FACULTAD DE ESTUDIOS ESTADÍSTICOS

GRADO EN ESTADÍSTICA APLICADA

Curso 2020/2021

Trabajo de Fin de Grado

Riesgo de quiebra: Análisis predictivo mediante modelos de aprendizaje supervisado

***Alumno*: José Manuel Mendoza Gómez**

***Tutor*: José Luis Valencia Delfa**

Madrid, septiembre de 2021

UNIVERSIDAD COMPLUTENSE
MADRID

RESUMEN

A través de la presente memoria, se intenta analizar la solvencia económica de las empresas taiwanesas desde un enfoque estadístico. Para este estudio, se analizó el conjunto de datos *Taiwan Bankruptcy*. Base de datos que constaba de 6819 empresas que operaban en el mercado taiwanés y que además llevaban 2 años consecutivos teniendo ingresos netos negativos.

La metodología empieza con la depuración de datos. Posteriormente se elimina la correlación entre las variables realizando la técnica del análisis factorial. Seguidamente, se realiza el muestreo necesario para solventar la problemática de las *"clases no balanceadas"* y culmina evaluando las técnicas de clasificación: Regresión logística, algoritmo KNN y bosques aleatorios.

Se eligieron estos algoritmos de clasificación debido a que proporcionan modelos robustos sin tener en cuenta diversos supuestos como el de normalidad. Los resultados que se obtuvieron fueron los que se fijaron previamente en los objetivos. Se eligió la regresión logística como la mejor técnica estadística.

Palabras clave: bancarrota, gestión de riesgos, fracaso financiero, regresión logística, KNN, Random forest.

ABSTRACT

Through this report, an attempt is made to analyse the solvency of Taiwanese companies from a statistical approach. To do this, the *Taiwan Bankruptcy* dataset was analyzed, which consisted of 6,819 firms that operated in the Taiwanese market and had had negative net income for 2 consecutive years.

The methodology begins with data cleansing. Subsequently, the correlation between the variables is eliminated by performing the factor analysis technique. After that, the necessary sampling is carried out to solve the problem of "*unbalanced classes"* and finally, the classification techniques are evaluated: Logistic regression, KNN algorithm and Random Forest.

These classification algorithms were chosen because they provided robust models without considering various assumptions such as normality. The results obtained were as desired. Logistic regression was chosen as the best technique.

Keywords: bankruptcy, risk management, financial failure, logistic regression, KNN, Random Forest.

ÍNDICE

ÍNDICE DE TABLAS

ÍNDICE DE FIGURAS

1. INTRODUCCIÓN

La etimología de la palabra bancarrota (*banca rotta*) surgió en Italia a mediados del siglo XV. Los banqueros ejercían su trabajo en una banca, donde colocaba parte de su patrimonio en oro para que fuera visible, mostrando así, síntomas de solvencia económica.

Cuando un banquero no podía hacer frente a la deuda con sus clientes, este debía romper su banca en público con el fin de advertir a los ciudadanos de que ya no goza de tal solvencia.

Actualmente, debido a que no existe una información teórica establecida como para definir el concepto de bancarrota, es necesario recabar información de empresas en quiebra y trabajar con ellas. Esta información es plasmada en forma de variables (Constand y Yazdipour, 2011)

La detección temprana de este problema es y será una tarea fundamental puesto que la bancarrota tiene la propiedad de ser un efecto contagioso (Doumpos & Zopoudinis, 1999). Es decir, existe una clara tendencia de los inversores en dudar sobre la reputación financiera de la empresa.

En el presente capítulo, se pone en contexto la situación histórica y actual sobre los aspectos más importantes en la investigación de empresas insolventes.

1.1. Situación histórica

El estudio de la solvencia económica de las empresas empezó a estudiarse con modelos paramétricos. Los modelos más comunes fueron la **regresión logística** y el modelo de **análisis discriminante múltiple**. Las variables de estos modelos estaban representadas mayormente por ratios financieros (Andar y Dar, 2006).

Los precursores en la investigación de esta cuestión fueron Fitzpatrick (1932) y Beaver (1936). Para ellos, los predictores más importantes en la detección de empresas en bancarrota fueron las variables relacionadas con los **ratios de liquidez**, los **ratios de deuda** y los **ratios de rotación**.

La siguiente etapa más importante fue introducida por Altman (1968) con el uso de modelos de **análisis discriminante múltiple**. El modelo se denominó **modelo Z-Score de Altman**. Tomó como predictores 22 ratios financieros de los cuales, 5 fueron significativos: **Rentabilidad**, **liquidez**, **solvencia**, **apalancamiento** y **actividad financiera**. Su estudio alcanzó una **precisión del 95%.**

1.2. Situación actual

Desde la década de los 90 hasta la actualidad, gracias a los avances tecnológicos se empezó a perfeccionar los **modelos no paramétricos**. De esta forma, los modelos crecieron en cuanto a complejidad y efectividad.

Los modelos más comunes fueron de **redes neuronales artificiales (ANN)**, **modelos Hazard**, **Híbridos**, etc. Estos modelos al contrario que los paramétricos, no parten de hipótesis preestablecidas.

Odom y Sharda (1990) fueron los pioneros en aplicar redes neuronales a este estudio. Estos modelos aumentaron la precisión debido a que detectaban relaciones no lineales, muy comunes en finanzas.

Un estudio reciente destacado fue el de Park y Hancer (2012). Su investigación consistió en la detección de empresas de restauración con problemas financieros a través de redes neuronales. El estudio consiguió una **sensibilidad del 97.5%** en la muestra de entrenamiento. La variable más importante fue el **ratio de endeudamiento** (cociente entre el pasivo total y el activo total).

2. OBJETIVOS DEL TRABAJO

2.1. Objetivo general

1. Construir modelos estadísticos que garantice la detección de empresas insolventes con un elevado porcentaje de eficacia. Las técnicas estadísticas fijadas son **la regresión logística**, **algoritmo KNN** y **árboles de decisión** (*Random Forest*).

2.2. Objetivos específicos

2. Hacer una comparación entre las tres técnicas predictivas y elegir la más adecuada a la base de datos de *Taiwan Economic Journal* (se busca aquel modelo con mayor sensibilidad).

3. Determinar cuáles son las perturbaciones más relevantes que afectan a la actividad financiera de una empresa, así como también, las de menor incidencia.

4. Comparar los resultados del estudio con los resultados históricos.

3. DESCRIPCIÓN DEL CONJUNTO DE DATOS

3.1. Naturaleza de los datos

Los datos utilizados en este proyecto fueron extraídos del portal *UCI Machine Learning Repository*. Y recopilados por la compañia *Taiwan Economic Journal* (TEJ), compañía proveedora de los datos más precisos y fiables sobre empresas de todo el continente asiático.

La base de datos consta de una muestra de 6819 empresas que operaban en el mercado taiwanés durante los años 1999 a 2009. Todas estas empresas obtuvieron ingresos netos negativos durante 2 años consecutivos.

Las variables explicativas son las que aportan información financiera sobre cada empresa, mientras que la variable respuesta (*Bankrupt_*) indica si la empresa se declaró en bancarrota o de lo contrario, no.

La mayor parte de las empresas no se declararon en bancarrota. En concreto, solo 220 empresas (3.23%) cayeron en quiebra. Las otras 6599 empresas (96.77%) se declararon solventes.

3.2. Análisis descriptivo de las variables

Salvo la variable respuesta (*Bankrupt_*), todas las variables empleadas en este estudio se presentan de forma continua y no estandarizadas.

El conjunto de datos consta de 96 variables. De ellas, se han usado 88 variables debido a que las restantes eran combinación lineal de otras. Esto ocasionaría en un futuro problemas ya que, al aplicar ciertas técnicas estadísticas, la matriz debe de ser regular (determinante distinto de cero).

A continuación, se presentan las variables explicativas que conforman la base de datos.

Variables explicativas	
_ROA_C__before_interest_and_depr	Rentabilidad de los activos totales (ROA) antes de intereses y depreciación antes de intereses.
_ROA_A__before_interest_and___af	Rentabilidad de los activos totales (ROA) antes de intereses y después de impuestos.
_ROA_B__before_interest_and_depr	Rentabilidad de los activos totales (ROA) antes de intereses y depreciación después de impuestos.
_Operating_Gross_Margin	Margen bruto operativo: Beneficio bruto / Ventas netas
_Realized_Sales_Gross_Margin	Margen bruto de ventas realizadas: Beneficio bruto realizado / Ventas netas
_Operating_Profit_Rate	Tasa de beneficio operativo: Ingresos operativos / Ventas netas
_Pre_tax_net_Interest_Rate	Tasa de interés neta antes de impuestos: Ingresos antes de impuestos / Ventas netas
_After_tax_net_Interest_Rate	Tasa de interés neta después de impuestos: Beneficios netos / Ventas netas
_Non_industry_income_and_expendi	Tasa de ingresos netos no operativos: (Gastos e ingresos no industriales) / Ingresos
_Continuous_interest_rate__after	Tasa de interés continua (después de impuestos): Ingresos netos (excluida ganancia por enajenación) / Ventas netas
_Operating_Expense_Rate	Tasa de gastos operativos: Gastos operativos / Ventas netas
_Research_and_development_expens	Tasa de gastos de investigación y desarrollo: (Gastos de investigación y desarrollo) / Ventas netas
_Cash_flow_rate	Tasa de flujo de caja: Flujo de caja operativo / Pasivos circulantes
_Interest_bearing_debt_interest	Tasa de interés de la deuda que devenga intereses: Deuda que devenga intereses / Capital total
_Tax_rate__A_	Tasa impositiva efectiva.
_Net_Value_Per_Share__B_	Valor neto por acción antes de intereses y depreciación después de impuestos.
_Net_Value_Per_Share__A_	Valor neto por acción antes de intereses y después de impuestos.
_Net_Value_Per_Share__C_	Valor neto por acción antes de intereses y depreciación antes de intereses.

Tabla 1.1. Variables explicativas de la base de datos, parte I.

