan varios modelos predictivos, evaluando su precisión y robustez mediante métricas como el Coeficiente de Gini y la Curva ROC.

En resumen, este trabajo ofrece una herramienta avanzada para mejorar la toma de decisiones financieras y la evaluación del riesgo crediticio en un contexto empresarial en evolución. Como también beneficiar a posibles empresas “emergentes” de la nación con mayores límites crediticios para potenciarlas.

## **Introducción**

Este proyecto tiene como objetivo desarrollar un modelo de Credit Scoring innovador utilizando técnicas avanzadas de análisis predictivo, incluida la Inteligencia Artificial. A su vez, anticipar el valor futuro de empresas emergentes en el mercado argentino para poder otorgarle mayores beneficios económicos. La metodología a trabajar será:

- Recopilación y preprocesamiento de datos financieros y no financieros.
- Exploración y análisis de datos.
- Selección de variables relevantes.
- Desarrollo de modelo de Credit Scoring con Machine Learning e Inteligencia Artificial.
- Interpretación y validación de modelos.

Al combinar métodos tradicionales con enfoques innovadores, como modelos de Machine Learning y Análisis de Regresión, se busca mejorar la capacidad de las instituciones financieras para evaluar el riesgo crediticio y tomar decisiones informadas sobre la asignación de crédito y la inversión en el mercado.

El trabajo se dividirá en capítulos de la siguiente manera:

### **Capítulo 1: Contexto y Marco Teórico**

En este capítulo se proporcionará una introducción al tema del Credit Scoring y se establecerá el marco teórico necesario para comprender el proyecto. Se abordarán los siguientes aspectos:

1. **Introducción al Credit Scoring:** Se explicará la importancia del Credit Scoring en la evaluación del riesgo crediticio y la toma de decisiones financieras en instituciones financieras.
2. **Limitaciones del enfoque convencional:** Se discutirán las limitaciones del enfoque convencional en el análisis de créditos, especialmente en entornos empresariales dinámicos como el argentino.
3. **Necesidad de un enfoque holístico:** Se argumentará la necesidad de una evaluación más holística que no solo considere el pasado, sino que también tenga en cuenta las proyecciones y el potencial futuro de las empresas.
4. **Teorías y herramientas relevantes:** Se presentarán las teorías y herramientas matemáticas que constituirán el núcleo del trabajo, incluyendo la teoría de Credit Scoring, la teoría de la Inteligencia Artificial y otras herramientas matemáticas utilizadas en análisis predictivo.

### **Capítulo 2: Desarrollo y Aplicación de Modelos de Credit Scoring**

Este capítulo se centrará en el desarrollo y la aplicación de los modelos de Credit Scoring propuestos. Se incluirán los siguientes aspectos:

1. **Base de datos y variables utilizadas:** Se describirá la base de datos de empresas argentinas utilizada en el estudio, así como las variables financieras relevantes y los datos históricos y del mercado.
2. **Metodología utilizada:** Se explicarán en detalle las técnicas avanzadas de análisis predictivo empleadas, incluyendo la Inteligencia Artificial, modelos de Machine Learning y Análisis de Regresión.

3. **Evaluación de los modelos:** Se presentarán las métricas utilizadas para evaluar la precisión y eficacia de los modelos, incluyendo el coeficiente de Gini, la curva ROC, el error cuadrático medio (MSE) y la validación cruzada.

### **Capítulo 3: Análisis del Modelo de Credit Scoring y Resultados**

Este capítulo se dedicará exclusivamente a la prueba del modelo de Credit Scoring desarrollado y la presentación de los resultados obtenidos. Se incluirán los siguientes aspectos:

1. **Descripción de la prueba del modelo:** Se explicará el proceso de prueba del modelo, incluyendo la selección de datos de prueba y las metodologías utilizadas para evaluar la precisión del modelo.
2. **Resultados obtenidos:** Se presentarán los resultados de la prueba del modelo, incluyendo las métricas de evaluación de la precisión del modelo y cualquier otro resultado relevante.
3. **Análisis de resultados:** Se realizará un análisis detallado de los resultados obtenidos, destacando las fortalezas y limitaciones del modelo y proporcionando recomendaciones para su aplicación práctica en la toma de decisiones financieras y de inversión.
4. **Conclusiones y recomendaciones:** Se resumirán las conclusiones del estudio y se proporcionarán recomendaciones para futuras investigaciones en el campo del Credit Scoring y la evaluación del riesgo crediticio en el mercado argentino.

## **Capítulo 1: Contexto y Marco Teórico**

### **Introducción al Credit Scoring**

En el ámbito financiero, la evaluación del riesgo crediticio es una tarea esencial para las instituciones financieras y los prestamistas. Las instituciones financieras y los prestamistas enfrentan el desafío constante de predecir la probabilidad de incumplimiento de los prestatarios, una tarea crucial para la toma de decisiones informadas sobre la concesión de créditos (Gujarati, 2010)<sup>1</sup>. El Credit Scoring ha surgido como una herramienta fundamental para abordar esta necesidad, al proporcionar un enfoque sistemático y cuantitativo para evaluar el riesgo crediticio de los prestatarios.

También conocido como puntuación de crédito, es un método estadístico que permite a las entidades financieras asignar una puntuación a cada individuo o empresa solicitante de un crédito. Esta puntuación, basada en datos y análisis estadísticos, refleja la probabilidad de que el prestatario cumpla con sus obligaciones de pago (Finlay, 2004)<sup>2</sup>. La metodología del Credit Scoring se basa en la premisa de que ciertas variables financieras y no financieras, como el historial crediticio, los ingresos, el patrimonio neto y el comportamiento de pago, pueden ser indicadores confiables del riesgo crediticio.

### **Contexto Histórico**

El desarrollo del Credit Scoring tiene sus raíces en el siglo XX, cuando los prestamistas comenzaron a utilizar modelos estadísticos para evaluar el riesgo crediticio de los solicitantes. Sin embargo, su uso generalizado y su desarrollo significativo se produjeron en las últimas décadas, impulsados por avances tecnológicos y una mayor disponibilidad de datos financieros. Durante este tiempo, el Credit Scoring ha evolucionado desde enfoques simples basados en reglas hasta modelos más sofisticados basados en técnicas estadísticas y de aprendizaje automático. Este avance ha sido fundamental para mejorar la precisión y la eficiencia de la evaluación del riesgo crediticio.

En Argentina, el desarrollo del Credit Scoring ha seguido una trayectoria similar, aunque con algunas particularidades. La liberalización financiera y la globalización de los mercados financieros han impulsado la necesidad de evaluar el riesgo crediticio de manera más precisa y eficiente, lo que ha llevado a un mayor uso y desarrollo de modelos de Credit Scoring más avanzados y efectivos en el contexto argentino.

### **Aplicaciones del Credit Scoring**

El Credit Scoring se utiliza en una amplia gama de aplicaciones dentro del sector financiero, abarcando desde la toma de decisiones sobre préstamos comerciales y de consumo hasta la evaluación de la solvencia de los clientes y la gestión del riesgo de crédito. Además, su uso se ha extendido a otros sectores, como el de seguros y el comercio electrónico, donde se emplea para determinar primas de seguros, evaluar el riesgo de fraude y gestionar carteras de clientes. Esta amplia aplicación del Credit Scoring destaca su importancia y relevancia en el ámbito financiero y más allá, y subraya la necesidad de su continua evolución y desarrollo.

### **Importancia del Credit Scoring**

La importancia del Credit Scoring radica en su capacidad para proporcionar una evaluación objetiva y cuantitativa del riesgo crediticio, basada en datos y análisis estadísticos. Esto permite a las instituciones financieras y a los prestamistas tomar

---

<sup>1</sup> Gujarati, D. N. (2010). *Econometría* (5a.ed.--.).

<sup>2</sup> Finlay, S. (2004). *Credit Scoring, Response Modeling, and Insurance Rating: A Practical Guide to Forecasting Consumer Behavior*.



decisiones informadas y fundamentadas sobre la concesión de créditos, reduciendo el riesgo de pérdidas financieras y garantizando la estabilidad del sistema financiero. En entornos empresariales dinámicos como Argentina, donde las condiciones del mercado pueden cambiar rápidamente, la capacidad de evaluar el riesgo crediticio de manera precisa y eficiente es especialmente crucial.

Entre los principales beneficios se destacan:

- **Mejora en la toma de decisiones crediticias:** El Credit Scoring permite a las entidades financieras tomar decisiones más informadas sobre la concesión de créditos, reduciendo el riesgo de impagos y mejorando la rentabilidad (Hastie, 2009)<sup>3</sup>.
- **Mayor inclusión financiera:** La implementación del Credit Scoring facilita el acceso al crédito para un mayor número de individuos y empresas, especialmente aquellos con historial crediticio limitado o inexistente.
- **Estabilidad del sistema financiero:** La evaluación precisa del riesgo crediticio contribuye a la estabilidad del sistema financiero al reducir la morosidad y los riesgos sistémicos.
- **Eficiencia y automatización:** El Credit Scoring automatiza en gran medida el proceso de evaluación del riesgo crediticio, reduciendo costos y aumentando la eficiencia operativa (Kuhn, 2013)<sup>4</sup>.

### **Evolución del Credit Scoring en Argentina**

La evolución del Credit Scoring en Argentina ha estado marcada por una serie de factores que han moldeado su desarrollo y aplicación en el contexto financiero del país. A lo largo de las décadas, desde sus primeras etapas hasta la actualidad, varias tendencias y cambios económicos han influido en la forma en que se ha implementado y utilizado el Credit Scoring en Argentina.

1. **Liberalización Financiera:** Durante las últimas décadas del siglo XX y principios del siglo XXI, Argentina experimentó procesos de liberalización financiera que transformaron el sector bancario y crediticio. La apertura de mercados financieros, la desregulación y la entrada de nuevas instituciones financieras extranjeras introdujeron competencia y nuevas prácticas en la evaluación del riesgo crediticio. Esto impulsó la necesidad de utilizar métodos más sofisticados para evaluar el riesgo crediticio, sentando las bases para el desarrollo de modelos de Credit Scoring más avanzados (Abadía, 2013)<sup>5</sup>.
2. **Crisis Económicas:** La historia económica reciente de Argentina está marcada por períodos de crisis económicas y financieras, que han afectado significativamente la estabilidad del sistema financiero. Estos eventos, como la crisis económica de principios de la década de 2000 y las crisis posteriores, han generado la necesidad de evaluar el riesgo crediticio de manera más precisa y eficiente para mitigar el impacto de posibles incumplimientos. Esto condujo a un

---

<sup>3</sup> Hastie, T. a. (2009). The Elements of Statistical Learning: Data Mining, Inference, and Prediction

<sup>4</sup> Kuhn, M. &. (2013). Applied Predictive Modeling.

<sup>5</sup> & <sup>6</sup> Abadía, J. &. (2013). El scoring crediticio en Argentina: Evolución, desafíos y perspectivas. 21(59), 145-174.\*\*.

mayor uso y desarrollo de modelos de Credit Scoring como herramientas para la toma de decisiones informadas en un entorno económico volátil (Abadía, 2013)<sup>6</sup>.

3. **Avances Tecnológicos:** El avance tecnológico ha sido un motor clave en la evolución del Credit Scoring en Argentina. El acceso a nuevas tecnologías de información y comunicación, así como el desarrollo de herramientas de análisis de datos y modelos estadísticos sofisticados, han permitido a las instituciones financieras mejorar sus métodos de evaluación del riesgo crediticio. La disponibilidad de grandes conjuntos de datos y la capacidad de procesarlos de manera eficiente han contribuido al desarrollo de modelos de Credit Scoring más precisos y personalizados (Abadía, 2013)<sup>7</sup>.
4. **Regulación y Supervisión:** La regulación y supervisión del sector financiero también han influido en la evolución del Credit Scoring en Argentina. La implementación de normativas y estándares internacionales, así como la supervisión por parte de entidades regulatorias locales, ha promovido la adopción de prácticas más rigurosas y estandarizadas en la evaluación del riesgo crediticio. Esto ha contribuido a la mejora de la transparencia y la confiabilidad de los modelos de Credit Scoring, fortaleciendo el sistema financiero en general (Banco Central de la República Argentina, 2017)<sup>8</sup>.
5. **Globalización y Competencia:** La globalización de los mercados financieros ha introducido nuevos desafíos y oportunidades para las instituciones financieras en Argentina. La competencia global ha incentivado la adopción de mejores prácticas y la búsqueda de modelos de Credit Scoring más eficientes y precisos para mantener la competitividad en el mercado. Esto ha impulsado la innovación y el desarrollo de modelos de Credit Scoring más adaptados a las necesidades específicas del mercado argentino (Abadía, 2013)<sup>9</sup>.

### **Limitaciones del Enfoque Convencional**

A pesar de su amplia aplicación y uso histórico, el enfoque convencional en el análisis de créditos presenta una serie de limitaciones significativas, especialmente en entornos empresariales dinámicos como el argentino. Este enfoque, que se basa principalmente en el análisis retrospectivo del historial financiero de los solicitantes, tiene varias deficiencias que pueden afectar su capacidad para proporcionar una evaluación precisa del riesgo crediticio y tomar decisiones financieras fundamentadas.

En primer lugar, el enfoque convencional tiende a centrarse exclusivamente en el pasado financiero del solicitante, utilizando variables como el historial crediticio, los ingresos y el patrimonio neto para determinar la elegibilidad y los términos del crédito. Esta mirada retrospectiva puede no capturar adecuadamente las condiciones actuales del mercado o las proyecciones futuras de la empresa. Como señala (Finlay, 2004)<sup>10</sup>, “en entornos empresariales dinámicos, donde las condiciones del mercado pueden

---

<sup>6</sup> Abadía, J. &. (2013). El scoring crediticio en Argentina: Evolución, desafíos y perspectivas.

<sup>7</sup> Abadía, J. &. (2013). El scoring crediticio en Argentina: Evolución, desafíos y perspectivas.

<sup>8</sup> Banco Central de la República Argentina. (2017). Capitales mínimos de las entidades financieras. Adecuaciones. <https://www.argentina.gob.ar/normativa/nacional/comunicaci%C3%B3n-6344-2017-318498/texto>

<sup>9</sup> Abadía, J. &. (2013). El scoring crediticio en Argentina: Evolución, desafíos y perspectivas.

<sup>10</sup> Finlay, S. (2004). Credit Scoring, Response Modeling, and Insurance Rating: A Practical Guide to Forecasting Consumer Behavior

cambiar rápidamente, esta falta de consideración hacia el futuro puede llevar a evaluaciones inexactas del riesgo crediticio”.

Además, el enfoque convencional puede ser susceptible a sesgos y errores de juicio. Dado que gran parte de la evaluación se basa en la interpretación subjetiva de los datos financieros por parte del analista de crédito, existe el riesgo de que los prejuicios personales y las percepciones incorrectas influyan en el proceso de toma de decisiones. Esta subjetividad puede dificultar la evaluación precisa del riesgo asociado a la concesión de créditos y aumentar el riesgo de decisiones erróneas que podrían resultar en pérdidas financieras para la institución prestamista.

Otra limitación importante del enfoque convencional es su falta de adaptabilidad a los cambios en el entorno empresarial y las condiciones del mercado. Dado que se basa en gran medida en datos históricos, puede no ser capaz de capturar cambios significativos en el comportamiento financiero de los prestatarios o en las tendencias del mercado que podrían afectar su capacidad de pago en el futuro. Esto puede llevar a evaluaciones obsoletas del riesgo crediticio y a decisiones desactualizadas que no reflejan adecuadamente la realidad del entorno empresarial actual.

En resumen, aunque el enfoque convencional en el análisis de créditos ha sido ampliamente utilizado durante décadas, presenta una serie de limitaciones que pueden afectar su eficacia en entornos empresariales dinámicos como el argentino. La falta de consideración hacia el futuro, la susceptibilidad a sesgos y errores de juicio, y la falta de adaptabilidad a los cambios en el entorno empresarial son solo algunas de las deficiencias que deben abordarse para mejorar la precisión y la eficacia de la evaluación del riesgo crediticio. En los próximos capítulos, exploraremos en detalle cómo los modelos de Credit Scoring pueden superar estas limitaciones y proporcionar una evaluación más precisa y completa del riesgo crediticio en el contexto argentino.

### **Necesidad de un Enfoque Holístico**

En el contexto empresarial dinámico de Argentina y otros entornos similares, la necesidad de un enfoque holístico en la evaluación del riesgo crediticio se vuelve cada vez más evidente. Si bien el enfoque convencional se ha utilizado ampliamente en el análisis de créditos, presenta limitaciones significativas que pueden no ser adecuadas para entornos empresariales dinámicos y cambiantes.

Una de las principales limitaciones del enfoque convencional radica en su excesivo apego al análisis del historial financiero pasado del solicitante. Este enfoque, centrado únicamente en variables financieras pretéritas como el historial crediticio, los ingresos y el patrimonio neto, descuida las condiciones actuales del mercado y las proyecciones futuras de la empresa. Esta miopía puede conducir a evaluaciones inexactas del riesgo crediticio, especialmente en el caso de empresas emergentes o sectores altamente dinámicos donde las condiciones del mercado cambian rápidamente (Abadía, 2013)<sup>11</sup>. Esto puede resultar en evaluaciones inexactas del riesgo crediticio, especialmente para empresas emergentes o en sectores altamente dinámicos donde las condiciones del mercado pueden cambiar rápidamente.

Además, “el enfoque convencional de evaluación del riesgo crediticio puede ser susceptible a sesgos y errores de juicio, ya que depende en gran medida de la interpretación subjetiva de los datos financieros por parte del analista de crédito”<sup>12</sup>

---

<sup>11</sup> Abadía, J. &. (2013). El scoring crediticio en Argentina: Evolución, desafíos y perspectivas

<sup>12</sup> Gujarati, D. N. (2010). Econometría (5a.ed.--.).

(Gujarati, 2010). Esta subjetividad puede dificultar la evaluación precisa del riesgo asociado a la concesión de créditos, lo que puede llevar a decisiones erróneas y pérdidas financieras. Como también, “la falta de adaptabilidad a los cambios en el entorno empresarial y las condiciones del mercado es otra limitación importante del enfoque convencional de evaluación del riesgo crediticio”<sup>13</sup> (Kuhn, 2013).

Para abordar estas limitaciones, es necesario adoptar un enfoque más holístico que no solo considere el pasado financiero del solicitante, sino que también tenga en cuenta las condiciones actuales del mercado y las proyecciones futuras de la empresa. Esto implicaría la incorporación de una variedad de variables y datos adicionales en el proceso de evaluación del riesgo crediticio, como información macroeconómica, datos de la industria, tendencias del mercado y proyecciones financieras de la empresa.

Al adoptar un enfoque holístico, las instituciones financieras y los prestamistas pueden obtener una visión completa y precisa del riesgo crediticio, lo que les permite tomar decisiones más informadas y fundamentadas sobre la concesión de créditos. Esto es especialmente importante en entornos empresariales dinámicos como el argentino, donde las condiciones del mercado pueden cambiar rápidamente y afectar la capacidad de pago del prestatario.

