

**Modelo de predicción de quiebra para compañías del sector de restaurante y
hotelería para 2019 y 2020**

Juan Martín Duran Martínez

Diego Alejandro Manosalva Arias

Colegio de Estudios Superiores de Administración CESA

Maestría en Finanzas Corporativas

Bogotá

2022

**Modelo de predicción de quiebra para compañías del sector de restaurante y
hotelería para 2019 y 2020**

Juan Martín Duran Martínez

Diego Alejandro Manosalva Arias

Tutor:

Javier Bernardo Cadena Lozano

Colegio de Estudios Superiores de Administración CESA

Maestría en Finanzas Corporativas

Bogotá

2022

Contenido

Planteamiento Del Problema	6
Pregunta De Investigación	9
Hipótesis	9
Objetivo General	10
Objetivos Específicos	10
Marco Teórico	11
Estado Del Arte	24
Muestra y datos	36
Indicadores evaluados	38
Resultados	46
Regresión por modelo Logit	46
Tablas de confusión	51
Análisis de variables atípicas	52
Regresión por Z-SCORE de Altman	56
Interpretación de resultados	58
Análisis de indicadores	61
Conclusión	65
Trabajos citados	67

Listado de tablas

Tabla 1 Lista de indicadores evaluados en el modelo	39
Tabla 2 Lista de indicadores depurados en el modelo.	41
Tabla 3 Número de observaciones por base	47
Tabla 4 Resultado base desbalanceada	48
Tabla 5 Prueba F	49
Tabla 6 SMOTE Dummy resultado	50
Tabla 7 Prueba de Hosmer – Lemeshow	51
Tabla 8 Tablas de confusión	52
Tabla 9 Estadística descriptiva	53
Tabla 10 Estadística descriptiva comparativa por percentil	53
Tabla 11 Proporción de observaciones	54
Tabla 12 Resultado regresión comparativa	55
Tabla 13 Prueba de Hosmer – Lemeshow comparativa	55
Tabla 14 Tablas de confusión comparativa	56
Tabla 15 Resultado de aciertos Altman y sensibilidad	57
Tabla 16 Resultados de predicción por tipo de base	59
Tabla 17 Resultados de predicción por autor bases balanceadas	59
Tabla 18 Resultados de predicción por autor	60
Tabla 19 Resultados de indicadores por estudio y base	61

Listado de Figuras

Figura 1 Descripción de métodos de remuestreo	22
Figura 2 SMOTE ejemplo de Oversampling	23
Figura 3 Gráficos de dispersión con cuatro indicadores de gestión de liquidez.	42
Figura 4 Gráficos de dispersión con cuatro indicadores de gestión de rentabilidad.	44
Figura 5 Gráficos de dispersión con cuatro indicadores de gestión de endeudamiento.	45

Planteamiento Del Problema

El sector industrial de Lodging (Hostelería¹) tiene un importante rol en la economía mundial, en el año 2018 aportó el 3,4% del PIB mundial manteniendo una tendencia desde el año 2010 (United Nations World Tourism Organization, 2019). Colombia, Ecuador y Perú son los países líderes en Suramérica teniendo un aporte representativo por el sector de turismo en su PIB.

A nivel local el sector turismo contribuye de manera importante a la economía, ya que su aporte al PIB mantuvo tasas por encima del 3% desde el 2005 hasta alcanzar tasas del 4,4% y 3.3% para 2019 y 2020 respectivamente. (DANE , 2020). En relación con la generación de empleo, su participación ha llegado a ser superior al 8% de acuerdo con el boletín de empleo del Centro de Pensamiento Turístico, una división de Cotelco² (Asociación Hotelera y Turística de Colombia; Fundación Universitaria CAFAM; Alcaldía Mayor de Bogotá, 2018).

Para estudiar la empleabilidad del sector, la medición considera los contratos suscritos en las entidades vinculadas formalmente a los sectores gremiales que las agrupan (Asociación Hotelera y Turística de Colombia; Fundación Universitaria CAFAM; Alcaldía Mayor de Bogotá, 2018); donde no es considerado la informalidad que giran en torno al turismo, la cual alcanza un 53% de las personas ocupadas legalmente (Leguizamón, 2016).

¹ La hostelería abarca aquellas actividades económicas relacionadas al servicio de alojamiento y alimentación esporádico. (Agency, 2021)

² Asociación hotelera y turística de Colombia – Cotelco.

Siguiendo la investigación realizada por Leguizamón, respecto a la informalidad en el sector turismo, el Ministerio CIT y el DANE, a través de la Gran Encuesta Integrada de Hogares, exponen que sector desde el año 2007 hasta 2018, pasaron de tener una informalidad del 58,2% a tener un 53,9%.

El turismo cuenta con una serie de subsectores dentro de los que se resaltan el de alojamiento (hoteles) y servicio de comidas-bebidas (restaurantes), los cuales son analizados en conjunto al ser vinculantes mutuamente, lo anterior debido a que cuando el consumidor esta fuera del hogar requerirá ambos servicios. Ambos subsectores de acuerdo al CIIU³ se clasifican en dos grupos: 55 para alojamiento y 56 para servicios de comida y bebidas, representando el 46,6% del total de ingreso per cápita percibido en el sector turismo para 2020:

Para el total 24 ciudades y áreas metropolitanas, el gasto promedio per cápita día de las personas que realizaron turismo interno fue \$78.559 pesos. El rubro en el que más gastaron los viajeros fue: alimentos y bebidas destinando \$21.851 diarios, seguido por alojamiento con \$14.761 y transporte terrestre (hacia y desde el destino) \$14.498 (DANE, 2021).

Respecto a la inversión extranjera directa⁴, el sector del turismo es uno de los mayores receptores en Colombia, de acuerdo con las cifras del Ministerio de Comercio Industria y

³ Clasificación internacional industrial uniforme (Cámara de comercio de Bogotá, 2021)

⁴ *“La inversión directa son los aportes de capital que cumplen con los siguientes dos requisitos: i) Existe una relación accionaria entre el inversionista y la empresa que reside en una economía distinta, ii) El inversionista tiene una influencia significativa en la toma de decisiones de la empresa”* (Banco de la República, 2021).

Turismo presentó desde 1995 un promedio del 9,8% del total de la Inversión directa que hubo en el país, presentando picos en el 2019 con un 18.4%. (DANE, 2021)

Para el año 2020, este sector fue uno de los más afectados a raíz de la emergencia sanitaria causada por el covid-19 y las restricciones en movilidad declaradas por los diferentes países con el fin de disminuir la posibilidad de contagio del virus. A nivel mundial, las pérdidas generadas al sector del turismo pueden llegar a totalizar 730 Billones de dólares americanos (United Nations World Tourism Organization, 2020). Para Colombia, la Asociación Colombiana de agencias de viajes y turismo, reportó a través de sus afiliados una disminución en el ingreso del 90% durante el primer semestre del 2020 (ANATO, 2020).

Puntualmente en los subsectores a analizar, para el 2020 se observó un incremento de empresas en proceso de reestructuración y liquidación del 172% (Superintendencia de Sociedad , 2021) pasando de 18 empresas en procesos de reestructuración a 49.

Según el DANE, para 2020 el sector donde se incluyen las actividades de alojamiento y restaurantes tuvieron un decrecimiento anual (2019 al 2020) del -15.1%, siendo el tercero más afectado luego de Mineras y Construcción. (DANE, 2021). En adición a lo anterior, este sector dejó de emplear alrededor de 266.000 personas, pasando de tener 1.655.000 personas empleadas en el 2019 a tener 1.389.000. (DANE, 2021).

En Colombia dadas las condiciones con las que cuenta el país respecto a geografía, variedad de cultura, gastronomía y alojamiento el sector de turismo es uno de los motores más

importantes de la economía nacional, por medio de la generación de empleo e inversión directa, convirtiéndolo en un sector primordial para el crecimiento del país; sin embargo hasta la fecha no se cuenta con un análisis financiero bajo la perspectiva de modelos de quiebra, que permita identificar practicas financieras y diseñar una hoja de ruta para un futuro estable y de constante crecimiento como la ha sido en los últimos años.

Pregunta De Investigación

¿Cuáles indicadores financieros y bajo que modelo estadístico se puede establecer la probabilidad de quiebra para las empresas de los subsectores de alojamiento (hoteles) y servicio de comidas-bebidas en Colombia, a partir de la emergencia económica del covid-19?

Hipótesis

La gestión de indicadores financieros de rentabilidad y liquidez son los que permiten tener mejores prácticas financieras prediciendo el potencial de quiebra para empresas del subsector de hotelería y restaurante garantizando la supervivencia de estas.

Objetivo General

Identificar cuáles son las variables e indicadores que mejor predicen la quiebra para empresas del sector de hotelería y restaurantes en Colombia a raíz de la emergencia sanitaria del COVID-19

Objetivos Específicos

- Realizar un análisis bibliográfico del concepto de quiebra, modelo de quiebra e indicadores financieros vinculados al sector.
- Seleccionar indicadores financieros relevantes para los diferentes modelos y para el sector a investigar.
- Establecer el modelo de quiebra para la investigación a través de una revisión bibliográfica.
- Analizar de los métodos estadísticos de predicción de quiebra cual es el más acertado para identificar el punto de quiebra.
- Contextualizar el subsector de hotelería - restaurantes en Colombia
- Revisar las causales de quiebra y supervivencia en el sector a analizar durante el año 2020 bajo la emergencia sanitaria causada por el covid-19
- Identificar oportunidades de mejora en la toma de decisiones por parte de los administradores de las compañías de estos subsectores sobre los indicadores de quiebra, así como su estructura financiera.

Marco Teórico

El estudio de los métodos predictivos de quiebra pasa desde la definición misma del término hasta el desglose y desarrollo vinculado a los diferentes métodos utilizados para sus estudios (Z-Score de Altman, Logit, etc); sin embargo, la base de estos considera la información contable de las organizaciones, que corresponden al reflejo de las estructuras de capital definidas por los administradores dentro de las compañías.

A lo largo de la investigación bibliográfica, los autores han usado diversos términos para determinar si una empresa está en quiebra o no; se encontraron conceptos como frágiles, bancarota, débiles, como sinónimos de empresas en quiebra y fuertes para empresas que no se encuentran quebradas.

La teoría de estructura de capital de acuerdo con (Rivera, 2002) busca definir e identificar el problema sobre la financiación para las empresas. El inicio de estas investigaciones fue desarrollado por parte de (Modigliani & Miller, 1958) quienes definieron, bajo el escenario de un mercado perfecto, la toma de decisiones de índole económico y financiero realizada, por parte de los administradores, no consideran la estructura de capital de una compañía, sin embargo, tal y como lo menciona (Aybar Arias, Casino Martínez, & Lopez García, 2003), *“esta idea es muy alejada de la realidad...e ignora las dificultades que puede poseer cualquier tipo de compañía”*.

Basado en esta posición, (Modiglian & Miller, 1963) realizan una corrección a su teoría inicial incluyendo la premisa de obtención de ventajas fiscales en la consecución y estructuración del capital de la compañía, a partir de un endeudamiento máximo por parte de estas.

La posición discutida por Modigliani, ha traído consigo un desarrollo académico del cual se resaltan las teorías del trade-off y pecking-order. La posición del óptimo financiero (o trade-off como su nombre en inglés), considera que existe una combinación igual de una estructura de capital entre los costos adquiridos en el endeudamiento y otros (relacionados a costos de quiebra, escudos fiscales y costos de agencia) que afianza la posición o no de adquirir deuda por parte del empresario. (Bradley, Tarrel, & Kim, 1984)

La teoría que contrapone el óptimo financiamiento está asociada a la jerarquización financiera (pecking-order) desarrollada por (Myers & Majluf, 1984), la cual concluye que la búsqueda de financiación genera la utilización de recursos propios de la compañía antes de acceder a mercados financieros, considerando la ventaja de información obtenida por parte de los administradores ante el mercado.

Si bien cómo se puede asociar en los marcos explicados anteriormente, la prevalencia de la deuda y sus costos relacionados para la toma de decisiones por parte de los administradores poseen una mayor prevalencia en la estructura de capital de una compañía, existen otros factores que la afectan y que tienen impacto en la obtención de recursos, tal como lo define (Mejía, 2013):

Las firmas con mayor nivel de activos fijos, mayor tamaño, mayor inversión en investigación y desarrollo... y trayectoria tienen mayor financiación que aquellas con las condiciones opuestas... las empresas con mayores niveles de rentabilidad tienen menores niveles de endeudamiento, circunstancia que puede involucrar una preferencia por la financiación con recursos propios.

Adicional al enfoque clásico de estructura de capital, existe otro que basa su comportamiento a partir de las estrategias corporativas que adopten las gerencias de cualquier tipo de compañía (Aybar Arias, Casino Martínez, & Lopez García, 2003, págs. 31-32). Dentro de este aspecto, se encontró que las condiciones relacionadas a una posición dominante del mercado, cadena productiva, así como el control ejercido por parte de la alta gerencia, proporciona una perspectiva que altera la capacidad productiva trayendo como consecuencias una alteración dentro de la estructura de capital definida. (Barton & Gordon, 1988).

Sobre las anteriores bases se pueden identificar algunas características básicas que afectan la estructura de capital y por ende son indicios en la posible búsqueda de riesgos de quiebra sobre empresas:

- La capacidad de aprovechamiento de estructuras fiscales a partir del financiamiento externo (Modiglian & Miller, 1963).
- La capacidad decisoria de la alta gerencia basada en métodos que priman la generación de recursos propios para su financiamiento (Myers & Majluf, 1984).

- Factores ajenos a la estructura de capital tradicional soportado en capacidad productiva y de mercado (Barton & Gordon, 1988).
- Condiciones de tamaño, capacidad de generación de activos fijos, rentabilidad de la sociedad (Mejía, 2013).

Para enmarcar los conceptos en los cuales se establecerán el método de predicción de quiebra, se deberá realizar una contextualización sobre la quiebra desde un enfoque normativo, expuesto en la Ley 1116 del 2006⁵ y complementarlo con los conceptos académicos que surgen alrededor de este término de quiebra.

Desde el enfoque normativo, se ha concebido la quiebra bajo el termino de insolvencia, esta puede presentarse de manera temporal, donde el ente empresarial reduce su capacidad de pago (o esta deja de existir) frente a sus diferentes acreedores. (Brachfield, 2021):

Una situación de iliquidez puede durar un par de semanas, de modo que el deudor podrá pasado este bache seguir atendiendo puntualmente sus obligaciones de pago. Sin embargo, puede tratarse de una situación crónica de iliquidez, provocada por la estructura financiera de la compañía, por lo que este tipo de clientes siempre tendrán dificultades en cumplir con sus pagos.

⁵ Ley 1116 del año 2006, Congreso de la República de Colombia.