_Persistent_EPS_in_the_Last_Four	Beneficio por acción en los últimos cuatro trimestres.
_Cash_Flow_Per_Share	Flujo de caja por acción
VAR22	Ingresos por acción (yuan ¥)
_Operating_Profit_Per_Share__Yua	Beneficio operativo por acción (yuan ¥)
_Realized_Sales_Gross_Profit_Gro	Tasa de crecimiento del beneficio bruto de las ventas realizadas.
_Operating_Profit_Growth_Rate	Tasa de crecimiento del beneficio operativo.
_After_tax_Net_Profit_Growth_Rat	Tasa de crecimiento del beneficio neto después de impuestos.
_Regular_Net_Profit_Growth_Rate	Tasa regular de crecimiento del beneficio neto: Ingresos de operaciones continuas después del crecimiento fiscal.
_Continuous_Net_Profit_Growth_Ra	Tasa crecimiento continua del beneficio neto: Crecimiento de la ganancia o pérdida por enajenación de los ingresos netos.
_Total_Asset_Growth_Rate	Tasa de crecimiento de activo total
_Net_Value_Growth_Rate	Tasa de crecimiento del capital total.
_Total_Asset_Return_Growth_Rate	Tasa de crecimiento del rendimiento del activs total.
_Cash_Reinvestment__	Tasa de reinversión en efectivo.
_Current_Ratio	Ratio actual: Activos circulantes / Pasivos circulantes
_Quick_Ratio	Test ácido: (Activos circulantes – Inventario) / Pasivos circulantes
_Interest_Expense_Ratio	Tasa de gastos por intereses: Gastos por intereses / Ingresos totales
_Total_debt_Total_net_worth	Pasivo total / Capital total neto
_Debt_ratio__	Tasa de endeudamiento: Pasivo total / Activo total
_Net_worth_Assets	Capital total / Activo total
_Long_term_fund_suitability_rati	Índice de idoneidad de fondos a largo plazo: (Pasivos a largo plazo + Capital) / Activos fijos
_Borrowing_dependency	Dependencia del endeudamiento: Costo de la deuda que devenga.
_Contingent_liabilities_Net_wort	Pasivos contingentes / Capital total neto
_Operating_profit_Paid_in_capita	Ingresos operativos / Capital
_Net_profit_before_tax_Paid_in_c	Beneficios netos antes de impuestos / Capital
_Inventory_and_accounts_receivab	(Inventario + Cuentas por cobrar) / Capital
_Total_Asset_Turnover	Rotación de activos totales
_Accounts_Receivable_Turnover	Rotación de cuentas por cobrar

Tabla 1.2. Variables explicativas de la base de datos, parte II.

_Average_Collection_Days	Días de cobro promedio: Días pendientes por cobrar
_Inventory_Turnover_Rate__times_	Tasa de rotación de inventario (veces)
_Fixed_Assets_Turnover_Frequency	Frecuencia de rotación de activos fijos
_Net_Worth_Turnover_Rate__times_	Tasa de rotación del patrimonio neto (veces): Rotación de acciones
_Revenue_per_person	Ingresos por persona: Ventas por empleado
_Operating_profit_per_person	Beneficio operativo por persona: Ingresos operativos por empleado
_Allocation_rate_per_person	Tasa de asignación por persona: Activos fijos por empleado
_Working_Capital_to_Total_Assets	Capital de trabajo / Activo total
_Quick_Assets_Total_Assets	Activos rápidos / Activo total
_Current_Assets_Total_Assets	Activos circulantes / Activo total
_Cash_Total_Assets	Efectivo / Activo total
_Quick_Assets_Current_Liability	Activos rápidos / Pasivos circulantes
_Cash_Current_Liability	Efectivo / Pasivos circulantes
_Current_Liability_to_Assets	Pasivos circulantes / Activo total
_Operating_Funds_to_Liability	Fondos operativos a responsabilidad
_Inventory_Working_Capital	Inventario / Capital de trabajo
_Inventory_Current_Liability	Inventario / Pasivos circulantes
_Current_Liabilities_Liability	Pasivos circulantes / Pasivo total
_Working_Capital_Equity	Capital de trabajo / Capital total
_Current_Liabilities_Equity	Pasivos circulantes / Capital total
_Long_term_Liability_to_Current	Pasivo a largo plazo / Activos circulantes
_Retained_Earnings_to_Total_Asse	Beneficios retenidos en activo total
_Total_income_Total_expense	Ingreso total / Gasto total
_Total_expense_Assets	Gasto total / activos
_Current_Asset_Turnover_Rate	Tasa de rotación de activos circulantes: activos circulantes a ventas
_Quick_Asset_Turnover_Rate	Tasa de rotación de activos rápidos: activos rápidos a ventas
_Working_capitcal_Turnover_Rate	Tasa de rotación del capital de trabajo: capital de trabajo a ventas
_Cash_Turnover_Rate	Tasa de rotación de efectivo: Efectivo a ventas
_Cash_Flow_to_Sales	Flujo de caja a ventas
_Fixed_Assets_to_Assets	Activos fijos a activos

Tabla 1.3. Variables explicativas de la base de datos, parte III.

_Current_Liability_to_Liability	Pasivos circulantes a pasivos
_Current_Liability_to_Equity	Pasivos circulantes a capital total
_Equity_to_Long_term_Liability	Capital total a pasivo a largo plazo
_Cash_Flow_to_Total_Assets	Flujo de caja a activo total
_Cash_Flow_to_Liability	Flujo de efectivo a pasivos
_CFO_to_Assets	CFO a activos
_Cash_Flow_to_Equity	Flujo de caja a capital total
_Current_Liability_to_Current_As	Pasivos circulantes a activos circulantes
_Liability_Assets_Flag	Indicador de pasivo
_Net_Income_to_Total_Assets	Ingresos netos a activos totales
_Total_assets_to_GNP_price	Activos totales a precio de Producto Nacional Bruto (PNB)
_No_credit_Interval	Intervalo sin crédito
_Gross_Profit_to_Sales	Beneficio bruto a ventas
_Net_Income_to_Stockholder_s_Equ	Beneficio neto del capital contable
_Liability_to_Equity	Pasivo a Capital total
_Degree_of_Financial_Leverage__D	Grado de apalancamiento financiero (DFL)
_Interest_Coverage_Ratio__Intere	Tasa de cobertura de intereses (gasto por intereses a EBIT)
_Net_Income_Flag	Indicador de ingreso neto: 1 si el ingreso neto es negativo durante los últimos dos años, 0 en caso contrario
_Equity_to_Liability	Capital a pasivo

Tabla 1.4. Variables explicativas de la base de datos, parte IV.

La variable respuesta y de interés es *Bankrupt_*. Esta variable indica si una empresa se declaró en bancarrota o no. Es una variable dicotómica y se encuentra codificada (0 si la empresa se declaró solvente o 1 si se declaró insolvente).

Bankrupt_	Frecuencia	Porcentaje
0	6599	96.77
1	220	3.23

Tabla 2. Tabla de frecuencias del conjunto de datos.

4. DEPURACIÓN DE LA BASE DE DATOS

El **preprocesamiento de los datos** es la actividad por la cual se transforma los datos que se encuentran de forma bruta a una más simple con el fin de que sea más entendible tanto para el programa como para los analistas de datos.

El preprocesamiento abarca 3 pasos que son inherentes: Limpieza de datos, transformación de datos y reducción de datos.

4.1. Limpieza de datos (*Data Cleaning*)

La **limpieza de datos** es el acto de depurar los datos completando observaciones con valores faltantes (*missings*), suavizando el ruido de los datos, y también identificando, corrigiendo y si es posible, eliminando registros con datos atípicos.

En nuestro conjunto de datos no se encontraron presencia de valores atípicos. Únicamente se encontraron observaciones influyentes y fue debido a las pocas empresas insolventes.

4.2. Transformación de datos (*Data Transformation*)

"*El propósito fundamental de una preparación de datos es manipular y transformar datos sin procesar para que el contenido de la información incluido en el conjunto de datos pueda exponerse o hacerse más accesible*." (Pyle, 1999).

La **transformación de datos** básicamente consiste en convertir los datos de formato bruto a otro formato más entendible con la finalidad de que el análisis se realice de una forma más cómoda y manejable. Una técnica común de transformar los datos es aplicar el escalamiento.

El escalamiento de los datos consiste en eliminar, a efectos de escala, las diferencias entre las variables dado un rango definido. Con esto se evita el sesgo debido a valores numéricos grandes.

En nuestro estudio se ha aplicado el escalamiento mediante estandarización. Todas las variables siguen una distribución normal con media 0 y desviación típica 1.

4.3. Reducción de datos (*Data Reduction*)

La **reducción de datos** es una transformación numérica o alfabética digitalizada que sirve para minimizar la cantidad de información de un conjunto de datos, pero que, al fin y al cabo, produce resultados analíticos de gran calidad.

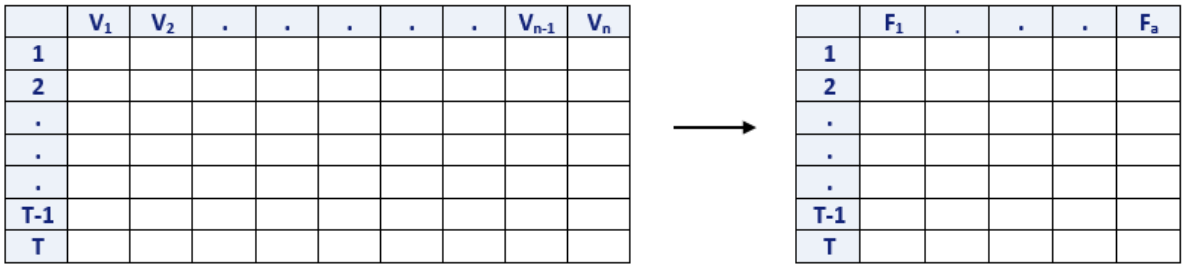


Figura 1. Representación de la reducción de dimensión.

El proceso de transformación de los datos (estandarización) y reducción (análisis factorial) son recogidas al inicio del apartado 5. Metodología.

5. METODOLOGÍA

5.1. Técnicas de preprocesamiento

5.1.1. Análisis factorial

El **análisis factorial** (Spearman,1904) es una técnica estadística de reducción de la dimensión cuya finalidad es explicar la relación existente entre las variables con una menor cantidad de ellas llamadas factores.

Las variables que tienen una alta correlación se agrupan entre sí. Al mismo tiempo, las variables presentan correlaciones bajas con los otros grupos. A partir de las correlaciones de cada grupo de variables, se forma una nueva variable oculta e inobservable directamente llamada factor.

Metodología

Antes de nada, se calcula la matriz de correlaciones R para verificar si existe un grado de asociación significativo. De no ser así, el realizar esta técnica carecería de interés. Los 4 métodos más destacados para examinar el grado de asociación son el test de esfericidad de Bartlett, el determinante R, el índice KMO y la medida de adecuación de la muestra. En este estudio se aplican los 2 últimos.

- Índice KMO de Kaiser-Meyer-Olkin:

  La prueba de adecuación de Kaiser-Meyer-Olkin evalúa conjuntamente el grado de asociación entre las variables. Este estadístico viene dado por:

$$KMO = \frac{\sum_{i \neq j} \sum_{j=1}^{p} r_{i,j}^2}{\sum_{i \neq j} \sum_{j=1}^{p} r_{i,j}^2 + \sum_{i \neq j} \sum_{j=1}^{p} rp_{i,j}^2}$$

  Donde $rp_{i,j}$ es la correlación parcial entre las variables i y j sin el impacto de las variables restantes.

  La distribución de este estadístico esta acotada entre los valores 0 y 1. Cuanto mayor sea el valor, existe una mayor relación entre las variables. Se recomienda un valor mínimo de 0.5 (Kaiser, 1970). Una medida más detallada para este estadístico es la siguiente:

$$\begin{cases} si\ KMO \leq 0.5\ se\ desaconseja\ analisis\ factorial \\ si\ 0.5 \leq KMO \leq 0.7\ valor\ mediocre \\ si\ 0.7 \leq KMO \leq 0.8\ valor\ bueno \\ si\ 0.8 \leq KMO \leq 0.9\ valor\ estupendo \\ si\ KMO \geq 0.9\ valor\ excelente \end{cases}$$

- Medida de adecuación de la muestra MSAj.