En los próximos capítulos, exploraremos cómo los modelos de Credit Scoring pueden integrar este enfoque holístico en el proceso de evaluación del riesgo crediticio, utilizando una variedad de variables y datos para proporcionar una evaluación más completa y precisa del riesgo crediticio en el contexto empresarial argentino.

### **Teorías y Herramientas Relevantes**

Para abordar la necesidad de un enfoque holístico en la evaluación del riesgo crediticio, es fundamental comprender y emplear una variedad de teorías y herramientas relevantes en el ámbito financiero. Estas teorías y herramientas proporcionan el marco conceptual y las técnicas necesarias para mejorar la precisión y la eficacia de los modelos de Credit Scoring. A continuación, profundizaremos en algunas de estas teorías y herramientas, destacando su importancia y aplicabilidad en el contexto de la evaluación del riesgo crediticio:

#### **Teoría de Credit Scoring**

La teoría de Credit Scoring es fundamental para comprender los principios subyacentes y los modelos utilizados en la evaluación del riesgo crediticio. Esta teoría, como lo señalan (Gujarati, 2010)<sup>14</sup>, “se basa en la premisa de que ciertas variables financieras y no financieras pueden utilizarse para predecir el riesgo crediticio de un individuo o entidad, permitiendo a las instituciones financieras asignar una puntuación de crédito objetiva y cuantificable. Esta teoría proporciona un marco conceptual sólido que guía el desarrollo y la aplicación de modelos de Credit Scoring, permitiendo a los analistas de crédito evaluar de manera más precisa el riesgo asociado a la concesión de créditos”.

#### **Teoría de la Inteligencia Artificial**

La Inteligencia Artificial ofrece un enfoque innovador para la predicción financiera y la evaluación del riesgo crediticio. Esta tecnología son capaces de identificar patrones complejos y relaciones no evidentes en los datos, lo que permite una mejor comprensión del comportamiento crediticio de los prestatarios. Esta capacidad para identificar patrones complejos ayuda a las instituciones financieras a identificar prestatarios de alto

---

<sup>13</sup> Kuhn, M. &. (2013). Applied Predictive Modeling

<sup>14</sup> Gujarati, D. N. (2010). Econometría (5a.ed.--.).

riesgo que podrían haber sido pasados por alto con métodos tradicionales. La IA se utiliza para crear modelos predictivos que estiman la probabilidad de que un prestatario incumpla con sus obligaciones crediticias. Estos modelos, que son más sofisticados que los métodos tradicionales basados en reglas, consideran una gama más amplia de factores y son capaces de adaptarse a cambios en el mercado. Esto permite a las instituciones financieras tomar decisiones crediticias más precisas y gestionar mejor el riesgo.

### **Herramientas Matemáticas Avanzadas**

Además de estas teorías, es fundamental utilizar como dice (Hastie, 2009)<sup>15</sup> "Las herramientas matemáticas avanzadas, como los modelos de Machine Learning, el Análisis de Regresión y otras técnicas de análisis predictivo, son fundamentales para desarrollar modelos de Credit Scoring más precisos y efectivos. Estas herramientas permiten una exploración exhaustiva de los datos y la identificación de patrones y relaciones complejas que no serían detectables mediante métodos convencionales. Al incorporar estas herramientas en el proceso de desarrollo de modelos, se puede mejorar significativamente la precisión y la capacidad predictiva de los modelos de Credit Scoring."

En conjunto, estas teorías y herramientas proporcionan los fundamentos teóricos y las técnicas prácticas necesarias para desarrollar modelos de Credit Scoring más avanzados y efectivos. Al integrar estas teorías y herramientas en el proceso de evaluación del riesgo crediticio, las instituciones financieras pueden mejorar la precisión y la eficacia de sus modelos, permitiéndoles tomar decisiones más informadas y fundamentadas sobre la concesión de créditos.

### **Modelos de Machine Learning**

Dentro del amplio espectro de herramientas matemáticas avanzadas, como se señala en el libro de (Gujarati, 2010)<sup>16</sup> "Los modelos de Machine Learning han ganado considerable atención en la evaluación del riesgo crediticio, debido a su capacidad para analizar grandes conjuntos de datos y extraer patrones complejos que permiten predecir el riesgo crediticio de manera más precisa que los métodos tradicionales. Algoritmos como Random Forest y Gradient Boosting son particularmente populares debido a su habilidad para manejar datos no lineales y capturar interacciones complejas entre variables."

### **Análisis de Regresión**

El Análisis de Regresión es otra herramienta fundamental en la evaluación del riesgo crediticio. Esta técnica, como lo señalan (Gujarati, 2010)<sup>17</sup>, permite identificar la relación entre una variable dependiente, como la probabilidad de incumplimiento, y un conjunto de variables independientes, como los ingresos, el historial crediticio y otros factores relevantes. Mediante el Análisis de Regresión, los analistas pueden cuantificar el impacto de cada variable en el riesgo crediticio y desarrollar modelos predictivos que estimen la probabilidad de incumplimiento de un prestatario.

### **Modelos Bayesianos**

Los modelos Bayesianos también desempeñan un papel importante en la evaluación del riesgo crediticio. Estos modelos, como lo señalan (Gelman, 2013)<sup>18</sup>, utilizan la teoría

---

<sup>15</sup> Hastie, T. a. (2009). The Elements of Statistical Learning: Data Mining, Inference, and Prediction.

<sup>16</sup> Gujarati, D. N. (2010). Econometría (5a.ed.--.).

<sup>17</sup> Gujarati, D. N. (2010). Econometría (5a.ed.--.).

<sup>18</sup> Gelman, A. C. (2013). Bayesian Data Analysis

de la probabilidad Bayesiana para estimar la probabilidad de ocurrencia de eventos futuros, como el impago de un préstamo. Al incorporar información previa y actualizar continuamente las estimaciones en función de nuevos datos, los modelos Bayesianos pueden proporcionar predicciones más precisas y robustas del riesgo crediticio en entornos cambiantes.

### **Herramientas de Validación y Evaluación**

Además de las técnicas de modelado, es crucial utilizar herramientas de validación y evaluación para garantizar la precisión y la fiabilidad de los modelos de Credit Scoring. Métricas como el coeficiente de Gini, la curva ROC y el error cuadrático medio (MSE), como lo señalan (Hand, 2001)<sup>19</sup>, son comúnmente utilizadas para evaluar el rendimiento de los modelos y comparar su capacidad predictiva. La validación cruzada también es fundamental para verificar la generalización de los modelos y su capacidad para funcionar bien en datos nuevos y no vistos.

Con esta exploración preliminar del contexto y el marco teórico del Credit Scoring, nos preparamos para abordar en el próximo capítulo la metodología utilizada para desarrollar y aplicar los modelos de Credit Scoring propuestos, así como para presentar los resultados obtenidos y analizar su relevancia en el contexto financiero argentino.

### **Empresas “Unicornio”**

Las empresas "unicornio" son aquellas startups de base tecnológica que han logrado alcanzar una valuación de mercado de al menos mil millones de dólares antes de su salida a bolsa o de ser adquiridas por otra empresa más grande. Este término fue acuñado por la inversora de capital de riesgo Aileen Lee en 2013, refiriéndose a lo raro y poco común que era en ese momento para una startup alcanzar una valuación tan alta.

"Las empresas unicornio son como el fenómeno del cisne negro en el mundo empresarial: son raras, difíciles de prever y desafían las expectativas convencionales". (Taleb, 2007)<sup>20</sup>

### **Características de las Empresas Unicornio:**

#### **1. Innovación Tecnológica**

Las empresas unicornio suelen destacarse por su enfoque en la innovación y la aplicación de tecnologías disruptivas en sus productos o servicios. Esto les permite diferenciarse en el mercado y atraer la atención de inversores y consumidores.

#### **2. Crecimiento Rápido**

Una característica distintiva de las empresas unicornio es su capacidad para experimentar un crecimiento acelerado en poco tiempo. Esto se debe en gran parte a su modelo de negocio escalable y la demanda de mercado de sus productos o servicios. "Las empresas unicornio no solo crecen rápido, sino que lo hacen de manera exponencial. Esto les permite superar a la competencia y capturar una cuota de mercado significativa."<sup>21</sup>

---

<sup>19</sup> Hand, D. J. (2001). Statistical methods in credit risk management. Oxford University Press, USA.

<sup>20</sup> Taleb, N. N. (2007). El Cisne Negro: El impacto de lo altamente improbable. Random House

<sup>21</sup> Peter Thiel, (cofundador de PayPal y Founders Fund)



### **3. Financiamiento de Capital de Riesgo**

Estas empresas a menudo dependen en gran medida del financiamiento de capital de riesgo para impulsar su crecimiento y expansión. Los inversores de capital de riesgo están dispuestos a invertir grandes sumas de dinero en estas startups a cambio de una participación en la empresa y la esperanza de obtener ganancias significativas en el futuro.

### **4. Disrupción de Industrias Tradicionales**

Las empresas unicornio suelen desafiar los modelos de negocio establecidos y las industrias tradicionales mediante la introducción de soluciones innovadoras y eficientes. Esto puede provocar cambios significativos en el mercado y generar nuevas oportunidades para el crecimiento y la rentabilidad.

### **5. Equipo Fundador Talentoso**

Detrás de cada empresa unicornio suele haber un equipo fundador altamente talentoso y visionario. Estos emprendedores están impulsados por la pasión, la creatividad y la determinación para llevar adelante su visión y hacer realidad el potencial de su empresa.

### **6. Enfoque en la Experiencia del Usuario**

Las empresas unicornio suelen priorizar la experiencia del usuario en el diseño y desarrollo de sus productos o servicios. Esto les permite generar lealtad entre los clientes y mantenerse a la vanguardia en un mercado altamente competitivo.

### **7. Internacionalización**

Muchas empresas unicornio tienen como objetivo expandirse a nivel internacional para aprovechar nuevas oportunidades de mercado y aumentar su base de clientes potenciales. Esto requiere una estrategia sólida y recursos significativos para adaptarse a las diferencias culturales y regulatorias en diferentes regiones del mundo.

### **8. Ecosistema y Red de Apoyo**

Las empresas unicornio a menudo se benefician de un ecosistema emprendedor vibrante y una red de apoyo que incluye mentores, inversores, aceleradoras y otros emprendedores exitosos. Esto les proporciona recursos adicionales, conocimientos y oportunidades de colaboración para impulsar su crecimiento y éxito continuo.

### **Z"-Score**

Desarrollada en los años 90, también conocida como la "Z"-Score Revisada", es una versión actualizada y refinada del modelo original de Altman, que fue creado por Edward I. Altman en la década de 1960. Esta versión revisada incorpora nuevas variables y coeficientes optimizados a través de análisis estadísticos actualizados y una base de datos más amplia. El objetivo de esta revisión era mejorar la precisión del modelo y su capacidad para predecir la probabilidad de quiebra de una empresa. Similar a la fórmula original, estos coeficientes se basan en un análisis estadístico de una muestra de empresas que han experimentado quiebras y empresas que han continuado operando de manera solvente. La Z"-Score revisada se interpreta de manera similar a la versión original, donde un valor mayor indica una menor probabilidad de quiebra, y viceversa.

Es importante destacar que la Z"-Score revisada se considera una mejora sobre la versión original, ya que incorpora una metodología más robusta y actualizada para evaluar el riesgo de quiebra de una empresa. Sin embargo, su utilidad y precisión dependen en gran medida de la calidad y relevancia de los datos utilizados en su cálculo, así como de la estabilidad y predictibilidad de las condiciones económicas y financieras en las que opera la empresa.

$$Z'' = \frac{EBIT}{Total Assets} * 6.72 + \frac{Total Equity}{Total Liabilities} * 1.05 + \frac{Working Capital}{Total Assets} * 6.56 + \frac{Retained Earnings}{Total Assets} * 3.26$$

A continuación, se describe cada modelo y su aplicación en el contexto del análisis de Credit Scoring:

### Modelo de Flujo de Caja Descontado (DCF)

El Modelo de Flujo de Caja Descontado (DCF) es una herramienta fundamental en la evaluación financiera que se utiliza para estimar el valor presente de los flujos de efectivo futuros de una empresa o proyecto. Este modelo se basa en el principio de que el valor de un activo está determinado por los flujos de efectivo que generará en el futuro, descontados a una tasa que refleje el costo de oportunidad de los inversores y el riesgo asociado con esos flujos de efectivo.

El proceso de aplicación del modelo de DCF implica varios pasos clave:

1. Proyección de flujos de efectivo futuros: El primer paso en el modelo de DCF es proyectar los flujos de efectivo que se espera que genere la empresa en el futuro. Estas proyecciones suelen abarcar un horizonte temporal de varios años y se basan en análisis exhaustivos de las operaciones comerciales, los ingresos, los costos y otras variables relevantes.
2. Determinación de la tasa de descuento: La tasa de descuento, también conocida como tasa de descuento o tasa de rendimiento requerida, representa el costo de capital de los inversores y el riesgo asociado con la inversión en la empresa. Esta tasa se utiliza para descontar los flujos de efectivo futuros a su valor presente. La determinación de la tasa de descuento puede basarse en varios factores, como la tasa libre de riesgo, el rendimiento esperado del mercado y el riesgo específico del activo.
3. Descuento de los flujos de efectivo futuros: Una vez proyectados los flujos de efectivo futuros y determinada la tasa de descuento adecuada, se aplican técnicas de descuento para calcular el valor presente de esos flujos de efectivo. Esto implica dividir cada flujo de efectivo futuro por  $(1 + \text{tasa de descuento})$  elevado al número de períodos en el futuro en que se espera que se produzca ese flujo de efectivo.
4. Estimación del valor terminal: Además de descontar los flujos de efectivo futuros, el modelo de DCF también puede incluir una estimación del valor terminal, que representa el valor de la empresa al final del horizonte de proyección. Esto se calcula generalmente aplicando un múltiplo de valoración (como el múltiplo de EBITDA) o utilizando un modelo de crecimiento perpetuo, como el modelo de Gordon.



5. Cálculo del valor total de la empresa: Una vez que se han descontado todos los flujos de efectivo futuros y se ha estimado el valor terminal, se suman estos valores para calcular el valor total de la empresa.

### **Modelo de Gordon-Shapiro**

El Modelo de Gordon-Shapiro, también conocido como el Modelo de Crecimiento de Dividendos, es una herramienta de valoración utilizada para estimar el valor de una empresa que distribuye dividendos a sus accionistas. Este modelo se basa en la premisa de que el valor de una acción es igual al valor presente de todos los dividendos futuros que se espera que la empresa pague a los accionistas, descontados a una tasa de descuento adecuada.

Como en nuestro caso teórico no trabajamos ni tenemos un histórico de las empresas, se decidió trabajar con una modificación sobre la fórmula de Gordon-Shapiro, utilizando una simulación del crecimiento de ventas para poder calcular el valor futuro.

Expresándose matemáticamente de la siguiente manera:

$$V_{gs} = \frac{EBITDA * Margen EBITDA * Periodo de Cobro}{Tasa de Descuento - Tasa de Crecimiento de Ventas}$$

### **Tasa de Crecimiento de Ventas**

Una de las piezas fundamentales del Modelo de Gordon-Shapiro es la tasa de crecimiento constante de las ventas. Esta tasa representa la tasa a la cual se espera que las ventas de la empresa crezcan en el futuro. En muchas aplicaciones del modelo, se asume que esta tasa es constante a perpetuidad. Sin embargo, es importante tener en cuenta que esta suposición puede no ser realista en todos los casos, especialmente para empresas en sectores de rápido crecimiento o sujetas a fluctuaciones económicas.

### **Tasa de Descuento (r)**

Otro componente crítico del Modelo de Gordon-Shapiro es la tasa de descuento (r), que representa el rendimiento requerido por los inversores para asumir el riesgo asociado con la inversión en la empresa. Esta tasa de descuento se utiliza para descontar los flujos de efectivo futuros (es decir, para nuestro caso los ingresos futuros) al valor presente. La selección de la tasa de descuento adecuada es crucial y puede variar según diversos factores, como el riesgo específico de la empresa, las condiciones del mercado y las tasas de interés vigentes.

### **Supuestos y Limitaciones**

Si bien el Modelo de Gordon-Shapiro proporciona un marco valioso para la valoración de empresas que distribuyen dividendos, también tiene algunas limitaciones importantes. Una de las principales limitaciones es la suposición de que la tasa de crecimiento de los dividendos es constante a perpetuidad. En la práctica, esta suposición puede no ser válida para todas las empresas, especialmente aquellas sujetas a cambios en el entorno económico o tecnológico.

### **Aplicación Práctica**

A pesar de sus limitaciones, el Modelo de Gordon-Shapiro sigue siendo una herramienta ampliamente utilizada en la valoración de empresas, especialmente para aquellas que tienen un historial estable de distribución de dividendos y se espera que continúen haciéndolo en el futuro. Los analistas financieros suelen combinar el Modelo de Gordon-Shapiro con otras técnicas de valoración, como el Modelo de Descuento de Flujos de Efectivo (DCF), para obtener una evaluación más completa y precisa del valor de una empresa.

### **Modelo de Credit Default Probabilities (CDP)**

El Modelo de CDP utiliza una variedad de variables financieras y no financieras para estimar la probabilidad de que un prestatario incumpla con sus obligaciones crediticias en un período de tiempo específico. Estas variables pueden incluir el historial crediticio del prestatario, la estabilidad financiera de la empresa, la situación económica del sector al que pertenece, entre otros. "Este modelo permite a los bancos y otras instituciones financieras tomar decisiones informadas sobre la concesión de préstamos, la fijación de precios y la gestión del riesgo crediticio."<sup>22</sup> (Corporate Finance Institute, s.f.)

El modelo utiliza técnicas avanzadas de análisis predictivo, como modelos de Machine Learning, análisis estadístico y Análisis de Regresión, para desarrollar una función predictiva que relaciona las variables de entrada con la probabilidad de incumplimiento. Estos modelos se entrenan utilizando datos históricos de incumplimientos y comportamiento crediticio de los prestatarios.

La precisión y eficacia del Modelo de CDP se evalúan mediante métricas de rendimiento como la curva ROC, el coeficiente de Gini y el error cuadrático medio (MSE). Estas métricas proporcionan información sobre la capacidad del modelo para discriminar entre prestatarios que cumplen y no cumplen con sus obligaciones crediticias.

El Modelo de CDP se utiliza en la industria financiera para asignar puntajes de crédito a los prestatarios y tomar decisiones informadas sobre la concesión de créditos. Al proporcionar una estimación precisa de la probabilidad de incumplimiento, este modelo ayuda a las instituciones financieras a evaluar y gestionar el riesgo crediticio de manera más efectiva.