(Brachfield, 2021) , propone entender los diferentes tipos de insolvencia que hay desde la óptica de un tercero conciliador entre las partes (acreedores y deudor), esto para moldear mejor la situación de cada ente económico. Los tipos son: i) Insolvencia sobrevenida: Aquella cuando una empresa por motivos internos y/o externos deja de tener liquidez; ii) Insolvencia fortuita: Aquella que está por encima de la voluntad del deudor dado a que se genera por factores no controlables e imprevistos; iii) Insolvencias transitorias: Falta de liquidez temporal de la empresa, puede abarcar semanas y hasta meses; iv) Insolvencia definitiva: Es un caso donde los pasivos excedieron los activos; v) Insolvencia de origen: Se da cuando los acreedores conceden créditos, sabiendo que el deudor se encuentra en una posición de liquidez delicada; vi) Insolvencias intencionales: Actos dolosos donde gerentes y/o propietarios vuelven a la empresa ilíquida para no hacer frente a sus deudas.

Con relación a la respuesta para identificar la quiebra desde el ámbito académico, el estudio de los modelos de quiebra posee sus bases en dos clases principales: Los métodos univariados y multivariados. El primer tipo de métodos consideran la evaluación de indicadores individuales comparados contra compañías consideradas como “fracasadas” versus aquellas compañías con éxito. Este método fue presentando en el año 1930 donde El Buró de Investigaciones de Negocios (BRR como su acrónimo en inglés) publicó un estudio de los ratios sobre aquellas compañías quebradas, resaltando que el capital de trabajo sobre el total de activos era un buen predictor de compañías con indicios de quiebras (Bellovary & Arkers, 2007). Entiéndase que el modelo univariable, solamente toma un indicador financiero para tratar de predecir la potencial quiebra de una empresa.

Dentro de los otros estudios bajo el modelo univariado, (FriskPatrick, P, 1932 citado en (Bellovary & Arkers, 2007) quien identificó que los indicadores financieros relacionados al patrimonio neto versus la deuda y utilidades netas versus patrimonio neto brindan más información sobre las compañías que son consideradas como quebradas.

Adicionalmente (Merwin, 1942) utiliza en su estudio la clasificación de cinco tipos de industrias dividiéndolas entre sanas y en bancarrota, centrando su rango de tiempo de estudio para el periodo de nueve años, comprendido entre 1926 y 1935, generando como resultado el uso de otros indicadores, tales como el capital neto sobre el total de activos, ratio de liquidez y patrimonio total sobre total de deuda.

Uno de los estudios que se puede considerar el más completo, respecto a las técnicas univariadas, es el de (Beaver, 1966), quien clasificó en seis grupos de 30 indicadores a las compañías seleccionadas a fin de determinar las características de una empresa fallida o no, generando como resultado de su evaluación que los indicadores relacionados a flujo de caja y deuda son los destacados para detectar las quiebras en compañías.

Como se puede observar dentro de los estudios citados, reinciden los mismos indicadores en los modelos univariados que están relacionados al indicador de deuda, efectivo y patrimonio neto.

El segundo tipo de modelo son los basados bajo el método estadístico de análisis multivariantes, dentro de este grupo podemos encontrar: análisis de multivariantes

discriminantes (MDA como su acrónimo en inglés), análisis Logit, análisis Probit y redes neuronales. (Bellovary & Arkers, 2007).

El estudio que dio el inicio al análisis con multivariantes es el Z-Score, elaborado por Altman en 1968. La novedad de este estudio y la elaboración de un modelo con múltiples variables, construido sobre aquellas que se consideran como relevantes de la situación financiera de la empresa.

El Z-Score de Altman es un modelo de predicción de insolvencia que se realiza con base en un análisis estadístico iterativo de discriminación múltiple, en el que se ponderan y suman cinco razones de medición para clasificar las empresas solventes o insolventes. (Grisales & Ospina, 2016)

El modelo de Altman ha tenido variaciones desde su publicación en 1968, debido a que sus factores varían de acuerdo con el sector, tamaño e incluso país de la muestra. Ampliando sobre la misma línea, Altman, resalta que hay tres Z-Score clásicos. El primer Z-Score (Z) conocido con cinco factores para ser aplicado en empresas públicas manufactureras, este fue el elaborado en 1968. Para el año 1983, Altman expande teóricamente su Z-Score para empresas no públicas (Z'), sustituyendo uno de los factores. Por último, en el mismo año, 1983, amplió su modelo para empresas tanto privadas como públicas reduciendo su fórmula a cuatro factores, teniendo además un Z-Score sensible al sector al que se vaya a usar, este sería Z-Score (Z''). (Altman, Iwanicz-Drozdowska, Laitinen, & Suvas, 2017)

A continuación, las tres fórmulas expuestas:

$$Z = 0.012 * \left(\frac{\text{Capital de trabajo}}{\text{Total de activos}} \right) + 0.014 * \left(\frac{\text{Utilidades retenidas}}{\text{Total activos}} \right) + 0.033 \\ * \left(\frac{\text{Utilidad operacional}}{\text{Total de activos}} \right) + 0.006 * \left(\frac{\text{Valor en el mercado del patrimonio}}{\text{Valor en libros de todos los pasivos}} \right) \\ + 0.999 * \left(\frac{\text{Ventas}}{\text{Total de activos}} \right).$$

$$Z' = 0.717 * \left(\frac{\text{Capital de trabajo}}{\text{Total de activos}} \right) + 0.847 * \left(\frac{\text{Utilidades retenidas}}{\text{Total activos}} \right) + 3.107 \\ * \left(\frac{\text{Utilidad operacional}}{\text{Total de activos}} \right) + 0.420 * \left(\frac{\text{Valor en el mercado del patrimonio}}{\text{Valor en libros de todos los pasivos}} \right) \\ + 0.998 * \left(\frac{\text{Ventas}}{\text{Total de activos}} \right).$$

$$Z'' = 3.25 + 6.56 * \left(\frac{\text{Capital de trabajo}}{\text{Total de activos}} \right) + 3.26 * \left(\frac{\text{Utilidades retenidas}}{\text{Total activos}} \right) + 6.72 \\ * \left(\frac{\text{Utilidad operacional}}{\text{Total de activos}} \right) + 1.05 * \left(\frac{\text{Valor en el mercado del patrimonio}}{\text{Valor en libros de todos los pasivos}} \right).$$

De acuerdo con Altman (Altman, Iwanicz-Drozowska, Laitinen, & Suvas, 2017), es el Z'' el que mayor uso tiene en investigaciones académicas dada su capacidad amplia de análisis y de abordar múltiples sectores y tipos de empresas.

Otro modelo utilizado por parte de los diferentes autores en la predicción de quiebra es el de regresión logística o Logit, dado que otorga una mayor flexibilidad en el uso de los supuestos estadísticos, respecto a los estudios desarrollados bajo análisis discriminantes; así mismo

condiciona para su evaluación la utilización de variables dependientes e independientes.
(Mossman, Bell, Swartz, & Turtle, 1998).

El desarrollo del modelo logístico clásico, mencionado por (Caro N. , 2016) y (Caro, Díaz, & Porporato, 2013) supone una variable de respuesta binaria que indica el grupo al que pertenece la variable observada, donde la variable respuesta asume un valor de 1 y su contraparte 0. Los problemas bajo selección dicotómica son modelados bajo la descripción que la variable respuesta y_i asume un valor de 1, suponiendo una covariable x_i cuyo predictor lineal es el siguiente:

$$\Pr\left(y_i = \frac{1}{x_i}\right) = h(\beta_1 + \beta_2 x_i) \quad \text{donde } i = (1, 2, \dots, n)$$

Donde:

x_i : variable dependiente

y_i : variable independiente

β : coeficientes de la regresión

n : cantidad de observaciones

La transformación del modelo a su función Logit es la siguiente:

$$\log it\{\Pr(y_i = 1/x_i)\} = \ln \left\{ \frac{\Pr(y_i = 1/x_i)}{1 - \Pr(y_i = 1/x_i)} \right\} = z_j.$$

Tal como lo describe (Caro, Díaz, & Porporato, 2013): “...el principal problema del modelo logístico clásico supone respuestas independientes dadas las covariables, resultando apropiado cuando los datos no tienen ningún tipo de agrupamiento”. (pág. 209)

Para la solución de este problema, se plantea el modelo logístico mixto en el que se distinguen efectos fijos y aleatorios, donde “Los fijos tienen como finalidad comparar los resultados de la variable dependiente para los distintos valores de las independientes, mientras que los efectos aleatorios analizan la variabilidad no explicada por la respuesta” (Caro N. , 2016, pág. 136)

La solución al anterior planteamiento se genera por la siguiente fórmula:

$$Y_{ij} = X_{ij}\beta + Z_{ij}b_j + \varepsilon_{ij}$$

Donde:

Y_{ij} : Efectos fijos

β_j : Parámetro que acompaña a cada variable observada X_{ij}

b_j : Efectos aleatorios

Z_{ij} : Matriz de diseño en función de que variable X_{ij} ha sido incluida como efecto aleatorio

ε_{ij} : Término de error.

Los anteriores términos poseen una distribución normal con una estructura de varianza (G) explicada por la siguiente expresión:

$$b_j \sim N(0, G)$$

De acuerdo con (Perez, Lopera, & Vásquez, 2019) los modelos logísticos cuando presentan más de dos categorías “la selección entre dos de ellas es independiente de las demás, lo cual significaría que los individuos tienen funciones de utilidad homogéneas” (pág. 10). Para tal fin se plantea un modelo logístico heterocedástico donde se asume:

$$var(\varepsilon_i) = [\exp(Z_i'Y)]^2$$

Donde Z_i' contiene variables que generan el problema de heterocedasticidad que puede encontrarse dentro de las variables X'_{ij} y Y_{ij} convirtiéndose en un vector de parámetros.

Otro problema relacionado con el modelo se encuentra vinculado al desarrollo de variables atípicas, por lo cual se puede aplicar un modelo logístico robusto. Es de resaltar que Y_{ij} solo puede tomar valores entre cero y uno, pero como lo menciona (Perez, Lopera, & Vásquez, 2019): “posible hablar de valores atípicos en las variables X'_{ij} que pueden hacer que Y_{ij} tome el valor cero cuando debía ser uno o viceversa.” (pág. 10). Para solucionar esto, el modelo logístico robusto, en su función de verosimilitud penaliza con cero y uno a las observaciones que presentan valores extremos en una o más variables explicativas, donde aquellas observaciones en X'_{ij} toman ponderaciones con valor cero cerca a la verosimilitud.

Dentro de la evaluación de datos a partir de la información tanto para modelos binarios como modelos MDA, pueden existir problemas de desbalanceo, que se consideran cuando cumplen algunos de los criterios descritos por (Gruszczyński, 2018), donde “n” es el número de muestras incluidas en el modelo:

1. Si la proporción de n_1 y n_2 en la muestra es diferente a la proporción 50:50, o
2. Si la proporción $p_1 = n_1 / n$ y $p_2 = n_2 / n$ son diferentes a las fracciones de quiebra y empresas en no condiciones de quiebra, entonces se considerará la muestra desbalanceada en términos de p_1 y p_2 .

Para solucionar el anterior problema, se han generado modelos que permitan el rebalanceo, denominado Oversampling y Undersampling, donde el primero corresponde al incremento del número de observaciones de la clase inferior con el fin de equilibrar la superior y el Undersampling, es generado a partir de remover observaciones de la clase superior con el fin de equilibrar la inferior (Liu, Ghosh, & Martin, 2007), tal como se detalla en la figura 1:

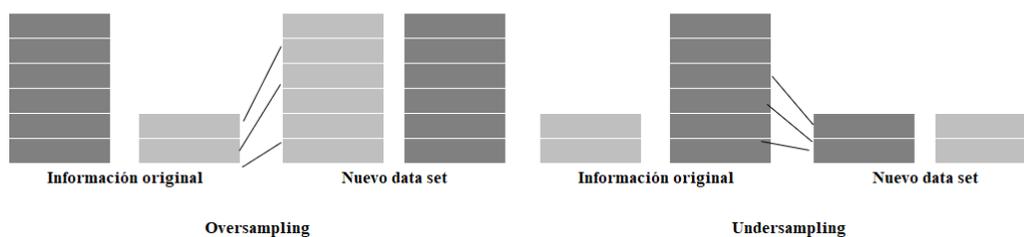


Figura 1 Descripción de métodos de remuestreo

Construcción propia a partir de (Mohammed, Rawashdeh, & Abdullah, 2020)

Los principales tipos de remuestreo utilizados tal como los menciona (Hernandez, Carrasco-Ochoa, & Martínez-Trinidad, 2013), son los siguientes:

1. Resample: Este método de Oversampling produce una clase uniforme de distribución, generado por medio de un Subsampling aleatorio a la clase mayoritaria, con reemplazo en la clase minoritaria.
2. Spread Subsampling: Esta técnica de Undersampling produce una submuestra de una base de datos. La distribución de clases se ajusta mediante una eliminación de observaciones de la clase mayoritaria.
3. Synthetic Minority Over-sampling Technique (SMOTE): Método de Oversampling que consiste en sintetizar elementos para la clase minoritaria, a partir de los que ya existen. Funciona seleccionando aleatoriamente un punto de la clase minoritaria y calculando los k vecinos más cercanos para este punto. Los puntos sintéticos se suman entre el punto elegido y sus vecinos. (Chawla, Bowyer, Hall, & Kegelmeyer, 2002). La representación gráfica de cómo opera SMOTE es el siguiente:

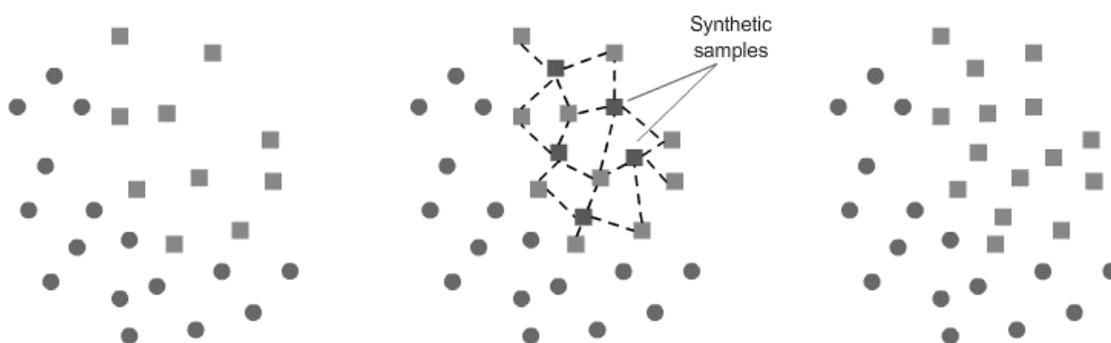


Figura 2 SMOTE ejemplo de Oversampling

(Alencar, 2018)

Estado Del Arte

Los estudios vinculados a modelos de predicción de quiebra para el sector hostelería, no se ha visto un análisis enfocado en la economía colombiana. Sin embargo, a continuación, se exponen algunos estudios desarrollados en el país sobre detección de quiebra:

(Rosillo, 2002) realizó uno de los primeros estudios enfocados al modelo de predicción de quiebra en Colombia, utilizando como base un modelo de análisis discriminante - MDA, donde evaluó la información de 106 empresas de diferentes sectores de los últimos 4 años, de las cuales 26 compañías fueron utilizadas para la validación del modelo propuesto y las 80 restantes para el diseño de este modelo, siendo clasificadas en dos categorías: Fuertes y débiles, a partir de la validación de 3 indicadores: prueba ácida, endeudamiento y rentabilidad de activo, donde aquellas compañías con la combinación de indicadores más bajos eran consideradas débiles. Adicionalmente, para la ejecución del modelo fueron evaluados 12 indicadores relacionados a liquidez, rentabilidad y apalancamiento. El resultado de este estudio obtuvo un 94% de acierto sobre las empresas consideradas como fuertes y 87% para las débiles, donde las compañías fuertes tienen un menor nivel de endeudamiento, mayor rentabilidad y mejor aspecto de cobertura deuda - patrimonio.