  Evaluado para cada variable. Consiste en eliminar sucesivamente del análisis aquellas variables con un MSA inferior a 0.5 (Tabachnick B y Fidell L, 2001).

Como anteriormente se realizó la estandarización de los datos, es indiferente utilizar la matriz de correlaciones o la matriz de covarianzas.

Una vez validado el uso del análisis factorial, es necesario determinar el método de extracción de los factores. Los dos tipos de extracción más destacables son el método de las componentes principales y el método de máxima verosimilitud.

- Extracción mediante componentes principales: extracción sucesiva de aquellos factores que explican la mayor parte de la varianza común. Este tipo de extracción es robusto a violaciones del supuesto de normalidad (Fabrigar et al, 1999)

- Extracción mediante el método de Máxima Verosimilitud: Útil cuando se asume que los datos siguen una distribución normal multivariada y carece de datos anómalos.

Una vez fijado el método de extracción, es necesario también fijar el método de rotación.

"El objetivo del investigador es encontrar aquella solución que proporcione una estructura simple." (Fabrigar et al., 1999).
La rotación factorial ayuda a realizar la interpretación de estos factores obtenidos. Según el tipo de rotación se pueden separar en dos clases: Rotación ortogonal (Quartimax, Varimax y Equimax) y rotación oblicua.

Es necesario también determina el número de factores a retener.
El método más frecuente es seguir la regla de Gutman-kaiser (Guttman, 1954). Esta regla sugiere en retener aquellos factores con valores propias mayores a 1.

5.1.2. Método de espera (*holdout method*)

El **método de espera *(holdout method)*** es una técnica de validación clásica y muy importante que sirve para mitigar las discrepancias y minimizar efectos del sobreajuste del modelo. Consiste en dividir el conjunto de datos en dos subconjuntos de forma aleatoria o lineal.

El bloque mayor corresponde al conjunto de entrenamiento y servirá para construir el modelo de clasificación. El bloque menor es el de validación y se encargará de validar el modelo.

En este estudio, se realiza un muestreo aleatorio simple estratificado. La variable de estratificación es la variable respuesta dicotómica *Bankrupt_*. El 70% de las observaciones corresponden a los datos de entrenamiento y el 30% restante al de validación.

Entrenamiento

Bankrupt_	Frecuencia	Porcentaje
0	4620	96.77
1	154	3.23

Validación

Bankrupt_	Frecuencia	Porcentaje
0	1979	96.77
1	66	3.23

Tablas 3.1 y 3.2. Tabla de frecuencias del fichero de entrenamiento y de validación respectivamente.

5.2. Técnicas de muestreo

5.2.1. Bootstrapping y submuestreo

La técnica de **Bootstraping** (Efron, 1979) es una técnica de muestreo usada para generalizar los resultados de un estudio a partir de distintas muestras representativas de la población. Las muestras extraídas son muestras con reemplazamiento del mismo tamaño.

"El bootstrap es una herramienta estadística extremadamente poderosa y de amplia aplicación que se puede utilizar para cuantificar la incertidumbre asociada con un estimador o método de aprendizaje estadístico determinado" (An Introduction to statistical Learning, 2013).

Realizar la técnica de Bootstraping ayuda a conseguir estimar mejor los modelos ya que al realizar el muestreo, se incluyen observaciones en el conjunto de entrenamiento que antes no estaban. De esta forma, al tener nuevos conjuntos de entrenamiento, se estimará también nuevos modelos.

En este estudio, debido al problema de las *clases no balanceadas*, no se puede tomar las muestras como si fueran representativas de la población. Esto es debido a que algunos algoritmos de clasificación se ven sesgados por la clase predominante.

Una manera de solucionar esta dificultad es combinar la técnica de bootstrapping con la **técnica de submuestreo (*undersampling*).**

Este método alternativo consiste en tomar una muestra con reemplazamiento de la clase predominante y unirlo al conjunto de datos de la clase minoritaria. El tamaño de la nueva clase predominante será el mismo tamaño de la clase minoritaria. Por otro lado, la clase minoritaria se queda en su forma original.

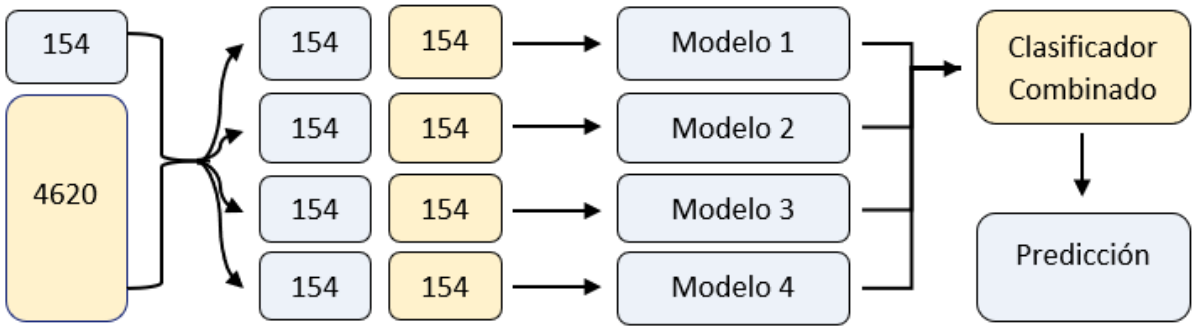


Figura 2. Técnicas muestrales aplicadas al estudio

5.3. Técnicas predictivas

5.3.1. Regresión logística

La **regresión logística** (D. Cox, 1958) es un modelo de regresión no lineal con variable dependiente discreta, que puede tomar dos o más categorías. Es un modelo robusto y como ventaja frente a otros algoritmos de clasificación es la simplicidad de su interpretación.

La regresión logística está basada en la función sigmoide:

$$f(z) = \frac{1}{1 + e^{-z}} = p_i$$

Expresado en términos del vector x=(x1,x2,…,xn) de variables explicativas:

$$f(x_1, \dots, x_n) = \frac{1}{1 + e^{-(B_0 + B_1' x_1 + \cdots + B_n' x_n)}}$$

Donde f(z) es una función de probabilidad y, además, cumple las siguientes propiedades:

$$\lim_{z \to -\infty} \frac{1}{1 + e^{-z}} = 0$$

$$\lim_{z \to +\infty} \frac{1}{1 + e^{-z}} = 1$$

El modelo logistico tambíen se puede expresar linealmente a partir de las siguientes transformaciones:

$$log\left(\frac{p_i}{1 - p_i}\right) = \frac{\frac{1}{1 + e^{-(B_0 + B_1' x_1 + \cdots + B_n' x_n)}}}{\frac{e^{-(B_0 + B_1' x_1 + \cdots + B_n' x_n)}}{1 + e^{-(B_0 + B_1' x_1 + \cdots + B_n' x_n)}}} = log\left(\frac{1}{e^{-(B_0 + B_1' x_1 + \cdots + B_n' x_n)}}\right) = B_0 + B_1' x_1 + \cdots + B_n' x_n$$

Aplicando la función sigmoide a este estudio, la probabilidad de que ocurra el evento Bancarrota tiene la forma:

$$P_i = P(Y = Bancarrota / X = x_i) = \frac{1}{1 + e^{-(B_0 + B_1' x_1 + \cdots + B_n' x_n)}}$$

La metodología de la regresión logística sigue la siguiente estructura:

1. Estimación
2. Diagnosis
3. Validación
4. Interpretación.

Metodología

Estimación del modelo

Estimación de los parámetros del modelo mediante la puntuación de Fisher (Fisher,1928).

Diagnosis del modelo

Se verifica la adecuación del modelo a la base de datos.

- Significatividad conjunta del modelo: Para todas las variables regresoras, contrasta si existe al menos una con un coeficiente igual a 0. Este contraste realiza la misma función que el contraste de la F de Snedecor para regresiones lineales. Las pruebas realizadas corresponden al likelihood ratio, Score y Wald.

$$H_0: B_0 = \ldots = B_n = 0 \qquad H_1: \exists i \,/\, B_i \neq 0$$

- Test de Homer-Lemeshow. Este test comprueba la bondad de ajuste verificando si los valores observados son similares a los esperados. Las hipótesis de esta prueba son:

$$H_0: El\ modelo\ se\ ajusta\ bien\ a\ los\ datos$$
$$H_1: El\ modelo\ no\ se\ ajusta\ bien\ a\ los\ datos$$

- Significatividad individual: Para cada variable regresora, contrasta si su coeficiente es igual a 0, es decir, contrasta individualmente la significación de las variables independientes. La prueba realizada corresponde al test de Wald del chi cuadrado.

Validación del modelo

- Curva ROC: Evalúa la capacidad discriminante del test comparando para distintos puntos de corte, las medidas de sensibilidad y especificidad. El área bajo la curva (AUC) puede tomar valores que oscilan entre 0,5 y 1. La prueba se resume en la siguiente tabla.

$$\begin{cases} si\ 0.5 < AUC \leq 0.7\ baja\ exactitud \\ si\ 0.7 < AUC \leq 0.9\ útiles\ para\ algunos\ propósitos \\ si\ 0.9 < AUC \leq 1\ alta\ exactitud \end{cases}$$

El punto de corte o valor umbral se fija por el analista dependiendo de la finalidad de la investigación. En este estudio, debido al problema de las *clases no balanceadas*, interesa tener una mayor sensibilidad antes que la especificidad. Es decir, es más conveniente saber si una empresa puede caer en quiebra que no saberlo.

La interpretación del AUC aplicada al estudio se podría definir como *"seleccionando dos empresas al azar, una en bancarrota y otra que no, la probabilidad de clasificar correctamente a cada una de ellas es del (valor del AUC) %".*

- Matriz de confusión: Matriz cuadrada donde cada fila representa el valor real de cada clase y cada columna su valor predicho. Sirve como otra forma de evaluar la capacidad predictiva del modelo. La precisión de la clasificación está condicionada por el valor umbral fijado.

<u>Interpretación del modelo</u>

- Odds ratio: Cuantifica la influencia de cada variable regresoras sobre la variable respuesta. Debido a que el cálculo de los efectos no es posible en el modelo logístico, se recurre al uso de los odds ratio o razones. Su interpretación es la siguiente:

  - Valores cercanos a 1 indican ausencia de asociación entre efecto del factor y el evento.
  - Valores inferiores a 1 indican relación inversa entre efecto del factor y el evento.
  - Valores superiores a 1 indican relación directa entre efecto del factor y el evento

5.3.2. Algoritmo KNN

El **algoritmo KNN** (Rosenblatt, 1956) cuyas siglas significan *K-Nearest Neighbors*, es un algoritmo de clasificación no paramétrico, ya que no presupone que los datos sigan una distribución concreta y basado en instancias, es decir, no aprende explícitamente del modelo, tan solo memoriza los datos del entrenamiento y su clase. Por ello, la duración del aprendizaje es trivial respecto al de clasificación.