Una de las principales ventajas del Modelo de CDP es su capacidad para capturar una amplia gama de factores que pueden influir en la probabilidad de incumplimiento, lo que lo hace más robusto y preciso que los enfoques tradicionales. Sin embargo, una limitación importante es la necesidad de datos históricos suficientes y representativos para entrenar el modelo de manera efectiva, lo que puede ser un desafío en entornos con poca disponibilidad de datos.

---

<sup>22</sup> <https://corporatefinanceinstitute.com/>

## **Capítulo 2: Desarrollo y Aplicación de Modelos de Credit Scoring**

### **Metodología Utilizada**

En esta sección, se detallarán las técnicas avanzadas de análisis predictivo empleadas en el desarrollo de los modelos de Credit Scoring. Estas técnicas se seleccionaron cuidadosamente para aprovechar al máximo la información disponible en la base de datos y generar modelos predictivos precisos y robustos. A continuación, se explicarán cada una de estas técnicas en detalle:

### **Inteligencia Artificial**

La IA ofrece herramientas y técnicas avanzadas para analizar grandes volúmenes de datos crediticios, lo que permite desarrollar modelos más precisos y eficientes para evaluar el riesgo crediticio. "La IA puede procesar grandes volúmenes de datos y encontrar patrones complejos, lo que la convierte en una herramienta poderosa para el análisis crediticio"<sup>23</sup> (Chen, 2020).

En el contexto específico del Credit Scoring, la IA se convierte en una herramienta fundamental para mejorar la precisión en la evaluación del riesgo de incumplimiento de préstamos. A través de algoritmos de aprendizaje automático y análisis predictivo, los analistas pueden identificar patrones complejos en los datos históricos de créditos y predecir con mayor exactitud la probabilidad de incumplimiento de un préstamo. "La IA puede identificar prestatarios de alto riesgo con mayor precisión, lo que reduce las pérdidas por morosidad y mejora la rentabilidad de las instituciones financieras"<sup>24</sup> (Jiang, 2020).

La IA también ofrece la capacidad de optimizar procesos y mejorar la eficiencia en la gestión de carteras de crédito. Mediante el uso de algoritmos de optimización y automatización, se pueden identificar áreas de riesgo y oportunidades de mejora en la cartera de créditos, lo que contribuye a una toma de decisiones más informada y a la reducción del riesgo de pérdidas. "La IA puede permitir a las instituciones financieras otorgar crédito a prestatarios que antes eran considerados de alto riesgo, expandiendo el acceso al crédito a una mayor parte de la población"<sup>25</sup> (Hu et al., 2021).

Además, la IA permite explorar diferentes escenarios y evaluar el impacto potencial de variables específicas en el riesgo crediticio. Esta capacidad de análisis avanzado proporciona una visión más completa y detallada del riesgo crediticio, lo que les permite desarrollar estrategias más efectivas para gestionar y mitigar el riesgo en las carteras de crédito.

### **Modelos de Machine Learning**

Los modelos de Machine Learning han revolucionado el campo del Credit Scoring al ofrecer un enfoque más avanzado y preciso para predecir el riesgo crediticio. Estos modelos utilizan algoritmos sofisticados que pueden analizar grandes volúmenes de datos y extraer patrones complejos, permitiendo una evaluación más precisa y dinámica del riesgo crediticio.

---

<sup>23</sup> Chen, M. &. (2020). Applications of Machine Learning in Credit Risk Management. In Machine Learning in Finance. Springer, Singapore.

<sup>24</sup> Jiang, S. &. (2020). The Application of Generative Adversarial Networks in Credit Scoring.

<sup>25</sup> Hu, Y., Chen, L., & Li, Y. (2021). Synthetic Data Generation for Credit Scoring: A Review

## 1. Random Forest

El Random Forest es uno de los modelos de Machine Learning más utilizados en Credit Scoring debido a su capacidad para manejar datos no lineales y capturar interacciones complejas entre variables. Este modelo se basa en la construcción de múltiples árboles de decisión y la combinación de sus predicciones para mejorar la precisión general del modelo. En el contexto del Credit Scoring, el Random Forest puede identificar relaciones no lineales entre las variables financieras y el riesgo crediticio, como lo señalan (Bontempi, 2014)<sup>26</sup>, permitiendo una evaluación más precisa del riesgo asociado a la concesión de créditos.

## 2. Gradient Boosting

El Gradient Boosting es otra técnica popular en el Credit Scoring que se utiliza para mejorar la precisión de los modelos predictivos. Este modelo funciona construyendo una serie de árboles de decisión secuenciales, donde cada árbol se enfoca en corregir los errores del modelo anterior. Esto permite mejorar gradualmente la precisión del modelo y capturar relaciones complejas entre las variables predictoras y el riesgo crediticio. En el contexto del Credit Scoring, el Gradient Boosting puede identificar patrones sutiles en los datos que pueden pasar desapercibidos para otros modelos, como lo señalan (Boosting, 2014)<sup>27</sup>, mejorando así la capacidad predictiva del modelo.

## 3. Redes Neuronales Artificiales

Las Redes Neuronales Artificiales (RNA) son modelos de Machine Learning inspirados en el funcionamiento del cerebro humano que han demostrado ser efectivos en la evaluación del riesgo crediticio. Estos modelos utilizan capas de neuronas interconectadas para aprender patrones complejos en los datos y hacer predicciones precisas sobre el riesgo crediticio de un prestatario. En el contexto del Credit Scoring, las RNA pueden identificar relaciones no lineales y no lineales complejas entre las variables financieras y el riesgo crediticio, como lo señalan (Zhang, 2018)<sup>28</sup>, permitiendo una evaluación más precisa del riesgo asociado a la concesión de créditos.

## 4. Support Vector Machines (SVM)

Las Support Vector Machines (Máquinas de Vectores de Soporte) son otro enfoque popular en el Credit Scoring que se utiliza para clasificar los prestatarios en función de su riesgo crediticio. Este modelo funciona construyendo un hiperplano en un espacio de características de alta dimensión que separa las instancias de diferentes clases. En el contexto del Credit Scoring, las SVM pueden identificar relaciones lineales y no lineales entre las variables predictoras y el riesgo crediticio, como lo señalan (Huang, Credit Scoring Using Support Vector Machines, 2014)<sup>29</sup>, permitiendo una evaluación precisa del riesgo asociado a la concesión de créditos.

## Análisis de Regresión

El Análisis de Regresión es una técnica estadística fundamental en el campo del Credit Scoring que se utiliza para identificar la relación entre variables independientes y una

---

<sup>26</sup> Bontempi, P. &. (2014). Credit Scoring Using Random Forest: A Comparative Analysis.

<sup>27</sup> Boosting, A. T. (2014). Chen & He

<sup>28</sup> Zhang, e. a. (2018). Application of Deep Learning to Credit Scoring

<sup>29</sup> Huang, e. a. (2014). Credit Scoring Using Support Vector Machines.

variable dependiente, como la probabilidad de incumplimiento de un préstamo. Esta técnica permite cuantificar el impacto de cada variable en el riesgo crediticio y desarrollar modelos predictivos que estimen la probabilidad de incumplimiento de un prestatario. Como lo señalan (Thomas, 2002)<sup>30</sup>, el Análisis de Regresión es una de las técnicas más utilizadas en Credit Scoring debido a su simplicidad, interpretabilidad y efectividad para modelar el riesgo crediticio.

### **1. Regresión Lineal**

La Regresión Lineal es la forma más simple de Análisis de Regresión y se utiliza comúnmente en el Credit Scoring para identificar relaciones lineales entre variables financieras y el riesgo crediticio. Este modelo asume una relación lineal entre las variables independientes y la variable dependiente y utiliza técnicas de optimización para ajustar una línea recta que minimice la suma de los errores cuadráticos. En el contexto del Credit Scoring, la Regresión Lineal puede identificar relaciones lineales simples entre variables financieras, como los ingresos y el historial crediticio, y el riesgo crediticio de un prestatario.

### **2. Regresión Logística**

La Regresión Logística es una extensión de la regresión lineal que se utiliza para modelar la relación entre variables independientes binarias o categóricas y una variable dependiente binaria, como la probabilidad de incumplimiento de un préstamo. Este modelo asume una Relación Logarítmica entre las variables independientes y la variable dependiente y utiliza técnicas de optimización para ajustar una curva sigmoide que se ajuste a los datos. En el contexto del Credit Scoring, "La capacidad de la Regresión Logística para identificar relaciones no lineales entre variables financieras y el riesgo crediticio, y predecir la probabilidad de incumplimiento, la hace ideal para una evaluación precisa del riesgo en la concesión de créditos" (Li, 2016)<sup>31</sup>.

### **3. Regresión de Árboles de Decisión**

La Regresión de Árboles de Decisión es una técnica de Análisis de Regresión que se utiliza para modelar relaciones no lineales entre variables independientes y una variable dependiente. Este modelo divide iterativamente el espacio de características en subconjuntos más pequeños, ajustando una función de regresión simple en cada subconjunto. En el contexto del Credit Scoring, la regresión de árboles de decisión puede identificar patrones complejos en los datos y capturar relaciones no lineales entre variables financieras y el riesgo crediticio, permitiendo una evaluación más precisa del riesgo asociado a la concesión de créditos.

### **4. Regresión Ridge y Lasso**

"La Regresión Ridge y Lasso son técnicas de regularización que se utilizan comúnmente en Credit Scoring para abordar la multicolinealidad y el sobreajuste, mejorando la estabilidad y el rendimiento de los modelos predictivos" (Huang, Regularization Techniques for Credit Scoring, 2010)<sup>32</sup>. Estos modelos introducen términos de regularización que penalizan los coeficientes de las variables, lo que ayuda a reducir la

---

<sup>30</sup> Thomas, e. a. (2002). A Review of Regression Techniques in Credit Scoring

<sup>31</sup> Li, e. a. (2016). A Survey of Machine Learning Techniques for Credit Scoring

<sup>32</sup> Huang, e. a. (2010). Regularization Techniques for Credit Scoring.

varianza del modelo y mejorar su capacidad predictiva. En el contexto del Credit Scoring, la Regresión Ridge y Lasso pueden mejorar la estabilidad y el rendimiento de los modelos predictivos, permitiendo una evaluación más precisa del riesgo crediticio de un prestatario.

### **Análisis Tornado de Variables**

"El análisis tornado de variables es una técnica complementaria al análisis de regresión que se utiliza en Credit Scoring para identificar y priorizar las variables más importantes en la predicción del riesgo crediticio." (Ayadi, 2011)<sup>33</sup>. "Esta técnica permite a los analistas de crédito evaluar el impacto relativo de cada variable en el modelo predictivo y determinar qué variables tienen el mayor efecto en la probabilidad de incumplimiento de un préstamo." (Huang, Credit Scoring Using Support Vector Machines, 2014)<sup>34</sup>

### **Proceso de Análisis Tornado**

El análisis tornado de variables comienza con la ejecución del modelo predictivo, que puede ser un modelo de Regresión Logística, Regresión de Árboles de Decisión u otra técnica de modelado predictivo. Una vez que se ha ejecutado el modelo, se calcula el impacto de cada variable en la predicción del riesgo crediticio. Este impacto se mide utilizando una métrica relevante, como los coeficientes de regresión en el caso de la Regresión Logística o la importancia de las características en el caso de la Regresión de Árboles de Decisión.

#### **1. Identificación de Variables Importantes**

Una vez que se ha calculado el impacto de cada variable, se identifican las variables más importantes utilizando un criterio predefinido, como el valor absoluto de los coeficientes de regresión o la importancia relativa de las características. Estas variables se consideran las más influyentes en la predicción del riesgo crediticio y son las que se analizan en detalle en el análisis tornado de variables.

#### **2. Visualización de Resultados**

Los resultados del análisis tornado de variables se presentan típicamente en forma de gráfico de barras o diagrama de tornado, que muestra la contribución relativa de cada variable a la predicción del riesgo crediticio. Las variables se ordenan según su impacto, con las más importantes en la parte superior del tornado y las menos importantes en la parte inferior.

#### **3. Interpretación de Resultados**

Una vez que se han visualizado los resultados, los analistas de crédito pueden interpretar la contribución relativa de cada variable y determinar qué variables tienen el mayor efecto en la probabilidad de incumplimiento de un préstamo. Esto puede ayudar a identificar áreas de riesgo potencial y guiar la toma de decisiones sobre la gestión de la cartera de préstamos.

---

<sup>33</sup> Ayadi, e. a. (2011). The Use of Tornado Analysis for Variable Importance Assessment in Credit Scoring.

<sup>34</sup> Huang, e. a. (2014). Credit Scoring Using Support Vector Machines



#### 4. Refinamiento del Modelo

Basándose en los resultados del análisis tornado de variables, los analistas de crédito pueden refinar el modelo predictivo, eliminando variables irrelevantes o poco importantes y centrando el modelo en las variables más influyentes. Esto puede mejorar la precisión y la eficacia del modelo predictivo y permitir una evaluación más precisa del riesgo crediticio.

Para predecir el valor futuro de las empresas en un horizonte temporal de 5 a 10 años, se utilizarán varios modelos de predicción a largo plazo. Estos modelos incluirán el modelo de flujo de caja descontado (DCF, por sus siglas en inglés) y el modelo de Gordon-Shapiro, ambos ampliamente utilizados en la valoración de empresas.

#### Evaluación de los Modelos

En esta sección, se presentarán las métricas utilizadas para evaluar la precisión y eficacia de los modelos de Credit Scoring desarrollados. Estas métricas fueron seleccionadas para medir diferentes aspectos del rendimiento de los modelos y garantizar una evaluación exhaustiva de su capacidad predictiva. A continuación, se describirán brevemente cada una de estas métricas:

#### Coeficiente de Gini

El coeficiente de Gini es una medida estadística utilizada en el análisis de Credit Scoring para evaluar la capacidad predictiva de un modelo de clasificación, especialmente en el contexto de la evaluación del riesgo crediticio. "El coeficiente de Gini se utiliza para comparar el rendimiento de diferentes modelos de Credit Scoring y para seleccionar el modelo que mejor se adapta a las necesidades específicas de una institución financiera. Un coeficiente de Gini de 0 indica que el modelo de Credit Scoring tiene una capacidad predictiva perfecta, mientras que un coeficiente de Gini de 1 indica que el modelo no tiene capacidad predictiva."<sup>35</sup> (Fair Isaac Corporation, s.f.). A continuación, se profundiza en el significado y la importancia del coeficiente de Gini en el contexto del Credit Scoring:

##### 1. Concepto del Coeficiente de Gini

El coeficiente de Gini es una medida de la desigualdad en la distribución de una variable, que puede ser aplicada en diversos contextos, incluido el análisis del riesgo crediticio. En el contexto del Credit Scoring, se utiliza para evaluar la capacidad del modelo de clasificación para distinguir entre los clientes que eventualmente incumplen sus obligaciones crediticias y aquellos que las cumplen.

##### 2. Cálculo del Coeficiente de Gini

El coeficiente de Gini se calcula a partir de la curva de Lorenz, que representa la relación entre la acumulación de la población y la acumulación de una variable, en este caso, las probabilidades pronosticadas por el modelo de Credit Scoring.

Curva de Lorenz:

$$L(x) = \frac{(\sum P_i * Y_i)}{(\sum Y_i)}$$

- L(x) representa el porcentaje acumulado de la riqueza o ingreso para un nivel de población dado (x).

---

<sup>35</sup> <https://www.fico.com/>

- $\Sigma$  representa la suma de todos los valores.
- $P_i$  representa la proporción de la población que pertenece al grupo  $i$ .
- $Y_i$  representa el ingreso o riqueza promedio del grupo  $i$ .

El coeficiente de Gini se calcula como el área entre la curva de Lorenz y la línea de igualdad perfecta dividida por el área total bajo la línea de igualdad perfecta.

### **3. Interpretación del Coeficiente de Gini**

El coeficiente de Gini varía entre 0 y 1, donde 0 representa una distribución perfectamente igualitaria (todas las probabilidades pronosticadas son iguales) y 1 representa una distribución perfectamente desigual (todas las probabilidades pronosticadas están concentradas en un solo grupo). En el contexto del Credit Scoring, un coeficiente de Gini cercano a 1 indica que el modelo es capaz de distinguir con precisión entre los buenos y malos pagadores, mientras que un coeficiente cercano a 0 indica que el modelo no tiene capacidad discriminativa.

### **4. Importancia del Coeficiente de Gini en el Credit Scoring**

El coeficiente de Gini es una medida fundamental en el análisis de Credit Scoring, ya que proporciona una evaluación cuantitativa de la capacidad predictiva del modelo de clasificación. Un coeficiente de Gini alto indica que el modelo es capaz de distinguir con precisión entre los prestatarios que eventualmente incumplirán sus obligaciones y aquellos que las cumplirán, lo que se traduce en una mejor toma de decisiones crediticias y una gestión más eficiente del riesgo crediticio.

### **5. Uso del Coeficiente de Gini en la Evaluación de Modelos**

En la práctica, el coeficiente de Gini se utiliza junto con otras métricas de evaluación, como la curva ROC (Receiver Operating Characteristic), el error cuadrático medio y la validación cruzada, para evaluar la precisión y la eficacia de los modelos de Credit Scoring. Al comparar el coeficiente de Gini de diferentes modelos o versiones del mismo modelo, los analistas de crédito pueden identificar el modelo más predictivo y tomar decisiones informadas sobre su implementación en la gestión del riesgo crediticio.

#### **Curva ROC**

La Curva ROC (Receiver Operating Characteristic) es una herramienta fundamental en el análisis de Credit Scoring que se utiliza para evaluar la capacidad predictiva de un modelo de clasificación binaria, como los utilizados para predecir el riesgo crediticio. La Curva ROC representa la relación entre la tasa de verdaderos positivos (sensibilidad) y la tasa de falsos positivos (1 - especificidad) para diferentes umbrales de clasificación del modelo. "Una Curva ROC cerca de la esquina superior izquierda indica un modelo de Credit Scoring con alta sensibilidad y especificidad, lo que significa que puede identificar correctamente a los prestatarios buenos y malos con un alto grado de precisión."<sup>36</sup> (Standard & Poor's, s.f.)

A continuación, se profundiza en el significado y la importancia de la Curva ROC en el contexto del Credit Scoring:

---

<sup>36</sup> <https://www.spglobal.com/ratings/en/>



## **1. Concepto de la Curva ROC**

La Curva ROC es una representación gráfica de la capacidad de un modelo de clasificación para distinguir entre las clases positiva y negativa. En el caso del Credit Scoring, las clases positivas suelen representar a los prestatarios que eventualmente incumplen sus obligaciones financieras, mientras que las clases negativas representan a aquellos que las cumplen. La Curva ROC se construye trazando la tasa de verdaderos positivos (sensibilidad) en el eje y frente a la tasa de falsos positivos ( $1 - \text{especificidad}$ ) en el eje x para diferentes valores de umbral de clasificación del modelo.

