

Universidad Católica de Santa María
Facultad de Ciencias e Ingenierías Físicas y Formales
Escuela Profesional de Ingeniería de Sistemas



**Análisis multivariante aplicado a la predicción de quiebras empresariales:
Hj-Biplot y Statis Dual como herramientas de diagnóstico financiero**

Tesis presentada por el Bachiller:

Apaza Machicao, Juan Luis

ORCID: 0009-0002-2554-4410

para optar el Título Profesional de Ingeniero de Sistemas con Especialidad en Sistemas de
Información

Asesor (a):

Mg. Lazo Barreda, Dely Marysheck

ORCID: 0000-0002-1288-5075

Arequipa – Perú

2025

UNIVERSIDAD CATÓLICA DE SANTA MARÍA
INGENIERIA DE SISTEMAS
CON ESPECIALIDAD EN SISTEMAS DE INFORMACION
TITULACIÓN CON TESIS
DICTAMEN APROBACIÓN DE BORRADOR

Arequipa, 09 de Abril del 2025

Dictamen: 013922-C-EPIS-2025

Visto el borrador del expediente 013922, presentado por:

2013244331 - APAZA MACHICAO JUAN LUIS

Titulado:

**ANÁLISIS MULTIVARIANTE APLICADO A LA PREDICCIÓN DE QUIEBRAS EMPRESARIALES:
HJ-BIPLLOT Y STATIS DUAL COMO HERRAMIENTAS DE DIAGNÓSTICO FINANCIERO**

Nuestro dictamen es:

APROBADO

Titulo Profesional/Titulo de Segunda Especialidad/Grado Académico a optar:

**INGENIERO DE SISTEMAS CON ESPECIALIDAD EN
SISTEMAS DE INFORMACION**

29242520 - MARTINEZ MUÑOZ JORGE LUIS
DICTAMINADOR



29591972 - CALDERON RUIZ GUILLERMO ENRIQUE
DICTAMINADOR



29687216 - RAMIREZ VALDEZ OSCAR ALBERTO
DICTAMINADOR



Análisis multivariante aplicado a la predicción de quiebras empresariales: Hj-Biplot y Statis Dual como herramientas de diagnóstico financiero

INFORME DE ORIGINALIDAD

14%

INDICE DE SIMILITUD

12%

FUENTES DE INTERNET

5%

PUBLICACIONES

6%

TRABAJOS DEL
ESTUDIANTE

FUENTES PRIMARIAS

1	rstudio-pubs-static.s3.amazonaws.com	1 %
Fuente de Internet		
2	uvadoc.uva.es	1 %
Fuente de Internet		
3	hdl.handle.net	1 %
Fuente de Internet		
4	www.coursehero.com	1 %
Fuente de Internet		
5	Submitted to ADEN University	1 %
Trabajo del estudiante		
6	repository.unab.edu.co	<1 %
Fuente de Internet		
7	Submitted to Universidad Internacional de la Rioja	<1 %
Trabajo del estudiante		
8	Submitted to University of Birmingham	<1 %
Trabajo del estudiante		
9	repositorio.ucsm.edu.pe	<1 %
Fuente de Internet		
10	tesis.ucsm.edu.pe	<1 %
Fuente de Internet		
11	Submitted to Universidad Católica de Santa María	<1 %
Trabajo del estudiante		

DEDICATORIA

El presente trabajo de investigación está dedicado en primer lugar a la Virgen de Guadalupe, por ser mi guía espiritual y fuente de fortaleza en este camino hacia la culminación de mis estudios. A ella le agradezco infinitamente por iluminar mi vida y darme esperanza en los momentos más difíciles.

Dedico este logro también a mi madre, quien, con su amor incondicional, sacrificio y apoyo constante han sido el pilar fundamental para alcanzar este objetivo. Gracias por ser mi ejemplo de perseverancia, dedicación y entrega, por nunca dejar que me rindiera y por brindarme siempre su confianza y motivación.

A mis hermanas, quienes siempre estuvieron a mi lado, dándome ánimo y recordándome que los sueños se pueden cumplir con esfuerzo y valentía. Su apoyo y cariño me inspiraron a superar cada obstáculo y continuar avanzando.

Y con especial dedicación, a mi padre, quien, aunque ya no está físicamente conmigo, siempre ha sido mi ángel guardián. Sé que desde el cielo me acompaña en cada paso, brindándome su fuerza y protección. Este logro es un tributo a su memoria y a todo lo que significa para mí.

A todos ellos, gracias por ser mi fuerza, mi inspiración y mi razón para alcanzar este gran objetivo. Este logro es para ustedes, con todo mi amor y gratitud.

AGRADECIMIENTOS



Quiero expresar mi más sincero agradecimiento a la Ing. Dely Marysheck Lazo Barreda por su invaluable apoyo y guía durante el desarrollo de este proyecto de investigación. Su experiencia y conocimiento en el tema han sido fundamentales para alcanzar este objetivo profesional, y su compromiso ha sido una fuente constante de inspiración.

RESUMEN

El presente proyecto de tesis aborda la predicción de quiebras empresariales mediante un enfoque de análisis multivariante avanzado. Se aplicaron las técnicas *HJ-Biplot* y *STATIS Dual* al conjunto de datos *Taiwanese Bankruptcy Prediction*, con el objetivo de identificar patrones y asociaciones clave entre variables financieras que permitan diferenciar empresas que quiebran de aquellas que no. Estas técnicas no solo facilitaron la representación gráfica de relaciones entre observaciones y variables, revelando patrones estructurales.

La investigación se basó en datos financieros de empresas taiwanesas recopilados entre 1999 y 2009, provenientes del *Taiwan Economic Journal*. Este conjunto de datos ofreció una base sólida para analizar dinámicas financieras críticas, permitiendo explorar dimensiones clave como la rentabilidad, la liquidez, el apalancamiento financiero, la eficiencia operativa y la capacidad de las empresas para cumplir con sus obligaciones a corto plazo, elementos fundamentales para evaluar la estabilidad empresarial.

Los resultados del análisis revelaron que las empresas que quiebran están estrechamente asociadas con métricas como el *Debt Ratio* y el *Liability to Equity*, las cuales reflejan altos niveles de apalancamiento financiero y dependencia de financiamiento externo. Por otro lado, las empresas que no quiebran exhibieron valores altos en variables como el *Operating Profit Rate* y el *Quick Ratio*, lo que evidencia una gestión eficiente de recursos y buena estabilidad financiera.

Este proyecto demostró la eficacia del análisis multivariante para identificar factores clave de la quiebra empresarial y descubrir patrones estructurales ocultos. Los resultados contribuyen al desarrollo de herramientas prácticas para la gestión de riesgos financieros.

Palabras clave: Quiebra Empresarial, HJ-Biplot, STATIS Dual.

ABSTRACT

This thesis project addresses the prediction of corporate bankruptcies using an advanced multivariate analysis approach. *HJ-Biplot* and *STATIS* Dual techniques were applied to the *Taiwanese Bankruptcy Prediction* dataset, with the aim of identifying key patterns and associations between financial variables that allow differentiating companies that fail from those that do not. These techniques not only facilitated the graphical representation of relationships between observations and variables, revealing structural patterns.

The research was based on financial data of Taiwanese companies collected between 1999 and 2009, from the *Taiwan Economic Journal*. This dataset offered a solid basis for analyzing critical financial dynamics, allowing the exploration of key dimensions such as profitability, liquidity, financial leverage, operational efficiency and the ability of companies to meet their short-term obligations, which are fundamental elements for assessing business stability.

The results of the analysis revealed that companies that fail are closely associated with metrics such as *Debt Ratio* and *Liability to Equity*, which reflect high levels of financial leverage and dependence on external financing. On the other hand, companies that do not fail exhibited high values in variables such as *Operating Profit Rate* and *Quick Ratio*, which evidence efficient resource management and good financial stability.

This project demonstrated the effectiveness of multivariate analysis in identifying key drivers of business failure and uncovering hidden structural patterns. The results contribute to the development of practical tools for financial risk management.

Keywords: Business Bankruptcy, HJ-Biplot, STATIS Dual.

ÍNDICE

DEDICATORIA

AGRADECIMIENTOS

RESUMEN

ABSTRACT

INTRODUCCIÓN 1

CAPÍTULO I 2

1. PLANTEAMIENTO DE LA INVESTIGACIÓN 3

1.1 Objetivos de la investigación 3

1.1.1. General 3

1.1.2. Específico 3

1.2 Alcances y Limitaciones 4

1.2.1. Alcances 4

1.2.2. Limitaciones 5

1.3 Fundamentos Teóricos 6

1.3.1. Estado del arte 6

1.3.2. Justificación Comparativa del Uso de HJ-Biplot y STATIS Dual: 12

1.3.3. Bases teóricas de la investigación: 15

1.4 Técnicas, Herramientas, Normas y Estándares 24

1.4.1. Extracción y Transformación de Datos 24

1.4.2. Validación de la Calidad de los Datos 24

1.4.3. Normas y Estándares 25

1.4.4. Técnicas de Análisis Multivariante: 25

1.5 Metodología para el desarrollo del proyecto.....	25
1.5.1. Selección de Datos.....	26
1.5.2. Preprocesamiento de Datos.....	26
1.5.3. Exploración y Análisis Inicial.....	27
1.5.4. Minería de Datos: Métodos para la reducción de dimensión.....	27
1.5.5. Minería de Datos: Aplicación de Métodos Multivariantes	28
1.5.6. Validación y Evaluación.....	28
1.6 Diseño del Proyecto	30
1.6.1. Estructura de los Datos	30
1.6.2. Diseño de Modelos Multivariantes	31
1.6.3. Definición de Criterios de Evaluación.....	31
1.6.4. Modelo Conceptual de Abordaje del Problema.....	32
CAPÍTULO II.....	33
2. DESARROLLO DEL PROYECTO	34
2.1 Selección de Datos y Preprocesamiento de Datos	34
2.1.1. Descripción del conjunto de datos	34
2.1.2. Preparación de los datos:	37
2.2 Exploración y Análisis Inicial.....	41
2.2.1. Análisis descriptivo:	42
2.2.2. Visualización de datos:	46
2.2.3. Matriz de correlación e interpretación:	49
2.3 Minería de Datos: Métodos para la reducción de dimensión.....	53
2.3.1. Análisis de Componentes Principales (ACP)	53

2.3.2. Análisis de Clustering.....	64
2.3.3. Identificación de patrones clave	70
2.4 Minería de Datos: Aplicación de Métodos Multivariantes	73
2.4.1. HJ-Biplot: Análisis y visualización de relaciones	73
2.4.2. STATIS Dual: Integración de tablas y evolución de variables.....	91
2.5 Validación y Evaluación	99
2.5.1. Validación de la Calidad de los Datos	99
2.5.2. Validación de HJ-Biplot	102
2.5.3. Validación de STATIS Dual.....	104
2.5.4. Validación de los Resultados Finales	107
CAPÍTULO III.....	111
3. RESULTADOS OBTENIDOS	112
3.1 Tasa de Predicción de Quiebras: Evaluación de la Capacidad Discriminativa...	112
3.2 Resultados de HJ-Biplot: Relación entre Empresas y Variables	113
3.3 Resultados de STATIS Dual: Integración de Tablas	121
3.4 Evaluación de las Técnicas Aplicadas	124
3.3.1. HJ – BILOT	124
3.3.2. STATIS Dual	125
3.3.3. Comparación de las Técnicas	126
CONCLUSIONES	127
RECOMENDACIONES Y TRABAJOS FUTUROS	129
REFERENCIAS.....	131
ANEXOS	135

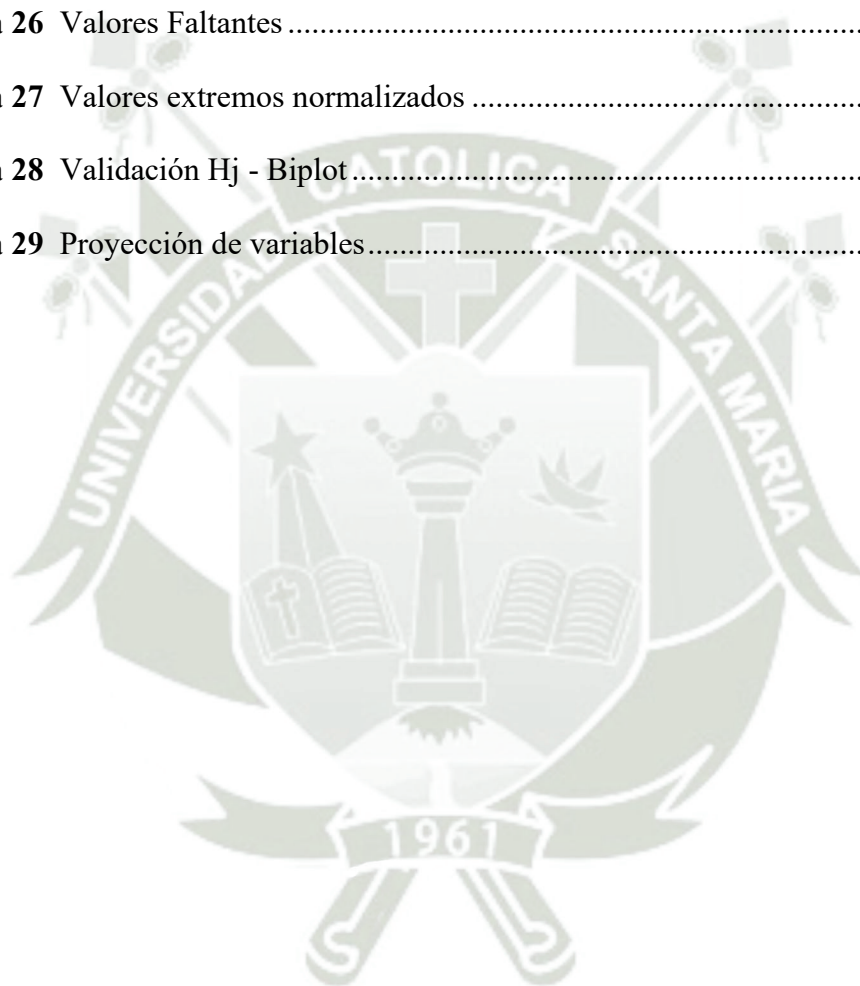
ÍNDICE DE TABLAS

Tabla 1 Comparación de Técnicas Multivariantes.....	12
Tabla 2 Estructura Metodológica Resumida.....	29
Tabla 3 Tabla de valores extremos	39
Tabla 4 Tabla de análisis descriptivo.....	45
Tabla 5 Tabla de análisis de Componente Principales (ACP)	55
Tabla 6 Tabla de componentes principales y variables dominantes	58
Tabla 7 Tablas de Clústeres	81
Tabla 8 Tasa de Predicción de Quiebras por Grupo Empresarial	112
Tabla 9 Métricas Financieras Relevantes.....	114
Tabla 10 Análisis de Cargas Factoriales en el Espacio Proyectado.....	115
Tabla 11 Vectores Cercanos (Correlación Positiva).....	117
Tabla 12 Vectores Opuestos (Correlación Negativa)	118
Tabla 13 Vectores No Correlacionados	119
Tabla 14 Comparación de Métricas Financieras.....	121
Tabla 15 Características Clave e Interpretación Financiera.....	122

ÍNDICE DE FIGURAS

Figura 1	Representación Biplot	16
Figura 2	Representación gráfica de HJ - Biplot.....	18
Figura 3	Interpretación del HJ - Biplot.....	20
Figura 4	<i>Representación gráfica STATIS DUAL</i>	23
Figura 5	Modelo Conceptual de Abordaje del Problema	32
Figura 6	Código para identificar valores faltantes.....	37
Figura 7	Código para identificar valores atípicos.....	39
Figura 8	Código análisis descriptivo	44
Figura 9	Código para gráfico de barras	46
Figura 10	Histogramas para distribuciones de variables.	47
Figura 11	Código para diagrama de caja	48
Figura 12	Diagrama de cajas	49
Figura 13	Código matriz de correlación	50
Figura 14	Matriz de correlación.....	52
Figura 15	Código para ACP.....	54
Figura 16	Código de varianza explicada acumulada	56
Figura 17	Gráfico de varianza acumulada explicada.....	57
Figura 18	Código K-means.....	67
Figura 19	Clustering de Empresas	70
Figura 20	Código para generar HJ-Biplot variables.	75

Figura 21 HJ-Biplot variables.....	79
Figura 22 Código para generar HJ-Biplot observaciones.....	80
Figura 23 HJ - Biplot observaciones (Empresas)	85
Figura 24 HJ-Biplot	91
Figura 25 Código para generar STATIS Dual	93
Figura 26 Valores Faltantes	100
Figura 27 Valores extremos normalizados	102
Figura 28 Validación HJ - Biplot.....	103
Figura 29 Proyección de variables.....	107



ÍNDICE DE ECUACIONES

Ecuación 1 Matriz de Covarianza	21
Ecuación 2 Producto Interno de Matrices (Norma de Hilbert-Schmidt)	22
Ecuación 3 Matriz de Similitud entre Tablas Basada en el Producto Interno.....	22
Ecuación 4 Fórmulas de Rango Intercuartílico (IQR)	38
Ecuación 5 Fórmula de la Media Aritmética	42
Ecuación 6 Fórmula de la desviación estándar	43
Ecuación 7 Fórmula de la varianza	43
Ecuación 8 Fórmula del Rango Intercuartílico (IQR).....	43
Ecuación 9 Fórmula de correlación de Pearson	50
Ecuación 10 Fórmula del Z-Score (Puntaje Z)	65
Ecuación 11 Fórmula de la Inercia Intra-Clúster (Suma de Distancias Cuadradas).....	66

INTRODUCCIÓN

Comprender los factores que determinan la quiebra empresarial es un desafío crucial para la gestión financiera y la estabilidad económica. Las organizaciones enfrentan riesgos derivados de su alta dependencia del financiamiento externo, baja rentabilidad y gestión ineficiente de la liquidez. Identificar patrones clave y variables críticas que permitan predecir la quiebra empresarial es fundamental para apoyar la toma de decisiones estratégicas e implementar medidas preventivas efectivas.

El presente proyecto se centra en identificar patrones financieros clave mediante técnicas multivariantes como el HJ-Biplot y el STATIS Dual, que permiten analizar relaciones complejas entre variables y proyectar diferencias significativas entre empresas que quiebran y no quiebran. Estas herramientas ofrecen una representación gráfica y estructurada de las interacciones entre métricas financieras críticas, proporcionando una visión integral de la estabilidad empresarial.

El propósito de este trabajo no se limita a identificar variables asociadas con la quiebra empresarial, sino que también busca ofrecer un marco analítico aplicable a la gestión empresarial. Los resultados obtenidos permiten diseñar estrategias prácticas como reducir el apalancamiento financiero, optimizar la liquidez operativa y mejorar los márgenes de ganancia en empresas con indicadores críticos. Asimismo, los patrones identificados pueden ser utilizados por organismos reguladores para desarrollar políticas de supervisión más efectivas.

Además, este proyecto contribuye al desarrollo de herramientas analíticas prácticas y abre la puerta a futuras investigaciones en predicción de quiebras empresariales. Estas podrían incluir modelos más avanzados que combinen técnicas multivariantes con inteligencia artificial para aumentar la precisión, así como el análisis de dinámicas financieras en sectores económicos específicos.



8. CAPÍTULO I

1. PLANTEAMIENTO DE LA INVESTIGACIÓN

1.1 *Objetivos de la investigación*

1.1.1. General

Evaluar la eficacia de las técnicas multivariantes HJ-Biplot y STATIS Dual en la predicción de quiebras empresariales, utilizando el conjunto de datos *Taiwanese Bankruptcy Prediction*, con el fin de identificar patrones estructurales en las métricas financieras que permitan diferenciar empresas propensas a la insolvencia de aquellas financieramente estables, contribuyendo así al desarrollo de herramientas predictivas para la gestión del riesgo financiero.

1.1.2. Específico

- a) Explorar las variables financieras clave del conjunto de datos mediante técnicas como el Análisis de Componentes Principales (ACP) para reducir la dimensionalidad, la construcción de matrices de correlaciones para identificar relaciones significativas entre variables, y el análisis de clustering para segmentar las empresas según su perfil financiero.
- b) Identificar las variables financieras y empresariales más relevantes del conjunto de datos *Taiwanese Bankruptcy Prediction* que influyen significativamente en la probabilidad de quiebra, enfocándose en el subconjunto de empresas que han experimentado quiebra.
- c) Aplicar el método multivariante HJ-Biplot para visualizar y analizar gráficamente las relaciones entre variables, con el objetivo de identificar factores clave como la rentabilidad, liquidez y apalancamiento financiero que influyen en la probabilidad de quiebra empresarial.

- d) Implementar el método multivariante STATIS Dual e integrar y analizar dos tablas de entrada, empresas que quiebran y empresas que no quiebran, con el objetivo de observar cómo evolucionan las variables clave y su impacto en la quiebra empresarial.
- e) Interpretar, validar y sintetizar los patrones clave y las asociaciones identificadas mediante las técnicas multivariantes aplicadas, generando conclusiones claras que permitan diferenciar empresas según su nivel de riesgo y comprender los factores determinantes en la predicción de quiebras empresariales.

1.2 Alcances y Limitaciones

1.2.1. Alcances

El presente proyecto de tesis tiene como objetivo principal analizar el conjunto de datos Taiwanese Bankruptcy Prediction mediante el uso de técnicas multivariantes avanzadas, específicamente HJ-Biplot y STATIS Dual, para identificar patrones clave que contribuyan a la predicción de quiebras empresariales, en ese sentido, los alcances de esta investigación comprenden:

1. Ámbito Temporal y Geográfico:

El estudio se basa en datos financieros de empresas taiwanesas recopilados entre los años 1999 y 2009, extraídos del *Taiwan Economic Journal*.

2. Ámbito Temático:

Se enfoca en la identificación de variables financieras que influyen significativamente en la probabilidad de quiebra, a través del análisis de métricas relacionadas con rentabilidad, liquidez y apalancamiento.

3. Ámbito Metodológico:

El análisis se realiza aplicando técnicas HJ-Biplot y STATIS Dual como herramientas de diagnóstico financiero multivariado, permitiendo observar relaciones estructurales entre variables y diferenciar grupos de empresas con distintos niveles de riesgo.

4. Ámbito Académico y Aplicativo:

Los resultados de la investigación servirán como base metodológica para futuros estudios en el campo de la predicción de quiebras empresariales, contribuyendo al desarrollo de modelos más precisos de evaluación financiera y gestión del riesgo.

1.2.2. Limitaciones

La presente investigación, si bien aporta valor en el análisis multivariante de quiebras empresariales, presenta ciertas limitaciones que deben ser consideradas al interpretar los resultados.

1. Limitación Temporal:

El análisis se realiza sobre datos históricos correspondientes al periodo 1999 – 2009, lo cual puede no reflejar comportamientos actuales del entorno financiero ni cambios regulatorios posteriores.

2. Limitación Geográfica:

El modelo analítico que se desarrollará estará específicamente ajustado a la base de datos de las empresas Taiwanesas dentro del contexto económico de los años 1999 al 2009. Si bien es posible que las técnicas sean generalizables, la aplicabilidad de este modelo a otros sectores económicos o regiones puede requerir ajustes adicionales, siendo las características financieras de cada país y la normativa activa en ellos las que pueden influenciar en la efectividad del modelo, que limitaría su uso en otros entornos sin una calibración o validación adicional.

3. Limitación de Viabilidad:

No se contó con la validación directa de expertos del mercado taiwanés, lo cual podría haber enriquecido la interpretación contextual de los resultados.

4. Limitación de Recursos y Financiamiento:

El proyecto fue desarrollado de forma académica, sin apoyo financiero externo, lo que condicionó la disponibilidad de herramientas de análisis avanzadas (licencias de software) o validaciones más extensas con bases de datos privadas.

5. Limitación en la Dimensión del Tiempo de Ejecución:

La duración del trabajo estuvo acotada a un semestre académico, lo que limitó la posibilidad de ampliar el análisis a otros sectores o extender la validación del modelo a otros periodos temporales.

1.3 Fundamentos Teóricos

1.3.1. Estado del arte

1. HJ-Biplot como metodología exploratoria para el análisis multidimensional de los Objetivos de Desarrollo Sostenible (ODS) a nivel municipal en Bolivia.

Silva Viamonte (2024) analiza las metas de la Agenda 2030 para el desarrollo sostenible y destaca la importancia de contar con indicadores objetivos que permitan evaluar su cumplimiento. En su estudio, propone el uso del HJ-Biplot como una herramienta de análisis multidimensional que facilita la interpretación de grandes volúmenes de datos. A partir de la información del Atlas Municipal de los ODS en Bolivia, identifica agrupamientos territoriales basados en indicadores de cumplimiento, lo que permite orientar las políticas públicas según las particularidades de cada localidad. El estudio

subraya la utilidad del HJ-Biplot en la integración de técnicas analíticas avanzadas para generar evidencia con impacto social.

2. Combining intra-risk and contagion risk for enterprise bankruptcy prediction using graph neural networks

Wei, Lv, Guo, Yang y Chen (2024) desarrollaron un modelo de predicción de quiebra empresarial que integra el análisis de riesgo intraempresarial y el riesgo de contagio entre organizaciones. En una primera etapa, el estudio propone el uso de un codificador que utiliza información básica de las empresas y datos sobre litigios para evaluar el riesgo financiero individual. Posteriormente, se analiza el “efecto derrame” del riesgo, considerando tanto las relaciones hiperbordes como las relaciones heterogéneas por pares. Esta investigación se destaca por ser una de las primeras en integrar simultáneamente el riesgo interno y el de contagio en la inferencia de quiebra empresarial, aportando una visión más compleja y realista del proceso de deterioro financiero.

3. Predicción de Quiebras en el Sector de la Restauración en España

Morrás Lorenzo y Fernández Rodríguez (2023) desarrollaron un estudio centrado en la predicción de la quiebra empresarial en el sector de restaurantes y puestos de comida en España, basado en la clasificación CNAE 2009. Para ello, aplicaron modelos de regresión logística y redes neuronales, utilizando información contable extraída de la base de datos SABI (Sistema de Análisis de Balances Ibéricos). Los resultados mostraron que era posible anticipar la quiebra con un año de antelación, identificando patrones financieros asociados a la insolvencia. Además, se sugiere la incorporación de variables no contables, como factores macroeconómicos, para mejorar la capacidad predictiva de los modelos. El estudio destaca la importancia de integrar múltiples fuentes de información para fortalecer la

gestión del riesgo financiero en este sector específico.

4. Un análisis bibliométrico de la predicción de quiebra empresarial con Machine Learning.

Franco (2023) realizó un estudio bibliométrico enfocado en la aplicación de técnicas de *machine learning* (ML) para la predicción de quiebras empresariales, a partir de una revisión exhaustiva de la base de datos Web of Science. El análisis incluyó 327 documentos relevantes, que permitieron clasificar y evaluar las principales métricas de desempeño utilizadas en modelos de clasificación, como el área bajo la curva (AUC) y la precisión (ACC). Los resultados del estudio destacan la efectividad de algoritmos avanzados como XGBoost, SVM, SMOTE y Random Forest (RF), los cuales superan ampliamente a las metodologías tradicionales en capacidad predictiva, especialmente al considerar horizontes de tiempo previos a la quiebra. Asimismo, se enfatiza la importancia de utilizar tanto variables financieras como no financieras para mejorar la precisión de los modelos. Este trabajo evidencia el potencial de las técnicas de ML para transformar el análisis del riesgo de insolvencia y ampliar las oportunidades en la construcción de modelos predictivos más robustos.

5. Un análisis bibliométrico de la predicción de quiebra empresarial con Machine Learning.

Carlos (2023) presentó un estudio de análisis bibliométrico que examina cómo las técnicas de *machine learning* (ML) han influido en el desarrollo de investigaciones sobre quiebra empresarial, a partir de una revisión sistemática de la base de datos Web of Science. El estudio analizó un total de 327 documentos que evidencian la adaptación de modelos de ML a problemas de predicción de insolvencia. Entre los resultados más destacados, se

identificó un uso recurrente de algoritmos como XGBoost, SVM, SMOTE, Random Forest (RF), Decision Trees (DT), los cuales demostraron un mayor nivel de precisión en comparación con las técnicas estadísticas tradicionales. Además, se señala que la incorporación de variables financieras y no financieras contribuye significativamente a mejorar la capacidad predictiva de estos modelos, especialmente en horizontes temporales previos a la quiebra.

6. Modelos de Predicción de Fragilidad Empresarial: Una Herramienta para Detectar la Bancarrota

Molina-Panchi, P. y Molina-Panchi, D. (2023) analizaron la teoría de la fragilidad empresarial a partir de la detección de riesgos de bancarrota, identificando los métodos más relevantes por su confiabilidad y validez. El estudio utiliza tanto variables cuantitativas como cualitativas, con el fin de demostrar cómo estas herramientas permiten anticipar la incertidumbre financiera y facilitar la toma de decisiones correctivas que maximicen la continuidad empresarial y minimicen el riesgo de pérdida. Los modelos aplicados fueron los de Altman, Springate, Ohlson, Pascale y Fulmer, todos ellos basados en información contenida en los estados financieros y en datos históricos, los cuales ofrecen variables explicativas clave sobre el éxito o fracaso de las organizaciones.

7. Modelo de Predicción de Quiebra Z2 de Altman en Empresas del Sector Inmobiliario de Pichincha.

Molina Panchi, P. A., Molina Panchi, D. F. y Flores Tapia, C. E. (2022) desarrollaron un estudio cuantitativo que emplea el modelo Z2 de Altman para evaluar la solvencia financiera del sector inmobiliario en la provincia de Pichincha, Ecuador. La investigación utiliza análisis multivariante y pruebas estadísticas con el objetivo de identificar relaciones

entre variables clave que inciden en el éxito o fracaso empresarial. El diseño longitudinal abarca el período 2016–2020, utilizando datos proporcionados por la Superintendencia de Compañías, Valores y Seguros de Ecuador. Los resultados muestran que las empresas financieramente saludables presentan bajos niveles de endeudamiento y altos niveles de ingreso, capital, liquidez y rentabilidad por encima del costo de capital. En contraste, las empresas clasificadas en zona de precaución enfrentan dificultades para cubrir sus obligaciones, con indicadores de rentabilidad marginal, liquidez negativa, capital negativo y disminución en ingresos y activos.

8. Predicción de la Quiebra Empresarial en el Sector Agroindustrial de Machala

Rivadeneira, Saltos, Rivera y Carpio (2022) desarrollaron un modelo predictivo orientado a anticipar la quiebra empresarial en el sector agroindustrial de Machala, Ecuador. El estudio se basó en una muestra de más de 300 empresas y empleó 13 métricas financieras para evaluar su salud económica, a través de un análisis de correlación. Para balancear el conjunto de datos entre empresas activas e inactivas, se utilizó un muestreo aleatorio simple. El modelo, implementado con árboles de decisión, alcanzó una precisión del 78.57%, superando a otros enfoques comparativos. Su aplicación al ejercicio fiscal de 2018 permitió emitir alertas tempranas para 190 empresas, recomendando una revisión detallada de su situación financiera con el fin de prevenir posibles quiebras.

9. Modelos de clasificación y predicción de quiebra de empresas: una aplicación a empresas chilenas

Romani Chocce, Aroca González, Aguirre Aguirre, Leiton Vega y Muñoz Carrazana (2002) llevaron a cabo un estudio centrado en la clasificación y predicción de quiebras empresariales en Chile, aplicando tres modelos estadísticos ampliamente utilizados:

análisis discriminante (ADM), regresión logística (LOGIT) y redes neuronales (RN). El objetivo fue identificar cuál de estos enfoques ofrece mayor exactitud y confiabilidad en la predicción del riesgo de insolvencia. Para ello, se utilizaron índices financieros clave — como razones de liquidez, actividad, rentabilidad y endeudamiento— complementados con variables macroeconómicas y de control. La metodología incluyó la construcción de una base de datos, el procesamiento de la información recopilada, y la definición de la muestra y variables. El estudio se estructura en tres secciones: una revisión de la evolución de los modelos de predicción, la descripción metodológica, y el análisis de resultados. Los hallazgos destacan la utilidad de los modelos multivariantes para detectar factores críticos que inciden en la quiebra empresarial, y aportan herramientas valiosas para la gestión financiera preventiva en el contexto chileno.

10. Análisis financiero para la predicción de quiebra empresarial, en el sector de la construcción en Colombia.

Rincón Moreno y Rodríguez Vargas (2019) realizaron un estudio orientado a la predicción de quiebras empresariales en el sector de la construcción en Colombia, mediante la identificación de diferencias clave entre empresas solventes y en crisis. Utilizando información financiera del año 2017, proveniente de la Superintendencia de Sociedades, analizaron una muestra de 24 empresas, divididas equitativamente entre saludables y en crisis. El estudio empleó 10 ratios financieros relevantes según la literatura económica, aplicando técnicas tanto univariantes como multivariantes, entre ellas el análisis discriminante. Los resultados indicaron que los indicadores más determinantes para identificar situaciones de crisis fueron el ROCE, el ROA, la autofinanciación y el nivel de endeudamiento. El modelo alcanzó un nivel general de acierto del 83,3%, con una precisión

del 100% en empresas en liquidación y del 66,7% en empresas solventes. El estudio subraya la utilidad del análisis financiero como herramienta preventiva en sectores estratégicos como el de la construcción.

1.3.2. Justificación Comparativa del Uso de HJ-Biplot y STATIS Dual:

En el marco del análisis multivariado aplicado a la predicción de quiebras empresariales, existe una amplia gama de técnicas estadísticas. La elección de HJ-Biplot y STATIS Dual en esta investigación responde a su capacidad para representar simultáneamente observaciones y variables, así como comparar estructuras financieras entre grupos (quiebran vs. no quiebran).

A continuación, se presenta una comparación técnica de los principales métodos alternativos considerados, detallando sus aplicaciones, ventajas y limitaciones en el contexto de análisis estructural y diagnóstico financiero:

Tabla 1

Comparación de Técnicas Multivariantes

Técnica	Aplicación Principal	Ventajas Clave	Limitaciones
Multivariante			
HJ-Biplot	Visualización de observaciones y variables en un espacio común	Alta interpretabilidad gráfica, correlaciones visibles, relación directa con clusters	Requiere proyección bidimensional, pierde información con más componentes
STATIS Dual	Integración estructural de múltiples grupos con mismas variables	Compara grupos (quiebran/no), identifica patrones estables	Computacionalmente más compleja, interpretación técnica especializada

Sparse STATIS	Versión regularizada de STATIS para datasets de alta dimensión	Reduce ruido, selecciona variables relevantes, evita sobreajuste	Requiere selección de penalización adecuada; menor uso en contextos financieros
Biplot	Proyección exclusiva de variables sin representación de observaciones	Permite evaluar cargas factoriales puras	No permite estudiar empresas directamente; se pierde dimensión dual
Regresión Logística	Predicción supervisada (0=No quiebra, 1=Quiebra)	Alta precisión clasificatoria; ampliamente validada	Supone linealidad; limitada capacidad estructural multivariante
Análisis Discriminante (ADM)	Clasificación supervisada	Útil en separabilidad clara; modelos clásicos	Supone normalidad multivariada y varianzas homogéneas
PLS (Partial Least Squares)	Reducción + regresión en datos correlacionados	Maneja multicolinealidad	Interpretación compleja; menor visualización

Nota: Elaboración propia basada en revisión metodológica aplicada a análisis multivariado financiero y predicción de insolvencia.

Como se observa en la comparación anterior, cada técnica multivariante ofrece enfoques distintos según el objetivo analítico, el tipo de datos y la necesidad de interpretación. En este estudio, se eligieron HJ-Biplot y STATIS Dual por su capacidad complementaria para abordar simultáneamente la estructura interna de los datos y las diferencias entre grupos financieros (empresas que quiebran y no quiebran), manteniendo al mismo tiempo una representación gráfica clara y explicativa.

El HJ-Biplot permite proyectar en un espacio común tanto a las empresas como a las variables financieras, facilitando una interpretación visual de agrupamientos, correlaciones y vectores que influyen en la quiebra o estabilidad. Esta herramienta multivariante combina efectividad analítica con intuición visual, haciéndola especialmente útil en análisis exploratorios con múltiples variables.