(Berrio & Cabeza de Vergara, 2003), desarrollaron un modelo de predicción de quiebra, basados en el estudio propuesto por Altman, para una muestra de 622 empresas de diversos sectores con información financiera de 1998. Para el manejo de la muestra se realizó una clasificación de las compañías en tres sectores: comercial, industrial y de servicios. Los

resultados de esta investigación presentaron un modelo donde se identificaron indicadores relevantes para la detección de quiebras en los sectores evaluados:

- Sector industrial: $\text{ventas} / \text{capital de trabajo neto}$, $\text{costos de ventas} / \text{inventario promedio}$, $\text{EBIT} / \text{ventas}$, $\text{deuda largo plazo} / (\text{deuda largo plazo} + \text{patrimonio})$ y $\text{utilidad retenida} / \text{activo total}$.
- Sector comercial y de servicios: $\text{EBITDA} / \text{total activos}$, $\text{patrimonio} / \text{deuda total}$, $\text{deuda largo plazo} / (\text{deuda largo plazo} + \text{patrimonio})$, $(\text{Activo} - \text{pasivos corrientes}) / \text{ventas y pasivo total} / \text{activo total}$.

(Romero, 2013), efectuó un estudio para la detección de quiebra en compañías PYMES basado en un modelo logístico. La muestra consistió en 2.622 empresas entre 2005 al 2011, catalogadas como fracasadas (aquellas que presentaron condiciones de liquidación judicial, concordato, reestructuración y reorganización) y no fracasadas, reportadas ante la Superintendencia de Sociedades. Así mismo fueron divididas en cinco sectores económicos: servicios, industrial, comercio, construcción y agropecuario. Para la validación del modelo utilizó un total de 40 indicadores financieros relacionados con liquidez y solvencia, relativas al flujo de efectivo, rotación, apalancamiento y aspectos estructurales de los estados financieros. El resultado de este estudio identificó que las empresas en estado de quiebra detectada presentan debilidad en indicadores de $\text{pasivo no corriente} / \text{activo total}$, $\text{pasivo total} (\text{activo corriente} - \text{inventario}) / \text{activo total}$, adicionalmente que las compañías con procesos de reestructuración con una longevidad de 22 a 30 años poseen una mayor probabilidad de fracaso.

(Pérez, González, & Lopera, 2013) desarrollaron un estudio sobre la predicción de quiebra sobre una muestra de 28.387 compañías que reportaron ante la Superintendencia de Sociedades y Financiera para el 2011 (antes denominada de Valores), clasificadas en dos categorías: frágiles y no frágiles, donde las primeras son aquellas que poseen un estado societario de: “reorganización, concordato, reestructuración y liquidación” (pág. 215). Para el estudio fue utilizado un modelo Logit y Probit, construido a partir de un total de 9 indicadores financieros relacionados con endeudamiento, liquidez, rentabilidad, apalancamiento y como variable dummy para el modelo logístico, se relacionaron los sectores económicos a los que pertenecían las empresas evaluadas, correspondientes a un total de 10. El resultado demostró que el modelo Logit, presentó una mejor predicción de fragilidad al alcanzar un 77.41% para las compañías clasificadas en esta categoría y 75.80% para el Probit, adicionalmente, los sectores más propensos a generar una quiebra fueron agricultura, ganadería y caza, así como el de industrias manufactureras, y los indicadores relacionados con rentabilidad del activo, rotación del activo y nivel de endeudamiento son los que predicen de mejor forma la quiebra.

(Arroyave, 2018) realizó un estudio sobre la predicción de quiebra corporativa en Colombia, a partir de la información de tres compañías, dos que se declararon en condiciones de quiebra (Electricaribe y Frontera Energy) y una compañía exitosa (Isagen), para periodo comprendido entre 2008 al 2015. Para tal fin utilizó modelos basados en el Z-Score de Altman, modelo de dos funciones de Korol, modelo P2 de Prusak así como un modelo logístico y árbol de decisión. Para la construcción de estos modelos fueron utilizados un total de diez indicadores financieros probados de manera anual. Como resultado del modelo, el autor indicó que para la

identificación de las quiebras, de las empresas colombianas, el modelo logístico y Korol son efectivos.

(Perez, Lopera, & Vásquez, 2019) desarrollaron un estudio sobre predicción de fragilidad a partir de variaciones del modelo logístico (como lo es el mixto, heterocedástico y robusto) de las empresas colombianas, con una muestra de 22.205 sociedades para el 2015, tomando como fuente la Superintendencia de Sociedades. En este estudio se realizó una clasificación de las empresas en dos categorías: frágiles y no frágiles, siendo las frágiles aquellas que se encuentran en estado de “reorganización, concordato, reestructuración o liquidación” (pág. 8). Para la modelación utilizaron un total de 11 indicadores financieros abarcando apalancamiento, rentabilidad y márgenes de utilidad. Los resultados de este estudio demostraron que los indicadores asociados a endeudamiento y rentabilidad del patrimonio son los más relevantes en las compañías clasificadas en condición de fragilidad y el modelo logístico mixto es óptimo en la determinación de la quiebra con un porcentaje promedio del 84.61%.

Respecto a investigaciones que hayan sido desarrolladas tomando el sector de hoteles y restaurantes, (Olsen, Bellas, & Ventrice, 1983) realizaron uno de los primeros estudios vinculados al sector, cuyo objetivo se enfocaba en detectar los momentos en que las compañías se declaraban en quiebra a partir de la revisión sobre la evolución de las cifras reportadas para un periodo de 24 meses (1981 – 1982), para ello compararon indicadores de liquidez, deuda de largo plazo, rentabilidad y rotación de activos sobre un listado de 19 compañías de las cuales siete se declararon en quiebra y 12 fueron exitosas. El resultado de este estudio encontró que los

indicadores asociados a activos corrientes / pasivos corrientes, capital de trabajo / activos totales, EBIT / total de activos, EBIT / ingresos, capital de trabajo / ingresos, predicen mejor la quiebra.

(Kwansa & Parsa, 1990), realizaron un estudio de eventos para la determinación de quiebra de 22 empresas en Estados Unidos en restaurantes para el periodo de 1970 a 1984, de los cuales el 50% se declararon en quiebra bajo ley norteamericana y el restante fueron exitosos. El resultado detectó que la posibilidad de impago de préstamos, reestructuración de deudas y pérdidas netas son las principales causales para la quiebra en este tipo de compañías.

(Gu & Gao, 2000) desarrollaron un estudio de predicción de quiebra sobre 28 compañías estadounidenses vinculadas al sector de hotelería y turismo para el periodo de 1987 a 1996, de las cuales 14 se declararon en quiebra y 14 como exitosas, utilizando un modelo de análisis discriminante (MDA) que toma como base 14 indicadores financieros: activos corrientes / pasivos corrientes, capital de trabajo / activo total, EBIT / pasivos corrientes, total de pasivos / total de activos, patrimonio total / pasivo total, pasivos de largo plazo / activos totales, utilidades retenidas / total de activos, EBIT / total de pasivos, EBIT / total de activos, margen bruto, margen neto, EBIT / pasivo más patrimonio de largo plazo, ventas sobre total de activos, ventas sobre activos fijos. Este estudio obtuvo un porcentaje de predicción sobre las compañías consideradas en quiebra equivalente al 93%, donde aquellas que tuvieron un nivel alto de deudas, así como pasivos de corto plazo están sujetas a declararse en quiebra durante el primer año.

(Gu Z. , 2002) realizó un estudio bajo el modelo de análisis discriminante (MDA), para la determinación de quiebra de compañías del sector de restaurantes, considerando el modelo Z-

Score desarrollado por Altman. Para este estudio la base fue construida a partir de 36 compañías listadas en la bolsa de valores de Nueva York (18 sanas y 18 en quiebra) para un periodo de 13 años comprendido entre 1986 a 1998. Igual al estudio efectuado por (Gu & Gao, 2000), el investigador consideró los mismos indicadores financieros. Del resultado se identificó un porcentaje de éxito en la predicción del 92% para el primer año, donde las condiciones asociadas a un alto nivel de deuda y un bajo EBITDA predicen aquellas compañías que tienen una alta probabilidad de caer en quiebra.

(Hyunjoon & Gu, 2006), realizaron un estudio de detección de quiebra para compañías del sector de restaurantes mediante la comparación de un modelo de análisis discriminante y un modelo logístico. Para ello, tomaron la muestra utilizada en 2002 por (Gu Z.) Los indicadores utilizados correspondieron a: liquidez, prueba ácida, EBIT / pasivos corrientes, pasivos totales / activos totales, utilidad neta / activos totales, margen bruto, margen neto, ventas sobre total de activos, ventas sobre activos fijos, EBIT sobre el total de activos, EBIT sobre el total de pasivos. El resultado del estudio generó una efectividad del 94% en el proceso de predicción del modelo logístico, respecto al modelo discriminante que alcanzó un 93%, sobre el cual las compañías que presentan un bajo EBIT, así como un alto nivel de pasivos, tienen mayor probabilidad de quiebra. Para reducir la quiebra, los autores sugieren manejar una estricta política de financiación, así como un mejor control de costos operativos.

(Gu & Kim, 2006) desarrollaron un estudio de detección de quiebra para el sector de hoteles, casinos y restaurantes a partir de la información de 16 compañías norteamericanas en el periodo de 1999 al 2004, mediante el uso de modelos logísticos en un periodo de evaluación de

dos años, utilizando 13 indicadores relacionados a liquidez, solvencia, rentabilidad y eficiencia operativa, este estudio considero una variable dummy para los años de mayor decrecimiento económico (2001 – 2002). Este modelo determinó un porcentaje de eficacia en la predicción para el primer año evaluado del 91% y el segundo del 84%, donde los indicadores asociados a solvencia (flujo de caja vinculado a deuda) y rentabilidad se encuentran vinculados a la determinación de la quiebra.

(Huo, 2006) realizó un estudio sobre PYMES en Estados Unidos relacionadas con el sector de restaurantes que se declararon en quiebra para el periodo de 1993 al 2003, para dicho estudio, fueron utilizados modelos de análisis discriminante basados en los descritos por Altman para industrias no manufactureras. Para la verificación de dichos modelos el autor consideró los siguientes indicadores financieros: capital de trabajo /total de activos, utilidades retenidas / activos totales, EBIT / total de activos, valor patrimonial / deuda, ventas / total de activos, utilidad antes de impuestos / pasivos totales, utilidad antes de impuestos / patrimonio, flujo de caja / total deuda, deuda /total de activos, pasivos corrientes / total de activos. Como resultado de la aplicación de este modelo, Altman y su variación sobre compañías no industriales, fue el más acertado, donde los indicadores vinculados a capital de trabajo, utilidades retenidas y EBIT son los indicadores que predicen con mayor facilidad la quiebra en dichas empresas.

(Youn & Gu, 2010) realizaron un estudio de predicción quiebra de los restaurantes en Estados Unidos, comparando un modelo logístico y una red neuronal (ANN como su acrónimo en inglés), para el periodo de 1996 al 2008 donde utilizaron una muestra de 42 empresas (21 declaradas como quebradas y 21 exitosas). Dentro del desarrollo de ambos modelos fueron

considerados un total de 10 variables, similares a las descritas por (Huo, 2006), (Gu & Gao, 2000) y (Hyunjoon & Gu, 2006), donde adicionalmente, se relacionaron indicadores de liquidez, prueba acida, EBITDA / pasivos corrientes, pasivos totales / activos totales, EBITDA / total de pasivos, EBIT / gastos de intereses, utilidad neta / ventas totales, utilidad neta / total de activos, costos de ventas / inventarios promedio, ingresos totales / promedio de activos totales.

El porcentaje de predicción para el modelo logístico correspondió a un 88.10% y 76.19% para el primer y segundo año, respectivamente, mientras que el modelo de redes neuronales sin una sensibilización de importancia a partir de los indicadores utilizados generó resultados similares al logístico (88.10% para el primer año y 78.57% para el segundo). Con el fin de dar una mayor efectividad a la red neuronal los autores dieron una escala de importancia sobre los indicadores evaluados, donde liquidez e inventarios son los de mayor relevancia. Con dicho resultado los autores demostraron una mayor efectividad en la predicción pasando al 95% y 80% para el primer y segundo año. El resultado final del estudio destacó que las firmas clasificadas como quebradas poseen condiciones similares en la predicción realizada por ambos modelos (sin sensibilización de los indicadores), donde EBITDA, ROA, deuda y costos influyen sobre las compañías que entran en quiebra en cualquiera de los dos años.

Otro estudio que compara las redes neuronales con modelos logísticos es el realizado por (Park & Hancer, 2012), en el cual utilizaron una base de 40 empresas estadounidenses de sectores hotelero, restaurantes y servicios turísticos para el periodo de 1990 al 2009 utilizando 16 indicadores relacionados a EBITDA, ROA, deuda, liquidez, entre otros.

Como resultado de los estudios, el modelo de redes neuronales obtuvo un mayor nivel de acierto en la predicción respecto al logístico, adicionalmente, los autores desarrollaron un modelo sobre los indicadores más relevantes en la predicción de quiebra donde ambos obtuvieron el mismo grado de predicción. Adicionalmente, los indicadores vinculados a liquidez y repago de deuda son los que tienen un mayor impacto en la predicción de quiebra de los restaurantes, basado en el apalancamiento financiero que sufren este tipo de compañías.

(Pacheco, 2015) desarrolló un estudio de predicción de quiebra, para el sector de hostelería, a partir de un modelo logístico para una muestra de 999 compañías PYMES portuguesas para el periodo de 2004 al 2014, clasificando 58 compañías como inactivas o en quiebra. Para la selección de las compañías PYMES, utilizaron tres filtros: cantidad de empleados, nivel de activos y ventas. Los indicadores utilizados fueron: ROA, ganancias en porcentaje de activos, deuda financiera sobre total de activos, pasivos no financieros en porcentaje de activos, efectivo en porcentaje del total de activos, patrimonio en porcentaje del total de activos. Los resultados de este estudio demuestran que factores asociados a las obligaciones financieras y patrimonio fueron los indicadores más acertados en la predicción.