## **2. Interpretación de la Curva ROC**

La Curva ROC proporciona una representación visual de la capacidad discriminativa del modelo de clasificación en diferentes umbrales de clasificación. Idealmente, se espera que el modelo tenga una alta sensibilidad (tasa de verdaderos positivos) y una baja tasa de falsos positivos para todos los umbrales de clasificación, lo que se traduce en una curva ROC que se acerca al vértice superior izquierdo del gráfico. Cuanto más cerca esté la curva ROC del vértice superior izquierdo, mejor será la capacidad predictiva del modelo.

## **3. Área bajo la Curva ROC (AUC)**

El Área bajo la Curva ROC (AUC) es una medida cuantitativa de la capacidad discriminativa del modelo de clasificación. Representa el área bajo la Curva ROC y varía entre 0 y 1, donde un valor de 1 indica una capacidad discriminativa perfecta y un valor de 0.5 indica un rendimiento aleatorio (sin capacidad discriminativa). Cuanto mayor sea el valor del AUC, mejor será la capacidad predictiva del modelo.

## **4. Importancia de la Curva ROC en el Credit Scoring**

La Curva ROC es una herramienta fundamental en el análisis de Credit Scoring, ya que proporciona una evaluación visual y cuantitativa de la capacidad predictiva del modelo de clasificación. Permite a los analistas de crédito comparar fácilmente diferentes modelos o versiones del mismo modelo y tomar decisiones informadas sobre su implementación en la gestión del riesgo crediticio. Un mayor AUC indica una mayor capacidad discriminativa del modelo y, por lo tanto, una mejor capacidad para predecir el riesgo crediticio.

## **5. Uso de la Curva ROC en la Evaluación de Modelos**

En la práctica, la Curva ROC se utiliza junto con otras métricas de evaluación, como el coeficiente de Gini, el error cuadrático medio y la validación cruzada, para evaluar la precisión y la eficacia de los modelos de Credit Scoring. Al comparar la forma y el AUC de las curvas ROC de diferentes modelos, los analistas de crédito pueden identificar el modelo más predictivo y tomar decisiones informadas sobre su implementación en la gestión del riesgo crediticio.

### **Error Cuadrático Medio (MSE)**

El Error Cuadrático Medio (MSE, por sus siglas en inglés Mean Squared Error) es una medida comúnmente utilizada en el análisis de Credit Scoring para evaluar la precisión de un modelo de regresión, especialmente en modelos predictivos donde se espera una estimación numérica del riesgo crediticio o de otra variable continua. A continuación,

profundizamos en el concepto y la importancia del MSE en el contexto del Credit Scoring:

### 1. Concepto del Error Cuadrático Medio (MSE)

El MSE es una medida de la discrepancia entre los valores predichos por un modelo y los valores reales observados en los datos. "Se calcula como la media de los cuadrados de las diferencias individuales entre las predicciones del modelo y los valores reales observados." <sup>37</sup> (Mathsisfun, s.f.) Matemáticamente, el MSE se define como:

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

Donde:

- $n \rightarrow$  Es el número total de observaciones en el conjunto de datos.
- $y_i \rightarrow$  Representa el valor real de la variable de interés para la  $i$ -ésima observación.
- $\hat{y}_i \rightarrow$  Representa la predicción del modelo para la  $i$ -ésima observación.

### 2. Interpretación del MSE

El MSE proporciona una medida de cuán cerca están las predicciones del modelo de los valores reales. Cuanto menor sea el valor del MSE, mejor será la precisión del modelo, ya que indica que las predicciones del modelo están más cerca de los valores reales. Por el contrario, un MSE más alto indica que el modelo tiene una mayor discrepancia con los datos reales, lo que sugiere una menor precisión predictiva.

### 3. Importancia del MSE en el Credit Scoring

En el contexto del Credit Scoring, donde se busca predecir el riesgo crediticio de los prestatarios, el MSE es una medida crucial de la precisión de los modelos de regresión utilizados para estimar la probabilidad de incumplimiento. Un MSE bajo indica que el modelo tiene una buena capacidad predictiva y puede proporcionar estimaciones precisas del riesgo crediticio, lo que permite a las instituciones financieras tomar decisiones informadas sobre la concesión de créditos.

### 4. Uso del MSE en la Evaluación de Modelos

El MSE se utiliza comúnmente junto con otras métricas de evaluación, como la Curva ROC y el coeficiente de Gini, para evaluar la precisión y la eficacia de los modelos de Credit Scoring. Al comparar el MSE de diferentes modelos o versiones del mismo modelo, los analistas de crédito pueden identificar el modelo con la mejor capacidad predictiva y tomar decisiones informadas sobre su implementación en la gestión del riesgo crediticio.

### Validación Cruzada

La validación cruzada es una técnica fundamental en el análisis de Credit Scoring que se utiliza para evaluar la capacidad predictiva de un modelo y estimar su rendimiento en datos no vistos. Consiste en dividir el conjunto de datos en subconjuntos de

---

<sup>37</sup> <https://www.mathsisfun.com/definitions/root-mean-square.html>

entrenamiento y prueba, y luego entrenar y evaluar el modelo en múltiples iteraciones, utilizando diferentes combinaciones de los datos. "También se utiliza para optimizar los parámetros de un modelo de Credit Scoring, como el número de variables o la profundidad del árbol de decisión." <sup>38</sup> (Moody's Investor Service:, s.f.)

A continuación, exploramos en detalle la importancia y el proceso de la validación cruzada en el contexto del Credit Scoring:

### 1. Importancia de la Validación Cruzada

La validación cruzada es esencial para evaluar la capacidad de generalización de un modelo de Credit Scoring. Permite estimar cómo se desempeñará el modelo en datos nuevos y no vistos, lo que es crucial para garantizar su fiabilidad y eficacia en la toma de decisiones financieras.

### 2. Proceso de la Validación Cruzada

El proceso de validación cruzada consta de varias etapas:

- **División de los Datos:** Se divide el conjunto de datos en un conjunto de entrenamiento y un conjunto de prueba. Tradicionalmente, se utiliza una proporción de 70-30 o 80-20, donde el conjunto de entrenamiento se utiliza para entrenar el modelo y el conjunto de prueba se reserva para evaluar su rendimiento.
- **Entrenamiento del Modelo:** Se entrena el modelo utilizando el conjunto de entrenamiento. Esto implica ajustar los parámetros del modelo para que se ajusten mejor a los datos de entrenamiento y minimicen el error de predicción.
- **Evaluación del Modelo:** Se evalúa el rendimiento del modelo utilizando el conjunto de prueba. Se calculan diversas métricas de evaluación, como el MSE, la Curva ROC y el coeficiente de Gini, para medir su capacidad predictiva y su ajuste a los datos de prueba.
- **Iteración del Proceso:** El proceso se repite varias veces, utilizando diferentes divisiones de los datos en cada iteración. Esto permite obtener estimaciones más estables y robustas del rendimiento del modelo, al evitar el sesgo introducido por una única división de los datos.
- **Promediado de Resultados:** Finalmente, se promedian los resultados de las diferentes iteraciones para obtener una estimación general del rendimiento del modelo en datos no vistos.

### 3. Beneficios de la Validación Cruzada

La validación cruzada ofrece varios beneficios clave en el análisis de Credit Scoring:

- Proporciona una estimación más precisa del rendimiento del modelo en datos no vistos, lo que ayuda a los analistas de crédito a evaluar su fiabilidad y eficacia en la toma de decisiones financieras.
- Permite identificar y evitar el sobreajuste del modelo, un fenómeno en el que el modelo se ajusta demasiado a los datos de entrenamiento y no generaliza bien a nuevos datos.

---

<sup>38</sup> <https://www.moody.com/>

- Facilita la comparación de diferentes modelos y la selección del mejor modelo en función de su capacidad predictiva y su ajuste a los datos de prueba.

### **Capítulo 3: Análisis del Modelo de Credit Scoring y Resultados**

#### **Armado de la base de datos**

Para realizar el desarrollo del modelo, se decidió trabajar con la simulación de 100 empresas para poder mostrar y reflejar las distintas realidades y posibles resultados.

En primer lugar, se llevó a cabo la configuración de todas nuestras variables, asignándoles valores de manera aleatoria. Esta medida se adoptó para evitar la representación de situaciones reales o la exposición de datos financieros de empresas en actividad. No obstante, se utilizaron parámetros reales para garantizar la adecuada demostración del modelo.

Aquí se aleatorizaron las siguientes variables:

1. Ventas Totales
2. Costo Total de ventas
3. Gastos Operativos Totales
4. Ingreso por Inversiones
5. Ingreso por Intereses
6. Gastos de Intereses
7. Ingreso Neto: Se calcula restando el Total de Ventas, el Costo Total de Ventas y los Gastos Operativos Totales, y luego sumando los Ingresos por Inversiones e Intereses, y restando las Pérdidas por Intereses.
8. Efectivo y Equivalentes de Efectivo
9. Inversiones a Corto Plazo
10. Inventario
11. Cuentas por Cobrar
12. Total de Activos Circulantes: Es la suma del Efectivo y Equivalentes de Efectivo, las Inversiones a Corto Plazo, las Cuentas por Cobrar y una cantidad aleatoria
13. Propiedad y Equipo
14. Inversiones
15. Amortización
16. Depreciación
17. Total de Activos No Circulantes: Se calcula sumando el Valor de la Propiedad y el Equipo, las Inversiones y otra cantidad aleatoria.
18. Intangibles Totales
19. Total de Activos: Es la suma del Total de Activos Circulantes, el Total de Activos No Circulantes y los Activos Intangibles.
20. Flujo de Efectivo Neto proporcionado utilizado por Actividades Operativas
21. Préstamos
22. Total de Pasivos Circulantes: Se obtiene sumando los Préstamos a Corto Plazo y otra cantidad aleatoria
23. Préstamos No Circulantes
24. Total de Pasivos No Circulantes: Es la suma de los Préstamos No Circulantes y otra cantidad aleatoria.
25. Total de Pasivos: Se calcula sumando el Total de Pasivos Circulantes y el Total de Pasivos No Circulantes.
26. Resultados Acumulados
27. Total de Patrimonio Neto: Se obtiene restando el Total de Pasivos del Total de Activos.

28. Ratio de liquidez corriente: Se calcula dividiendo el Total de Activos Circulantes entre el Total de Pasivos Circulantes.
29. Capital de Trabajo: Es la diferencia entre el Total de Activos Circulantes y el Total de Pasivos Circulantes.
30. Deuda sobre Patrimonio Tangible
31. EBITDA: Es la suma del Ingreso Neto, los Ingresos por Intereses, la Depreciación y la Amortización.
32. Margen EBITDA: Se calcula dividiendo el EBITDA entre el Total de Ventas y multiplicándolo por 100.
33. Palanca de Crecimiento: Es el Margen EBITDA dividido entre el Total de Activos.
34. Z-Score
35. Período de Cobro: Se calcula multiplicando las Cuentas por Cobrar entre las Ventas Totales y luego multiplicándolo por 365 días
36. Productividad del KTNO (PKT): Es el Capital de Trabajo Neto dividido entre las Ventas Totales y multiplicado por 100.
37. Rentabilidad del Patrimonio (ROE): Se calcula dividiendo el Ingreso Neto entre el Patrimonio del Propietario
38. Utilidad Operativa Después de Impuestos (UODI): Se obtiene restando los Gastos Operativos Totales del Ingreso Neto.
39. Activos Netos de Operación (ANDEO): Es la suma de los Activos Circulantes y No Circulantes, menos los Pasivos Circulantes.
40. Rentabilidad del Activo Neto (RAN): Se calcula dividiendo la UODI entre los ANDEO.
41. Prueba Ácida: Se calcula restando el Inventario del Total de Activos Circulantes y luego dividiendo este resultado entre los Pasivos Circulantes.
42. Máxima Altura de Mora
43. Contador de Mora a 15 días
44. Reestructuraciones
45. Central de Riesgo: Se establece a partir de la variable Máxima Altura de Mora.
46. Posición Competitiva
47. Planes de Crecimiento
48. Diversificación de Ingresos
49. Calidad de Administración
50. Gestión Financiera y de Riesgos
51. Innovaciones Tecnológicas
52. Políticas de Control y Auditoría
53. Poder de Negociación de Clientes
54. Poder de Negociación de Proveedores
55. Necesidades de Inversión
56. Transparencia y Calidad de Información
57. Certificación de Analista
58. Antigüedad de Estados Financieros
59. Riesgo País
60. Nivel de Riesgo País
61. Estructura del Sector
62. Competencia
63. Riesgos Propios de la Actividad

Además, con el fin de presentar de manera efectiva la distribución de las variables numéricas, se empleó un gráfico de cajas (boxplot), mientras que para las variables categóricas se utilizó un gráfico de barras.

Inicialmente, al aplicar diversas fórmulas y procesos en nuestro modelo, nos encontramos con una variable que denominamos Z"-Score. Esta variable es un indicador que nos muestra la probabilidad de quiebra de una empresa, donde un valor mayor indica una menor probabilidad de quiebra, y viceversa.

Esta variable arrojó valores mínimos de -25 y máximos de 21. Esta amplia gama de valores llevó a tomar la decisión de normalizar la variable para asegurar una interpretación más intuitiva y coherente en nuestro análisis. La normalización de la variable implicó reescalar sus valores para que oscilaran entre 0 y 3, lo que permitió una clasificación más clara y significativa de las empresas en términos de riesgo financiero.

Al establecer estos nuevos límites, se enfatizó la importancia de esta clasificación, dividiendo las empresas en tres categorías fundamentales:

- Alto Riesgo: Z"-Score menor a 1.1
- Neutral: Z"-Score entre 1.1 y 2.6
- Bajo Riesgo: Z"-Score mayor a 2.6

Este proceso de normalización no solo facilitó la interpretación de los resultados, sino que también proporcionó una base sólida para la toma de decisiones informadas en la evaluación del riesgo financiero de las empresas.

Después de haber completado el proceso de normalización de todas nuestras variables, incluido el Z"-Score, procedimos a desarrollar el modelo de Gordon-Shapiro. Este modelo, de gran relevancia en el análisis financiero, nos permitió proyectar el valor presente de las ganancias futuras de cada empresa.

Formula de Gordon-Shapiro:

$$V_{gs} = \frac{EBITDA * Margen EBITDA * Periodo de Cobro}{Tasa de Descuento - Tasa de Crecimiento de Ventas}$$

Para llevar a cabo este proceso, inicialmente aleatorizamos la tasa de crecimiento potencial de las ventas utilizando una distribución normal con un desvío estándar de 0.1 y una media de 0. Esta aleatorización permitió capturar la variabilidad inherente en el crecimiento de las ventas y generar escenarios realistas para su proyección a futuro.

Además, se estableció una tasa de descuento del 5%, que es una práctica común en la valoración de activos financieros. Esta tasa nos permitió descontar adecuadamente las ganancias futuras para obtener su valor presente.

Posteriormente, se definieron umbrales específicos para identificar aquellas empresas con mayor probabilidad de ser consideradas "unicornios". Estos umbrales incluyeron un crecimiento de ventas superior al 20%, un valor presente específico, un margen EBITDA del 5%, un ROE del 5% y una relación deuda-patrimonio de 0.7.

Aquí está la justificación para cada uno de los umbrales:

1. Tasa de Crecimiento de Ventas (20%):

Este umbral se establece en el 20%, lo que indica un crecimiento sólido pero no excesivo en las ventas. Un crecimiento demasiado alto podría ser insostenible o indicar una burbuja temporal en lugar de un crecimiento real y sostenido.

2. Valor Presente (10,000,000):



El umbral del valor presente se establece en 10,000,000, lo que indica que se están considerando empresas con un valor significativo. Este valor se utiliza como un punto de corte para asegurarse de que las empresas seleccionadas tengan un valor suficiente para justificar la inversión.

3. Margen EBITDA (5%):

Este umbral del 5% para el margen EBITDA se selecciona como un indicador de la rentabilidad operativa de la empresa. Un margen EBITDA más bajo puede indicar que la empresa está teniendo dificultades para generar beneficios a partir de sus operaciones.

4. ROE (5%):

El umbral del ROE se establece en el 5%, lo que indica una rentabilidad razonable sobre el patrimonio neto. Un ROE más alto sugiere que la empresa está generando más ingresos con menos inversión, lo que es favorable para los inversionistas.

5. Deuda a Patrimonio (0.7):

Este umbral se establece en 0.7, lo que indica que se prefieren empresas con niveles moderados de endeudamiento en relación con su patrimonio. Un nivel de endeudamiento demasiado alto puede ser una señal de riesgo financiero y puede afectar la estabilidad a largo plazo de la empresa.

Estos umbrales son razonables para identificar empresas con características que podrían ser similares a las de las "unicornio". Sin embargo, es importante recordar que el éxito de una empresa no solo depende de estos indicadores individuales, sino de una combinación de factores, incluidos el modelo de negocio, la innovación, el equipo directivo y las condiciones del mercado.

Al aplicar estos criterios, pudimos identificar a una empresa que destacaban por su buena liquidez, márgenes saludables y otros indicadores financieros positivos.

Después de haber identificado a la empresa que cumplían con los criterios para ser consideradas como posibles "unicornios", continuo nuestro análisis examinando la correlación entre las variables, pero no eliminando todas las variables, manteniendo el total de ventas, el ingreso neto y el capital de trabajo. Este paso fue crucial para identificar y eliminar aquellas variables que estaban altamente correlacionadas entre sí. Sin embargo, mantuvimos constantes dos variables fundamentales: nuestro Z-Score, que es de gran relevancia en nuestro modelo, y la variable binomial "Central de Riesgo", la cual será utilizada en las futuras regresiones y análisis predictivos.

Además, se llevó a cabo la identificación de valores atípicos en las variables numéricas. Estos valores atípicos pueden distorsionar nuestro análisis y afectar la precisión de nuestros modelos. Para abordar este problema:

Se calcularon los límites superiores e inferiores para identificar valores atípicos en todas las variables numéricas, excepto en la variable Z\_Score. Estos límites se calcularon utilizando el método del rango intercuartílico (IQR) junto con el criterio de 1.5 veces el IQR. Se identificaron los valores atípicos en todas las variables numéricas, mediante la comparación con los límites calculados en el paso anterior. Se convirtieron las variables categóricas en variables dummy para facilitar el tratamiento de los valores atípicos. Se reemplazaron los valores atípicos identificados en todas las columnas, con NA (valores faltantes) para su posterior corrección. Y finalmente, se utilizó el método de



winsorización para ajustar los valores atípicos al reemplazar los extremos de la distribución de datos con valores menos extremos pero aún plausibles. Los valores faltantes resultantes de la identificación y corrección de valores atípicos se completaron utilizando la mediana de cada columna, para mantener la integridad y coherencia de los datos.

Al hacerlo, se pudo corregir las distorsiones causadas por los valores atípicos y garantizar que nuestros datos estuvieran dentro de rangos aceptables y realistas.

Para optimizar nuestro análisis y mitigar posibles sesgos, se tomó la decisión de trabajar con los valores en logaritmo.

