



COMILLAS
UNIVERSIDAD PONTIFICIA

Jter

FACULTAD DE CIENCIAS ECONOMICAS Y EMPRESARIALES.

MODELOS DE SCORING. DESARROLLO DE UN MODELO LOGIT
APLICADO A EMPRESAS NO FINANCIERAS.

Autor: Rocío Santos González

Director: Álvaro Caballo Trébol

MADRID, MARZO DE 2023

RESUMEN

El presente trabajo de fin de grado profundiza en la revisión bibliográfica del estado de la cuestión sobre los tipos de modelos scoring utilizados en la evaluación del riesgo crediticio de las empresas, así como desarrolla un modelo logit con el programa SPSS utilizando una base de empresas de datos no financieros.

En la primera parte del trabajo, se realiza una revisión exhaustiva existente sobre los tipos de modelos de scoring, incluyendo todo tipo de modelos basados en el riesgo, en probabilidades y puntuaciones. Se analizan las ventajas e inconvenientes de cada uno de estos y se comparan distintas técnicas de scoring utilizadas en la actualidad. Además, se discute la relevancia y aplicabilidad de estas técnicas, en particular de la toma de decisiones al predecir el riesgo de impago de compañías.

En la segunda parte del trabajo, se lleva a cabo el desarrollo de un modelo logit utilizando el programa SPSS con una base de datos de empresas no financieras. El modelo se crea a partir de una base de datos que permite evaluar el riesgo crediticio de las empresas. Se detalla el proceso de selección de variables y la metodología utilizada para la construcción del modelo, así como los resultados obtenidos. Por último, se discutirán las implicaciones y limitaciones del modelo desarrollado, así como su aplicabilidad en la práctica.

En conjunto, este trabajo de fin de grado busca contribuir al conocimiento de los modelos de scoring y su aplicación en la evaluación del riesgo crediticio de las empresas. Este trabajo tiene relevancia práctica y académica, ya que se discuten las implicaciones y limitaciones de los modelos scoring en la práctica, así como la aplicabilidad de modelos desarrollado en la toma de decisiones de crédito.

Palabras clave: modelos scoring, modelo logit, riesgo crediticio, riesgo de impago, SPSS.

ABSTRACT

This final degree thesis examines in depth the literature review of the state of the art on the types of scoring models used in the assessment of corporate credit risk and develops a logit model with the SPSS software using a database of non-financial data companies.

In the first part of the paper, an exhaustive review of the existing types of scoring models is carried out, including all types of models based on risk, probabilities and scores. The advantages and disadvantages of each of these are analyzed and different scoring techniques currently used are compared. Furthermore, the relevance and applicability of these techniques, in particular for decision making in predicting the default risk of companies, is discussed.

In the second part of the paper, a logit model is developed using SPSS with a database of non-financial companies. The model is created from a database that allows the credit risk of companies to be assessed. The variable selection process and the methodology used to build the model are detailed, as well as the results obtained. Finally, the implications and limitations of the model developed, as well as its applicability in practice, will be discussed.

Overall, this thesis aims to contribute to the knowledge of scoring models and their application in the assessment of corporate credit risk. This work has practical and academic relevance, as it discusses the implications and limitations of scoring models in practice, as well as the applicability of models developed in credit decision making.

Key words: credit risk, credit scoring models, logit model, credit risk, default risk, SPSS.

INDICE

1. INTRODUCCION	5
1.1 Justificación del interés del tema	5
1.2 Objetivos.....	6
1.3 Metodología.....	6
2. PARTE I. REVISION BIBLIOGRAFICA DEL ESTADO DE LA CUESTION.....	8
2.1 Modelos de credit scoring: Beaver, Altman, Ohlson	8
2.2 Modelos paramétricos y no paramétricos.....	11
2.3 Modelos scoring de empresas no financieras	12
2.4 Clasificación de modelos de riesgo.....	15
2.4.1 Análisis discriminante	15
2.4.2 Modelos de regresión lineal.....	16
2.4.3 Modelos logit regresión logística	17
2.4.4 Modelos de programación matemática.....	19
2.4.5 Árboles de decisión	20
2.4.6 Redes neuronales.....	22
2.4.7 Modelos variables en el tiempo	23
3. PARTE II. CREACION DE UN MODELO DE RIESGO DE IMPAGO.....	25
3.1. Resumen	25
3.2 Metodología.....	25
3.3 SPSS (Statistical Package for the Social Sciences)	25
3.4 Discriminación de variables	27
3.5 Construcción del modelo.....	28
3.5.1 Descripción de la muestra.....	28
3.5.2 Variables dependientes y explicativas del modelo.....	28
3.5.3 Modificaciones en las variables explicativas del modelo.....	32
3.6 Obtención del modelo	33
3.6.1 Pasos previos.....	33
3.6.2 Modelo definitivo.....	33
4. CONCLUSIONES	37
6. BIBLIOGRAFIA	40

1.INTRODUCCION

1.1 Justificación del interés del tema

Los modelos de scoring son de gran interés en la actualidad puesto que las empresas buscan constantemente formas de aumentar su rentabilidad y minimizar sus riesgos financieros y estratégicos. Estos modelos son herramientas eficaces para evaluar el riesgo crediticio y la calidad de los clientes, lo que puede ayudar a las empresas a tomar decisiones informadas sobre la concesión de créditos o de realizar pagos a crédito a otras empresas.

Los modelos de scoring son especialmente útiles en el sector financiero. Sin embargo, también pueden aplicarse en diversos campos, como la evaluación de riesgo de seguros, la identificación de posibles fraudes y la selección de clientes potenciales.

Cada vez más, se estudia la necesidad de que las empresas cuenten con herramientas de predicción de riesgos para ser capaces de conocer con mayor precisión su desempeño en el futuro. Muchos autores han investigado la relación existente entre el riesgo de quiebra y la eficiencia productiva de las empresas. En “Bankruptcy risk and productive efficiency in manufacture firms” de L Bochetti y J Sierra, se estudia como el riesgo conocido de las empresas afecta a su productividad, medida por su eficiencia técnica. Se confirma que las empresas con mayores riesgos de impago, quiebra o generalmente financieros tienen una eficiencia productiva más baja. Por ello, los modelos de predicción de riesgos para las empresas son importantes para llevar a cabo decisiones tanto económicas como estratégicas. Además, estos modelos predicen la rentabilidad de las empresas a largo plazo, lo que facilita la toma de decisiones estratégicas, así como la evaluación de la viabilidad de la empresa y la gestión de esta.

El desarrollo de modelos de riesgo de impago efectivos y precisos es crucial para mejorar la toma de decisiones y reducir los riesgos asociados con el otorgamiento de créditos. Un modelo de scoring bien diseñado puede ayudar a las instituciones financieras a identificar a los clientes con mayor probabilidad de cumplir con sus obligaciones de pago, lo que puede ayudar a minimizar las pérdidas y mejorar la rentabilidad. Además, el desarrollo de modelos de scoring también es importante desde una perspectiva más amplia, ya que puede contribuir a mejorar la eficiencia y la transparencia en el proceso de toma de decisiones. Al utilizar modelos de scoring basados en datos, las instituciones financieras pueden evitar los sesgos y las decisiones subjetivas que pueden estar presentes en los procesos de evaluación tradicionales.

En este contexto, el trabajo de fin de grado proporciona una revisión crítica de los diferentes tipos de modelos utilizados por las empresas y cómo se aplican en diferentes contextos. Esto permitirá comprender mejor las ventajas y limitaciones de cada modelo y cómo pueden ser utilizados de manera efectiva en diferentes situaciones. Se analizará la utilidad de los modelos de scoring para empresas no financieras y cómo estas pueden beneficiarse de la implementación de estos modelos.

En resumen, este trabajo sobre modelos de scoring para empresas no financieras resulta relevante e interesante porque los modelos de scoring son herramientas útiles para evaluar el riesgo crediticio y la calidad de los clientes, lo que puede ayudar a las empresas a tomar decisiones informadas sobre la concesión de créditos y a mejorar su rentabilidad. Además, el trabajo contribuye al avance del conocimiento en este campo al proporcionar una revisión crítica y una evaluación de las prácticas actuales y futuras de los modelos de riesgo de impago.

1.2 Objetivos

El trabajo tiene como principal objetivo el estudio de los modelos de scoring aplicados a empresas no financieras. Estos modelos sirven de gran utilidad para muchas compañías ya que predicen la probabilidad de impago y nos aportan información acerca del riesgo de morosidad cuando se realizan pagos a crédito.

La primera parte del trabajo se centrará en el estudio de los modelos scoring existentes. De esta forma, se explicará tanto la utilidad de estos modelos como su origen y la razón de su importancia para muchas empresas. Además, se analizarán los principales modelos de scoring utilizados en la actualidad, examinando como han sido creados, las variables utilizadas y su calidad a la hora de predecir sucesos futuros.

Una vez realizado un análisis bibliográfico del estudio de la cuestión, se presentará el modelo llamado *Riskforecast*, un modelo de riesgo de impago cuyo objetivo es reducir la morosidad de las empresas no financieras analizando balances, cuentas de pérdidas y ganancias y estados de flujos de caja. En esta sección, se desarrollará paso a paso el modelo, añadiendo variables significativas con el objetivo de conseguir un modelo de riesgos de impago con la máxima precisión posible.

1.3 Metodología

La metodología utilizada en el TFG combina un análisis bibliográfico del estado de la cuestión con el desarrollo de un modelo logit de predicción. Por ello, se basa tanto en datos secundarios como primarios. En primer lugar, se realizará una revisión exhaustiva de la literatura existente en el campo de estudio, con el objetivo de recopilar información relevante sobre los temas clave y las variables implicadas. Se reunirá información recopilada por otros autores que han tratado el tema a estudiar. Para ello, se utilizarán bases de datos como Google Scholar, Jstor o la biblioteca de la Universidad Pontificia de Comillas. A continuación, se llevará a cabo el desarrollo de un modelo de scoring para el cual se utilizarán métodos de aprendizaje automático y estadístico, así como una amplia gama de datos históricos y predictores relevantes. En esta segunda parte, se

detallarán las diferentes etapas del proceso de desarrollo del modelo, como la selección y preparación de los datos, el ajuste del modelo y la evaluación del rendimiento del modelo con el conjunto de datos seleccionado.

Finalmente, se realizará una evaluación crítica de los resultados obtenidos, discutiendo las implicaciones teóricas y prácticas de los hallazgos, así como las posibles limitaciones del estudio y las recomendaciones para futuras investigaciones.

2. PARTE I. REVISION BIBLIOGRAFICA DEL ESTADO DE LA CUESTION

2.1 Modelos de credit scoring: Beaver, Altman, Ohlson

Los modelos de credit scoring son modelos de clasificación, en los que a partir de una muestra a priori, se crea un modelo capaz de predecir un resultado a posteriori. “Se denomina credit scoring a todo sistema de evaluación crediticia que permite valorar de forma automática el riesgo asociado a cada solicitud de crédito” (Bonilla et al., 2003). El objetivo de los modelos de scoring es reducir el riesgo de impago. Según Gutiérrez Girault “son algoritmos que de manera automática evalúan el riesgo de crédito de una entidad” (2007). Cuanto mayor es el score del modelo, menor es el riesgo, entendiendo como score la puntuación numérica que se utiliza para predecir la probabilidad de que un cliente o deudor no cumpla con sus obligaciones de pago en el futuro.

Estos modelos son de gran utilidad para las entidades puesto que conocer el riesgo de impago de los clientes puede ser decisivo a la hora de prestar un crédito o servicio, en el caso de empresas no financieras. De esta manera, la entidad asegura que el cliente va a pagar sus deudas. Por el contrario, podría afectar no solo a sus costes financieros, sino también a costes de administración y gestión.

Los modelos de credit scoring cada vez son más utilizados tanto en empresas financieras como no financieras. En un principio, los únicos criterios utilizados para calcular el riesgo de impago de un cliente eran el patrimonio y la capacidad de pago. Sin embargo, aparecieron tres inconvenientes: la subjetividad, secuencialidad del análisis y el hecho de ser un proceso lento y costoso (Bonilla et al., 2003). Como consecuencia, en los años 70 comenzaron a surgir modelos de riesgo de impago gracias a los avances tecnológicos y estadísticos. Durante los años 90 estos avances comenzaron a ser generalmente aceptados. Con los avances de la tecnología y el aumento de la disponibilidad de datos, los modelos de scoring se han vuelto cada vez más sofisticados y precisos. Hoy en día, estos modelos de scoring utilizan una variedad de técnicas estadísticas y de aprendizaje automático para analizar grandes cantidades de datos y proporcionar predicciones precisas y detalladas sobre el comportamiento del cliente.