Funcionamiento

La clasificación se fundamenta en la cercanía respecto a las otras instancias, es decir, la instancia se clasifica en función de las k-instancias de entrenamiento más cercanas a ella.. Si la mayor parte de las k-instancias pertenecen a una determinada categoría cuyas características son semejantes entre sí, entonces la instancia analizada también pertenece a esa categoría.

Metodología

La metodología de esta técnica es la siguiente:

1. Para cada observación del conjunto de datos de validación se calcula la distancia con todas las observaciones del conjunto de entrenamiento.
2. Las distancias entre las observaciones se colocan de forma ascendente.
3. Se seleccionan las k-instancias más cercanas.
4. Se evalúa la frecuencia de cada categoría de las k-instancias seleccionadas.
5. Se realiza la votación y se clasifica la instancia de prueba a la categoría predominante.

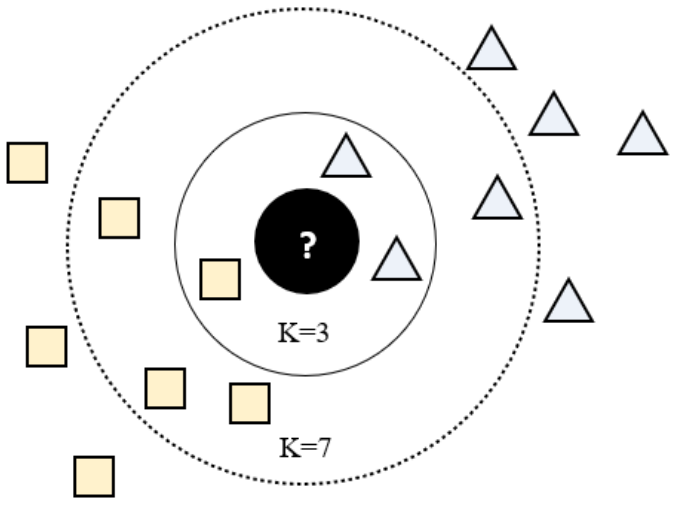


Figura 3. Ejemplo de Algoritmo KNN. En esta figura, si el parámetro es k=3, la clase predicha es un triángulo. En cambio, si el parámetro es k=7, la clase predicha es un cuadrado.

Elección de distancia

La elección de la distancia adecuada dependerá del tipo de variables existente. En este estudio debido a que se está utilizando una variable numérica, se empleó la distancia euclídea.

La distancia euclídea es la distancia más común y representa, en forma de línea, la distancia existente entre dos puntos de un espacio euclídeo. Su fórmula matemática es:

$$d(x, y) = \sqrt{\sum_i (x_i - y_i)^2}$$

Validación

Una vez elegida la distancia, se valida el modelo a través de las matrices de confusión.

- Resustitución: El fichero de entrenamiento sirve también como fichero de validación.

- Validación cruzada: El fichero de entrenamiento se divide en P particiones de igual tamaño. Se estima el modelo en función de P-1 particiones. La partición restante sirve como conjunto de validación. Se repite el procedimiento hasta que todas las particiones actúen alguna vez como fichero de validación. Al finalizar, se promedian los resultados para obtener una sola estimación.

- Método de espera (*Hold-Out method*): El conjunto de datos es separado en dos subconjuntos. El de entrenamiento y el de validación. El modelo se construye en función de los datos de entrenamiento y se evalua con los datos de validación.

5.3.3. Bosques aleatorios (*Random Forest*)

El algoritmo de bosques aleatorios (*Random forest*) es otra técnica estadística muy popular desde su introducción en 2001 (Breiman). Su notoriedad se debe mayoritariamente a la robustez que proporciona.

El bosque aleatorio es tan solo una agrupación de árboles de decisión. El resultado final del algoritmo es el promedio de los resultados de los árboles de decisión.

Metodología

La creación de cada árbol de decisión consta de dos partes:

1. Selección aleatoria de variables predictoras: De todas las variables que conforman la base de datos, se seleccionan aleatoriamente P variables (*sin contar la variable dependiente*). La variable de interés se incluye también.

   Un método común para determinar el número de variables a elegir en cada selección P es la regla del pulgar (*rule of thumb*). Esta regla establece el valor de N como la raíz cuadrada del número de variables totales.

2. Selección aleatoria de observaciones: De todas las observaciones que conforman la base de datos, se seleccionan aleatoriamente y con reposición N observaciones. El tamaño del nuevo conjunto de datos es de la misma dimensión que la base de datos original.

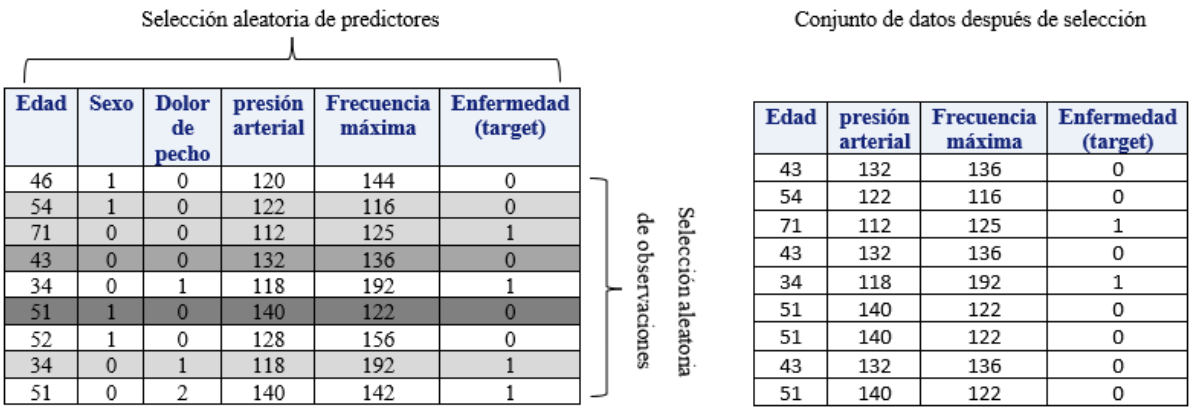


Figura 4. Ejemplo de selección de observaciones y variables aleatorias con reemplazamiento en el algoritmo Random Forest. La escala de grises indica cuantas veces se ha seleccionado cada observación.

Una vez completado estos dos pasos, se elige el nodo raíz (*primer nodo*) de entre todas las variables. La elección se evalúa en función de la impureza de Gini. Aquella variable que tenga menor impureza en el nodo indicará una mayor homogeneidad. Entendiendo homogeneidad como comportamiento semejante entre observaciones.

Para cada variable seleccionada, se crean nodos de decisión. El nodo padre o raíz (*root node*) está formado por aquella variable que genere el subconjunto de datos lo más homogéneo posible.
Fijado el nodo raíz, se seleccionan nuevamente un conjunto de variables para establecer el nodo intermedio (*branch node*).

Este procedimiento no finaliza hasta que se construye todo el árbol de decisión. Una vez terminada la construcción del árbol, se repite el procedimiento según el valor fijado en la opción NTREE=A (Número de árboles).

Un método eficaz para determinar el hiperparámetro NTREE es escoger aquel número de árboles tal que la tasa de clasificación errónea para las observaciones OOB sea la más inferior.

Una vez construido todos los árboles de decisión que conforman el bosque aleatorio, se define el resultado final a través de la votación por mayoría ("*majority voting ensemble"*) si la variable dependiente es categórica o se hace el promedio de los resultados si la variable dependiente es continua.

La votación por mayoría consiste en combinar el resultado de los múltiples árboles de decisión.

<u>Validación</u>
Existen 2 formas de validar el algoritmo de bosques aleatorios: OOB y método de espera

- OOB: Normalmente 1/3 de las observaciones que conforman la base de datos no son empleadas en los árboles de decisión. Estas observaciones se conocen como observaciones fuera de bolsa *("Out-Of Bag"*).

  Estas observaciones pueden ser utilizadas como conjunto de validación y así, verificar la precisión del algoritmo.

- Método de espera: Método ya mencionado en los anteriores apartados.

6. RESULTADOS

6.1. Resultados de análisis factorial

Debido a la enorme cantidad de variables que tiene la base de datos (88 variables), se procedió a reducir la dimensión de los datos. Todas las variables se encontraban de forma continua, ninguna era dicotómica, por lo tanto, cumplían uno de los supuestos del análisis factorial (Collins, 2002).

Se fijó el valor 0.5 como punto de corte para la medida de adecuación de la muestra (MSA_j). De esta forma, se eliminaron iterativamente las variables de menor MSA_j hasta conseguir que todas las variables tuvieran un valor de MSA mayor o igual a 0.5.

Se utilizó el índice KMO como medida de adecuación de los datos para realizar el análisis factorial y fue evaluado según las medidas de Kaiser.

La extracción de los factores se realizó mediante el método de las componentes principales ya que es robusta ante los posibles supuestos de no normalidad de los datos.

Se eligió la rotación QUARTIMAX debido a su eficacia cuando el número de factores a retener es elevado.

Se utilizó la regla de Gutman-kaiser para determinar el número de factores a retener.

Para finalizar, se comprobó la ausencia de multicolinealidad a través de la matriz de correlación de Pearson y se interpretó cada factor que se retuvo.