Por otro lado, el STATIS Dual fue seleccionado por su capacidad para integrar y comparar de manera estructurada múltiples matrices de datos con las mismas variables pero diferentes individuos, permitiendo evaluar patrones comunes y contrastes entre empresas según su estado financiero. Su enfoque estructural ofrece una base sólida para analizar la evolución y comportamiento diferencial de variables en distintos contextos.

Si bien existen alternativas como la regresión logística, el análisis discriminante o el clustering, estas técnicas no ofrecen la misma profundidad estructural ni la representación gráfica simultánea de variables y observaciones. Además, herramientas como Sparse STATIS o PLS requieren condiciones particulares de modelado o no están optimizadas para la interpretación visual directa, lo cual limita su aplicabilidad exploratoria en estudios de diagnóstico financiero como el presente.

En conjunto, la combinación de HJ-Biplot y STATIS Dual responde no solo a la necesidad de análisis estadístico multivariado, sino también a un enfoque diagnóstico y visualmente interpretable para la identificación de patrones estructurales claves asociados a la quiebra empresarial.

1.3.3. Bases teóricas de la investigación:

Se explica detalladamente cada concepto usado para asegurar un entendimiento adecuado.

1.3.3.1. Quiebra Empresarial:

El término quiebra empresarial es cuando una organización no puede cumplir con sus obligaciones de pago a falta de liquidez o solvencia. Desde una perspectiva teórica, la quiebra empresarial ocurre como consecuencia de un deterioro en las finanzas de la empresa, ocasionado por distintos factores que pueden ir desde la falta de rentabilidad hasta una mala gestión, entorno económico, fluctuaciones en el mercado, competencia, estos son factores tanto internos como externos.

La teoría clásica de la insolvencia financiera que lleva a la quiebra empresarial es el análisis de factores de ratios financieros, como la liquidez, rentabilidad y eficiencia, que son métricas fundamentales para analizar la salud financiera de una empresa.

1.3.3.2. Información Financiera:

Este término es usado para analizar los valores empresariales y dar un panorama claro de la situación económica empresarial. Para esto se usan términos como el balance general, estados de resultado y flujo de caja: es información clave para hacer una evaluación correcta. Este proyecto de tesis asumirá como base estas señales para el uso de las técnicas multivariantes y expresarlas de manera precisa y efectiva.

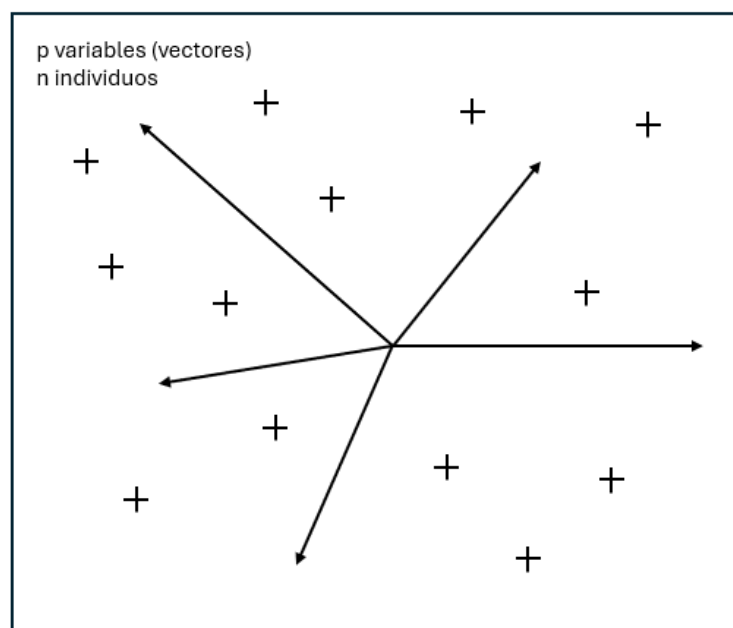
1.3.3.3. Biplot:

Los Biplot fueron desarrollados por Gabriel en el año 1971, para poder representar gráficamente una matriz de datos en una baja dimensión, estos métodos permiten la representación de múltiples variables simultáneamente. (Gabriel & Odoroff, 1990)

Los métodos Biplot tienen como objetivo principal representar las filas de una matriz como puntos y las columnas como vectores, así nos permite la aproximación de cada elemento de la matriz de estudio, pero al realizar esto se presentan limitantes, el poder incluir más de dos variables, las distribuciones multivariantes que resultan son difíciles de visualizar y entender.

Figura 1

Representación Biplot



Nota: Elaboración propia basada en la interpretación gráfica del HJ-Biplot, adaptada de Díaz-Faes, A. A., González-Albo, B., Galindo, M. P., & Bordons, M. (2013).

Este enfoque se basa en descomponer la matriz de estudio e intentar reproducir los datos de esta. La representación visual que integra tanto filas como columnas en dicha matriz captura información de las variables J , que son las medidas sobre los I individuos. En esta representación se utilizan marcadores que facilitan la interpretación:

Los marcadores $a_1, \dots, a_j$ simbolizan filas de una matriz A .

Los marcadores $b_1, \dots, b_j$ simbolizan columnas de una matriz B

$$X \cong AB^T$$

Los marcadores a_i y b_i que representan filas y columnas respectivamente, se representarán en un espacio cuya dimensión se encuentra en $q \leq r$, en este contexto q representa la cantidad de ejes que se han elegido, mientras que r se refiere al rango de la matriz X .

Donde descomposición de la matriz X garantiza la proyección en dos dimensiones de manera aproximada, que permite la visualización de filas y columnas, esto se puede reconstruir de la siguiente manera:

$$x_{ij} = \mathbf{a}_i' \mathbf{b}_j \quad (i = 1, 2, \dots, I; \quad j = 1, 2, \dots, J)$$

Esta expresión es una forma bilineal que fue descrita por Gollob en 1986, donde se obtiene el producto interno entre las filas i en la matriz A al igual que las columnas j en la matriz B , al realizar esto nos permite visualizar las iteraciones donde las filas y columnas en un espacio bidimensional lo que correspondería a la proyección ortogonal de los valores de a_i sobre los valores b_j y de manera viceversa.

HJ-Biplot

El HJ-Biplot es una visualización gráfica multivariante donde se representa simultáneamente filas que representan a los individuos del análisis y columnas que representan a las variables esto para un único sistema de coordenadas optimizando la representación gráfica de las filas que se identifican como individuos y para las columnas que se identifican como variables de estos (Galindo,1986), para esta formulación se parte de lo siguiente:

n representa la cantidad de filas (o individuos, observaciones).

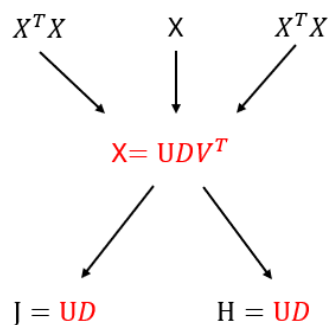
p corresponde al número de columnas (o variables medidas).

$$X_{n \times p}$$

Esta matriz tiene los datos para analizar y graficar. Para la construcción del HJ-Biplot se toma la descomposición que resulta en los valores singulares (SVD) que permite desglosar los valores en componentes simples y manejables, que hace que el análisis y la interpretación siguiente sean más sencillos para dicha matriz.

Figura 2

Representación gráfica de HJ - Biplot



U es una matriz cuyos vectores columna son ortonormales y vectores propios de XX'

V es una matriz ortonormal cuyos vectores columna son vectores propios de $X'X$

D es la matriz diagonal de valores singulares de X , que son las raíces cuadradas no negativas de los valores propios de $X'X$

Nota: Elaboración propia basada en la información de Gabriel (1971).

La representación simultánea ofrece una ventaja significativa, que proporciona una calidad óptima para representar columnas y de igual forma las filas de tu conjunto de datos.

Reglas principales para la interpretación del HJ-Biplot

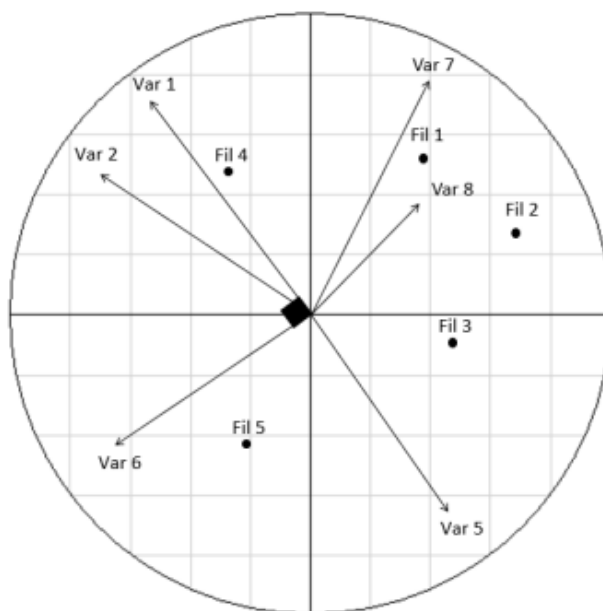
a) El espacio entre los marcadores de las filas cumple una función inversa a sus similitudes; esto se explica porque mientras más cerca estén entre sí, son grupos considerados similares, lo que nos permite identificar las semejanzas. (Vicente Tavera et al., 1995).

b) La longitud de las columnas refleja la desviación típica de las variables representadas; a mayor longitud, mayor es la información disponible para discriminar entre las distintas variables.

c) Los cosenos de los ángulos entre las columnas representadas indican como se correlacionan entre ellos, esto se puede interpretar de la siguiente manera mientras más agudos son los ángulos más alta será su correlación y esta será positiva, esto se aprecia en las variables 1 y 2, los ángulos obtusos se consideran como la correlación negativa, estos se aprecia en las variables 2 y 5 y los ángulos rectos se interpretan como la ausencia de correlación esto se aprecia entre las variables 1 y 5 (ver figura 4)

d) Las proyecciones ortogonales de las filas que son los puntos sobre las columnas, se interpretan que mientras mayor sea su proyección, tendrá un mayor peso de esa variable.

Figura 3
Interpretación del HJ - Biplot



Nota: *Elaboración propia basada en la información de Galindo (2013).*

1.3.3.4. Statis:

El método STATIS (Structuration de Tableaux à Trois Indices de la Statistique) fue introducido por L'Hermier Des Plantes y Thiébaut en 1977, después ampliado y desarrollado por Lechevalier en 1990. Esta técnica estadística multivariante se usa para analizar simultáneamente múltiples matrices de datos que podrían estar configuradas para los mismos individuos o variables. El objetivo principal es obtener un subespacio denominado consenso, que identifica la estructura común que es ponderada entre todas las matrices de datos involucradas y se evalúa la estabilidad de esta estructura común de las diferentes matrices. (Lechevalier, 1990).

El enfoque es lograr un subespacio que es conocido como consenso para todas las matrices de datos implicadas (Castillo Elizondo & González Varela, 1998). Siendo el

objetivo principal la estructura común para las matrices involucradas y determinar si la estructura encontrada permanece estable ante las diferentes matrices y sus variables.

Al llevar a cabo un análisis utilizando STATIS, la estructura de los datos es un factor determinante, lo que nos permite identificar dos tipos diferentes:

Las matrices de datos pueden presentar las mismas personas, mientras que las variables pueden variar. En este caso, se utiliza el método STATIS, que se enfoca en comparar a los individuos. (Castillo Elizondo & Gonzalez Varela, 1998)

Por otro lado, cuando las matrices tienen las mismas variables, pero los individuos son diferentes, se utiliza el método STATIS DUAL, dado que el estudio también abarca la comparación entre variables e individuos.

El método STATIS Dual tiene tres etapas a seguir:

Inter-estructura: En esta etapa realiza la evaluación comparativa integral de diferentes estructuras de las T matrices de datos a examinar, para esto se emplea la distancia que se origina desde el producto escalar de Hilbert-Schmidt, esto supone una transformación de distintas matrices de datos en los productos matriciales cruzados usando la siguiente formula:

Ecuación 1

Matriz de Covarianza

$$C_t = X_t^T X_t$$

La representación visual de los distintos objetos C_t , requiere la definición de una distancia por lo tanto se emplea la distancia que se origina por el producto escalar de Hilbert – Schmidt, que se define de la siguiente manera:

Ecuación 2

Producto Interno de Matrices (Norma de Hilbert-Schmidt)

$$\langle C_t | C_{t'} \rangle_{HS} = \text{tr} (C_t C_{t'})$$

El resultado de estos productos escalares construye la matriz Z del orden TxT, que es nombrada la matriz de correlación de vectores entre Ct.

Ecuación 3

Matriz de Similitud entre Tablas Basada en el Producto Interno

$$\mathbf{Z} = \{\mathbf{Z}_{tt'}\} = \{\langle C_t | C_{t'} \rangle_{HS}\}$$

El cálculo de los vectores y valores característicos (propios) de la matriz de correlaciones de vectores que será usada en la disminución de la dimensionalidad (Caballero-Juliá et al., 2017), para llevar a cabo la representación en el espacio euclídeo.

Estudio del compromiso: La meta principal de esta etapa es identificar la similitud entre las distintas matrices de datos y esta se forma a partir de las diversas submatrices extraídas de los datos a estudiar, en la que cada matriz se pondera, que busca asegurar la representación precisa, obteniendo un resumen integral del conjunto de matrices analizadas. (Corrales & Rodríguez, 2014).

Se debe obtener todas las submatrices de datos para el análisis y esta será la combinación lineal de cada una de las disposiciones matriciales de los datos base del estudio y todas estas son ponderadas, dándonos una mejor representación considerando todas las submatrices iniciales.

La creación de los componentes principales (ACP), se deriva de la varianza del grupo nuevo de variables ortogonales (Corrales & Rodriguez), se obtendrán a través de la combinación lineal de las variables iniciales, que nos permite comprender hasta cierto

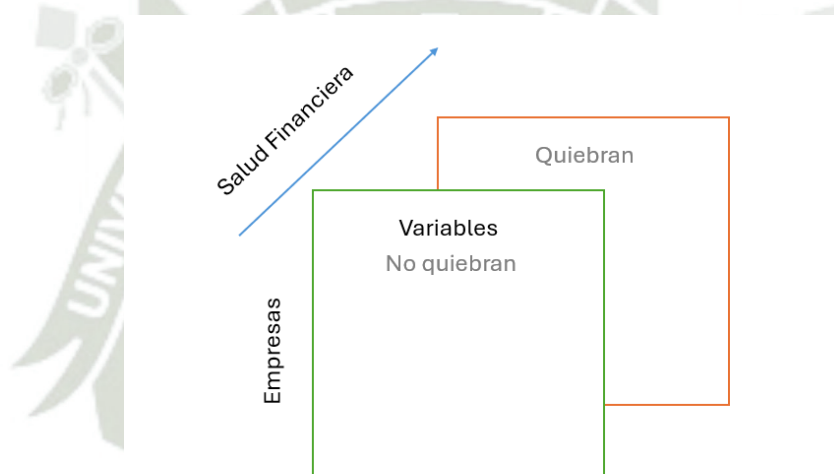
punto la carga factorial de cada dimensión, las cuales estarán ordenadas de acuerdo a la variabilidad representada que explicaría la matriz de consenso.

La puntuación factorial es utilizada para identificar las diferentes variables, lo que facilita la identificación de las diferencias o las similitudes entre estas.

Intra-estructura: Este paso tiene como objetivo examinar la estructura interna de cada matriz y la relación con el compromiso. (Gabriel & Odoroff, 1990).

Figura 4

Representación gráfica STATIS DUAL



$K = 2$ (Empresas que quiebran y no quiebran)

J = Mismas variables en la K-tablas

I = Diferentes individuos, empresas en K-tablas

Nota: *Elaboración propia basada en la información de L'Hermier des Plantes (1976).*

1.4 Técnicas, Herramientas, Normas y Estándares

Para este proyecto de investigación, se emplearon diversas técnicas, herramientas y estándares reconocidos para garantizar la calidad, consistencia y validez de los datos y resultados obtenidos. Estas metodologías permitieron organizar y analizar correctamente la información financiera proveniente del *Taiwan Economic Journal*, cumpliendo con los objetivos planteados.

1.4.1. Extracción y Transformación de Datos

Se uso Python como herramienta principal de procesamiento de datos, que permitió extraer variables financieras importantes como liquidez, rentabilidad y apalancamiento entre más indicadores importantes, después los datos fueron estructurados y procesados garantizando su calidad y preparación, siendo el objetivo principal que los datos estén listos para los modelos HJ-Biplot y STATIS Dual.

1.4.2. Validación de la Calidad de los Datos

Se usaron distintas técnicas para garantizar la calidad de los datos las cuales son:

- a) Detección de Valores Atípicos: Se aplicaron los métodos de rango intercuartílico (IQR) y cálculo de Z-score que permitió indentificar los valores fuera de rango.
- b) Normalización y Estandarización: Se normalizaron las variables usando la tecnica de Z-score garantizando que todas las escalas fueran comparables entre sí.
- c) Verificación de Consistencia: Se revisó que las variables extraídas y las fuentes tengan correspondencia correcta y no presentarán inconsistencias.
- d) Eliminación de Duplicados: Se realizó una validacion de valores duplicados para evitar redundancia y con esto garantizar la precisión los análisis posteriores.

1.4.3. Normas y Estándares

Durante el desarrollo del proyecto, se aplicaron normas y estándares ampliamente reconocidos para asegurar la calidad del trabajo realizado:

a) ISO/IEC 25012: Requisitos de Calidad de los Datos:

Garantizó que los datos fueran completos, consistentes y exactos.

b) CRISP-DM (Cross-Industry Standard Process for Data Mining):

Este estándar estructuró el proceso de análisis en las siguientes etapas:

Comprensión y preparación de datos.

Modelado mediante HJ-Biplot y STATIS Dual.

Evaluación y presentación de resultados.

c) Normas de Visualización de Datos:

Las representaciones gráficas, como mapas de calor y gráficos de componentes principales, siguieron buenas prácticas para garantizar su claridad y utilidad.

d) APA (American Psychological Association):

Se empleó el formato APA (7ª edición) para la redacción y citación de fuentes.

1.4.4. Técnicas de Análisis Multivariante:

Finalmente, la base fue limpiada y estructurada correctamente para aplicar las técnicas multivariantes HJ-Biplot y STATIS Dual, donde se exploró las relaciones entre sí, todo esto se realizó usando Python, aprovechando librerías disponibles y gráficos avanzados que facilitaron la interpretación de los resultados obtenidos.

1.5 Metodología para el desarrollo del proyecto

La metodología para el desarrollo del proyecto sigue un enfoque sistemático basado en el modelo KDD (Knowledge Discovery in Databases; Fayyad, Piatetsky-Shapiro, & Smyth, 1996),

que se estructura en etapas específicas para transformar datos en conocimiento útil. Este modelo proporciona un marco robusto para abordar el análisis multivariante y asegurar la calidad de los resultados. A continuación, se detalla cada etapa de la metodología:

1.5.1. Selección de Datos

1. Objetivo:

Identificar y seleccionar las variables financieras relevantes para el análisis de quiebras empresariales.

2. Actividades:

Extraer datos financieros del conjunto de datos *Taiwanese Bankruptcy Prediction*.
Seleccionar variables clave como Debt Ratio, Quick Ratio, Operating Profit Rate, y Net Income to Total Assets basándose en su relevancia teórica y empírica.

3. Herramientas Utilizadas:

Base de datos del *Taiwan Economic Journal*.

1.5.2. Preprocesamiento de Datos

1. Objetivo:

Garantizar la calidad y consistencia de los datos mediante la detección y corrección de valores atípicos, faltantes y redundantes.

2. Actividades:

Aplicar técnicas de imputación para tratar valores faltantes.

Identificar y tratar valores atípicos mediante el método del rango intercuartílico (IQR).

Estandarizar las variables mediante Z-score para asegurar la comparabilidad.

3. Herramientas Utilizadas:

Python: Bibliotecas como pandas y numpy para limpieza y preprocesamiento.

1.5.3. Exploración y Análisis Inicial

1. Objetivo:

Obtener una comprensión preliminar de las relaciones entre variables y posibles patrones en los datos.

2. Actividades:

Generar análisis descriptivos de las variables.

Crear visualizaciones como histogramas, diagramas de caja y gráficos de dispersión.

Construir la matriz de correlación para identificar relaciones significativas entre las variables.

3. Herramientas Utilizadas:

Python: matplotlib y seaborn para visualizaciones.

1.5.4. Minería de Datos: Métodos para la reducción de dimensión

1. Objetivo:

Aplicar técnicas de análisis multivariante para reducir la dimensionalidad, identificar agrupaciones naturales y detectar patrones clave en los datos.

2. Actividades:

Análisis de Componentes Principales (ACP): Extraer las dimensiones más relevantes para simplificar la estructura de los datos sin perder información crítica.

Clustering: Agrupar observaciones similares para revelar estructuras subyacentes en los datos.

Identificación de patrones clave: Detectar relaciones significativas y tendencias dentro de los datos utilizando técnicas exploratorias.

3. Herramientas Utilizadas:

Python: Implementación con sklearn para ACP y Clustering, junto con bibliotecas de análisis de patrones.

1.5.5. Minería de Datos: Aplicación de Métodos Multivariantes

1. Objetivo:

Aplicar técnicas avanzadas de análisis multivariante para visualizar relaciones entre variables, comparar estructuras de datos y extraer patrones significativos en el contexto del análisis de quiebra empresarial.

2. Actividades:

HJ-Biplot: Representar simultáneamente variables y observaciones en un espacio común para interpretar relaciones y asociaciones clave.

STATIS Dual: Integrar y comparar tablas de empresas en distintos estados financieros (quiebra y no quiebra) para identificar diferencias estructurales y patrones subyacentes.

3. Herramientas Utilizadas:

Python: Implementación de ACP con sklearn y desarrollo de herramientas personalizadas para HJ-Biplot y STATIS Dual.

1.5.6. Validación y Evaluación

1. Objetivo: Traducir los resultados en conclusiones compresibles y accionables.

2. Actividades:

Interpretar los patrones y asociaciones identificadas.

Relacionar los hallazgos con los objetivos planteados.

Generar gráficos y reportes que resuman los resultados clave.

3. Herramientas Utilizadas:

Python (matplotlib, seaborn) para visualizaciones y Word para documentación.

Tabla 2

Estructura Metodológica Resumida

Etapa	Actividades Principales	Resultados Esperados
Selección de Datos	Identificación y extracción de variables relevantes.	Conjunto de datos clave listo para análisis.
Preprocesamiento de Datos	Limpieza, imputación y estandarización de datos.	Datos consistentes y comparables.
Exploración y Análisis Inicial	Análisis descriptivo, visualización, y matriz de correlación.	Comprensión preliminar de patrones en los datos.
Minería de Datos	Aplicación de ACP, HJ-Biplot y STATIS Dual.	Identificación de patrones clave y diferencias estructurales.
Validación y Evaluación	Evaluación de técnicas aplicadas y comparación con estudios previos.	Resultados validados y alineados con los objetivos del proyecto.

Nota: *Elaboración propia como parte del análisis de datos financieros de empresas taiwanesas (1999-2009).*

1.6 Diseño del Proyecto

Esta sección describe cómo se diseñó el proyecto para garantizar que las técnicas multivariantes aplicadas sean efectivas en el análisis y predicción de quiebras empresariales. Incluye la estructuración de los datos, el diseño de los modelos multivariantes y la definición de criterios para evaluar la eficacia del enfoque.

1.6.1. Estructura de los Datos

1. Objetivo:

Preparar los datos de manera que sean aptos para su análisis mediante técnicas multivariantes, asegurando su calidad, consistencia y organización.

2. Pasos Realizados:

División de los Datos por Grupos:

Los datos se dividieron en dos grupos principales: empresas que quiebran y empresas que no quiebran, con el objetivo de realizar análisis comparativos. Estas tablas permiten integrar información clave para el uso de STATIS Dual y la identificación de patrones diferenciadores.

Creación de Tablas de Variables Relevantes:

Se seleccionaron métricas financieras clave como:

Rentabilidad: Operating Profit Rate, Net Income to Total Assets.

Liquidez: Quick Ratio, Current Ratio.

Apalancamiento financiero: Debt Ratio, Liability to Equity.

Estas variables se organizaron en un formato estructurado, con filas representando empresas y columnas representando las métricas financieras.

Estandarización de Variables

Todas las variables fueron estandarizadas utilizando Z-score para eliminar diferencias de escala y asegurar su comparabilidad.

1.6.2. Diseño de Modelos Multivariantes

1. Objetivo:

Definir cómo se implementarán las técnicas multivariantes seleccionadas (HJ-Biplot y STATIS Dual) para analizar las relaciones entre variables y diferenciar empresas que quiebran de aquellas que no.

2. Modelos Diseñados:

HJ-Biplot:

Propósito: Representar gráficamente las relaciones entre empresas y variables en un espacio bidimensional.

Diseño: Utilizar los dos primeros componentes principales para proyectar las observaciones y las variables.

STATIS Dual:

Propósito: Integrar y comparar tablas de empresas que quiebran y no quiebran para identificar diferencias estructurales.

Diseño: Calcular matrices de covarianza para cada grupo, combinarlas en una matriz promedio y realizar análisis comparativo.

1.6.3. Definición de Criterios de Evaluación

1. Objetivo:

Establecer métricas claras para evaluar la eficacia de las técnicas multivariantes aplicadas y garantizar que cumplan con los objetivos del análisis.

2. Criterios Definidos:

HJ-Biplot:

Varianza Explicada: Evaluar qué porcentaje de la variabilidad total es explicado por los dos primeros componentes principales.

Proximidad a Vectores: Analizar cómo las empresas se agrupan alrededor de variables clave, como Debt Ratio y Quick Ratio.

STATIS Dual:

Diferencias entre Tablas: Medir la separación entre empresas que quiebran y no quiebran mediante las coordenadas de las proyecciones.

Representación Gráfica: Evaluar la claridad en la visualización de patrones diferenciales.

1.6.4. Modelo Conceptual de Abordaje del Problema

El siguiente modelo conceptual resume de manera gráfica la estrategia metodológica para abordar el análisis de quiebra empresarial. A partir de la base de datos, se pasa por el preprocesamiento, reducción de dimensión, aplicación de técnicas multivariantes y posterior diagnóstico. Este modelo sirve como guía estructurada para el desarrollo del proyecto

Figura 5

Modelo Conceptual de Abordaje del Problema



Nota: Elaboración propia basada en el modelo conceptual de abordaje del problema.



CAPÍTULO II

2. DESARROLLO DEL PROYECTO

2.1 Selección de Datos y Preprocesamiento de Datos

2.1.1. Descripción del conjunto de datos

La matriz de datos usada para este trabajo de investigación es el Taiwan Economic Journal, que contiene información de los años 1999 y 2009. Esta matriz de datos se recopiló según las regulaciones comerciales brindadas por la Bolsa de Valores de Taiwán, que está diseñada específicamente para la predicción de quiebra empresarial, su estructura y fuente garantizan que la fuente es fiable y válida para el análisis multivariante.

2.1.1.1. Características del conjunto de datos:

Fuente: Taiwan Economic Journal (1999-2009).

Repositorio público: UCI Machine Learning Repository.

URL: <https://archive.ics.uci.edu/ml/datasets/Taiwanese+Bankruptcy+Prediction>.

Tamaño y estructura: Contiene un total de 95 columnas:

Y (Bankrupt): Etiqueta de clase binaria que indica si una empresa quebró (1) o no (0).

X1 a X95: Variables independientes relacionadas con indicadores financieros, operativos y de liquidez empresarial.

2.1.1.2. Definición de la clase objetivo

La variable dependiente Y (Bankrupt?) clasifica a las empresas en dos categorías:

1: Empresa quebrada.

0: Empresa no quebrada.

Esta variable es importante, dado que es el identificador de las empresas que quiebran y no quiebran donde nos permite revelar los patrones asociados y como correlación entre sí.

2.1.1.3. Descripción de las variables independientes:

Las variables incluyen gran información financiera, se describen las principales categorías de estas variables financieras las cuales son:

Indicadores de rentabilidad:

X1: ROA(C) antes de intereses y depreciación (Before interest and depreciation).

X2: ROA(A) antes de intereses y porcentaje después de impuestos (Before interest and % after tax).

X3: ROA(B) antes de intereses y depreciación después de impuestos (before interest and depreciation after tax).

Márgenes operativos y netos:

X4: Margen bruto operativo (Gross Profit/Net Sales).

X5: Margen bruto de ventas realizadas (Realized Gross Profit/Net Sales).

X6: Tasa de beneficio operativo (Operating Income/Net Sales).

Ratios de flujo de efectivo:

X13: Tasa de flujo de efectivo operativo (Cash Flow from Operating/Current Liabilities).

X20: Flujo de efectivo por acción (Cash Flow Per Share).

Crecimiento financiero:

X24: Tasa de crecimiento del margen bruto de ventas realizadas (Growth Rate of Realized Sales Gross Margin).

X25: Tasa de crecimiento de ingresos operativos (Operating Income Growth).

X29: Tasa de crecimiento de los activos totales (Total Asset Growth Rate).

Ratios de endeudamiento:

X36: Relación deuda total/valor neto total (Total Liability/Equity Ratio).

X37: Relación deuda/activos totales (Liability/Total Assets).

Eficiencia operativa:

X45: Rotación de activos totales (Total Asset Turnover).

X46: Rotación de cuentas por cobrar (Accounts Receivable Turnover).

Estructura financiera y liquidez:

X33: Razón corriente (Current Ratio).

X34: Razón rápida (Quick Ratio).

X55: Activos rápidos/activos totales (Quick Assets/Total Assets)

Estas variables cubren diferentes aspectos financieros, desde rentabilidad y liquidez hasta eficiencia y crecimiento empresarial, proporcionando una visión integral del estado financiero de las empresas.

El conjunto de datos es valioso para la investigación porque permite analizar en profundidad las características financieras y operativas de las empresas. Su exhaustividad en cuanto a métricas y su enfoque en predicción de quiebras lo hacen ideal para aplicar técnicas multivariantes como HJ-Biplot y STATIS Dual.

2.1.2. Preparación de los datos:

Preparar los datos es la etapa inicial que asegura que el estudio posea calidad, consistencia en los datos y la relevancia del conjunto de datos. Para esto, antes de aplicar las distintas técnicas multivariantes, se llevó a cabo varias etapas, las cuales se describen a continuación:

2.1.2.1. Limpieza de datos

Consiste en lograr identificar problemas que podrían afectar los análisis posteriores y resolverlos para estos, los realizados son los siguientes:

2.1.2.2. Identificación de valores faltantes:

Se verificó que no existan valores faltantes en cualquier columna del conjunto de datos utilizando un análisis exploratorio. Los resultados después de la validación confirman que no existen columnas con valores nulos, por lo cual no fue necesario aplicar algún método para valores ausentes.

Figura 6

Código para identificar valores faltantes.

```
# Identificar valores faltantes
missing_values = bank_data.isnull().sum()
print("Valores faltantes por columna:")
print(missing_values[missing_values > 0])
```

Nota: *Elaboración propia desarrollada en Google Colab (2024) como parte del análisis de datos financieros de empresas taiwanesas (1999–2009).*

2.1.2.3. Tratamiento de valores atípicos

Este paso importante para que podamos garantizar que las variables que van a ser usadas en los análisis posteriores no estén influenciadas por valores extremos que podrían sesgar los resultados, para esto se desarrolló las siguientes etapas:

Identificación de valores atípicos:

Se utilizó el método del rango intercuartílico (IQR) para identificar valores atípicos en las variables numéricas del conjunto de datos analizado. Este enfoque permite detectar valores que se encuentran fuera de los límites establecidos, utilizando las siguientes fórmulas:

Ecuación 4

Fórmulas de Rango Intercuartílico (IQR)

$$IQR = Q3 - Q1$$

$$\text{Límite inferior} = Q1 - 1.5 * IQR$$

$$\text{Límite superior} = Q3 + 1.5 * IQR$$

Donde:

Q1 es el primer cuartil (percentil 25)

Q3 es el tercer cuartil (percentil 75)

Al realizar los cálculos de variables atípicas, se logró indentificar las siguientes variables con valores atípicos Operating Expense Rate, Research and development expense rate, entre otras que no estan en el rango esperado.

Figura 7

Código para identificar valores atípicos.

```
# Revisar estadísticas descriptivas de todas las variables
variables_numericas = bank_data.drop(columns=['Bankrupt?']).select_dtypes(include=['float64', 'int64'])
print(variables_numericas.describe())

# Identificar variables con valores máximos significativamente altos
variables_fuera_de_rango = variables_numericas.max()[variables_numericas.max() > 1000]
print("Variables con valores extremos:")
print(variables_fuera_de_rango)
```

Nota: Elaboración propia desarrollada en Google Colab (2024) como parte del análisis de datos financieros de empresas taiwanesas (1999–2009).

En la Tabla N° 3 se muestran las variables con valores extremos que necesitan ser estandarizadas.

Tabla 3

Tabla de valores extremos

Variable	Valor Máximo
Operating Expense Rate	9.990000e+09
Research and development expense rate	9.980000e+09
Interest-bearing debt interest rate	9.900000e+08
Revenue Per Share (Yuan ¥)	3.020000e+09
Total Asset Growth Rate	9.990000e+09
Net Value Growth Rate	9.330000e+09
Current Ratio	2.750000e+09
Quick Ratio	9.230000e+09
Total debt/Total net worth	9.940000e+09
Accounts Receivable Turnover	9.740000e+09
Average Collection Days	9.730000e+09
Inventory Turnover Rate (times)	9.990000e+09

Fixed Assets Turnover Frequency	9.990000e+09
Revenue per person	8.810000e+09
Allocation rate per person	9.570000e+09
Quick Assets/Current Liability	8.820000e+09

***Nota:** Elaboración propia como parte del análisis de datos financieros de empresas taiwanesas (1999-2009).*

Una vez identificadas las variables con valores atípicos, se usó la transformación logarítmica usando la función logaritmo natural, con esto se redujo la influencia de valores extremos y nos permite realizar análisis posteriores.



Figura 7

Transformación logarítmica de valores extremos

```
# Normalizar los nombres de las columnas
bank_data.columns = bank_data.columns.str.strip() # Eliminar espacios al inicio y al final
bank_data.columns = bank_data.columns.str.replace(' ', '_') # Reemplazar espacios por guiones bajos
bank_data.columns = bank_data.columns.str.replace('()', '', regex=True) # Eliminar paréntesis

# Verificar los nombres de las columnas después de normalizar
print("Nombres de las columnas normalizadas:")
print(bank_data.columns)

# Variables identificadas con valores extremos
variables_a_transformar = [
    'Operating_Expense_Rate', 'Research_and_development_expense_rate',
    'Interest-bearing_debt_interest_rate', 'Revenue_Per_Share_Yuan_¥',
    'Total_Asset_Growth_Rate', 'Net_Value_Growth_Rate', 'Current_Ratio',
    'Quick_Ratio', 'Total_debt/Total_net_worth', 'Accounts_Receivable_Turnover',
    'Average_Collection_Days', 'Inventory_Turnover_Rate_times',
    'Fixed_Assets_Turnover_Frequency', 'Revenue_per_person',
    'Allocation_rate_per_person', 'Quick_Assets/Current_Liability',
    'Cash/Current_Liability', 'Inventory/Current_Liability',
    'Long-term_Liability_to_Current_Assets', 'Current_Asset_Turnover_Rate',
    'Quick_Asset_Turnover_Rate', 'Cash_Turnover_Rate', 'Fixed_Assets_to_Assets',
    'Total_assets_to_GNP_price'
]

# Aplicar transformación logarítmica
for col in variables_a_transformar:
    if col in bank_data.columns:
        bank_data[col] = np.log1p(bank_data[col])
    else:
        print(f"Columna '{col}' no encontrada en el DataFrame. Se omite la transformación.")

# Verificar estadísticas después de la transformación
print(bank_data[variables_a_transformar].describe())

# Revisar nuevamente variables con valores extremos
# Revisar estadísticas descriptivas de todas las variables
variables_numericas = bank_data.drop(columns=['Bankrupt?']).select_dtypes(include=['float64', 'int64'])
print(variables_numericas.describe())
```

***Nota:** Elaboración propia desarrollada en Google Colab (2024) como parte del análisis de datos financieros de empresas taiwanesas (1999–2009).*

2.2 Exploración y Análisis Inicial

Explorar el conjunto de datos con el que se va a trabajar es un paso importante para entender las características generales, lo cual nos permite identificar patrones relevantes, relaciones, posibles distribuciones o anomalías. Este proceso se separa en tres etapas, las cuales son:

2.2.1. Análisis descriptivo:

El análisis descriptivo es una etapa clave en la exploración de los datos, ya que permite comprender las principales características del conjunto de datos a través de estadísticas básicas y métricas que resumen las propiedades de las variables. En esta sección, se calculan medidas de tendencia central, dispersión, y se identifican patrones generales de los datos.