Desde que surgieron estas técnicas de predicción de riesgo de impago, fueron muchos especialistas que publicaron sus estudios y realizaron investigaciones sobre el tema. Fair Isaac Corporation (FICO) introduce el primer modelo de scoring en la década de 1950. Este modelo utilizaba información personal y financiera del cliente para calcular una puntuación de crédito con el objetivo de evaluar el riesgo crediticio del solicitante. En el artículo “Scorecards as devices for Consumer Credit: the Case of Fair Isaac & Company incorporated”, Marta Poon explica como

FICO desarrolló y utilizó scorecards para evaluar el riesgo crediticio de los solicitantes de crédito y cómo estas puntuaciones son utilizadas por los prestamistas para tomar decisiones de aprobación de crédito y para establecer los términos y condiciones del préstamo. La autora afirma que es necesario evaluar constantemente la efectividad de los scorecards para asegurar una evaluación justa y precisa del riesgo crediticio (Poon, 2007). Además, se concluye que el uso de estos modelos resulta beneficioso no solo para los prestamistas sino también para los solicitantes de créditos, mejorando la eficacia y la seguridad del mercado crediticio en general.

Beaver (1966) en *Financial Ratios as Predictors of Failure* argumentó que los ratios financieros pueden utilizarse como indicadores para predecir la probabilidad de quiebra de una empresa. Beaver (1966, 1968) utilizó el análisis univariante para ratios seleccionados y descubrió que algunos tenían un poder predictivo muy bueno. El autor afirma que la inclusión tanto de las variables de mercado como de los ratios financieros en el modelo de predicción puede mejorar significativamente su precisión y capacidad predictiva. Ambos son complementarios en la predicción, ya que proporcionan valiosa información sobre diferentes aspectos de un negocio. El autor se centró en las variables financieras que más influyen en la clasificación. Entre ellas destacó los flujos de caja de las empresas analizadas haciendo hincapié en las reservas de efectivo, el flujo de efectivo operativo, apalancamiento y gastos operativos. En particular, propuso el uso del ratio de flujo de efectivo sobre activos (Cash Flow to Assets) como una medida de la capacidad de la empresa en cuestión para generar efectivo a partir de sus activos. Con el análisis de estas variables era posible evaluar la salud financiera de una empresa y detectar signos tempranos de problemas financieros de la misma. (Charpentier et al., 2013).

Al hablar de modelos de riesgo de impago, es esencial mencionar a Edward I. Altman por su destacada contribución del modelo Z-Score a finales de la década de 1960. Desde el punto de vista de Altman, los análisis univariantes pueden tener variables no significativas que en el análisis multivariante sí pueden ser significativas, puesto que se combinan con otras variables (Ayús, 2016). Por ello, el autor apuesta por un análisis discriminante múltiple en su modelo. El modelo se ha convertido en una de las herramientas más utilizadas en el análisis del riesgo de crédito. Aunque luego ha tenido variaciones, el modelo de Z score empezó siendo un método estadístico de análisis discriminante lineal un modelo que tenía el objetivo de predecir la probabilidad de bancarrota de una empresa (Altman et al., 2017).

La puntuación Z de Altman es una herramienta importante en el ámbito del análisis del riesgo de crédito. Su sencillez, amplia aplicabilidad y fiabilidad demostrada la convierten en una herramienta valiosa para los inversores, las agencias de calificación crediticia y otras partes interesadas que necesitan evaluar la salud financiera de las empresas. Sin embargo, es importante recordar que el Altman Z-Score es sólo una herramienta entre muchas otras, y que debe utilizarse

junto con otra información y análisis para tomar decisiones informadas sobre el riesgo de crédito. (Calandro, 2007).

El modelo de Z-Score se basa en cinco ratios financieros, que se combinan para calcular una puntuación que indica el nivel de dificultades financieras de una empresa. Los cinco ratios financieros utilizados en el modelo Altman Z-Score son: (1) capital circulante/activo total, (2) beneficios no distribuidos/activo total, (3) beneficios antes de intereses e impuestos/activo total, (4) valor de mercado de los fondos propios/valor contable del pasivo total y (5) ventas/activo total. Estos ratios se eligen para ofrecer una imagen completa de la salud financiera de una empresa, y se utilizan para calcular una puntuación que va de -infinito a 10, donde se considera que una puntuación superior a 2,99 indica que es poco probable que una empresa quiebre, mientras que una puntuación inferior a 1,81 indica una alta probabilidad de quiebra. (Altman, 1968). Este se basa en una combinación lineal de ratios que se presentan en la siguiente formula:

$Z=1.2X1 + 1.4X2 + 3.3X3 + 0.6X4 + 1.0X5$. donde:

X1=ratio de capital circulante sobre activos totales. Mide los activos líquidos en relación con el tamaño de la compañía.

X2=ratio de ganancias retenidas sobre activos totales. Mide rentabilidad y refleja la edad de la compañía y la capacidad de obtener ganancias.

X3=ratio de ganancias antes de impuestos sobre activos totales. Mide la eficiencia operativa excluyendo los impuestos y los factores de apalancamiento. Reconoce la importancia de las ganancias operativas para la viabilidad a largo plazo.

X4=ratio de capitalización bursátil sobre el valor contable del pasivo total. Añade una dimensión de mercado que puede mostrar una fluctuación del precio de la acción como una señal de alarma.

X5=ratio de ventas sobre activos totales. Medida estándar de la rotación de activos total (varía bastante con el tipo de industria). (Altman et al., 2017).

La puntuación Z de Altman tiene varias ventajas sobre otras herramientas de análisis del riesgo de crédito. En primer lugar, es fácil de calcular y comprender, lo que lo hace accesible a una amplia gama de usuarios. En segundo lugar, utiliza datos financieros fáciles de obtener, por lo que es ampliamente aplicable a diferentes tipos de empresas. Este modelo ha sido ampliamente probado y validado, y se ha demostrado que es un indicador fiable de la probabilidad de quiebra de una empresa.

A pesar de sus muchas ventajas, la puntuación Z de Altman no está exenta de limitaciones. Una de sus principales limitaciones es que se basa en datos financieros históricos, por lo que no tiene en cuenta acontecimientos futuros que puedan afectar a la salud financiera de una empresa.

Además, el modelo se basa en un conjunto fijo de ratios financieros y no tiene en cuenta ni los cambios en el panorama financiero ni los diferentes modelos de negocio.

Gracias a la investigación de Ohlson (1980), el modelo de regresión logística o también conocido como modelo logit pasó a ser mundialmente aceptado para predecir el riesgo de quiebra o de impago de un crédito. Ohlson estudia como dos factores afectan directamente al riesgo de impago de la compañía. Estos son: el valor de los activos de la empresa y el nivel de sus pasivos. El modelo supone que los activos y pasivos de una empresa siguen un camino aleatorio, y que el valor de los activos de la empresa disminuirá gradualmente con el tiempo hasta llegar a cero, momento en el que la empresa quebrará. El modelo también supone que el ritmo al que disminuyen los activos de la empresa es proporcional al nivel de sus pasivos.

El modelo de Ohlson se utiliza ampliamente en el sector de la gestión del riesgo de crédito y se ha investigado y perfeccionado mucho a lo largo de los años. Es una herramienta valiosa para los analistas de crédito y los inversores, que pueden utilizarlo para evaluar la solvencia de una empresa y tomar decisiones de inversión con conocimiento de causa. (Molano, 2018).

2.2 Modelos paramétricos y no paramétricos

Existen dos tipos de modelos estadísticos; los modelos paramétricos y los modelos no paramétricos. Por un lado, los modelos paramétricos asumen que los datos siguen una distribución específica, y que los parámetros de esta distribución pueden ser estimados a partir de los datos. Por ejemplo, un modelo de regresión lineal asume que la relación entre las variables dependientes e independientes es lineal y que los errores de predicción siguen una distribución normal. Los modelos paramétricos son útiles cuando se dispone de información previa o conocimiento sobre la distribución subyacente de los datos. Entre las principales ventajas que poseen los modelos paramétricos destaca su flexibilidad para adaptarse a distintas distribuciones de datos, así como su precisión predictiva. (Bonilla et al., (2003). Además, los modelos paramétricos son muy eficientes computacionalmente puesto que requieren número limitado de parámetros a estimar y permiten el uso de pruebas de hipótesis e intervalos de confianza para estimar la incertidumbre en las predicciones. También se caracterizan por ser fáciles de interpretar y requieren menos datos para ser ajustados que los modelos no paramétricos. Sin embargo, puede resultar una tarea compleja la elección de un modelo paramétrico concreto adecuado a las necesidades ya que hay una gran variedad. Por ello, analizaremos en mayor profundidad distintos tipos de modelos y su aplicación al riesgo de impago de la empresa (De la Ossa Ciodaro, 2016).

Por otro lado, los modelos no paramétricos no hacen suposiciones sobre la distribución subyacente de los datos. A diferencia de los anteriores, estos modelos se ajustan a los datos sin asumir la forma funcional de la relación entre las variables. Por ejemplo, un modelo de árbol de decisión divide iterativamente los datos en subconjuntos más pequeños utilizando criterios como

la ganancia de la información o la reducción de la entropía para identificar las características más importantes en la predicción de la variable objetivo. Cabe mencionar que los modelos no paramétricos tienen una mayor productividad, siendo más precisos en la predicción de valores extremos y situaciones desconocidas ya que no están limitados por supuestos sobre la distribución de los datos. Además, hacen menos suposiciones sobre los datos lo que hace a este tipo de modelos tengan menos probabilidades de fallar si no se cumplen las suposiciones (Medina, 2000).

La elección entre modelos paramétricos y no paramétricos depende de la naturaleza de los datos y de la pregunta de investigación. Los modelos paramétricos son más apropiados cuando tenemos conocimiento previo sobre la distribución subyacente de los datos. Sin embargo, los modelos no paramétricos son más flexibles y pueden ser útiles cuando no se dispone de información previa o cuando la relación entre las variables no es lineal.

2.3 Modelos scoring de empresas no financieras

Como ya hemos visto, los modelos de scoring son utilizados por las empresas como herramientas de análisis de riesgos para evaluar el comportamiento de solvencia de clientes o usuarios en términos de su capacidad para cumplir sus compromisos de pago. Aunque estos modelos son muy utilizados en la industria financiera, también se pueden aplicar en empresas no financieras. Al aplicar estos modelos, las empresas pueden identificar oportunidades de crecimiento y mejorar la experiencia del cliente. Además, pueden gestionar los riesgos empresariales y tomar decisiones informadas en términos de calidad de producto o servicio. Existen diversos modelos utilizados en este ámbito, como, por ejemplo, los modelos de puntuación de clientes, modelos de riesgo empresarial o modelos de calidad del producto.

Los modelos de puntuación de clientes se utilizan para evaluar el riesgo de impago o la probabilidad de cancelación de un cliente. Para ello, se analizan diversos factores como el comportamiento de pagos, el tipo de producto o servicio adquirido, el perfil demográfico del cliente, entre otros. La ventaja para la empresa de este modelo es que permite segmentar a sus clientes y ofrecer productos o servicios personalizados para satisfacer sus necesidades. De esta manera se aumenta la fidelidad de los clientes y la retención de estos en la empresa. Este modelo se basa en datos históricos de los clientes y utiliza técnicas estadísticas y de aprendizaje automático para predecir el comportamiento futuro de los clientes en términos de sus patrones de consumo, preferencias y lealtad a la marca. Con esta información, la empresa puede desarrollar estrategias de marketing personalizadas y mejorar la experiencia del cliente (Willey, 1992).

La aplicación más común del modelo de puntuación de clientes es en la industria de servicios, como en compañías de telecomunicaciones, compañías de seguros o servicios públicos. Estas empresas utilizan este modelo para identificar a los clientes que tienen más probabilidad de cancelar sus servicios o incumplir con sus pagos, y tomar medidas proactivas para reducir el

riesgo y mejorar la rentabilidad. Otra aplicación del modelo de puntuación de clientes se puede encontrar en el comercio minorista. En este área, el modelo se utiliza para analizar el comportamiento de compra de los clientes y clasificarlos en grupos según su historial de compras. Esta información permite personalizar ofertas y promociones, mejorar la experiencia del cliente y fidelizarlo. El modelo de puntuación de clientes también se utiliza en la industria bancaria, donde se utiliza para evaluar el riesgo crediticio y determinar su capacidad de pago. Los bancos pueden identificar a los clientes que son más propensos a incumplir con sus pagos y tomar medidas preventivas para reducir el riesgo. En general, el modelo de puntuación de clientes es una herramienta muy útil para cualquier negocio que tenga una base de clientes grande y diversa. Con este modelo, la empresa puede segmentar a sus clientes, identificar a aquellos que tienen más probabilidades de no cumplir con sus compromisos de pago y ajustar estrategias personalizadas para retenerlos. A medida que avanza la tecnología, el aumento de la competencia en todas las industrias aumenta las tasas de abandono o rotación, lo que aumenta la importancia que las empresas otorgan a la retención de clientes. La relación entre la satisfacción y la lealtad del cliente y el desempeño del negocio es más fuerte en industrias con mayor competencia y menores costes de cambio (Morgan, 2006). Por lo tanto, es importante controlar métricas como el índice de satisfacción del cliente o el índice de lealtad del cliente.