Trabajar con valores en logaritmo nos ofrece, en primer lugar, a estabilizar la varianza de nuestros datos, especialmente cuando estos muestran una distribución sesgada o heterocedasticidad. Al aplicar el logaritmo, reducimos la escala de las diferencias entre valores extremos, lo que facilita la interpretación y el análisis de los datos.

Además, trabajar con valores en logaritmo puede hacer que las relaciones entre variables sean más lineales, lo que puede simplificar los modelos estadísticos y mejorar su precisión predictiva. Esto es especialmente útil cuando se trabaja con modelos de regresión, donde se asume una relación lineal entre las variables.

Otro beneficio importante es que la transformación logarítmica puede ayudar a reducir la influencia de los valores atípicos en nuestros análisis. Al suavizar la distribución de los datos, los valores extremos tienen menos impacto en las estimaciones de parámetros y en las conclusiones del análisis.

### **Análisis y elección del modelo más adecuado**

El primer análisis se centró en la aplicación de un **Árbol de Decisión** para predecir la variable de respuesta "Central\_de\_Riesgo". El proceso se desglosó en las siguientes etapas: Para ello, se dividió el conjunto de datos en conjuntos de entrenamiento y prueba. En este caso, se asignó el 70% de los datos al conjunto de entrenamiento y el 30% restante al conjunto de prueba.

Luego, se convirtió la variable de respuesta a factor en los datos de entrenamiento. Esto se hizo para asegurarse de que R interprete la variable de respuesta como categórica y no como numérica.

Se definió un esquema de validación cruzada utilizando el método de validación cruzada con 5 repeticiones para evaluar la robustez ante diferentes divisiones de los datos. Este método se utiliza para mejorar la generalización del modelo y reducir el sobreajuste. Se ajustó el modelo de árbol de decisión utilizando los datos de entrenamiento y la validación cruzada.

Posteriormente, se realizaron predicciones en los datos de prueba utilizando el modelo ajustado. Las predicciones se convirtieron a factores con los mismos niveles que los datos de prueba. Finalmente, se calculó la matriz de confusión para evaluar el rendimiento del modelo en los datos de prueba.

### **Modelo y validación del modelo de Árbol de Decisión**

Con base en los resultados obtenidos, podemos sacar las siguientes conclusiones:

1. **Exactitud (Accuracy):** El modelo presenta una exactitud del 70%, lo que indica que clasifica correctamente el 70% de las observaciones en los datos de prueba.

2. **Intervalo de Confianza (CI):** El intervalo de confianza del 95% para la exactitud sugiere una precisión moderada del modelo, con una incertidumbre asociada a la estimación.
3. **Kappa:** El coeficiente kappa es de 0.3836, lo que indica un acuerdo moderado más allá del azar.
4. **Sensibilidad y Especificidad:** La sensibilidad es del 87.50% y la especificidad es del 50%. Esto indica que el modelo identifica correctamente el 87.50% de los casos positivos y el 50% de los casos negativos.
5. **Valor Predictivo Positivo y Negativo:** Tanto el valor predictivo positivo como el negativo son del 66.67% y 77.78%, respectivamente. Esto indica que el 66.67% de las predicciones positivas y el 77.78% de las predicciones negativas son correctas.
6. **Prevalencia y Tasa de Detección:** La prevalencia de la clase positiva en los datos de prueba es del 53.33%. La tasa de detección es del 46.67%, lo que indica que el modelo identifica correctamente el 46.67% de los casos positivos.
7. **Precisión Balanceada (Balanced Accuracy):** La precisión balanceada es del 68.75%, lo que indica un rendimiento moderado del modelo en términos de sensibilidad y especificidad.
8. **Índice de Gini:** El índice de Gini es de 0.538, indica que el modelo tiene un rendimiento razonable en la clasificación de las observaciones en las clases positiva y negativa, aunque hay margen para mejoras.

En resumen, aunque el modelo presenta una capacidad moderada para clasificar las observaciones en los datos de prueba, los resultados sugieren que aún hay margen para mejorar su rendimiento y generalización a nuevos datos.

### **Modelo y validación del Modelo Gradient Boosting**

Por lo tanto, se realiza un **Gradient Boosting** donde también se dividieron en conjunto de entrenamiento y prueba, asignando el 70% de los datos al conjunto de entrenamiento y 30% al restante conjunto de prueba.

Dándonos estos resultados:

1. **Sensibilidad:** La sensibilidad es del 50%, lo que indica que el modelo identifica correctamente el 50% de los casos positivos.
2. **Especificidad:** La especificidad es del 54.55%, lo que indica que el modelo identifica correctamente el 54.55% de los casos negativos.
3. **Índice de Gini:** El índice de Gini es de 0.5827, lo que sugiere que el modelo tiene un rendimiento razonable en la clasificación de las observaciones en las clases positiva y negativa.
4. **Curva ROC:** La curva ROC muestra un AUC (Área bajo la curva) de 0.518, lo que indica un rendimiento moderado del modelo en términos de discriminar entre clases positivas y negativas.

### **Modelo y validación del Modelo Redes Neuronales**

Teniendo en cuenta que mediante este modelo no pudo mejorar nuestro análisis, continuamos con el análisis de **Redes Neuronales**, donde también se repitió la misma forma de modelar los datos de entrenamiento y prueba, el cual nos brinda lo siguiente:

1. **Sensibilidad:** La sensibilidad es del 50%, lo que indica que el modelo identifica correctamente el 50% de los casos positivos.
2. **Especificidad:** La especificidad es del 56.25%, lo que indica que el modelo identifica correctamente el 56.25% de los casos negativos.
3. **Curva ROC de las Redes:** La curva ROC muestra un AUC (Área bajo la curva) de 0.531, lo que indica un rendimiento moderado del modelo en términos de discriminar entre clases positivas y negativas.
4. **Índice de Gini:** Aunque técnicamente es posible calcular el índice de Gini para una red neuronal, no es una práctica común debido a la naturaleza y la complejidad de las redes neuronales.

Aunque el modelo de redes neuronales exhibe cierta capacidad para clasificar las observaciones en los datos de prueba, los resultados indican que aún existen oportunidades para mejorar su rendimiento y capacidad de generalización a nuevos datos.

### Modelo y validación de Support Vector Machines

Luego se prosiguió analizando mediante **Support Vector Machines**.

Para el modelo de Support Vector Machines (SVM), se obtuvieron los siguientes resultados:

1. **Sensibilidad:** La sensibilidad es del 100%, lo que indica que el modelo identifica correctamente el 100% de los casos positivos.
2. **Especificidad:** La especificidad es del 100%, lo que indica que el modelo identifica correctamente el 100% de los casos negativos.
3. **Índice de Gini:** Dado que el índice de Gini no se calcula en el contexto de los modelos SVM, no se proporciona un valor para esta métrica.
4. **Curva ROC:** La curva ROC muestra un AUC (Área bajo la curva) de 0.527, lo que indica un rendimiento moderado del modelo en términos de discriminar entre clases positivas y negativas.

Estos resultados muestran que el modelo SVM tiene una capacidad perfecta para clasificar correctamente todas las observaciones en ambos grupos, lo que sugiere un ajuste adecuado del modelo a los datos de prueba. Sin embargo, el AUC relativamente bajo indica que aún hay margen para mejorar el rendimiento del modelo en términos de discriminar entre las clases positivas y negativas.

Y por último, se realizó el último análisis:

### Modelo y validación de Ridge y Lasso

Para los modelos de Regresión **Ridge y Lasso**, se realizaron las siguientes acciones y se obtuvieron los siguientes resultados:

- Preprocesamiento de datos:
  - a) Se eliminaron las filas con valores faltantes.
  - b) Todas las columnas se convirtieron a variables numéricas.
- División de datos: Los datos se dividieron en conjuntos de entrenamiento y prueba en una proporción del 70% para entrenamiento y el resto para prueba.
- Entrenamiento de modelos:
  - a) Se ajustó un modelo de Regresión Ridge utilizando validación cruzada para encontrar el mejor valor de penalización.
  - b) Se ajustó un modelo de Regresión Lasso también utilizando validación cruzada para determinar el mejor valor de penalización.
- Predicciones en datos de prueba: Se realizaron predicciones en los datos de prueba utilizando los modelos Ridge y Lasso ajustados.
- Evaluación del rendimiento: Se calculó el error cuadrático medio (MSE) para las predicciones en los datos de prueba para ambos modelos.

Los resultados obtenidos fueron los siguientes:

### 1. Regresión Ridge:

MSE en datos de prueba: 0.313

### 2. Regresión Lasso:

MSE en datos de prueba: 0.330

El MSE (Error Cuadrático Medio) más bajo obtenido con el modelo de Regresión Ridge (0.3133) indica que este modelo tiene un mejor rendimiento en la predicción de la variable objetivo (Central de Riesgo) en comparación con el modelo de Regresión Lasso (0.3301). Esto sugiere que el modelo Ridge puede estar mejorando la precisión de las predicciones al penalizar ciertos coeficientes de manera más efectiva.

## Modelo y validación de Regresión Logística

Finalmente se realiza un modelo de **Regresión Logística** para poder compararse con el modelo de árbol de decisión que fue el que mejor convergía. En la comparativa pudimos ver estos resultados:

- **Devianza nula y residual:** La devianza nula es la devianza del modelo nulo (un modelo sin variables predictoras), mientras que la devianza residual es la devianza del modelo ajustado. En el modelo de Regresión Logística, la devianza residual es esencialmente cero, lo que indica un ajuste muy bueno del modelo a los datos. En el modelo de árbol de decisión, la devianza residual es más alta, lo que sugiere un ajuste menos preciso del modelo a los datos.
- **AIC (Criterio de Información de Akaike):** El AIC es una medida de la calidad del modelo que penaliza la complejidad del modelo. Un valor de AIC más bajo indica un mejor modelo. En este caso, el valor de AIC en el modelo de Regresión Logística (90) es más bajo que el valor de AIC en el modelo de árbol de decisión. Esto sugiere que el modelo de Regresión Logística es más preferible según el criterio de AIC.
- **Número de iteraciones de puntuación Fisher:** Indica cuántas iteraciones se necesitaron para que el algoritmo de optimización convergiera. Un número más alto de iteraciones puede indicar que el modelo necesitó más tiempo para converger. En el modelo de Regresión Logística, el número de iteraciones de

puntuación Fisher es 27, lo que indica que el algoritmo de optimización convergió después de 27 iteraciones. En comparación, el modelo de árbol de decisión no tiene un número de iteraciones asociado, ya que no se aplica el mismo algoritmo de optimización.

Basándonos en estas métricas, podríamos concluir que el modelo de Regresión Logística tiene un mejor ajuste a los datos en comparación con el modelo de árbol de decisión.

Para poder decidir cuál es el mejor modelo hay que basarnos en estos puntos clave:

- Rendimiento predictivo: Evalúa cómo se desempeña cada modelo en datos nuevos y no vistos.
- Interpretabilidad: Hay modelos que son más fáciles de interpretar porque ofrecen una comprensión clara de cómo cada variable afecta la predicción. Mientras que otros pueden ser más difíciles de interpretar debido a su complejidad.
- Robustez: Considera la capacidad del modelo para manejar datos ruidosos o atípicos.
- Capacidad para manejar características no lineales o interacciones: Busca que se puedan capturar relaciones no lineales e interacciones entre variables de forma más efectiva.
- Velocidad de entrenamiento y predicción: Se busca que tiendan a ser más rápidos de entrenar y predecir en conjuntos de datos grandes.
- Regularización y control de sobreajuste: Se busca que pueden ser regularizados fácilmente para evitar el sobreajuste, mientras que otros modelos pueden requerir ajustes adicionales, como la profundidad máxima del árbol o el número de estimadores, para evitar el sobreajuste.

### **Desarrollo del modelo**

Luego de elegir el mejor modelo, el cual también refleja la realidad ya que nos indica que las variables en las cuales se basa y con las que mejor predice son: Ventas Totales; Ingreso Neto; Capital de Trabajo; Flujo de Efectivo Neto proporcionado utilizado por Actividades Operativas y Deuda Sobre Patrimonio Tangible. Se comenzó calculando las probabilidades de incumplimiento utilizando un modelo de Regresión Logística.

Las probabilidades más cercanas a 1 se traducirían en calificaciones más altas, como por ejemplo "AAA", lo que indicaría una menor probabilidad de incumplimiento y, por ende, un menor riesgo crediticio. Por otro lado, las probabilidades más bajas reflejarían una mayor posibilidad de incumplimiento y, por lo tanto, se asociarían con calificaciones más bajas, como "BB" o "CCC". Este análisis nos permite evaluar el nivel de riesgo crediticio de cada caso específico, ayudando así en la toma de decisiones financieras fundamentadas.

Posteriormente, se desarrolló una función llamada **asignar\_calificacion()** que asigna calificaciones de riesgo en función de las probabilidades de incumplimiento calculadas. Estas calificaciones se basan en una escala de AAA a D, donde AAA representa el menor riesgo y D el mayor. La función asigna las calificaciones de manera que las probabilidades más altas de incumplimiento correspondan a calificaciones más bajas y viceversa.

Estas escalas se basan en la metodología de calificación crediticia de Fitch Ratings, una de las principales agencias de calificación crediticia a nivel mundial. Estas calificaciones

se utilizan para evaluar el riesgo crediticio de los emisores de deuda y los instrumentos financieros.

- **AAA:** La calificación AAA se otorga a los emisores o instrumentos financieros que tienen la mayor calidad crediticia y el menor riesgo de incumplimiento. Se considera que estos tienen una capacidad excepcional para cumplir con sus obligaciones financieras incluso en condiciones económicas adversas.
- **AA:** Las calificaciones AA se asignan a aquellos emisores o instrumentos financieros que tienen una calidad crediticia muy alta y un riesgo de incumplimiento muy bajo. Aunque ligeramente inferior a AAA, siguen siendo considerados como de muy alta calidad y con una alta capacidad de cumplimiento de sus obligaciones.
- **A:** Las calificaciones A se otorgan a emisores o instrumentos financieros con una calidad crediticia alta y un riesgo de incumplimiento moderadamente bajo. Estos emisores o instrumentos financieros tienen una buena capacidad para cumplir con sus obligaciones, pero pueden estar expuestos a ciertos factores de riesgo.
- **BBB:** Las calificaciones BBB se asignan a emisores o instrumentos financieros con una calidad crediticia buena y un riesgo de incumplimiento moderado. Aunque considerados de grado de inversión, estos emisores o instrumentos financieros pueden ser más susceptibles a los cambios económicos y financieros.
- **BB a D:** Las calificaciones desde BB hasta D se asignan a emisores o instrumentos financieros con una calidad crediticia baja y un riesgo de incumplimiento más alto. Estos emisores o instrumentos financieros pueden tener dificultades para cumplir con sus obligaciones financieras, y las calificaciones más bajas indican un mayor riesgo de incumplimiento, siendo D la calificación más baja, que indica incumplimiento o alta probabilidad de incumplimiento.

Para fines prácticos cada punto de corte se realizó en base a las variables aleatorizadas, pero estos pueden variar según el contexto y el mercado el cual se esté analizando.

Una vez asignadas las calificaciones de riesgo, procedimos a analizar la relación entre los puntajes Z y estas calificaciones mediante un gráfico. Los puntajes Z son una medida estadística que indica cuántas desviaciones estándar un valor particular está por encima o por debajo de la media. La visualización de esta relación nos ayuda a comprender cómo los puntajes Z están distribuidos en función de las calificaciones de riesgo asignadas, lo que puede proporcionar información sobre la relación entre ambos.

Finalmente, realizamos un análisis de varianza (ANOVA) para determinar si hay una diferencia significativa entre los puntajes Z en función de las calificaciones de riesgo asignadas. El ANOVA nos permite evaluar si las diferencias observadas en los puntajes Z son estadísticamente significativas o si pudiesen atribuirse simplemente al azar. Sin embargo, basándonos en el análisis de varianza (ANOVA) realizado, se observó que no hay una diferencia significativa en los puntajes Z entre las diferentes calificaciones de riesgo asignadas (valor  $p = 0.415$ ), es decir, las calificaciones de riesgo pueden no ser útiles para predecir los puntajes Z y viceversa.

Una vez asignadas las calificaciones de riesgo, nuestro siguiente paso consistió en evaluar la elegibilidad de cada cliente para recibir crédito. Para llevar a cabo esta



evaluación, se establecieron categorías de riesgo aprobadas, que comprenden los niveles AAA, AA, A y BBB. En el caso de que un cliente posea una calificación dentro de estas categorías, se considera que el crédito está aprobado para él. Por otro lado, si la calificación del cliente no se encuentra dentro de estas categorías, se considera que el crédito no está aprobado para él.

Para gestionar estos datos de manera más organizada y accesible, creamos un marco de datos denominado "resultados". Este marco de datos almacena información relevante sobre cada cliente, incluyendo el nombre de la empresa, el valor de su puntaje Z, su calificación de riesgo y si se le aprobó o no el crédito. De esta manera, se puede tener una visión completa de la situación crediticia de cada cliente.

Además, para determinar los montos de crédito que pueden ser asignados a cada cliente de manera justa y precisa, se empleó la librería "truncnorm". Esta librería permitió generar montos de crédito ajustados para cada cliente, teniendo en cuenta una serie de factores importantes. Para poder trabajar y presentar resultados, establecimos un límite máximo de USD 100,000 en el monto a fin de reflejar el trabajo realizado. Estos factores incluyen el puntaje Z del cliente, su calificación de riesgo y si la empresa a la que pertenece es considerada un "unicornio" en el ámbito financiero (es decir, una empresa con una valoración de mercado de al menos mil millones de dólares). En el caso de que la calificación del cliente se encuentre dentro de las categorías aprobadas y sea un unicornio, se le asigna el monto máximo de crédito permitido. Sin embargo, si la calificación del cliente no cumple con estos criterios, el monto de crédito se ajusta en función de su calificación y puntaje Z, asegurando así una asignación justa y equitativa.

Finalmente, para facilitar el análisis y la visualización de los resultados, desarrollamos algunas funciones específicas. Estas funciones nos permiten filtrar y examinar los resultados por empresa, lo que nos brinda una comprensión más detallada del desempeño del modelo en diferentes escenarios. Además, estas funciones también nos proporcionan información sobre las variables más predictivas del modelo para cada empresa, lo que nos ayuda a evaluar la asignación de crédito de manera más efectiva y precisa. En resumen, estas herramientas nos permiten realizar un análisis exhaustivo y completo de la asignación de crédito para cada cliente, lo que resulta fundamental para la toma de decisiones financieras informadas y efectivas.

### Resultados y muestras:

Dándonos así 3 posibles resultados en nuestro análisis (obviamente variando la calificación):

Empresa **sin límite** de crédito:

Empresa	Valor_Scoring	Calificación	Resultado	Unicornio	Monto del crédito
YYKWZ	0.8953113	BB	Crédito No Aprobado	FALSE	0

De las 100 empresas trabajadas hubo un total de 53 empresas con el Crédito no Aprobado.