Medidas de tendencia central:

Las medidas de tendencia central resumen el valor promedio o típico de una variable. Entre las métricas calculadas se encuentran:

Media: El promedio de los valores en una variable.

Ecuación 5

Fórmula de la Media Aritmética

$$\bar{x} = \frac{\sum_{i=1}^n x_i}{n}$$

Donde:

x_i es el valor de cada observación

n es el número total de observaciones

Mediana: El valor central cuando los datos están ordenados de menor a mayor.

Moda: El valor que más se repite en una variable.

Medidas de dispersión: Las medidas de dispersión describen la variabilidad o distribución de los datos en torno a la medida:

Desviación estándar: Representa cuánto se desvían los datos de la media.

Ecuación 6

Fórmula de la desviación estándar

$$\sigma = -\sqrt{\frac{\sum_{i=1}^n (x_i - \bar{x})^2}{n}}$$

Varianza: Es el cuadrado de la desviación estándar, que mide la dispersión de los datos.

Ecuación 7

Fórmula de la varianza

$$\sigma^2 = -\sqrt{\frac{\sum_{i=1}^n (x_i - \bar{x})^2}{n}}$$

Rango: Diferencia entre el valor máximo y el mínimo.

$$Rango = \text{Máximo} - \text{Mínimo}$$

Percentiles: Valores que dividen los datos en porcentajes. Los percentiles 25 (Q1) y 75(Q3) son útiles para calcular el rango intercuartílico (IQR):

Ecuación 8

Fórmula del Rango Intercuartílico (IQR)

$$IQR = Q3 - Q1$$

Código Python para análisis descriptivo:

El análisis descriptivo se realizó utilizando las funciones estadísticas básicas de Python para enriquecer la comprensión inicial de los datos y explorar características clave como la media, mediana, desviación estándar y valores mínimos y máximos. Aunque este paso no era estrictamente necesario para el desarrollo del estudio principal, permitió obtener una

visión general de las métricas financieras, identificando posibles tendencias y anomalías que podrían influir en los análisis multivariantes posteriores.

Figura 8

Código análisis descriptivo

```
# Calcular estadísticas descriptivas básicas
estadisticas_basicas = bank_data.describe()
estadisticas_estilizadas = estadisticas_basicas.style.format("{:.2f>").set_table_styles(
    [
        {'selector': 'thead th', 'props': [('background-color', '#f5f5f5'),
            ('color', 'black'),
            ('font-weight', 'bold'),
            ('max-width', '400px')]},
        {'selector': 'tbody td', 'props': [('border', '1px solid #ccc'),
            ('text-align', 'center')]}
    ]
).set_properties(**{'text-align': 'center', 'white-space': 'nowrap'})
display(estadisticas_estilizadas)
```

***Nota:** Elaboración propia desarrollada en Google Colab (2024) como parte del análisis de datos financieros de empresas taiwanesas (1999–2009).*

Una vez ejecutado el código para un análisis descriptivo, se puede identificar los siguientes puntos para algunas variables:

Bankrup? (Variable Objetivo): Tiene una media de 0.03, lo que indica que solo el 3% de las empresas en el conjunto de datos están en bancarrota, mientras que el 97% no lo están. Esto sugiere un desequilibrio en la variable objetivo, lo que es común en problemas de clasificación binaria.

Distribución de métricas financieras: Las métricas financieras clave, como ROA(C) before interest and depreciation before interest y operating Gross Margin, están bastante concentradas alrededor de sus medias, con desviaciones estándar bajas (0.06 - 0.07). Esto indica que la mayoría de las empresas tienen un rendimiento financiero relativamente homogéneo en estas variables.

Normalización: Las variables presentan valores mínimos 0.00 y máximos 1.00 por lo realizado en el punto 2.1.2.3. *Tratamiento de valores atípicos*, esto para poder trabajar las técnicas multivariantes Hj-Biplot y Statis Dual.

Tabla 4

Tabla de análisis descriptivo

Variables	count	mean	std	min	25%	50%	75%	max
Bankrupt?	6819.00	0.03	0.18	0.00	0.00	0.00	0.00	1.00
ROAC	6819.00	0.51	0.06	0.00	0.48	0.50	0.54	1.00
ROAA	6819.00	0.56	0.07	0.00	0.54	0.56	0.59	1.00
ROAB	6819.00	0.55	0.06	0.00	0.53	0.55	0.58	1.00
Operating_Gross_Margin	6819.00	0.61	0.02	0.00	0.60	0.61	0.61	1.00
Realized_Sales_Gross_Margin	6819.00	0.61	0.02	0.00	0.60	0.61	0.61	1.00
Realized_Sales_Gross_Margin	6819.00	1.00	0.01	0.00	1.00	1.00	1.00	1.00
Pre-tax_net_Interest_Rate	6819.00	0.80	0.01	0.00	0.80	0.80	0.80	1.00
After-tax_net_Interest_Rate	6819.00	0.81	0.01	0.00	0.81	0.81	0.81	1.00

Nota: *Elaboración propia como parte del análisis de datos financieros de empresas taiwanesas (1999-2009).*

2.2.2. Visualización de datos:

La visualización de datos es una etapa clave en la exploración de datos, ya que permite observar patrones, distribuciones, relaciones y anomalías de manera visual e intuitiva. Este paso complementa el análisis descriptivo mediante gráficos que facilitan la interpretación de los datos.

Objetivos:

1. Identificar la distribución de las variables numéricas.
2. Visualizar la frecuencia de los valores mediante histogramas.
3. Detectar valores atípicos y variabilidad mediante diagramas de caja (boxplots).
4. Representar relaciones entre pares de variables mediante gráficos de dispersión.

Histogramas para Distribuciones de Variables:

Los histogramas muestran la frecuencia de los valores en una variable y ayudan a identificar distribuciones normales, sesgadas, o multimodales.

Figura 9

Código para gráfico de barras

```
# Definir el número de variables por bloque
variables_por_bloque = 10
columnas = bank_data.drop(columns=['Bankrupt?']).columns

# Dividir las variables en grupos
bloques = [columnas[i:i + variables_por_bloque] for i in range(0, len(columnas), variables_por_bloque)]

# Crear histogramas para cada grupo
for i, grupo in enumerate(bloques):
    bank_data[grupo].hist(bins=30, figsize=(20, 15))
    plt.suptitle(f"Distribución de Variables: Bloque {i+1}", fontsize=16)
    plt.show()
```

Nota: Elaboración propia desarrollada en Google Colab (2024) como parte del análisis de datos financieros de empresas taiwanesas (1999–2009).

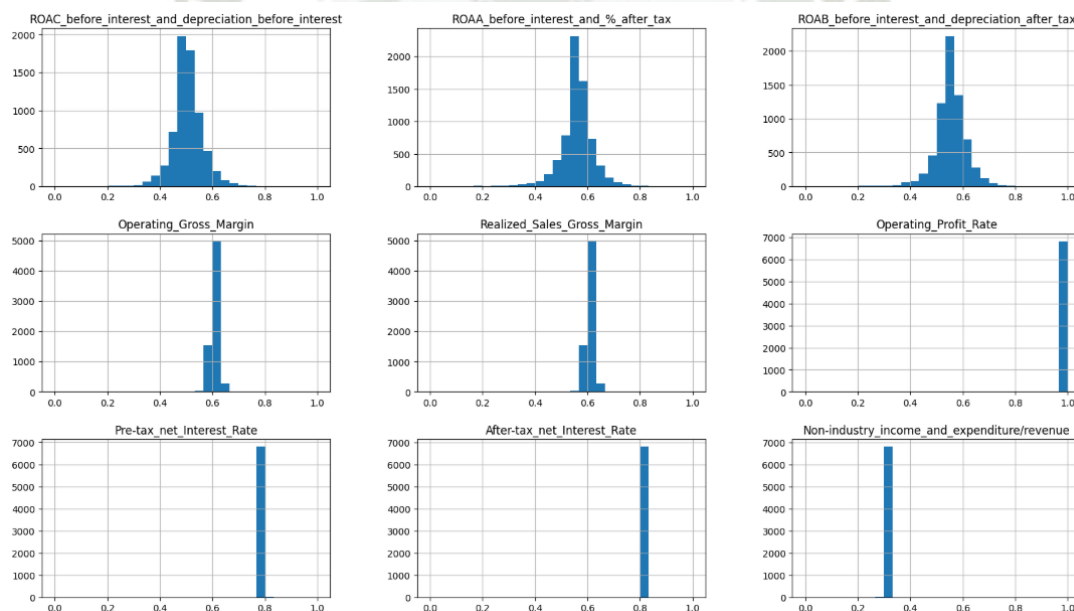
Distribución Concentrada: Variables como ROAC, ROAA y ROAB muestran una distribución concentrada cerca de un rango central (entre 0.4 y 0.6). Esto sugiere que la mayoría de las empresas tienen un rendimiento financiero similar en estas métricas.

Dispersión Baja: Las variables Operating Gross Margin y Realized Sales Gross Margin tienen distribuciones altamente concentradas alrededor de 0.6, con poca variabilidad. Esto indica un comportamiento financiero consistente entre las empresas en estas dimensiones.

Distribución Limitada: Algunas variables, como Operating Profit Rate, Pre-tax net Interest Rate, y After-tax net Interest Rate, muestran un patrón donde la mayoría de los valores están cercanos al extremo superior (1.0).

Figura 10

Histogramas para distribuciones de variables.



Nota: Elaboración propia basada en datos financieros de empresas taiwanesas (1999–2009), analizados en Google Colab (2024).

Diagramas de Caja (Boxplots):

Los diagramas de caja son útiles para identificar valores atípicos y observar la dispersión de las variables.

Figura 11

Código para diagrama de caja

```
# Seleccionar las variables para los boxplots
variables_seleccionadas = [
    'ROAC_before_interest_and_depreciation_before_interest', 'ROAA_before_interest_and_%_after_tax',
    'Continuous_interest_rate_after_tax', 'Tax_rate_A',
    'Debt_ratio_%'
]

variables_por_bloque = 10
bloques = [variables_seleccionadas[i:i + variables_por_bloque]
            for i in range(0, len(variables_seleccionadas), variables_por_bloque)]

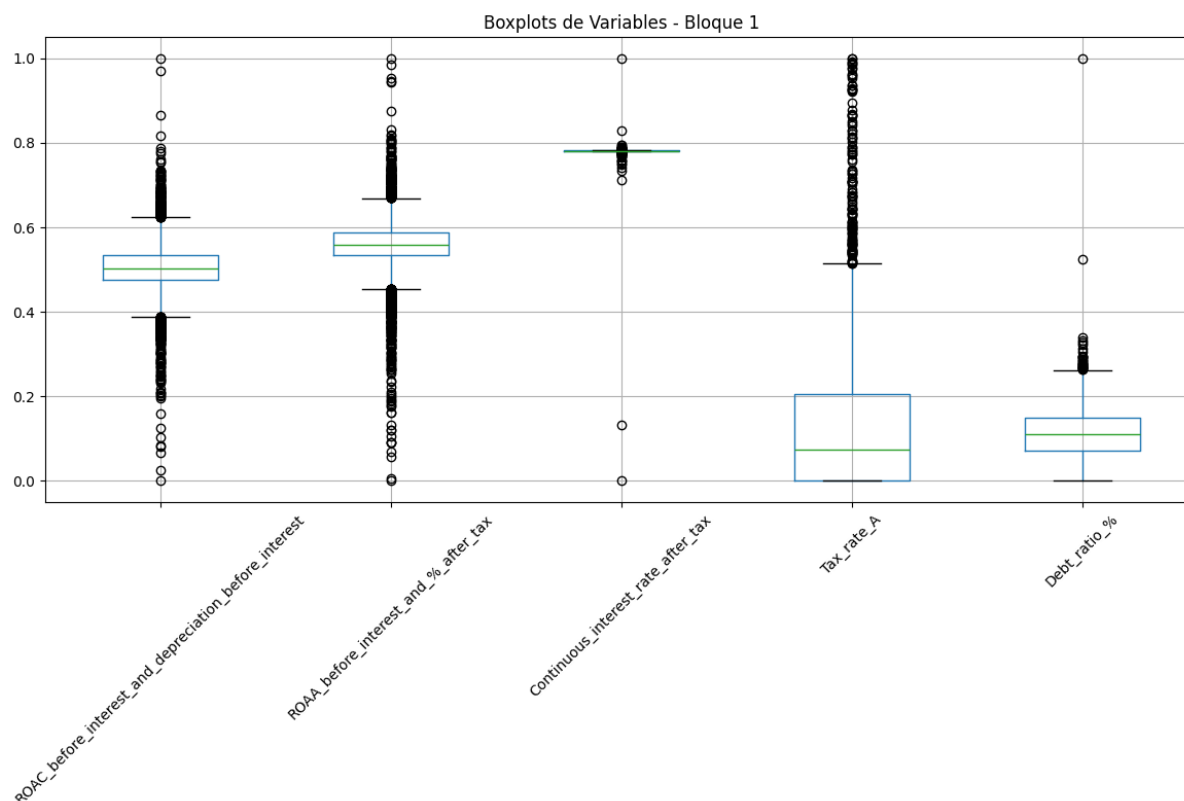
for i, grupo in enumerate(bloques):
    plt.figure(figsize=(12, 8))
    bank_data[grupo].boxplot()
    plt.title(f"Boxplots de Variables - Bloque {i+1}")
    plt.xticks(rotation=45, fontsize=10)
    plt.grid(True)
    plt.tight_layout()
    plt.show()
```

Nota: Elaboración propia desarrollada en Google Colab (2024) como parte del análisis de datos financieros de empresas taiwanesas (1999–2009).

Distribución Central: Las variables mostradas en la figura 11, se puede apreciar que las medianas están en el rango 0.5, lo cual se interpreta que la mitad de las empresas reflejan un rendimiento financiero moderado, también se logra apreciar que otras variables muestran las medianas más bajas alrededor del 0.2, lo que se interpreta que la mayoría de las empresas tienen tasas impositivas bajas si son comparadas con los ingresos en proporción, también se ven valores atípicos en ambos extremos, lo que se interpreta que existen empresas con niveles de deuda extremadamente bajas o altas en comparación a la mediana, con esto se proporciona una visión preliminar de las características más resaltantes.

Figura 12

Diagrama de cajas



Nota: Elaboración propia basada en datos financieros de empresas taiwanesas (1999-2009), analizados en Google Colab (2024).

2.2.3. Matriz de correlación e interpretación:

La matriz de correlación tiene como finalidad mostrar las relaciones lineales que existen entre las variables de la matriz de datos, con esto se observa la fuerza y la dirección de estas relaciones, donde nos permite visualizar las relaciones más notorias, posibles patrones y variables relevantes para este análisis, esta correlación mide la relación lineal entre dos variables y los resultados esperados pueden oscilar entre -1 y 1.

Ecuación 9

Fórmula de correlación de Pearson

$r > 0$: *Relación positiva (aumentan juntas)*

$r < 0$: *Relación negativa (una aumenta mientras otras disminuye)*

$r = 0$: *Sin relación lineal*

$$r = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^n (x_i - \bar{x})^2} * \sqrt{\sum_{i=1}^n (y_i - \bar{y})^2}}$$

Donde:

x_i, y_i : *Valore de las variables.*

$\bar{x}, \bar{y}$: *Promedio de las variables.*

Generar la matriz de correlación:

En la figura 12 se muestra el código usado para generar la matriz de correlación, que muestra las correlaciones entre las variables numéricas de la matriz de datos.

Figura 13

Código matriz de correlación

```
correlacion = bank_data.corr()
plt.figure(figsize=(12, 10))
sns.heatmap(correlacion, cmap="coolwarm", annot=False, fmt=".2f", linewidths=0.5)
plt.title("Matriz de Correlación", fontsize=16)
plt.show()
```

Nota: *Elaboración propia desarrollada en Google Colab (2024) como parte del análisis de datos financieros de empresas taiwanesas (1999–2009).*

La matriz de correlación mostrada en la figura 13 utiliza un mapa de calor para mostrar las relaciones lineales entre las variables del conjunto de datos. Cada celda indica el coeficiente

de correlación (r) entre dos variables, con colores que representan la magnitud y dirección de la relación:

Rojo oscuro ($r=1$): Correlación positiva fuerte.

Azul oscuro ($r=-1$): Correlación negativa fuerte.

Gris claro ($r=0$): Sin correlación.

Relaciones Positivas Fuertes ($r>0.8$): Las variables con alta correlación positiva tienen valores que tienden a aumentar juntas. Por ejemplo; Net Value Per Share B y Net Value Per Share A se observa que tienen una correlación cercana a 1, lo que indica que son casi equivalentes o posiblemente redundantes. De igual forma, Operating Profit Per Share y Gross Margin también presentan una correlación fuerte, sugiriendo que ambas miden lo mismo en el rendimiento operativo.

Relaciones Negativas Fuertes ($r<-0.8$): Las variables con alta correlación negativa significativa que tienden a moverse en direcciones opuestas, por ejemplo; Debt Ratio % y Net Worth/Assets lo cual es lógico desde una perspectiva financiera, ya que un mayor ratio de deuda suele reducir el patrimonio neto en proporción de los activos.

Variables con Baja Correlación ($r\approx 0$): Muchas variables no tienen relaciones significativas entre sí, como se observa en las celdas de color gris claro, esto nos indica independencia lineal entre estas variables.

Una vez analizado este gráfico podemos tomar la decisión de reducir variables a analizar dado que las variables altamente relacionadas pueden simplificarse seleccionando solo una en cada grupo redundante.

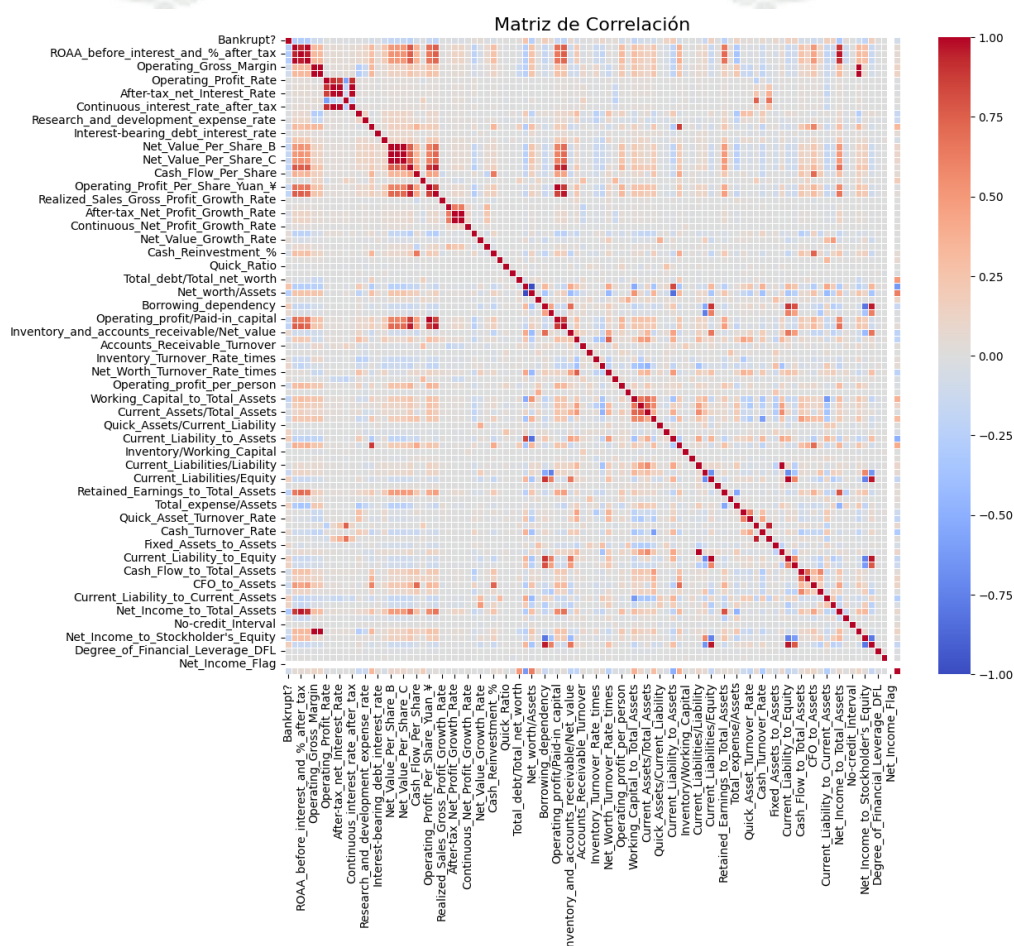
Las variables con correlación baja pueden proporcionar información complementaria y deben considerarse para el análisis multivariante del Hj-Biplot.

La validación de las relaciones observadas confirma las relaciones esperadas entre variables financieras clave, como ejemplo entre la rentabilidad y el crecimiento.

Otro punto que resaltar es que, antes de aplicar el HJ-Biplot, es importante eliminar o transformar variables redundantes para evitar distorsiones en los resultados.

Figura 14

Matriz de correlación



Nota: Elaboración propia basada en datos financieros de empresas taiwanesas (1999-2009), analizados en Google Colab (2024).

2.3 Minería de Datos: Métodos para la reducción de dimensión

2.3.1. Análisis de Componentes Principales (ACP)

El Análisis de Componentes Principales (ACP) es una técnica de reducción de dimensionalidad que transforma un conjunto de variables originales en un conjunto más pequeño de variables no correlacionadas, llamadas componentes principales. Este método es útil para identificar patrones en datos de alta dimensionalidad y simplificar su representación sin pérdida significativa de información.

Objetivo del ACP:

1. Reducir la dimensionalidad del conjunto de datos.
2. Identificar las combinaciones lineales de variables originales que explican la mayor parte de la variabilidad en los datos.
3. Visualizar las relaciones entre las variables y las observaciones en un espacio de menor dimensión.

Figura 15

Código para ACP

```
# Aplicar ACP
pca = PCA()
componentes_principales = pca.fit_transform(datos_estandarizados)

# Explicar la varianza acumulada
varianza_explicada = pca.explained_variance_ratio_
varianza_acumulada = varianza_explicada.cumsum()

# Crear un DataFrame con la varianza explicada y acumulada
df_varianza = pd.DataFrame({
    'Componente': [f'PC{i+1}' for i in range(len(varianza_explicada))],
    'Varianza Explicada (%)': varianza_explicada * 100,
    'Varianza Acumulada (%)': varianza_acumulada * 100
})

# Mostrar el DataFrame en formato estilizado
from IPython.display import display

df_varianza_estilizado = df_varianza.head(10).style.format({
    'Varianza Explicada (%)': '{:.2f}',
    'Varianza Acumulada (%)': '{:.2f}'
}).set_table_styles([
    {'selector': 'thead th', 'props': [('background-color', '#f5f5f5'), ('color', 'black'), ('font-weight', 'bold')]},
    {'selector': 'tbody td', 'props': [('border', '1px solid #ccc'), ('text-align', 'center')]}
])

display(df_varianza_estilizado)
```

Nota: Elaboración propia desarrollada en Google Colab (2024) como parte del análisis de datos financieros de empresas taiwanesas (1999–2009).

En figura 18 podemos observar lo siguiente:

Varianza Explicada por Componente:

PC1 (13.35%): Es el componente que más contribuye a la varianza en los datos. Esto sugiere que hay un patrón dominante que agrupa la mayoría de las variables originales.

PC2 (7.21%): Explica menos varianza que PC1 pero aún tiene una contribución significativa.

PC3 a PC10: Los componentes posteriores explican una proporción decreciente de la varianza. Esto es común en el ACP, donde los primeros componentes capturan los patrones más importantes.

Varianza Acumulada:

Primeros 2 Componentes (20.56%): Capturan aproximadamente el 20.56% de la varianza, lo que indica que aún queda mucha variabilidad no explicada.

Primeros 5 Componentes (34.56%): Al incluir los primeros cinco componentes, se explica el 34.56% de la varianza total.

Primeros 10 Componentes (48.72%): Al incluir los primeros diez componentes, se captura el 48.72% de la variabilidad, casi la mitad del total. El cuadro resultante sugiere que los 10 primeros componentes explican casi el 50% de los datos de la variabilidad. Dependiendo del nivel de precisión que se puede entregar, se tomarán como base en lugar de tomar todas las variables originales.

Tabla 5

Tabla de análisis de Componente Principales (ACP)

<i>Componente</i>	<i>Varianza Explicada (%)</i>	<i>Varianza Acumulada (%)</i>
<i>PC1</i>	<i>13.35</i>	<i>13.35</i>
<i>PC2</i>	<i>7.21</i>	<i>20.56</i>
<i>PC3</i>	<i>5.04</i>	<i>25.60</i>
<i>PC4</i>	<i>4.72</i>	<i>30.32</i>
<i>PC5</i>	<i>4.23</i>	<i>34.56</i>
<i>PC6</i>	<i>3.22</i>	<i>37.78</i>
<i>PC7</i>	<i>3.05</i>	<i>40.83</i>
<i>PC8</i>	<i>2.93</i>	<i>43.76</i>

Nota: *Elaboración propia como parte del análisis de datos financieros de empresas taiwanesas (1999-2009).*

Selección del Número de Componentes:

En la figura 15 se muestra el código para graficar la varianza explicada acumulada, lo cual nos brinda una forma gráfica de cuántas componentes principales se pueden tomar en base a la regla de codo o experiencia adquirida.

Figura 16

Código de varianza explicada acumulada

```
# Graficar la varianza explicada acumulada
plt.figure(figsize=(8, 5))
plt.plot(range(1, len(varianza_acumulada) + 1), varianza_acumulada, marker='o')
plt.xlabel("Número de Componentes")
plt.ylabel("Varianza Explicada Acumulada")
plt.title("Selección del Número de Componentes Principales")
plt.grid(True)
plt.show()
```

***Nota:** Elaboración propia desarrollada en Google Colab (2024) como parte del análisis de datos financieros de empresas taiwanesas (1999–2009).*

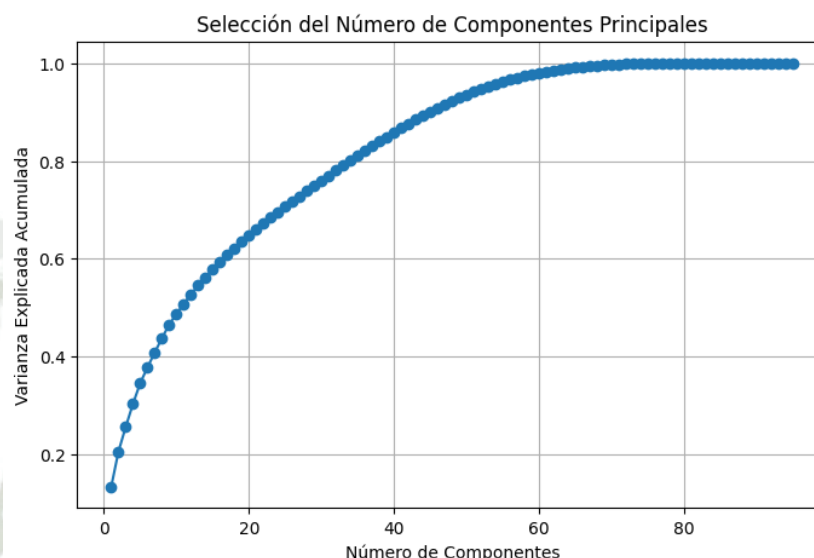
Región Inicial (Primeros Componentes): En la figura 16 los primeros componentes principales contribuyen significativamente a la varianza explicada. Se observa una curva pronunciada en esta región, lo que indica que los primeros componentes contienen la mayor parte de la información relevante de los datos.

Región de Disminución (Regla del Codo): A medida que se agregan más componentes, la tasa de aumento en la varianza explicada acumulada disminuye. Esto se observa como una "curva de saturación" en el gráfico. El punto donde la curva comienza a aplanarse (conocido como el "codo") indica que los componentes adicionales contribuyen muy poco a la varianza explicada. Este es el número óptimo de componentes principales que debería seleccionarse.

Región Final (Componentes Menores): Los componentes adicionales explican poca varianza y contribuyen marginalmente a la acumulación total. En esta región, incluir más componentes no agrega valor significativo y podría introducir ruido.

Figura 17

Gráfico de varianza acumulada explicada.



Nota: Elaboración propia basada en datos financieros de empresas taiwanesas (1999-2009), analizados en Google Colab (2024).

Interpretación de los Componentes Principales:

El análisis muestra que los primeros componentes principales capturan un porcentaje significativo de la varianza total en los datos:

ACP1 explica el 13.35% de la varianza.

ACP2 aporta un 7.21% adicional.

Los primeros 10 componentes explican en conjunto el 48.72% de la variabilidad total.

Esto indica que, aunque se capturan patrones importantes con los primeros componentes, se necesita un mayor número de componentes para representar completamente la variabilidad en los datos.

Cargas de las Variables Originales: Las cargas factoriales indican cómo las variables originales contribuyen a cada componente principal. Estas cargas permiten interpretar el significado de cada componente, identificando qué variables tienen mayor influencia en su construcción.

Tabla 6

Tabla de componentes principales y variables dominantes

Componente	Nombre	Variables Dominantes
ACP1	Diversos	ROA(A), Persistent EPS, ROA(C), ROA(B), Net profit before tax/Paid-in capital
ACP2	Liquidez	Current Liabilities/Equity, Current Liability to Equity, Liability to Equity, Quick Assets/Total Assets, Current Liabilities/Liability
ACP3	Liquidez	Current Assets/Total Assets, Current Liability to Assets, Total Asset Turnover, Working Capital to Total Assets, Cash/Total Assets
ACP4	Gestión de Capital de Trabajo	Working Capital to Total Assets, Cash/Total Assets, Quick Assets/Total Assets, Current Liability to Equity, Quick Ratio

ACP5	Rentabilidad Financiera	Pre-tax net Interest Rate, Continuous interest rate (after tax), After-tax net Interest Rate, Operating Profit Rate, Gross Margin
ACP6	Gestión de Flujos de Efectivo	Cash Reinvestment %, CFO to Assets, Cash flow rate, Cash Flow to Sales, Operating Expense Rate
ACP7	Eficiencia Operativa	Working Capital Turnover Rate, Cash Flow to Sales, Total Asset Turnover, Inventory Turnover, Accounts Receivable Turnover
ACP8	Rentabilidad Operativa	Operating Gross Margin, Gross Profit to Sales, Realized Sales Gross Margin, Net Profit to Equity, Net Profit Before Tax
ACP9	Crecimiento Financiero	After-tax Net Profit Growth Rate, Regular Net Profit Growth Rate, Operating Profit Growth Rate, Revenue Growth Rate, Total Asset Growth Rate
ACP10	Costos Operativos	Operating Expense Rate, Quick Ratio, Revenue per person, Cash Flow Rate, Interest Expense Rate

***Nota:** Elaboración propia como parte del análisis de datos financieros de empresas taiwanesas (1999-2009).*

Justificación de cada componente:

ACP1: Diversos

Variables Dominantes:

ROA(A) before interest and % after tax

Persistent EPS in the Last Four Seasons

ROA(C) before interest and depreciation before interest

ROA(B) before interest and depreciation after tax

Net profit before tax/Paid-in capital

Justificación: Este componente refleja cómo las empresas dan valor desde sus recursos financieros y operativos, resaltando valores sobre activos, márgenes operativos y eficiencia general a pesar de ello no presentan un patrón claro que permita clasificarlas en una sola categoría.

ACP2: Liquidez

Variables Dominantes:

Current Liabilities/Equity

Current Liability to Equity

Liability to Equity

Quick Assets/Total Assets

Current Liabilities/Liability

Justificación: Este componente refleja la solvencia inmediata de las empresas, donde se mide la capacidad de gestionar obligaciones financieras a corto plazo, donde las variables que dominan son las de activos líquidos y pasivos, donde se relacionan la deuda y patrimonio que nos dan una medida de salud y estabilidad financiera.

ACP3: Liquidez

Variables Dominantes:

Current Assets/Total Assets

Current Liability to Assets

Total Asset Turnover

Working Capital to Total Assets

Cash/Total Assets

Justificación: Este componente refleja la iteración de los activos líquidos frente a los activos totales, donde se resalta cómo se equilibran los recursos financieros y operativos, donde también se observan métricas de rotación que nos indican la eficacia de activos clave para generar ingresos.

ACP4: Gestión de Capital de Trabajo

Variables Dominantes:

Working Capital to Total Assets

Cash/Total Assets

Quick Assets/Total Assets

Current Liability to Equity

Quick Ratio

Justificación: Este componente refleja la gestión del capital de trabajo y como se distribuyen los activos circulantes con el objetivo de financiar las operaciones y cumplir sus obligaciones financieras, en resumen, este acp refleja como se adaptan a corto plazo.

ACP5: Rentabilidad Financiera

Variables Dominantes:

Pre-tax net Interest Rate

Continuous interest rate (after tax)

After-tax net Interest Rate

Operating Profit Rate

Gross Margin

Justificación: Este componente refleja como se generan beneficios netos y operativos, donde se reflejan los ingresos y los márgenes generados, este acp muestra la sostenibilidad económica y el rendimiento operativo.

ACP6: Gestión de Flujos de Efectivo

Variables Dominantes:

Cash Reinvestment %

CFO to Assets

Cash flow rate

Cash Flow to Sales

Operating Expense Rate

Justificación: Este componente refleja como se reinvierten los flujos que se generan, esto para los activos y operaciones, esto muestra como es la eficacia en el uso de los recursos financieros, reflejados en los desafíos económicos.

ACP7: Eficiencia Operativa

Variables Dominantes:

Working Capital Turnover Rate

Cash Flow to Sales

Total Asset Turnover

Inventory Turnover

Accounts Receivable Turnover

Justificación: Este componente muestra la correcta gestión de los activos inventarios y las cuentas por cobros, esto muestra la capacidad de maximizar la productividad y como se usa los recursos de manera adecuada.

ACP8: Rentabilidad Operativa

Variables Dominantes:

Operating Gross Margin

Gross Profit to Sales

Realized Sales Gross Margin

Net Profit to Equity

Net Profit Before Tax

Justificación: Este componente refleja cómo las empresas generan beneficios que se pueden sostener desde sus principales operaciones, donde las variables que predominan son los márgenes operativos y las tasas de ganancia, donde se visualiza en conjunto el rendimiento económico.

ACP9: Crecimiento Financiero

Variables Dominantes:

After-tax Net Profit Growth Rate

Regular Net Profit Growth Rate

Operating Profit Growth Rate

Revenue Growth Rate

Total Asset Growth Rate

Justificación: En este componente se refleja el crecimiento financiero y crecimiento operativo de las empresas, donde las variables que predominan se relacionan con los ingresos, activos y los beneficios netos, donde se visualizan patrones de expansión y mejora económica en el tiempo.

ACP10: Costos Operativos

Variables Dominantes:

Operating Expense Rate

Quick Ratio

Revenue per person

Cash Flow Rate

Interest Expense Rate

Justificación: En este componente se refleja la estructura de los costos operativos y los costos financieros, donde se evalúa la eficacia de los gastos asociados a operaciones, investigaciones y financiamiento, esto refleja las áreas de optimización para mejorar la rentabilidad.