El resultado final del uso de estos modelos es una mejora en la rentabilidad y una mayor satisfacción del cliente. Este modelo se utiliza para evaluar el riesgo de la empresa y se basa en el análisis de diversos factores como la estabilidad financiera, la solvencia, la rentabilidad, la posición competitiva en el mercado. El principal objetivo es identificar los riesgos a los que se enfrenta la compañía y tomar decisiones informadas para reducirlos, mejorar la gestión de la empresa y maximizar su rentabilidad.

Para las empresas que desean evaluar y comprender los riesgos relacionados con su sector, un modelo de evaluación de riesgos empresariales es una herramienta crucial. La posibilidad de que una empresa se enfrente a un acontecimiento de riesgo, como una interrupción de la cadena de suministro, una caída de las ventas o incluso un impago financiero, se calcula utilizando diversas variables y factores (Jarrow, 2009). Existen varios modelos para evaluar el riesgo empresarial, cada uno con su propio conjunto de parámetros y procedimientos. El análisis DAFO, que evalúa los puntos fuertes y débiles, las oportunidades y las amenazas de la empresa, es uno de los métodos más populares. El análisis PESTEL, que evalúa los factores políticos, económicos, sociales, tecnológicos, medioambientales y jurídicos que pueden afectar a una empresa, es otro modelo muy conocido.

Además de estos modelos generales, existen modelos para evaluar los riesgos empresariales que son específicos de determinados sectores o industrias. Por ejemplo, un modelo de evaluación de

riesgos para la industria sanitaria o un determinado sector bancario. Estos modelos de evaluación de riesgos empresariales, en general, son un recurso útil para cualquier empresa que pretenda comprender los riesgos relacionados con sus operaciones y tomar decisiones bien informadas para mitigar esos riesgos. En la actualidad, existe incertidumbre tanto a nivel macroeconómico como comercial. En consecuencia, disminuir o gestionar estos riesgos y mitigar los efectos negativos puede lograrse de forma significativa empleando técnicas y la optimización para mitigar estos riesgos o crear escenarios potenciales (Sweeting, 2007). Estos modelos ayudan a las empresas a aumentar su capacidad para anticiparse y responder a los riesgos, lo que favorece la estabilidad financiera y la reputación corporativa a largo plazo.

Es importante también destacar los modelos de puntuación de calidad de producto o servicio puesto que estos ayudan a las empresas a evaluar sus productos o servicios ofrecidos por las empresas. Este modelo puede ayudar a identificar las áreas en las que la empresa necesita mejorar su oferta y la satisfacción del cliente, y también a detectar oportunidades para desarrollar nuevos productos o servicios (Chacón, 2018).

Se trata de herramientas de análisis que se utiliza en las empresas para evaluar la calidad de los productos o servicios que ofrecen. Este modelo se basa en la opinión de los clientes y su grado de satisfacción con el producto o servicio que han adquirido. Para implementar el modelo de puntuación de calidad. Del producto o servicio, se realizan encuestas o cuestionarios a los clientes para evaluar su nivel de satisfacción en diferentes aspectos del producto o servicio. Estos aspectos pueden incluir la calidad del producto, el servicio post-venta, la rapidez en la entrega, la facilidad de uso, la relación calidad-precio, entre otros. La utilidad de este modelo para la empresa es que le permite detectar áreas de mejora y oportunidades de innovar en su oferta, lo que puede conducir a un aumento de la satisfacción del cliente y, por ende, a un aumento de las ventas.

La experiencia del cliente de principio a fin es un factor clave para la retención y satisfacción de este. Por ello, debe existir proactividad por parte de las empresas para la mejora constante de estos factores. (Lemon, 2016). Estos modelos recopilan y analizan datos sobre la experiencia del cliente, como, por ejemplo, la escucha en la red, análisis de sentimientos o la medición del Net Promoter Score (NPS).

Una vez que se han recopilado las respuestas de los clientes, se asigna una puntuación a cada uno de los aspectos evaluados y se calcula la puntuación total del producto o servicio. Esta puntuación se puede utilizar para comparar la calidad de los productos o servicios ofrecidos por la empresa con los de su competencia y para identificar áreas de mejora.

La aplicación más común del modelo de puntuación de calidad del producto o servicio es en la industria minorista. En este sector, las empresas utilizan este modelo para evaluar la calidad de sus productos o servicios y para identificar áreas de mejora. La puntuación obtenida a partir de

este modelo puede utilizarse para desarrollar estrategias que permitan mejorar la calidad del producto o servicio y aumentar la satisfacción del cliente. Otra aplicación del modelo de puntuación de calidad del producto o servicio es en la industria de servicios, como en compañías de telecomunicaciones, servicios de salud, entre otros. En este sector, el modelo se utiliza para evaluar la calidad del servicio ofrecido y para identificar áreas de mejora en la experiencia del cliente.

En resumen, el modelo de puntuación de calidad del producto o servicio es una herramienta muy útil para cualquier empresa que ofrezca productos o servicios. Al utilizar este modelo, la empresa puede evaluar la calidad de sus productos o servicios, identificar áreas de mejora y desarrollar estrategias que permitan mejorar la calidad y la satisfacción del cliente. El resultado final es una mejora en la rentabilidad y una mayor fidelidad del cliente.

Tras haber analizado los distintos modelos de scoring, podemos concluir que estos pueden ser muy útiles para empresas no financieras, ya que permiten analizar datos y obtener información valiosa que puede ser utilizada para tomar decisiones informadas y mejorar su gestión. Los modelos scoring pueden ayudar a las segmentar a los clientes, identificar oportunidades de mejora, detectar riesgos y mejorar la rentabilidad de la empresa.

2.4 Clasificación de modelos de riesgo

Existen distintos métodos para calcular el riesgo crediticio de una empresa. Debemos de tener en cuenta que no existe un método mejor o peor, sino que cada caso, dependerá de distintos factores como, por ejemplo, las características de los datos y el objetivo concreto que se quiera alcanzar (Rayo Cantón et al., 2010). También, es importante mencionar que no todos los modelos sirven para el mismo tipo de empresa, puesto que las variables, datos a analizar y aspectos a tener en cuenta pueden variar entre unas y otras. Por ello, cada empresa debe escoger el modelo que más se adapte a sus necesidades y características específicas (Hand et al., 1997). A continuación, analizaremos los modelos más utilizados por las empresas para calcular el riesgo crediticio.

2.4.1 Análisis discriminante

El análisis discriminante es un método estadístico que tiene como objetivo estudiar si las variables independientes utilizadas en el modelo son representativas, es decir, si clasifican correctamente a los objetos estudiados. Utilizando este análisis, se busca encontrar las variables que mejor explican el modelo.

Este análisis se caracteriza por ser multivariante, teniendo una variable dependiente categórica y variables independientes continuas. Las variables independientes determinan a que grupo pertenecen cada una de las observaciones. El objetivo es clasificar cada observación en el grupo

correcto, según el valor de la variable clasificadora, disminuyendo el error de clasificación (de Lema, 1995).

El análisis discriminante puede aplicarse a la mayoría de los modelos como demostró Durand (1941), puesto que es útil tanto para separar linealmente grupos en función de la variable dependiente tanto como para manejar múltiples grupos en el caso de clasificación multiclase. Además, otra de las ventajas que proporciona un análisis discriminante es que los resultados son fáciles de interpretar y pueden utilizarse para hacer predicciones futuras.

Sin embargo, autores como Breiman, L (2001) dudan de este método y critican distintos aspectos como la distribución de las variables, los errores de clasificación y la significatividad individual de las variables. El análisis discriminante requiere suposiciones sobre la distribución de las variables, que no pueden cumplirse siempre. Por otro lado, este análisis es sensible a los valores atípicos que pueden afectar en gran medida a los resultados. Además, la interpretación de este modelo se vuelve más compleja cuando existen múltiples discriminantes.

2.4.2 Modelos de regresión lineal

Los modelos de regresión lineal son una herramienta utilizada para predecir relaciones entre variables. Los modelos de regresión lineal han ido evolucionando, desde la regresión lineal simple, que implican dos variables, una dependiente y otra independiente. Después, surge la regresión lineal múltiple que permite tener en cuenta múltiples variables independientes en la predicción de una variable dependiente (del Pino, 2018).

A finales del siglo XX, Orgler (1970) utilizó el análisis de regresión en un modelo para préstamos comerciales en la que se evaluaron los distintos préstamos. Se muestra como estos modelos se podrían aplicar tanto para fines científicos como económicos. Además, nos muestra como los modelos de regresión lineal tienen resultado sencillos y fáciles de interpretar, mostrando claramente las relaciones entre las variables, requiriendo también poca preparación de datos en comparación con otros modelos de aprendizaje automático. Sin embargo, estos modelos se limitan a relaciones lineales entre variables además de suponer una distribución normal de los errores que no siempre se cumple en la práctica.

Cabe destacar que este tipo de modelos se utilizan habitualmente en la calificación crediticia y las finanzas para predecir la probabilidad de impago y determinar la solvencia. En el contexto de calificación crediticia, los modelos de regresión lineal pueden utilizarse para analizar la relación entre diversos factores financieros y demográficos, como los ingresos, el ratio-deuda ingresos y la antigüedad del historial crediticio y la probabilidad de impago de un préstamo.

2.4.3 Modelos logit regresión logística

El modelo logit también se conoce como modelo de regresión logística y se utiliza ampliamente en campos como la economía, las finanzas, las ciencias sociales, la investigación médica, el marketing o las ciencias políticas.

El origen de los modelos logit se remonta al campo de la bioestadística en la década de 1950. El modelo logit fue propuesto por primera vez por el estadístico Joseph Berkson en "Application of the logistic function to bio-assay. Journal of the American statistical association". Más tarde, David Cox realizó contribuciones importantes a la teoría y aplicación del modelo. El autor reconoció que el modelo de regresión lineal tradicional no podía utilizarse para analizar datos binarios. Se impuso rápidamente en el campo de la bioestadística y pronto empezó a utilizarse también en otros campos, como la economía, la psicología y sociología.

Muchos autores han ido desarrollando y aportando novedades al modelo logit. JS Cramer (2003) discute en detalle la modelización de distintos ámbitos como las elecciones de consumo, la ocupación laboral, los transportes y las viviendas, entre otros temas. McFadden, por ejemplo, introdujo el modelo logit condicional, que ha sido utilizado en una amplia variedad de campos. Se basa utiliza para estimar los parámetros de una función de utilidad, lo que permite la predicción de probabilidades (McFadden, 1973). Para cada ámbito y necesidad han ido surgiendo distintos modelos como, por ejemplo, el logit Jerárquico, el logit multinomial, el logit mixto, o el logit de valor extremo (Rodríguez, 2008).

El modelo logit se utiliza para calcular la probabilidad de un resultado binario basado en una o mas variables predictoras. Suele ajustarse a los datos mediante la estimación de máxima verosimilitud. Una vez ajustado el modelo, puede utilizarse para predecir la probabilidad del resultado binario para nuevos datos.

La utilidad de estos reside en su capacidad para modelizar la relación entre una variable de resultado binaria y una o más variables independientes. Son especialmente útiles cuando la variable de resultado es binaria y las variables independientes son continuas o categóricas. Pueden utilizarse para estimar la probabilidad de un resultado binario, como la probabilidad de éxito o fracaso. La variable de resultado suele denominarse variable dependiente y puede adoptar uno de dos valores, como "éxito" o "fracaso". Las variables independientes, también conocidas como variables predictoras, pueden ser cualquier combinación de variables continuas o categóricas. También pueden utilizarse para identificar los factores más importantes a la hora de determinar el resultado (Rodríguez, 2008).

Es una función monótona creciente que esta acotada en el intervalo $[0,1]$. Es un modelo de elección binaria. Estos modelos se caracterizan por tener solamente dos alternativas en las que

implica tener solamente dos actuaciones, en este caso, es si la empresa ha pagado (0 o 1) y si la empresa ha impagado (0 o 1).