Para el caso español bajo un modelo de Kaplan – Meier y regresión de COX (Gémar, Moninche, & Morales, 2016) evaluaron un total de 1.033 hoteles para el periodo de 1997 al 2009 realizando validación de indicadores financieros, dentro de las que se encuentran: liquidez, rentabilidad, ROA, solvencia y no financieras como distancias de hotel, calificación y categoría del hotel, entre otros. Del resultado, los autores identificaron que variables vinculadas al margen

de utilidad y el ciclo operacional de ingresos tienen mayor incidencia en las compañías que son evaluadas para quiebra.

(Lado - Sestayo & Viel - Bua, 2018) realizaron un estudio sobre el sector de hotelería y turismo gallego sobre un total de 1.495 compañías para el periodo 2005 al 2011. Para el desarrollo de este estudio utilizaron los modelos Probit y Logit, mediante la evaluación de 27 indicadores relacionados con rentabilidad, liquidez y endeudamiento. Como resultado del estudio los indicadores de endeudamiento y liquidez fueron los determinantes para la quiebra en España.

(Becerra-Vicario, Alaminos, Aranda, & Fernandez, 2020) efectuaron un estudio sobre 460 empresas españolas del sector de restaurantes, para el periodo de 2008 al 2017 en el cual utilizaron la técnica de redes neuronales basados en Deep Learning⁶, a partir de 28 variables que abarcan liquidez, rentabilidad y solvencia; también, adoptó variables no financieras asociadas a la antigüedad de la compañía, uso de franquicias y certificados de calidad. Los resultados del modelo de la red neuronal muestran una superioridad al modelo logístico al presentar un nivel de éxito en la predicción del 95% a la muestra total y 93.5% sobre las variables focalizadas, mientras que Logit conserva su porcentaje en alrededor del 84.10%, todo esto relacionado con la quiebra del primer año. La evaluación del segundo año posee porcentajes bajo el mismo rango, este estudio muestra que los factores de liquidez, rentabilidad y solvencia como variables financieras son las que tienen un mayor índice en la predicción de quiebra y en relación con la información no financiera la calidad de los restaurantes como variable principal.

⁶ Deep Recurrent Convolutional Neural Network

(Paez & Cabrer-Borrás, 2020) realizaron un análisis de la supervivencia de las empresas del sector de hotelería y restaurantes, tomando como base la información del periodo 2009 al 2015, generando una muestra total de 7.988 firmas. Para este estudio los autores desarrollaron un modelo de elección binaria desarrollado bajo las metodologías de descomposición de Blinder – Oaxaca y el modelo de regresión de COX. Para la definición de variables se tomaron: ROA, deuda / patrimonio neto, variables no financieras asociadas a ventas y número de empleados, tamaño de la compañía (grande o pequeña) y como variable dicotómica el flujo de caja generado (donde el flujo positivo toma valor 1 y el negativo 0). El resultado del modelo demostró una efectividad del 93% sobre la totalidad de la muestra, PYMES de un 92,46% y grandes compañías asociado a un 95.45%. Este estudio demostró la alta probabilidad de quiebra para las compañías clasificadas como PYMES con porcentajes que rodean al 60%, así como las que no involucran un alto número de empleados en su funcionamiento.

Es de resaltar que desde 2020 a partir del impacto generado por la emergencia sanitaria causada por el covid-19, se han venido desarrollando estudios que han expuesto la fragilidad existente en los diferentes sectores por las restricciones impuestas por parte de los diferentes gobiernos.

(Rafel Crespí-Cladera, 2021), realizaron una predicción de fragilidad financiera de las empresas del sector de hostelería en España, tomando una muestra del periodo comprendido de 2006 al 2019. El resultado de este estudio demostró que este sector al poseer un alto nivel de gastos operativos combinado con una alta financiación a través del sector financiero, durante el

covid-19, al tener restricciones operativas, limita la capacidad de atender sus obligaciones, así como el acceso a la banca, el cual se limita por dicha incapacidad de pago.

(Wieczorek-Kosmala, 2021), expuso la deficiencia en la preparación de las empresas del sector de hostelería ante crisis. La muestra fue construida a partir de los siguientes países: Hungría, República Checa, Polonia y Eslovaquia, para tal fin por medio de la prueba Kruskal-Wallis, identificaron que las compañías con mejores indicadores de rentabilidad (retorno sobre activos, retorno sobre patrimonio y retorno sobre ventas) y mejores capacidades de consecución de recursos financieros (medida a través de niveles de liquidez) tienen una mayor capacidad de preparación ante escenarios de crisis; por otro lado, se expone que el nivel de las ventas no tiene ninguna correlación con la capacidad de preparación ante crisis.

Metodología

El modelo de determinación de quiebra para las empresas colombianas de los sectores de hotelería y restaurantes se desarrollará por medio de un análisis estadístico sobre los indicadores más significativos, así como por medio de la regresión logística – LOGIT y el uso del Z-SCORE de Altman para empresas no manufactureras.

Para el modelo de Logit se tendrá en cuenta el riesgo de curtosis y asimetría en las observaciones tomadas, por lo cual se tendrá dentro de la metodología la winsorización de variables dado el caso.

La winzorización “(...) no es más que reemplazar valores extremos bajos o altos en el conjunto de datos con el fin de controlar los valores atípicos, los mismos que impactarán en la distribución de los datos.” (Ventura - León, 2020)

Y el método de trafico de caja sirve para “(...) forma de presentación estadística destinada, fundamentalmente, a resaltar aspectos de la distribución de las observaciones en una o más series de datos cuantitativos. Reemplaza, en consecuencia, al histograma y a la curva de distribución de frecuencias sobre los que tiene ventajas en cuanto a la información que brinda y a la apreciación global que surge de la lectura.” (Palladino, 2011)

A través del BoxPlot se puede determinar la concentración de observaciones en cada uno de los indicadores escogidos y así proceder con la winsorización de los mismos. Y tener así unos modelos más consistentes evitando tomar en cuenta valores atípicos y/o extremos.

Muestra y datos

Se obtuvo la base por medio de la plataforma SIIS – Sistema de Integrado de información Societaria de la Superintendencia de Sociedades, para el periodo 2019-2020, de las compañías incluidas dentro de los subsectores de hotelería y restaurantes, clasificadas mediante los códigos CIU versión 4 del DANE, con los siguientes subcódigos:

Grupo 551 – Actividades de alojamiento de estancias cortas

Grupo 561 – Actividades de restaurantes, cafeterías y servicio móvil de comidas.

Con el fin de clasificar las compañías en condición de quiebra, si bien algunos investigadores como (Perez, Lopera, & Vásquez, 2019), (Pérez, González, & Lopera, 2013) y (Romero, 2013) las han catalogado a partir del estatus reportado en la Superintendencia de Sociedades denominado como “proceso de reorganización, concordato, reestructuración o liquidación”, para objeto de la depuración realizada se determinó que una empresa se consideraba en quiebra cuando habiendo reportado estados financieros en 2019, no lo hicieron en 2020, generando una variable Dummy, donde 1 es igual a la empresa que no presentó estados financieros en 2020 y 0 si los presentó. Adicionalmente, ha sido incluida una variable Dummy para diferenciar las compañías hoteleras y de restaurantes, con el fin de probar su significancia estadística. Para manejar una consistencia respecto a la presentación de información financiera histórica se tomaron solo las empresas que hubieran presentado estados financieros de manera consistente desde 2015 hasta 2019.

Uno de los principales retos en la construcción de la base de datos, provino de los valores perdidos, los cuales generaron errores en la construcción de los ratios a evaluar, que no permitió incluir la totalidad de indicadores identificados como significativos por parte de los autores.

Realizada esta clasificación se identificaron 242 empresas, de las cuales 37 (equivalente al 15%) se catalogaron como “compañía sin presentar estados financieros” y 205 (equivalente al 85%) “como compañías con presentación de estados financieros”.

Dado de que el propósito es realizar un modelo de pronóstico de quiebra y contemplando que la base de datos se encuentra desbalanceada, se decide realizar el modelo con cuatro

propuestas: datos desbalanceados, remuestreo Undersampling, remuestreo Oversampling y remuestreo SMOTE.

Considerando el desbalance identificado por la depuración de datos anterior, se utilizó la siguiente metodología aplicable para el modelo logístico:

1. Se generó una regresión logística considerando la totalidad de datos, para determinar las variables más significativas.
2. Se estimó un modelo bajo la metodología de Undersampling, Oversampling y SMOTE.
3. Se compararon los resultados de la predicción por cada metodología aplicada, con el fin de identificar la técnica con menor error de predicción.

Para cada modelo evaluado, se verificó los indicadores y sus signos, con respecto a los estudios incluidos en el estado del arte a fin de verificar la consistencia del modelo determinado.

Indicadores evaluados

Los indicadores escogidos fueron los utilizados por los diversos autores en los listados iniciales de variables, con los cuales se realizará la primera revisión de la predictibilidad a partir de los modelos. A continuación, se incluye una descripción de los indicadores obtenidos de la investigación

Tabla 1 *Lista de indicadores evaluados en el modelo*

Indicador	Autores
Capital De Trabajo / Activo Total	Gu & Gao, 2000; Gu, 2002; Park & Hancer, 2012
Capital De Trabajo /Total De Activos	Gu & Gao, 2000; Gu, 2002; Park & Hancer, 2012
Ebit / Pasivo Más Patrimonio De Largo Plazo	Gu & Gao, 2000; Gu, 2002; Park & Hancer, 2012
Ebit / Pasivos Corrientes	Gu & Gao, 2000; Gu, 2002; Park & Hancer, 2012
Ebit / Total De Pasivos	Gu & Gao, 2000; Gu, 2002; Park & Hancer, 2012
Margen Bruto	Gu & Gao, 2000; Gu, 2002; Park & Hancer, 2012
Margen Neto	Gu & Gao, 2000; Gu, 2002; Park & Hancer, 2012
Pasivos De Largo Plazo / Activos Totales	Gu & Gao, 2000; Gu, 2002; Park & Hancer, 2012
Patrimonio Total / Pasivo Total	Gu & Gao, 2000; Gu, 2002; Park & Hancer, 2012
Total De Pasivos / Total De Activos	Gu & Gao, 2000; Gu, 2002; Park & Hancer, 2012
Utilidades Retenidas / Activos Totales	Gu & Gao, 2000; Gu, 2002; Park & Hancer, 2012
Utilidades Retenidas / Total De Activos	Gu & Gao, 2000; Gu, 2002; Park & Hancer, 2012
Ventas Sobre Activos Fijos	Gu & Gao, 2000; Gu, 2002; Park & Hancer, 2012
Ventas Sobre Total De Activos	Gu & Gao, 2000; Gu, 2002; Park & Hancer, 2012
Deuda /Total De Activos	Huo, 2006; Youn & Gu, 2010; Park & Hancer, 2012
Ebitda / Pasivos Corrientes	Huo, 2006; Youn & Gu, 2010; Park & Hancer, 2012
Flujo De Caja / Total Deuda	Huo, 2006; Youn & Gu, 2010; Park & Hancer, 2012
Pasivos Corrientes / Total De Activos.	Huo, 2006; Youn & Gu, 2010; Park & Hancer, 2012
Pasivos Totales / Activos Totales	Huo, 2006; Youn & Gu, 2010; Park & Hancer, 2012
Prueba Acida	Huo, 2006; Youn & Gu, 2010; Park & Hancer, 2012
Utilidad Antes De Impuestos / Pasivos Totales	Huo, 2006; Youn & Gu, 2010; Park & Hancer, 2012

Indicador	Autores
Utilidad Antes De Impuestos / Patrimonio	Huo, 2006; Youn & Gu, 2010; Park & Hancer, 2012
Valor Patrimonial / Deuda	Huo, 2006; Youn & Gu, 2010; Park & Hancer, 2012
Ventas / Total De Activos	Huo, 2006; Youn & Gu, 2010; Park & Hancer, 2012
Activos Corrientes / Pasivos Corrientes	Olsen, Bellas, & Ventrice, 1983
Activos Totales	Olsen, Bellas, & Ventrice, 1983
Capital De Trabajo / Ingresos	Olsen, Bellas, & Ventrice, 1983
Ebit / Ingresos	Olsen, Bellas, & Ventrice, 1983
Ebit / Total De Activos	Olsen, Bellas, & Ventrice, 1983
Efectivo En Porcentaje Del Total De Activos	Pacheco, 2015
Pasivos No Financieros En Porcentaje De Activos	Pacheco, 2015
Patrimonio En Porcentaje Del Total De Activos.	Pacheco, 2015
Deuda Financiera Sobre Total De Activos	Youn & Gu, 2010
Ganancias En Porcentaje De Activos	Youn & Gu, 2010
Costos De Ventas / Inventarios Promedio	Youn & Gu, 2010
Ebit / Gastos De Intereses	Youn & Gu, 2010
Ebitda / Total De Pasivos	Youn & Gu, 2010
Ingresos Totales / Promedio De Activos Totales	Youn & Gu, 2010
Roa	Youn & Gu, 2010
Utilidad Neta / Total De Activos	Youn & Gu, 2010
Utilidad Neta / Ventas Totales	Youn & Gu, 2010

Nota. Elaborada a partir de los resultados de la investigación validado en el estado del

arte

De la totalidad de indicadores evaluados, como se describió previamente, existieron algunos que por la falta de datos no fue posible su construcción. A continuación, se incluye el detalle de los indicadores depurados:

Tabla 2 Lista de indicadores depurados en el modelo.

Cuentas	Autores
Ventas Sobre Activos Fijos	Gu & Gao, 2000; Gu, 2002; Park & Hancer, 2012
Valor Patrimonial / Deuda	Huo, 2006; Youn & Gu, 2010; Park & Hancer, 2012
Ebitda / Pasivos Corrientes	Huo, 2006; Youn & Gu, 2010; Park & Hancer, 2012
Flujo De Caja / Total Deuda	Huo, 2006; Youn & Gu, 2010; Park & Hancer, 2012
Ebit / Gastos De Intereses	Youn & Gu, 2010
Costos De Ventas / Inventarios Promedio	Youn & Gu, 2010
Ebitda / Total De Pasivos	Youn & Gu, 2010

Nota. Elaborada a partir de los resultados de la investigación validado en el estado del

arte

En adición a lo anterior, se expone los diferentes grupos de indicadores existentes y sus definiciones, para una mejor comprensión de los resultados de los modelos y un análisis descriptivo del sector. Esta agrupación es tomada de (Anaya, 2015), la cual agrupa en cuatro los indicadores financieros:

- **Indicadores de liquidez:** *“Estos indicadores surgen de la necesidad de medir la capacidad que tienen las empresas para cancelar sus obligaciones de corto plazo”.* (Anaya, 2015).
- **Indicadores de actividad:** *“...llamados también indicadores de rotación tratan de medir la eficiencia con la cual una empresa utiliza sus activos, según la velocidad de recuperación de los valores aplicado en ellos.”* (Anaya, 2015).
- **Indicadores de rentabilidad:** *“Los indicadores de rentabilidad, denominados también de rendimiento o lucratividad, sirven para medir la efectividad de la*

administración de la empresa para controlar los costos y gastos, y de esta manera convertir las ventas en utilidades.” (Anaya, 2015).