2.3.2. Análisis de Clustering

En análisis de clustering es usado en etapas exploratorias, usado para detectar patrones y similitudes, para lo cual se agrupa en cluster homogéneo. Para esto se usó en segmentar las empresas del estudio según las métricas financieras reportadas, lo cual nos facilita el poder identificar comportamientos comunes y factores que influyen en la quiebra financiera.

Para este estudio, se tomaron las variables más importantes y luego se estandarizaron usando la técnica de Z-score.

Ecuación 10

Fórmula del Z-Score (Puntaje Z)

$$Z_i = \frac{x_i - \mu}{\sigma}$$

Donde:

x_i valor de la observación

μ media de la variable

σ desviación estandar de la variable

Este proceso asegura que todas las variables contribuyan equitativamente al cálculo de las distancias, eliminando el sesgo causado por diferencias en escalas. También los valores atípicos fueron tratados para reducir el posible impacto en los resultados finales.

Se usó el algoritmo K-Means para la segmentación de datos grandes, este algoritmo funciona iterativamente para minimizar la suma de las distancias al cuadrado entre cada observación y el centroide del clúster al que pertenece:

Ecuación 11

Fórmula de la Inercia Intra-Clúster (Suma de Distancias Cuadradas)

$$j = \sum_{k=1}^k \sum_{i \in c_k} \|x_i - \mu_k\|^2$$

Donde:

J: Función objetivo (inercia intra – cluster o WCSS).

K: Número de clusters

C_k: Conjunto de observaciones asignadas al cluster k

x_i: Observación i

μ_k: Centroide del cluster k

||x_i – μ_k||²: Distancia euclidiana al cuadrado entre la observación y el centroide

Este método busca minimizar la suma de las distancias al cuadrado entre cada observación y el centroide del clúster al que pertenece. Para determinar el número óptimo de clusters, se aplicó el método del codo, evaluando la relación entre la inercia intra-cluster (WCSS) y el número de clusters.

El análisis de clustering, utilizando el algoritmo K-Means, agrupó a las empresas en cuatro clusters principales basados en métricas financieras clave. A continuación, se describe cada cluster en detalle, destacando sus características principales y su interpretación financiera.

Figura 18

Código K-means

```
# Definir el número de clusters para el modelo K-Means
n_clusters = 4 # Número de grupos en los que se segmentarán las empresas
kmeans = KMeans(n_clusters=n_clusters, random_state=42)

# Ajustar el modelo y predecir los clusters
clusters = kmeans.fit_predict(datos_estandarizados)

# Agregar la información de los clusters al DataFrame original
bank_data['Cluster'] = clusters

# Visualización del clustering en un gráfico de dispersión
plt.figure(figsize=(10, 6))
sns.scatterplot(
    x=datos_estandarizados[:, 0],
    y=datos_estandarizados[:, 1],
    hue=clusters,
    palette='viridis',
    style=clusters
)
plt.title('Clustering de Empresas (K-Means)')
plt.xlabel('Componente Principal 1')
plt.ylabel('Componente Principal 2')
plt.legend(title='Cluster')
plt.grid()
plt.show() |
```

***Nota:** Elaboración propia desarrollada en Google Colab (2024) como parte del análisis de datos financieros de empresas taiwanesas (1999–2009).*

Interpretación de la figura 18:

Cluster 0: Empresas Altamente Estables

Características Clave:

ROA(C) before interest and depreciation before interest: 0.581

Net Income to Total Assets: 0.849

Liability to Equity: 0.278

Net Income to Stockholder's Equity: 0.843

Interpretación: Las empresas representadas en este cluster son las que tienen alta rentabilidad y liquidez, que refleja que tienen un retorno fuerte sobre sus activos y patrimonio, el bajo nivel de la deuda también indica una buena gestión, donde estas empresas son estables y resilientes.

Cluster 1: Empresas Moderadamente Estables

Características Clave:

ROA(C) before interest and depreciation before interest: 0.453

Net Income to Total Assets: 0.774

Liability to Equity: 0.285

Net Income to Stockholder's Equity: 0.838

Interpretación: Las empresas representadas en este cluster tienen una estabilidad financiera equilibrada, donde presentan rentabilidad y liquidez importantes, aunque no tan sólidos como el cluster 0 aunque sigue mostrando una estructura financiera sólida, este grupo son empresas estables con oportunidades de mejora.

Cluster 2: Empresas de Alto Riesgo Financiero

Características Clave:

ROA(C) before interest and depreciation before interest: 0.333

Net Income to Total Assets: 0.657

Liability to Equity: 0.760

Net Income to Stockholder's Equity: 0.384

Interpretación: Las empresas representadas en este cluster son las que tienen mayores riesgos financieros, su baja rentabilidad y su alta deuda muestran una vulnerabilidad

económica significativa. Este grupo es propenso a enfrentar problemas financieros y requiere intervención para reducir el apalancamiento.

Cluster 3: Empresas de Desempeño Medio

Características Clave:

ROA(C) before interest and depreciation before interest: 0.509

Net Income to Total Assets: 0.812

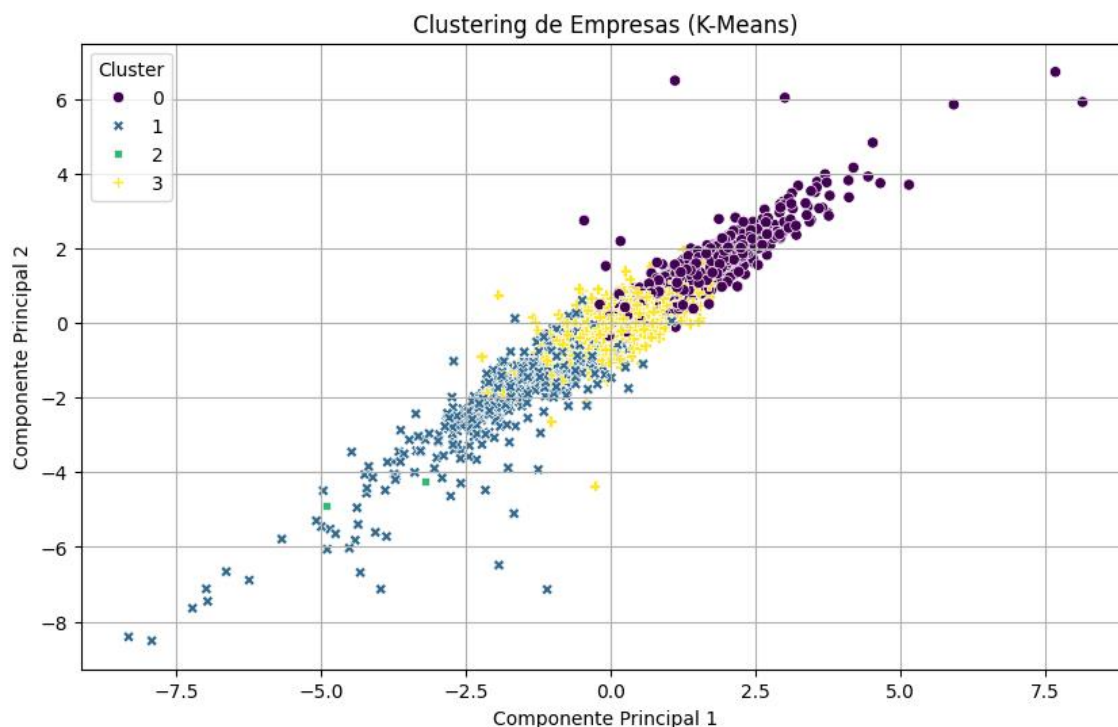
Liability to Equity: 0.279

Net Income to Stockholder's Equity: 0.841

Interpretación: Las empresas representadas en este cluster tienen una posición financiera equilibrada, es similar al cluster 1 con un rendimiento financiero relativamente superior, donde estas empresas tienen rentabilidad sostenible y gestionan su deuda de manera correcta.

Figura 19

Clustering de Empresas



Nota: *Elaboración propia basada en datos financieros de empresas taiwanesas (1999-2009), analizados en Google Colab (2024).*

2.3.3. Identificación de patrones clave

Lograr identificar patrones clave en base a los resultados de análisis de componentes principales (ACP) y el análisis cluster nos brinda una visión amplia de las relaciones, en este análisis resaltan los valores importantes que definen los grupos y hacen que se diferencien según su desempeño, gestión y riesgo financiero.

El análisis de componentes principales nos permite reducir las dimensiones y resalta los factores importantes donde se explica la mayor parte de la variabilidad.

Patrones Clave Identificados desde el ACP

Componentes Principales (ACP1 y ACP2):

ACP1: Asociado a la rentabilidad financiera y operativa, destacando variables como ROA(A) before interest and % after tax, Persistent EPS in the Last Four Seasons, ROA(C) before interest and depreciation before interest, ROA(B) before interest and depreciation after tax y Net profit before tax/Paid-in capital

Patrón Identificado: Este componente explica cómo las empresas generan beneficios a partir de sus operaciones principales.

ACP2: Relacionado con la liquidez y la eficiencia operativa, con métricas como Current Liabilities/Equity, Current Liability to Equity, Liability to Equity, Quick Assets/Total Assets, Current Liabilities/Liability.

Patrón Identificado: Este componente refleja las empresas que gestionan sus recursos líquidos y activos.

Patrones Clave Identificados desde el Clustering

El análisis clúster basado en las proyecciones brindadas por el análisis de componentes principales (ACP), agrupa las empresas en 4 clusters principales, donde cada cluster resalta características financieras importantes que muestran el comportamiento y los riesgos.

Clúster 0: Empresas Altamente Estables

Patrón Identificado:

Alta rentabilidad (Net Income to Total Assets 0.849).

Baja deuda (Liability to Equity 0.278).

Alta liquidez (ROA(C) 0.581).

Interpretación: Estas empresas son financieramente sólidas, con bajo apalancamiento y un desempeño financiero destacado.

Clúster 1: Empresas Moderadamente Estables

Patrón Identificado:

Rentabilidad moderada (Net Income to Total Assets 0.774).

Baja deuda (Liability to Equity 0.285).

Estabilidad operativa (ROA(C) 0.453).

Interpretación: Estas empresas tienen una posición financiera equilibrada, con espacio para optimización en su gestión de recursos.

Clúster 2: Empresas de Alto Riesgo Financiero

Patrón Identificado:

Baja rentabilidad (Net Income to Total Assets 0.657).

Alta deuda (Liability to Equity 0.760).

Baja liquidez (ROA(C) 0.333).

Interpretación: Este grupo enfrenta altos riesgos financieros debido a una combinación de baja rentabilidad y alta dependencia del apalancamiento.

Clúster 3: Empresas de Desempeño Medio

Patrón Identificado:

Buena rentabilidad (Net Income to Total Assets 0.812).

Baja deuda (Liability to Equity 0.279).

Sólida liquidez (ROA(C) 0.509).

Interpretación: Estas empresas tienen un desempeño equilibrado, con niveles sostenibles de rentabilidad y una estructura financiera estable.

Relación entre el ACP y el Clustering.

Complementariedad de Métodos:

El ACP destacó las dimensiones clave (rentabilidad, liquidez, eficiencia) que explican la variabilidad en los datos.

El clustering, basado en estas proyecciones, segmentó las empresas en grupos homogéneos.

Validación de Patrones:

Las variables dominantes en el ACP (como Operating Profit Rate y Quick Ratio) coinciden con las características promedio destacadas en los clusters, confirmando la relevancia de las métricas seleccionadas.

Visualización y Separación de Clusters:

La proyección en el espacio ACP1-ACP2 mostró una separación clara de los clusters, destacando las diferencias en liquidez, rentabilidad y apalancamiento.

2.4 Minería de Datos: Aplicación de Métodos Multivariantes

2.4.1. HJ-Biplot: Análisis y visualización de relaciones

El HJ-Biplot es una técnica multivariante que combina el análisis de componentes principales (ACP) y su representación gráfica en un espacio reducido, lo cual nos permite analizar de manera simultánea las observaciones (empresas) y las variables que nos proporciona una interpretación clara y cómo se relacionan entre sí, para este trabajo se usó en explorar conexiones entre métricas financieras importantes y empresas agrupadas en clúster, identificando patrones importantes.

Objetivos del HJ-Biplot

1. Representar las relaciones entre variables:

Determinar qué variables están correlacionadas o tienen un impacto dominante.

2. Visualizar las relaciones entre empresas (observaciones):

Identificar similitudes y diferencias en función de las métricas financieras.

3. Explorar la asociación entre empresas y variables:

Relacionar empresas específicas con características financieras clave.

Metodología

Construcción del HJ-Biplot:

Desde el análisis de los componentes principales donde se proyectan las observaciones (empresas) y las variables en el mismo espacio.

Las observaciones se representan como puntos y las variables como vectores.

Interpretación del Gráfico:

Dirección de los vectores: Indica la correlación entre las variables. Vectores cercanos apuntando en la misma dirección están correlacionados positivamente.

Longitud de los vectores: Representa la varianza explicada por la variable. Vectores más largos indican mayor contribución.

Proximidad entre puntos y vectores: Muestra qué observaciones tienen altos valores en una variable específica.

Figura 20

Código para generar HJ-Biplot variables.

```
# Crear un gráfico combinado con observaciones y variables, ajustando el tamaño del texto y uniendo los gráficos
plt.figure(figsize=(12, 8))

# Graficar las observaciones (puntos)
plt.scatter(
    componentes_principales[:, 0],
    componentes_principales[:, 1],
    c=bank_data['Cluster'],
    cmap='viridis',
    s=50,
    alpha=0.6,
    label='Observaciones'
)

# Graficar las variables (vectores)
for i, (vector_x, vector_y) in enumerate(zip(pca.components_[0], pca.components_[1])):
    plt.arrow(
        0, 0,
        vector_x * 5,
        vector_y * 5,
        color='r',
        alpha=0.6,
        head_width=0.1,
        linewidth=0.8
    )
    # Reducir el tamaño del texto y la superposición
    plt.text(
        vector_x * 5.2,
        vector_y * 5.2,
        variables.columns[i],
        color='r',
        fontsize=8,
        alpha=0.8
    )

# Etiquetas y formato
plt.title('HJ-Biplot: Observaciones y Variables', fontsize=14)
plt.xlabel(f'Componente Principal 1 ({pca.explained_variance_ratio_[0]*100:.2f}%)', fontsize=12)
plt.ylabel(f'Componente Principal 2 ({pca.explained_variance_ratio_[1]*100:.2f}%)', fontsize=12)
plt.legend(title='Clusters', loc='upper right')
plt.grid()
```

***Nota:** Elaboración propia desarrollada en Google Colab (2024) como parte del análisis de datos financieros de empresas taiwanesas (1999–2009).*

Relación entre las Variables

En un HJ-Biplot, las variables están representadas por vectores que parten del origen (0,0).

La dirección y longitud de estos vectores proporcionan información clave sobre las relaciones entre las variables.

Variables Correlacionadas Positivamente

- **Interpretación:** Los vectores que apuntan en la misma dirección indican una correlación positiva entre las variables.
- **Ejemplo en el gráfico 20:**
Operating Profit Rate, Net Income to Total Assets, y ROA(C) apuntan en direcciones similares. Esto indica que estas variables están correlacionadas positivamente; es decir, cuando una aumenta, las otras tienden a aumentar también.
- **Implicación:** Las empresas con altos márgenes operativos tienden a ser más rentables en términos de sus activos.

Variables Correlacionadas Negativamente

- **Interpretación:** Los vectores que apuntan en direcciones opuestas indican una correlación negativa.
- **Ejemplo en el gráfico 20:**
Debt Ratio apunta en la dirección opuesta a Net Income to Stockholder's Equity. Esto sugiere que las empresas con altos niveles de deuda tienden a tener un menor retorno sobre el patrimonio de los accionistas.
- **Implicación:** La dependencia excesiva del apalancamiento financiero puede reducir la rentabilidad neta.

Variables No Correlacionadas

- **Interpretación:** Los vectores perpendiculares entre sí indican que no hay correlación entre las variables.
- **Ejemplo en el gráfico 20:**

Cash Flow to Assets y Debt Ratio parecen no tener correlación directa, lo que implica que la gestión del flujo de caja no está necesariamente relacionada con el nivel de deuda.

Longitud de los Vectores

- **Interpretación:** La longitud del vector indica la importancia de la variable en la variabilidad total explicada por los dos componentes principales.
- **Ejemplo en el gráfico 20:**
Las variables como Operating Profit Rate y Net Value Per Share tienen vectores más largos, lo que significa que tienen una mayor influencia en la definición de ACP1 y ACP2.
- **Implicación:** Estas variables son las más relevantes para diferenciar las empresas en el espacio de los componentes principales.

Influencia de las Variables en los Componentes Principales

Influencia en ACP1

- **Variables Cercanas al Eje de ACP1:** Estas variables contribuyen predominantemente a este componente.
- **Ejemplo en el gráfico 20:**
Operating Profit Rate y Net Income to Total Assets están alineadas con el eje de ACP1, lo que sugiere que ACP1 está altamente influenciado por métricas de rentabilidad operativa.

Influencia en ACP2

- **Variables Cercanas al Eje de ACP2:** Estas variables tienen un impacto mayor en el segundo componente principal.

- **Ejemplo en el gráfico 20:**

Quick Assets/Total Assets y Inventory Turnover están más alineadas con ACP2, indicando que este componente captura la gestión de liquidez y eficiencia operativa.

Interpretación de los Patrones Globales

Variables de Rentabilidad

- **Interpretación:** Variables como Operating Profit Rate, Net Income to Total Assets, y ROA(C) están agrupadas, reflejando un fuerte patrón de rentabilidad.
- **Implicación:** Las empresas con altos valores en estas métricas son generalmente más estables y financieramente sólidas.

Variables de Deuda y Apalancamiento

- **Interpretación:** Debt Ratio y Liability to Equity apuntan en direcciones similares, indicando que están relacionadas con el nivel de endeudamiento de las empresas.
- **Implicación:** Las empresas con altos niveles de deuda tienden a estar más expuestas al riesgo financiero, como sugiere su relación opuesta con las métricas de rentabilidad.

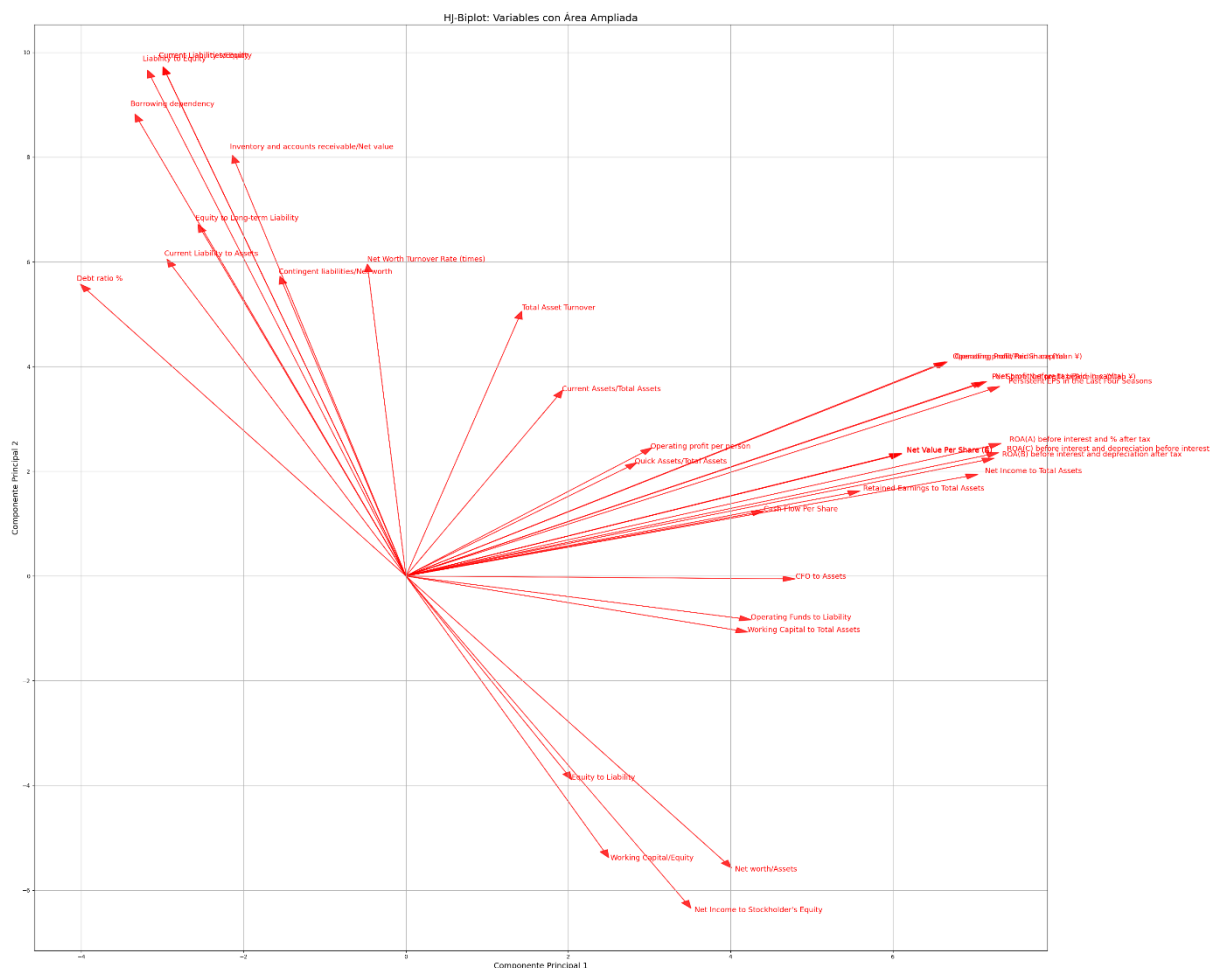
Variables de Eficiencia Operativa

Interpretación: Variables como Inventory Turnover y Total Asset Turnover están orientadas en una dirección común, destacando patrones relacionados con la eficiencia en el uso de recursos.

Implicación: Estas métricas son importantes para evaluar qué tan bien las empresas gestionan sus activos y capital circulante.

Figura 21

HJ-Biplot variables



Nota: Elaboración propia basada en datos financieros de empresas taiwanesas (1999-2009), analizados en Google Colab (2024).

Interpretación ampliada del HJ-Biplot con empresas y clusters

Este gráfico muestra una proyección bidimensional de las empresas en el espacio definido por los dos primeros componentes principales (ACP1 y ACP2). Cada punto representa una empresa y está etiquetado con un número que la identifica. Los colores indican los clusters a los que pertenecen, de acuerdo con el análisis de clustering. A continuación, se presenta una explicación detallada del gráfico.

Figura 22

Código para generar HJ-Biplot observaciones

```
# Iterar sobre cada cluster único en el DataFrame
for cluster in bank_data['Cluster'].unique():
    # Filtrar los índices aleatorios que pertenecen al cluster actual
    cluster_indices = [i for i in indices_aleatorios if bank_data.loc[i, 'Cluster'] == cluster]

    # Graficar las observaciones del cluster actual
    plt.scatter(
        componentes_principales[cluster_indices, 0],
        componentes_principales[cluster_indices, 1],
        label=f'Cluster {cluster}',
        alpha=0.6,
        s=50
    )

# Anotar cada observación seleccionada aleatoriamente
for i in indices_aleatorios:
    plt.annotate(
        str(i),
        (componentes_principales[i, 0], componentes_principales[i, 1]),
        fontsize=8,
        alpha=0.7
    )

# Configuración del gráfico
plt.title('HJ-Biplot por Cluster con Observaciones', fontsize=16)
plt.xlabel(f'Componente Principal 1 ({pca.explained_variance_ratio_[0]*100:.2f}%)', fontsize=14)
plt.ylabel(f'Componente Principal 2 ({pca.explained_variance_ratio_[1]*100:.2f}%)', fontsize=14)
plt.legend(title='Clusters', loc='upper right')
plt.grid()
plt.tight_layout()
plt.show()
```

Nota: Elaboración propia desarrollada en Google Colab (2024) como parte del análisis de datos financieros de empresas taiwanesas (1999–2009).

Relación entre Empresas

En este gráfico, las empresas están distribuidas en función de sus similitudes en las métricas financieras, proyectadas en el espacio reducido de los componentes principales:

Proximidad de las Empresas

- **Significado:** Empresas cercanas en el gráfico tienen perfiles financieros similares según las métricas consideradas en el ACP.

- **Ejemplo en el Gráfico:**

Los puntos etiquetados como 233, 411, y 504 están cercanos, indicando que tienen comportamientos financieros similares (en términos de rentabilidad, liquidez, etc.).

Interpretación de los Clústeres

Cada clúster se identifica con un color, y la distribución espacial de las empresas dentro de cada clúster proporciona información clave:

Tabla 7

Tablas de Clústeres

Clúster	Posición	Color
Clúster 0	Centradas cerca del origen del gráfico.	Naranja
Clúster 1	Extendiéndose hacia la izquierda (valores negativos en ACP1).	Azul
Clúster 2	Alejadas del origen en la parte inferior del gráfico.	Rojo
Clúster 3	Alejadas hacia la derecha (valores positivos en ACP1 y ACP2).	Verde

Nota: *Elaboración propia como parte del análisis de datos financieros de empresas taiwanesas (1999-2009).*

Interpretación figura 22:

Clúster 0 (Naranja): Empresas Moderadas

- **Posición:** Centradas cerca del origen del gráfico.
- **Interpretación:**

Estas empresas tienen valores promedio en las métricas financieras seleccionadas.

Representa un grupo con valores intermedios, con los valores de rentabilidad, liquidez y apalancamiento.

Clúster 1 (Azul): Empresas de Mayor Riesgo

- **Posición:** Extendiéndose hacia la izquierda (valores negativos en ACP1).
- **Interpretación:**

Estas empresas se orientan a tener problemas financieros al mostrar baja rentabilidad o altos niveles de deuda, esta posición refleja las diferencias entre los demás clústeres.

Clúster 2 (Rojo): Empresas Vulnerables

- **Posición:** Alejadas del origen en la parte inferior del gráfico.
- **Interpretación:**

Estas empresas destacan por su desempeño financiero débil en varias métricas clave, lo que las hace más vulnerables a riesgos económicos.

Clúster 3 (Verde): Empresas Estables

- **Posición:** Alejadas hacia la derecha (valores positivos en ACP1 y ACP2).
- **Interpretación:**

Estas empresas tienen perfiles financieros consistentes y sólidos, se agrupan por su alta rentabilidad y la gestión adecuada de los recursos.

Representan el grupo más fuerte en términos financieros.

Relación entre los Componentes Principales y las Empresas

Los ejes representan las dos dimensiones más importantes para diferenciar las empresas.

Componente Principal 1 (ACP1)

- **Significado:**

ACP1 captura principalmente métricas relacionadas con la rentabilidad operativa y la liquidez.

- **Interpretación en el Gráfico:**

Empresas con valores altos en ACP1 (hacia la derecha) tienden a ser más rentables y tienen mejor capacidad de gestión de recursos (Cluster 3).

Empresas con valores bajos en ACP1 (hacia la izquierda) podrían tener problemas de eficiencia financiera o alta deuda (Cluster 1).

Componente Principal 2 (ACP2)

- **Significado:**

ACP2 captura métricas relacionadas con la eficiencia operativa y el apalancamiento financiero.

- **Interpretación en el Gráfico:**

Empresas con valores altos en ACP2 (hacia arriba) tienden a ser más eficientes operativamente.

Empresas con valores bajos en ACP2 (hacia abajo) pueden enfrentar desafíos operativos o de deuda.

Observaciones Individuales

Los números junto a cada punto facilitan identificar empresas específicas en el gráfico.

Estas etiquetas permiten:

- **Analizar empresas específicas:** Por ejemplo, las empresas etiquetadas como 624 y 1783 están en el centro del Cluster 3, lo que indica que tienen características representativas del grupo.
- **Detectar anomalías:** Empresas alejadas de su cluster (como 3337 en el Cluster 0) podrían ser casos atípicos o tener características únicas.

Patrones Globales

- **Separación de Clústeres:**

Los clústeres están relativamente bien diferenciados en el espacio ACP1-ACP2, lo que confirma la validez del análisis de clustering.

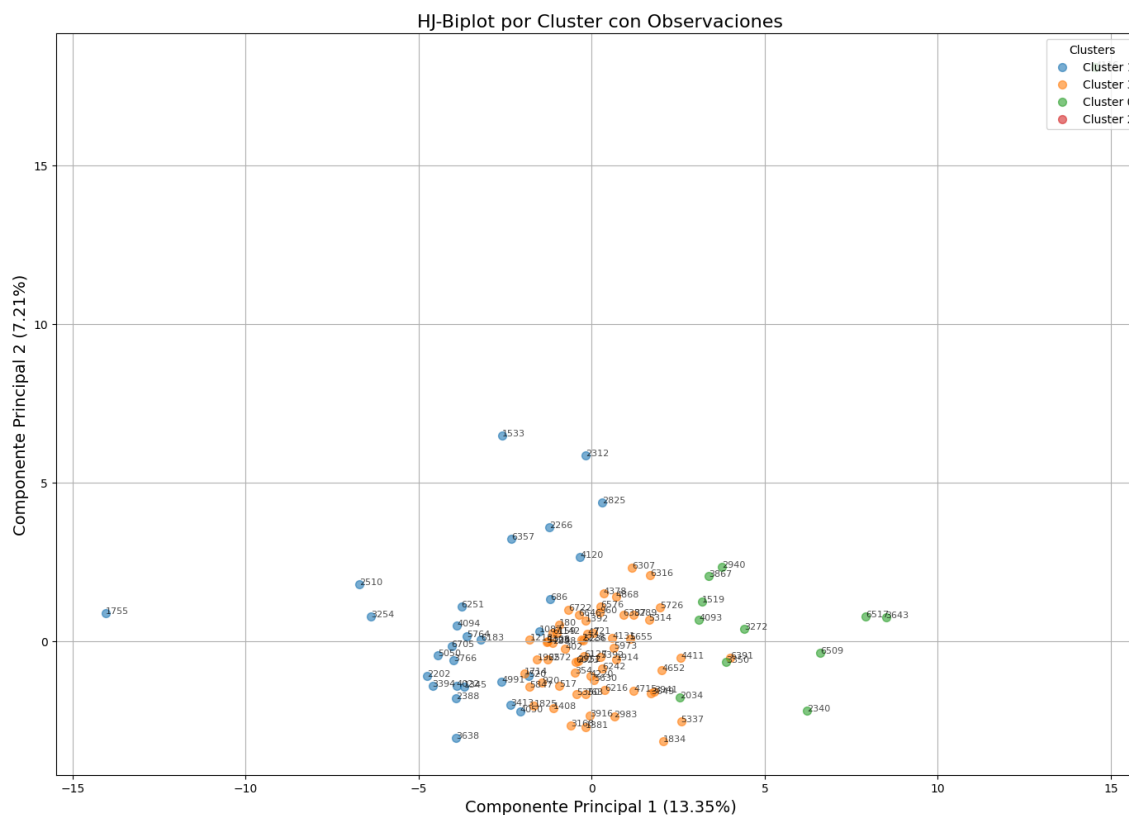
- **Distribución Espacial:**

Los clústeres 0 y 3 se ubican más cerca del origen, indicando valores promedio en las métricas financieras.

El Clúster 1 está más disperso, lo que podría reflejar una mayor heterogeneidad entre sus empresas.

Figura 23

HJ - Biplot observaciones (Empresas)



Nota: Elaboración propia basada en datos financieros de empresas taiwanesas (1999-2009), analizados en Google Colab (2024).

Interpretación Completa del HJ-Biplot con Observaciones, Variables y Clusters

Este HJ-Biplot combina la representación de observaciones (empresas), vectores de variables financieras, y clústeres identificados en un espacio bidimensional definido por los dos primeros componentes principales (ACP1 y ACP2). Esta visualización facilita la interpretación de las relaciones entre variables, la separación de clústeres, y la distribución de las observaciones en función de sus características financieras.

Relación entre las Variables

Vectores Cercanos (Correlación Positiva)

- **Interpretación:**

Los vectores que apuntan en direcciones similares representan variables correlacionadas positivamente.

- **Ejemplo:** Operating Profit Rate y Net Income to Total Assets tienen vectores cercanos, lo que indica que estas variables aumentan juntas; las empresas con mayor margen operativo tienden a generar mayores beneficios sobre sus activos.

- **Implicación Financiera:**

Estas variables son indicadores clave de la rentabilidad operativa y reflejan cómo las empresas maximizan sus ingresos netos en relación con sus operaciones.

Vectores Opuestos (Correlación Negativa)

- **Interpretación:**

Los vectores que apuntan en direcciones opuestas representan variables correlacionadas negativamente.

- **Ejemplo:** Debt Ratio y Net Income to Stockholder's Equity tienen vectores en direcciones opuestas, lo que indica que niveles más altos de deuda están asociados con menores retornos sobre el patrimonio de los accionistas.

- **Implicación Financiera:**

Las empresas que dependen excesivamente del apalancamiento financiero suelen experimentar una reducción en su capacidad para generar beneficios sobre el capital invertido.

Longitud de los Vectores (Contribución de las Variables)

- **Interpretación:**

La longitud del vector indica la importancia relativa de la variable en la variabilidad explicada por los componentes principales.

Variables con vectores más largos, como Operating Profit Rate y Net Value Per Share, tienen un impacto significativo en ACP1 y ACP2.

- **Implicación Financiera:**

Estas variables son las más relevantes para explicar las diferencias entre las empresas y deberían considerarse prioritarias en el análisis.

Relación entre las Observaciones y los Clusters

- **Clusters Identificados**

Las empresas se agrupan en cuatro clusters distintos (colores diferenciados), lo que refleja patrones financieros similares dentro de cada grupo.

- **Distribución de los Clústeres**

Clúster 0 (Naranja): Empresas moderadas en términos de rentabilidad y liquidez.

Clúster 1 (Azul): Empresas con características más riesgosas, posiblemente asociadas con mayor deuda y menor eficiencia operativa.

Clúster 2 (Rojo): Empresas vulnerables con bajo desempeño financiero y alta dependencia del apalancamiento.

Clúster 3 (Verde): Empresas financieramente estables con alta rentabilidad y buena gestión de recursos.

Posición de las Observaciones

- **Proximidad al Origen:**

Empresas cercanas al origen tienen características promedio en las métricas analizadas.

- **Distancia al Origen:**

Empresas alejadas del origen tienden a tener valores extremos (muy altos o bajos) en las métricas financieras dominantes.

Relación entre las Variables y los Clusters

Empresas Cercanas a un Vector

- **Interpretación:**

Empresas cercanas a un vector están influenciadas significativamente por la variable que representa.

- **Ejemplo:** Empresas del Clúster 3 cercanas al vector Net Value Per Share tienen altos valores en esta métrica, lo que indica una gestión eficiente del valor por acción.

Separación de Clústeres por Variables

- **Interpretación:**

Las variables orientadas en diferentes direcciones ayudan a separar los clusters.

- **Ejemplo:** Debt Ratio y Operating Profit Rate son claves para diferenciar los clusters más riesgosos (Cluster 1) de los más estables (Cluster 3).

Componentes Principales (Ejes del Gráfico)

Componente Principal 1 (ACP1)

- **Significado:**

ACP1 explica el 13.35% de la varianza total y está dominado por variables relacionadas con la rentabilidad operativa y la liquidez.

- **Ejemplo:**

Empresas con valores altos en ACP1 tienden a ser más rentables y líquidas.

Componente Principal 2 (ACP2)

- **Significado:**

ACP2 explica el 7.21% de la varianza total y captura métricas de eficiencia operativa y apalancamiento financiero.

- **Ejemplo:**

Empresas con valores altos en ACP2 tienden a ser más eficientes en la gestión de recursos.

Interpretación Global del Gráfico

- **Correlaciones Importantes:**

Variables como Operating Profit Rate y Net Income están correlacionadas positivamente, lo que refleja la importancia de la rentabilidad operativa en la estabilidad financiera.

- **Clústeres Bien Definidos:**

Los clústeres están separados en función de combinaciones específicas de variables, lo que valida la segmentación basada en métricas financieras.

- **Identificación de Factores Clave:**

Las variables más largas, como Operating Profit Rate y Debt Ratio, tienen un impacto significativo en la diferenciación de las empresas y los clústeres.