Medida de Kaiser de suficiencia muestral: MSA total = 0.80302881			
_After_tax_Net_Profit_Growth_Rat	0.60772855	*_Net_Value_Per_Share__B_*	0.96481752
_After_tax_net_Interest_Rate	0.72992781	*_Net_Value_Per_Share__C_*	0.88701341
_Allocation_rate_per_person	0.56277114	*_Net_Worth_Turnover_Rate__times_*	0.65353893
_Borrowing_dependency	0.9059217	*_Net_profit_before_tax_Paid_in_c*	0.89652122
_CFO_to_Assets	0.82408036	*_No_credit_Interval*	0.64846996
_Cash_Current_Liability	0.66168959	*_Operating_Expense_Rate*	0.7782615
_Cash_Flow_Per_Share	0.88584465	*_Operating_Funds_to_Liability*	0.79274014
_Cash_Flow_to_Equity	0.51445189	*_Operating_Gross_Margin*	0.724109
_Cash_Flow_to_Liability	0.73081373	*_Operating_Profit_Growth_Rate*	0.95050991
_Cash_Flow_to_Total_Assets	0.67738339	*_Operating_Profit_Per_Share__Yua*	0.85757464
_Cash_Reinvestment__	0.76114039	*_Operating_Profit_Rate*	0.74298767
_Cash_Total_Assets	0.8631139	*_Operating_profit_Paid_in_capita*	0.85394073
_Cash_Turnover_Rate	0.80968981	*_Operating_profit_per_person*	0.88891438
_Cash_flow_rate	0.79501435	*_Per_Share_Net_profit_before_tax*	0.93852053
_Contingent_liabilities_Net_wort	0.69923984	*_Persistent_EPS_in_the_Last_Four*	0.89651049
_Continuous_Net_Profit_Growth_Ra	0.91279432	*_Pre_tax_net_Interest_Rate*	0.76265267
_Continuous_interest_rate__after	0.8481896	*_Quick_Asset_Turnover_Rate*	0.80125117
_Current_Asset_Turnover_Rate	0.77170175	*_Quick_Assets_Total_Assets*	0.81201029
_Current_Assets_Total_Assets	0.60347769	*_Quick_Ratio*	0.51989364
_Current_Liabilities_Equity	0.70134381	*_ROA_A__before_interest_and___af*	0.89635654
_Current_Liability_to_Current_As	0.85502749	*_ROA_B__before_interest_and_depr*	0.83890225
_Debt_ratio__	0.65844752	*_ROA_C__before_interest_and_depr*	0.86952877
_Degree_of_Financial_Leverage__D	0.80431161	*_Realized_Sales_Gross_Margin*	0.72343775
_Equity_to_Liability	0.7635441	*_Realized_Sales_Gross_Profit_Gro*	0.5627707
_Equity_to_Long_term_Liability	0.63084065	*_Regular_Net_Profit_Growth_Rate*	0.6089032
_Fixed_Assets_Turnover_Frequency	0.87374042	*_Research_and_development_expens*	0.68702513
_Fixed_Assets_to_Assets	0.51975794	*_Retained_Earnings_to_Total_Asse*	0.95051183
_Interest_Expense_Ratio	0.51231549	*_Tax_rate__A_*	0.73877991
_Interest_bearing_debt_interest	0.72903283	*_Total_Asset_Growth_Rate*	0.61046643
*_Inventory_Turnover_Rate__times_*	0.69480767	*_Total_Asset_Turnover*	0.71340427
_Inventory_and_accounts_receivab	0.61086805	*_Total_assets_to_GNP_price*	0.71717892
_Liability_to_Equity	0.71671368	*_Total_debt_Total_net_worth*	0.56500994
_Long_term_fund_suitability_rati	0.54325884	*_Total_expense_Assets*	0.74457032
_Net_Income_to_Stockholder_s_Equ	0.80021925	*_Working_Capital_Equity*	0.65297949
_Net_Income_to_Total_Assets	0.88595599	*_Working_Capital_to_Total_Assets*	0.71535975
*_Net_Value_Per_Share__A_*	0.86251919		

Tabla 4. Medidas de adecuación muestral de las variables que entraron al estudio.

Las variables que consiguieron entrar en el estudio se encuentran en la tabla 5, todas ellas tuvieron un valor superior a 0.5. En consecuencia, estas variables seleccionadas tuvieron una mayor correlación con los factores que las no seleccionadas.

El valor del índice KMO fue de 0.8030, un valor alto, luego fue apropiado realizar la reducción de dimensión mediante el análisis factorial.

Autovalores de la matriz de correlación: Total = 71 Promedio = 1				
	Autovalor	Diferencia	Proporción	Acumulada
1	11.953158	6.3998074	0.1684	0.1684
2	5.5533502	1.6125022	0.0782	0.2466
3	3.940848	0.161067	0.0555	0.3021
4	3.779781	0.0726072	0.0532	0.3553
5	3.7071739	0.9957985	0.0522	0.4075
6	2.7113754	0.1560353	0.0382	0.4457
7	2.5553401	0.482461	0.036	0.4817
8	2.072879	0.1638082	0.0292	0.5109
9	1.9090709	0.1111534	0.0269	0.5378
10	1.7979175	0.0821084	0.0253	0.5631
11	1.715809	0.2587153	0.0242	0.5873
12	1.4570937	0.1000966	0.0205	0.6078
13	1.3569971	0.1331716	0.0191	0.6269
14	1.2238255	0.087022	0.0172	0.6441
15	1.1368035	0.016047	0.016	0.6602
16	1.1207565	0.071072	0.0158	0.6759
17	1.0496845	0.0076273	0.0148	0.6907
18	1.0420572	0.0332173	0.0147	0.7054
19	1.0088399	0.0064541	0.0142	0.7196
20	1.0023858	0.0105794	0.0141	0.7337
21	0.9918065	0.0117203	0.014	0.7477

Tabla 5. Autovalores de los factores.

Hubo 20 autovalores mayores a la unidad. Por lo tanto, siguiendo la regla de Gutman-Kaiser se retuvo la misma cantidad de factores. **Estos 20 factores retenidos lograron explicar un 73.37% de la variabilidad total de los datos.**

Reducir la dimensión de esta base de datos ocasionó una pérdida de información del 26,63%.
Algo totalmente natural ya que al principio del estudio se estaba trabajando con 88 variables y realizar el análisis factorial consiguió reducir el número de variables a más del 75% (El número de variables final fue 20).

Modelo factorial de rotación					
	Factor1	Factor2	Factor3	Factor4	Factor5
_Persistent_EPS_in_the_Last_Four	0.95664	-0.0566	0.00626	0.0713	0.06603
_Per_Share_Net_profit_before_tax	0.93884	-0.05778	0.00681	0.0884	0.05638
_Net_profit_before_tax_Paid_in_c	0.93794	-0.05493	0.0063	0.09078	0.06357
_Operating_Profit_Per_Share__Yua	0.87684	-0.02954	0.00291	0.17707	0.13582
_Operating_profit_Paid_in_capita	0.87432	-0.0284	0.00273	0.17655	0.13753
_Net_Value_Per_Share__B_	0.86532	-0.02478	0.00888	-0.08935	-0.05878
_Net_Value_Per_Share__A_	0.865	-0.02581	0.00884	-0.08849	-0.05935
_Net_Value_Per_Share__C_	0.86489	-0.02576	0.00888	-0.08852	-0.05954
_ROA_A__before_interest_and___af	0.74868	-0.09073	0.02208	0.10264	0.13798
_ROA_C__before_interest_and_depr	0.74624	-0.09347	0.0249	0.06087	0.21641
_ROA_B__before_interest_and_depr	0.73234	-0.08938	0.02518	0.04363	0.2082
_Net_Income_to_Total_Assets	0.68093	-0.11274	0.01835	0.08109	0.12508
_Operating_profit_per_person	0.38462	0.03203	0.00712	0.05411	-0.15561
_Liability_to_Equity	-0.06451	0.97456	0.00092	0.04547	-0.02596
_Borrowing_dependency	-0.09005	0.96607	0.00085	-0.04759	-0.06786
_Current_Liabilities_Equity	-0.04879	0.91732	0.00171	0.10443	-0.05959
_Equity_to_Long_term_Liability	-0.07181	0.81187	-0.00195	-0.08432	0.05774
_Inventory_and_accounts_receivab	-0.01678	0.69766	0.0097	0.28644	-0.14033
_Net_Income_to_Stockholder_s_Equ	0.17821	-0.81457	0.00723	0.05013	0.02459
_Pre_tax_net_Interest_Rate	0.02685	-0.00111	0.99484	0.01057	0.00458
_Continuous_interest_rate__after	0.02438	0.00014	0.99429	0.00891	0.00464
_After_tax_net_Interest_Rate	0.02355	-0.00001	0.97963	0.01427	0.00358
_Operating_Profit_Rate	0.0156	0.00003	0.94133	0.00904	0.01082
_Current_Assets_Total_Assets	0.12427	0.04877	0.02309	0.86562	-0.11013
_Quick_Assets_Total_Assets	0.15192	-0.02698	0.01385	0.8087	0.06872
_Total_Asset_Turnover	0.18834	0.04132	0.02107	0.6878	0.05351
_Working_Capital_to_Total_Assets	0.20181	-0.10056	0.01339	0.61922	-0.00362
_Net_Worth_Turnover_Rate__times_	0.05236	0.22994	0.01408	0.53478	0.11602
_Fixed_Assets_Turnover_Frequency	-0.10087	0.02189	0.02159	-0.54418	0.09821
_Cash_Reinvestment__	0.1253	-0.11174	0.00795	0.02147	0.84456
_CFO_to_Assets	0.29374	-0.03716	0.01327	-0.04764	0.82753
_Cash_Flow_Per_Share	0.43107	-0.00785	0.00217	-0.02948	0.75452

Tabla 6.1. Matriz de factores rotados, parte I.

1. El primer factor esta correlacionado con las variables relacionadas con la rentabilidad.

   El término *Per Share* se usa para describir los beneficios netos/brutos según una determinada operación (comercio, bienes y raíces,) dividido por el número de acciones.

   El termino *Net Value Per Share* es utilizado para expresar el valor del patrimonio neto por cada acción. Valores bajos indican que la empresa no tiene un buen sustento económico.

   El término ROA es una fórmula matemática que evalúa los retornos de activos (*Return On Assets*). Este indicador mide la rentabilidad en función de los recursos que posea. Valores bajos indican menor rentabilidad.

2. El segundo factor esta correlacionado con las variables relacionadas a las deudas de la empresa. El termino *Liabilities* significa pasivos u obligaciones.

3. El tercer factor esta correlacionado con las variables relacionadas a las tasas de interés. Es el ratio de los ingresos dividido entre las ventas. Un valor bajo indica que se han realizado operaciones de forma errónea.

4. El cuarto factor está representado por las variables relacionadas a los activos de la empresa. El termino *Assets* significa bienes. Es conveniente que estos valores sean altos.

5. El quinto factor está representado por las variables relacionadas al flujo de caja, en concreto, al flujo de entrada.
   El *Cash Flow* o flujo de caja es un indicador que sirve para ver la cantidad de dinero que se mueve dentro y fuera de la empresa.

Modelo factorial de rotación					
	Factor6	**Factor7**	**Factor8**	**Factor9**	**Factor10**
_Debt_ratio__	0.79158	-0.01648	-0.13338	0.02962	-0.06875
_Current_Liability_to_Current_As	0.62049	-0.01815	-0.00822	0.0017	-0.06806
_Equity_to_Liability	-0.65076	0.00129	0.02375	-0.01535	-0.02871
_After_tax_Net_Profit_Growth_Rat	-0.00554	0.9619	0.01623	0.00398	0.03933
_Regular_Net_Profit_Growth_Rate	-0.0056	0.96107	0.0148	0.00359	0.04064
_Operating_Profit_Growth_Rate	-0.01661	0.80799	-0.00499	-0.00457	-0.01563
_Realized_Sales_Gross_Margin	-0.09663	0.02284	0.92428	0.03492	0.01867
_Operating_Gross_Margin	-0.09649	0.02289	0.92388	0.03475	0.01964
_Operating_Expense_Rate	0.07359	-0.00266	-0.31333	0.01951	0.15856
_Cash_Flow_to_Total_Assets	-0.03891	0.01199	0.00033	0.88637	0.09346
_Cash_Flow_to_Liability	-0.04283	0.00865	0.07714	0.78324	0.0391
_Cash_Flow_to_Equity	0.0715	-0.01801	-0.04144	0.71474	-0.00726
_Cash_Total_Assets	-0.42732	0.00361	0.14726	0.43602	-0.10979
_Retained_Earnings_to_Total_Asse	-0.1207	0.00398	-0.01714	0.0501	0.57266
_Total_expense_Assets	0.00631	-0.0026	0.38193	-0.10785	-0.6028

Tabla 6.2 Matriz de factores rotados, parte II.