- **Indicadores de endeudamiento:** “...tienen por objeto medir en qué grado u de qué forma participan los acreedores dentro del financiamiento de la empresa.” (Anaya, 2015).

Adicionalmente, se realizó un análisis descriptivo del sector para entender de una mejor manera cómo se encontraban las empresas en 2019, esto con el fin de dar un mayor contexto a los resultados de los modelos ejecutados.

Indicadores de liquidez:

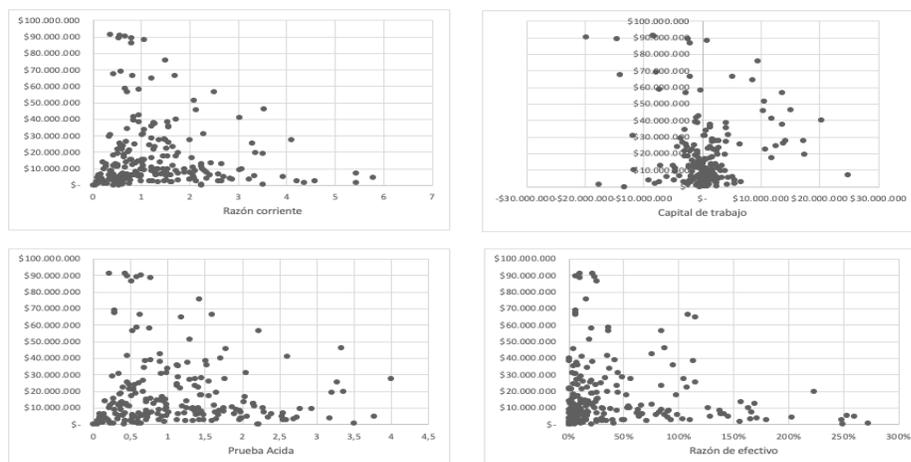


Figura 3 Gráficos de dispersión con cuatro indicadores de gestión de liquidez.

Elaboración propia

La figura 3, expone un sector que cuenta con cierta liquidez, ya que los indicadores de razón corriente y prueba acida presentan varias empresas por encima del valor de uno (1,0). A su vez es un sector altamente dependiente de su inventario y eso se puede apreciar en cómo se reduce el eje X en prueba acida en comparación a la razón corriente.

Adicionalmente, aspectos como el capital y razón de efectivo presentan una alta concentración en los valores bajos, donde se puede determinar que las compañías pertenecientes a estos subsectores no generan una solidez que les permita subsistir en escenarios desafiantes como los presentados en pandemia.

Indicadores de actividad : Para el análisis de indicadores de actividad, la base no permitió calcular valores propios de las fórmulas dificultando su lectura y revisión, esto se debe a que las cuentas precisas sobre clientes y proveedores solo se podrían sacar depurando cuentas, lo cual lo hace aproximado y para la rigurosidad del análisis no genera ningún valor agregado.

Indicadores de rentabilidad

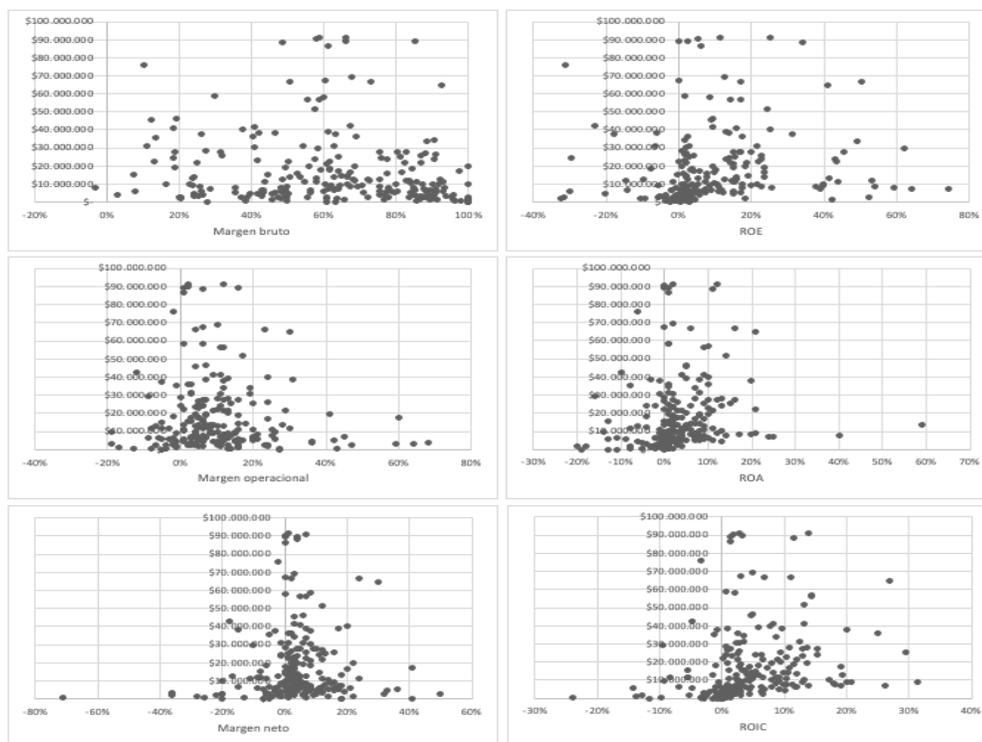


Figura 4 Gráficos de dispersión con cuatro indicadores de gestión de rentabilidad.

Elaboración propia.

La gráfica 4 presenta un sector con una buena gestión de costos en general, sin embargo, con estructuras administrativas muy robustas (disminución general entre margen bruto al margen operativo). En relación con el margen neto se observa la debilidad de las compañías por el pago de intereses a raíz de las obligaciones financieras de las compañías. El análisis comparado entre ROA y ROE, nos confirma que estamos ante un sector altamente apalancado, al haber un mayor retorno sobre el patrimonio, habla de un patrimonio pequeño versus un apalancamiento

financiero amplio, lo cual se puede corroborar con un ROA mucho más alto y diferente visualmente al ROE.

Indicadores de endeudamiento

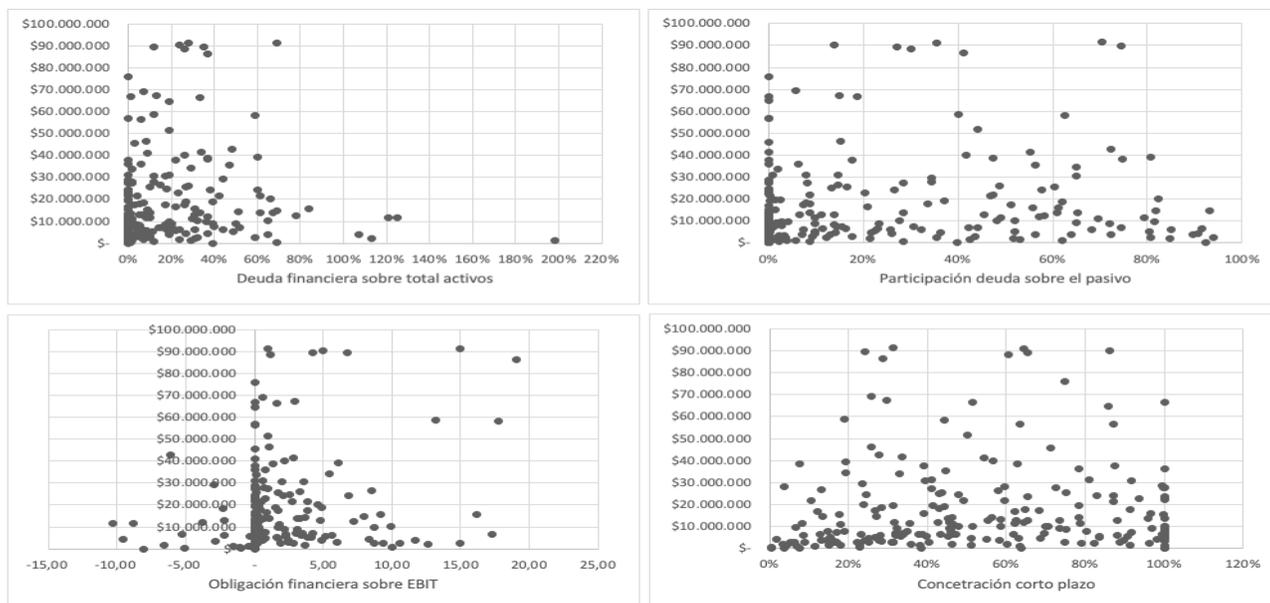


Figura 5 Gráficos de dispersión con cuatro indicadores de gestión de endeudamiento.

Elaboración propia.

La figura 5 nos demuestran un sector que se apalanca en sus alternativas de financiación, si bien se pueden ver algunas empresas con un bajo nivel de endeudamiento financiero, la imagen del sector a nivel general presenta una alta concentración de empresas con más de cinco años de EBIT (utilidad operacional), sin embargo, la porción de pago está relacionada con el corto plazo, lo cual puede estar asociado a las obligaciones con los proveedores así como a créditos de capital de trabajo para cubrir ciclos operacionales propios del sector.

Resultados

Regresión por modelo Logit

Con el fin de identificar cuales variables explican la quiebra de las empresas clasificadas en el sector de hotelería y restaurantes en la base sin balancear, se estimó un modelo considerando la siguiente expresión de acuerdo con (Caro N. , 2016) y (Caro, Díaz, & Porporato, 2013):

$$\Pr\left(y_i = \frac{1}{x_i}\right) = h(\beta_1 + \beta_2 x_i) \quad \text{donde } i = (1, 2, \dots, n)$$

Donde:

x : variable dependiente en las que se encuentran relacionadas las siguientes:

- x1 Ventas___total_de_activos_
- x2 Utilidades_retenidas___total_de_activos_
- x3 Utilidad_antes_de_impuestos___patrimonio_
- x4 Utilidad_antes_de_impuestos___pasivos_totales_
- x5 Roa
- x6 Prueba_acida_
- x7 Patrimonio_total___pasivo_total_
- x8 Pasivos_de_largo_plazo___activos_totales_
- x9 Pasivos_corrientes___total_de_activos__
- x10Margen_operativo
- x11Margen_neto_
- x12Margen_bruto
- x13Ebit___total_de_pasivos_

- x14Ebit__total_de_activos_
- x15Ebit__pasivos_corrientes_
- x16Ebit__pasivo_mas_patrimonio_de_largo_plazo_
- x17Capital_de_trabajo__ingresos_
- x18Capital_de_trabajo__activo_total_
- x19Activos_corrientes__pasivos_corrientes_
- x20_patrimonio_en_porcentaje_del_total_de_activos__
- x21_pasivos_no_financieros_en_porcentaje_de_activos_
- x22_efectivo_en_porcentaje_del_total_de_activos_
- x23_Deuda_Financiera_Sobre_Total_De_Activos_
- x24Utilidad_operacion__total de activos

y_i : Variable independiente asociada la quiebra de las compañías

β : coeficientes de la regresión

n : cantidad de observaciones resumidas a continuación tipo de base efectuada efectuado:

Tabla 3 Número de observaciones por base

Variable dummy de clasificación de quiebra	Tipo de base			
	M1	M2	M3	M4
0 = Empresa que presento estados financieros	205	37	205	205
1= Empresa que no presento estados financieros	37	37	205	205
Total, de observaciones	242	74	410	410

Fuente: Elaboración propia a partir de las bases utilizadas

Donde:

- **M1:** Base desbalanceada
- **M2:** Undersampling
- **M3:** Oversampling
- **M4:** Smote

Generada la regresión por pasos para las cuatro propuestas, se eliminaron las variables no significativas:

Tabla 4 Resultado base desbalanceada

Variable	M1	M2	M3	M4
x1	-0.7** -0.28	-0.79** -0.35	-0.74*** -0.18	-0.88*** -0.2
x2	NA	NA	1.92*** -0.5	1.72*** -0.51
x4	-9.1*** -2.9	NA	-13.55*** -2.48	-15.85*** -2.86
x5	NA	NA	-6.86*** -2.48	-6.68*** -2.25
x8	-4.96*** -1.48	NA	-5.02*** -1	-87.98*** -25.12
x9	NA	NA	NA	-83.84*** -25.19
x10	NA	7.35* -4.57	NA	NA
x11	NA	-6.89* -4.14	0.91** -0.45	0.95*** -0.31
x12	1.82** -0.81	NA	3.33*** -0.56	2.97*** -0.61
x13	7.69***	NA	13.48***	14.69***

	-2.72		-2.4	-2.66
x15	NA	-1.17*	-0.67***	-0.41*
		-0.64	-0.25	-0.22
x16	0.58*	NA	NA	NA
	-0.33			
x17	NA	NA	NA	-0.28***
				-0.09
x18	2.48**	NA	2.59***	2.77***
	-1		-0.75	-0.96
x21	2.72***	NA	4.55***	87.98***
	-1.05		-0.86	-25.2
x22	NA	6.71** (2.95)	NA	3.63**
				-1.74
x23	2.32*	NA	3.72***	86.97***
	-1.19		-1	-25.11

Nota: *** Significativo al 1%; ** Significativo al 5%; *Significativo al 10%

Fuente: Elaboración propia a partir de resultados de regresiones

Adicionalmente, considerando que la base analizada en todos los escenarios presenta compañías del subsector hotelería y restaurantes, a fin de determinar si existe incidencia estadística del subsector, se ha incluido una variable dummy relacionada a cada subsector donde 1=hotelería y 2=restaurantes. Para probar lo anterior, se verificará la significancia estadística por medio de la prueba F, que dispone lo siguiente:

$$RHo = Pvalue < N$$

Tabla 5 Prueba F

	Base ejecutada			
	M1	M2	M3	M4
F-Statistic	0.9594	0.9424	0.5922	0.0001

Fuente: Elaboración propia a partir de resultados de regresiones

Del resultado previo, se identifica que la variable dummy (relacionada con la variable x25) asociada al subsector, para la base bajo metodología SMOTE, es estadísticamente significativa a un nivel de significancia al 5% (incluyéndose como M5 para futuras referencias respecto a la base ejecutada) Bajo lo anterior, el resultado de la regresión final es la siguiente:

Tabla 6 SMOTE Dummy resultado

Variable	Valor
x1	-0.42* (0.22)
x2	1.51*** (0.54)
x4	-14.36*** (2.88)
x5	-6.75*** (2.29)
x8	-85.46*** (26.22)
x9	-81.08*** (26.21)
x11	0.85*** (0.32)
x12	2.17*** (0.63)
x13	13.87*** (2.67)
x15	-0.49** (0.22)
x17	-0.25** (0.1)
x18	3.4*** (0.88)
x21	85.74*** (26.29)
x23	85.15*** (26.23)
x25	1.81*** (0.41)

*Nota: *** Significativo al 1%; ** Significativo al 5%; *Significativo al 10%*

Fuente: Elaboración propia a partir de resultados de regresiones

Prueba de Hosmer – Lemeshow

Con el fin de verificar la validez del modelo LOGIT como parte del pronóstico, se efectúa la prueba de Hosmer – Lemeshow, donde:

$$NRHo = Pvalue > NS$$

Tabla 7 Prueba de Hosmer – Lemeshow

	Base Logit				
	M1	M2	M3	M4	M5
Prob. Chi-Sq(8):	0.2184	0.9395	0	0.5569	0.1195

Fuente: Elaboración propia a partir de resultados de regresiones

De acuerdo con resultado previo, el único modelo que no es válido para ejecutar una regresión bajo método LOGIT, es el que utiliza la metodología Oversampling, toda vez que el P-value es menor al nivel de significancia del 5% establecida, toda vez que se rechaza la hipótesis nula, razón por la cual no es considerado para la verificación de las tablas de confusión.