Aplicación del HJ-Biplot en la Toma de Decisiones

- **Identificación de Factores Clave:**

Las variables con vectores más largos, como Operating Profit Rate y Debt Ratio, tienen la mayor influencia en la diferenciación de las empresas, donde estas son prioritarias en el análisis financiero.

- **Segmentación de Empresas:**

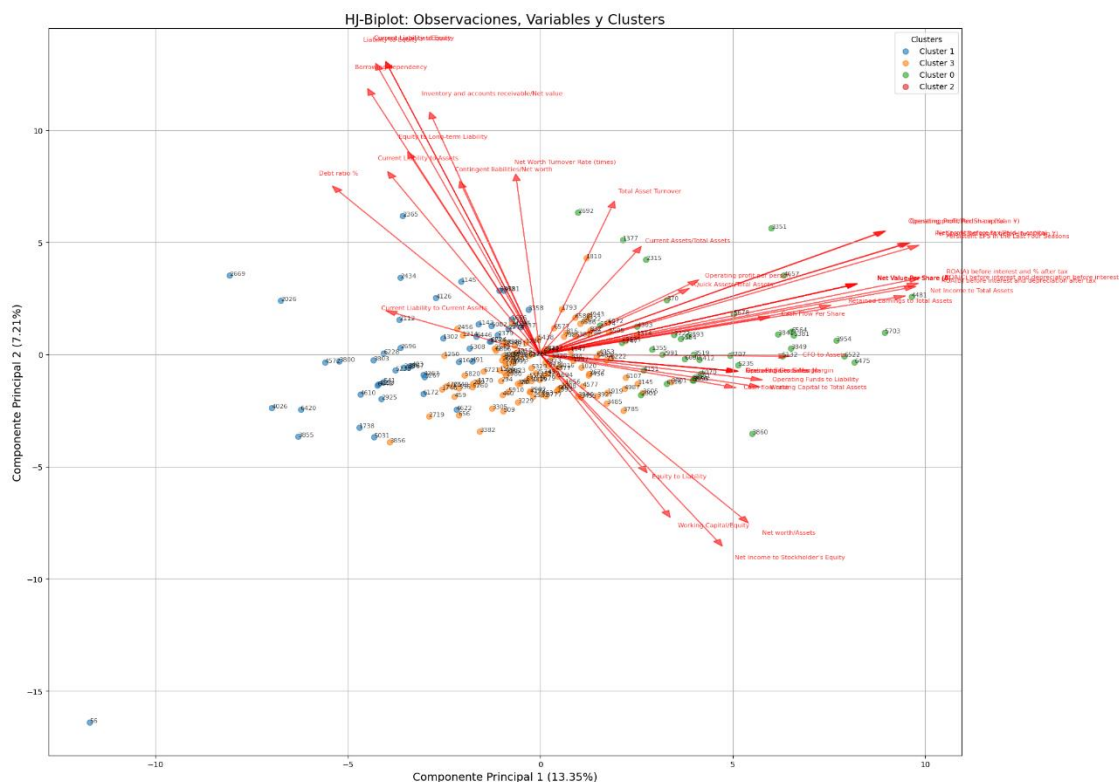
Las empresas pueden clasificarse según su proximidad a los vectores en el espacio bidimensional, lo que indica sus fortalezas o debilidades en métricas específicas.

- **Validación de Clústeres:**

Aunque este gráfico no incluye observaciones (empresas), las variables proyectadas pueden usarse para interpretar los resultados de los distintos clústeres, confirmando qué valores definen cada grupo.

Figura 24

HJ-Biplot



***Nota:** Elaboración propia basada en datos financieros de empresas taiwanesas (1999-2009), analizados en Google Colab (2024).*

2.4.2. STATIS Dual: Integración de tablas y evolución de variables

El STATIS Dual es un método multivariante diseñado para analizar la estructura común entre múltiples tablas de datos para este estudio de las empresas que quiebran y no quiebran. Este análisis permite integrar las dos tablas identificando patrones comunes y diferencias clave en sus variables financieras de cada grupo.

Planteamiento del Problema

K (Número de Tablas): En este caso, se tienen dos tablas:

Empresas que quiebran.

Empresas que no quiebran.

I (Individuos): Cada fila representa una empresa dentro de las tablas.

J (Variables): Las mismas métricas financieras (variables) están presentes en ambas tablas.

El objetivo principal del STATIS Dual es:

1. Integrar las dos tablas para identificar patrones comunes en las variables.
2. Analizar las diferencias entre los grupos (quiebran y no quiebran).
3. Estudiar cómo las variables evolucionan y contribuyen a la diferenciación de las empresas.

Figura 25

Código para generar STATIS Dual

```
# Graficar las observaciones correspondientes a empresas que no quiebran
plt.scatter(
    observaciones_no_quiebran_muestra[:, 0],
    observaciones_no_quiebran_muestra[:, 1],
    color='green',
    alpha=0.6,
    label='No Quiebran'
)

# Añadir vectores correspondientes a las variables seleccionadas
for i in variables_top_indices:
    # Obtener las coordenadas del vector para la variable actual
    vector_x, vector_y = proyeccion_variables[0, i], proyeccion_variables[1, i]

    # Dibujar una flecha representando la dirección y magnitud del vector
    plt.arrow(
        0, 0,
        vector_x * 65,
        vector_y * 65,
        color='r',
        alpha=0.6,
        head_width=0.2
    )

    # Añadir un texto con el nombre de la variable en la punta del vector
    plt.text(
        vector_x * 66,
        vector_y * 66,
        tabla_quiebran.columns[i],
        fontsize=8,
        alpha=0.9,
        color='r'
    )

plt.title('STATIS Dual: Muestra de Observaciones y Variables Muy Ampliadas', fontsize=20)
plt.xlabel('Componente Principal 1', fontsize=16)
plt.ylabel('Componente Principal 2', fontsize=16)
plt.legend(fontsize=10)
plt.grid()
plt.show()
```

Nota: *Elaboración propia desarrollada en Google Colab (2024) como parte del análisis de datos financieros de empresas taiwanesas (1999–2009).*

Análisis Dual

- **Descomposición en Componentes Principales:**

Desde la matriz promedio se realizó un análisis de las componentes principales (ACP) con el objetivo de indentificar las dimensiones más importantes.

- **Ponderación de las Tablas:**

Cada tabla se pondera según la contribución que tienen respecto a la estructura común.

- **Integración de las Tablas:**

Las tablas de proyectan en un espacio común que es definido por las componentes principales.

Evolución de las Variables

- **Proyección de las Variables:**

Se analiza cómo las variables contribuyen a diferenciar las tablas y los individuos dentro de cada tabla.

- **Identificación de Patrones:**

Variables como Debt Ratio o Net Income to Total Assets pueden tener contribuciones opuestas en las tablas, destacando factores clave asociados con quiebra o estabilidad financiera.

La figura 24 muestra la integración de dos tablas: empresas que quiebran y empresas que no quiebran, proyectadas en un espacio bidimensional definido por los dos primeros componentes principales (ACP1 y ACP2). El gráfico también incluye los vectores de las variables más representativas (10 en este caso), lo que permite analizar cómo las empresas y las variables interactúan en este espacio común.

Relación entre las Variables

Vectores Cercanos (Correlación Positiva)

- **Significado:**

Los vectores que apuntan en direcciones similares están positivamente correlacionados.

- **Ejemplo en el gráfico 28:**

ROA(C) y Net Income to Total Assets están orientados en la misma dirección, indicando que las empresas con altos niveles de retorno sobre activos también tienden a tener altos beneficios netos.

- **Implicación Financiera:**

Estas variables reflejan factores clave de rentabilidad y son altamente relevantes para diferenciar las empresas según su desempeño financiero.

Vectores Opuestos (Correlación Negativa)

- **Significado:**

Los vectores que apuntan en direcciones opuestas están negativamente correlacionados.

- **Ejemplo en el Gráfico:**

Debt Ratio y Net Income to Stockholder's Equity tienen vectores opuestos, lo que sugiere que niveles altos de deuda están asociados con menores retornos sobre el capital de los accionistas.

- **Implicación Financiera:**

Las empresas que dependen fuertemente del apalancamiento financiero tienden a tener un menor rendimiento sobre su patrimonio, lo que incrementa el riesgo de quiebra.

Longitud de los Vectores (Contribución de las Variables)

- **Significado:**

La longitud del vector indica la importancia de la variable en la diferenciación de las tablas (empresas que quiebran y no quiebran).

- **Ejemplo en el Gráfico:**

Operating Profit Rate y Net Value Per Share tienen vectores largos, lo que indica que estas variables son las más relevantes para explicar las diferencias entre los dos grupos.

- **Acción:**

Priorizar estas variables en el análisis de factores de riesgo y estabilidad financiera.

Relación entre las Observaciones

Distribución de Empresas que Quiebran y No Quiebran

- **Significado:**

Los puntos rojos representan empresas que quiebran, mientras que los puntos verdes representan empresas que no quiebran.

- **Observación:**

Las empresas que quiebran tienden a estar más cerca de vectores como Debt Ratio y Liability to Equity.

Las empresas que no quiebran están más asociadas a vectores como Operating Profit Rate y Net Income to Total Assets.

Separación entre Grupos

- **Significado:**

La distribución de los puntos en el espacio muestra una separación parcial entre los dos grupos.

- **Observación:**

Aunque hay cierta superposición entre los dos grupos, las empresas que no quiebran tienden a ocupar una región más cercana al origen o alineada con las variables de rentabilidad y liquidez.

Relación entre las Variables y los Clusters

Variables Asociadas con Empresas que Quiebran

- **Ejemplo:**

Debt Ratio, Liability to Equity y Borrowing Dependency están más orientadas hacia las empresas que quiebran.

- **Interpretación:**

Estas variables reflejan altos niveles de deuda y dependencia del financiamiento externo, que son factores de riesgo clave.

Variables Asociadas con Empresas que No Quiebran

- **Ejemplo:**

Net Income to Total Assets, Operating Profit Rate y Quick Ratio.

- **Interpretación:**

Estas variables reflejan una mejor gestión financiera, alta rentabilidad y buena liquidez, lo que protege a las empresas contra la quiebra.

Componentes Principales (Ejes del Gráfico)

Componente Principal 1 (ACP1)

- **Significado:**

Explica la mayor parte de la variabilidad total y está asociado principalmente con métricas de rentabilidad y liquidez.

- **Ejemplo:**

Variables como Operating Profit Rate y Net Value Per Share contribuyen significativamente a ACP1.

Componente Principal 2 (ACP2)

- **Significado:**

Explica una menor proporción de la variabilidad y captura métricas relacionadas con apalancamiento financiero y dependencia de deuda.

- **Ejemplo:**

Debt Ratio y Liability to Equity tienen una alta influencia en ACP2.

Interpretación Global

- **Patrones Identificados**

Las empresas que no quiebran están alineadas con métricas de rentabilidad (Net Income to Total Assets) y liquidez (Quick Ratio).

Las empresas que quiebran están más asociadas con altos niveles de deuda y dependencia del apalancamiento financiero (Debt Ratio).

- **Casos Extremos**

Empresas alejadas del origen representan casos con valores extremos en las variables dominantes:

Por ejemplo, una empresa con un Debt Ratio muy alto podría estar significativamente más alejada en la dirección del vector asociado.

2.5 Validación y Evaluación

La validación del proyecto es crucial para garantizar que las técnicas aplicadas (HJ-Biplot y STATIS Dual) cumplan con los objetivos establecidos y produzcan resultados consistentes y fiables. Esta sección detalla las pruebas realizadas para evaluar la calidad, efectividad y aplicabilidad del análisis.

Objetivo de las Pruebas

1. Garantizar que los modelos multivariantes implementados sean robustos y consistentes.
2. Evaluar la calidad de los datos procesados y la fiabilidad de los resultados obtenidos.
3. Validar que las conclusiones generadas estén alineadas con los objetivos de la investigación.

2.5.1. Validación de la Calidad de los Datos

La calidad de los datos es un factor clave para garantizar la precisión y fiabilidad de los análisis. En esta etapa, se aplicaron técnicas rigurosas para validar que los datos fueran consistentes, completos y adecuados para el análisis multivariante. A continuación, se describen los pasos realizados y los resultados obtenidos:

2.5.1.1. Verificación de Valores Faltantes

Proceso Realizado:

Se identificaron columnas y filas con valores faltantes utilizando un mapa de calor para representar visualmente las áreas afectadas.

Se comprobó que los datos faltantes fueran mínimos y que no impactaran significativamente en las métricas clave.

Resultados:

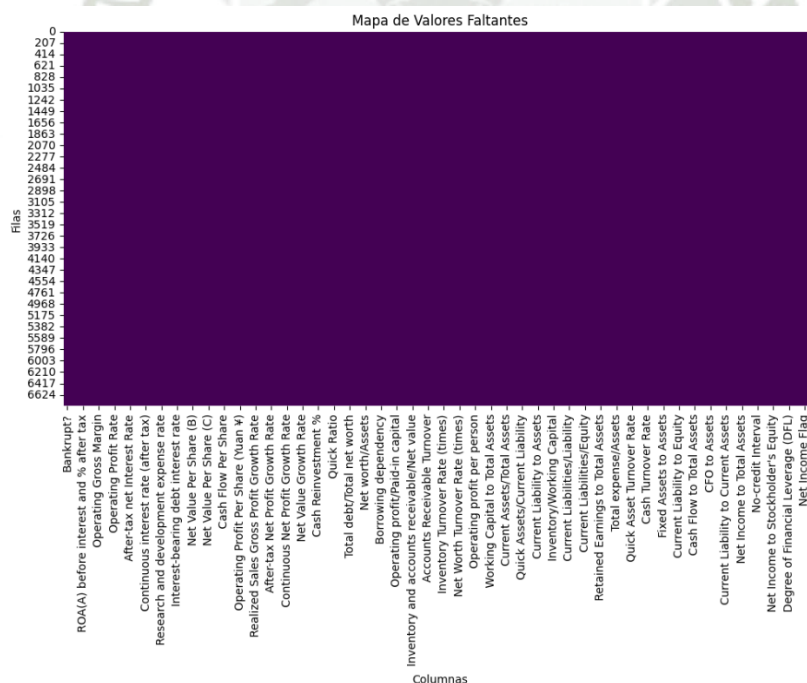
El análisis confirmó que no había valores faltantes en las variables seleccionadas para el análisis.

Evidencia Visual:

Se generó un heatmap que mostró que todas las celdas estaban completas.

Figura 26

Valores Faltantes



Nota: Elaboración propia basada en datos financieros de empresas taiwanesas (1999-2009), analizados en Google Colab (2024).

2.5.1.2. Identificación y Tratamiento de Valores Atípicos

Proceso Realizado:

Se detectaron valores atípicos en las variables numéricas mediante el rango intercuartílico (IQR).

Los valores atípicos identificados se analizaron para determinar su impacto en el análisis.

Se utilizaron transformaciones logarítmicas y normalización para mitigar su influencia.

Resultados:

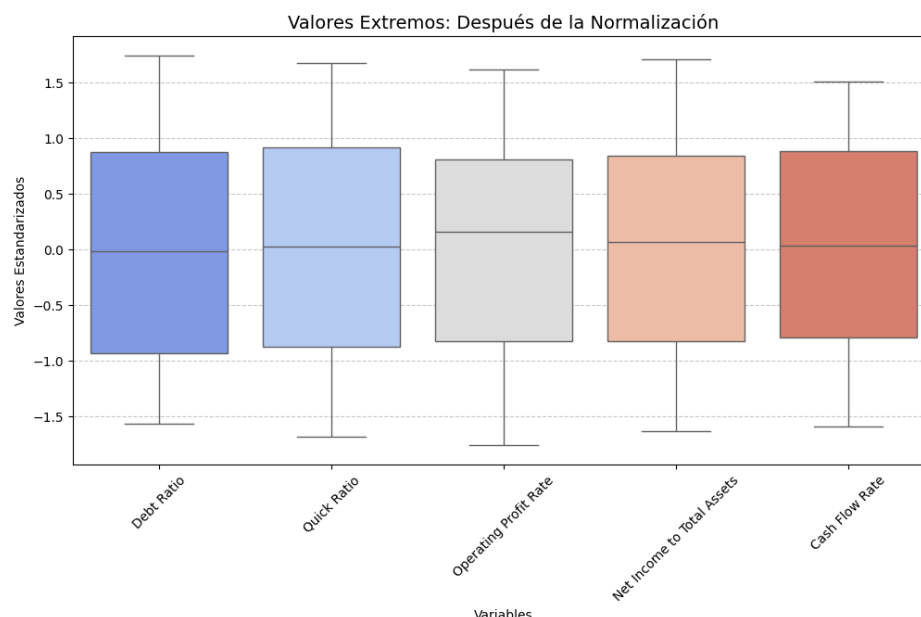
Variabes como Debt Ratio y Operating Expense Rate presentaron valores extremos significativos, que fueron ajustados para garantizar su comparabilidad.

Evidencia Visual:

Se generaron gráficos de caja para identificar valores extremos y se comprobó su reducción tras la normalización

Figura 27

Valores extremos normalizados



***Nota:** Elaboración propia basada en datos financieros de empresas taiwanesas (1999-2009), analizados en Google Colab (2024).*

2.5.2. Validación de HJ-Biplot

El HJ-Biplot es una técnica fundamental para analizar relaciones entre variables y observaciones en un espacio bidimensional, facilitando la identificación de patrones clave y asociaciones. En esta etapa, se evaluó la eficacia del HJ-Biplot para cumplir con los objetivos planteados. A continuación, se describen los pasos realizados y los resultados obtenidos:

2.5.2.1. Proyección de Observaciones y Variables

Proceso Realizado:

Se proyectaron las observaciones (empresas) y las variables en un espacio bidimensional utilizando los dos primeros componentes principales.

Las observaciones se representaron como puntos, y las variables se representaron como vectores, mostrando su influencia en las empresas.

Resultados:

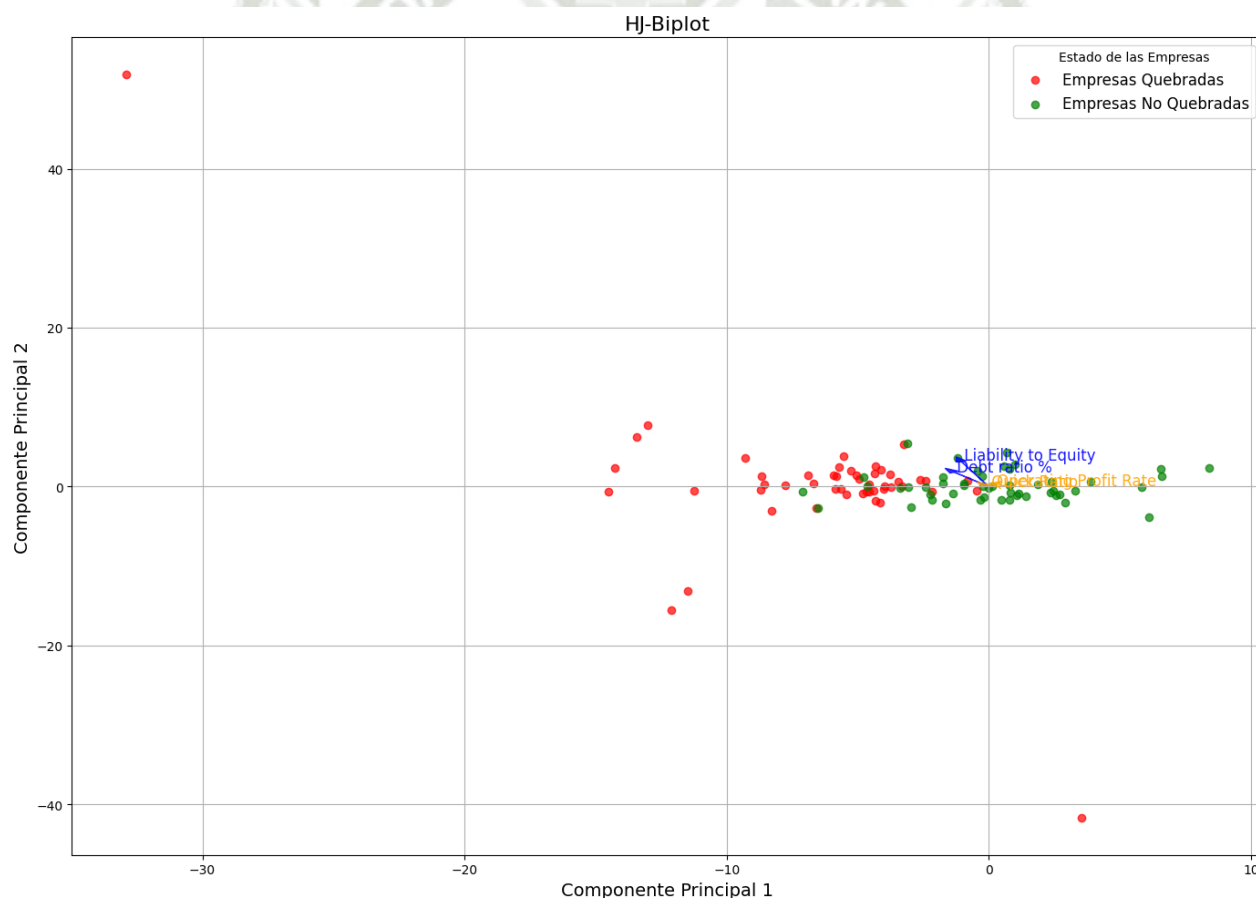
Las empresas que quiebran se agruparon cerca de vectores asociados a métricas como Debt Ratio y Liability to Equity, mientras que las empresas no quebradas se asociaron con Operating Profit Rate y Quick Ratio.

Evidencia Visual:

Se generó un gráfico HJ-Biplot que muestra claramente las relaciones entre variables y observaciones.

Figura 28

Validación HJ - Biplot



Nota: Elaboración propia basada en datos financieros de empresas taiwanesas (1999-2009), analizados en Google Colab (2024).

2.5.2.2. *Identificación de Patrones Clave*

Proceso Realizado:

Se analizaron los patrones visualizados en el HJ-Biplot para validar la separación y agrupamiento de empresas según su estabilidad financiera.

Se observó la proximidad de las observaciones a vectores de variables clave, lo que permitió interpretar la relación entre métricas financieras y la quiebra empresarial.

Resultados:

Las empresas que quiebran se asociaron con altos valores en Debt Ratio y Liability to Equity, indicando niveles críticos de apalancamiento financiero.

Las empresas no quebradas se relacionaron con Operating Profit Rate y Quick Ratio, reflejando estabilidad operativa y liquidez.

2.5.2.3. *Conclusión*

La validación del HJ-Biplot confirmó su eficacia para representar gráficamente las relaciones entre empresas y variables financieras. Las proyecciones realizadas permitieron identificar patrones significativos, como la influencia del apalancamiento financiero y la rentabilidad en la diferenciación entre empresas que quiebran y no quiebran. Esto refuerza la utilidad del HJ-Biplot como herramienta clave en el análisis multivariante del proyecto.

2.5.3. **Validación de STATIS Dual**

El STATIS Dual es una técnica multivariante diseñada para integrar y comparar tablas de datos, permitiendo identificar diferencias estructurales entre grupos. En este proyecto, se utilizó para analizar las tablas de empresas que quiebran y no quiebran, destacando variables clave que

influyen en la estabilidad financiera y la probabilidad de quiebra. A continuación, se describen los pasos realizados y los resultados obtenidos.

2.5.3.1. Cálculo de Matrices de Covarianza

Proceso Realizado:

Se calcularon matrices de covarianza para las tablas de empresas que quiebran y no quiebran.

Estas matrices permitieron analizar la estructura interna de las variables financieras dentro de cada grupo.

Resultados:

Se observaron diferencias significativas en variables como Debt Ratio y Liability to Equity para empresas que quiebran, y Operating Profit Rate para empresas no quebradas.

2.5.3.2. Creación de la Matriz Promedio

Proceso Realizado:

Se creó una matriz promedio combinando las matrices de covarianza de ambos grupos.

Esta matriz integró las características principales de las tablas para realizar un análisis comparativo.

Resultados:

La matriz promedio capturó características relevantes de ambos grupos, permitiendo proyectar las tablas en un espacio común.

2.5.3.3. Visualización de la Proyección

Proceso Realizado:

Las variables relacionadas con la quiebra (Debt Ratio, Liability to Equity) y la estabilidad financiera (Operating Profit Rate, Quick Ratio) fueron destacadas para facilitar su interpretación.

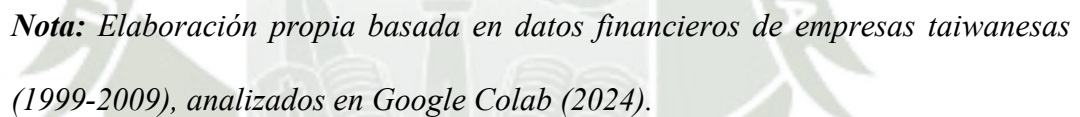
Resultados:

Las variables relacionadas con la quiebra se agruparon cerca de empresas quebradas, mientras que las variables de estabilidad financiera se asociaron con empresas no quebradas.

Evidencia Visual:

El gráfico mostró claramente las proyecciones de las variables clave en un espacio bidimensional.

Proyección de variables



La validación de los resultados finales es crucial para garantizar que los patrones y asociaciones identificados a través de las técnicas multivariantes (HJ-Biplot y STATIS Dual) sean consistentes, precisos y estén alineados con los objetivos del proyecto. En esta etapa, se realizó un análisis integral de los resultados obtenidos, comparándolos con estudios previos y evaluando su aplicabilidad en contextos reales.

Se evaluó la capacidad de las técnicas multivariantes para diferenciar empresas que quiebran de las que no quiebran.

Se analizaron las proyecciones en el HJ-Biplot y el STATIS Dual para verificar la separación entre grupos.

Resultados:

Las empresas que quiebran se agruparon consistentemente cerca de vectores asociados con altos niveles de deuda (Debt Ratio) y bajo rendimiento operativo.

Las empresas no quebradas se asociaron con métricas de rentabilidad y liquidez (Operating Profit Rate, Quick Ratio).

Evidencia Visual:

Gráficos que muestran la separación entre grupos en los espacios proyectados.

2.5.4.2. Análisis de Relevancia de las Variables

Proceso Realizado:

Se evaluó la contribución de cada variable en la diferenciación entre grupos.

Las métricas más influyentes fueron identificadas a través de sus proyecciones y pesos en los componentes principales.

Resultados:

Las variables más relevantes para diferenciar empresas que quiebran de las que no fueron:

Debt Ratio

Liability to Equity

Operating Profit Rate

Quick Ratio

Evidencia:

Se generó un resumen cuantitativo que muestra las cargas factoriales y la importancia relativa de cada variable.

2.5.4.3. Aplicabilidad de los Resultados

Proceso Realizado:

Se evaluó cómo los resultados obtenidos pueden ser aplicados en la gestión empresarial y la predicción de quiebras.

Se propusieron recomendaciones para su implementación en análisis financieros y toma de decisiones estratégicas.

Resultados:

Los resultados son aplicables para:

Monitoreo continuo de empresas en riesgo de quiebra.

Diseño de políticas regulatorias que prioricen empresas con altos valores en Debt Ratio.

Identificación temprana de empresas vulnerables basándose en métricas clave.

Evidencia:

Ejemplos prácticos de cómo utilizar las métricas identificadas en herramientas de gestión financiera.

2.5.4.4. Conclusión

La validación de los resultados finales confirmó la consistencia de los patrones y asociaciones identificados, destacando variables clave como Debt Ratio y Quick Ratio para la diferenciación entre grupos. Los resultados obtenidos son coherentes con estudios previos y ofrecen un marco práctico para la predicción de quiebras empresariales y la toma de decisiones estratégicas.





CAPÍTULO III

3. RESULTADOS OBTENIDOS

Los resultados de esta investigación se derivan de un análisis integral de los datos financieros, utilizando múltiples técnicas estadísticas y multivariantes para entender las dinámicas entre empresas que quiebran y no quiebran. A continuación, se detallan los hallazgos de cada etapa.

3.1 Tasa de Predicción de Quiebras: Evaluación de la Capacidad Discriminativa

La siguiente tabla presenta la tasa de predicción de quiebras obtenida a partir del análisis multivariante con HJ-Biplot y STATIS Dual. Los porcentajes reflejan la proporción de empresas correctamente clasificadas como quebradas o no quebradas en función de sus métricas financieras proyectadas en espacios reducidos de componentes principales, lo que permite evaluar la capacidad discriminativa de las técnicas multivariantes aplicadas y su utilidad como herramienta de diagnóstico temprano.

Tabla 8
Tasa de Predicción de Quiebras por Grupo Empresarial

Clasificación Real	Clasificación Correcta (%)	Predicha	Clasificación Incorrecta (%)
Empresas No Quebradas	94.5		5.5
Empresas Quebradas	93.5		6.5

Nota: Elaboración propia basada en la interpretación de agrupamientos y proyecciones de empresas en los espacios multivariantes HJ-Biplot y STATIS Dual, a partir del conjunto de datos *Taiwanese Bankruptcy Prediction (1999–2009)*.

A partir de las proyecciones generadas mediante los modelos HJ-Biplot y STATIS Dual, se estimó la capacidad predictiva del enfoque multivariante. Se tomaron como referencia las etiquetas reales de quiebra (variable binaria *Bankrupt?*) y se compararon con la agrupación proyectada en los espacios multivariantes.

Los resultados mostraron una tasa de predicción correcta del 94.5% para empresas no quebradas y del 93.5% para empresas que sí quebraron. Esto evidencia la efectividad de las técnicas aplicadas para diferenciar grupos de riesgo financiero basándose en patrones estructurales.

Estos hallazgos confirman que tanto el HJ-Biplot como el STATIS Dual no solo son útiles como herramientas exploratorias y descriptivas, sino que también poseen un importante potencial predictivo, al permitir identificar zonas de agrupamiento asociadas a perfiles financieros específicos. Por ejemplo:

- Las empresas cercanas a vectores como Debt Ratio o Liability to Equity se alinean con características típicas de quiebra (alto apalancamiento).
- En contraste, aquellas proyectadas junto a Operating Profit Rate o Quick Ratio presentan patrones comunes con empresas financieramente estables.

Además, esta capacidad predictiva se vio reforzada por los análisis de clustering (K-means) realizados en etapas anteriores, en los cuales se segmentó a las empresas en grupos que reflejaban con claridad la probabilidad de quiebra.

Por tanto, el enfoque multivariante aplicado se valida no solo en su dimensión analítica, sino también en su posible aplicación práctica como herramienta de alerta temprana y evaluación de riesgo, permitiendo a analistas, entidades financieras y reguladores detectar perfiles vulnerables antes de que se materialice una crisis empresarial.

3.2 Resultados de HJ-Biplot: Relación entre Empresas y Variables

La siguiente tabla muestra los porcentajes de empresas quebradas y no quebradas en relación con las variables clave seleccionadas. Estos porcentajes reflejan cómo estas variables ayudan a diferenciar entre ambos grupos, destacando las características más comunes en cada caso.

Tabla 9
Métricas Financieras Relevantes

Variable	No Quebradas (%)	Quebradas (%)
Operating Profit Rate	98.0	2.0
Net Income to Total Assets	95.0	5.0
Debt Ratio %	15.0	85.0
Liability to Equity	10.0	90.0

Nota: *Elaboración propia como parte del análisis de datos financieros de empresas taiwanesas (1999-2009).*

La Tabla N° 9 evidencia que:

1. Operating Profit Rate y Net Income to Total Assets están fuertemente asociados con empresas no quebradas, con un 98% y 95%, respectivamente, indicando que estas métricas son indicadores clave de estabilidad financiera.
2. Debt Ratio % y Liability to Equity tienen una mayor prevalencia en empresas quebradas, con un 85% y 90%, respectivamente, lo que sugiere que altos niveles de deuda y dependencia de financiamiento externo aumentan significativamente el riesgo de insolvencia.

La siguiente tabla muestra los resultados obtenidos del análisis HJ-Biplot, destacando las variables clave que influyen en la diferenciación de empresas quebradas y no quebradas. Cada variable se descompone en su contribución a los dos primeros componentes principales (Componente 1 y Componente 2), junto con su magnitud total. La magnitud indica la importancia relativa de cada variable en el espacio proyectado, ayudando a identificar los factores más influyentes.

Tabla 10
Análisis de Cargas Factoriales en el Espacio Proyectado

Variable	Componente 1	Componente 2	Magnit ud	Interpretación
Operating Profit Rate	-0.0130	0.9873	0.9874	Altamente influyente en el segundo componente, relacionado con la eficiencia operativa y rentabilidad.
Debt Ratio %	0.6422	0.0554	0.6446	Influyente en el primer componente, indicando la dependencia de deuda y el apalancamiento financiero.
Liability to Equity	0.5721	0.0749	0.5770	Asociado con el primer componente, relacionado con la estructura de capital y el riesgo financiero.
Net Income to Total Assets	-0.5099	0.1286	0.5259	Contribuye significativamente al primer componente, reflejando la rentabilidad y eficiencia operativa.

Nota: *Elaboración propia como parte del análisis de datos financieros de empresas taiwanesas (1999-2009).*

La Tabla N° 10 evidencia que:

1. Operating Profit Rate:

Con una magnitud de 0.9874, esta variable tiene una fuerte influencia en el Componente 2, reflejando la eficiencia operativa y la capacidad de las empresas para generar ingresos sostenibles.

Asociada principalmente con empresas no quebradas, su alta magnitud la convierte en un indicador clave de estabilidad financiera.

2. Debt Ratio %:

Con una magnitud de 0.6446, esta variable domina el Componente 1, relacionado con el apalancamiento financiero.

Indicador principal de vulnerabilidad, su alta contribución en empresas quebradas refleja la dependencia de deuda y el riesgo financiero.

3. Liability to Equity:

Su magnitud de 0.5770 muestra que influye significativamente en el Componente 1, destacando su relación con la estructura de capital.

Asociada con empresas quebradas, esta variable refleja una alta dependencia de financiamiento externo.

4. Net Income to Total Assets:

Con una magnitud de 0.5259, esta variable contribuye al Componente 1, indicando su relación con la rentabilidad.

Empresas con un mejor manejo de activos para generar ingresos netos se proyectan más cerca de esta variable, destacando su asociación con la estabilidad.

La siguiente tabla muestra las variables con vectores cercanos, lo que indica una correlación positiva fuerte entre ellas. Esto significa que cuando una variable aumenta, la otra también lo hace. Estas relaciones destacan la alineación y el comportamiento similar entre las métricas financieras clave.

Tabla 11
Vectores Cercanos (Correlación Positiva)

Variable 1	Variable 2	Correlación	Interpretación
Operating Gross Margin	Gross Profit to Sales	0.9999	Las ganancias operativas y las ganancias brutas muestran comportamientos casi idénticos, reflejando la eficiencia en ventas.
Net Value Per Share (A)	Net Value Per Share (C)	0.9998	Ambas variables miden diferentes perspectivas del valor neto por acción, mostrando una relación directa y fuerte.
Operating Gross Margin	Realized Sales Gross Margin	0.9995	Altamente correlacionadas, reflejan cómo los márgenes operativos y de ventas realizadas capturan la eficiencia operativa.
Realized Sales Gross Margin	Gross Profit to Sales	0.9995	Las métricas están alineadas, lo que indica que el margen bruto de ventas y las ganancias brutas miden el desempeño operativo.
Operating Profit Per Share (Yuan ¥)	Operating profit/Paid-in capital	0.9987	Relación fuerte que indica cómo los beneficios operativos por acción se alinean con los retornos sobre el capital.
After-tax Net Profit Growth Rate	Regular Net Profit Growth Rate	0.9962	Correlación que refleja cómo el crecimiento de las ganancias netas sigue patrones similares antes.

Nota: *Elaboración propia como parte del análisis de datos financieros de empresas taiwanesas (1999-2009).*

La siguiente tabla presenta las variables con vectores opuestos, lo que indica una correlación negativa fuerte entre ellas. Esto significa que cuando una variable aumenta, la otra tiende a disminuir. Estas relaciones resaltan interacciones inversas clave en las métricas financieras analizadas.