6. El sexto factor esta correlacionado con las variables relacionadas al riesgo. El riesgo también es conocido como apalancamiento.

   El apalancamiento es un mecanismo financiero que consiste en obtener financiación (a menudo contrayendo una deuda) para destinar ese dinero a una inversión. De este modo existe la posibilidad de obtener más beneficios o de lo contrario más perdidas.

7. El séptimo factor está relacionado con las variables referentes a la tasa de crecimiento de la empresa.

   El término *Growth Rate* sirve para expresar en términos anuales, la tasa de crecimiento de una variable. Es conveniente que estos valores sean altos.

8. El octavo factor se relaciona con las variables referentes al margen bruto.

   El termino *Gross Margin* es la diferencia entre los ingresos menos los costos en los que pueda incurrir.

9. El noveno factor está relacionado con las variables referentes al flujo de caja, pero esta vez, con el flujo de caja de salida.

10. El décimo factor está correlacionado con las variables relacionadas al colchón financiero de la empresa, tanto actual, como pasada. Es decir, el dinero destinado a imprevistos.

Modelo factorial de rotación					
	Factor11	**Factor12**	**Factor13**	**Factor14**	**Factor15**
_Quick_Asset_Turnover_Rate	0.79061	-0.01918	0.0009	0.02619	0.03635
_Current_Asset_Turnover_Rate	0.78605	-0.03361	0.01002	-0.03176	-0.04964
_Total_debt_Total_net_worth	-0.05167	0.8071	-0.01378	-0.00271	-0.02741
_Cash_flow_rate	0.03863	0.64429	0.01412	0.04238	0.0178
_Operating_Funds_to_Liability	0.01467	0.58287	0.01383	0.01875	0.01804
_Long_term_fund_suitability_rati	0.02965	0.07467	0.82483	0.02311	0.04216
_Fixed_Assets_to_Assets	-0.039	-0.03038	0.69691	0.02769	0.0236
_Allocation_rate_per_person	0.02026	-0.02622	0.64365	-0.04277	-0.05736
_Working_Capital_Equity	-0.03254	0.02107	0.01005	0.71697	0.00613
_Contingent_liabilities_Net_wort	-0.02926	0.00881	0.01315	-0.68847	0.0256
_Quick_Ratio	0.00759	-0.00639	-0.04285	-0.01662	0.6865
_Cash_Current_Liability	-0.0203	-0.02095	0.02839	-0.00114	0.67635

Tabla 6.3. Matriz de factores rotados, parte III.

11. El décimo primer factor esta correlacionado con las variables referentes al índice de rotación.

   El término *Turnover Rate* es un índice que mide la capacidad de la empresa para generar ventas a través de sus bienes.

12. El décimo segundo factor esta correlacionado con las variables relacionadas a ratios que miden la solvencia de la empresa.

_Total_debt_Total_net_worth es una variable que indica el cociente entre el pasivo total y el capital total.

_Cash_flow_rate es variable mide la capacidad de cubrir los pasivos con los flujos de caja de la empresa

_Operating_Funds_to_Liability hace referencia al presupuesto destinado a los gastos esperados de la empresa.

13. El décimo tercer factor es correlacionado con las variables relacionadas al activo fijo de la empresa.

14. El décimo cuarto factor esta correlacionado con las variables relacionadas a cocientes donde intervienen el patrimonio.

    En primer lugar, la variable *_Working_Capital_Equity* es un cociente entre el capital de trabajo y el patrimonio.

    En segundo lugar y opuestamente, se encuentra la variable *_Contingent_liabilities_Net_wort*, que es un cociente entre el pasivo contingente (obligación si se cumple una condición requerida) y el patrimonio.

15. El décimo quinto factor esta correlacionado con las variables relacionadas a la liquidez a corto plazo.

    La liquidez a corto plazo mide la capacidad de cumplir los pasivos a corto plazo con sus activos más líquidos.

Modelo factorial de rotación					
	Factor16	**Factor17**	**Factor18**	**Factor19**	**Factor20**
_Total_Asset_Growth_Rate	0.40953	0.0752	-0.35048	-0.12381	0.04461
_Continuous_Net_Profit_Growth_Ra	0.20767	-0.18844	-0.0223	0.05902	-0.02379
_No_credit_Interval	-0.38221	-0.29797	-0.17153	-0.25524	-0.25035
_Total_assets_to_GNP_price	-0.43814	0.0503	-0.07968	-0.00094	-0.01144
_Inventory_Turnover_Rate__times_	-0.05852	0.73106	-0.06476	0.05266	-0.01533
_Cash_Turnover_Rate	0.21855	0.32485	-0.11053	-0.27172	-0.14264
_Realized_Sales_Gross_Profit_Gro	0.07717	-0.06916	0.59325	0.00756	0.09877
_Research_and_development_expens	0.31532	0.046	0.40375	-0.34655	-0.28378
_Tax_rate__A_	0.14142	-0.11851	-0.33547	0.07711	0.02702
_Interest_bearing_debt_interest	0.02129	0.03667	-0.00628	0.77212	-0.13461
_Degree_of_Financial_Leverage__D	0.07472	0.13044	0.11915	0.0141	0.69702
_Interest_Expense_Ratio	-0.03698	-0.15269	-0.0681	-0.12047	0.54146

Tabla 6.4. Matriz de factores rotados, parte IV.

16. El décimo sexto factor esta correlacionado con las variables relacionadas a las tasas de crecimiento de los activos y las ganancias en contraposición al indicador de liquidez a corto plazo. *_No_credit_Interval* es un indicador que mide cuanto tiempo la empresa puede seguir subsistiendo si dejan de tener ingresos.

17. El décimo séptimo factor esta correlacionado con las variables relacionadas a la tasa de rotación de sus activos, es decir, cada cuanto se renuevan sus activos.
    La variable *_Inventory_Turnover_Rate__times_* es un indicador que se utiliza para saber cada cuanto tiempo se renueva el inventario

    La variable *_Cash_Turnover_Rate* es un indicador que sirve para ver cada cuanto tiempo se gasta el efectivo de la empresa.

18. El décimo octavo factor esta correlacionado con las variables relacionadas al crecimiento debido a las ventas y como consecuencia, la tasa impositiva efectiva a la que se enfrenta.
    La tasa impositiva efectiva es el impuesto sobre la renta de una empresa sobre sus ingresos.

19. El décimo noveno factor está formado por una única variable. Esto quiere decir que no se ha encontrado una gran correlación entre esta variable y las demás.
    La variable de este factor es *Interest_bearing_debt_interest*, y sirve para evaluar el monto de deuda pendiente que enfrenta respecto al patrimonio que posee.

20. El vigésimo factor esta correlación con las variables relacionadas al grado de inversión/subsistencia de la empresa con sus propios recursos.

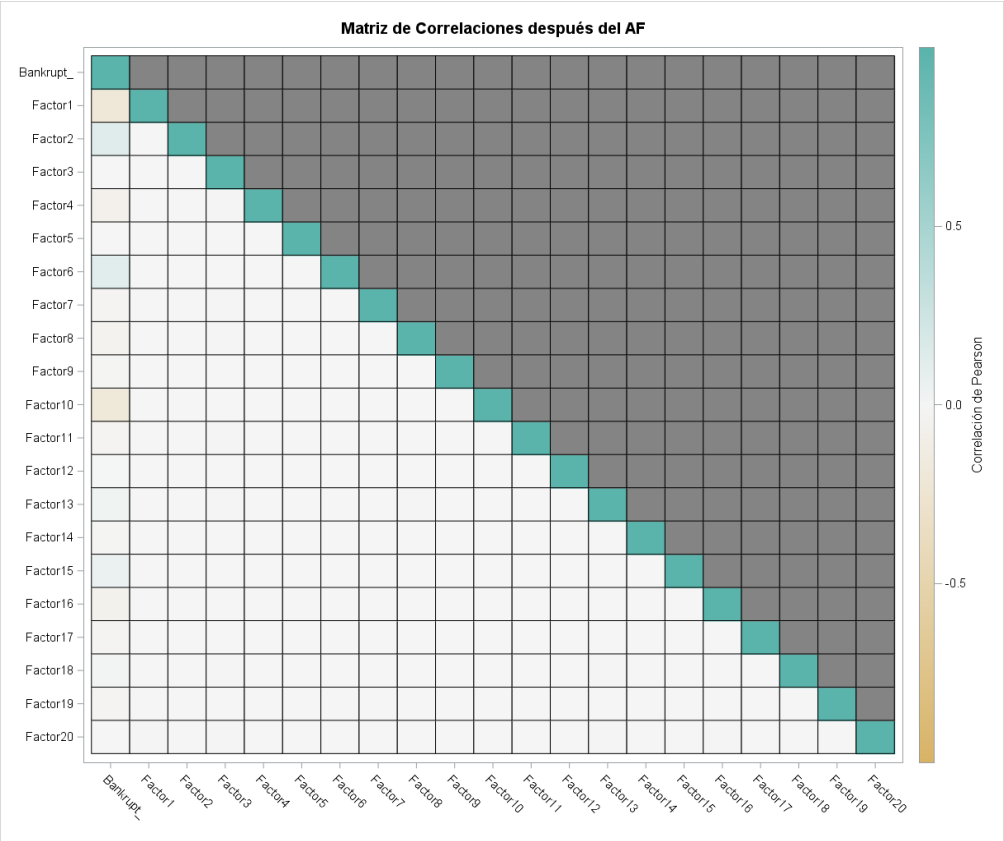


Figura 5. Matriz de correlación lineal de Pearson.

Las variables que conformaron la matriz de coeficientes de correlación de Pearson de la figura 4 fueron los factores retenidos y la variable dependiente *Bankrupt_*.

El análisis factorial consiguió eliminar cualquier indicio de interdependencia lineal entre los factores.

En cuanto a la relación existente entre los factores y la variable dependiente se percibe que las variables con mayor fuerza de asociación son los *factores 1,2,4,6 y 10.*

Estos factores pueden catalogarse en dos categorías.

- Por un lado, están los factores que presentan asociación positiva, que son los factores que tienen una **relación directa** con la variable bancarrota y son los ***factores 2 y 6***. Como se mencionó anteriormente, estos factores están relacionados con la **deuda** y el **apalancamiento**.

- Por otro lado, están los factores que presentan asociación negativa, que son los factores que tienen una **relación inversa** con la variable de interés y son los ***factores 1, 4 y 10***. Como se mencionó anteriormente, estos factores están relacionados con la **rentabilidad**, los **activos** y el **colchón financiero** de la empresa.

6.2. Resultados de bootstrapping y submuestreo

Una vez realizado el análisis factorial se llevó a cabo la técnica estadística de bootstrapping combinado con el submuestreo.