La función logística tiene la forma de $1/(1+e^{(-x)})$ donde x es la entrada de la función. Este modelo utiliza esta función para modelizar la probabilidad de un resultado binario en función de una o más variables independientes. La fórmula de un modelo logit es:

$$P(Y = 1|X) = e^{(\beta_0 + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_n X_n)} / (1 + e^{(\beta_0 + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_n X_n)})$$

Donde $P(Y = 1|X)$ es la probabilidad de que la variable de resultado (Y) sea igual a 1, dadas las variables independientes (X), e es la base del logaritmo natural, y $\beta_0, \beta_1, \beta_2, \dots, \beta_n$ son los parámetros del modelo.

Una de las principales ventajas de los modelos logit es que pueden utilizarse para modelizar la relación entre una variable de resultado binaria y una o más variables independientes, incluso cuando la relación no es lineal. Esto es importante porque muchos fenómenos del mundo real no son lineales y los modelos de regresión lineal tradicionales no son apropiados para estas situaciones. También pueden utilizarse para modelizar interacciones entre variables independientes, lo que puede ser útil para comprender relaciones complejas.

Además, los modelos logit son relativamente fáciles de estimar e interpretar. Pueden estimarse utilizando software estadístico estándar y los resultados pueden interpretarse fácilmente en términos de *odds ratio* que mide la relación entre la variable de resultado y las variables independientes. Puede utilizarse para comparar las probabilidades de que el resultado se produzca en diferentes grupos. Por último, los modelos de regresión logística son flexibles y pueden utilizarse para analizar una amplia gama de tipos de datos y preguntas de investigación. De ahí su aplicación en distintos campos de investigación. (Eguídanos, 1994).

En el contexto de la calificación crediticia, las variables predictoras pueden incluir el historial crediticio del prestatario, su historial laboral, sus ingresos y otra información financiera. El modelo logit utiliza estas variables para estimar la probabilidad de que el prestatario incumpla el préstamo. Estos modelos tienen la capacidad para manejar datos complejos y captar relaciones no lineales entre las variables predictoras y el resultado. Por ejemplo, el modelo logit puede manejar situaciones en las que el impacto de una variable predictiva concreta sobre la probabilidad de impago no es constante en todos los prestatarios, sino que varía en función de otros factores. Otra ventaja de los modelos logit en la aplicación al credit scoring es su facilidad de interpretación. Los coeficientes de las variables predictoras en un modelo logit pueden interpretarse como el cambio en las probabilidades logarítmicas de impago para un cambio de una unidad en la variable predictora, manteniendo constantes todas las demás variables. Esto permite a los analistas

identificar qué variables predictoras tienen mayor impacto en la probabilidad de impago y tomar decisiones informadas sobre qué variables incluir en el modelo.

Sin embargo, utilizar un modelo logit para predecir un riesgo de impago puede tener algunos inconvenientes. Como hemos mencionado anteriormente uno de los principales inconvenientes es la suposición de que las variables predictoras son independientes, lo que no suele ser el caso en los datos del mundo real. Otra limitación es que los modelos logit pueden ser sensibles a los valores atípicos y pueden verse influidos por los valores extremos de los datos (Laffarga, 1987).

En conclusión, los modelos logit se utilizan ampliamente en la calificación crediticia debido a su capacidad para manejar datos complejos, captar relaciones no lineales y proporcionar resultados fáciles de interpretar. A pesar de sus limitaciones, los modelos logit proporcionan información valiosa sobre la probabilidad de impago de los prestatarios potenciales y son una herramienta importante para las instituciones financieras en sus procesos de evaluación de riesgos. Combinando los modelos logit con otras técnicas estadísticas, como los árboles de decisión y las redes neuronales, las entidades financieras pueden comprender mejor los factores que influyen en el riesgo de crédito y tomar decisiones más informadas sobre la concesión de préstamos.

2.4.4 Modelos de programación matemática

La programación matemática es una rama de las matemáticas que se ocupa de la optimización de sistemas complejos. Consiste en formular un problema como un modelo matemático y utilizar algoritmos y técnicas matemáticas para encontrar la solución óptima. En los últimos años, la programación matemática ha encontrado aplicaciones en una amplia gama de campos, como las finanzas, la ingeniería y la investigación operativa. Existen varios tipos de métodos de programación matemática, como la programación lineal, la programación entera mixta y la programación no lineal. Estos métodos pueden utilizarse para resolver problemas de optimización, toma de decisiones y evaluación de riesgos (Lieberman, 1991).

La programación lineal es uno de los métodos de programación matemática más utilizados. Consiste en formular un problema como una función lineal, con restricciones representadas como ecuaciones o desigualdades lineales. El objetivo es encontrar los valores de las variables que maximizan o minimizan la función sujeta a las restricciones (Cauas, 1966). La programación lineal es una potente herramienta para optimizar sistemas lineales y tiene una amplia gama de aplicaciones, como la optimización de carteras, la planificación de la producción y la asignación de recursos.

La programación entera mixta es un tipo de programación matemática en la que intervienen variables continuas y discretas. Este tipo de método es útil cuando el problema incluye variables que sólo pueden tomar ciertos valores (como los valores enteros o binarios), así como variables

que pueden tomar cualquier valor dentro de un rango (como los números reales). La programación entera mixta se utiliza en diversas aplicaciones, como la programación, el diseño de redes y la planificación logística.

La programación no lineal es un tipo de programación matemática que trata problemas con restricciones u objetivos no lineales. Los métodos de programación no lineal se utilizan para optimizar funciones no lineales y pueden emplearse para resolver problemas de optimización complejos en diversos campos, como la ingeniería, las finanzas y la economía (Espinosa-Paredes, 2016). Los métodos de programación no lineal también pueden utilizarse para resolver problemas con objetivos múltiples, en los que el objetivo es encontrar la solución óptima que equilibre múltiples objetivos en conflicto.

En resumen, los métodos de programación matemática son herramientas poderosas para resolver problemas complejos de optimización. Ya se trate de programación lineal, programación entera mixta o programación no lineal, estos métodos pueden utilizarse para encontrar las mejores soluciones para una amplia gama de aplicaciones, desde las finanzas a la ingeniería y más allá. Gracias a su capacidad para manejar sistemas complejos, los métodos de programación matemática son una herramienta importante para tomar decisiones informadas y mejorar la eficiencia y eficacia de muchas industrias.

2.4.5 Árboles de decisión

El particionamiento recursivo es una técnica de aprendizaje automático utilizada para construir árboles de decisión para problemas de clasificación y regresión. La idea principal del particionamiento recursivo es dividir el espacio de características en regiones cada vez más pequeñas que contengan variables objetivo homogéneas. El proceso comienza con todo el espacio de características y lo divide en regiones más pequeñas basándose en la mejor característica y su valor umbral que maximiza la separación entre las variables objetivo (Kingsford et al., 2008). El proceso se repite recursivamente en cada región hasta que se cumple un criterio de parada.

El origen del particionamiento recursivo se remonta a principios de la década de 1960, cuando los investigadores empezaron a utilizar árboles de decisión para problemas de clasificación. Los primeros algoritmos utilizados para el particionamiento recursivo eran no recursivos, es decir, el espacio de características se dividía en regiones utilizando un número fijo de divisiones (Høyland, 2001). Con la llegada de los ordenadores, la partición recursiva se hizo posible, y los primeros algoritmos para construir árboles de decisión utilizando la partición recursiva se introdujeron en la década de 1980.

Los modelos de partición recursiva son fáciles de interpretar y se pueden aplicar a relaciones no lineales entre las variables de entrada y la variable objetivo. Además, es robusto ante valores

atípicos y no requiere suposiciones sobre la distribución de las características de entrada. Sin embargo, debemos tener en cuenta que este método puede resultar en sobreajuste, creando árboles demasiado complejos poco generalizables e inestables. Además, en muchas ocasiones los resultados pueden estar sesgado con características con muchos valores únicos lo que dificulta la interpretación y clasificación de los datos (Shu, 2020).

En el ámbito de la clasificación crediticia los árboles de decisión son herramientas muy potentes para predecir el riesgo de impago. Como su propio nombre indica, este método se basa en datos históricos de prestatarios anteriores y utiliza una estructura en forma de árbol para hacer predicciones sobre los nuevos solicitantes de préstamos. El proceso de utilizar árboles de decisión para la clasificación crediticia implica varios pasos clave, como la recopilación y el procesamiento de datos, la construcción del modelo de árbol y el uso del modelo para evaluar a los nuevos solicitantes de préstamos.

El primer paso en el uso de árboles de decisión para la clasificación crediticia es recopilar y procesar los datos históricos de los prestatarios. Estos datos suelen incluir información como ingresos, historial laboral, historial crediticio y otra información financiera relevante. A continuación, los datos se procesan y transforman en un formato que pueda utilizarse para construir el modelo de árbol de decisión. Esto puede implicar, como en muchos otros modelos, la normalización de los datos, la eliminación de valores atípicos o la realización de otros procedimientos estadísticos para garantizar que los datos son adecuados para el modelado (Hernández, 2004).

La clave del proceso en la construcción del árbol es dividir los datos en varias ramas basadas en criterios específicos, como los ingresos, el historial laboral, el historial crediticio, etc. Cada rama representa un punto de decisión, y el resultado al final de cada rama representa el impago o no impago del préstamo. El árbol continúa dividiéndose en ramas adicionales, basadas en criterios cada vez más específicos, hasta llegar a una predicción final. El árbol puede construirse utilizando varios algoritmos, como los algoritmos ID3, C4.5 o CART, entre otros, y puede optimizarse utilizando técnicas como la poda o la validación cruzada.

Una vez construido el modelo de árbol de decisión, puede utilizarse para evaluar nuevos solicitantes de préstamos introduciendo sus datos y siguiendo las ramas hasta una predicción de impago o no impago. En algunos casos, el modelo de árbol de decisión puede combinarse con otros modelos, como la regresión logística o las redes neuronales, para crear un sistema de puntuación crediticia más sólido y preciso.

2.4.6 Redes neuronales

Las redes neuronales son un tipo de modelo de aprendizaje automático inspirado en la estructura y el funcionamiento del cerebro humano. Están formadas por nodos interconectados, llamados neuronas artificiales, que procesan información y aprenden de los datos de entrada para producir resultados (Cheng, 1994).

Las redes neuronales tienen el potencial de revolucionar el campo de la calificación crediticia. Al procesar grandes cantidades de datos y aprender de ellos, pueden hacer predicciones más precisas sobre la solvencia de un prestatario y conducir a un proceso de calificación crediticia más justo y eficiente (Fadlalla, 2001). Sin embargo, hay que tener cuidado al garantizar que los datos utilizados para entrenar la red sean de alta calidad y que las decisiones tomadas por la red sean transparentes y comprensibles.

En el campo de la calificación crediticia, las redes neuronales pueden utilizarse para analizar y hacer predicciones sobre la solvencia de un prestatario. Esto implica utilizar una gran cantidad de datos financieros, como el historial crediticio, los ingresos y la situación laboral, para entrenar a la red neuronal a identificar patrones y tomar decisiones sobre si es probable que una persona devuelva un préstamo.

Una de las ventajas de utilizar redes neuronales para la calificación crediticia es su capacidad para manejar relaciones complejas y no lineales entre variables. Esto es importante en la calificación crediticia, ya que puede haber muchos factores que influyan en la capacidad de un prestatario para devolver un préstamo, y estos factores no siempre tienen una relación directa. Las redes neuronales también son capaces de identificar y ponderar automáticamente la importancia de cada variable, lo que puede hacer que el proceso de calificación crediticia sea más eficiente y preciso. (Hinton, 1992). Otra ventaja de las redes neuronales es su capacidad de aprender y mejorar con el tiempo. A medida que la red neuronal procesa más datos, puede afinar sus predicciones y tomar decisiones más precisas. Esto puede conducir a un proceso de calificación crediticia más justo y equitativo, ya que la red neuronal puede aprender de los errores del pasado y ajustar su toma de decisiones en consecuencia.

Sin embargo, el uso de redes neuronales para la calificación crediticia tiene algunas limitaciones. Una de ellas es la necesidad de grandes cantidades de datos para entrenar la red, así como la necesidad de datos de alta calidad que reflejen con precisión las variables que influyen en la solvencia (Ripley, 1994). Además, puede haber dudas sobre la interpretabilidad de los modelos de redes neuronales, ya que las decisiones tomadas por la red a veces pueden ser difíciles de entender o explicar.