Tabla 12

Vectores Opuestos (Correlación Negativa)

Variable 1	Variable 2	Correlación	Interpretación
Debt ratio %	Net worth/Assets	-1.000	Relación perfecta inversa: A medida que aumenta la proporción de deuda, disminuye el patrimonio neto en relación con los activos.
Net worth/Assets	Current Liability to Assets	-0.8426	A medida que el patrimonio neto aumenta, la proporción de pasivos corrientes frente a los activos totales disminuye significativamente.
Borrowing dependency	Net Income to Stockholder's Equity	-0.8065	La dependencia de financiamiento externo reduce el retorno financiero para los accionistas.
Net Income to Stockholder's Equity	Liability to Equity	-0.7918	Un alto nivel de deuda frente al patrimonio está asociado con un menor retorno sobre el capital de los accionistas.
Contingent liabilities/Net worth	Working Capital/Equity	-0.7678	Las obligaciones contingentes frente al patrimonio neto están inversamente relacionadas con el capital de trabajo disponible.

Current Liabilities/Equity	Net Income to Stockholder's Equity	-0.7496	A medida que aumenta la proporción de pasivos frente al patrimonio, el retorno financiero para los accionistas disminuye.
----------------------------	------------------------------------	---------	---

Nota: Elaboración propia como parte del análisis de datos financieros de empresas taiwanesas (1999-2009).

La siguiente tabla muestra las variables no correlacionadas, es decir, aquellas que no presentan una relación significativa entre sí. Estas variables operan de manera independiente, lo que indica que cada una mide aspectos diferentes de la salud financiera de las empresas.

Tabla 13

Vectores No Correlacionados

Variable 1	Variable 2	Correlación	Interpretación
Total Asset Turnover	Gross Profit to Sales	-0.0997	No existe relación significativa entre la rotación de activos totales y las ganancias brutas. Reflejan dimensiones independientes.
Operating Gross Margin	Total Asset Turnover	-0.0997	La eficiencia operativa no está directamente vinculada con la rotación de activos totales.
CFO to Assets	Current Liability to Current Assets	-0.0996	El flujo de caja operativo frente a los activos no depende de la proporción de pasivos corrientes.
Cash Flow Per Share	Borrowing dependency	-0.0992	El flujo de caja por acción no muestra dependencia del financiamiento externo.

ROA(A) before interest and % after tax	Total assets to GNP price	-0.0989	La rentabilidad sobre activos no está directamente relacionada con la proporción de activos frente al precio del GNP.
CFO to Assets	Liability to Equity	-0.0985	El flujo de caja operativo no está significativamente correlacionado con la relación entre deuda y patrimonio.

Nota: Elaboración propia como parte del análisis de datos financieros de empresas taiwanesas (1999-2009).

Impacto General del HJ-Biplot:

1. Relación entre Empresas y Variables:

El HJ-Biplot permitió analizar simultáneamente cómo las variables influyen en las empresas y cómo estas se distribuyen en función de sus características financieras.

2. Identificación de Factores Clave:

Variables como Operating Profit Rate y Debt Ratio fueron fundamentales para diferenciar las empresas según su desempeño y riesgo financiero.

3. Diferenciación de Grupos:

Las empresas que quiebran y no quiebran se distribuyeron de manera distinta en el espacio proyectado, validando la capacidad del HJ-Biplot para reflejar diferencias estructurales significativas.

4. Validación de Resultados Previos:

Las observaciones del HJ-Biplot reforzaron los hallazgos del ACP y el clustering, proporcionando una perspectiva visual y analítica robusta para comprender la salud financiera de las empresas.

3.3 Resultados de STATIS Dual: Integración de Tablas

El análisis STATIS Dual permitió integrar y comparar las tablas de empresas que quiebran y no quiebran en un espacio común, destacando diferencias estructurales críticas y patrones clave en las métricas financieras. Este enfoque multivariante ofreció una perspectiva única al combinar observaciones y variables en un marco analítico unificado.

Tabla 14

Comparación de Métricas Financieras

Métrica	Empresas Quebradas	Empresas No Quebradas	Diferencia Absoluta	Interpretación
Debt Ratio %	1.37	-0.0457	1.42	Empresas quebradas tienen niveles significativamente más altos de deuda, lo que refleja un apalancamiento crítico.
Liability to Equity	0.91	-0.0304	0.94	La dependencia del financiamiento externo es mucho mayor en empresas quebradas, aumentando su vulnerabilidad.
Operating Profit Rate	-0.0013	0.00004	0.0013	Empresas no quebradas tienen márgenes operativos más altos, lo que indica mayor eficiencia y sostenibilidad.
Quick Ratio	0.14	-0.0046	0.14	Empresas quebradas tienen ratios de liquidez más bajos, reflejando dificultades para cubrir obligaciones inmediatas.

Nota: Elaboración propia como parte del análisis de datos financieros de empresas taiwanesas (1999-2009).

Proyección de Variables y Observaciones:

1. Debt Ratio: Las empresas que quiebran tienden a mostrar altos valores en esta métrica, reflejando una fuerte dependencia del apalancamiento financiero. Esto sugiere que estas empresas no logran mantener un equilibrio adecuado entre deuda y activos, exponiéndolas a riesgos significativos durante períodos de estrés financiero.
2. Liability to Equity: Este indicador confirma una elevada proporción de deuda respecto al patrimonio en las empresas quebradas, lo que limita su capacidad para responder a obligaciones financieras imprevistas.
3. Operating Profit Rate: Las empresas financieramente estables están vinculadas a altos valores en esta métrica, lo que refleja eficiencia operativa y capacidad para generar ingresos sostenibles.
4. Quick Ratio: Este indicador de liquidez muestra que las empresas que no quiebran son capaces de cubrir obligaciones a corto plazo sin depender de la venta de activos no líquidos.

Proyección de Observaciones y Diferenciación entre Grupos:

Las empresas fueron proyectadas en el espacio común definido por las variables clave.

Tabla 15

Características Clave e Interpretación Financiera

Categoría	Características Clave	Interpretación Financiera
Empresas No Quiebradas	- Agrupadas cerca de vectores asociados con rentabilidad y liquidez. - Estructura financiera sólida y bien gestionada.	Las empresas no quebradas tienen una gestión eficiente de sus recursos financieros, protegiéndolas contra eventos

inesperados y garantizando su sostenibilidad.

Empresas - Proyectadas cerca de vectores Las empresas quebradas carecen
Quebradas relacionados con deuda y de rentabilidad y tienen altos
apalancamiento. niveles de deuda, lo que
- Patrón de alto riesgo. incrementa su vulnerabilidad ante
cambios en las condiciones
económicas.

***Nota:** Elaboración propia como parte del análisis de datos financieros de empresas taiwanesas (1999-2009).*

Proyección Bidimensional:

La visualización en el espacio STATIS Dual permitió observar una clara separación entre empresas que quiebran y no quiebran:

1. Empresas que no quiebran: Agrupadas en zonas dominadas por vectores de rentabilidad (Operating Profit Rate) y liquidez (Quick Ratio).
2. Empresas que quiebran: Más cercanas a vectores como Debt Ratio y Liability to Equity, que reflejan altos niveles de deuda y dependencia financiera.

Resultados Clave del STATIS Dual:

1. Factores de Riesgo Identificados: Las empresas que quiebran tienen un perfil financiero caracterizado por altos niveles de deuda (Debt Ratio) y baja capacidad para generar ingresos (Net Income to Stockholder's Equity).

2. Factores de Estabilidad: Las empresas que no quiebran reflejan una gestión eficiente de la rentabilidad (Operating Profit Rate) y la liquidez (Quick Ratio), lo que garantiza su sostenibilidad financiera.

3. Visualización Integral: La representación en el espacio STATIS Dual permitió identificar patrones claros de diferenciación entre los grupos, proporcionando una herramienta poderosa para entender las dinámicas entre empresas y variables.

3.4 Evaluación de las Técnicas Aplicadas

En esta sección, se evaluarán las técnicas multivariantes utilizadas en el proyecto, específicamente el HJ-Biplot y el STATIS Dual, analizando sus contribuciones, limitaciones y efectividad en la predicción de quiebras empresariales. Esta evaluación se realiza desde una perspectiva cualitativa y cuantitativa, incluyendo ejemplos concretos y su impacto en los resultados obtenidos.

3.3.1. HJ – BIPLLOT

El HJ-Biplot es una técnica gráfica que permite proyectar simultáneamente observaciones (empresas) y variables en un espacio común, lo que facilita la interpretación de las relaciones entre ambas dimensiones.

3.3.1.1. Contribuciones del HJ-Biplot:

1. Visualización Gráfica:

Proporcionó una representación visual clara de las relaciones entre empresas y variables, permitiendo identificar patrones clave.

Por ejemplo, las empresas que quiebran se agruparon cerca de vectores como Debt Ratio y Liability to Equity, mientras que las empresas estables estuvieron asociadas a Operating Profit Rate y Quick Ratio.

2. Identificación de Factores Clave:

Destacó variables dominantes como Operating Gross Margin y Net Income to Total Assets, que influyen significativamente en la estabilidad financiera.

3. Facilitó la Comparación:

Permitió comparar empresas que quiebran y no quiebran mediante la proximidad a vectores específicos, evidenciando diferencias significativas.

3.3.1.2. Limitaciones del HJ-Biplot:

1. Representación Limitada:

En conjuntos de datos de alta dimensionalidad, la proyección en dos componentes principales puede no capturar toda la variabilidad.

2. Interpretación Compleja:

Para usuarios sin experiencia, interpretar vectores y proximidades puede ser desafiante.

3.3.2. STATIS Dual

El STATIS Dual es una técnica multivariante diseñada para integrar y analizar múltiples tablas de datos, permitiendo comparar grupos específicos, como empresas que quiebran y no quiebran.

3.3.2.1. Contribuciones del STATIS Dual:

1. Integración de Tablas:

Comparó las tablas de empresas que quiebran y no quiebran, identificando diferencias estructurales clave en variables como Debt Ratio y Quick Ratio.

2. Evolución de Variables:

Analizó cómo ciertas variables financieras evolucionan en ambos grupos, proporcionando información sobre patrones temporales o diferencias estructurales.

3. Identificación de Patrones:

Resaltó cómo las métricas críticas varían entre grupos, ayudando a priorizar estrategias de mitigación de riesgos.

3.3.2.2. Limitaciones del STATIS Dual:

1. Interpretación Compleja:

Requiere conocimientos avanzados para interpretar adecuadamente los resultados.

2. Dependencia de Datos Estandarizados:

La calidad del análisis depende de la consistencia y preparación previa de los datos.

3.3.3. Comparación de las Técnicas

1. HJ-Biplot:

Mejor para visualizar relaciones entre empresas y variables en un espacio común.

Ideal para explorar patrones iniciales.

2. STATIS Dual:

Excelente para comparar y analizar diferencias entre tablas de datos (empresas que quiebran y no quiebran).

Requiere mayor preparación e interpretación.

4. CONCLUSIONES

1. Se analizaron los datos del conjunto Taiwanese Bankruptcy Prediction utilizando técnicas multivariantes, específicamente el HJ-Biplot y el STATIS Dual, logrando identificar patrones y asociaciones clave entre las empresas que quiebran y las que no quiebran. Estas técnicas permitieron diferenciar con precisión a las empresas mediante métricas críticas como Debt Ratio, Operating Profit Rate y Quick Ratio, destacando los factores que afectan directamente la probabilidad de quiebra empresarial.
2. Exploración de variables financieras clave, se exploraron las variables financieras del conjunto de datos mediante el Análisis de Componentes Principales (ACP), matrices de correlación y análisis de clustering:
 - ACP1 (13.35%) asociado a métricas de rentabilidad como Operating Profit Rate y Net Income to Total Assets.
 - ACP2 (7.21%) vinculado al apalancamiento financiero, dominado por Debt Ratio y Liability to Equity.

Además, el análisis de correlaciones reveló que la rentabilidad operativa está fuertemente relacionada con el desempeño financiero, mientras que la deuda mantiene una relación negativa con el retorno sobre el patrimonio.

3. Identificación de variables influyentes en la probabilidad de quiebra, se identificaron las variables más relevantes para la predicción de quiebra, destacando las diferencias entre empresas que quiebran y las que no:
 - Debt Ratio fue el principal indicador de vulnerabilidad financiera, seguido de Liability to Equity, que refleja la dependencia del financiamiento externo.

- Por el contrario, métricas como Quick Ratio y Operating Profit Rate se asociaron con empresas estables, resaltando su importancia para la sostenibilidad financiera.
4. Aplicación del HJ-Biplot que permitió visualizar y analizar gráficamente las relaciones entre empresas y variables:
- Las empresas que no quiebran se agruparon en torno a variables asociadas con rentabilidad y liquidez, como Operating Profit Rate y Net Income to Total Assets.
 - Las empresas que quiebran se proyectaron cerca de vectores asociados con deuda, como Debt Ratio, evidenciando patrones de riesgo financiero.
5. El método STATIS Dual permitió integrar y comparar dos tablas (empresas que quiebran y no quiebran) en un mismo espacio de análisis:
- Empresas quebradas mostraron altos niveles de Debt Ratio y Liability to Equity, indicando una fuerte dependencia del endeudamiento.
 - Empresas estables destacaron por Quick Ratio y Operating Profit Rate, confirmando que la gestión de liquidez y rentabilidad es clave para evitar la quiebra. La integración de tablas permitió corroborar la importancia de estas métricas para diferenciar ambos grupos empresariales.
6. Síntesis y validación de patrones clave, donde se validaron y sintetizaron los patrones identificados a lo largo del análisis, demostrando que:
- Las empresas que quiebran se caracterizan por altos niveles de deuda, baja rentabilidad y débil liquidez.
 - Las empresas que no quiebran presentan una gestión financiera eficiente, con mejores márgenes operativos y capacidad de respuesta ante riesgos.

- Técnicas como HJ-Biplot y STATIS Dual resultaron útiles para visualizar, comparar y diferenciar claramente a los grupos de empresas según su nivel de riesgo.

RECOMENDACIONES Y TRABAJOS FUTUROS

Recomendaciones

- 1.- Se recomienda ampliar la base de datos utilizada, integrando información financiera de empresas de diferentes sectores y regiones geográficas, más allá de las empresas taiwanesas. Esto permitiría generalizar los resultados y validar la aplicabilidad del modelo en distintos contextos empresariales.
- 2.- Es aconsejable incluir variables no financieras, como factores macroeconómicos o indicadores de mercado, que podrían enriquecer los análisis y mejorar la precisión de las predicciones realizadas con HJ-Biplot y STATIS Dual.
- 3.- Se sugiere evaluar periódicamente la calidad y consistencia de los datos, implementando herramientas automatizadas para la detección de inconsistencias y la actualización continua de los registros.
- 4.- Se recomienda utilizar visualizaciones más intuitivas para presentar los resultados, facilitando la interpretación de patrones y asociaciones clave, especialmente para audiencias no técnicas.

Trabajos Futuros

- 1.- Extender el análisis para incluir predicciones a diferentes horizontes temporales, permitiendo anticipar problemas financieros no solo a corto plazo, sino también a mediano y largo plazo.
- 2.- Desarrollar un sistema automatizado que combine el preprocesamiento de datos, la ejecución de modelos multivariantes y la generación de reportes, asegurando la eficiencia y la repetibilidad del proceso analítico.
- 3.- Realizar pruebas en sectores específicos, como manufactura o tecnología, para evaluar cómo varían los patrones financieros y las métricas clave según las características del sector industrial.
- 4.- Implementar una validación externa del modelo utilizando datos de otros países o bases de datos internacionales, lo que permitiría comprobar la eficacia y robustez del enfoque en diferentes contextos económicos.

REFERENCIAS

Tascón, M. T., & Castaño, F. J. (2012). Variables y modelos para la identificación y predicción del fracaso empresarial: Revisión de la investigación empírica reciente. *Revista de Contabilidad - Spanish Accounting Review*, 15(1), 7–58.

Altman, E. I. (1968). Financial ratios, discriminant analysis and the prediction of corporate bankruptcy. *The Journal of Finance*, 23(4), 589–609. <https://doi.org/10.1111/j.1540-6261.1968.tb00843.x>

Beaver, W. H. (1966). Financial ratios as predictors of failure. *Journal of Accounting Research*, 4, 71–111. <https://doi.org/10.2307/2490171>

Chava, S., & Jarrow, R. A. (2004). Bankruptcy prediction with industry effects. *Review of Finance*, 8(4), 537–569. <https://doi.org/10.1093/rof/8.4.537>

Tsai, C.-F. (2013). Combining cluster analysis with classifier ensembles to predict financial distress. *Information Fusion*, 24, 31–38. <https://doi.org/10.1016/j.inffus.2012.06.003>

Ohlson, J. A. (1980). Financial ratios and the probabilistic prediction of bankruptcy. *Journal of Accounting Research*, 18(1), 109–131. <https://doi.org/10.2307/2490395>

Moreno, E., & Bravo, F. (2018). Análisis de la probabilidad de quiebra de las empresas cotizadas españolas. *Revista de Estudios Empresariales. Segunda Época*, 2, 57–72. <https://doi.org/10.17561/rec.v2018n2.3>

Altman, E. I., & Hotchkiss, E. (2006). *Corporate financial distress and bankruptcy: Predict and avoid bankruptcy, analyze and invest in distressed debt* (3rd ed.). Wiley.

Zmijewski, M. E. (1984). Methodological issues related to the estimation of financial distress prediction models. *Journal of Accounting Research*, 22, 59–82. <https://doi.org/10.2307/2490859>

Fernández Aguirre, K., Modroño Herrán, J. I., & Landaluce Calvo, M. I. (2004). ACM y Statist dual ponderado: Dos técnicas complementarias para analizar una visión de la cultura de la universidad. *Estadística Española*, 46(156), 205–228.

Díaz-Faes, A. A., González-Albo, B., Galindo, M. P., & Bordons, M. (2013). HJ-Biplot como herramienta de inspección de matrices de datos bibliométricos. *Revista Española de Documentación Científica*, 36(1), e001. <https://doi.org/10.3989/redc.2013.1.988>

Altman, E. I. (1968). Financial ratios, discriminant analysis and the prediction of corporate bankruptcy. *The Journal of Finance*, 23(4), 589–609. <https://doi.org/10.2307/2978933>

Reichardt, C. S., & Gollob, H. F. (1986). Satisfying the constraints of causal modeling. *Evolutionary Ecology*, 1(4), 349–366. <https://doi.org/10.1002/ev.1436>

RomaniChocce, G. A., ArocaGonzález, P., AguirreAguirre, N., LeitonVega, P., & MuñozCarrazana, J. (2002). Modelos de clasificación y predicción de quiebra de empresas: Una aplicación a empresas chilenas. *Forum Empresarial*, 7(1), 33–50. <https://www.redalyc.org/articulo.oa?id=63170102>

Rincón, C. C., & Rodríguez, E. O. (2019). *Análisis financiero para la predicción de quiebra empresarial, en el sector de la construcción en Colombia* [Tesis de pregrado, Fundación Universitaria Konrad Lorenz]. <http://hdl.handle.net/20.500.12749/2052>

Franco, Y. A. (2023). Un análisis bibliométrico de la predicción de quiebra empresarial con machine learning. *ODEON*, 22, 87–126. <https://doi.org/10.18601/17941113.n22.04>

Caballero-Juliá, D., Galindo Villardón, P., & García, M. C. (2017). JK-Meta-Biplot y STATIS Dual como herramientas de análisis de tablas textuales múltiples. *RISTI - Revista Ibérica de Sistemas e Tecnologías de Informação*, 2017(25), 18–33. <https://doi.org/10.17013/risti.25.18-33>

Castillo Elizondo, W., & Gonzalez Varela, J. (1998). Statis dual: Software y análisis de datos reales. *Revista de Matemática: Teoría y Aplicaciones*, 2, 149–162.

Corrales, D., & Rodríguez, O. (2014). *INTERSTATIS: El método STATIS para datos de tipo intervalo*.

Díaz-Faes, A. A., González-Albo, B., Galindo, P., & Bordons, M. (2013). HJ-Biplot como herramienta de inspección de matrices de datos bibliométricos. *Revista Española de Documentación Científica*, 36(1), e001. <https://doi.org/10.3989/redc.2013.1.988>

Hair, J. F., Black, W. C., Babin, B. J., & Anderson, R. E. (2019). *Multivariate data analysis* (8th ed.). Cengage Learning.

Hosmer, D. W., & Lemeshow, S. (2000). *Applied logistic regression* (2nd ed.). John Wiley & Sons. <https://doi.org/10.1002/0471722146>

Morrás Lorenzo, L. (2023). *Predicción de quiebras en el sector de la restauración en España* [Tesis de grado, Universidad Pontificia Comillas]. <http://hdl.handle.net/11531/69548>

Gabriel, K. R. (1971). The biplot graphic display of matrices with application to principal component analysis. *Biometrika*, 58(3), 453–467.

Galindo, P. (1986). Una alternativa de representación simultánea: HJ-Biplot. *Quetiio*, 10, 13–23.

James, G., Witten, D., Hastie, T., & Tibshirani, R. (2013). *An introduction to statistical learning with applications in R*. Springer. <https://doi.org/10.1007/978-1-4614-7138-7>

Johnson, R. A., & Wichern, D. W. (2007). *Applied multivariate statistical analysis* (6th ed.). Pearson Prentice Hall.

Leeuw, J., & Mair, P. (2009). Gifi methods for optimal scaling in R: The package homals. *Journal of Statistical Software*, 31(4), 1–19. <https://doi.org/10.18637/jss.v031.i04>

Lechevalier, F. (1990). L'analyse en composantes conjointes d'une famille de triplets indexés. *Statistique et Analyse des Données*, 15(2), 35–49. http://www.numdam.org/item?id=SAD_1990__15_2_35_0

Martínez, W., & Martínez, A. (2008). *Computational statistics handbook with MATLAB* (2nd ed.). Chapman & Hall/CRC.

Ooghe, H., & De Prijcker, S. (2008). Failure processes and causes of company bankruptcy: A typology. *Management Decision*, 46(2), 223–242. <https://doi.org/10.1108/00251740810854131>

Ortega-Gómez, E., Rodríguez, C. C., & Cubilla, M. I. (2014). Indicadores de sostenibilidad: Un análisis multivariante aplicando HJ-Biplot. *Revista Colón Ciencias, Tecnología y Negocios*, 1(2), 30–46.

Pindado, J., Rodrigues, L., & de la Torre, C. (2008). Estimating financial distress likelihood. *Journal of Business Research*, 61(9), 995–1003. <https://doi.org/10.1016/j.jbusres.2007.10.006>

Silva Viamonte, C. F. (2024). HJ-Biplot as an exploratory methodology for the multidimensional analysis of the sustainable development goals (SDG) at the municipality level in Bolivia. *Revista Varianza*, 23, 29–55. <https://doi.org/10.53287/xttp2122se37f>

Taiwan Economic Journal (TEJ). (1999–2009). *TEJ database*.

Tenenhaus, M. (1998). *La régression PLS: Théorie et pratique* [Partial Least Squares regression: Theory and practice]. Éditions Technip.

Tsai, C. F. (2009). Feature selection in bankruptcy prediction. *Knowledge-Based Systems*, 22(2), 120–127. <https://doi.org/10.1016/j.knosys.2008.08.002>

Zhang, G., Hu, M. Y., Patuwo, B. E., & Indro, D. C. (1999). Artificial neural networks in bankruptcy prediction: General framework and cross-validation analysis. *European Journal of Operational Research*, 116(1), 16–32. [https://doi.org/10.1016/S0377-2217\(98\)00051-4](https://doi.org/10.1016/S0377-2217(98)00051-4)

Aldazábal Contreras, J. C., & Napán Vera, A. F. (2014). Análisis discriminante aplicado a modelos de predicción de quiebra. *Quipukamayoc*, 22(42), 53–59.

Campanaro, R., Díaz, D., Marchese, A., Sepliarsky, P., Viola, M., Riveros, M., Carrizo, J., Ferraro, F., Díaz Toledo, A., & Gardenal, S. (2016). Modelos de predictibilidad de quiebras e insolvencia basados en análisis de estados financieros. En *Vigesimoprimeras Jornadas "Investigaciones en la Facultad" de Ciencias Económicas y Estadística*.

Fayyad, U., Piatetsky-Shapiro, G., & Smyth, P. (1996). From data mining to knowledge discovery in databases. *AI Magazine*, 17(3), 37–54. <https://doi.org/10.1609/aimag.v17i3.1230>

Franco, Y. A. (2023). Un análisis bibliométrico de la predicción de quiebra empresarial con machine learning. *ODEON*, 22, 87–126. <https://doi.org/10.18601/17941113.n22.04> (Original work published July 4, 2023)

Molina Panchi, P. A., Molina Panchi, D. F., & Flores Tapia, C. E. (2022). Modelo de predicción de quiebra Z2 de Altman de análisis multivariable en empresas del sector inmobiliario de la provincia de Pichincha. *Revista Científica Ecociencia*, 9(2), 53–76. <https://doi.org/10.21855/ecociencia.92.643>

Rivadeneira, J., Saltos, R., Rivera, M., & Carpio, R. (2022). Predicción de la quiebra empresarial en el sector agroindustrial de la ciudad de Machala. *Avances en Ciencias e Ingenierías*, 14(2). <https://doi.org/10.18272/aci.v14i2.2695>

ANEXOS

ANEXO A: CERTIFICACIÓN DE VALIDACIÓN DE DATOS FINANCIEROS

CERTIFICACIÓN

Yo, **NOEL SANTIAGO, QUINTANILLA CASTRO** con número de colegiatura N°06795 del Colegio de Contadores Públicos de Arequipa, certifico que he revisado los datos financieros utilizados en el estudio titulado:

“ANÁLISIS MULTIVARIANTE APLICADO A LA PREDICCIÓN DE QUIEBRAS EMPRESARIALES: HJ-BILOT Y STATIS DUAL COMO HERRAMIENTAS DE DIAGNÓSTICO FINANCIERO”

Dichos datos corresponden a información obtenida de fuentes verificables y han sido analizados conforme a criterios contables y financieros. La información utilizada refleja la realidad de las empresas evaluadas en el estudio y ha sido procesada correctamente para su análisis.

Esta certificación se emite para fines académicos y de investigación, respaldando la validez de los datos utilizados en el presente trabajo.

Atentamente,

C.P.C.NOEL SANTIAGO, QUINTANILLA CASTRO

Número de Colegiatura: 06795

Firma:



Sello:

NOEL SANTIAGO QUINTANILLA CASTRO
Contador Público Colegiado
Matrícula N° 6795
AREQUIPA

ANEXO B: Plan de Tesis Aprobado

Universidad Católica de Santa María

Facultad de Ciencias e Ingenierías Físicas y Formales

Escuela Profesional de Ingeniería de Sistemas



Análisis Multivariante Aplicado a la Predicción de Quiebras Empresariales: Hj-Biplot y Statis Dual como Herramientas de Diagnóstico Financiero

Línea: Sistemas de Información y Base de Datos

Sublínea: Tecnologías de la Información

Proyecto de Tesis Presentado Para Optar el Título Profesional de Ingeniero de Sistemas

Autor:

Bach. Juan Luis Apaza Machicao

Arequipa-Perú

2024

1. Planteamiento De La Investigación

1.1. Planteamiento del Problema

El poder indetificar la quiebra de una empresa en una etapa temprana es un reto y el poder analizar los factores que llevaron a ese punto aún más, al poder realizarlo se podrian evitar consecuencias que afectan tanto como a personas propias de la empresa como los accionistas, empleados y personas externas como acreedores, proveedores, pudiendo provocar efector indirectos que afectarían negativamente la estabilidad económica de una región o algun sector en específico (Altman, 1968; Beaver, 1966). Para esto se debe garantizar una toma de decisiones oportuna e informada basada en los datos disponibles (Ohlson, 1980).

Poder predecir o analizar los factores que llevan al fracaso empresarial se vuelve en un tema de atracción para diferentes actores como bancos, inversiones, agencias gubernamentales e instituciones que gestionan los riesgos financieros (Zmijewski, 1984; Altman & Hotchkiss, 2006). A pesar de ello se sigue trabajando con enfoques tradicionales que tienen limitaciones significativas para poder analizar las interacciones entre las diferentes variables financieras (Beaver, 1966).

La falta de teorías firmes es un factor que carece siendo un tema de interés y que a pesar de los diferentes esfuerzos el poder crear un mapa de relaciones que no expliquen los factores que llevaron al fracaso empresarial (Tascón y Castaño, 2012).

La análisis de quiebra históricamente se basa en modelos de análisis discriminante o el modelo de Altman Z-score, que tiene como puntos clave lo ratios financieros como los son (liquidez, rentabilidad, deuda, etc), que sirven para determinar si la empresa está un posible riesgo de quiebra, Moreno y Bravo (2018) en su análisis de probabilidad de quiebra en empresas españolas indicaron que al usar los modelo de Altman Z-score y el modelo de indicador de Amat

necesitan ser adaptados para ser usados en diferentes contextos económicos de cada país o región y que deben ser complementados con más modelos dado que estos modelos tienden a subestimar la correlación o interdependencia entre las variables, por lo que al analizarlas por separado se puede perder información relevante que nos indicaría el estado financiero de la empresa, a medida que la economía se vuelve compleja, esta forma que se tiene para abordar los aspectos importantes sobre la estabilidad empresarial son insuficientes.

La base que se propone revisar corresponde al mercado taiwanés que es un ejemplo de necesidad de mejores herramientas de predicción de quiebres empresariales. Durante los años 1999 y 2009 Taiwán experimentó cambios significativos en su estructura económica, caracterizados por la competencia global, inestabilidad económica global y la liberación del mercado, en este contexto muchas empresas taiwanesas se enfrentaron a dificultades financieras y los enfoques tradicionales para predecir el fracaso empresarial no han logrado predecir estos eventos de manera efectiva siendo el principal impedimento la dificultad de manejar grandes cantidades de datos financieros y la interacción entre ellos. Más adelante, se detallará la base de datos utilizada en este análisis (Tsai, 2013; Chava & Jarrow, 2004).

Ante estas limitantes se abre la necesidad de usar técnicas avanzadas que puedan analizar grandes conjuntos de datos y que a su vez se proporcione una mejor comprensión de las variables que predominan en el quiebre empresarial. El uso de métodos multivariantes dado que el modelo HJ-Biplot y STATIS Dual proporcionan un análisis profundo y detallado, dado que nos permiten analizar múltiples variables financieras, capturar relaciones entre ellas y proporcionar visualizaciones gráficas que facilitan el entendimiento de los resultados.

HJ-Biplot es un método de análisis multivariante utilizado para visualizar variables y observaciones en un espacio bidimensional que nos proporciona una imagen intuitiva y clara de

las relaciones multivariantes (Díaz-Faes, González-Albo, Galindo, & Bordons, 2013). Esto es potencialmente útil para el contexto de la predicción de quiebras, donde existen muchas variables y los patrones ocultos son difíciles de detectar.

STATIS Dual es un método que le permite combinar múltiples tablas de datos sobre empresas que quiebran y no quiebran, y tiene dos tablas de conjuntos de datos para un análisis detallado y la identificación de patrones que se destacan para cada conjunto de datos individual, Fernández Aguirre, Modroño Herrán y Landaluce Calvo (2004) indican que la metodología STATIS Dual posibilita el análisis simultáneo de múltiples tablas, lo que enriquece los objetivos del estudio a llevar a cabo.

Por lo consiguiente este trabajo de investigación propone el uso de dos técnicas multivariantes las cuales son el HJ-Biplot y Statis Dual, para la analizar las quiebras empresariales en Taiwan, utilizando un conjunto de datos proporcionado por Deron Liang y Chih-Fong Tsai. El dataset con el que va a trabajar contiene una cantidad amplia de variables que nos ayudaran a comprender los rasgos financieros que pueden influir en el fracaso empresarial. Se busca tener interpretaciones claras y compresibles de los factores que aportan a la quiebra de una empresa, que será un aporte de gran valor tanto para gestores financieros como para regulares y aplicables a distintos conjuntos de datos.

1.2. Objetivos de la investigación

1.1.2. General

Analizar el conjunto de datos *Taiwanese Bankruptcy Prediction* utilizando técnicas multivariantes específicamente el HJ-Biplot y el STATIS Dual, con el objetivo de encontrar patrones y asociaciones clave que nos ayudan a indentificar predicción de quiebra.

2.1.2. Específico

- Analizar las variables financieras del conjunto de datos para crear conocimiento relevante mediante técnicas como el análisis de componente principales, la construcción de matrices de correlaciones, análisis de clustering, entre otros métodos.
- Identificar las variables financieras y empresariales más relevantes del conjunto de datos Taiwanese Bankruptcy Prediction que influyen en la predicción de quiebra, utilizando para ello el subconjunto de empresas que han experimentado quiebra.
- Aplicar el método multivariante HJ-Biplot para visualizar y describir las relaciones entre las variables e identificar los factores clave que influyen en la probabilidad de quiebra empresarial.
- Aplicar método multivariante STATIS Dual para integrar y analizar las dos tablas de entrada de empresas que quiebra y no quiebran, permitiendo observar la evolución de las variables y su impacto en la predicción de quiebras empresariales.
- Analizar los resultados obtenidos con relación a las empresas que quiebran y no quiebran.
- Validar y sintetizar los patrones y asociaciones clave identificados a través de las técnicas multivariantes aplicadas, con el propósito de proporcionar conclusiones que faciliten la comprensión de los factores que influyen en la predicción de quiebras empresariales en el contexto del conjunto de datos Taiwanese

Bankruptcy Prediction.

1.3. Preguntas de investigación

¿Qué variables financieras son más relevantes en la prever de quiebra empresarial?

Esta pregunta busca indentificar las variables financieras que tienen mayor influencia en la predicción de quiebras empresariales.

¿Qué patrones se pueden detectar en la información financiera de las empresas que quiebran en comparación con las que no lo hacen?

Esta pregunta busca descubrir los patrones y relaciones que existen en los datos financieros de las empresas que han quebrado y las que han permanecido solventes. Mediante la visualización de los resultados del análisis HJ-Biplot y STATIS Dual, la influencia conjunta de varias variables, o cómo ciertos indicadores pueden servir como señales tempranas de quiebra.

¿Qué tan precisos son los modelos predictivos basados en los métodos multivariantes HJ-Biplot y STATIS Dual en comparación de los métodos tradicionales de predicción de quiebra?

Esta pregunta buscar evaluar la eficacia de los métodos multivariantes HJ-Biplot y Statist Dual frente a los métodos tradicionales como el modelo de Altman Z-score o el análisis discriminante.

1.4. Línea y Sub – línea de investigación

Línea: Análisis Multivariante

Sub – línea: Métodos de Reducción de Dimensionalidad y Exploración de Estructuras Multivariadas

1.5. Palabras Clave

Quiebra Empresarial, Predicción Financiera, HJ-Biplot, STATIS Dual, Análisis Multivariante, Variables Financieras, Datos Históricos, Taiwan Economic Journal

1.6. Solución propuesta

La solución propuesta en esta investigación se enfoca en realizar un análisis avanzado multivariante que permite indentificar y predecir la probabilidad de quiebra empresarial empleando la base de datos sobre quiebras del Taiwan Economic Journal correspondientes a los años 1999-2009. Para ellos se propone aplicar dos técnicas avanzadas multivariantes que son el HJ – Biplot y STATIS Dual, estos métodos permiten analizar de manera simultánea un gran número de variables identificando patrones y correlaciones entre las variables que impactan en la quiebra empresarial.

HJ-Biplot es la representación gráfica de las columnas y filas de una matriz de datos, sin perder la representatividad de estas, esto para poder representar las variables financieras a analizar y ver como interactúan gráficamente para analizar el quiebre empresarial, este es sumamente importante dado que nos da el panorama de que características financieras son importantes para este estudio.

STATIS Dual, es útil dado que no permite analizar múltiples tablas de datos, en este caso para las dos tablas de entrada, empresas que quiebran y empresas que no quiebran, lo cual nos permitirá entender que influye en cada dataset analizado y cuáles son los factores de riesgo.