Tablas de confusión

Efectuada la prueba, el siguiente paso consiste en verificar que el modelo es válido para realizar el pronóstico, utilizando como puntos de corte el 50%. Como se observa en la tabla 8, la regresión para la base original presenta el mejor nivel de predicción con un 84.3% de aciertos,

respecto a la utilización de metodologías de rebalanceo, SMOTE con un 77.56% es la más acertada. Si bien la utilización de una variable dummy para discriminar el subsector no es determinante a nivel estadístico es de resaltar que la base SMOTE con dicho indicador no presenta una diferencia significativa en su porcentaje de predicción alcanzando un 76.1% , como se detalla a continuación:

Tabla 8 Tablas de confusión

	Base utilizada											
	M1			M3			M4			M5		
	0	1	Total	0	1	Total	0	1	Total	0	1	Total
% Correcto	98.54	5.41	84.3	70.27	70.27	70.27	73.66	81.46	77.56	73.66	78.54	76.1
% Incorrecto	1.46	94.59	15.7	29.73	29.73	29.73	26.34	18.54	22.44	26.34	21.46	23.9

Fuente: Elaboración propia a partir de resultados de regresiones

Análisis de variables atípicas

Como parte del proceso de revisión de si las bases utilizadas presentan valores atípicos que distorsionen los resultados de la regresión, se realiza un análisis estadístico descriptivo de las variables respectivas, detallado a continuación:

Tabla 9 Estadística descriptiva

	x1	x4	x8	x12	x13	x18	x21	x23
Media	1.12	0.18	0.3	0.57	0.23	0	0.43	0.17
Mediana	0.72	0.07	0.16	0.61	0.12	0.01	0.26	0.06
Máximo	6.46	2.64	18.32	1	2.66	0.94	19.67	1.98
Mínimo	0	-0.41	0	-7.38	-0.37	-4.68	-0.47	0
Desviación estándar	1.19	0.4	1.18	0.57	0.4	0.43	1.31	0.26
Asimetría	1.68	3.09	14.53	-11.22	3.07	-6.52	13.2	2.74
Curtosis	6.18	15.67	220.88	156.73	15.16	65.33	191.56	14.32
Jarque-Bera	216.59	2006.83	487231	243401.1	1874.75	40896.12	365551	1598.01

Fuente: Elaboración propia a partir de resultados de regresiones

Para la validación de muestras atípicas tanto la revisión del coeficiente de curtosis, así como de asimetría son los estadísticos precisos de evaluación, toda vez que los valores en distribuciones normales tienden del 0 al 3. Verificando que existen variables que presentan datos que se desplazan hasta 220 para el caso de la curtosis y 14.53 de asimetría, con el fin de reducir los datos atípicos, se realiza un proceso de winsorización al percentil 1% y 99% así como el 5% y 95%, generando el siguiente resultado:

Tabla 10 Estadística descriptiva comparativa por percentil

	Base original		1% - 99%		5% - 95%	
	Asimetría	Curtosis	Asimetría	Curtosis	Asimetría	Curtosis
x1	1.68	6.18	1.36	4.44	1.04	3.17
x4	3.09	15.67	2.99	15.43	1.38	4.82
x8	14.53	220.88	1	3.61	0.6	2.51

x12	-11.22	156.73	-0.41	2.27	-0.26	2.07
x13	3.07	15.16	2.9	14.62	1.38	5.05
x18	-6.52	65.33	-0.21	5.84	0.31	3.49
x21	13.2	191.56	0.78	2.79	0.72	2.48
x23	2.74	14.32	1.65	6.03	1.12	3.38

Fuente: Elaboración propia a partir de resultados de regresiones

Como se observa bajo el escenario del percentil 5% al 95%, respecto a su par del 1% - 99%, presenta una mejora representativa en los coeficientes evaluados, se valida que el número de observaciones previstas bajo los dos escenarios de winsorización, tal como se detalla en la tabla 11, presenta una reducción respecto a la base inicial pasando de 242, 215 con percentil 1% - 99% y 143 con percentil 5% - 95%, conservando en cierta medida las proporciones por cada tipo de variable:

Tabla 11 Proporción de observaciones

	Número de observaciones			Proporción		
	Original	1% - 99%	5% - 95%	Original	1% - 99%	5% - 95%
0	205	184	125	84.7	85.6	87.4
1	37	31	18	15.3	14.4	12.6
Total	242	215	143	100	100	100

Fuente: Elaboración propia a partir de resultados de regresiones

Seguido a esto se verifica la capacidad de predicción del modelo winsorizado en los escenarios del percentil 1% - 99% y 5% - 95%, en comparación con la base desbalanceada original, a continuación, se compara tanto las variables analizadas, la prueba de Hosmer – Lemeshow así como las tablas de confusión, siguiendo las hipótesis presentadas de manera previa:

Tabla 12 Resultado regresión comparativa

Variable	M1_Original	M1_P 1% - 99%	M1_P 5% - 95%
x1	-0.7** -0.28	NA	-0.78* 0.43
x4	-9.1*** -2.9	-1.65* 0.97	NA
x8	-4.96*** -1.48	NA	-3.23* 1.89
x12	1.82** -0.81	2.14*** 0.92	2.82** 1.40
x13	7.69*** -2.72	NA	NA
x16	0.58* -0.33	NA	NA
x18	2.48** -1	NA	NA
x21	2.72*** -1.05	NA	3.71*** 1.49
x23	2.32* -1.19	NA	NA

Nota: *** Significativo al 1%; ** Significativo al 5%; *Significativo al 10%

Fuente: Elaboración propia a partir de resultados de regresiones

Tabla 13 Prueba de Hosmer – Lemeshow comparativa

	Base Logit		
	M1_Original	M1_P 1% - 99%	M1_P 5% - 95%
Prob. Chi-Sq(8):	0.2184	0.7172	0.5555

Fuente: Elaboración propia a partir de resultados de regresiones

De acuerdo con los resultados precedentes por las regresiones realizadas, en la tabla 12 se observa que, siguiendo la línea de los indicadores depurados a los niveles de significancia, presentan la misma interpretación respecto al signo. Es de resaltar que el modelo al percentil 5%

- 95% da un mayor nivel de indicadores respecto a su par del 1%. Así mismo todos los modelos como se demuestra en la tabla 13 son válidos para el pronóstico al rechazar la hipótesis nula prevista.

Tabla 14 Tablas de confusión comparativa

	Base utilizada								
	M1_Original			M1_P 1% - 99%			M1_P 5% - 95%		
	0	1	Total	0	1	Total	0	1	Total
% Correcto	98.54	5.41	84.3	100	0	85.58	99.2	5.56	87.41
% Incorrecto	1.46	94.59	15.7	0	100	14.42	0.8	94.44	12.59

Fuente: Elaboración propia a partir de resultados de regresiones

En lo que respecta al nivel de precisión de los modelos de pronóstico ambos escenarios no presentan diferencias representativas en los que respecta al nivel de precisión de la predicción, variando porcentualmente en 1% y 3% respecto a la base original.

Regresión por Z-SCORE de Altman

Para la regresión a través del Z-Score de Altman, se tomó en cuenta la variación al modelo inicial, tomando el Z-Score para empresas no públicas, no manufactureras (Z''). En la utilización de este modelo se puede tomar una base de datos con observaciones desbalanceadas, acorde con Altman (Altman, Iwanicz-Drozowska, Laitinen, & Suvas, 2017).

Las variables tomadas para la ejecución del Z'' de Altman son las siguientes:

$$Z'' = 3.25 + 6.56 * \left(\frac{\text{Capital de trabajo}}{\text{Total de activos}} \right) + 3.26 * \left(\frac{\text{Utilidades retenidas}}{\text{Total activos}} \right) + 6.72 * \left(\frac{\text{Utilidad operacional}}{\text{Total de activos}} \right) + 1.05 * \left(\frac{\text{Valor en el mercado del patrimonio}}{\text{Valor en libros de todos los pasivos}} \right).$$

Sobre el valor obtenido del Altman se realiza una regresión logística, la cual consiste en las siguientes fórmulas

$$P(Y = 1|X) = \frac{e^{X'\beta}}{1+e^{X'\beta}} = \frac{\exp(\beta_0 + \beta_1 x_1 + \dots + \beta_k x_k)}{1 + (\exp(\beta_0 + \beta_1 x_1 + \dots + \beta_k x_k))}; \text{ (Hernández - Cruz, 2022)}$$

Y en el dado caso que el resultado sea mayor al umbral escogido, en este ejercicio fue de 50%, se da un valor de 0 (no quiebra) y si no se le da un valor de 1 (quiebra)

Tabla 15 Resultado de aciertos Altman y sensibilidad

Altman			
Estimada			
Evidenciada	0	1	Total
0	170	35	205
1	28	9	37
Total, general	198	44	242
Tasa correcta de aciertos	73.97%		
Umbral	50%		

Sensibilidad del umbral	
Umbral	74%
10%	75.62%
20%	74.79%
30%	74.38%

40%	73.97%
50%	73.97%
60%	73.55%
70%	71.90%
80%	70.25%
90%	67.77%
100%	15.29%

Fuente: Elaboración propia a partir de cálculos realizados.

Con un umbral del 50% se puede obtener hasta un 73.97% de capacidad de acierto en predicción de quiebra, por lo cual se vuelve una de las opciones sencillas a nivel de procedimiento y con una capacidad de predicción competitiva en cuestión a los demás esquemas de regresión socializados en la investigación.

Interpretación de resultados

El estudio realizado tomo como base de información, los datos reportados por las compañías a la Superintendencia de Sociedades, que por su propia naturaleza presentó un desbalanceo entre compañías consideradas como quebradas y no quebradas. A partir de ello, y tomando como base de validación, se realizó rebalanceo de la información, por medio de los métodos descritos previamente, así mismo se efectuaron revisión de estadísticos con el fin de validar variables atípicas en las bases verificadas. Como resultado, se observa que la base sin ningún tipo de rebalanceo presenta el mejor nivel de predicción alcanzando un 84.3%, respecto a las bases balanceadas la utilizada por medio del SMOTE es la que brinda el porcentaje más alto del 77.56% como se detalla a continuación:

Tabla 16 Resultados de predicción por tipo de base

	Base utilizada			
	M1	M2	M4	M5
% Correcto	84.3%	70.27%	77.56%	76.1%
% Incorrecto	15.7%	29.73%	22.44%	23.9%

Fuente: Elaboración propia a partir de resultados de regresiones

Como se demuestra la utilización de metodologías de rebalanceo de la data para el estudio efectuado, no genera un porcentaje de predicción más alto respecto a la base original, sin embargo, discrepa con lo descrito dentro de los estudios desarrollados por (Sisodia & Verma, 2018), (Alswiti, Faris, Aljawazneh, Safi, & Castillo, 2018), (Tuong, Mi Young, Jun Ryeol, & Sung Wook, 2018) y (Hui, Jia, Pei-Chann, & Jie, 2013), como se detalla a continuación:

Tabla 17 Resultados de predicción por autor bases balanceadas

Autor	Resultado	Base Original	Oversampling	Undersampling
(Sisodia & Verma, 2018)	% Correcto	50.7	74.2	84.7
	% Incorrecto	49.3	25.8	15.3
(Alswiti, Faris, Aljawazneh, Safi, & Castillo, 2018)	% Correcto	77.4	92.8	
	% Incorrecto	22.6	7.18	
(Tuong, Mi Young, Jun Ryeol, & Sung Wook, 2018)	% Correcto	82.4	84.1	84.1
	% Incorrecto	17.6	15.9	15.9
(Hui, Jia, Pei-Chann, & Jie, 2013)	% Correcto	91.9	97.1	
	% Incorrecto	8.1	2.9	

Fuente: Elaboración propia a partir de bibliografía

La anterior tabla refleja que una mejor manera de realizar predicciones de modelos de quiebra se encuentra al solucionar los problemas de desbalanceo de las datas, generando una

mayor capacidad de predicción y menor nivel de errores. Lo anterior comparado con el estudio realizado, puede estar asociado a la calidad de la información reportada por las entidades evaluadas, así como a el bajo número de empresas reportantes.

En lo que respecta a la precisión del modelo desarrollado en comparación con los autores incluidos en estudios previos, a continuación exponemos un resumen a partir del resultado para el primer año de quiebra de las compañías analizadas, donde el modelo de regresión logística presenta una mayor efectividad en la predicción, respecto a su par de MDA, a medida que los modelos son más precisos y se utilizan base de datos más amplias, coincidiendo con los resultados entregados por el estudio realizado en esta investigación:

Tabla 18 Resultados de predicción por autor

Autor	Gu & Gao (2000)	Gu (2002)	Young & Gu (2010)	Hui, Li Et. Al (2013)	Hui, Li Et. Al (2013)	Becerra Et. Al (2020)
Modelo	MDA	MDA	Logit	MDA	Logit	Logit
% Precisión	93%	92%	88.10%	98.61%	99.88%	84.10%

Fuente: Elaboración propia a partir de datos incluidos de los autores

En lo que respecta a la verificación de las ocurrencias atípicas es de resaltar que si bien la precisión de pronóstico, así como los valores generados para los indicadores de curtosis y asimetría evaluados bajo los escenarios winsorizados al percentil 1% - 99% y 5% - 95% han mejorado con relación a la base original, existe una pérdida de empresas observadas en especial para el escenario del 5%-95% y una disminución en el número de variables incluidas en el modelo, lo cual puede ocasionar un sesgo en la interpretación de los resultados de pronóstico,

considerando toda vez que el sector de hostelería prevalece en gran medida de un mayor número de PYMES respecto a empresas No PYMES y en los modelos evaluados no se introdujo una variable DUMMY de control por tamaño de empresa y considerando que el objetivo de esta investigación es generar un modelo de carácter general para cualquier tipo de compañía del subsector, el análisis así como interpretación son realizados sin ningún tipo de optimización efectuada.