2.4.7 Modelos variables en el tiempo

Los modelos variables en el tiempo aplicados al credit scoring, tienen en cuenta la naturaleza cambiante de la situación financiera de un prestatario a lo largo del tiempo. A diferencia de los modelos tradicionales de calificación crediticia, que utilizan información estática para hacer predicciones, los modelos variables en el tiempo incorporan información dinámica, como cambios en los ingresos, la situación laboral y el historial crediticio, para hacer predicciones más precisas (Beck, 1983).

Los modelos variables en el tiempo se caracterizan por su capacidad de incorporar cambios en la situación financiera de un prestatario a lo largo del tiempo. Los modelos tradicionales de calificación crediticia, como muchos de los explicados anteriormente, utilizan información estática para hacer predicciones, lo que puede resultar problemático en situaciones en las que la situación financiera del prestatario cambia significativamente.

En cambio, los modelos variables en el tiempo pueden incorporar cambios en los ingresos, la situación laboral y el historial crediticio, lo que permite realizar predicciones más precisas sobre la solvencia de un prestatario. Esta mayor precisión puede conducir a decisiones de aprobación de préstamos más informadas y, en última instancia, reducir el riesgo de impago. Además, tienen capacidad para adaptarse a los cambios en las condiciones del mercado. Por ejemplo, durante las recesiones económicas, el riesgo de impago puede aumentar, y los modelos variables en el tiempo pueden tener esto en cuenta y ajustar sus predicciones en consecuencia. Esta capacidad de adaptación a los cambios en las condiciones del mercado hace que los modelos variables en el tiempo sean especialmente útiles en el campo de la calificación crediticia, donde las condiciones del mercado pueden tener un impacto significativo en la capacidad de un prestatario para devolver un préstamo.

A pesar de sus ventajas, los modelos variables en el tiempo también tienen algunas desventajas. Pueden ser más complejos de construir y aplicar que los modelos tradicionales de calificación crediticia. Esto se debe a que estos requieren la recopilación y el procesamiento de más datos, y los propios modelos suelen ser más complejos, lo que puede dificultar su comprensión e interpretación. Además, los modelos variables en el tiempo pueden requerir más recursos informáticos para su ejecución, lo que puede encarecer su aplicación.

A pesar de estas desventajas, los modelos variables en el tiempo se han aplicado con éxito en el campo de la calificación crediticia de diversas maneras. Por ejemplo, algunas instituciones utilizan modelos variables en el tiempo para incorporar cambios en el historial crediticio a sus modelos de calificación crediticia, lo que puede mejorar la precisión de sus predicciones (Kolar, 2010). Otras instituciones utilizan modelos variables en el tiempo para incorporar cambios en las

condiciones del mercado en sus modelos, lo que les permite adaptarse a los cambios en las condiciones del mercado y tomar decisiones de aprobación de préstamos más informadas.

En conclusión, los modelos variables en el tiempo ofrecen varias ventajas sobre los modelos tradicionales de calificación crediticia, como su capacidad para incorporar cambios en la situación financiera del prestatario a lo largo del tiempo y su capacidad para adaptarse a los cambios en las condiciones del mercado. A pesar de sus desventajas, como su complejidad y sus mayores requisitos computacionales, estos modelos se han aplicado con éxito en el campo de la calificación crediticia y tienen el potencial de mejorar la precisión de las decisiones de aprobación de préstamos y reducir el riesgo de impago.

3. PARTE II. CREACION DE UN MODELO DE RIESGO DE IMPAGO

3.1. Resumen

Tras haber realizado una revisión bibliográfica de los modelos de riesgo, en este apartado del trabajo crearemos un modelo de riesgos llamado *Riskforecast*. El principal objetivo del modelo es predecir la probabilidad de impago de una empresa en función de distintas variables que van a ser creadas durante el proceso. Esto se realizará mediante un modelo estadístico que nos permitirá predecir con fiabilidad la probabilidad que tiene una empresa determinada de impagar. Como explicaremos posteriormente, hemos utilizado un modelo de regresión logit. Además, hemos partido de una base de datos de más de 1778 empresas no financieras que nos ha servido para ajustar el modelo y hacerlo lo más significativo posible. A medida que vamos avanzando, veremos cómo esta base de datos se va reduciendo puesto que muchas empresas carecen de algunos datos.

La finalidad durante el proceso de construcción del modelo es minimizar los errores de tipo I y de tipo II. El error de tipo uno, o también llamado “falso positivo”, se produce cuando se rechaza la hipótesis nula cuando esta es en realidad cierta. El error de tipo II, también conocido como “falso negativo”, se produce cuando se acepta incorrectamente la hipótesis nula. En este caso el error de tipo I sería catalogar a una empresa que no puede pagar como una que no puede hacerlo y el error de tipo II sería clasificar una empresa que si puede pagar como una que no puede hacerlo). Por ello, hemos intentado, en la medida de lo posible, reducir ambos errores para que nuestro modelo de regresión sea lo más significativo posible y nos ofrezca una buena predicción a partir de nuestra base de datos inicial.

3.2 Metodología

El modelo presentado es un modelo de Regresión Logística o “Logit”. En este caso, se trata de un modelo Logit dicotómico (dos posibles alternativas excluyentes entre sí). Este modelo se basa en la probabilidad de impago de una entidad en un corto periodo de tiempo y, a partir de esto, se pretende normalizar la respuesta de aceptación o rechazo de créditos. Como ya hemos analizado anteriormente, el uso de un modelo logit presenta ventajas como su flexibilidad y facilidad de uso, ya que la respuesta obtenida es binaria. Sin embargo, una de las limitaciones que presenta este tipo de modelos es el error en la previsión, ya que no es posible que estos modelos ofrezcan una fiabilidad del 100% en los resultados obtenidos.

3.3 SPSS (Statistical Package for the Social Sciences)

SPSS (Statistical Package for the Social Sciences), es un programa de software de análisis y gestión de datos. Se trata de un paquete de software potente y ampliamente utilizado para el

análisis y la gestión de datos. A finales de los años 60, Norman Nie, Dale Bent y C. Hadlai Hull crearon SPSS. Su objetivo principal era ser un instrumento para académicos. Sin embargo, debido a su facilidad de uso y a su gama de habilidades estadísticas, rápidamente se hizo popular en otros sectores. SPSS fue adquirido por IBM en 2009, y desde entonces se ha convertido en uno de los paquetes de software de análisis de datos más utilizados del mundo.

El programa se caracteriza por su interfaz fácil de usar ya que tiene un diseño muy intuitivo. Además, posee una amplia gama de técnicas estadísticas que lo convierten en una herramienta ideal para investigadores, analistas y profesionales de diversos campos. SPSS analiza datos de clientes, realiza estudios de mercado o desarrolla modelos predictivos.

El análisis descriptivo, las pruebas estadísticas, el análisis multivariante y el análisis visual son algunas de las técnicas estadísticas que ofrece SPSS. Estos métodos pueden utilizarse para estudiar datos de diversos sectores. SPSS dispone de potentes herramientas para la gestión de datos que facilitan a los clientes la manipulación y limpieza de enormes conjuntos de datos. Además, el programa permite realizar gráficos y figuras que pueden ser de gran utilidad en el ámbito empresarial.

La interfaz de SPSS es una de sus principales ventajas. En concreto, ofrece una vista de los datos introducidos que permite visualizarlos rápidamente. Además, contiene opciones desplegadas que simplifican el uso de la herramienta y permiten al usuario analizar rápidamente los datos. Ofrece una interfaz fácil de usar y una gran variedad de enfoques estadísticos, lo que lo convierte en la herramienta perfecta para el análisis de datos.

El programa permite a los usuarios importar datos de diversas fuentes, como hojas de cálculo, bases de datos y datos de encuestas. Una vez importados los datos, pueden manipularse, limpiarse y transformarse según sea necesario para el análisis. SPSS proporciona una serie de herramientas gráficas para visualizar los datos, como histogramas, diagramas de dispersión y diagramas de caja, que permiten a los usuarios identificar rápidamente patrones y relaciones en los datos.

SPSS también proporciona una amplia gama de procedimientos estadísticos para realizar análisis complejos. Por ejemplo, SPSS incluye herramientas para la comprobación de hipótesis, análisis de regresión, análisis factorial y análisis de conglomerados, que pueden utilizarse para comprobar las relaciones entre variables e identificar patrones en los datos. SPSS también incluye una serie de herramientas de análisis predictivo, como árboles de decisión y redes neuronales, que pueden utilizarse para realizar predicciones basadas en datos históricos.

En el ámbito empresarial, SPSS se ha utilizado para analizar datos de clientes, realizar estudios de mercado y desarrollar modelos predictivos para predecir las ventas y el comportamiento de los clientes. Es importante también mencionar que el programa no solo se utiliza en el ámbito

financiero sino también en diversos campos, como la empresa, la educación, la sanidad y la administración pública.

3.4 Discriminación de variables

En el proceso de construcción del modelo llevaremos a cabo, en algunos casos, la discriminación de variables, la acotación de los valores tanto por arriba como por abajo y la asignación de valores según intervalos. En cada una de las variables, si se da el caso, especificaremos si se ha discriminado la variable en cuestión.

Esta discriminación se lleva a cabo por distintos motivos. La eliminación de ciertas variables se lleva a cabo si estas resultan irrelevantes para el modelo, es decir, si la variable no tiene una correlación significativa con la variable dependiente o no contribuye a explicar la variabilidad de la variable dependiente. En este caso, la permanencia de la variable en el modelo puede afectar negativamente a la precisión y estabilidad del modelo.

Además, en muchos casos, se producirá la multicolinealidad. En este caso la variable en cuestión está altamente correlacionada con otras variables incluidas en el modelo por lo que también puede afectar negativamente al desempeño del modelo.

Debemos de tener en cuenta también tanto el ruido como el sesgo del modelo. El ruido se refiere a la variabilidad aleatoria o no sistemática que se encuentra en los datos y que no está relacionada con la variable objetivo o dependiente que se desea predecir. En otras palabras, el ruido es el error no explicado que se encuentra en el conjunto de datos que se utiliza para entrenar un modelo. El ruido puede ser causado por factores externos no controlados, errores de medición, datos incompletos entre otros. Los modelos que están demasiado enfocados en el ruido pueden ser menos precisos y útiles para predecir nuevos datos (Sapatinas, 2014).

Sesgo se refiere a la tendencia sistemática del modelo a predecir valores que se desvían del valor real. el sesgo se produce cuando el modelo esta simplificado o mal especificado y, por lo tanto, no puede representar adecuadamente la complejidad de los datos. los modelos con un alto sesgo ceden no ser capaces de capturar patrones y relaciones importantes en los datos. lo que lleva a predicciones inexactas (Sapatinas, 2014).

3.5 Construcción del modelo

3.5.1 Descripción de la muestra

La base de datos contiene la siguiente información para cada una de las entidades:

- Información con frecuencia mensual sobre si las entidades han impagado o no desde febrero de 2012 hasta abril de 2013. Si la empresa ha impagado será clasificada como un 1 y si la empresa ha pagado será clasificada con un 0.

- Información anual sobre la situación contable de cada entidad desde 2008 hasta 2011. No obstante, en algunas entidades no está disponible la información contable para los cuatro años.

De manera resumida, los pasos que hemos seguido para la obtención del modelo son:

1) Limpiar la base de datos. Eliminamos las empresas para cuales no había información contable de los cuatro años.

2) Obtención de la variable dependiente (Y) la cual toma valores 1 o 0 según si la entidad ha impagado o no.

3) Construcción de variables explicativas (X) del modelo a partir de la información contable de cada empresa.

4) Empleo del programa SPSS para construir el modelo de probabilidad de impago, que modeliza la determinación de la probabilidad de impago de cada entidad.

5) Discusión de los resultados obtenidos.

Una vez realizada la limpieza de la base de datos, la muestra contiene información sobre 1778 empresas. El número total de impagos durante los años 2012 y 2013 ha sido de 120 entidades, es decir, un 6,75 % del total de las entidades.

3.5.2 Variables dependientes y explicativas del modelo

La variable dependiente del modelo es la variable que predice el impago de las empresas, basándose en las variables dependientes que hemos ido construyendo y seleccionando tras un análisis detallado.

Las variables independientes o explicativas del modelo se han construido con el objetivo de que aporten información con sentido financiero sobre la situación en la que se encuentra cada entidad. Estas variables explican características tales como rentabilidad, liquidez o estructura de la deuda. A continuación, se explican cada una de las variables seleccionadas con detalle:

1) **Margen BAIT**

Es el resultado de dividir BAIT/Ventas y nos indica el resultado de explotación que se obtiene por cada unidad de venta. El margen BAIT, o también conocido como EBIT, mide la rentabilidad de una empresa según las ganancias obtenidas antes del pago de tributos e intereses. A la hora de interpretar la variable margen BAIT/Ventas, podemos afirmar que si se mantienen las ventas constantes y el margen disminuye, crece la probabilidad de impago. Debemos de tener en cuenta no solo la variación, sino también el valor absoluto para ver si el riesgo aumenta o disminuye.