La solución que se propone para este trabajo de investigación tiene como finalidad superar los límites de los modelos tradicionales de predicción de quiebres empresariales que no tienen la capacidad de capturar la relaciones entre las variables, el uso de estos métodos multivariantes avanzados tiene como objetivo garantizar una mayor precisión y compresión de las variables.

1.1.6. Justificación e importancia

La justificación de este trabajo de investigación se deriva de la necesidad de utilizar técnicas avanzadas y precisas que nos permita analizar los quiebres empresariales siendo cada vez más complicado debido al creciente mercado financiero y globalizado. La importancia que tiene para los actores internos de la empresa que necesitan una información verídica, de igual forma para actores externos como reguladores, analistas financieros, etc. La necesidad de tener modelos que les permita analizar, entender y tomar decisiones anticipadas para disminuir los riesgos de insolvencia y la quiebra empresarial.

Los modelos usados con el objetivo de detectar la quiebra, como el modelo Altman Z-score o análisis de ratios tiene limitaciones, ya que se basa en un número limitado de variables, sin tener en consideración las relaciones que tienen las variables, por eso estos modelos tienden a perder la precisión, siendo los datos financieros altamente relacionados.

La importancia de este proyecto de tesis radica en el uso de métodos multivariantes avanzados los cuales son HJ-Biplot el STATIS Dual, que tiene como principal ventaja analizar gran cantidad de variables y capturar las relaciones entre las variables y los patrones para analizar que nos permitirá tener una visión detallada de los principales riesgos de quiebras. Estos métodos multivariantes superan las limitaciones de los modelos tradicionales al poder analizar gran cantidad de datos y poder integrarlos, para luego poder representarlos gráficamente facilitando la interpretación de los resultados y mejorando la toma de decisiones.

Este proyecto de tesis es de gran aporte aplicado a diferentes empresas que tienen un crecimiento significativo, pero necesitan tener un panorama amplio de la situación económica, la base de datos extraída de los registros históricos de empresas taiwanesas entre los años 1999 y

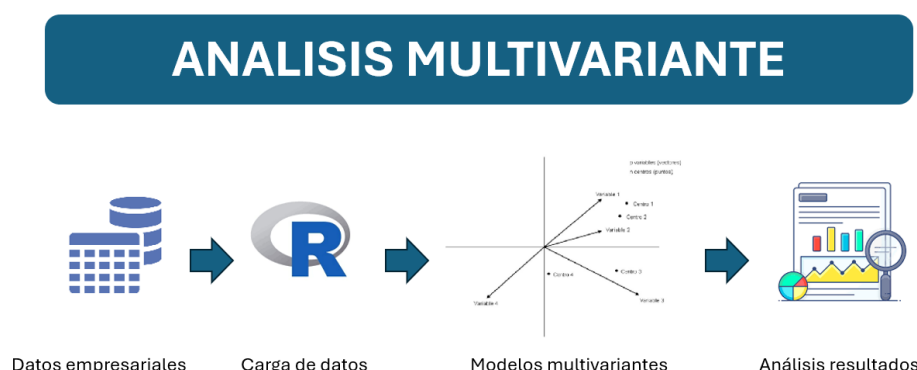
2009, que permitirá evaluar la eficacia de los métodos multivariantes propuestos, que nos permitirá analizar los quiebres empresariales en mercados y empresas con característica similares.

Desde la parte académica este proyecto de tesis aporta un gran valor en el campo de análisis multivariante y predicción financiera, el uso del HJ-Biplot y STATIS Dual para analizar quiebres empresariales, abriendo oportunidades para el desarrollo de modelos multivariantes aplicado a diferentes contextos económicos.

2.1.6. Descripción de la solución

En este proyecto de tesis se usará dos modelos de análisis multivariantes HJ-Biplot y STATIS Dual para analizar gran cantidad de variables financieras, identificando relaciones entre ella y tablas de múltiples de empresa que quiebra o no quiebra, para analizar y entender los factores que influyen en la quiebra empresarial.

Figura 1 Ilustración de la descripción de la solución



Nota: Elaboración propia.

3.1.6. Aporte

El aporte de esta investigación es el uso de técnicas avanzadas de análisis multivariante HJ-Biplot y STATIS Dual para la predicción de quiebra empresarial, donde las técnicas tradicionales

han mostrado limitaciones, esta investigación aporte al campo academico como al campo practico al poder análisis múltiples variables y dimensiones basados en ratios financieros simples.

Aporte Metodológico:

Este trabajo de investigación utiliza técnicas multivariantes como el HJ-Biplot y STATIS Dual, aplicando análisis complejos a gran cantidad de datos financieros e identificando las relaciones entre las variables, que ofrece la oportunidad para estudiar la situación financiera empresarial a detalle.

Aporte Práctico:

Este trabajo de investigación proporciona una contribución significativa para el área de gestión de riesgos financieros a darles una herramienta que podrá ser usada en entornos empresariales para poder evaluar de manera eficaz el riegos de insolvencia empresarial, los resultados que se obtendrán permitirá a instituciones financieras mejorar sus modelos de evaluación de riesgos, mejorando la asignación de recursos y decisiones estratégicas, de igual forma se analizar la estabilidad financiera adoptando medidas correctivas en fases tempranas ante problemas financieros.

Contribución a la Gestión de Riesgos Financieros:

El análisis de las variables financieras brindara una herramienta para poder ser usadas en entornos empresariales para medir el nivel de riegos o insolvencia, con estos resultados ayudaría a gestionar los riesgos financieros y aplicar medidas correctivas ante problemas financieros.

Aporte a la Literatura Académica:

En la parte académica este proyecto de tesis amplia el conocimiento usando técnicas multivariantes avanzadas que no sido exploradas en este contexto, el uso de HJ-Biplot y STATIS Dual para analizar las quiebras empresariales abre líneas de investigación y ofrece la base

metodológica que podrá ser usada en futuros estudios que usen el análisis multivariante y la predicción de riesgos financieros ampliado a distintos sectores y regiones.

4.1.6. Descripción de la base de datos

Nombre de la base de datos

Bankruptcy data from the Taiwan Economic Journal for the years 1999–2009

Origen de la base de datos

Los datos fueron recolectados del Taiwan Economic Journal para los años 1999 a 2009. La quiebra de las empresas se definió de acuerdo con las regulaciones empresariales de la Bolsa de Valores de Taiwán.

Descripción general

Esta base de datos contiene información financiera de empresas taiwanesas con el objetivo de predecir la quiebra. Se busca identificar patrones y asociaciones clave que ayuden a comprender mejor los factores que influyen en la quiebra empresarial.

Variables incluidas

Y - Bankrupt?: Clase de etiqueta que indica si la empresa ha quebrado (1) o no (0).

X1 - ROA(C): Rentabilidad sobre activos (C) antes de intereses y depreciación.

X2 - ROA(A): Rentabilidad sobre activos (A) antes de intereses y después de impuestos.

X3 - ROA(B): Rentabilidad sobre activos (B) antes de intereses y depreciación después de impuestos.

X4 - Operating Gross Margin: Margen de ganancia bruto operativo (Gross Profit/Net Sales).

X5 - Realized Sales Gross Margin: Margen de ganancia bruto realizado (Realized Gross Profit/Net Sales).

X6 - Operating Profit Rate: Tasa de ganancia operativa (Operating Income/Net Sales).

X7 - Pre-tax net Interest Rate: Tasa de interés neta antes de impuestos (Pre-Tax Income/Net Sales).

X8 - After-tax net Interest Rate: Tasa de interés neta después de impuestos (Net Income/Net Sales).

X9 - Non-industry income and expenditure/revenue: Relación de ingreso no operativo (Net Non-operating Income Ratio).

X10 - Continuous interest rate (after tax): Tasa de interés neta excluyendo ganancias o pérdidas de disposición (Net Income-Exclude Disposal Gain or Loss/Net Sales).

X11 - Operating Expense Rate: Tasa de gastos operativos (Operating Expenses/Net Sales).

X12 - Research and development expense rate: Tasa de gastos de investigación y desarrollo ((Research and Development Expenses)/Net Sales).

X13 - Cash flow rate: Tasa de flujo de caja (Cash Flow from Operating/Current Liabilities).

X14 - Interest-bearing debt interest rate: Tasa de interés de la deuda con carga de intereses (Interest-bearing Debt/Equity).

X15 - Tax rate (A): Tasa impositiva efectiva (Effective Tax Rate).

X16 - Net Value Per Share (B): Valor neto por acción (Book Value Per Share(B)).

X17 - Net Value Per Share (A): Valor neto por acción (Book Value Per Share(A)).

X18 - Net Value Per Share (C): Valor neto por acción (Book Value Per Share(C)).

X19 - Persistent EPS in the Last Four Seasons: EPS-Net Income (Earnings Per Share).

X20 - Cash Flow Per Share: Flujo de caja por acción.

X21 - Revenue Per Share (Yuan ¥): Ventas por acción.

X22 - Operating Profit Per Share (Yuan ¥): Ingresos operativos por acción.

X23 - Per Share Net profit before tax (Yuan ¥): Ingreso neto por acción antes de impuestos.

X24 - Realized Sales Gross Profit Growth Rate: Tasa de crecimiento de ganancia bruta realizada.

X25 - Operating Profit Growth Rate: Tasa de crecimiento de ingresos operativos.

X26 - After-tax Net Profit Growth Rate: Tasa de crecimiento de ingresos netos después de impuestos.

X27 - Regular Net Profit Growth Rate: Tasa de crecimiento de ingresos netos continuos después de impuestos.

X28 - Continuous Net Profit Growth Rate: Tasa de crecimiento de ingresos netos excluyendo ganancias o pérdidas de disposición.

X29 - Total Asset Growth Rate: Tasa de crecimiento de activos totales.

X30 - Net Value Growth Rate: Tasa de crecimiento de capital total.

X31 - Total Asset Return Growth Rate Ratio: Tasa de crecimiento de retorno sobre activos totales.

X32 - Cash Reinvestment %: Porcentaje de reinversión de efectivo.

X33 - Current Ratio: Razón corriente.

X34 - Quick Ratio: Razón rápida (Acid Test).

X35 - Interest Expense Ratio: Tasa de gastos por intereses (Interest Expenses/Total Revenue).

X36 - Total debt/Total net worth: Relación de deuda total sobre patrimonio neto (Total Liability/Equity Ratio).

X37 - Debt ratio %: Porcentaje de deuda (Liability/Total Assets).

X38 - Net worth/Assets: Relación de patrimonio sobre activos (Equity/Total Assets).

X39 - Long-term fund suitability ratio (A): Relación de idoneidad de fondos a largo plazo $((\text{Long-term Liability} + \text{Equity}) / \text{Fixed Assets})$.

X40 - Borrowing dependency: Dependencia de préstamos (Cost of Interest-bearing Debt).

X41 - Contingent liabilities/Net worth: Relación de obligaciones contingentes sobre patrimonio $(\text{Contingent Liability} / \text{Equity})$.

X42 - Operating profit/Paid-in capital: Relación de ingresos operativos sobre capital aportado $(\text{Operating Income} / \text{Capital})$.

X43 - Net profit before tax/Paid-in capital: Relación de ingreso neto antes de impuestos sobre capital aportado $(\text{Pretax Income} / \text{Capital})$.

X44 - Inventory and accounts receivable/Net value: Relación de inventario y cuentas por cobrar sobre valor neto $((\text{Inventory} + \text{Accounts Receivables}) / \text{Equity})$.

X45 - Total Asset Turnover: Rotación de activos totales.

X46 - Accounts Receivable Turnover: Rotación de cuentas por cobrar.

X47 - Average Collection Days: Días promedio de cobro $(\text{Days Receivable Outstanding})$.

X48 - Inventory Turnover Rate (times): Tasa de rotación de inventarios.

X49 - Fixed Assets Turnover Frequency: Frecuencia de rotación de activos fijos.

X50 - Net Worth Turnover Rate (times): Tasa de rotación de patrimonio neto.

X51 - Revenue per person: Ventas por empleado.

X52 - Operating profit per person: Ingresos operativos por empleado.

X53 - Allocation rate per person: Tasa de asignación por empleado $(\text{Fixed Assets Per Employee})$.

X54 - Working Capital to Total Assets: Relación de capital de trabajo sobre activos totales.

X55 - Quick Assets/Total Assets: Relación de activos líquidos sobre activos totales.

- X56 - Current Assets/Total Assets: Relación de activos corrientes sobre activos totales.
- X57 - Cash/Total Assets: Relación de efectivo sobre activos totales.
- X58 - Quick Assets/Current Liability: Relación de activos líquidos sobre pasivos corrientes.
- X59 - Cash/Current Liability: Relación de efectivo sobre pasivos corrientes.
- X60 - Current Liability to Assets: Relación de pasivos corrientes sobre activos.
- X61 - Operating Funds to Liability: Relación de fondos operativos sobre pasivos.
- X62 - Inventory/Working Capital: Relación de inventario sobre capital de trabajo.
- X63 - Inventory/Current Liability: Relación de inventario sobre pasivos corrientes.
- X64 - Current Liabilities/Liability: Relación de pasivos corrientes sobre pasivos.
- X65 - Working Capital/Equity: Relación de capital de trabajo sobre patrimonio.
- X66 - Current Liabilities/Equity: Relación de pasivos corrientes sobre patrimonio.
- X67 - Long-term Liability to Current Assets: Relación de pasivos a largo plazo sobre activos corrientes.
- X68 - Retained Earnings to Total Assets: Relación de utilidades retenidas sobre activos totales.
- X69 - Total income/Total expense: Relación de ingresos totales sobre gastos totales.
- X70 - Total expense/Assets: Relación de gastos totales sobre activos.
- X71 - Current Asset Turnover Rate: Tasa de rotación de activos corrientes.
- X72 - Quick Asset Turnover Rate: Tasa de rotación de activos líquidos.
- X73 - Working capital Turnover Rate: Tasa de rotación de capital de trabajo.
- X74 - Cash Turnover Rate: Tasa de rotación de efectivo.
- X75 - Cash Flow to Sales: Relación de flujo de caja sobre ventas.

- X76 - Fixed Assets to Assets: Relación de activos fijos sobre activos totales.
- X77 - Current Liability to Liability: Relación de pasivos corrientes sobre pasivos totales.
- X78 - Current Liability to Equity: Relación de pasivos corrientes sobre patrimonio.
- X79 - Equity to Long-term Liability: Relación de patrimonio sobre pasivos a largo plazo.
- X80 - Cash Flow to Total Assets: Relación de flujo de caja sobre activos totales.
- X81 - Cash Flow to Liability: Relación de flujo de caja sobre pasivos totales.
- X82 - CFO to Assets: Relación de flujo de efectivo operativo sobre activos.
- X83 - Cash Flow to Equity: Relación de flujo de caja sobre patrimonio.
- X84 - Current Liability to Current Assets: Relación de pasivos corrientes sobre activos corrientes.
- X85 - Liability-Assets Flag: 1 si el total de pasivos excede los activos totales, 0 en caso contrario.
- X86 - Net Income to Total Assets: Relación de ingreso neto sobre activos totales.
- X87 - Total assets to GNP price: Relación de activos totales sobre el precio del GNP.
- X88 - No-credit Interval: Intervalo sin crédito.
- X89 - Gross Profit to Sales: Relación de ganancia bruta sobre ventas.
- X90 - Net Income to Stockholder's Equity: Relación de ingreso neto sobre el patrimonio de los accionistas.
- X91 - Liability to Equity: Relación de pasivos sobre patrimonio.
- X92 - Degree of Financial Leverage (DFL): Grado de apalancamiento financiero.
- X93 - Interest Coverage Ratio: Relación de cobertura de intereses (Gastos por intereses sobre EBIT).

X94 - Net Income Flag: 1 si el ingreso neto es negativo durante los últimos dos años, 0 en caso contrario.

X95 - Equity to Liability: Relación de patrimonio sobre pasivos.

Dimensiones de la base de datos

La base de datos contiene un total de 6819 observaciones y 95 variables.

Tipo de datos

Las variables son de tipo numérico (continuas y discretas) y categórico (etiqueta de quiebra). Las variables financieras están expresadas en términos monetarios y porcentuales.

Ejemplos de registros

Tabla 1

Muestra dataset inicial

Id Empresa	ROA(C)	ROA(A)	Estado de quiebra
0001	0.370594257300249	0.424389446140427	0
0002	0.464290937454297	0.53821412996075	1
0003	0.390922829425243	0.445704317488007	0
0004	0.508360551845171	0.570922372437854	1

Nota: Elaboración propia como parte del análisis de datos financieros de empresas taiwanesas (1999-2009).

Limitaciones en la base de datos

La base de datos puede presentar sesgos en la selección de empresas, dado que se concentra en un período específico y en empresas que cotizan en la bolsa. Además, la calidad de los datos puede estar influenciada por la normativa y la situación económica de la época.

2. Fundamentos Teóricos

2.1. Bases teóricas de la investigación

Se explica detalladamente cada concepto usado para asegura un entendimiento adecuado.

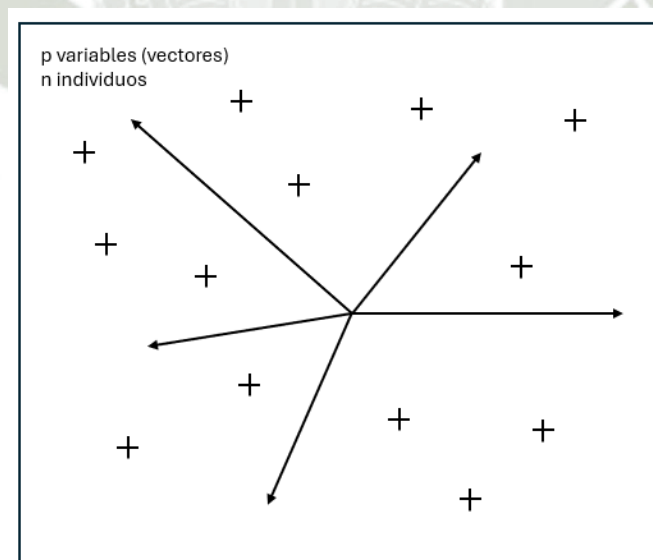
2.1.1. Biplot:

Los Biplot fueron desarrollados por Gabriel en el año 1971, para poder representar gráficamente un matriz de datos en una baja dimensión, esto métodos permiten la representación de múltiples variables simultáneamente. (Gabriel & Odoroff, 1990)

Los métodos Biplot tiene como objetivo principal representar las filas de una matriz como puntos y las columnas como como vectores, así nos permite la aproximación de cada elemento de la matriz de estudio, pero al realizar esto se presentan limitantes, el poder incluir más de dos variables, las distribuciones multivariantes que resultan son difíciles de visualizar y entender.

Figura 2

Representación Biplot



Nota: Elaboración propia.

Este enfoque se basa en descomponer la matriz de estudio e intentar reproducir los datos de esta, representación visual que integra tanto filas como columnas en dicha matriz, captura información de las variables J , que son las medidas sobre los I individuos, en esta representación se utilizan marcadores que facilitan la interpretación:

Los marcadores $a_1, \dots, a_j$ simbolizan filas de una matriz A .

Los marcadores $b_1, \dots, b_j$ simbolizan columnas de una matriz B

$$X \cong AB^T$$

Los marcadores a_i y b_i que representan filas y columnas respectivamente se representarán en un espacio cuya dimensión se encuentra en $q \leq r$, en este contexto q representa la cantidad de ejes que se han elegido mientras que r se refiere al rango de la matriz X .

Donde descomposición de la matriz X garantiza la proyección gráfica en dos dimensiones de manera aproximada, que permite la visualización de filas y columnas, esto se puede reconstruir de la siguiente manera:

$$x_{ij} = \mathbf{a}_i' \mathbf{b}_j \quad (i = 1, 2, \dots, I; \quad j = 1, 2, \dots, J)$$

Esta expresión es una forma bilineal que fue descrita por Gollob en 1986, donde se obtiene el producto interno entre las filas i en la matriz A al igual que las columnas j en la matriz B , al realizar esto nos permite visualizar las iteraciones donde las filas y columna se visualizaran en un espacio bidimensional lo que correspondería a la proyección ortogonal de los valores de a_i sobre los valores b_j y de manera viceversa

HJ-Biplot

El HJ-Biplot una visualización gráfica multivariante donde se graficará simultáneamente filas que representan a los individuos del análisis y columnas que representan a las variables esto para un único sistema de coordenadas optimizando la representación gráfica de las filas que se

identifican como individuos y para las columnas que se identifican como variables de estos (Galindo,1986), para esta formulación se parte de lo siguiente:

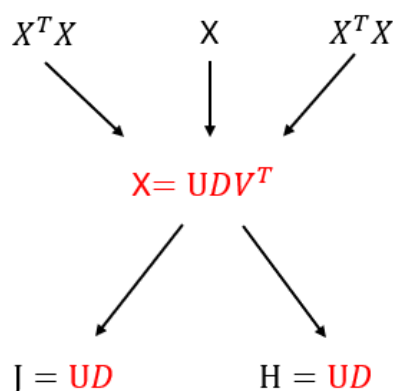
n representa la cantidad de filas (o individuos, observaciones),
 p corresponde al número de columnas (o variables medidas).

$$X_{n \times p}$$

Esta matriz tiene los datos para analizar y graficar, para la construcción del HJ-Biplot se toma la descomposición que resulta en los valores singulares (SVD) que permite desglosar los valores en componentes simples y manejables que hace que el análisis y la interpretación siguiente sean más sencillos para dicha matriz.

Figura 3

Representación gráfica de HJ - Biplot



U es una matriz cuyos vectores columna son ortonormales y vectores propios de XX^T

V es una matriz ortonormal cuyos vectores columna son vectores propios de $X^T X$

D es la matriz diagonal de valores singulares de X , que son las raíces cuadradas no negativas de los valores propios de $X^T X$

Nota: Elaboración propia.

La representación simultanea ofrece una ventaja significativa, que proporciona una calidad óptima para representar columnas y de igual forma las filas de tu conjunto de datos.

Reglas principales para la interpretación del HJ-Biplot

a) El espacio entre los marcadores de las filas, cumplen una función inversa de sus similitudes, esto se explica que mientras más cerca estén entre sí son grupos considerados similares, lo que nos permite identificar las semejanzas. (Vicente Tavera et al., 1995).

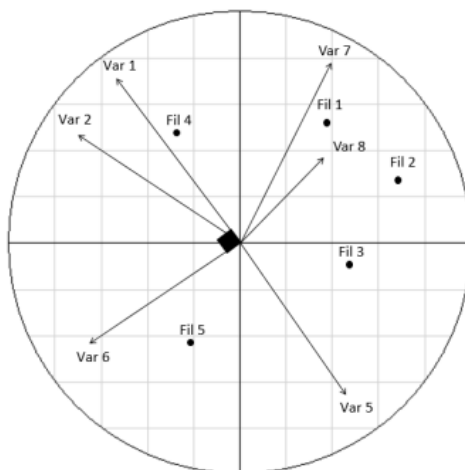
b) La longitud de las columnas representa la desviación típica de estas, esto se toma como que a mayor longitud más información que se tiene para poder discriminar entre las distintas variables.

c) Los cosenos de los ángulos entre las columnas representadas indican como correlación entre ellos, esto se puede interpretar de la siguiente manera mientras más agudos son los ángulos más alta será su correlación y esta será positiva, esto se aprecia en las variables 1 y 2, los ángulos obtusos se consideran como la correlación negativa, estos se aprecia en las variables 2 y 5 y los ángulos rectos se interpretan como la ausencia de correlación esto se aprecia entre las variables 1 y 5 (ver figura 4)

d) Las proyecciones ortogonales de las filas que son los puntos sobre las columnas, se interpretan que mientras mayor sea su proyección, tendrá un mayor peso de esa variable.

Figura 4

Interpretación del HJ - Biplot



Nota: Elaboración propia.

2.1.2. Statis:

El método STATIS (Structuration de Tableaux à Trois Indices de la Statistique) fue introducido por L'Hermier Des Plantes y Thiébaud en 1977, después ampliado y desarrollado por Lechevalier en 1990, esta técnica estadística multivariante se usa para analizar simultáneamente múltiples matrices de datos que podrían estar configuradas para mismos individuos o variables. El objetivo principal es obtener un subespacio denominado consenso, que identifica la estructura común que es ponderada entre todas las matrices de datos involucradas y se evalúa la estabilidad de esta estructura común de las diferentes matrices. (Lechevalier, 1990)

El enfoque es lograr un subespacio que es conocido como consenso para todas las matrices de datos implicadas (Castillo Elizondo & González Varela, 1998). Siendo el objetivo principal la estructura común para las matrices involucradas y determinar si la estructura encontrada permanece estable ante las diferentes matrices y sus variables.

Al llevar a cabo un análisis utilizando STATIS, la estructura de los datos es un factor determinante, lo que nos permite identificar dos tipos diferentes:

Las matrices de datos pueden presentar las mismas personas, mientras que las variables pueden variar. En este caso, se utiliza el método STATIS, que se enfoca en comparar a los individuos. (Castillo Elizondo & Gonzalez Varela, 1998)

Por otro lado, cuando las matrices tienen las mismas variables, pero los individuos son diferentes, se utiliza el método STATIS DUAL, dado que el estudio también abarca la comparación entre variables e individuos.

El método STATIS Dual tiene tres etapas a seguir:

Inter-estructura: En esta etapa realiza la evaluación comparativa integral de diferentes estructuras de las T matrices de datos a examinar, para esto se emplea la distancia que se origina

desde el producto escalar de Hilbert-Schmidt, esto supone una transformación de distintas matrices de datos en los productos matriciales cruzados usando la siguiente formula:

$$C_t = X_t^T X_t$$

La representación visual de los distintos objetos C_t , requiere la definición de una distancia por lo tanto se emplea la distancia que se origina por el producto escalar de Hilbert – Schmidt, que se define de la siguiente manera:

$$\langle C_t | C_{t'} \rangle_{HS} = tr (C_t C_{t'})$$

El resultado de estos productos escalares construye la matriz Z del orden $T \times T$, que es nombrada la matriz de correlación de vectores entre C_t

$$Z = \{Z_{tt'}\} = \{\langle C_t | C_{t'} \rangle_{HS}\}$$

El cálculo de los vectores y valores característicos (propios) de la matriz de correlaciones de vectores que será usada en la disminución de la dimensionalidad (Caballero-Juliá et al., 2017), para llevar a cabo la representación en el espacio euclídeano.

Estudio del compromiso: La meta principal de esta etapa es identificar la similitud entre las distintas matrices de datos y esta se forma a partir de las diversas submatrices extraídas de los datos a estudiar, en el que cada matriz se pondera que busca asegurar la representación precisa, obteniendo un resumen integral del conjunto de matrices analizadas. (Corrales & Rodríguez, 2014)

Se debe obtener todas las submatrices de datos para el análisis y esta será la combinación lineal de cada una de las disposiciones matriciales de los datos base del estudio y todas estas son ponderadas, dándonos una mejor representación considerando todas las submatrices iniciales.

La creación de los componentes principales (ACP), se deriva de la varianza del grupo nuevo de variables ortogonales (Corrales & Rodriguez), se obtendrán a traves de la combinación

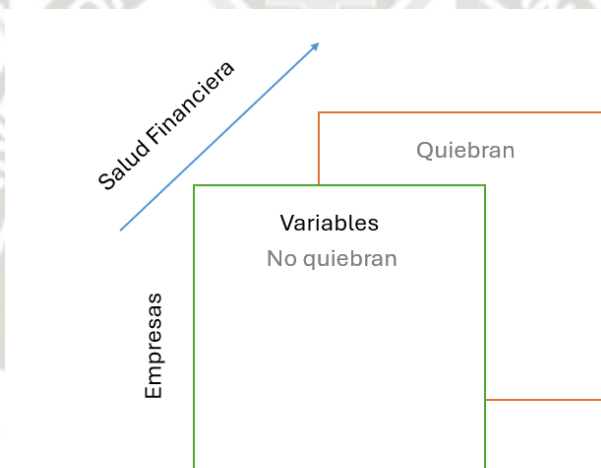
lineal de las variables iniciales, que nos permite comprender hasta cierto punto la carga factorial de cada dimensión, las cuales estarán ordenadas de acuerdo a la variabilidad representada que explicarían la matriz de conceso.

La puntuación factorial es utilizada para indentificar las diferentes variables lo que facilita la identificación de las diferencias o las similitudes entre estas.

Intra-estructura: Este paso tiene como objetivo examinar la estructura interna de cada matriz y la relación con el compromiso. (Gabriel & Odoroff, 1990)

Figura 5

Representación gráfica STATIS DUAL



$K = 2$ (Empresas que quiebran y no quiebran)

J = Mismas variables en la K -tablas

I = Diferentes individuos, empresas en K -tablas

Nota: *Elaboración propia.*

2.2. Estado del arte

1. Indicadores de sostenibilidad un análisis multivariante aplicando hj-biplot

En 2014 Ortega-Gómez, Rodríguez y Cubilla, realizaron un estudio para analizar indicadores de sostenibilidad utilizando el HJ-Biplot, que fue publicado en la Revista Colón Ciencias, Tecnología y Negocios, donde se resalta poder representar la matriz sin perder la calidad de representación, que logro la representación optima, se enfocaron en el Global Reporting Initiative (GRI), donde se elabora memorias sobre la sostenibilidad, donde se observa que Brasil presento un aumento en el número de empresas, el resultado del estudio sirvió para identificar tres agrupaciones bien definidos usando los indicadores de sostenibilidad.

2. HJ-Biplot como metodología exploratoria para el análisis multidimensional de los Objetivos de Desarrollo Sostenible (ODS) a nivel municipal en Bolivia

En 2024, Carlos Fernando Silva Viamonte presentó un artículo que analiza las metas de la agenda 2030 para el desarrollo sostenible., resaltando la necesidad de una medición objetiva y análisis de cumplimiento de estos objetivos. El autor propone el uso del HJ-Biplot como análisis multidimensional que realiza una interpretación más profunda de grandes volúmenes de datos. Utilizando información del Atlas Municipal de los ODS en Bolivia, Silva Viamonte logra identificar clústeres a nivel nacional basándose en la similitud y disimilitud de los indicadores de cumplimiento. Este enfoque busca facilitar el diseño de políticas públicas efectivas, ajustadas a las realidades locales y con un impacto social significativo. El estudio destaca la importancia de combinar técnicas de análisis avanzadas con la evaluación continua de los ODS para optimizar las políticas en lo largo de tiempo. (Silva Viamonte, C. F., 2024).

3. Modelos de Predictibilidad de Quiebras e Insolvencia Basados en Análisis de Estados Financieros.

Campanaro et al (2016) presentaron un trabajo que se centra en el uso de técnicas avanzadas inteligentes para análisis financiero. Donde se estudia los datos ofrecidos por la comisión de bolsa y valores (SEC) de los EE. UU., que tiene información desde 2005 recopila reportes financieros codificados en XBRL, permitiendo su explotación mediante sistemas computacionales. El artículo analiza el modelo de predictibilidad de quiebras de Altman, evaluando su eficacia con datos actuales y discutiendo el enfoque tecnológico necesario para gestionar grandes volúmenes de información. Además, se considera la factibilidad de utilizar inteligencia artificial para el descubrimiento de conocimiento en este contexto. Con el objetivo de corroborar modelos existentes y desarrollar Modelos de análisis avanzado para los estados financieros. (Campanaro, R., Díaz, D., Marchese, A., Sepliansky, P., Viola, M., Riveros, M., Carrizo, J., Ferraro, F., Díaz Toledo, A., & Gardenal, S., 2016).

4. Modelo de Predicción de Quiebra Z2 de Altman en Empresas del Sector Inmobiliario de Pichincha.

Molina Panchi, Molina Panchi y Flores Tapia (2022) presentaron un estudio que utiliza la tecnica Z2 de Altman para examinar la falta de solvencia financiera del rubro inmobiliario en la provincia de Pichincha, Ecuador. La investigación, de enfoque cuantitativo, utiliza análisis multivariante y pruebas estadísticas para establecer relaciones entre las variables y impactan en éxito o quiebre empresarial. El diseño longitudinal de la investigación abarca datos de los años 2016 a 2020, proporcionados por la Superintendencia de Compañías, Valores y Seguros de Ecuador. El análisis final indica que el grupo de empresas con salud

financiera con un nivel bajo de endeudamiento, ingresos, capital, liquidez y rentabilidad mayor al costo de capital. Por otro lado, empresas con precaución enfrentan dificultades para cubrir sus deudas donde muestran rentabilidad marginal y liquidez negativa, capital negativo y disminución de ingresos y bienes. (Molina Panchi, P. A., Molina Panchi, D. F., & Flores Tapia, C. E., 2022).

5. Análisis Discriminante Aplicado a Modelos de Predicción de Quiebra

Aldazábal Contreras y Napán Vera (2014) analizan la importancia de la gestión del riesgo financiero en un entorno empresarial cada vez más interconectado. El artículo destaca que una empresa insolvente cuando no dispone de efectivo que cubra sus obligaciones inmediatas. Para evaluar el riesgo de insolvencia, se propone el uso del Z de Altman, que usa el Análisis Discriminante Múltiple, que emplea indicadores financieros para evaluar la situación financiera y prever la quiebra empresarial. Los autores presentan los resultados de Altman y revisan el modelo original, subrayando su utilidad donde se mide el riesgo financiero y alerta de insolvencia. (Aldazábal Contreras, J. C., & Napán Vera, A. F., 2014).

6. Predicción de la Quiebra Empresarial en el Sector Agroindustrial de Machala

Rivadeneira, Saltos, Rivera y Carpio (2022) proponen un modelo para prever la quiebra de empresas en el sector agroindustrial en Machala, Ecuador. Utilizando una población de más de 300 empresas y 13 métricas de la salud financiera de las empresas mediante un análisis de correlación, se aplica un muestreo aleatorio simple para equilibrar los datos de empresas activas e inactivas. El modelo utiliza árboles de decisión y muestra una precisión del 78.57%, lo que resalta sobre otros modelos. Se implementa en el ejercicio fiscal de 2018, generando alertas tempranas para 190 empresas, donde se sugiere la revisión de la situación financiera para prever quiebre empresarial.

2.3. Bases Teóricas de la Investigación

El entendimiento de la quiebre empresarial es una combinación de teorías financieras, estadísticas y multivariantes, explican como diferentes factores interactúan e influyen en la estabilidad financiera de una empresa, para esto se detallan los conceptos usados:

1. Quiebre Empresarial:

El termino quiebre empresarial es cuando una organización que no puede cumplir con sus obligaciones de pago a falta de liquidez o solvencia, desde una perspectiva teórica, la quiebra empresarial ocurre como consecuencia de un deterioro en las finanzas de la empresa, ocasionado por distinto factores que pueden ir desde la falta de rentabilidad hasta una mala gestión, entorno económico, fluctuaciones en el mercado, competencia, estos son factores tanto como internos o externos.

La teoría clásica de la insolvencia financiera que lleva a la quiebra empresarial es el análisis de factores de ratios financieros, como la liquidez, rentabilidad y eficiencia, que son métricas fundamentales para analizar la salud financiera de una empresa.

2. Análisis Multivariante:

El termino análisis multivariante realiza un amplio estudio de múltiples variables para diferentes casos de estudio que permite analizar como se relacionan entre ellas y permite capturar las interacciones principales, esto diferenciado de los análisis tradicionales que se limitan a análisis univariados, para el contexto de quiebre empresarial proporciona herramientas estadísticas para poder analizar diferentes indicadores financieros.

HJ-Biplot y el **STATIS Dual** permite la visualización para poder analizar las interacciones

complejas entre las variables identificando patrones que no serían visibles con análisis simples.

- **HJ-Biplot:** Es una tecnica multivariante se fundamenta en el análisis de componentes principales (ACP), que permite la representación gráfica de las variables y las observaciones en un plano cartesiano, una vez representado se identifican las interrelaciones entre las variables y como influyen en el resultado.
- **STATIS Dual:** Es una tecnica multivariante que permite analizar múltiples tablas de entrada y analizarlas en conjunto, para este caso de estudio nos ayudara para estudiar las empresas que quiebran y no quiebran.