Empresa **con límite** de crédito:

Empresa	Valor_Scoring	Calificación	Resultado	Unicornio	Monto del crédito
UWSBX	2.620475	AAA	Crédito Aprobado	FALSE	86470.58

De las 100 empresas trabajadas hubo un total de 47 empresas con el Crédito Aprobado, dentro de las cuales se encuentran las empresas catalogadas como Unicornios.

Empresa **con límite** de crédito y **Unicornio**:

Empresa	Valor_Scoring	Calificación	Resultado	Unicornio	Monto del crédito
MLYMT	2.907626	AAA	Crédito Aprobado	TRUE	100000

De las 100 empresas trabajadas hubo un total de 2 empresas que resultaron Unicornio.



### **Conclusiones**

Este modelo de Credit Scoring sin historial pasado finalmente logra cumplir su objetivo de evaluar de manera crediticia a un potencial deudor para indicar la solvencia del cliente y poder asignarle un puntaje indicador de su situación/probabilidad de pago para poder tener una anticipación del futuro. Como también logra identificar a posibles “unicornios” para otorgarle mayores límites crediticios y potenciarlos económicamente.

A pesar de las posibles objeciones respecto a su carencia de datos históricos, esta característica le confiere una versatilidad y adaptabilidad excepcionales ante los cambios del entorno. Al prescindir de información previa, evita los sesgos asociados a la extrapolación de tendencias pasadas hacia el futuro, posicionándose como una opción sumamente útil en contextos de transformación rápida y disruptiva, como los mercados financieros y la innovación tecnológica.

Además, su independencia de registros históricos le otorga la capacidad de identificar señales incipientes y tendencias emergentes que podrían pasar desapercibidas para enfoques más convencionales. Su habilidad para adaptarse y aprender de manera continua lo convierte en un recurso valioso para la toma de decisiones estratégicas en un mundo cada vez más complejo y dinámico.

Es crucial destacar que la efectividad de este modelo radica en su capacidad para analizar una amplia diversidad de datos en tiempo real y extraer patrones significativos de forma autónoma. Al combinar técnicas avanzadas de aprendizaje automático con un profundo entendimiento del contexto y la dinámica del mercado, este modelo puede proporcionar predicciones precisas y oportunas que permitan a las organizaciones anticipar y adaptarse proactivamente a los cambios en su entorno competitivo.

En consecuencia, el modelo sin historial pasado representa una evolución significativa en la ciencia de la predicción, ofreciendo una alternativa poderosa y efectiva a los enfoques tradicionales basados en datos históricos. Su capacidad para anticipar el futuro de manera precisa y oportuna lo convierte en un recurso invaluable para la toma de decisiones estratégicas en una amplia variedad de contextos, desde las finanzas hasta la innovación tecnológica. Adoptar este enfoque innovador permite a las organizaciones posicionarse de manera más efectiva para enfrentar los desafíos del futuro y capitalizar las oportunidades emergentes con confianza y éxito.

En este sentido, la intención fundamental de este trabajo va más allá de la mera exploración teórica. Se aspira a convertir este modelo sin historial pasado en una herramienta profesional de gran relevancia, disponible para instituciones financieras u otras entidades interesadas en su aplicación. Se pretende no solo perfeccionar y adaptar el modelo a cada modelo de negocio y mercado específico, sino también fomentar su continuo desarrollo y mejora.

La visión es que este modelo se convierta en un recurso dinámico y adaptable, capaz de evolucionar junto con las necesidades y desafíos cambiantes del entorno económico y empresarial. Su potencial para proporcionar predicciones precisas y oportunas podría ser de gran valor para una amplia gama de industrias y sectores, desde instituciones financieras hasta empresas tecnológicas y organizaciones de investigación.

Al promover la continuidad de este trabajo y su transformación en una herramienta práctica y accesible, se busca contribuir significativamente al avance de la ciencia de la predicción y al fortalecimiento de la capacidad de las organizaciones para tomar decisiones estratégicas fundamentadas en información sólida y perspicaz.

**Bibliografía**

- (s.f.). Obtenido de Corporate Finance Institute: <https://corporatefinanceinstitute.com/>
- (s.f.). Obtenido de Fair Isaac Corporation: <https://www.fico.com/>
- (s.f.). Obtenido de Mathsisfun: <https://www.mathsisfun.com/definitions/root-mean-square.html>
- (s.f.). Obtenido de Moody's Investor Service: <https://www.moody.com/>
- (s.f.). Obtenido de Standard & Poor's: <https://www.spglobal.com/ratings/en/>
- Abadía, J. &. (2013). El scoring crediticio en Argentina: Evolución, desafíos y perspectivas. 21(59), 145-174.\*\*.
- Argentina, B. C. (s.f.). Obtenido de <https://www.bcra.gob.ar/>
- Ayadi, e. a. (2011). The Use of Tornado Analysis for Variable Importance Assessment in Credit Scoring.
- Banco Central de la República Argentina. (2017). *Capitales mínimos de las entidades financieras. Adecuaciones*. BANCO CENTRAL DE LA REPUBLICA ARGENTINA (B.C.R.A). Obtenido de Argentina.gob.ar: <https://www.argentina.gob.ar/normativa/nacional/comunicaci%C3%B3n-6344-2017-318498/texto>
- Bontempi, P. &. (2014). Credit Scoring Using Random Forest: A Comparative Analysis.
- Boosting, A. T. (2014). Chen & He.
- Bruce, P. &. (2017). *Practical Statistics for Data Scientists: 50 Essential Concepts*.
- Chen, M. &. (2020). *Applications of Machine Learning in Credit Risk Management*. In *Machine Learning in Finance*. Springer, Singapore.
- Finlay, S. (2004). *Credit Scoring, Response Modeling, and Insurance Rating: A Practical Guide to Forecasting Consumer Behavior*.
- Gelman, A. C. (2013). *Bayesian Data Analysis*.
- Gujarati, D. N. (2010). *Econometría (5a.ed.--.)*.
- Hand, D. J. (2001). *Statistical methods in credit risk management*. Oxford University Press, USA.
- Hastie, T. a. (2009). *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*.
- Hu et al. (2021). Synthetic Data Generation for Credit Scoring: A Review.
- Huang, e. a. (2010). Regularization Techniques for Credit Scoring.
- Huang, e. a. (2014). Credit Scoring Using Support Vector Machines.
- Jiang, S. &. (2020). The Application of Generative Adversarial Networks in Credit Scoring.
- Kuhn, M. &. (2013). *Applied Predictive Modeling*.
- Li, e. a. (2016). A Survey of Machine Learning Techniques for Credit Scoring.
- Taleb, N. N. (2007). *El Cisne Negro: El impacto de lo altamente improbable*. Random House.
- Thomas, e. a. (2002). A Review of Regression Techniques in Credit Scoring.
- Valores:, C. N. (s.f.). Obtenido de <https://www.argentina.gob.ar/cnv>
- Zhang, e. a. (2018). Application of Deep Learning to Credit Scoring.

### **Agradecimientos**

Este trabajo da fin a un recorrido de casi 6 años donde estoy profundamente agradecido a quienes me formaron y acompañaron en este largo camino.

En primer lugar, agradecerle a mi familia por el apoyo incondicional y a mi pareja que a pesar de no comprender acerca de mi carrera su contención fue crucial para llegar hasta acá.

Deseo también agradecer a mi tutor, Chichel, Ezequiel Cristian, cuya orientación experta, apoyo constante y sabios consejos fueron fundamentales en cada etapa de este proceso de investigación. Su dedicación y compromiso fueron una inspiración para mí y han sido cruciales en la culminación de este trabajo.

También quiero extender mi sincero agradecimiento a Alejandra Di Rado. Aunque no fue mi tutora oficial, su valiosa contribución y apoyo brindaron una perspectiva única y enriquecedora a mi investigación. Su disposición para compartir su conocimiento y experiencia como especialista en estadística matemática y Data Analyst fue un verdadero regalo que ha dejado una huella indeleble en este trabajo.

"El aprendizaje es un tesoro que se lleva consigo toda la vida." (Lao-Tse)

**Anexo:**

El presente estudio emplea el entorno de desarrollo integrado RStudio para llevar a cabo el análisis del modelo de Credit Scoring.

A continuación, se presenta el código utilizado en el análisis del modelo de Credit Scoring:

```
# Establecer la semilla aleatoria
```

```
set.seed(123)
```

```
if (!require(dplyr)) {  
  install.packages("dplyr")  
}  
library(dplyr)
```

```
# Creación de las variables de forma aleatorias con 100 filas
```

```
datos <- data.frame(  
  Total_Sales = runif(n = 100, min = 10000, max = 500000),  
  Total_Cost_of_Goods = runif(n = 100, min = 5000, max = 450000)  
)
```

```
# Generar más variables
```

```
n <- 100 # Número de filas
```

```
datos$Total_Operating_Expenses <- runif(n, min = 10000, max = 50000)
```

```
datos$Investment_Income <- runif(n, min = 500, max = 5000)
```

```
datos$Interest_Income <- runif(n, min = 200, max = 2000)
```

```
datos$Less_Interest_Expense <- runif(n, min = 100, max = 1000)
```

```
datos$Net_Income <- datos$Total_Sales - datos$Total_Cost_of_Goods -  
datos$Total_Operating_Expenses + datos$Investment_Income +  
datos$Interest_Income - datos$Less_Interest_Expense
```

```
datos$Cash_and_Cash_Equivalents <- runif(n, min = 1000, max = 10000)
```

```
datos$Short_Term_Investments <- runif(n, min = 5000, max = 20000)
```

```
datos$Inventario <- runif(n, min = 1000, max = 5000)
```

```
datos$Accounts_Receivable <- runif(n, min = 500, max = 1000)
```

```
datos$Total_Current_Assets <- datos$Cash_and_Cash_Equivalents +  
datos$Short_Term_Investments + datos$Accounts_Receivable + runif(n, min = 1000,  
max = 5000)
```

```
datos$Property_and_Plant_Equipment <- runif(n, min = 50000, max = 200000)
```

```
datos$Investments <- runif(n, min = 2000, max = 10000)
```

```
datos$Amortization <- runif(n, min = 0, max = 5000)
```

```
datos$Depreciation <- runif(n, min = 0, max = 5000)
```

```

datos$Total_Non_Current_Assets <- datos$Property_and_Plant_Equipment +
datos$Investments + runif(n, min = 1000, max = 5000)

datos$Total_Intangibles <- runif(n, min = 1000, max = 5000)

datos$Total_Assets <- datos$Total_Current_Assets + datos$Total_Non_Current_Assets
+ datos$Total_Intangibles

datos$Net_Cash_provided_used_by_Operating_Activities <- runif(n, min = -5000, max
= 5000)

datos$Notes_Payable <- runif(n, min = 10000, max = 50000)

datos$Total_Current_Liabilities <- datos$Notes_Payable + runif(n, min = 1000, max =
5000)

datos$Notes_Payable_Non_current <- runif(n, min = 5000, max = 20000)

datos$Total_Non_Current_Liabilities <- datos$Notes_Payable_Non_current + runif(n,
min = 1000, max = 5000)

datos$Total_Liabilities <- datos$Total_Current_Liabilities +
datos$Total_Non_Current_Liabilities

datos$Retained_Earnings <- runif(n, min = 1000, max = 5000)

datos$Total_Owner_s_Equity <- datos$Total_Assets - datos$Total_Liabilities

datos$Current_Ratio <- datos$Total_Current_Assets / datos$Total_Current_Liabilities

datos$Net_Working_Capital <- datos$Total_Current_Assets -
datos$Total_Current_Liabilities

datos$Debt_to_Tangible_Equity <- runif(n, min = 0.1, max = 1)

datos$EBITDA <- datos$Net_Income + datos$Interest_Income + datos$Depreciation +
datos$Amortization

datos$Margen_EBITDA <- (datos$EBITDA / datos$Total_Sales) * 100

datos$PDC <- datos$Margen_EBITDA / datos$Total_Assets

datos$Z_Score <-
(datos$EBITDA/datos$Total_Assets)*6.72+1.05*(datos$Total_Owner_s_Equity/datos$
Total_Liabilities)+6.56*(datos$Net_Working_Capital/datos$Total_Assets)+3.26*(datos$
Retained_Earnings/datos$Total_Assets)

datos$Periodo_de_Cobro <- (datos$Accounts_Receivable/datos$Total_Sales) * 365

datos$PKT <- datos$Net_Working_Capital / datos$Total_Sales * 100

datos$ROE <- datos$Net_Income / datos$Total_Owner_s_Equity

datos$UODI <- datos$Net_Income - datos$Total_Operating_Expenses

datos$ANDEO <- (datos$Total_Current_Assets + datos$Total_Non_Current_Assets) -
datos$Total_Current_Liabilities

datos$RAN <- datos$UODI / datos$ANDEO

datos$Prueba_Acida <- (datos$Total_Current_Assets -datos$Inventario) /
datos$Total_Current_Liabilities

datos$Maxima_Altura_de_Mora <- runif(n, min = 0, max = 365)

datos$Contador_de_Mora_15_dias <- sample(0:5, n, replace = TRUE)

datos$Reestructuraciones <- sample(0:2, n, replace = TRUE)

```

```
datos$Central_de_Riesgo <- ifelse(datos$Maxima_Altura_de_Mora > 180, 1, 0)
```

```
datos$Posicion_Competitiva <- factor(sample(c("Líder", "Competidor", "Seguidor"), n,  
replace = TRUE))
```

```
levels(datos$Posicion_Competitiva) <- c("1", "2", "3")
```

```
datos$Planes_Crecimiento <- factor(sample(c("Corto Plazo", "Mediano Plazo", "Largo  
Plazo"), n, replace = TRUE))
```

```
levels(datos$Planes_Crecimiento) <- c("1", "2", "3")
```

```
datos$Diversificacion_Ingresos <- factor(sample(c("Geográfica", "Por Productos", "Por  
Líneas de Negocio"), n, replace = TRUE))
```

```
levels(datos$Diversificacion_Ingresos) <- c("1", "2", "3")
```

```
datos$Calidad_Administracion <- factor(sample(c("Baja", "Media", "Alta"), n, replace =  
TRUE))
```

```
levels(datos$Calidad_Administracion) <- c("1", "2", "3")
```

```
datos$Gestion_Financiera_Riesgos <- factor(sample(c("Débil", "Aceptable", "Fuerte"),  
n, replace = TRUE))
```

```
levels(datos$Gestion_Financiera_Riesgos) <- c("1", "2", "3")
```

```
datos$Innovaciones_Tecnologicas <- factor(sample(c("Baja", "Media", "Alta"), n, replace  
= TRUE))
```

```
levels(datos$Innovaciones_Tecnologicas) <- c("1", "2", "3")
```

```
datos$Políticas_Control_Auditoria <- factor(sample(c("Débil", "Aceptable", "Fuerte"), n,  
replace = TRUE))
```

```
levels(datos$Políticas_Control_Auditoria) <- c("1", "2", "3")
```

```
datos$Poder_Negociacion_Clientes <- factor(sample(c("Débil", "Aceptable", "Fuerte"),  
n, replace = TRUE))
```

```
levels(datos$Poder_Negociacion_Clientes) <- c("1", "2", "3")
```

```
datos$Poder_Negociacion_Proveedores <- factor(sample(c("Débil", "Aceptable",  
"Fuerte"), n, replace = TRUE))
```

```
levels(datos$Poder_Negociacion_Proveedores) <- c("1", "2", "3")
```

```
datos$Necesidades_Inversion <- factor(sample(c("Baja", "Media", "Alta"), n, replace =  
TRUE))
```

```
levels(datos$Necesidades_Inversion) <- c("1", "2", "3")
```

```
datos$Transparencia_Calidad_Informacion <- factor(sample(c("Baja", "Media", "Alta"),  
n, replace = TRUE))
```

```
levels(datos$Transparencia_Calidad_Informacion) <- c("1", "2", "3")
```

```
datos$Certificacion_Analista <- factor(sample(c("Sí", "No"), nrow(datos), replace =  
TRUE))
```

```
datos$Antiguedad_Estados_Financieros <- factor(sample(c("Anual", "Cuatrimestral",  
"Mensual", "Trimestral"), nrow(datos), replace = TRUE))
```

```
datos$Riesgo_Pais <- runif(n, min = 1, max = 1700) # Valores posibles entre 1 y 3000  
para reflejar la realidad de América Latina
```

```
datos$Nivel_Riesgo_Pais <- cut(datos$Riesgo_Pais, breaks = c(0, 450, 900, 1700),  
labels = c("Bajo", "Medio", "Alto"), include.lowest = TRUE)
```

```
datos$Estructura_Sector <- factor(sample(c("Concentrado", "Competitivo",  
"Fragmentado"), n, replace = TRUE))
```

```
levels(datos$Estructura_Sector) <- c("1", "2", "3")
```

```
datos$Competencia <- factor(sample(c("Baja", "Media", "Alta"), n, replace = TRUE))
```

```
levels(datos$Competencia) <- c("1", "2", "3")
```

```
datos$Riesgos_Propios_Actividad <- factor(sample(c("Bajos", "Moderados", "Altos"), n,  
replace = TRUE))
```

```
levels(datos$Riesgos_Propios_Actividad) <- c("1", "2", "3")
```

```
# Normalizar la columna de Score para que esté en un rango entre 0 y 3
```

```
# Definir la función para normalizar y asignar categorías de riesgo
```

```
normalizar_a_rango_0_3 <- function(Z_Score) {
```

```
  # Función para asignar la categoría de riesgo
```

```
  asignar_categoria <- function(x) {
```

```
    if (x < 1.10) {
```

```
      return("High Risk")
```

```
    } else if (x <= 2.60) {
```

```
      return("Neutral")
```

```
    } else {
```

```
      return("Low Risk")
```

```
    }
```



```
}
```

```
# Normalizar los valores de Z_Score al rango de 0 a 3
valores_normalizados <- numeric(length(Z_Score))
for (i in seq_along(Z_Score)) {
  if (Z_Score[i] < 1.10) {
    valores_normalizados[i] <- (Z_Score[i] - min(Z_Score)) * (1 / (1.10 - min(Z_Score)))
  } else if (Z_Score[i] <= 2.60) {
    valores_normalizados[i] <- (Z_Score[i] - 1.10) * (1 / (2.60 - 1.10)) + 1
  } else {
    valores_normalizados[i] <- (Z_Score[i] - 2.60) * (1 / (max(Z_Score) - 2.60)) + 2.6
  }
}
```

```
# Asignar categoría de riesgo a cada valor
categorias_riesgo <- sapply(Z_Score, asignar_categoria)

return(list(valores_normalizados, categorias_riesgo))
}
```

```
resultado <- normalizar_a_rango_0_3(datos$Z_Score)
valores_normalizados <- resultado[[1]]
categorias_riesgo <- resultado[[2]]
```

```
# Asegurar que los valores normalizados estén dentro del rango de 0 a 3
valores_normalizados[valores_normalizados < 0] <- 0
valores_normalizados[valores_normalizados > 3] <- 3
```

```
# Mostrar los valores normalizados y las categorías de riesgo
print(valores_normalizados)
print(categorias_riesgo)
```

```
print(datos$Z_Score)
```

```
# Reemplazar los Z-Score originales con los nuevos valores normalizados
datos$Z_Score <- valores_normalizados
```

```
# Verificar que se haya realizado correctamente el reemplazo  
print(datos$Z_Score)
```

```
# Convertir la matriz o array a un data.frame  
datos <- as.data.frame(datos)
```

```
sapply(datos, class)
```

```
# Convertir todas las columnas a numéricas  
datos_numericos <- sapply(datos, as.numeric)
```

```
#GRAFICAR FUNCIONES  
#VARIABLES NUMÉRICAS
```

```
colnames(datos)
```

```
boxplot(datos$Total_Sales)  
boxplot(datos$Total_Cost_of_Goods)  
boxplot(datos$Total_Operating_Expenses)  
boxplot(datos$Investment_Income)  
boxplot(datos$Interest_Income)  
boxplot(datos$Less_Interest_Expense)  
boxplot(datos$Net_Income)  
boxplot(datos$Cash_and_Cash_Equivalents)  
boxplot(datos$Short_Term_Investments)  
boxplot(datos$Inventario)  
boxplot(datos$Accounts_Receivable)  
boxplot(datos$Total_Current_Assets)  
boxplot(datos$Property_and_Plant_Equipment)  
boxplot(datos$Investments)  
boxplot(datos$Amortization)  
boxplot(datos$Depreciation)  
boxplot(datos$Total_Non_Current_Assets)  
boxplot(datos$Total_Intangibles)  
boxplot(datos$Total_Assets)
```

```

boxplot(datos$Net_Cash_provided_used_by_Operating_Activities)
boxplot(datos$Notes_Payable)
boxplot(datos$Total_Current_Liabilities)
boxplot(datos$Notes_Payable_Non_current)
boxplot(datos$Total_Non_Current_Liabilities)
boxplot(datos$Total_Liabilities)
boxplot(datos$Retained_Earnings)
boxplot(datos$Total_Owner_s_Equity)
boxplot(datos$Current_Ratio)
boxplot(datos$Net_Working_Capital)
boxplot(datos$Debt_to_Tangible_Equity)
boxplot(datos$EBITDA)
boxplot(datos$Margen_EBITDA)
boxplot(datos$PDC)
boxplot(datos$Z_Score)
boxplot(datos$Periodo_de_Cobro)
boxplot(datos$PKT)
boxplot(datos$ROE)
boxplot(datos$UODI)
boxplot(datos$ANDEO)
boxplot(datos$RAN)