2) **Margen BN**

Es el resultado entre BAIT/VENTAS. Nos permite saber el resultado de explotación que genera la empresa. La empresa tendrá más probabilidad de impagos a medida que las ventas permanezcan constantes y el margen de BN vaya disminuyendo.

3) **COSTES FIJOS**

Hemos visto como es imprescindible tener en cuenta la estructura de costes fijos de una empresa, puesto que, a mayores costes fijos, más riesgo de impago posee la empresa. Como ya sabemos, los costes fijos representan una cantidad invariable y son independientes de la actividad de producción, al menos durante un periodo de tiempo determinado.

4) **ROA I**

ROA o también llamado *Return on Assets* se utiliza para evaluar la rentabilidad de una empresa en relación con sus activos totales. Esta medida financiera nos indica cuánto genera una empresa por cada unidad monetaria invertida en activos. Es importante mencionar que esa medida es muy útil para comparar la eficiencia de las empresas en la utilización de sus activos para generar ganancias. Lo hemos calculado haciendo BAIT/Total Activo. Esta variable la hemos capado entre 0,3 y - 0,3. El ROA nos indica cuantas unidades monetarias (€) obtengo por cada inversiónn de 100 euros en mis activos.

5) **ROA DISCRIMINADO**

Hemos discriminado la variable del ROA de tal forma que, si es mayor que 0,12, lo clasificamos como un 4, si es mayor de 0,05 como un 3, si es mayor que 0 con un 2 y si es menor que 0, como un 1.

6) **Rentabilidad real vs Esperada**

En esta variable hemos comparado la rentabilidad que se espera obtener en un futuro con la rentabilidad que se ha obtenido realmente. En primer lugar, la rentabilidad esperada es la ganancia que se espera obtener de una inversión en función de las proyecciones del mercado y la

información disponible en el momento de la inversión. Este tipo de rentabilidad se calcula utilizando modelos matemáticos y estadísticos que tienen en cuenta variables como el riesgo, la tasa de interés, la volatilidad del mercado, entre otros. Por otro lado, la rentabilidad real es la ganancia real cuyo resultado se obtiene después de haber realizado la inversión, teniendo en cuenta los costes y los rendimientos reales que se han obtenido. Esta rentabilidad es más precisa. Ambas rentabilidades no siempre coinciden por lo que hemos creado una variable que compara ambas.

7) LIQUIDEZ

Es la facilidad que tienen los activos de la empresa en ser convertidos en ingresos monetarios de manera rápida. Queremos que el índice de liquidez sea mayor a 1, es decir, que la empresa sea capaz de convertir los activos en efectivo en un periodo relativamente corto. Si, por el contrario, el índice de liquidez es menor que 1, la empresa tiene problemas con la liquidez.

8) APALANCAMIENTO

El apalancamiento es la relación entre el crédito y el capital propio invertido. Consiste en usar la deuda para financiar una operación, por ello, en este caso cuando mayor apalancamiento, mayor riesgo de impago. Si dividimos el pasivo entre fondos propios y nos da cercano a 1, significa que la empresa tiene las mismas cantidades de recursos ajenos y las deudas. Si es menor que 1 significa que la empresa no quiere asumir grandes riesgos con su deuda.

9) ROE_IA

El ROE o también llamado Return on Equity, es una medida financiera utilizada para evaluar la rentabilidad de una empresa en relación con el capital invertido por sus accionistas. Esta medida nos indica cuántas unidades monetarias genera una empresa por cada unidad monetaria invertida por sus accionistas. El ROE es una medida importante para evaluar la eficiencia con la que una empresa utiliza el capital invertido por sus accionistas para generar ganancias. El ROE o "Return on Equity" tiene que ser mayor, porque el apalancamiento es ROE/ROA, nos interesa que el índice de apalancamiento sea mayor que 1.

10) Índice Apalancamiento

Es el resultado de ROE/ROA, por lo que el resultado nos aporta información sobre la estructura de financiación de la empresa. Si es igual a 1, significa que la totalidad del activo ha sido financiado con fondos propios. Si el resultado es mayor a 1, significa que el pasivo tiene efectos positivos. Sin embargo, si es menor que 1, el endeudamiento tendrá efectos negativos.

11) ACID TEST

Se calcula como $(\text{Activo Corriente} - \text{Existencias}) / \text{Pasivo Corriente}$. Se calcula para estudiar la liquidez de cada entidad, es decir, la capacidad de hacer frente a las obligaciones corrientes sin tener en cuenta la venta de inventarios. De esta manera, si el resultado es menor a 1 significa que la entidad tiene problemas de liquidez (de sus saldos de efectivo) y, si es mayor a 1, lo contrario, disminuyendo la probabilidad de impago. Esta variable está capada por arriba en el valor de 4.

12) FCL

Flujo de caja libre Es el resultado de: $\text{Resultado de explotación} - (0,25 * \text{Resultado de explotación}) - \text{Amortización inmovilizado} + \text{CAPEX} + \text{Variación NOF}$. El CAPEX son las inversiones en activos fijos. El FCL representa lo que genera la empresa con la explotación de su negocio, es decir, el dinero disponible una vez se han cubierto las necesidades de reinversión en activos fijos y necesidades operativas de fondos, considerando que no existe deuda y no hay cargas financieras.

13) FCL Sin Capex

Se ha calculado como $\text{FCL} - \text{CAPEX}$. La variable calculada representa lo que genera la explotación del negocio en el corto plazo. Se ha calculado esta variable ya que una entidad puede tener un FCL negativo por haber hecho una gran inversión en activos a largo plazo lo cual no es necesariamente negativo. No obstante, si esta variable también es negativa, la empresa sí que tendrá una alta probabilidad de impago.

14) Deuda Financiera Neta

Esta variable es el resultado de $\text{deuda financiera c/p} + \text{gastos financieros} - \text{tesorería} - \text{ingresos financieros}$. De esta manera, si la variable es negativa, significa que la entidad puede hacer frente a sus deudas financieras con la tesorería e ingresos financieros, es decir, dispone de más dinero del que debe a las entidades financieras (bancos). Por lo tanto, cuanto menor sea esta variable, mejor.

15) FCL/ Deuda Financiera Neta

Una vez tenemos el FCL, hemos calculado la deuda financiera neta, pero debemos tener cuidado al dividir ambos resultados ya que podemos obtener distintos escenarios:

- FCL – y la deuda neta es -, valor absoluto
- FCL + y la deuda + calculamos el ratio normal, porque iba a salir positivo
- FCL – y deuda +, ratio normal

- FCL+ y deuda -, hacemos también el ratio con valor absoluto

16) FCL/ Deuda financiera c/p

Esta variable se interpreta de la siguiente forma: Si es negativa significa que el FCL es negativo, lo que supone que la entidad no puede hacer frente a sus deudas y, por tanto, incrementa la probabilidad de impago. Si es menor que 1, significa que la deuda financiera c/p es mayor que el FCL y, por tanto, incrementa también la probabilidad de impago. Por último, si es mayor a 1, significa que el FCL es mayor que la deuda financiera c/p, es decir, los fondos que genera la entidad son mayores que sus deudas contraídas lo cual disminuye la probabilidad de impago. Sin embargo, hay que tener en cuenta que el FCL puede ser negativo por una gran inversión en activos a largo plazo, por lo que en estos casos habrá que analizar también la variable FCL sin capex / Deuda financiera c/p para confirmar el buen o mal estado financiero de cada entidad.

17) FCL sin Capex / Deuda financiera neta

La interpretación de esta variable es la misma que la de FCL / Deuda financiera c/p.

3.5.3 Modificaciones en las variables explicativas del modelo

Para obtener variables que sean significativas, hemos modificado algunas de ellas. Tanto elevarlas al cuadrado como al cubo nos han permitido hacer nuestro modelo un mejor predictor que sin ellas. Además, la variación del logaritmo también fue un paso importante a la hora de transformar variables construyendo nuevas. Además, hemos transformado algunas de las variables que hemos creado inicialmente con el objetivo de conseguir nuevas variables significativas. Las que hemos añadido al modelo son:

- ROA1_cuadrado
- ROA3_cuadrado
- ROAyANCD4_cuadrado
- Liquidez3_cuadrado
- Apalancamiento3_cuadrado
- Variación del LN de los costes fijos.

Por último, quitamos la constante, puesto que la habíamos quitado anteriormente y nos daba error, por ello dejamos este paso para el final.

3.6 Obtención del modelo

3.6.1 Pasos previos

Para construir el modelo empleamos SPSS. En primer lugar, hicimos una regresión logística binaria con todas las variables explicadas en el apartado 3. Para ello, empleamos un punto de corte de 0,0675 por ser este el porcentaje de impagados de la muestra. Sin embargo, en vez de meter todas las variables al mismo tiempo, las fuimos añadiendo al modelo poco a poco de tal forma que pudiésemos observar la evolución de la performance del modelo (R cuadrado y porcentaje de impagados correctamente clasificados) según qué variables incluíamos en el modelo. De esta manera, las que empeoraban el modelo las eliminamos.

En otras palabras, al ir añadiendo variables nos fijábamos en si estas eran significativas o no. Aquellas que no lo eran, fueron eliminadas. Algunas de estas fueron, por ejemplo, el índice de Apalancamiento del año 3, el Flujo de Caja Libre sin Capex del año 4, y el Flujo de Caja Libre entre Deuda de los años 3 y 4. Una vez probamos con todas las variables “iniciales”, comenzamos a introducir cuadrados y cubos de las variables más significativas, considerando como tales a las variables con significación inferior a 0,8 y betas muy cercanas a 0. Además, realizamos otras transformaciones como el logaritmo neperiano. De esta manera, incrementamos el R cuadrado, es decir, aumentamos el porcentaje de la variabilidad que el modelo es capaz de explicar. Además, también aumentó el porcentaje de impagados correctamente clasificados. Finalmente, quitamos la constante lo cual nos llevó a incrementar el R cuadrado hasta un 0,8 aproximadamente.

3.6.2 Modelo definitivo

Como hemos explicado en el apartado anterior, el modelo definitivo se obtuvo eliminando aquellas variables menos significativas, introduciendo transformaciones de las más significativas y finalmente quitando la constante.

Tabla 1. Resumen del modelo

Resumen del modelo			
Paso	Logaritmo de la verosimilitud -2	R cuadrado de Cox y Snell	R cuadrado de Nagelkerke
1	400,749 ^a	,640	,853

a. La estimación ha terminado en el número de iteración 20 porque se ha alcanzado el máximo de iteraciones. La solución final no se puede encontrar.

Fuente: Elaboración propia.

Como bien se muestra en la tabla de resumen, el R cuadrado obtenido es 0,853, es decir, el modelo explica el 85,3% de la variabilidad de los datos.

Tabla 2. Tabla de clasificación

Tabla de clasificación^a

Observado		Pronosticado		
		Variable Dependiente		Porcentaje correcto
		0	1	
Paso 1	Variable Dependiente 0	773	247	75,8
	1	14	64	82,1
Porcentaje global				76,2

a. El valor de corte es ,068

Fuente: Elaboración propia.

En la tabla de clasificación podemos ver que el porcentaje de empresas que pagan el préstamo son predichas por el modelo correctamente en un 75.8% de las veces. Además, el porcentaje de las empresas que impagan es predicho por el modelo correctamente un 82.1%.

Por último, en las tablas 3,4,5 y 6 se muestran las variables explicativas empleadas en el modelo, así como las respectivas betas, el error estándar y la significación.