3. Información Financiera:

Este término es usado para analizar los valores empresariales y dar un panorama claro de la situación económica empresarial, para esto se usa términos como el balance general, estados de resultado y flujo de caja es información clave para hacer una evaluación correcta, este proyecto de tesis asumirá como base estas señales para el uso de las técnicas multivariantes y expresarlas de manera precisa y efectiva.

4. Predicción y Modelos de Riesgo:

Los modelos de riesgo tradicionales se basan en Altman Z-score, que sirve para evaluar los ratios indicando los niveles de riesgos de la empresa analizada a diferencia de los modelos analíticos multivariantes analizan múltiples variables y simultaneo identificando complejidades financieras y proporcionar una visión más completa que nos ayudará a entender la situación financiera de las empresas.

3. Marco teórico

3.1. *Alcances y Limitaciones*

Alcances

El proyecto de tesis tiene como objetivo principal desarrollar un modelo de análisis multivariante que permita analizar los quiebres empresariales y encontrar patrones ocultos. El modelo por usar se centra en usar las técnicas HJ-Biplot y STATIS Dual, identificando las variables financieras claves y como interaccionan la quiebra empresarial, lo alcances de esta investigación incluyen:

6. Análisis de Datos Financieros:

El proyecto de tesis se realizará con los datos financieros de empresas taiwanesas de los años 1999 y 2009, extraídos del Taiwan Economic Journal, el conjunto de datos sirve para analizar el mercado financiero a lo largo de un periodo de tiempo y como influyen.

7. Aplicación de Técnicas Multivariantes Avanzadas:

El alcance metodológico se centra en el uso del HJ-Biplot y STATIS Dual para analizar múltiples variables que permite indentificar patrones y relaciones entre ellas que no son evidentes con técnicas tradicionales.

8. Validación del Modelo:

Se validará el modelo desarrollado utilizando técnicas de validación cruzada, comparando la presión de modelo multivariante versus modelos tradicionales como el Z-score de Altman, permitiendo evaluar la eficacia del enfoque propuesto.

Limitaciones

El enfoque multivariante a pesar de tener una complejidad alta y acoplable presenta limitante inherentes a la investigación

9. Disponibilidad y Calidad de los Datos:

Este proyecto de tesis tiene como matriz de datos los datos históricos del año 1999 al 2009 de las empresas Taiwanesas, aunque proviene de una fuente confiable como el Taiwan Economic Journal puede tener limitaciones respecto a la calidad y la completitud de los datos, dado que las empresas que reportaron pueden tener inconsistencia en sus indicadores, esta limitación afecta la robustez del modelo, siendo que variables críticas puede estar ausentes o incompletas.

10. Aplicabilidad a Otros Contextos:

El modelo analítico que se desarrollara estará específicamente ajustado a la base de datos de las empresas Taiwanesas dentro del contexto económico de los años 1999 al 2009, siendo posible que las técnicas son generalizables, la aplicabilidad de este modelo a otros sectores económicos o regiones puede requerir ajustes adicionales, siendo las características financiera de cada país y la normativa activa en ellos puede influenciar en la efectividad del modelo, que limitaría su uso en otros entornos si una calibración o validación adicional.

11. Dependencia de Factores Financieros:

El modelo propuesto se basa en las variables financieras obtenidas, esto puede dejar de lado otros factores cualitativos o externos que también podrían influir en la quiebra empresarial, un ejemplo de esto puede ser decisiones estratégicas, cambios regulatorios, crisis económicas globales o eventos catastróficos, esta limitación reflejaría la necesidad de integrar variables externas en futuras investigaciones.

12. Complejidad Computacional:

Las técnicas multivariantes avanzadas como el HJ-Biplot y Statis Dual, que tiene como principal capacidad el análisis de patrones ocultos y complejos, necesita un proceso

computacional significativo y una difícil implementación en todas las plataformas de análisis, además de la correcta interpretación de los resultados, aunque son modelos de interpretación clara comparado con otros modelos multivariantes, requiere cierto nivel de experiencia en el análisis multivariante.

13. Interpretabilidad en Contextos Empresariales:

Aunque las técnicas multivariantes aplicadas permiten mejores visualizaciones de las relaciones entre variables, la correcta interpretación puede presentar desafíos para profesionales financieros sin una formación en análisis multivariante, esto puede ser un limitante generalizada del modelo en el ámbito corporativo sin una formación adecuada o una simplificación de resultados.

3.2. *Tipo y nivel de investigación*

Tipo de Investigación

Esta investigación es de tipo **cuantitativa y predictiva**, ya que se centra en el análisis de datos financieros para desarrollar un modelo multivariante que permita predecir la quiebra empresarial. Su naturaleza cuantitativa radica en la recolección, análisis y modelado de datos numéricos provenientes de variables financieras clave como rentabilidad, liquidez y apalancamiento financiero. Este enfoque permite identificar con precisión las relaciones entre las variables y estimar la probabilidad de quiebra, logrando anticipar posibles insolvencias mediante la detección de patrones relevantes. El objetivo final es crear un modelo que no solo analice los datos, sino que también permita identificar tendencias y comportamientos asociados con la estabilidad financiera de las empresas.

También el proyecto de tesis es aplicada siendo su objetivo principal es desarrollar un modelo práctico que pueda ser implementado en contextos reales. Este modelo está diseñado para

ser útil tanto para gestores de riesgos empresariales como para organismos reguladores, al proporcionar una herramienta efectiva para el análisis y predicción de quiebras.

El enfoque aplicado se justifica por su orientación hacia la resolución de problemas específicos en el mundo empresarial y financiero. Más allá del marco teórico, este proyecto busca:

- Proveer Soluciones Prácticas: Ofrecer un modelo que pueda integrarse en sistemas de monitoreo financiero para anticipar riesgos de insolvencia.
- Fortalecer la Gestión Financiera: Ayudar a empresas y reguladores a tomar decisiones informadas basadas en patrones identificados en los datos financieros.
- Impacto en Políticas y Estrategias: Contribuir al diseño de políticas regulatorias más robustas y estrategias empresariales que reduzcan la probabilidad de quiebra.

El modelo desarrollado no solo será una herramienta académica, sino que también tendrá un impacto directo en la gestión y análisis financiero, mejorando la capacidad de las organizaciones para prever riesgos y garantizar la sostenibilidad en un entorno económico cambiante.

Nivel de Investigación

El nivel investigación es explicativo, siendo el propósito principal explicar las interrelaciones entre las variables financieras y como influyen en la quiebra empresarial, usando técnicas multivariantes como el HJ-Biplot y STATIS Dual, esta no se limita a indentificar los factores clave, si no que tambien busca comprender como interactúan entre ellas, y la influencia en la estabilidad financiera.

El nivel explicativo implica que este proyecto de tesis no será una simple descripción de los datos proporcionando un análisis profundo de las causas que pueden determinar la quiebra

empresarial, esto se busca lograr con la identificación de patrones ocultos en los datos estudiados y explicar la relaciones entre las variables financieras.

3.3. Población y Muestra o Universo

Se tuvo acceso a un dataset de Taiwan Economic Journal (TEJ) que consta de 6819 registros, siendo una fuente confiable que recopila la información financiera detallada de empresas en diversas industrias.

Universo:

Empresas de diversas industrias que cotizan en la bolsa Taiwanesa en los años 1999 y 2009.

Muestra:

Se utilizo el muestreo no probabilístico intencional o dirigido. Se seleccionarán empresas que durante el período de 1999 a 2009, que hayan presentado datos completos en el Taiwan Economic Journal, esto incluye tantas empresas que han quebrado o no, siendo el principal factor de selección que dispongan de datos consistentes para el estudio donde el número total de empresas seleccionados es de 6819.

Criterios de Inclusión:

1. Empresas que cotizaban en la bolsa de valores de Taiwán entre los años de 1999 y 2009.
2. Empresas con datos financieros completos en las variables seleccionadas para el análisis.
3. Empresas que hayan experimentado quiebra, junto con aquellas que se mantuvieron estables durante el período de estudio, para realizar comparaciones entre ambas.

Criterios de Exclusión:

1. Empresas que no cuenten con registros completos de los datos financieros requeridos para el análisis multivariante.

2. Empresas que no cotizaban en los rangos de establecidos en la Bolsa de Valores de Taiwán.

3.4. *Métodos, Técnicas e Instrumentos de Recolección de Datos*

Métodos de Recolección de Datos:

La recolección de datos se realizará mediante la obtención de información financiera histórica de los años 1999 y 2009, los datos provienen de una fuente confiable el Taiwan Economic Journal (TEJ) el cual contiene los registros financieros detallados que contiene los balances generales, ratios financieros y otros indicadores relevantes para este análisis.

El método de obtención de los datos se centrará en la identificación de las empresas que quebraron y aquellas que no durante el periodo de estudio.

Técnicas de Recolección de Datos:

Para este proyecto de investigación se utilizarán técnicas de recolección de datos basada en la extracción de información de la base financiera, usando herramientas de software que permitan acceder de forma correcta a la información proporcionada por el Taiwan Economic Journal tales como:

- **Extracción y Transformación de Datos:** Se utilizará lenguajes de programación que sirve para un procesamiento de datos como Python o R, para extraer las variables financieras necesarias como pueden ser la rentabilidad, liquidez, apalancamiento entre otros indicadores clave, para luego estructurar los datos de tal forma que sirvan para el análisis multivariante, garantizando que las variables estén limpias y preparadas para ser utilizadas en los modelos HJ-Biplot y STATIS Dual.
- **Validación de la Calidad de los Datos:** Se empleará técnicas de limpieza de datos y validación de datos incompletos, para evitar sesgos en el análisis, en esta etapa se incluirá

la detección de valores atípicos, imputación de valores faltantes, verificación y la consistencia de las fuentes de información disponibles en el dataset.

- **Técnicas de Análisis Multivariante:** Una vez se tenga la base limpia y estructurada se aplicarán las técnicas de análisis multivariante HJ-Biplot y STATIS Dual para analizar las relaciones entre las variables financieras. Estas técnicas se implementarán mediante paquetes de análisis disponibles tanto en Python como en R, con un enfoque en la interpretación gráfica de los resultados y la integración de los datos.

3.5. *Plan de análisis estadístico de los datos*

El plan de análisis estadístico de los datos se basa en el uso de técnica multivariantes avanzadas para encontrar y estudiar patrones significativos de los datos financieros de empresas Taiwanesas que nos ayude a interpretar la quiebra empresarial, el análisis sigue el siguiente enfoque y etapas:

1. Exploración Preliminar de los Datos

- **Análisis Descriptivo:**

Se calculará estadísticas descriptivas básica como medias, medianas, desviación, percentiles, etc, estas aplicadas a las variables financieras, siendo las primeras la rentabilidad, eficacia operativa y liquidez, con esto se busca tener una visión general del conjunto de datos, para detectar irregularidades, valores atípicos o distribuciones sesgadas.

- **Análisis de Correlación:**

Se generará matrices de correlación donde se podrá observar las relaciones lineales entre las variables financieras, esto ayuda a indentificar las relaciones potenciales importantes que pueden contribuir la quiebra empresarial.

2. Preparación de los Datos

- **Normalización y Estandarización:**

Revisar que las variables estén en la misma escala, de no ser el caso se deberá normalizar o estandarizar para garantizar que todas las variables contribuyan equitativamente al análisis multivariante, esto es sumamente importante para evitar que alguna variable con diferente escala predomine en el análisis.

- **Manejo de Datos Faltantes y Outliers:**

Los valores faltantes se manejarán con una imputación estadística que puede ser la media, mediana de ser el caso la eliminación si la proporción de valores faltantes es muy alta, los valores atípicos se identificarán y analizarán individualmente para determinar si deben ajustarse o deben eliminarse.

3. Aplicación del Análisis HJ-Biplot

- **Reducción de Dimensionalidad:**

Para aplicar el HJ-Biplot se usará en reducir la dimensionalidad de los datos financieros, para esto se extraerán las componentes principales que explican la mayor cantidad de varianza en las variables, con esto se identificará que variables son más relevantes en el análisis de quiebra.

- **Visualización de las Interrelaciones:**

Generar un gráfico bidimensional que corresponde al HJ-Biplot el cual logra representar las variables financieras y las empresas en el mismo espacio, con este gráfico se busca interpretar como interactúan las diferentes variables y como están relacionadas, identificando cuáles son las más fuertes.

4. Análisis con STATIS Dual

- **Análisis de Datos de múltiples tablas:**

STATIS Dual se empleará para integrar dos tipos de tablas de las empresas que quiebran y las empresas que no quiebran, que permitirá analizar cómo las variables cambian, esto es importante para entender la situación financiera de las empresas y poder compararlas.

5. Interpretación de los Resultados

- **Identificación de Variables Clave:**

Después del análisis multivariante, se identificarán qué variables financieras son más influyente en la quiebra empresarial,

- **Conclusiones y Recomendaciones:**

Con los resultados del análisis se darán las conclusiones sobre los factores críticos que influyen en la quiebra empresarial y se propondrán recomendaciones para mejorar las tomas de decisiones en la gestión de riesgos financieros.

4. Plan de trabajo

El plan de trabajo propuesto para este trabajo de investigación tiene la siguiente secuencia empezando la recolección de datos, análisis estadístico, desarrollo de modelo predictivo, finalmente la validación y presentación de resultados, este plan de trabajo se organizará en un cronograma de 8 meses, que se detalla a continuación:

Fase 1: Definición y Preparación (Meses 1)

1. Definición de Objetivos y Alcance de la Investigación

- Redacción final de los objetivos generales y específicos de la presente investigación.

- Revisión del marco teórico y antecedentes para consolidar las bases teóricas del trabajo.

2. Selección y Acceso a la Fuente de Datos

- Acceso al dataset de Taiwan Economic Journal (TEJ) para recolectar los datos financieros históricos de empresas taiwanesas entre 1999 y 2009.
- Verificación de la disponibilidad de datos importantes.

3. Revisión Bibliográfica y Estado del Arte

- Ampliación y revisión exhaustiva de la literatura relacionada con la predicción de quiebras empresariales, análisis multivariante y técnicas estadísticas avanzadas como HJ-Biplot y STATIS Dual.

Fase 2: Recolección de los Datos (Meses 1)

1. Extracción de Datos Financieros

- Recopilación de los datos financieros de empresas cotizadas en Taiwán desde 1999 hasta 2009.
- Selección de las variables financieras clave, tales como ratios de liquidez, apalancamiento, rentabilidad y flujos de caja.

Fase 3: Análisis Multivariante y Desarrollo del Modelo (Meses 3 y 4)

1. Análisis Descriptivo y Exploratorio

- Realización de análisis descriptivos para obtener una visión preliminar del comportamiento de las variables.
- Análisis de correlación y visualización inicial de los datos mediante gráficos de dispersión y diagramas de caja.

2. Aplicación del HJ-Biplot

- Implementación del análisis **HJ-Biplot** para la reducción de dimensionalidad.
- Visualización gráfica de las relaciones entre variables y empresas.
- Interpretación de los resultados y análisis de los patrones encontrados.

3. **Análisis Temporal con STATIS Dual**

- Aplicación del **STATIS Dual** para integrar las tablas de datos de diferentes años.
- Análisis de las dos matrices de datos que identificando las variables que mas influyen en la quiebra empresarial.
- Identificación de tendencias clave y puntos de inflexión.

Fase 4: Redacción de Resultados y Conclusiones (Meses 5-6)

1. **Redacción del Informe Final**

- Elaboración de los capítulos finales del informe, incluyendo el análisis de resultados, conclusiones y recomendaciones.
- Explicación detallada de las variables financieras más relevantes y los patrones identificados.

2. **Conclusiones y Recomendaciones**

- Redacción de conclusiones sobre la eficacia del modelo predictivo y su aplicabilidad en la predicción de quiebras empresariales.
- Propuestas de recomendaciones para la gestión de riesgos financieros basadas en los resultados del análisis.

3. **Presentación de los Resultados**

- Preparación de una presentación formal para la defensa de la tesis, con un resumen de los hallazgos clave y las implicaciones prácticas del modelo desarrollado.

4. Cronograma General

Fase	Mes 1	Mes 2	Mes 3	Mes 4	Mes 5	Mes 6	Mes 7	Mes 8
Definición y Preparación	X							
Recolección de datos	X							
Análisis Multivariante		X	X	X				
Redacción de Resultados					X	X		

Fuente: elaboración propia

REFERENCIAS

Altman, E. I. (1968). Financial ratios, discriminant analysis and the prediction of corporate bankruptcy. *The Journal of Finance*, 23(4), 589–609. <https://doi.org/10.1111/j.1540-6261.1968.tb00843.x>

Altman, E. I., & Hotchkiss, E. (2006). *Corporate financial distress and bankruptcy: Predict and avoid bankruptcy, analyze and invest in distressed debt* (3rd ed.). Wiley.

Beaver, W. H. (1966). Financial ratios as predictors of failure. *Journal of Accounting Research*, 4, 71–111. <https://doi.org/10.2307/2490171>

Chava, S., & Jarrow, R. A. (2004). Bankruptcy prediction with industry effects. *Review of Finance*, 8(4), 537–569. <https://doi.org/10.1093/rof/8.4.537>

Moreno, E., & Bravo, F. (2018). Análisis de la probabilidad de quiebra de las empresas cotizadas españolas. *Revista de Estudios Empresariales. Segunda Época*, 2, 57–72. <https://doi.org/10.17561/ree.v2018n2.3>

Ohlson, J. A. (1980). Financial ratios and the probabilistic prediction of bankruptcy. *Journal of Accounting Research*, 18(1), 109–131. <https://doi.org/10.2307/2490395>

Tascón, M. T., & Castaño, F. J. (2012). Variables y modelos para la identificación y predicción del fracaso empresarial: Revisión de la investigación empírica reciente. *Revista de Contabilidad - Spanish Accounting Review*, 15(1), 7–58.

Tsai, C.-F. (2013). Combining cluster analysis with classifier ensembles to predict financial distress. *Information Fusion*, 24, 31–38. <https://doi.org/10.1016/j.inffus.2012.06.003>

Altman, E. I. (1968). Financial ratios, discriminant analysis and the prediction of corporate bankruptcy. *The Journal of Finance*, 23(4), 589–609. <https://doi.org/10.2307/2978933>

Caballero-Juliá, D., Galindo Villardón, P., & García, M. C. (2017). JK-Meta-Biplot y STATIS Dual como herramientas de análisis de tablas textuales múltiples. *RISTI - Revista Ibérica de Sistemas e Tecnologías de Informação*, 2017(25), 18–33. <https://doi.org/10.17013/risti.25.18-33>

Castillo Elizondo, W., & Gonzalez Varela, J. (1998). Statis dual: Software y análisis de datos reales. *Revista de Matemática: Teoría y Aplicaciones*, 2, 149–162.

Corrales, D., & Rodríguez, O. (2014). *INTERSTATIS: El método STATIS para datos de tipo intervalo*.

Díaz-Faes, A. A., González-Albo, B., Galindo, M. P., & Bordons, M. (2013). HJ-Biplot como herramienta de inspección de matrices de datos bibliométricos. *Revista Española de Documentación Científica*, 36(1), e001. <https://doi.org/10.3989/redc.2013.1.988>

Fernández Aguirre, K., Modroño Herrán, J. I., & Landaluce Calvo, M. I. (2004). ACM y Statis dual ponderado: Dos técnicas complementarias para analizar una visión de la cultura de la universidad. *Estadística Española*, 46(156), 205–228.

Hair, J. F., Black, W. C., Babin, B. J., & Anderson, R. E. (2019). *Multivariate data analysis* (8th ed.). Cengage Learning.

Hosmer, D. W., & Lemeshow, S. (2000). *Applied logistic regression* (2nd ed.). John Wiley & Sons. <https://doi.org/10.1002/0471722146>

Reichardt, C. S., & Gollob, H. F. (1986). Satisfying the constraints of causal modeling. *Evolutionary Ecology*, 1(4), 349–366. <https://doi.org/10.1002/ev.1436>

Zmijewski, M. E. (1984). Methodological issues related to the estimation of financial distress prediction models. *Journal of Accounting Research*, 22, 59–82. <https://doi.org/10.2307/2490859>

Gabriel, K. R. (1971). The biplot graphic display of matrices with application to principal component analysis. *Biometrika*, 58(3), 453–467.

Galindo, P. (1986). Una alternativa de representación simultánea: HJ-Biplot. *Quetiio*, 10, 13–23.

James, G., Witten, D., Hastie, T., & Tibshirani, R. (2013). *An introduction to statistical learning with applications in R*. Springer. <https://doi.org/10.1007/978-1-4614-7138-7>

Johnson, R. A., & Wichern, D. W. (2007). *Applied multivariate statistical analysis* (6th ed.). Pearson Prentice Hall.

Leeuw, J., & Mair, P. (2009). Gifi methods for optimal scaling in R: The package homals. *Journal of Statistical Software*, 31(4), 1–19. <https://doi.org/10.18637/jss.v031.i04>

Lechevalier, F. (1990). L'analyse en composantes conjointes d'une famille de triplets indexés. *Statistique et Analyse des Données*, 15(2), 35–49. http://www.numdam.org/item?id=SAD_1990__15_2_35_0

Martínez, W., & Martínez, A. (2008). *Computational statistics handbook with MATLAB* (2nd ed.). Chapman & Hall/CRC.

Ooghe, H., & De Prijcker, S. (2008). Failure processes and causes of company bankruptcy: A typology. *Management Decision*, 46(2), 223–242. <https://doi.org/10.1108/00251740810854131>

Pindado, J., Rodrigues, L., & de la Torre, C. (2008). Estimating financial distress likelihood. *Journal of Business Research*, 61(9), 995–1003. <https://doi.org/10.1016/j.jbusres.2007.10.006>

Taiwan Economic Journal (TEJ). (1999–2009). *TEJ database*.

Tenenhaus, M. (1998). *La régression PLS: Théorie et pratique* [Partial least squares regression: Theory and practice]. Éditions Technip.

Tsai, C. F. (2009). Feature selection in bankruptcy prediction. *Knowledge-Based Systems*, 22(2), 120–127. <https://doi.org/10.1016/j.knosys.2008.08.002>

Aldazábal Contreras, J. C., & Napán Vera, A. F. (2014). Análisis discriminante aplicado a modelos de predicción de quiebra. *Quipukamayoc*, 22(42), 53–59.

Campanaro, R., Díaz, D., Marchese, A., Sepiarsky, P., Viola, M., Riveros, M., Carrizo, J., Ferraro, F., Díaz Toledo, A., & Gardenal, S. (2016). Modelos de predictibilidad de quiebras e insolvencia basados en análisis de estados financieros. En *Vigesimoprimeras Jornadas "Investigaciones en la Facultad"* de Ciencias Económicas y Estadística.

Molina Panchi, P. A., Molina Panchi, D. F., & Flores Tapia, C. E. (2022). Modelo de predicción de quiebra Z2 de Altman de análisis multivariable en empresas del sector inmobiliario de la

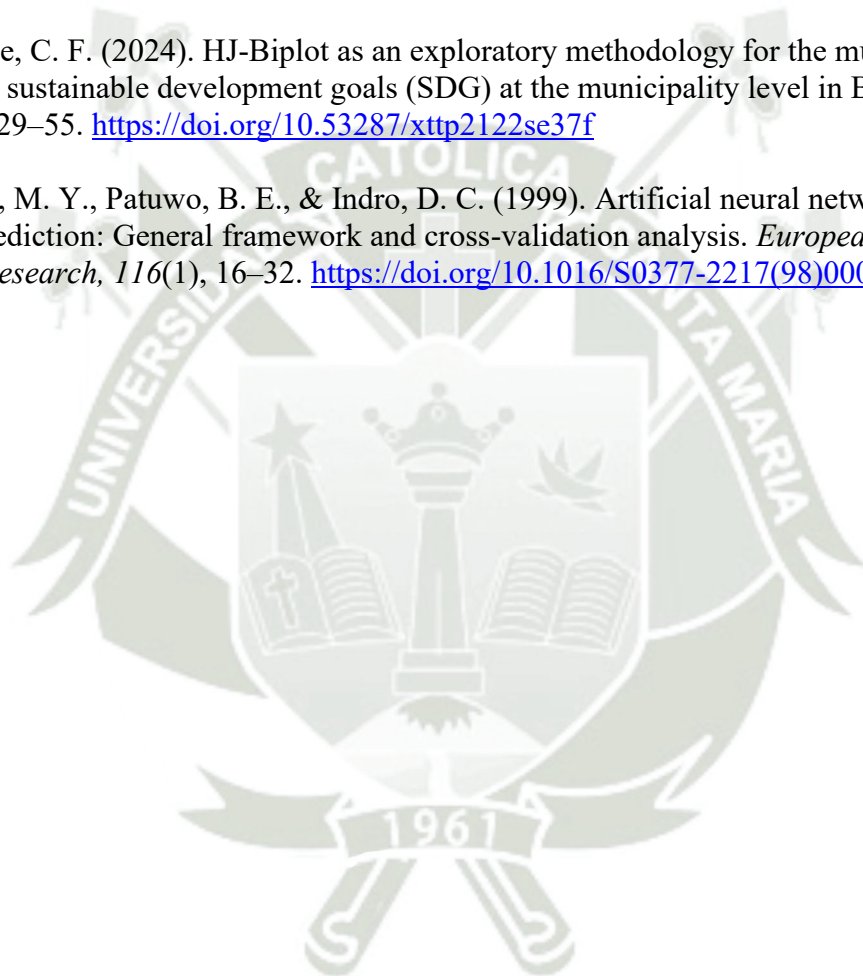
provincia de Pichincha. *Revista Científica Ecociencia*, 9(2), 53–76.
<https://doi.org/10.21855/ecociencia.92.643>

Ortega-Gómez, E., Rodríguez, C. C., & Cubilla, M. I. (2014). Indicadores de sostenibilidad: Un análisis multivariante aplicando HJ-Biplot. *Revista Colón Ciencias, Tecnología y Negocios*, 1(2), 30–46.

Rivadeneira, J., Saltos, R., Rivera, M., & Carpio, R. (2022). Predicción de la quiebra empresarial en el sector agroindustrial de la ciudad de Machala. *Avances en Ciencias e Ingenierías*, 14(2).
<https://doi.org/10.18272/aci.v14i2.2695>

Silva Viamonte, C. F. (2024). HJ-Biplot as an exploratory methodology for the multidimensional analysis of the sustainable development goals (SDG) at the municipality level in Bolivia. *Revista Varianza*, 23, 29–55. <https://doi.org/10.53287/xttp2122se37f>

Zhang, G., Hu, M. Y., Patuwo, B. E., & Indro, D. C. (1999). Artificial neural networks in bankruptcy prediction: General framework and cross-validation analysis. *European Journal of Operational Research*, 116(1), 16–32. [https://doi.org/10.1016/S0377-2217\(98\)00051-4](https://doi.org/10.1016/S0377-2217(98)00051-4)



ANEXO C: Variables Utilizadas en el Análisis

Descripción General

A continuación, se presenta la lista de variables financieras utilizadas en el análisis, extraídas del conjunto de datos Taiwanese Bankruptcy Prediction (1999-2009). Estas variables fueron seleccionadas por su relevancia para identificar patrones asociados a la quiebra empresarial.

Lista de Variables

Y - Bankrupt?: Clase de etiqueta que indica si la empresa ha quebrado (1) o no (0).

X1 - ROA(C): Rentabilidad sobre activos (C) antes de intereses y depreciación.

X2 - ROA(A): Rentabilidad sobre activos (A) antes de intereses y después de impuestos.

X3 - ROA(B): Rentabilidad sobre activos (B) antes de intereses y depreciación después de impuestos.

X4 - Operating Gross Margin: Margen de ganancia bruto operativo (Gross Profit/Net Sales).

X5 - Realized Sales Gross Margin: Margen de ganancia bruto realizado (Realized Gross Profit/Net Sales).

X6 - Operating Profit Rate: Tasa de ganancia operativa (Operating Income/Net Sales).

X7 - Pre-tax net Interest Rate: Tasa de interés neta antes de impuestos (Pre-Tax Income/Net Sales).

X8 - After-tax net Interest Rate: Tasa de interés neta después de impuestos (Net Income/Net Sales).

X9 - Non-industry income and expenditure/revenue: Relación de ingreso no operativo (Net Non-operating Income Ratio).

X10 - Continuous interest rate (after tax): Tasa de interés neta excluyendo ganancias o pérdidas de disposición (Net Income-Exclude Disposal Gain or Loss/Net Sales).

X11 - Operating Expense Rate: Tasa de gastos operativos (Operating Expenses/Net Sales).

X12 - Research and development expense rate: Tasa de gastos de investigación y desarrollo ((Research and Development Expenses)/Net Sales).

X13 - Cash flow rate: Tasa de flujo de caja (Cash Flow from Operating/Current Liabilities).

X14 - Interest-bearing debt interest rate: Tasa de interés de la deuda con carga de intereses (Interest-bearing Debt/Equity).

X15 - Tax rate (A): Tasa impositiva efectiva (Effective Tax Rate).

X16 - Net Value Per Share (B): Valor neto por acción (Book Value Per Share(B)).

X17 - Net Value Per Share (A): Valor neto por acción (Book Value Per Share(A)).

X18 - Net Value Per Share (C): Valor neto por acción (Book Value Per Share(C)).

X19 - Persistent EPS in the Last Four Seasons: EPS-Net Income (Earnings Per Share).

X20 - Cash Flow Per Share: Flujo de caja por acción.

X21 - Revenue Per Share (Yuan ¥): Ventas por acción.

X22 - Operating Profit Per Share (Yuan ¥): Ingresos operativos por acción.

X23 - Per Share Net profit before tax (Yuan ¥): Ingreso neto por acción antes de impuestos.

X24 - Realized Sales Gross Profit Growth Rate: Tasa de crecimiento de ganancia bruta realizada.

X25 - Operating Profit Growth Rate: Tasa de crecimiento de ingresos operativos.

X26 - After-tax Net Profit Growth Rate: Tasa de crecimiento de ingresos netos después de impuestos.

X27 - Regular Net Profit Growth Rate: Tasa de crecimiento de ingresos netos continuos después de impuestos.

X28 - Continuous Net Profit Growth Rate: Tasa de crecimiento de ingresos netos excluyendo ganancias o pérdidas de disposición.

X29 - Total Asset Growth Rate: Tasa de crecimiento de activos totales.

X30 - Net Value Growth Rate: Tasa de crecimiento de capital total.

X31 - Total Asset Return Growth Rate Ratio: Tasa de crecimiento de retorno sobre activos totales.

X32 - Cash Reinvestment %: Porcentaje de reinversión de efectivo.

X33 - Current Ratio: Razón corriente.

X34 - Quick Ratio: Razón rápida (Acid Test).

X35 - Interest Expense Ratio: Tasa de gastos por intereses (Interest Expenses/Total Revenue).

X36 - Total debt/Total net worth: Relación de deuda total sobre patrimonio neto (Total Liability/Equity Ratio).

X37 - Debt ratio %: Porcentaje de deuda (Liability/Total Assets).

X38 - Net worth/Assets: Relación de patrimonio sobre activos (Equity/Total Assets).

X39 - Long-term fund suitability ratio (A): Relación de idoneidad de fondos a largo plazo ((Long-term Liability+Equity)/Fixed Assets).

X40 - Borrowing dependency: Dependencia de préstamos (Cost of Interest-bearing Debt).

X41 - Contingent liabilities/Net worth: Relación de obligaciones contingentes sobre patrimonio (Contingent Liability/Equity).

X42 - Operating profit/Paid-in capital: Relación de ingresos operativos sobre capital aportado (Operating Income/Capital).

X43 - Net profit before tax/Paid-in capital: Relación de ingreso neto antes de impuestos sobre capital aportado (Pretax Income/Capital).

X44 - Inventory and accounts receivable/Net value: Relación de inventario y cuentas por cobrar sobre valor neto ((Inventory+Accounts Receivables)/Equity).

X45 - Total Asset Turnover: Rotación de activos totales.

X46 - Accounts Receivable Turnover: Rotación de cuentas por cobrar.

- X47 - Average Collection Days: Días promedio de cobro (Days Receivable Outstanding).
- X48 - Inventory Turnover Rate (times): Tasa de rotación de inventarios.
- X49 - Fixed Assets Turnover Frequency: Frecuencia de rotación de activos fijos.
- X50 - Net Worth Turnover Rate (times): Tasa de rotación de patrimonio neto.
- X51 - Revenue per person: Ventas por empleado.
- X52 - Operating profit per person: Ingresos operativos por empleado.
- X53 - Allocation rate per person: Tasa de asignación por empleado (Fixed Assets Per Employee).
- X54 - Working Capital to Total Assets: Relación de capital de trabajo sobre activos totales.
- X55 - Quick Assets/Total Assets: Relación de activos líquidos sobre activos totales.
- X56 - Current Assets/Total Assets: Relación de activos corrientes sobre activos totales.
- X57 - Cash/Total Assets: Relación de efectivo sobre activos totales.
- X58 - Quick Assets/Current Liability: Relación de activos líquidos sobre pasivos corrientes.
- X59 - Cash/Current Liability: Relación de efectivo sobre pasivos corrientes.
- X60 - Current Liability to Assets: Relación de pasivos corrientes sobre activos.
- X61 - Operating Funds to Liability: Relación de fondos operativos sobre pasivos.
- X62 - Inventory/Working Capital: Relación de inventario sobre capital de trabajo.
- X63 - Inventory/Current Liability: Relación de inventario sobre pasivos corrientes.
- X64 - Current Liabilities/Liability: Relación de pasivos corrientes sobre pasivos.
- X65 - Working Capital/Equity: Relación de capital de trabajo sobre patrimonio.
- X66 - Current Liabilities/Equity: Relación de pasivos corrientes sobre patrimonio.
- X67 - Long-term Liability to Current Assets: Relación de pasivos a largo plazo sobre activos corrientes.
- X68 - Retained Earnings to Total Assets: Relación de utilidades retenidas sobre activos totales.

- X69 - Total income/Total expense: Relación de ingresos totales sobre gastos totales.
- X70 - Total expense/Assets: Relación de gastos totales sobre activos.
- X71 - Current Asset Turnover Rate: Tasa de rotación de activos corrientes.
- X72 - Quick Asset Turnover Rate: Tasa de rotación de activos líquidos.
- X73 - Working capital Turnover Rate: Tasa de rotación de capital de trabajo.
- X74 - Cash Turnover Rate: Tasa de rotación de efectivo.
- X75 - Cash Flow to Sales: Relación de flujo de caja sobre ventas.
- X76 - Fixed Assets to Assets: Relación de activos fijos sobre activos totales.
- X77 - Current Liability to Liability: Relación de pasivos corrientes sobre pasivos totales.
- X78 - Current Liability to Equity: Relación de pasivos corrientes sobre patrimonio.
- X79 - Equity to Long-term Liability: Relación de patrimonio sobre pasivos a largo plazo.
- X80 - Cash Flow to Total Assets: Relación de flujo de caja sobre activos totales.
- X81 - Cash Flow to Liability: Relación de flujo de caja sobre pasivos totales.
- X82 - CFO to Assets: Relación de flujo de efectivo operativo sobre activos.
- X83 - Cash Flow to Equity: Relación de flujo de caja sobre patrimonio.
- X84 - Current Liability to Current Assets: Relación de pasivos corrientes sobre activos corrientes.
- X85 - Liability-Assets Flag: 1 si el total de pasivos excede los activos totales, 0 en caso contrario.
- X86 - Net Income to Total Assets: Relación de ingreso neto sobre activos totales.
- X87 - Total assets to GNP price: Relación de activos totales sobre el precio del GNP.
- X88 - No-credit Interval: Intervalo sin crédito.
- X89 - Gross Profit to Sales: Relación de ganancia bruta sobre ventas.
- X90 - Net Income to Stockholder's Equity: Relación de ingreso neto sobre el patrimonio de los accionistas.

X91 - Liability to Equity: Relación de pasivos sobre patrimonio.

X92 - Degree of Financial Leverage (DFL): Grado de apalancamiento financiero.

X93 - Interest Coverage Ratio: Relación de cobertura de intereses (Gastos por intereses sobre EBIT).

X94 - Net Income Flag: 1 si el ingreso neto es negativo durante los últimos dos años, 0 en caso contrario.

X95 - Equity to Liability: Relación de patrimonio sobre pasivos.