```

#### #VARIABLES CATEGORICAS

```
library(ggplot2)
```

# Lista de variables categóricas

```

variables_categoricas <- c("Prueba_Acida", "Maxima_Altura_de_Mora",
"Contador_de_Mora_15_dias",
"Reestructuraciones", "Central_de_Riesgo", "Posicion_Competitiva",
"Planes_Crecimiento", "Diversificacion_Ingresos",
"Calidad_Administracion",
"Gestion_Financiera_Riesgos", "Innovaciones_Tecnologicas",
"Políticas_Control_Auditoria",
"Poder_Negociacion_Clientes", "Poder_Negociacion_Proveedores",
"Necesidades_Inversion",
"Transparencia_Calidad_Informacion", "Certificacion_Analista",
"Antigüedad_Estados_Financieros",
"Riesgo_Pais", "Nivel_Riesgo_Pais", "Estructura_Sector",
"Competencia",

```

"Riesgos\_Propios\_Actividad")

```
# Crear un gráfico de barras para cada variable categórica
```

```
for (variable in variables_categoricas) {
```

```
  # Crear un dataframe con la frecuencia de cada categoría
```

```
  frecuencias <- data.frame(table(datos[[variable]]))
```

```
  # Ordenar las categorías por frecuencia descendente
```

```
  frecuencias <- frecuencias[order(-frecuencias$Freq), ]
```

```
  # Crear un gráfico de barras
```

```
  plot <- ggplot(frecuencias, aes(x = Var1, y = Freq)) +
```

```
    geom_bar(stat = "identity", fill = "skyblue") +
```

```
    labs(title = variable, x = "Categoría", y = "Frecuencia")
```

```
  # Mostrar el gráfico
```

```
  print(plot)
```

```
}
```

```
#MODELO GORDON-SHAPIRO
```

```
set.seed(123)
```

```
# Generar 100 muestras de la tasa de crecimiento aleatoria utilizando una distribución normal
```

```
datos_gordon <- data.frame(
```

```
  Tasa_Crecimiento_Ventas = rnorm(n, mean = 0, sd = 0.10),
```

```
  Total_Sales = datos$Total_Sales,
```

```
  Net_Income = datos$Net_Income,
```

```
  Total_Assets = datos$Total_Assets,
```

```
  Total_Owner_s_Equity = datos$Total_Owner_s_Equity,
```

```
  Current_Ratio = datos$Current_Ratio,
```

```
  Debt_to_Tangible_Equity = datos$Debt_to_Tangible_Equity,
```

```
  EBITDA = datos$EBITDA,
```

```
  Margen_EBITDA = datos$Margen_EBITDA,
```

```
  Periodo_de_Cobro = datos$Periodo_de_Cobro,
```

```
  ROE = datos$ROE,
```

```
  Central_de_Riesgo = datos$Central_de_Riesgo,
```

```
Nombres_Empresas = replicate(100, paste(sample(LETTERS, 5, replace = TRUE),  
collapse = ""))
```

```
)
```

```
print(datos_gordon$Nombres_Empresas)
```

```
# Defino la tasa de descuento
```

```
tasa_descuento <- 0.05 # 5%
```

```
# Calcular el valor presente de las ganancias futuras para cada empresa
```

```
valor_presente_gordon_shapiro <- function(total_sales, tasa_crecimiento,  
margen_ebitda, periodo_cobro, tasa_descuento) {
```

```
  # Utilizar el modelo de Gordon-Shapiro para calcular el valor presente
```

```
  valor_presente <- (total_sales * margen_ebitda * periodo_cobro) / (tasa_descuento -  
tasa_crecimiento)
```

```
  # Tomar el valor absoluto del resultado para asegurar que sea positivo
```

```
  valor_presente <- abs(valor_presente)
```

```
  return(valor_presente)
```

```
}
```

```
# Calcular el valor presente de las ganancias futuras para cada empresa en la muestra
```

```
valor_presente_gordon_shapiro <-  
valor_presente_gordon_shapiro(datos_gordon$Total_Sales,  
datos_gordon$Tasa_Crecimiento_Ventas, datos_gordon$Margen_EBITDA,  
datos_gordon$Periodo_de_Cobro, tasa_descuento)
```

```
# Mostrar los resultados
```

```
print(valor_presente_gordon_shapiro)
```

```
sapply(valor_presente_gordon_shapiro, class)
```

```
datos_gordon$ValorGordon_Shapiro <- valor_presente_gordon_shapiro
```

```
print(datos_gordon)
```

```
# Establecer los umbrales para cada métrica
```

```
umbral_crecimiento_ventas <- 0.2 # 20%
```

```
umbral_valor_presente <- 10000000
```

```
umbral_margen_ebitda <- 10 # 10%
umbral_roe <- 0.5 # 5%
umbral_deuda_patrimonio <- 0.7

# Filtrar las empresas que cumplen con los criterios
empresas_unicornio <- datos_gordon[
  datos_gordon$Tasa_Crecimiento_Ventas > umbral_crecimiento_ventas &
  datos_gordon$ValorGordon_Shapiro > umbral_valor_presente &
  datos_gordon$Margen_EBITDA > umbral_margen_ebitda &
  datos_gordon$ROE > umbral_roe &
  datos_gordon$Debt_to_Tangible_Equity < umbral_deuda_patrimonio,
]

# Mostrar las empresas que cumplen con los criterios
print(empresas_unicornio)

if (!require(caret)) {
  install.packages("caret")
}
library(caret)

colnames(datos)

# ELIMINAR VARIABLES ALTAMENTE CORRELACIONADAS
datos_cor <- cor(datos_numericos)
highly_correlated <- findCorrelation(datos_cor, cutoff = 0.7)
print(highly_correlated)

# Mantener Z_Score en las variables altamente correlacionadas si está presente
if ("Z_Score" %in% rownames(datos_cor)) {
  highly_correlated <- setdiff(highly_correlated, which(rownames(datos_cor) ==
"Central_de_Riesgo"))
}

# Mantener Central_de_Riesgo en las variables altamente correlacionadas si está
presente
if ("Central_de_Riesgo" %in% rownames(datos_cor)) {
```

```
highly_correlated <- setdiff(highly_correlated, which(rownames(datos_cor) ==  
"Z_Score"))  
}
```

```
# Asegurar que se mantengan estas variables
```

```
variables_a_mantener <- c("Total_Sales",  
                          "Net_Income",  
                          "Net_Cash_provided_used_by_Operating_Activities",  
                          "Net_Working_Capital",  
                          "Z_Score",  
                          "Central_de_Riesgo")
```

```
# Eliminar las variables altamente correlacionadas excepto las requeridas
```

```
highly_correlated <- setdiff(highly_correlated, which(colnames(datos_cor) %in%  
variables_a_mantener))  
datos_filtered <- datos[, -highly_correlated]
```

```
# Agregar las variables a mantener al conjunto de datos filtrados
```

```
variables_a_agregar <- setdiff(variables_a_mantener, colnames(datos_filtered))  
datos_filtered <- cbind(datos_filtered, datos[, variables_a_agregar])  
print(datos_filtered)
```

```
# Convertir todas las columnas a numéricas
```

```
datos <- as.data.frame(lapply(datos_filtered, as.numeric))
```

```
# Verificar si la conversión fue exitosa
```

```
if (all(sapply(datos_numericos, is.numeric))) {  
  print("Todas las columnas se han convertido correctamente a numéricas.")  
} else {  
  print("Ha ocurrido un error en la conversión. Revisa tus datos.")  
}
```

```
sapply(datos, class)
```

```
datos <- as.data.frame(datos)
```

```
# Calcular los límites para identificar valores atípicos en todas las variables numéricas
```



```
limites_atipicos <- apply(datos[, !names(datos) %in% "Z_Score"], 2, function(x) {  
  Q1 <- quantile(x, 0.25)  
  Q3 <- quantile(x, 0.75)  
  IQR <- Q3 - Q1  
  limite_superior <- Q3 + 1.5 * IQR  
  limite_inferior <- Q1 - 1.5 * IQR  
  return(list(limite_superior = limite_superior, limite_inferior = limite_inferior))  
})
```

```
# Identificar valores atípicos en todas las variables numéricas excepto Z_Score  
valores_atipicos <- lapply(1:ncol(datos[, !names(datos) %in% "Z_Score"]), function(i) {  
  x <- datos[, !names(datos) %in% "Z_Score"][, i]  
  limite_superior <- limites_atipicos[[i]]$limite_superior  
  limite_inferior <- limites_atipicos[[i]]$limite_inferior  
  return(x[x < limite_inferior | x > limite_superior])  
})
```

```
print(valores_atipicos)
```

```
# Convertir factores a variables dummy
```

```
datos_dummy <- as.data.frame(model.matrix(~., datos[, !names(datos) %in%  
"Z_Score"]))
```

```
# Identificar valores atípicos
```

```
limites_atipicos <- lapply(datos_dummy, function(x) {  
  Q1 <- quantile(x, 0.25, na.rm = TRUE)  
  Q3 <- quantile(x, 0.75, na.rm = TRUE)  
  IQR <- Q3 - Q1  
  limite_superior <- Q3 + 1.5 * IQR  
  limite_inferior <- Q1 - 1.5 * IQR  
  list(limite_superior = limite_superior, limite_inferior = limite_inferior)  
})
```

```
# Reemplazar los valores atípicos con NA en todas las columnas excepto Z_Score
```

```
for (i in 1:ncol(datos_dummy)) {  
  limite_superior <- limites_atipicos[[i]]$limite_superior  
  limite_inferior <- limites_atipicos[[i]]$limite_inferior
```

```
datos_dummy[datos_dummy[, i] > limite_superior | datos_dummy[, i] < limite_inferior, i]  
<- NA  
}
```

```
# Instalar y cargar el paquete zoo si no está instalado
```

```
if (!require(zoo)) {  
  install.packages("zoo")  
}  
library(zoo)
```

```
# Aplicar la técnica de winsorización excepto a Z_Score
```

```
datos_winsorizados <- datos_dummy
```

```
# Obtener el índice de la columna de Z_Score
```

```
z_score_index <- which(colnames(datos) == "Z_Score")
```

```
for (i in 1:ncol(datos_dummy)) {
```

```
  Q1 <- quantile(datos_dummy[, i], 0.25, na.rm = TRUE)
```

```
  Q3 <- quantile(datos_dummy[, i], 0.75, na.rm = TRUE)
```

```
  IQR <- Q3 - Q1
```

```
  limite_superior <- Q3 + 1.5 * IQR
```

```
  limite_inferior <- Q1 - 1.5 * IQR
```

```
  if (i != z_score_index) {
```

```
    datos_winsorizados[, i][datos_dummy[, i] > limite_superior] <- limite_superior
```

```
    datos_winsorizados[, i][datos_dummy[, i] < limite_inferior] <- limite_inferior
```

```
  }
```

```
}
```

```
# Reasignar la columna de Z_Score
```

```
datos_winsorizados <- cbind(datos_winsorizados, Z_Score = datos[, "Z_Score"])
```

```
# Completar los valores NA con la mediana de cada columna excepto Z_Score
```

```
datos_winsorizados <- na.aggregate(datos_winsorizados, FUN = median, na.rm =  
FALSE)
```

```
# Imprimir los datos winsorizados
```

```
print(datos_winsorizados)
```

```
# Convertir todas las columnas a numéricas
```

```
datos <- as.data.frame(lapply(datos_winsorizados, as.numeric))
```

```
# Verificar si la conversión fue exitosa
```

```
if (all(sapply(datos_numericos, is.numeric))) {
```

```
  print("Todas las columnas se han convertido correctamente a numéricas.")
```

```
} else {
```

```
  print("Ha ocurrido un error en la conversión. Revisa tus datos.")
```

```
}
```

```
sapply(datos, class)
```

```
datos <- as.data.frame(datos)
```

```
colnames(datos)
```

```
# Definir las variables a transformar
```

```
variables_a_transformar <- c("Total_Sales", "Total_Cost_of_Goods",  
"Total_Operating_Expenses", "Investment_Income",
```

```
"Interest_Income", "Less_Interest_Expense", "Net_Income",  
"Cash_and_Cash_Equivalents",
```

```
"Short_Term_Investments", "Inventario", "Accounts_Receivable",
```

```
"Property_and_Plant_Equipment", "Investments", "Amortization",
```

```
"Depreciation", "Total_Intangibles",
```

```
"Net_Cash_provided_used_by_Operating_Activities",
```

```
"Notes_Payable",
```

```
"Notes_Payable_Non_current", "Retained_Earnings",
```

```
"Net_Working_Capital", "Debt_to_Tangible_Equity",
```

```
"Periodo_de_Cobro", "PKT", "ROE", "Contador_de_Mora_15_dias",  
"Reestructuraciones",
```

```
"Posicion_Competitiva", "Planes_Crecimiento",  
"Diversificacion_Ingresos",
```

```
"Calidad_Administracion", "Gestion_Financiera_Riesgos",  
"Innovaciones_Tecnologicas",
```

```
"Políticas_Control_Auditoria", "Poder_Negociacion_Clientes",  
"Poder_Negociacion_Proveedores",
```

```
"Necesidades_Inversion", "Transparencia_Calidad_Informacion",  
"Certificacion_Analista",
```

```
"Antigüedad_Estados_Financieros", "Estructura_Sector",  
"Competencia",
```