Tabla 3. Variables en la ecuación

Variables en la ecuación

		B	Error estándar	Wald	gl	Sig.	Exp(B)
Paso 1 ^a	MARGEN BAIT 1	-1,734	4,213	,169	1	,681	,177
	MARGEN BAIT 2	-,968	5,067	,036	1	,849	,380
	MARGEN BAIT 3	-3,218	4,709	,467	1	,494	,040
	MARGEN BAIT 4	2,797	3,459	,654	1	,419	16,397
	MARGEN BN 1	3,436	3,087	1,238	1	,266	31,061
	MARGEN BN 2	,701	3,096	,051	1	,821	2,017
	MARGEN BN 3	1,590	3,061	,270	1	,604	4,902
	MARGEN BN 4	-5,904	2,453	5,790	1	,016	,003
	ROA 1	-,616	6,618	,009	1	,926	,540
	ROA 2	8,809	6,523	1,824	1	,177	6692,510
	ROA 3	-6,463	6,200	1,087	1	,297	,002
	ROA 4	4,764	4,697	1,029	1	,310	117,206
	ROA DISCRIMINADO 1			,459	3	,928	
	ROA DISCRIMINADO 1(1)	,503	1,795	,078	1	,780	1,653
	ROA DISCRIMINADO 1(2)	,505	1,115	,205	1	,651	1,656
	ROA DISCRIMINADO 1(3)	,191	,898	,045	1	,831	1,211
	ROA DISCRIMINADO 2			1,820	3	,611	
	ROA DISCRIMINADO 2(1)	1,803	1,807	,996	1	,318	6,065
	ROA DISCRIMINADO 2(2)	1,436	1,143	1,578	1	,209	4,205
	ROA DISCRIMINADO 2(3)	,833	,942	,783	1	,376	2,301
	ROA DISCRIMINADO 3			14,299	3	,003	
	ROA DISCRIMINADO 3(1)	-2,462	1,761	1,955	1	,162	,085
	ROA DISCRIMINADO 3(2)	-3,673	1,123	10,692	1	,001	,025
	ROA DISCRIMINADO 3(3)	-3,158	,926	11,622	1	,001	,043
	ROA DISCRIMINADO 4			8,329	3	,040	
	ROA DISCRIMINADO 4(1)	-4,998	2,754	3,293	1	,070	,007
	ROA DISCRIMINADO 4(2)	3,047	1,497	4,142	1	,042	21,043
	ROA DISCRIMINADO 4(3)	2,375	1,404	2,861	1	,091	10,753

Fuente: Elaboración propia.

Tabla 4. Variables en la ecuación

ANC DISCRIMINADO 2(1)	,847	,982	,744	1	,388	2,332
ANC DISCRIMINADO 3(1)	-,192	,904	,045	1	,831	,825
ANC DISCRIMINADO 4(1)	-2,071	,726	8,139	1	,004	,126
ROAyANC 1	,017	,436	,001	1	,970	1,017
ROAyANC 2	,229	,437	,274	1	,600	1,257
ROAyANC 3	,479	,436	1,204	1	,272	1,614
ROAyANC 4	-2,646	1,039	6,481	1	,011	,071
ROE 1	1,191	1,861	,409	1	,522	3,289
ROE 2	-2,028	1,821	1,241	1	,265	,132
ROE 3	-1,786	1,502	1,415	1	,234	,168
ROE 4	,753	1,197	,396	1	,529	2,123
Liquidez 1	,065	,755	,007	1	,931	1,067
Liquidez 2	-1,565	,881	3,152	1	,076	,209
Liquidez 3	3,594	1,393	6,662	1	,010	36,383
Liquidez 4	-1,027	,740	1,926	1	,165	,358
Apalancamiento 1	-,124	2,542	,002	1	,961	,883
Apalancamiento 2	1,806	3,040	,353	1	,553	6,083
Apalancamiento 3	5,346	4,465	1,433	1	,231	209,747
IA 2	,217	,135	2,580	1	,108	1,242
ACID TEST 1	-,241	,812	,088	1	,766	,786
ACID TEST 2	1,060	,979	1,171	1	,279	2,885
ACID TEST 3	-2,105	1,135	3,435	1	,064	,122
ACID TEST 4	1,144	,791	2,091	1	,148	3,140
Flujo de caja libre 2	,000	,000	,285	1	,594	1,000
Flujo de caja libre 3	,000	,000	,620	1	,431	1,000
Flujo de caja libre 4	,000	,000	6,330	1	,012	1,000

Fuente: Elaboración propia.

Tabla 5. Variables en la ecuación

Flujo de cja libre sin capex 3	,000	,000	1,408	1	,235	1,000
Flujo de cja libre sin capex 2	,000	,000	5,059	1	,024	1,000
FCLDI 2	-,001	,021	,001	1	,979	,999
FCLDF 3	,000	,000	2,481	1	,115	1,000
FCLDF 4	,000	,000	,052	1	,819	1,000
FCL (sin capex)/DF 2	,000	,000	3,536	1	,060	1,000
FCL (sin capex)/DF 4	,000	,000	,201	1	,654	1,000
DI Vabsoluto 1	,000	,000	,008	1	,929	1,000
DI Vabsoluto 3	,000	,000	,044	1	,835	1,000
FCLDI 4	,000	,002	,017	1	,895	1,000
ROA1_Cuadrado	-9,022	20,634	,191	1	,662	,000
ROA3_Cuadrado	-10,452	19,174	,297	1	,586	,000
ROAyANCD4_Cuadrado			7,615	2	,022	
ROAyANCD4_Cuadrado (2)	-5,728	2,076	7,615	1	,006	,003
ROAyANCD4_Cuadrado (3)	-23,383	4003,201	,000	1	,995	,000
Liquidez3_Cuadrado	-,243	,193	1,582	1	,208	,785
Apalancamiento3_Cuadrado	-2,188	2,736	,640	1	,424	,112
IA 1	-,178	,128	1,939	1	,164	,837
IA 4	-,052	,090	,334	1	,563	,949
ANC DISCRIMINADO 1(1)	-,413	,853	,235	1	,628	,661

Fuente: Elaboración propia.

Tabla 6. Variables en la ecuación

Costes fijos 1	,000	,001	,000	1	,999	1,000
Costes fijos 2	,000	,001	,000	1	,999	1,000
Costes fijos 3	,000	,000	,000	1	1,000	1,000
Costes fijos 4	,000	,000	,000	1	,999	1,000
LN. Costes_fijos_variacion14	,260	,122	4,542	1	,033	1,296

a. Variables especificadas en el paso 1: MARGEN BAIT 1, MARGEN BAIT 2, MARGEN BAIT 3, MARGEN BAIT 4, MARGEN BN 1, MARGEN BN 2, MARGEN BN 3, MARGEN BN 4, ROA 1, ROA 2, ROA 3, ROA 4, ROA DISCRIMINADO 1, ROA DISCRIMINADO 2, ROA DISCRIMINADO 3, ROA DISCRIMINADO 4, ANC DISCRIMINADO 1, ANC DISCRIMINADO 2, ANC DISCRIMINADO 3, ANC DISCRIMINADO 4, ROA y ANC 1, ROA y ANC 2, ROA y ANC 3, ROA y ANC 4, ROE 1, ROE 2, ROE 3, ROE 4, Liquidez 1, Liquidez 2, Liquidez 3, Liquidez 4, Apalancamiento 1, Apalancamiento 2, Apalancamiento 3, IA 2, ACID TEST 1, ACID TEST 2, ACID TEST 3, ACID TEST 4, Flujo de caja libre 2, Flujo de caja libre 3, Flujo de caja libre 4, Flujo de cja libre sin capex 3, Flujo de cja libre sin capex 2, FCL/DI 2, FCL/DI 3, FCL/DI 4, FCL (sin capex)/DF 2, FCL (sin capex)/DF 4, DI Vabsoluto 1, DI Vabsoluto 3, FCL/DI 2, FCL/DI 4, ROA1_Cuadrado, ROA3_Cuadrado, ROAyANC4_Cuadrado, Liquidez3_Cuadrado, Apalancamiento3_Cuadrado, IA 1, IA 4, ANC DISCRIMINADO 1, Costes fijos 1, Costes fijos 2, Costes fijos 3, Costes fijos 4, LN.Costes_fijos_variacion14.

Fuente: Elaboración propia.

4. CONCLUSIONES

A modo de cierre de este trabajo de fin de grado, podemos señalar que hemos realizado un estudio completo de los modelos de riesgo de impago. Al haber analizado en detalle los distintos tipos de modelos que se utilizan para la predicción de riesgos, podemos concluir que estos son muy variados y tienen distintas finalidades. Cada uno de ellos tiene sus ventajas e inconvenientes y está creado para alcanzar distintos objetivos. Por ello, es importante seleccionar con cierto criterio el modelo que se va a utilizar dependiendo de los objetivos que se quieren alcanzar.

También, tras haber hecho un estudio de los modelos de riesgo desde su inicio podemos concluir que tienen un impacto claramente positivo en las empresas que los utilizan. En nuestro caso, al tratarse de empresas no financieras, con el uso de los modelos, estas pueden calcular la probabilidad de que sus clientes no cumplan con las obligaciones de pago. La utilización de los modelos no solo puede ayudar a tomar medidas preventivas para evitar pérdidas financieras, sino que también ayuda a las empresas a tener una mejor gestión del crédito.

Teniendo información precisa de que un cliente cumpla sus obligaciones de pago, la empresa puede ajustar sus políticas de créditos y llevar una mejor organización. Podemos afirmar que las empresas aprovechan estos modelos tanto para reducir costes asociados con el cobro de deuda y gestión de morosidad como para mejorar la rentabilidad al permitir un enfoque mayor en aquellos clientes que son más rentables y que tienen menos riesgo. Todo ello, se traduce en un ahorro de tiempo y recursos que pueden destinarse a otras áreas del negocio haciendo que la empresa crezca y ganando posición competitiva en el mercado.

Además, tras haber analizado la revisión bibliográfica a cerca de los modelos de riesgo de impago para empresas, observamos como los modelos de credit scoring están en constante evolución y mejora. Hemos visto como a lo largo de los años cada vez son más los especialistas en matemáticas y estadística los que introducen novedades a los modelos, consiguiendo que estos sean cada vez más rigurosos a la hora de predecir sucesos futuros. La rápida evolución de la tecnología hace posible la optimización de los modelos, el desarrollo de programas que los realicen y que cada vez la intervención humana sea menor en el proceso. Con todo esto, es importante a la hora de crear modelos de riesgo tener en cuenta las actualizaciones sobre los últimos avances en la materia para poder utilizar las mejores técnicas y herramientas en cada proceso.

El resultado más relevante de este trabajo de fin de grado es el modelo de riesgo de impago creado con la herramienta de SPSS. Es importante entender las medidas de precisión de los modelos para evaluar la calidad de estos. El R cuadrado o también llamado coeficiente de determinación, es una medida utilizada en el campo de la estadística para medir la calidad de un modelo de

regresión, representando la variabilidad en la variable dependiente que puede ser explicada por el modelo (Cox, 1989). Esta medida predice un resultado numérico a partir de una o más variables predictoras, siendo un número entre 0 y 1. Si el R cuadrado está muy cerca del uno, entonces nuestro modelo se ajusta mejor a los datos. Por el contrario, si se acerca a cero, el modelo no se ajusta adecuadamente a los datos. Por ello, el R cuadrado nos proporciona información valiosa sobre la calidad un modelo de regresión. Sin embargo, muchos estadísticos aconsejan que esta medida debe ser utilizada junto con otras con el fin de tener un modelo con la máxima precisión posible. Como hemos visto en el análisis de nuestro modelo, la herramienta de SPSS nos ha proporcionado dos tipos de R cuadrado para nuestro modelo.

El R cuadrado de Cox y Snell se basa en calcular como el cociente entre la probabilidad logarítmica de un modelo completo y la probabilidad logarítmica de un modelo nulo (Cox, 1989). Por otro lado, el R cuadrado de Nagelkerke es la versión ajustada del R cuadrado de Cox y Snell, se calcula como el cociente entre la varianza explicada por el modelo y la varianza total de los datos (Nagelkerke, 1991). Ambas medidas varían entre 0 y 1 y se interpretan de la misma manera. Sin embargo, tienen algunas diferencias. mientras que el R cuadrado de Cox y Snell es utilizado en modelos de regresión lineal, el R cuadrado de Nagelkerke se usa en los modelos de regresión logística, como es nuestro caso.

Cabe destacar también, que el modelo desarrollado tiene una amplia variedad de aplicaciones en distintas áreas relacionadas con la evaluación de riesgo de créditos. El modelo de riesgo de impagos puede ser utilizado, como en nuestro caso por empresas no financieras que quieran conocer el riesgo de proveedores, fabricantes, vendedores o clientes, entre otros, utilizando este modelo para tomar decisiones sobre los plazos de pago y las condiciones de crédito. Además, el modelo puede ser utilizado también para consumidores con el objetivo de conocer sus comportamientos de pagos y cobros y evaluar su riesgo crediticio. Otra de las aplicaciones en las que el modelo construido puede servir de gran utilidad es en las startups. Se podría utilizar para determinar si estas son válidas para obtener financiación, así como para inversores que pueden obtener decisiones a cerca de la inversión y su estrategia.