Análisis de indicadores

A continuación, se incluye una ayuda visual de los indicadores evaluados en el proceso de modelación, aquellas celdas sombreadas en la tabla 19 son las que presentaron coeficiente negativo y las que están en blanco coeficiente positivo:

Tabla 19 Resultados de indicadores por estudio y base

Variable	Z-SCORE	Desbalanceada	Undersampling	SMOTE	SMOTE con Dummy
x1		X	X	X	X
x2	X			X	X
x4		X		X	X
x5				X	X
x7	X				
x8		X		X	X
x9				X	X
x10			X		
x11			X	X	X
x12		X		X	X
x13		X		X	X
x15			X	X	X
x16		X			
x17				X	X
x18	X	X		X	X
x21		X		X	X
x22			X	X	
x23		X		X	X

x24	X				
X25					X

Fuente: Elaboración propia a partir de los resultados

Con el fin de presentar una mejor interpretación de la relación sobre los indicadores, aquellos con coeficiente positivo, presentan una relación directa con la probabilidad de quiebra, entre mayor sean, mayor riesgo de quiebra. En el caso de coeficiente negativos, es una relación inversa, entre mayor sean, menor riesgo de quiebra.

Hay indicadores que aparecen en más de tres regresiones con el mismo signo del coeficiente positivo, los cuales proporcionan una estrategia financiera en pro de la mitigación de quiebra.

En adición a lo anterior, el análisis se realizará con base al modelo con mayor nivel de predicción, que fue tomando la base desbalanceada, 84.3%. Los indicadores fueron los siguientes:

- Indicadores con un nivel de significancia del 1% o menor.
 - Utilidad antes de impuestos sobre pasivo total (x4): Entre mayor sea, menor probabilidad de quiebra. Hace referencia a un indicador de endeudamiento, donde muestra nuestra capacidad de pago basada en la utilidad de la empresa.
 - Pasivo de largo plazo sobre activos totales (x8): Entre mayor sea, menor probabilidad de quiebra. Hace referencia a la concentración de la deuda. El

modelo expone que aquellas empresas que logren estructurar sus deudas en un plazo más amplio podrían reducir riesgo de quiebra.

- EBIT sobre total pasivos (x13): Entre menor sea, menor probabilidad de quiebra. Va en contraposición al primer indicador expuesto, x4, siendo el EBIT, la utilidad operacional. Es posible que se detecten fuentes de ingreso en otras actividades que pueden robustecer las utilidades antes de impuestos y que claramente en la utilidad operacional aún no se aprecia.
 - Pasivos no financieros sobre activos (x21): Entre menor sea, menor probabilidad de quiebra. El indicador hace referencia al apalancamiento operativo (proveedores, laborales).
- Indicadores con un nivel de significancia igual o menor al 5% y mayor al 1%.
- Ventas sobre total de activos (x1): Entre mayor sea, menor probabilidad de quiebra. Capacidad de generación de ingresos sobre la inversión que tiene la empresa sobre sus activos.
 - Margen bruto (x12): Entre menor sea, menor probabilidad de quiebra. Los resultados a nivel de costos y gastos no son concluyentes para la gestión de quiebra, debido en ningún escenario práctico, una empresa con bajos márgenes pueda mitigar el riesgo de quiebra.
 - Capital de trabajo sobre total activos (x18): Entre menor sea, menor probabilidad de quiebra. En relación a la liquidez que tiene la empresa con

respecto a sus activos en general, esto se puede dar debido a lo robusto que pueden ser los activos en este sector para poder operar.

- Indicadores con un nivel de significancia de igual o menor al 10% y mayor al 5%. Los de menor relevancia de los indicadores expuestos en el modelo de base desbalanceada.
 - o EBIT sobre patrimonio más pasivo de largo plazo: Entre menor sea, menor probabilidad de quiebra. Es un indicador de financiamiento, hablando de la capacidad de pago. En relación con el indicador (X13***), la utilidad operacional no resulta ser concluyente para la gestión de quiebra en una empresa.
 - o Deuda financiera sobre total de activos: Entre menor sea, menor probabilidad de quiebra. Es de los indicadores que más se repiten en las bibliografías que han sido expuestas. Donde una deuda financiera robusta puede condicionar la capacidad de operación de una empresa en momentos de crisis.

Con relación a los resultados autores que basan sus estudios en el mismo sector (Kwansa & Parsa, 1990) (Gu & Gao, 2000), (Youn & Gu, 2010) (Park & Hancer, 2012) (Pacheco, 2015) (Becerra-Vicario, Alaminos, Aranda, & Fernandez, 2020) presentaron como resultados de sus modelos altos niveles de endeudamiento ya sea por pasivos financieros o pasivos operacionales, como indicadores claves de quiebra, lo cual va en línea con los resultados de los modelos ejecutados en esta investigación.

Otros autores como lo son (Gu & Kim, 2006), (Huo, 2006), (Park & Hancer, 2012), (Lado - Sestayo & Viel - Bua, 2018), (Becerra-Vicario, Alaminos, Aranda, & Fernandez, 2020) si bien tienen semejanzas en indicadores de operación y rentabilidad, enfocan los indicadores asociados a liquidez como factor de quiebra, los cuales para los modelos evaluados no tienen tanto peso como los niveles de deuda.

Comparado con el caso colombiano sin discriminación del sector en específico, autores como (Rosillo, 2002) (Berrio & Cabeza de Vergara, 2003) (Berrio & Cabeza de Vergara, 2003) y (Perez, Lopera, & Vásquez, 2019) coincidieron con el estudio para el sector de hotelería y restaurantes que el nivel de deuda es un indicador vital para la predicción de quiebra.

Se puede apreciar que el estudio realizado muestra consistencia con aquellos expuestos en el marco teórico, donde el nivel de endeudamiento, en especial en el corto plazo puede afectar de manera definitiva la estabilidad de la empresa y es un grupo de indicadores que los gerentes deben de tener en cuenta.

Considerado lo anterior, la hipótesis planteada de manera inicial es rechazada al comprobarse con los modelos la ejecución respectiva.

Conclusión

Las compañías del sector de hotelería y restaurantes fueron unos de los más afectados a raíz de las medidas tomadas por parte del gobierno nacional como parte de la mitigación de los

efectos del covid-19, toda vez que gran parte de su flujo operativo depende de la actividad diaria que puedan ejercer las compañías.

Bajo este escenario se evaluaron un conjunto de indicadores asociados a endeudamiento, liquidez y rentabilidad. Sin embargo, como resultado del análisis efectuado, los indicadores de endeudamiento son los que detectan de mejor manera la quiebra para las compañías evaluadas, invitando a una mejor gestión y estructura de la financiación, enfocándose en primera instancia en la capacidad de apalancamiento, acompañándolo con una gestión clara hacia la rentabilidad y liquidez.

Respecto a la precisión de predicción de los modelos LOGIT bajo una base desbalanceada presenta un mayor nivel del 84.3% respecto a las otras metodologías incluidas el Z-SCORE de Altman a nivel global. En lo que respecta a la tasa de detección de empresas quebradas la base SMOTE posee la capacidad de predicción más alta al alcanzar el 77.56% a nivel global y 81.46% de predicción sobre empresas quebradas. Es de resaltar que el incluir un elemento diferenciador como es el subsector, no influye de manera significativa en la predicción efectuada.

Con base a lo expuesto en los indicadores que predicen las quiebras de las empresas en el sector de hostelería, restaurantes y hoteles; y en línea con los autores que han sido expuestos en este texto, los indicadores de endeudamiento, gestión de deuda, son los claves para mantener vigente las empresas en el sector. Evitar la concentración de deuda en el corto plazo y tener una

empresa capaz de tener unos múltiplos de deuda en relación con sus utilidades netas bastante controladas (con respecto a los indicadores resultantes, entre mayor sea esta relación, mejor).

Para futuras investigaciones se sugiere realizar una conexión de los resultados de los indicadores identificados con la evolución posterior a la pandemia para identificar si una mejora en los mismos impacta la estructura financiera y operacional de las compañías, así como la introducción de una variable Dummy por tipo de empresa (pequeña, mediana y grande) que permita controlar de mejor manera los estadísticos y variables atípicas así como un análisis adicional para la facilidad de adquisición de créditos, el costo de la deuda y su inherencia en la predicción de quiebra.

Trabajos citados

Agency, U. S. (12 de Mayo de 2021). *EPA*. Obtenido de EPA:

https://archive.epa.gov/airquality/community/web/html/lodging_sector_addl_info.html

Alencar, R. (2018). *Kaggle*. Recuperado el 15 de Febrero de 2022, de Resampling strategies for imbalanced datasets: <https://www.kaggle.com/code/rafjaa/resampling-strategies-for-imbalanced-datasets/notebook>

Alswiti, W., Faris, H., Aljawazneh, H., Safi, S., & Castillo, M. (Septiembre de 2018). Empirical evaluation of advanced oversampling. *Proc. Int. Conf. Time Ser. Forecasting (ITISE)* , 1495 - 1506. Recuperado el 15 de Febrero de 2022, de https://www.researchgate.net/profile/Hossam-Faris/publication/327780896_Empirical_evaluation_of_advanced_oversampling_methods_for_improving_bankruptcy_prediction/links/5ba41c3045851574f7dabdb6/Empirical-evaluation-of-advanced-oversampling-methods-for-impr

Altman, E. I., Iwanicz-Drozdowska, M., Laitinen, E. K., & Suvas, A. (2017). Financial Distress Prediction in an International Context: A Review and Empirical Analysis of Altman's Z-Score Model. *Journal of International Financial Management & Accounting*, 131-171.

Altman, E., & Gabriele, S. (2007). Modeling credit risk for SMEs: Evidence form the US market. *A Journal of Accounting, Finance and Business Studies*, 43(3), 332-357. Recuperado el 26 de Marzo de 2021, de <https://onlinelibrary.wiley.com/doi/abs/10.1111/j.1467-6281.2007.00234.x>

ANATO. (2020). *IV Encuesta de impacto Covid-19 Agencias de Viajes*. Bogotá: ANATO.

Anaya, H. (2015). *Análisis financiero aplicado*. Bogotá: Universidad Externado de Colombia.

Arroyave, J. (2018). A comparative analysis of the effectiveness of corporate bankruptcy prediction models based on financial ratios: Evidence from Colombia. *Journal of International Studies*, 11(1), 273 - 287. doi:10.14254/2071-8330.2018/11-1/21

- Asociación Hotelera y Turística de Colombia; Fundación Universitaria CAFAM; Alcaldía Mayor de Bogotá. (2018). *El empleo en el sector turismo: análisis de los indicadores laborales para Colombia 2007 - 2017*. Bogotá: Fundación Universitaria CAFAM.
- Aybar Arias, C., Casino Martínez, A., & Lopez García, J. (2003). Estrategia y estructura de capital en la PYME. *Estudios de Economía Aplicada*, 21(1), 27-52. Recuperado el 10 de Marzo de 2021, de <https://www.redalyc.org/articulo.oa?id=30121108>
- Banco de la República. (22 de Mayo de 2021). *Sector Externo*. Obtenido de Inversión Indirecta: <https://www.banrep.gov.co/es/estadisticas/inversion-directa>
- Banco Santander. (22 de Mayo de 2021). *Glosario*. Obtenido de Balanza Comercial: <https://www.bancosantander.es/glosario/balanza-comercial>
- Barton, S., & Gordon, P. (1988). Corporate Strategy and Capital Structure. *Strategic Management Journal*, 9(6), 623 - 632. Recuperado el 11 de Marzo de 2021, de <https://www.jstor.org/stable/2486694>
- Beaver, W. H. (1966). Financial ratios as predictors of failure. *Journal of Accounting Research*, 4, 71-111. Recuperado el 12 de Marzo de 2021, de <https://www.jstor.org/stable/2490171>
- Becerra-Vicario, R., Alaminos, D., Aranda, E., & Fernandez, M. (2020). Deep Recurrent Convolutional Neural Network for Bankruptcy Prediction: A Case of the Restaurant Industry. *Sustainability*, 12(12), 1-12. Recuperado el 31 de Marzo de 2021, de <https://ideas.repec.org/a/gam/jsusta/v12y2020i12p5180-d376089.html>
- Bellovary, J. L., & Arkers, M. D. (2007). A Review of Bankruptcy Prediction Studies: 1930 to Present. *Journal of Financial Education*, 33, 1-42. Recuperado el 1 de Marzo de 2021, de <https://www.jstor.org/stable/41948574>

- Ben Amor, S. (2009). Modèle prévisionnel de la défaillance financière des PME québécoises emprunteuses. *Journal of Small Business & Entrepreneurship*, 22(4), 517-534.
Recuperado el 16 de Marzo de 2021, de <https://www.tandfonline.com/doi/abs/10.1080/08276331.2009.10593469>
- Berrio, D., & Cabeza de Vergara, L. (2003). Verificación y adaptación del modelo de ALTMAN a la Superintendencia de Sociedades de Colombia. *Pensamiento y gestión: revista de la División de Ciencias Administrativas de la Universidad del Norte*, 15, 26-51. Recuperado el 1 de Marzo de 2021, de <https://dialnet.unirioja.es/servlet/articulo?codigo=2497285>
- Brachfield, P. (2021). Las distintas clases de insolvencia empresarial y la tipología de los deudores. *Gestión financiera*, 145 - 151.
- Bradley, M., Tarrel, G., & Kim, H. (1984). On the Existence of an Optimal Capital Structure: Theory and Evidence. *The Journal of Finance*, 39(3), 857-878. Recuperado el 10 de Marzo de 2021, de <https://www.jstor.org/stable/2327950>
- Cámara de comercio de Bogotá. (12 de Mayo de 2021). *Cámara de comercio de Bogotá*.
Obtenido de Recursos: http://recursos.ccb.org.co/ccb/instructivos/acerca_CIIU/01.html
- Caro, N. (2016). Predicción de fracaso empresarial en empresas de Argentina, Chile y Perú a través de indicadores contables. *Revista de Dirección y Administración de Empresas*(23), 130-147. Recuperado el 29 de Mayo de 2021, de <https://addi.ehu.es/bitstream/handle/10810/20341/8.%20P.Caro.pdf?sequence=1&isAllowed=y>
- Caro, N. P., Diaz, M., & Porporato, M. (2013). Predicción de quiebras empresariales en economías emergentes: Uso de un modelo logístico mixto. *Revista de Métodos*