En el fichero de entrenamiento había solamente 154 empresas que se habían declarado en bancarrota. Al realizar el submuestreo con bootstrapping, se creó 4 nuevos ficheros de entrenamiento con 154 observaciones por cada categoría. En total, 616 observaciones.

El muestreo para eliminar las observaciones de la clase mayoritaria fue un muestreo con reemplazamiento. Mientras que, a la clase minoritaria, no se le aplicó ningún tipo de muestreo. Es decir, se tomaron las observaciones en su forma original.

El fichero de validación se mantuvo intacto. Es decir, siguió siendo muy no balanceado.

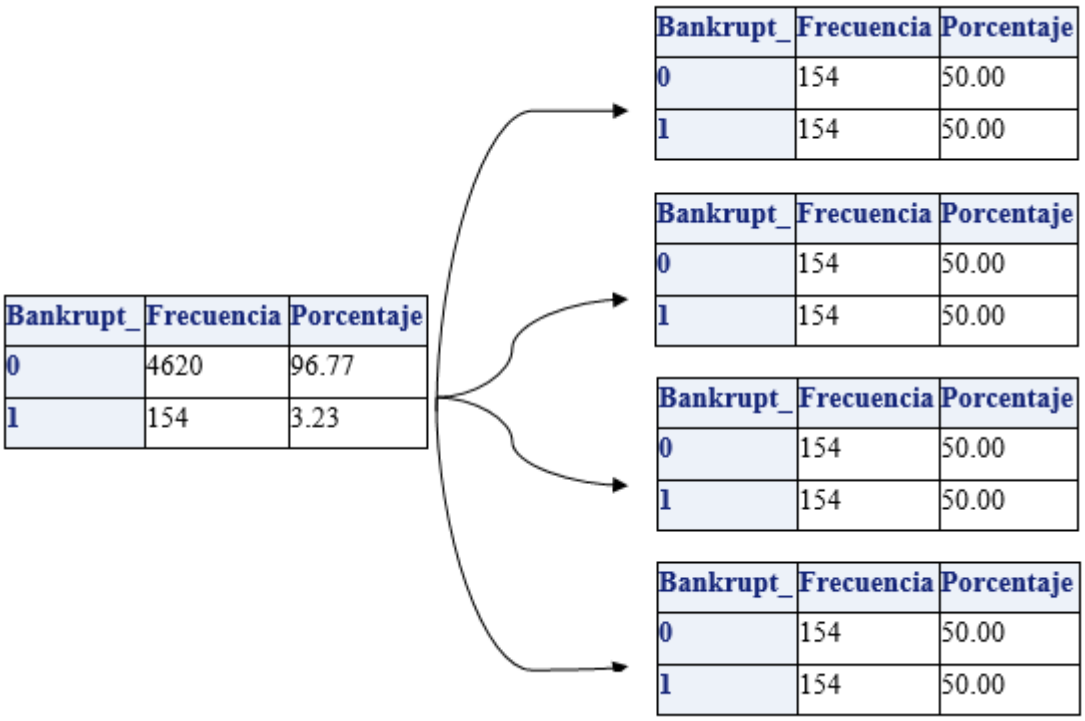


Figura 6. División del conjunto de datos.

6.3. Resultados de regresión logística

Una vez realizado el muestreo, se llevó a cabo las técnicas de clasificación.
Se realizó la regresión logística a partir de los factores retenidos y cuya variable dependiente era Bankrupt_.

Como criterio de selección de variables, se usó el método Stepwise. Los p-valores de entrada y salida fueron los de defecto del software SAS, 0.05 para ambos.

Estimación del modelo

Las 4 muestras originaron 4 modelos cada uno diferente del otro.

Los factores que fueron comunes a todos los modelos eran los ***factores 1, 2, 6***. Estos factores como se mencionó anteriormente representaban el nivel de **rentabilidad**, **deuda** y **riesgo** de la empresa respectivamente.

Resumen de selección de paso a paso

Paso	Efecto Introducido	Efecto Eliminado	DF	Número en	Chi-cuadrado de puntuación	Chi-cuadrado de Wald	Pr > ChiSq
1	Factor1		1	1	89.5269		<.0001
2	Factor6		1	2	22.0393		<.0001
3	Factor4		1	3	16.9194		<.0001
4	Factor12		1	4	8.3095		0.0039
5	Factor2		1	5	8.6507		0.0033
6	Factor5		1	6	8.3076		0.0039
7	Factor15		1	7	6.0220		0.0141
8	Factor10		1	8	5.3072		0.0212

Resumen de selección de paso a paso

Paso	Efecto Introducido	Efecto Eliminado	DF	Número en	Chi-cuadrado de puntuación	Chi-cuadrado de Wald	Pr > ChiSq
1	Factor1		1	1	88.3433		<.0001
2	Factor6		1	2	28.4307		<.0001
3	Factor2		1	3	21.8039		<.0001
4	Factor12		1	4	22.7518		<.0001
5	Factor15		1	5	3.8508		0.0497
6		Factor2	1	4		2.6719	0.1021
7		Factor15	1	3		1.7180	0.1899
8	Factor2		1	4	91.8086		<.0001

Resumen de selección de paso a paso

Paso	Efecto Introducido	Efecto Eliminado	DF	Número en	Chi-cuadrado de puntuación	Chi-cuadrado de Wald	Pr > ChiSq
1	Factor1		1	1	93.5003		<.0001
2	Factor6		1	2	31.9852		<.0001
3	Factor2		1	3	21.8840		<.0001
4	Factor12		1	4	31.1342		<.0001
5	Factor4		1	5	4.1197		0.0424

Resumen de selección de paso a paso

Paso	Efecto Introducido	Efecto Eliminado	DF	Número en	Chi-cuadrado de puntuación	Chi-cuadrado de Wald	Pr > ChiSq
1	Factor1		1	1	101.9593		<.0001
2	Factor6		1	2	17.6891		<.0001
3	Factor2		1	3	11.9569		0.0005
4	Factor5		1	4	17.6273		<.0001

Tablas 7.1, 7.2, 7.3 y 7.4. Variables que entraron al modelo logístico 1, 2, 3 y 4 respectivamente..

Diagnosis

Se analizó la significatividad conjunta del modelo mediante las pruebas de ratio de verosimilitud, puntuación de Fisher y estadístico de Wald.

Probar hipótesis nula global: BETA=0			
Test	Chi-cuadrado	DF	Pr > ChiSq
Ratio de verosim	220.1828	8	<.0001
Puntuación	141.4358	8	<.0001
Wald	70.7778	8	<.0001

Probar hipótesis nula global: BETA=0			
Test	Chi-cuadrado	DF	Pr > ChiSq
Ratio de verosim	203.1007	4	<.0001
Puntuación	117.5124	4	<.0001
Wald	68.8074	4	<.0001

Probar hipótesis nula global: BETA=0			
Test	Chi-cuadrado	DF	Pr > ChiSq
Ratio de verosim	231.9393	5	<.0001
Puntuación	134.3047	5	<.0001
Wald	69.1564	5	<.0001

Probar hipótesis nula global: BETA=0			
Test	Chi-cuadrado	DF	Pr > ChiSq
Ratio de verosim	204.1405	4	<.0001
Puntuación	126.0339	4	<.0001
Wald	67.5683	4	<.0001

Tablas 8.1, 8.2, 8.3 y 8.4. Significatividad conjunta del modelo logístico 1, 2, 3 y 4 respectivamente.

Las hipótesis de estos 4 test es la siguiente:

$$H_0: B_0 = \ldots = B_n = 0 \qquad H_1: \exists i \,/\, B_i \neq 0$$

En las 4 muestras se rechazó la hipótesis nula con un p-valor inferior a 0.0001. De esta forma, se concluyó entonces que al menos, uno de los coeficientes del modelo ajustado era estadísticamente significativo (distinto de cero) y, por lo tanto, el modelo era estadísticamente significativo.

Se analizó también la adecuación del modelo a los datos mediante el Test de Hosmer y Lemeshow.

Test de bondad de ajuste de Hosmer y Lemeshow		
Chi-cuadrado	DF	Pr > ChiSq
2.7011	8	0.9517

Test de bondad de ajuste de Hosmer y Lemeshow		
Chi-cuadrado	DF	Pr > ChiSq
8.7756	8	0.3616

Test de bondad de ajuste de Hosmer y Lemeshow		
Chi-cuadrado	DF	Pr > ChiSq
3.5195	8	0.8977

Test de bondad de ajuste de Hosmer y Lemeshow		
Chi-cuadrado	DF	Pr > ChiSq
7.2855	8	0.5062

Tablas 9.1, 9.2, 9.3 y 9.4. Test de Hosmer Lemeshow del modelo logístico 1, 2, 3 y 4 respectivamente.

Las hipótesis del test de Hosmer y Lemeshow son:

H_0: El modelo se ajusta bien a los datos *H_1: El modelo no se ajusta bien a los datos*

Para las 4 muestras, no se encontraron indicios como para poder rechazar la hipótesis nula. Por lo tanto, los valores esperados eran similares a los observados y se concluyó que el modelo se ajustaba bien a los datos.

La significatividad individual se ha contrastado a través del estadístico de Chi-cuadrado de Wald.

Análisis del estimador de máxima verosimilitud

Parámetro	DF	Estimador	Error estándar	Chi-cuadrado de Wald	Pr > ChiSq	Estimador estandarizado	Exp(Est)
Intercept	1	-2.1599	0.3533	37.3770	<.0001		0.115
Factor1	1	-1.6942	0.4573	13.7261	0.0002	-0.9528	0.184
Factor2	1	0.6360	0.1785	12.6938	0.0004	1.0894	1.889
Factor4	1	-0.5878	0.1693	12.0506	0.0005	-0.3747	0.556
Factor5	1	-0.5625	0.2049	7.5337	0.0061	-0.4017	0.570
Factor6	1	1.6707	0.3112	28.8195	<.0001	1.1488	5.316
Factor10	1	-0.6831	0.3200	4.5557	0.0328	-0.5513	0.505
Factor12	1	-1.8375	0.7766	5.5981	0.0180	-0.5286	0.159
Factor15	1	1.7589	0.5461	10.3749	0.0013	1.0217	5.806

Análisis del estimador de máxima verosimilitud

Parámetro	DF	Estimador	Error estándar	Chi-cuadrado de Wald	Pr > ChiSq	Estimador estandarizado	Exp(Est)
Intercept	1	-2.3314	0.3378	47.6264	<.0001		0.097
Factor1	1	-2.3307	0.3218	52.4706	<.0001	-1.3819	0.097
Factor2	1	0.9049	0.2414	14.0498	0.0002	1.5496	2.472
Factor6	1	1.9252	0.3291	34.2302	<.0001	1.2247	6.856
Factor12	1	-2.0375	0.5932	11.7988	0.0006	-0.5986	0.130

Análisis del estimador de máxima verosimilitud

Parámetro	DF	Estimador	Error estándar	Chi-cuadrado de Wald	Pr > ChiSq	Estimador estandarizado	Exp(Est)
Intercept	1	-2.5501	0.3660	48.5505	<.0001		0.078
Factor1	1	-2.7296	0.3786	51.9840	<.0001	-1.6373	0.065
Factor2	1	1.0332	0.1986	27.0698	<.0001	1.7679	2.810
Factor4	1	-0.3900	0.1944	4.0263	0.0448	-0.2304	0.677
Factor6	1	1.9783	0.3283	36.3124	<.0001	1.3095	7.230
Factor12	1	-2.0063	0.5546	13.0875	0.0003	-0.5927	0.134

Análisis del estimador de máxima verosimilitud

Parámetro	DF	Estimador	Error estándar	Chi-cuadrado de Wald	Pr > ChiSq	Estimador estandarizado	Exp(Est)
Intercept	1	-2.1599	0.3027	50.9147	<.0001		0.115
Factor1	1	-3.3551	0.4413	57.8043	<.0001	-1.8334	0.035
Factor2	1	0.6187	0.1691	13.3831	0.0003	1.0598	1.857
Factor5	1	-0.6972	0.1721	16.4074	<.0001	-0.4714	0.498
Factor6	1	0.9490	0.1978	23.0082	<.0001	0.6450	2.583

Tablas 10.1, 10.2, 10.3 y 10.4. Significatividad individual del modelo logístico 1, 2, 3 y 4 respectivamente.