Aunque el modelo desarrollado en este trabajo tiene múltiples aplicaciones, es importante mencionar que también puede presentar algunas limitaciones. En la creación del modelo, solamente hemos utilizado variables financieras por lo que no se han tenido en cuenta factores externos a estas que pueden afectar al riesgo de impago de una compañía. Por ello, aunque sean más difíciles de cuantificar, resultaría interesante incluir en futuros trabajos variables como la reputación de la empresa, el riesgo del sector en el que opera o riesgos operacionales, entre otros, que pueden afectar al impago. Además, puede mejorar la calidad del modelo incluir en el modelo variables como el historial crediticio tanto de compañías como de personas físicas.

En cuanto a futuras líneas de investigación que podrían ser de gran utilidad relacionadas con el estudio desarrollado, se deberían de tener en cuenta la incorporación de técnicas de aprendizaje profundo, como por ejemplo las redes neuronales analizadas anteriormente. A pesar de su complejidad, las técnicas de aprendizaje profundo pueden resultar muy productivas a la hora de construir modelos puesto que funcionan correctamente con datos no estructurados. Además, facilitan la automatización de los procesos, son escalables y se adaptan a los cambios de los datos. Por todo esto, resulta interesante profundizar en este tipo de técnicas con el objetivo de mejorar la precisión de los modelos o incluso encontrar riesgos que las técnicas utilizadas en este estudio no han detectado.

Además, podemos concluir que el sector financiero no es el único que puede beneficiarse de las ventajas que proporcionan los modelos de negocio. Son numerosas las empresas que pueden aplicar a sus negocios estos modelos. Para los bancos, minoristas y compañías de seguros, entre otros, es fundamental conocer el riesgo de impago de sus clientes. También, sectores como la sanidad, los servicios, la hostelería, los transportes y logística o el textil obtendrían numerosos beneficios con estos modelos. Todo esto supone la adaptación de todo el proceso de creación del modelo al sector en cuestión, teniendo en cuenta las variables, características del sector y clientes específicos. Por ello, es importante seleccionar el modelo que mas se ajuste a los objetivos que se quieran lograr.

6. BIBLIOGRAFIA

REFERENCIAS

- Altman, E. I. (1968). FINANCIAL RATIOS, DISCRIMINANT ANALYSIS AND THE PREDICTION OF CORPORATE BANKRUPTCY. *Journal of Finance*, 23(4), 589-609. <https://doi.org/10.1111/j.1540-6261.1968.tb00843.x>
- Altman, E. I., Iwanicz-Drozowska, M., Laitinen, E. K., & Suvas, A. (2017). Financial Distress Prediction in an International Context: A Review and Empirical Analysis of Altman's Z- Score Model. *Journal of International Financial Management and Accounting*, 28(2), 131-171. <https://doi.org/10.1111/jifm.12053>
- Ayús, A. L. T. (2016). El método popperiano en la estimación de la probabilidad de incumplimiento de un deudor. *Sinapsis*, 8(2), 22-33.
- Beaver, W. H. (1966). Financial Ratios As Predictors of Failure. *Journal of Accounting Research*, 4, 71. <https://doi.org/10.2307/2490171>
- Beaver, W. H. (1968). Market Prices, Financial Ratios, and the Prediction of Failure. *Journal of Accounting Research*, 6(2), 179–192. <https://doi.org/10.2307/2490233>.
- Beck, N. (1983). Time-Varying Parameter Regression Models. *American Journal of Political Science*, 27(3), 557–600. <https://doi.org/10.2307/2110985>
- Berkson, J. (1944). Application of the logistic function to bio-assay. *Journal of the American statistical association*, 39(227), 357-365.
- Bonilla, M., Olmeda, I., & Puertas, R. (2003). Modelos paramétricos y no paramétricos en problemas de credit scoring. *Spanish Journal of Finance and Accounting/Revista Española de Financiación y Contabilidad*, 32(118), 833-869. <http://www.jstor.org/stable/42781528>
- Breiman, L. (2001). Statistical Modeling: The Two Cultures. *Statistical Science*, 16(3), 199–215. <http://www.jstor.org/stable/2676681>
- Calandro, J. (2007). Considering the utility of Altman's Z-score as a strategic assessment and performance management tool. *Strategy & Leadership*, 35(5), 37-43. <https://doi.org/10.1108/10878570710819206>
- CAUAS, J. (1966). PROGRAMACION LINEAL Y TEORIA ECONOMICA. *Cuadernos de Economía*, 3(8), 99–106. <http://www.jstor.org/stable/41950819>
- Chacón, J., & Rugel, S. (2018). Artículo de revisión. Teorías, modelos y sistemas de gestión de calidad. *Revista espacios*, 39(50).

- Charpentier, J. A. V., Gómez, M. B., & Rojas, J. M. C. (2013). Modelos para la prevención de bancarrotas empresariales utilizados por el sector empresarial costarricense. *Tec Empresarial*, 7(3), 43-49.
- Cheng, B., & Titterington, D. M. (1994). Neural Networks: A Review from a Statistical Perspective. *Statistical Science*, 9(1), 2–30. <http://www.jstor.org/stable/2246275>
- Cox, D. R. (1958). Two further applications of a model for binary regression. *Biometrika*, 45(3/4), 562-565.
- Cox, D. R., & Snell, E. J. (1974). The Choice of Variables in Observational Studies. *Journal of the Royal Statistical Society. Series C (Applied Statistics)*, 23(1), 51–59. <https://doi.org/10.2307/2347053>
- Cramer, J. S. (2003). *Logit models from economics and other fields*. Cambridge University Press
- DE LA OSSA CIODARO, M. F. (2016). *Diseño de modelos financieros* (1st ed.). Universidad del Externado. <https://doi.org/10.2307/j.ctv18msptf>
- de Lema, D. G. P., Pérez, A. A., & Segura, A. C.-F. (1995). UN MODELO DISCRIMINATE PARA EVALUAR EL RIESGO BANCARIO EN LOS CREDITOS A EMPRESAS. *Revista Española de Financiación y Contabilidad*, 24(82), 175–200. <http://www.jstor.org/stable/42780274>
- del Pino, L. A. V. (2018). Modelos de regresión lineal y logística. In *MÉTODOS BIOESTADÍSTICOS.: Segunda edición* (2nd ed., pp. 207–246). Ediciones UC. <http://www.jstor.org/stable/j.ctvvngkg.10>
- Dent, T. H. S., Wright, C. F., Stephan, B. C. M., Brayne, C., & Janssens, A. C. J. W. (2012). Risk Prediction Models: A Framework for Assessment. *Public Health Genomics*, 15(2), 95–105. <https://www.jstor.org/stable/26686603>
- Enguíanos, A. M. (1994). LOS MODELOS DE PREDICCIÓN DEL FRACASO EMPRESARIAL: UNA APLICACIÓN EMPÍRICA DEL LOGIT. *Revista Española de Financiación y Contabilidad*, 23(78), 203–233. <http://www.jstor.org/stable/42781063>
- Espinosa-Paredes, G., & Rodríguez, A. V. (2016). *Aplicaciones de programación no lineal*. OmniaScience.
- Fadlalla, A., & Lin, C.-H. (2001). An Analysis of the Applications of Neural Networks in Finance. *Interfaces*, 31(4), 112–122. <http://www.jstor.org/stable/25062724>
- FICO Poon, M. (2007). Scorecards as devices for consumer credit: The case of Fair, Isaac & Company Incorporated. *The sociological review*, 55(2_suppl), 284-306.

- Grinold, R. C. (1972). Mathematical Programming Methods of Pattern Classification. *Management Science*, 19(3), 272–289. <http://www.jstor.org/stable/2629510>
- Gutierrez Girault, M. A. (2007). Modelos de credit scoring: qué, cómo, cuándo y para qué.
- Hand, D. J., & Henley, W. E. (1997). Statistical classification methods in consumer credit scoring: a review. *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, 160(3), 523-541.
- Hernández, P. A. C. (2004). Aplicación de árboles de decisión en modelos de riesgo crediticio. *Revista Colombiana de estadística*, 27(2), 139-151.
- Hinton, G. E. (1992). How Neural Networks Learn from Experience. *Scientific American*, 267(3), 144–151. <http://www.jstor.org/stable/24939221>
- Høyland, K., & Wallace, S. W. (2001). Generating Scenario Trees for Multistage Decision Problems. *Management Science*, 47(2), 295–307. <http://www.jstor.org/stable/2661576>
- Jarrow, R. A. (2009). Credit Risk Models. *Annual Review of Financial Economics*, 1, 37–68. <http://www.jstor.org/stable/42939933>
- Kingsford, C., & Salzberg, S. L. (2008). What are decision trees?. *Nature biotechnology*, 26(9), 1011-1013. Kingsford, C., & Salzberg, S. L. (2008). What are decision trees?. *Nature biotechnology*, 26(9), 1011-1013.
- Kolar, M., Song, L., Ahmed, A., & Xing, E. P. (2010). ESTIMATING TIME-VARYING NETWORKS. *The Annals of Applied Statistics*, 4(1), 94–123. <http://www.jstor.org/stable/27801581>
- Laffarga Briones, J., Martín Marín, J. L., & Vázquez Cueto, M. J. (1987). Predicción de la crisis bancaria en España: comparación entre el análisis logit y el análisis discriminante. *Cuadernos de Ciencias Económicas y Empresariales*, 18, 49-57.
- Lemon, K. N., & Verhoef, P. C. (2016). Understanding Customer Experience Throughout the Customer Journey. *Journal of Marketing*, 80(6), 69–96. <http://www.jstor.org/stable/44134974>
- Lieberman, E. R. (1991). Soviet Multi-Objective Mathematical Programming Methods: An Overview. *Management Science*, 37(9), 1147–1165. <http://www.jstor.org/stable/2632331>
- Logit models. From economics and other fields) Cramer, J. S. (2003). *Logit models from economics and other fields*. Cambridge University Press
- McFadden, D. (1973). Conditional logit analysis of qualitative choice behavior.

- Medina, R. P. (2000). MODELOS NO PARAMÉTRICOS EN PROBLEMAS DE CLASIFICACIÓN FINANCIERA. *Revista Española de Financiación y Contabilidad*, 29(105), 857–861. <http://www.jstor.org/stable/42782276>
- Molano Ramírez, M. C. (2018). Modelo de riesgo aplicado para el otorgamiento de crédito comercial masivo, para micropymes y pymes colombianas: laboratorio aplicado Banco de Bogotá. No se si borrala es de OHLSON
- Morgan, N. A., & Rego, L. L. (2006). The Value of Different Customer Satisfaction and Loyalty Metrics in Predicting Business Performance. *Marketing Science*, 25(5), 426–439. <http://www.jstor.org/stable/40057034>
- Nagelkerke, N. J. D. (1991). A Note on a General Definition of the Coefficient of Determination. *Biometrika*, 78(3), 691–692. <https://doi.org/10.2307/2337038>
- Ohlson, J. A. (1980). Financial Ratios and the Probabilistic Prediction of Bankruptcy. *Journal of Accounting Research*, 18(1), 109. <https://doi.org/10.2307/2490395>,
- Orgler, Y. E. (1970). A credit scoring model for commercial loans. *Journal of money, Credit and Banking*, 2(4), 435-445
- Rayo Cantón, S., Lara Rubio, J., & Camino Blasco, D. (2010). Un Modelo de Credit Scoring para instituciones de microfinanzas en el marco de Basilea II. *Journal of Economics, Finance and Administrative Science*, 15(28), 89-124.
- Ripley, B. D. (1994). Neural Networks and Related Methods for Classification. *Journal of the Royal Statistical Society. Series B (Methodological)*, 56(3), 409–456. <http://www.jstor.org/stable/2346118>
- Rodríguez, E. M. (2008). Logit Model como modelo de elección discreta: origen y evolución. *Anuario jurídico y económico escurialense*, (41), 469-484.
- Sapatinas, T. (2004). [Review of The Elements of Statistical Learning, by T. Hastie, R. Tibshirani, & J. Friedman]. *Journal of the Royal Statistical Society. Series A (Statistics in Society)*, 167(1), 192–192. <http://www.jstor.org/stable/3559815>
- Shu, X. (2020). CLASSIFICATION AND DECISION TREES. In *Knowledge Discovery in the Social Sciences: A Data Mining Approach* (1st ed., pp. 175–190). University of California Press. <https://doi.org/10.2307/j.ctvw1d683.11>
- Sweeting, P. J. (2007). MODELLING AND MANAGING RISK. *British Actuarial Journal*, 13(3), 579–636. <http://www.jstor.org/stable/41141736>

Wiley, A. L. (1992). Customer Satisfaction Measurement. *Technical Communication*, 39(2), 308–310. <http://www.jstor.org/stable/43090302>