- Cuantitativos para la Economía y la Empresa*, 16, 200-215. Recuperado el 16 de Marzo de 2021, de <http://hdl.handle.net/10419/113865>
- Caro, N., Díaz, M., & Porporato, M. (2013). Predicción de quiebras empresariales en economías emergentes: Uso de un modelo logístico mixto. *Revista de Métodos Cuantitativos para la Economía y la Empresa*, 16, 200 - 215. Recuperado el 29 de Mayo de 2021, de <https://www.econstor.eu/bitstream/10419/113865/1/774536799.pdf>
- Chawla, N., Bowyer, K., Hall, L., & Kegelmeyer, P. (2002). SMOTE: Synthetic Minority Over-sampling Technique. *Journal of Artificial Intelligence*, 16, 321-357. Recuperado el 15 de Febrero de 2022, de Resampling strategies for imbalanced datasets: <http://www.jair.org/index.php/jair/article/view/10302>
- Ciampi, F., & Gordini, N. (2008). Using Economic-Financial Ratios for Small Enterprise Default Prediction Modeling: an Empirical Analysis. *Oxford Business & Economics Conference Proceedings, Association for Business and Economics Research*, 1, 1-21. Recuperado el 14 de Marzo de 2021, de https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2230035
- DANE . (2020). *Cuentas Nacionales Trimestrales*. Bogotá: DANE.
- DANE. (2021). *Boletín Técnico - Gran encuesta integrada de hogares*. Bogotá: DANE.
- DANE. (2021). *Encuesta de Gasto Interno en Turismo EGIT*. Bogotá: DANE.
- DANE. (13 de Mayo de 2021). *Producto Interno Bruto -PIB- nacional trimestral*. Obtenido de Información IV trimestre y año 2020: <https://www.dane.gov.co/index.php/estadisticas-por-tema/cuentas-nacionales/cuentas-nacionales-trimestrales/pib-informacion-tecnica>
- Gémar, G., Moninche, L., & Morales, A. (2016). Survival analysis of the Spanish hotel industry. *Tourism Management*, 54, 428 - 438. Recuperado el 12 de Abril de 2021, de <https://www.sciencedirect.com/science/article/pii/S0261517715300637>

- Girón, H., García, J., & Herrera, R. (2016). Determinantes de la quiebra empresarial en las empresas ecuatorianas en el año 2016. *Revista Publicando*, 4(13), 108-126. Recuperado el 16 de Marzo de 2021, de <https://revistapublicando.org/revista/index.php/crv/article/view/780>
- Grisales, R. B., & Ospina, A. T. (2016). *¿Es el modelo Z-Score de Altman un buen predictor de la situación financiera de las Pymes en Colombia?* Obtenido de Repositorio EAFIT: https://repository.eafit.edu.co/bitstream/handle/10784/11575/Andres_TrujilloOspina_Rosmery_BelalcazarGrisales_2016.pdf?sequence=2&isAllowed=y
- Gruszczyński, M. (2018). On Unbalanced Sampling in Bankruptcy Prediction. *Institute of Econometrics*, 7(2), 2-13. Recuperado el 27 de Noviembre de 2021, de <https://www.mdpi.com/2227-7072/7/2/28/htm>
- Gu, Z. (2002). Analyzing bankruptcy in the restaurant industry: A multiple discriminant model. *International Journal of Hospitality Management*, 21(1), 25 - 42. Recuperado el 13 de Abril de 2021, de <https://www.sciencedirect.com/science/article/pii/S0278431901000135>
- Gu, Z., & Gao, L. (2000). A multivariable model for predicting business failures of hospitality firms. *Tourism and Hospitality Research*, 2(1), 37 - 49. Recuperado el 1 de Abril de 2021, de <https://journals.sagepub.com/doi/abs/10.1177/146735840000200108#articleCitationDownloadContainer>
- Gu, Z., & Kim, H. (2006). A LOGISTIC REGRESSION ANALYSIS FOR PREDICTING BANKRUPTCY IN THE HOSPITALITY INDUSTRY. *The Journal of Hospitality Financial Management*, 14(1), 17 - 34. Recuperado el 10 de Abril de 2021, de <https://www.tandfonline.com/doi/abs/10.1080/10913211.2006.10653812>

- Hernández - Cruz, L. V. (02 de Febrero de 2022). *Crisis empresarial en Colombia Probabilidad de entrar en proceso de insolvencia: 2016-2019*. Obtenido de Universidad de los Andes: <https://repositorio.uniandes.edu.co/bitstream/handle/1992/55745/25554.pdf?sequence=1>
- Hernandez, J., Carrasco-Ochoa, J., & Martínez-Trinidad, J. (2013). An Empirical Study of Oversampling and Undersampling for Instance Selection Methods on Imbalance Datasets. *Iberoamerican Congress on Pattern Recognition*, 262 - 269. Recuperado el 27 de Noviembre de 2021, de https://link.springer.com/chapter/10.1007/978-3-642-41822-8_33
- Hopwood, W., McKeown, J., & Mutchler, J. (1989). A Test of the Incremental Explanatory Power of Opinions Qualified for Consistency and Uncertainty. *The Accounting Review*, 64(1), 28-48. Recuperado el 12 de Marzo de 2021, de <https://www.jstor.org/stable/248127>
- Hui, L., Jia, L., Pei-Chann, C., & Jie, S. (2013). Parametric prediction on default risk of Chinese listed tourism companies by using random oversampling, isomap, and locally linear embeddings on imbalanced samples. *International Journal of Hospitality Management*, 35, 141 - 151. Recuperado el 15 de Febrero de 2022, de <https://www.sciencedirect-com.cvirtual.cesa.edu.co/science/article/abs/pii/S0278431913000893>
- Huo, Y. H. (2006). Bankruptcy Situation Model in Small Business: The Case of Restaurant Firms. *Hospitality Review*, 24(2), 49-58. Recuperado el 31 de Marzo de 2021, de <https://digitalcommons.fiu.edu/hospitalityreview/vol24/iss2/5>
- Hyunjoon, K., & Gu, Z. (2006). Predicting Restaurant Bankruptcy: A Logit Model in Comparison with a Discriminant Model. *Journal of Hospitality & Tourism*, 30(4), 474-493. Recuperado el 2 de Abril de 2021, de

<https://journals.sagepub.com/doi/abs/10.1177/1096348006290114#articleCitationDownloadContainer>

Kwansa, F. A., & Parsa, H. (1990). Business failure analysis: An events approach. *Journal of Hospitality & Tourism Research*, 14(2), 23-34. Recuperado el 2 de Abril de 2021, de

<https://journals.sagepub.com/doi/pdf/10.1177/109634809001400204#articleCitationDownloadContainer>

Lado - Sestayo, R., & Viel - Bua, M. (2018). DIAGNOSIS DE QUIEBRA HOTELERA Y DESTINO TURÍSTICO. *Lurralde: Investigación y espacio*, 41, 149 - 174. Recuperado el 10 de Abril de 2021, de <https://dialnet.unirioja.es/servlet/articulo?codigo=7252546>

Leguizamón, M. (2016). EMPLEO FORMAL E INFORMAL EN EL SECTOR TURÍSTICO EN COLOMBIA. *Turismo y Sociedad*, 179 - 189.

Liu, A., Ghosh, J., & Martin, C. (2007). Generative Oversampling for Mining Imbalanced Datasets. *DMIN*, 66-72. Recuperado el 27 de Noviembre de 2021, de <https://www.uni-muenster.de/FB10/u/lammers/EDU/ws07/Softcomputing/Literatur/4-DMI5467.pdf>

Mejía, A. (2013). La estructura de capital en la empresa: su estudio contemporáneo. *Revista Finanzas y Política Económica*, 5(2), 141-160. Recuperado el 11 de Marzo de 2021, de <http://www.redalyc.org/articulo.oa?id=323529880008>

Merwin, C. (1942). *Financing Small Corporations in Five Manufacturing*. New York: NBER. Recuperado el 12 de Marzo de 2021, de <https://www.nber.org/books-and-chapters/financing-small-corporations-five-manufacturing-industries-1926-36>

Modiglian, F., & Miller, M. (1963). Corporate Income Taxes and the Cost of Capital: A Correction. *The American Economic Review*, 53(3), 433-443. Recuperado el 10 de Marzo de 2021, de <https://www.jstor.org/stable/1809167>

- Modigliani, F., & Miller, M. (1958). The Cost of Capital, Corporation Finance and the Theory of Investment. *The American Economic Review*, 48(3), 261 - 297. Recuperado el 1 de Marzo de 2021, de <https://www.jstor.org/stable/1809766>
- Mohammed, . R., Rawashdeh, J., & Abdullah, M. (2020). Machine Learning with Oversampling and Undersampling Techniques: Overview Study and Experimental Results. *11th International Conference on Information and Communication Systems (ICICS)*, 243 - 248. Recuperado el 27 de Noviembre de 2021, de <https://ieeexplore.ieee.org/abstract/document/9078901/>
- Mossman, C., Bell, G., Swartz, L. M., & Turtle, H. (1998). An empirical comparison of bankruptcy. *The Financial Review*, 33, 35-54. Recuperado el 27 de Marzo de 2021, de <https://search.proquest.com/openview/39d804d566aefb9989cec91e34f5b0ab/1?pq-origsite=gscholar&cbl=37348>
- Myers, S., & Majluf, N. (1984). Corporate financing and investment decisions when firms have information that investors do not have. *Journal of Financial Economics*, 13(2), 187-221. Recuperado el 11 de Marzo de 2021, de <https://www.sciencedirect.com/science/article/pii/0304405X84900230>
- Olsen, M., Bellas, C., & Ventrice, L. (1983). Improving the prediction of restaurant failure through ratio analysis. *International Journal of Hospitality Management*, 2(4), 187-193. Recuperado el 2 de Abril de 2021, de <https://www.sciencedirect.com/science/article/pii/0278431983900191>
- Olshon, J. (1980). Financial Ratios and the Probabilistic Prediction of Bankruptcy. *Journal of Accounting Research*, 18(1), 109-131. Recuperado el 11 de Marzo de 2021, de <https://www.jstor.org/stable/2490395>

- Pacheco, L. (2015). SMEs probability of default: the case of the hospitality sector. *Tourism & Management Studies*, 11(1), 153 - 159. Recuperado el 10 de Abril de 2020, de <http://hdl.handle.net/11328/1080>
- Paez, B., & Cabrer-Borrás, B. (2020). Size and survival of the hospitality industry: The case of Spain. *Tourism Economics*, 1, 1-19. doi:10.1177/1354816620909997
- Palladino, A. (2011). Gráfico de Caja. *ATENCIÓN PRIMARIA DE SALUD, EPIDEMIOLOGÍA E INFORMÁTICA II*, 1-4. Recuperado el 01 de Mayo de 2022, de <https://med.unne.edu.ar/sitio/multimedia/imagenes/ckfinder/files/files/aps/GR%C3%81FICO%20DE%20CAJA.pdf>
- Park, S. -S., & Hancer, M. (2012). A comparative study of logit and artificial neural networks in predicting bankruptcy in the hospitality industry. *Tourism Economics*, 311-338. Recuperado el 31 de Marzo de 2021, de <https://journals-sagepub-com.cvirtual.cesa.edu.co/doi/abs/10.5367/te.2012.0113>
- Pérez, J., González, K., & Lopera, M. (2013). Modelos de predicción de la fragilidad empresarial: aplicación al caso colombiano para el año 2011. *Perfil de Coyuntura Económica*, 22, 205-228. Recuperado el 10 de Mayo de 2021, de <https://www.redalyc.org/pdf/861/86131758010.pdf>
- Perez, J., Lopera, M., & Vásquez, F. (2019). Una breve aplicación a la predicción de la fragilidad de empresas colombianas, mediante el uso de modelos estadísticos. *Borradores departamento de economia*, 84, 3-19. Recuperado el 4 de Abril de 2021, de <https://econpapers.repec.org/paper/col000196/017525.htm>
- Rafel Crespí-Cladera, A. M.-O.-F. (2021). Financial distress in the hospitality industry during the Covid-19 disaster. *Tourism Management*.

- Rivera, J. A. (Septiembre de 2002). Teoría sobre la estructura de capital. *Estudios gerenciales*, 84, 31-59. Recuperado el 1 de Marzo de 2021, de <https://www.redalyc.org/articulo.oa?id=21208402>
- Romero, F. (2013). Variables financieras determinantes del fracaso empresarial para la pequeña y mediana empresa en Colombia: análisis bajo modelo Logit. *Pensamiento y gestión*, 34, 235-277. Recuperado el 10 de Marzo de 2021, de http://www.scielo.org.co/scielo.php?pid=S1657-62762013000100012&script=sci_abstract&tlng=es
- Rosillo, J. (2002). Modelo de predicción de quiebras de las empresas colombianas. *INNOVAR. Revista de Ciencias Administrativas y Sociales*, 19, 109-124. Recuperado el 20 de Marzo de 2021, de <http://www.redalyc.org/articulo.oa?id=81801908>
- Sisodia, D., & Verma, U. (Junio de 2018). The Impact of Data Re-Sampling on Learning Performance of Class. *International Journal on Electrical Engineering and Informatics*, 10(2), 433 - 446. Recuperado el 15 de Febrero de 2022, de <http://www.ijeei.org/docs-14703946875bfe55d628fbb.pdf>
- Superintendencia de Sociedad . (13 de Mayo de 2021). *Insolvencia en Colombia: Datos y cifras*. Obtenido de Informes y publicaciones: https://www.supersociedades.gov.co/delegatura_insolvencia/Paginas/publicaciones.aspx
- Superintendencia de Sociedades. (s.f.).
- Tuong, L., Mi Young, L., Jun Ryeol, P., & Sung Wook, B. (2018). Oversampling Techniques for Bankruptcy Prediction:. *Symmetry*, 10(4). Recuperado el 15 de Febrero de 2022, de <https://www.mdpi.com/275218>

- United Nations World Tourism Organization. (2019). *International Tourism Highlights*. Madrid: United Nations World Tourism Organization.
- United Nations World Tourism Organization. (2020). *World Tourism Barometer*. Madrid: United Nations World Tourism Organization.
- Ventura - León, J. (Septiembre de 2020). Pearson Winsorizado: Un coeficiente robusto para las correlaciones con muestras pequeñas. *Revista chilena de pediatría*, 91(4). Recuperado el 01 de Mayo de 2022, de https://www.scielo.cl/scielo.php?script=sci_arttext&pid=S0370-41062020000400642&lang=pt
- Wieczorek-Kosmala, M. (2021). COVID-19 impact on the hospitality industry: Exploratory study of financial-slack-driven risk preparedness. *International Journal of Hospitality Management*.
- Youn, H., & Gu, Z. (2010). Predict US restaurant firm failures: The artificial neural network model versus logistic regression model. *Tourism and Hospitality Research*, 10(3), 171-187. Recuperado el 2 de Abril de 2021, de <https://journals.sagepub.com/doi/abs/10.1057/thr.2010.2>