"Riesgos\_Propios\_Actividad")

```
# Ajustar los valores negativos o ceros
for (variable in variables_a_transformar) {
  # Obtener el valor mínimo de la variable
  min_value <- min(datos[[variable]])

  # Verificar si el valor mínimo es negativo o cero
  if (min_value <= 0) {
    # Sumar 1 más el valor absoluto del valor mínimo para que todos los valores sean
    positivos
    datos[[variable]] <- datos[[variable]] - min_value + 1
  }
}

print(datos)

# Aplicar transformación logarítmica a las variables numéricas
for (variable in variables_a_transformar) {
  if (is.numeric(datos[[variable]])) {
    datos[[paste0(variable, "_log")]] <- log(datos[[variable]])
  }
}

print(datos)

# Obtener los índices de las columnas que tienen "_log" al final de su nombre
indices_log <- grep("_log$", colnames(datos))

# Conservar la columna "Central_de_Riesgo"
indices_central_de_riesgo <- grep("Central_de_Riesgo", colnames(datos))

# Combinar los índices de las columnas "_log", "Z_Score" y "Central_de_Riesgo"
indices_a_mantener <- c(indices_log, indices_central_de_riesgo,
  which(colnames(datos) == "Z_Score"))
```

```
# Seleccionar solo las columnas que cumplen con los criterios
datos_filtrados <- datos[, indices_a_mantener]

# Verificar el resultado
print(datos_filtrados)

datos <- datos_filtrados

print(datos)

datos <- as.data.frame(datos)

if (!require(dplyr)) {
  install.packages("dplyr")
}
library(dplyr)

if (!require(rpart)) {
  install.packages("rpart")
}
if (!require(rpart.plot)) {
  install.packages("rpart.plot")
}

library(rpart)
library(rpart.plot)

#ARBOL DE DECISION
set.seed(123)

# Dividir los datos en conjuntos de entrenamiento y prueba
indices_entrenamiento <- createDataPartition(datos$Central_de_Riesgo, p = 0.7, list =
FALSE)
datos_entrenamiento <- datos[indices_entrenamiento, ]
datos_prueba <- datos[-indices_entrenamiento, ]

# Convertir la variable de respuesta a factor en ambos conjuntos de datos
```

```
datos_entrenamiento$Central_de_Riesgo
factor(datos_entrenamiento$Central_de_Riesgo, levels = c("0", "1"))

datos_prueba$Central_de_Riesgo <- factor(datos_prueba$Central_de_Riesgo, levels =
c("0", "1"))

# Calculamos los pesos de clase
pesos_clase <- ifelse(datos_entrenamiento$Central_de_Riesgo == "0", 0.7, 0.3)

# Definir el esquema de validación cruzada
control <- trainControl(method = "repeatedcv",
                        number = 5,      # Número de repeticiones de validación cruzada
                        repeats = 1)     # Número de repeticiones del proceso completo de
validación cruzada

# Ajustar el modelo de árbol de decisión con los datos de entrenamiento y validación
cruzada
modelo_arbol <- train(Central_de_Riesgo ~ .,
                      data = datos_entrenamiento,
                      method = "rpart",  # Método de ajuste del modelo: árbol de decisión
                      trControl = control, # Especificar el esquema de validación cruzada
                      tuneGrid = expand.grid(cp = seq(0.01, 0.5, by = 0.01)), # Rango de valores
para el parámetro cp
                      weights = pesos_clase) # Asignar pesos de clase

# Realizar predicciones en los datos de prueba
predicciones <- predict(modelo_arbol, newdata = datos_prueba)

# Calcular la matriz de confusión nuevamente
conf_matrix <- confusionMatrix(predicciones, datos_prueba$Central_de_Riesgo)
conf_matrix

# Extraer la sensibilidad y especificidad de la matriz de confusión
sensibilidad <- conf_matrix$byClass['Sensitivity']
especificidad <- conf_matrix$byClass['Specificity']

# Imprimir los resultados
print(paste("Sensibilidad:", sensibilidad))
print(paste("Especificidad:", especificidad))
```

```
if (!require(DescTools)) {  
  install.packages("DescTools")  
}  
library(DescTools)  
  
# Convertir las predicciones a numérico  
predicciones_numeric <- as.numeric(as.character(predicciones))  
  
# Convertir los datos de prueba a numérico  
datos_prueba_numeric <- as.numeric(as.character(datos_prueba$Central_de_Riesgo))  
  
# Calcular el índice de Gini  
gini_arbol <- Gini(ifelse(predicciones == "1", 1, 0),  
  ifelse(datos_prueba$Central_de_Riesgo == "1", 1, 0))  
print(paste("Índice de Gini del Árbol de Decisiones:", gini_arbol))  
  
# Evaluar el rendimiento del modelo (Curvas ROC)  
if (!require(pROC)) {  
  install.packages("pROC")  
}  
library(pROC)  
  
# Construir la curva ROC  
roc_curve <- roc(ifelse(predicciones == "1", 1, 0),  
  as.numeric(datos_prueba$Central_de_Riesgo))  
  
# Plotear la curva ROC  
plot(roc_curve, main = "Curva ROC", print.auc = TRUE)  
  
#Gradient Boosting  
  
set.seed(123)  
if (!require(gbm)) {  
  install.packages("gbm")  
}  
library(gbm)
```



```
# Dividir los datos en conjuntos de entrenamiento y prueba

indices_entrenamiento_gbm <- createDataPartition(datos$Central_de_Riesgo, p = 0.7,
list = FALSE)

datos_entrenamiento_gbm <- datos[indices_entrenamiento_gbm, ]
datos_prueba_gbm <- datos[-indices_entrenamiento_gbm, ]

# Entrenar el modelo de Gradient Boosting

modelo_gbm <- gbm(Central_de_Riesgo ~ .,
  data = datos_entrenamiento_gbm,
  distribution = "bernoulli",
  n.trees = 100,
  interaction.depth = 3,
  shrinkage = 0.1) # Ajusta la tasa de aprendizaje según sea necesario

# Obtener las predicciones del modelo en los datos de prueba

predicciones_gbm <- predict(modelo_gbm, datos_prueba_gbm, n.trees = 100, type =
"response")

# Convertir las predicciones a clases (0 o 1)

predicciones_clases_gbm <- ifelse(predicciones_gbm > 0.5, 1, 0)

# Calcular la matriz de confusión

conf_matrix_gbm <- table(predicciones_clases_gbm,
  datos_prueba_gbm$Central_de_Riesgo)

# Calcular la sensibilidad y especificidad

sensibilidad_gbm <- conf_matrix_gbm[2, 2] / sum(conf_matrix_gbm[2, ])
especificidad_gbm <- conf_matrix_gbm[1, 1] / sum(conf_matrix_gbm[1, ])

# Imprimir los resultados

print(paste("Sensibilidad:", sensibilidad_gbm))
print(paste("Especificidad:", especificidad_gbm))

# Calcular el índice de Gini

gini <- Gini(as.numeric(datos_prueba_gbm$Central_de_Riesgo), predicciones_gbm)
```

```
# Imprimir el índice de Gini
print(paste("Índice de Gini Gradient:", gini))

# Calcular la curva ROC del Gradient
roc_curve_gbm      <-      roc(as.numeric(datos_prueba_gbm$Central_de_Riesgo),
as.numeric(predicciones_clases_gbm))

# Graficar la curva ROC del Gradient
plot(roc_curve_gbm, main = "Curva ROC", print.auc = TRUE)

#Redes Neuronales

set.seed(123)
if (!require(neuralnet)) {
  install.packages("neuralnet")
}
library(neuralnet)

# Dividir los datos en conjuntos de entrenamiento y prueba
indices_entrenamiento_ann <- createDataPartition(datos$Central_de_Riesgo, p = 0.7,
list = FALSE)

datos_entrenamiento_ann <- datos[indices_entrenamiento_ann, ]
datos_prueba_ann <- datos[-indices_entrenamiento_ann, ]

# Entrenar la red neuronal artificial
modelo_ann <- neuralnet(Central_de_Riesgo ~ ., data = datos_entrenamiento_ann,
hidden = c(5, 2), linear.output = FALSE)

# Realizar predicciones en los datos de prueba
predicciones_ann <- predict(modelo_ann, datos_prueba_ann)

# Convertir las predicciones a clases (0 o 1)
predicciones_clases_ann <- ifelse(predicciones_ann > 0.5, 1, 0)

# Calcular la matriz de confusión
conf_matrix_ann      <-      table(predicciones_clases_ann,
datos_prueba_ann$Central_de_Riesgo)
```

```
conf_matrix_ann
```

```
# Calcular la sensibilidad y especificidad
```

```
sensibilidad_ann <- conf_matrix_ann[2, 2] / sum(conf_matrix_ann[2, ])
```

```
especificidad_ann <- conf_matrix_ann[1, 1] / sum(conf_matrix_ann[1, ])
```

```
# Imprimir los resultados
```

```
print(paste("Sensibilidad:", sensibilidad_ann))
```

```
print(paste("Especificidad:", especificidad_ann))
```

```
# Calcular la curva ROC de las Redes
```

```
roc_curve_ann <- roc(as.numeric(datos_prueba_ann$Central_de_Riesgo),  
as.numeric(predicciones_clases_ann))
```

```
# Graficar la curva ROC de las Redes
```

```
plot(roc_curve_ann, main = "Curva ROC", print.auc = TRUE)
```

```
#Support Vector Machines (SVM)
```

```
set.seed(123)
```

```
if (!require(e1071)) {
```

```
  install.packages("e1071")
```

```
}
```

```
library(e1071)
```

```
# Dividir los datos en conjuntos de entrenamiento y prueba
```

```
indices_entrenamiento_svm <- createDataPartition(datos$Central_de_Riesgo, p = 0.7,  
list = FALSE)
```

```
datos_entrenamiento_svm <- datos[indices_entrenamiento_svm, ]
```

```
datos_prueba_svm <- datos[-indices_entrenamiento_svm, ]
```

```
# Entrenar el modelo SVM
```

```
modelo_svm <- svm(Central_de_Riesgo ~ ., data = datos_entrenamiento_svm, kernel =  
"radial")
```

```
# Realizar predicciones en los datos de prueba
```

```
predicciones_svm <- predict(modelo_svm, datos_prueba_svm)
```

```
# Calcular la matriz de confusión
conf_matrix_svm <- table(predicciones_svm, datos_prueba_svm$Central_de_Riesgo)
conf_matrix_svm

# Calcular la sensibilidad y especificidad
sensibilidad_svm <- conf_matrix_svm[2, 2] / sum(conf_matrix_svm[2, ])
especificidad_svm <- conf_matrix_svm[1, 1] / sum(conf_matrix_svm[1, ])

# Imprimir los resultados
print(paste("Sensibilidad:", sensibilidad_svm))
print(paste("Especificidad:", especificidad_svm))

# Calcular la curva ROC de SVM
roc_curve_svm <- roc(as.numeric(datos_prueba_svm$Central_de_Riesgo),
as.numeric(predicciones_svm))

# Graficar la curva ROC de SVM
plot(roc_curve_svm, main = "Curva ROC", print.auc = TRUE)

#REGRESION RIDGE & LASSO

set.seed(123)
if (!require(glmnet)) {
  install.packages("glmnet")
}
library(glmnet)

# Eliminar filas con valores faltantes y convertir todas las columnas a numéricas
datos_sin_na <- na.omit(datos)
datos_numericos <- sapply(datos_sin_na, as.numeric)

# Dividir los datos en conjuntos de entrenamiento y prueba
indices_entrenamiento_r_l <- createDataPartition(datos$Central_de_Riesgo, p = 0.7,
list = FALSE)
datos_entrenamiento_r_l <- datos[indices_entrenamiento_r_l, ]
datos_prueba_r_l <- datos[-indices_entrenamiento_r_l, ]
```

```
# Preparar los predictores y la variable respuesta
x <- as.matrix(datos_entrenamiento_r_l[, !colnames(datos_entrenamiento_r_l) %in%
"Central_de_Riesgo"])
y <- datos_entrenamiento_r_l$Central_de_Riesgo

# Ajustar un modelo de regresión Ridge
modelo_ridge <- cv.glmnet(x, y, alpha = 0)

# Ajustar un modelo de regresión Lasso
modelo_lasso <- cv.glmnet(x, y, alpha = 1)

# Realizar predicciones en los datos de prueba
x_prueba <- as.matrix(datos_prueba_r_l[, !colnames(datos_prueba_r_l) %in%
"Central_de_Riesgo"])
predicciones_ridge <- predict(modelo_ridge, s = "lambda.min", newx = x_prueba)
predicciones_lasso <- predict(modelo_lasso, s = "lambda.min", newx = x_prueba)

# Calcular el error cuadrático medio (MSE) en los datos de prueba
mse_ridge <- mean((predicciones_ridge - datos_prueba_r_l$Central_de_Riesgo)^2)
mse_lasso <- mean((predicciones_lasso - datos_prueba_r_l$Central_de_Riesgo)^2)

# Imprimir los resultados
print("Regresión Ridge:")
print(paste("MSE en datos de prueba:", mse_ridge))
print("\nRegresión Lasso:")
print(paste("MSE en datos de prueba:", mse_lasso))

# Ajustar un modelo de regresión logística
set.seed(123)
modelo_logit <- glm(Central_de_Riesgo ~ ., data = datos_entrenamiento, family =
"binomial",
                    control = glm.control(maxit = 1000))

# Imprimir el modelo de regresión logística
print("\nModelo de Regresión Logística:")
```

```
print(summary(modelo_logit))

# Obtener los datos del Z-Score
Z_Score <- datos$Z_Score
Z_Score

# Calcular las probabilidades de incumplimiento basadas en el modelo de regresión
logística
set.seed(123)
modelo_logit_completo <- glm(Central_de_Riesgo ~ .,
                             data = datos, family = binomial(link = "logit"),
                             control = glm.control(maxit = 1000))

probabilidades_incumplimiento <- predict(modelo_logit_completo, type = "response")

# Desactivar la notación científica para imprimir números
options(scipen = 999)

# Mostrar las primeras probabilidades de incumplimiento calculadas
print(probabilidades_incumplimiento)

# Definir la función asignar_calificacion() con los umbrales adecuados
asignar_calificacion <- function(probabilidad_incumplimiento) {
  if (probabilidad_incumplimiento >= 0.9999999999999997779554) {
    return("AAA")
  } else if (probabilidad_incumplimiento >= 0.99999999999957617236035) {
    return("AA")
  } else if (probabilidad_incumplimiento >= 0.9999999999942537076691) {
    return("A")
  } else if (probabilidad_incumplimiento >= 0.9999999999901710845407) {
    return("BBB")
  } else if (probabilidad_incumplimiento >= 0.0000000000051490913657) {
    return("BB")
  } else if (probabilidad_incumplimiento >= 0.0000000000043316507051) {
    return("B")
  } else if (probabilidad_incumplimiento >= 0.0000000000024820614702) {
```

```
    return("CCC")
  } else if (probabilidad_incumplimiento >= 0.00000000000010797496919) {
    return("CC")
  } else if (probabilidad_incumplimiento >= 0.00000000000001702553529) {
    return("C")
  } else {
    return("D")
  }
}

# Aplicar la función asignar_calificacion() a las probabilidades de incumplimiento
calificaciones <- sapply(probabilidades_incumplimiento, asignar_calificacion)

# Mostrar las calificaciones asignadas
print(calificaciones)

print(datos$Z_Score)

# Filtrar datos para eliminar NA
datos_filtrados <- datos[!is.na(datos$Z_Score) & !is.na(calificaciones), ]

# Graficar los datos de puntajes Z y las calificaciones de riesgo
plot(datos_filtrados$Z_Score, datos_filtrados$calificaciones,
      xlab = "Calificación de Riesgo", ylab = "Puntaje Z",
      main = "Relación entre Puntaje Z y Calificación de Riesgo")

if (!require(stats)) {
  install.packages("stats")
}

# Cargar la biblioteca necesaria para ANOVA
library(stats)

# ANOVA
resultado_anova <- aov(Z_Score ~ calificaciones, data = datos)
```



```
# Mostrar el resumen del ANOVA
summary(resultado_anova)

# Calcular los residuos del modelo
residuos <- resid(modelo_logit, type = "deviance")

# Graficar los residuos
plot(residuos, type = "p", main = "Gráfico de Residuos", xlab = "Índice de Observación",
ylab = "Residuos de Deviance")
abline(h = 0, col = "red")

# Asignar las calificaciones de Moody's a los datos
datos$Calificacion_Moodys <- calificaciones

# Mostrar las primeras filas de los datos actualizados
print(datos)

set.seed(123)
datos$Nombres_Empresas <- datos_gordon$Nombres_Empresas

# Definir las categorías de riesgo aprobadas
categorias_aprobadas <- c("AAA", "AA", "A", "BBB")

# Asignar el crédito aprobado o no aprobado
credito_aprobado <- ifelse(calificaciones %in% categorias_aprobadas, "Crédito
Aprobado", "Crédito No Aprobado")

# Instalar y cargar la librería 'truncnorm'
if (!require(truncnorm)) {
  install.packages("truncnorm")
}

library(truncnorm)

# Crear un data frame con los resultados y reemplazar los números de cliente con los
nombres de las empresas
```

```
resultados <- data.frame(Empresa = datos$Nombres_Empresas,  
                          Valor_Scoring = Z_Score,  
                          Calificación = calificaciones,  
                          Resultado = credito_aprobado)  
  
# Mostrar los resultados en forma de tabla  
print(resultados)  
  
# Obtener los nombres de las empresas unicornio  
nombres_empresas_unicornio <- empresas_unicornio$Nombres_Empresas  
  
# Marcar las empresas en resultados como unicornio si están en la lista de nombres de  
empresas unicornio  
resultados$Unicornio <- resultados$Empresa %in% nombres_empresas_unicornio  
  
# Mostrar los resultados  
print(resultados)  
  
set.seed(123)  
# Definir una función para asignar montos de crédito con ajuste  
asignar_monto_credito_ajustado <- function(z_score, calificaciones, unicornio) {  
  # Definir los límites de crédito máximo y mínimo  
  limite_credito_maximo <- 100000 # $100,000  
  
  # Factor de ajuste para reflejar mejor la relación entre el puntaje de crédito y el monto  
  de crédito  
  factor_ajuste <- 0.75  
  
  # Verificar si la calificación es aprobada (AAA, AA o A)  
  if (calificaciones %in% c("AAA", "AA", "A", "BBB")) {  
    # Calcular el monto de crédito inicial basado en el Z-Score  
    monto_credito <- z_score / 3 * limite_credito_maximo  
  
    # Ajustar el monto de crédito en función de la calificación  
    if (calificaciones %in% c("AAA", "AA")) {  
      monto_credito <- monto_credito * 1.5 # Aumentar el monto de crédito para  
      calificaciones más altas
```

```
}

# Aplicar el factor de ajuste
monto_credito <- monto_credito * factor_ajuste

# Ajustar el monto de crédito si es unicornio
if (unicornio) {
  monto_credito <- min(monto_credito, limite_credito_maximo) # Asegurar que no
exceda el límite máximo
}

# Generar un monto de crédito ajustado usando una distribución beta truncada
if (unicornio) {
  monto_credito_ajustado <- monto_credito
} else {
  monto_credito_ajustado <- rtruncnorm(1, a = 0, b = limite_credito_maximo, mean =
monto_credito, sd = monto_credito * 0.2)
}
} else {
  # Si la calificación no está aprobada, asignar un monto de crédito de $0
  monto_credito_ajustado <- 0
}

return(monto_credito_ajustado)
}

# Aplicar la función a los datos, pasando el parámetro 'unicornio' para las empresas
resultados$Monto_del_Credito <- mapply(asignar_monto_credito_ajustado,
resultados$Valor_Scoring, resultados$Calificación, resultados$Unicornio)

# Verificar si hay algún monto de crédito negativo y ajustarlo a $0 si es necesario
resultados$Monto_del_Credito[resultados$Monto_del_Credito < 0] <- 0

# Mostrar los resultados
print(resultados)

# Aplicar la función a los datos, pasando el parámetro 'unicornio' para las empresas
```

```
resultados$Monto_del_Credito <- mapply(asignar_monto_credito_ajustado,  
resultados$Valor_Scoring, resultados$Calificación, resultados$Unicornio)
```

```
# Verificar si hay algún monto de crédito negativo y ajustarlo a $0 si es necesario
```

```
resultados$Monto_del_Credito[resultados$Monto_del_Credito < 0] <- 0
```

```
# Establecer opciones para mostrar números sin notación científica
```

```
options(scipen = 999)
```

```
# Mostrar los resultados
```

```
print(resultados)
```

```
# Función para filtrar los resultados por nombre de empresa
```

```
filtrar_resultados_por_empresa <- function(nombre_empresa) {
```

```
  # Filtrar los resultados para la empresa específica
```

```
  resultados_empresa <- resultados[resultados$Empresa == nombre_empresa, ]
```

```
# Mostrar los resultados en forma de tabla
```

```
print(resultados_empresa)
```

```
# Obtener las variables más predictivas del modelo
```

```
variables_predictivas_importantes <- coef(modelo_logit_completo)
```

```
top_variables <- names(sort(abs(variables_predictivas_importantes), decreasing =  
TRUE))[1:5]
```

```
print(top_variables)
```

```
}
```

```
#PRUEBA CON UNO SIN CREDITO
```

```
nombre_empresa <- "YYKWZ" # Cambiar por el nombre de la empresa deseada
```

```
filtrar_resultados_por_empresa(nombre_empresa)
```

```
#PRUEBA CON UNO CON CREDITO
```

```
nombre_empresa <- "UWSBX" # Cambiar por el nombre de la empresa deseada
```

```
filtrar_resultados_por_empresa(nombre_empresa)
```

```
#PRUEBA CON UN UNICORNIO
```

```
nombre_empresa <- "MLYMT" # Cambiar por el nombre de la empresa deseada
```

filtrar\_resultados\_por\_empresa(nombre\_empresa)