Las hipótesis del test de significatividad individual son:

$$H_0: \beta_i = 0 \qquad H_1: \beta_i \neq 0$$

Esta prueba contrasta individualmente la significancia de las variables dependientes, que, en este estudio, son los factores. Aunque en la salida también se muestra el contraste para la constante.

Como indican las tablas 12, los p-valores asociados a cada significatividad individual de todas las muestras fueron muy bajos, en consecuencia, todos los factores y pendientes de las muestras fueron significativos, es decir, cada factor y cada pendiente aporta información significativa al modelo.

Validación

Una forma de validación de los modelos es la representación de la curva ROC. Esta área evalúa el poder discriminatorio del modelo entre las dos clases. La curva se puede calcular tanto en el fichero de entrenamiento como en el de validación. En este caso, se tomaron con mayor consideración el trabajar con el fichero de prueba puesto que no se utilizó en la obtención del modelo.

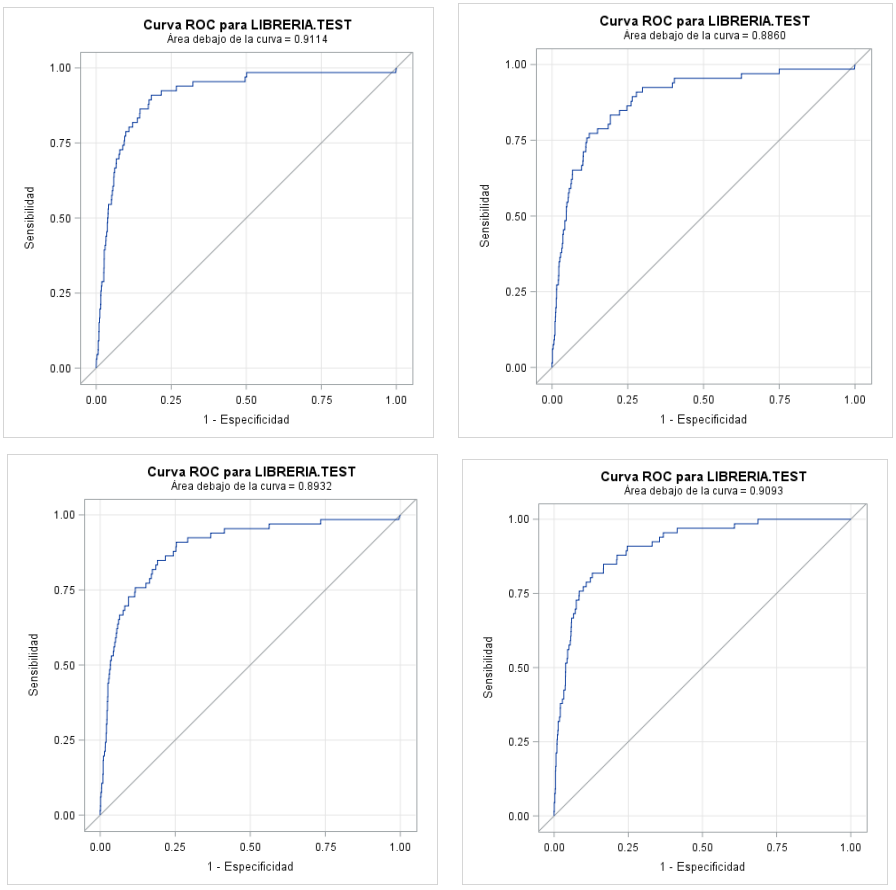


Figura 7. Curva ROC del modelo logístico 1, 2, 3 y 4 respectivamente.

Las 4 muestras presentaron un área bajo la curva superior al 88%. Un porcentaje muy elevado lo que indica un clasificador preciso.

La interpretación aplicada a este estudio se podría traducir como: *"**Seleccionando dos empresas al azar, una en bancarrota y otra no, la probabilidad de que el modelo las clasifique correctamente es en media superior a un 88%**"*

Sin embargo, cuando existe el problema de las *clases no balanceadas*, el calcular la curva ROC no toma demasiado interés puesto que es probable que el alto valor del área bajo la curva se deba a la correcta clasificación de la clase mayoritaria.

Un reducido número de observaciones de la clase minoritaria no aportará demasiada información al modelo y ocasionará que la mayor parte de ellas se clasifiquen erróneamente. Este problema se solventa dándole mayor importancia a la sensibilidad que a la especificidad.

Un método más eficaz que la curva ROC para evaluar la correcta clasificación es a través de la matriz de confusión. Su clasificación puede variar dependiendo del punto de corte que se fije.

El criterio que se fijó fue el de conseguir al menos, un 90% de sensibilidad media entre las 4 muestras de entrenamiento. Los puntos de corte se obtuvieron con la opción *ctable* de SAS y se cambiaron con el procedimiento *plm.*.

El punto de corte óptimo fue el punto 0.38. En este punto, la sensibilidad media que se obtuvo fue de un 90.25%, mientras que la especificidad fue de un 76.45%.

Tabla de clasificación con punto de corte 0.38		
Muestra	**Sensibilidad**	**Especificidad**
Muestra 1	89.6	77.3
Muestra 2	92.2	74.0
Muestra 3	90.9	79.2
Muestra 4	88.3	75.3

Tabla 11. Sensibilidad y especificidad combinada en el punto de corte 0.38.

Fijado el punto de corte, se evaluo la matriz de confusión sobre el fichero de prueba. Los resultados mostrados se calcularon haciendo la media de las 4 muestras. Por lo tanto, la matriz de esta expresada en términos medios.

El fichero de validación, como se mencionó anteriormente, permaneció intacto. Había 1979 empresas que no estaban en bancarrota y 66 que si lo estaban.

Tabla de Bankrupt por I Bankrupt				
		I Bankrupt (A: Bankrupt_)		
		0	1	Total
Bankrupt				
0	Frecuencia media	1510.5	468.5	1979
	Pct fila	76.32	23.68	
1	Frecuencia media	7.5	58.5	66
	Pct fila	11.36	88.64	
Total	Frecuencia media	1518	527	2045

Tabla 12. Matriz de confusión combinada entre las 4 muestras.

Las filas expresan los valores observados mientras que las columnas los valores predichos. Todo ello en términos medios.

Fijar una mínima sensibilidad media del 90% en el conjunto de entrenamiento ocasionó un 88.64% de sensibilidad media en el conjunto de validación**.** Es decir, **en media, el 88.64% de las empresas del fichero de validación que se declararon insolventes fueron detectados por el modelo**

En media, el 76.32% de las empresas del fichero de validación que no cayeron en bancarrota fueron detectados por el modelo de clasificación.

Para finalizar el estudio, se examinó los Odds Ratio estimados junto con su intervalo de confianza.

El Odds Ratio es una medida de asociación entre dos variables cuantificada. En este estudio, los odds ratio cuantifican la relación existente entre los factores y la variable *Bankrupt_*.

Tal medida solo tiene sentido interpretarla si se sabe de antemano que existe una asociación entre ambas variables. Para saberlo, se recurre a los intervalos de confianza.

Una vez verificada la asociación, se interpreta el valor del indicador. Mientras mayor sea el valor estimado del Odds Ratio (superior a 1), mayor incidencia en la probabilidad de entrar en bancarrota, de lo contrario, un valor pequeño (inferior a 1), tiene una menor incidencia.

Debido a que se trabajó con 4 muestras distintas, era de esperar que se obtuvieran resultados distintos.

Estimadores de cocientes de disparidad;			
Efecto	Estimador del punto	Límites de confianza al 95% de Wald	
Factor1	0.184	0.075	0.450
Factor2	1.889	1.331	2.680
Factor4	0.556	0.399	0.774
Factor5	0.570	0.381	0.851
Factor6	5.316	2.889	9.784
Factor10	0.505	0.270	0.946
Factor12	0.159	0.035	0.730
Factor15	5.806	1.991	16.932

Estimadores de cocientes de disparidad;			
Efecto	Estimador del punto	Límites de confianza al 95% de Wald	
Factor1	0.097	0.052	0.183
Factor2	2.472	1.540	3.967
Factor6	6.856	3.598	13.067
Factor12	0.130	0.041	0.417

Estimadores de cocientes de disparidad;			
Efecto	Estimador del punto	Límites de confianza al 95% de Wald	
Factor1	0.065	0.031	0.137
Factor2	2.810	1.904	4.147
Factor4	0.677	0.463	0.991
Factor6	7.230	3.799	13.759
Factor12	0.134	0.045	0.399

Estimadores de cocientes de disparidad;			
Efecto	Estimador del punto	Límites de confianza al 95% de Wald	
Factor1	0.035	0.015	0.083
Factor2	1.857	1.333	2.586
Factor5	0.498	0.355	0.698
Factor6	2.583	1.753	3.807

Tablas 13.1, 13.2, 13.3 y 13.4. Odds ratio del modelo logísticos 1, 2, 3 y 4 respectivamente.

Para todas las muestras, se detectó que existia asociación enter todos los factores que entraron al modelo y la variable *Bankrupt_*. Esto se supo puesto que sus respectivos intervalos de confianza no contenian al valor 1.

Desglosando las muestras, se analizó el riesgo de tener una empresa en bancarrota con los factores más y menos influyentes junto con sus respectivas expresiones matemáticas. Los resultados fueron los siguientes:

Muestra 1

- *Factor 15* (estimador = 5.806): Por cada incremento en una unidad del factor liquidez, la probabilidad de que se produzca bancarrota frente a que no lo haga aumenta en un 580%.

- *Factor 1* (estimador = 0.184): Por cada incremento en una unidad del factor rentabilidad, la probabilidad de que se produzca bancarrota frente a que no lo haga se reduce en un 81.6%.

$$z = 0.184F_1 + 1.889F_2 + 0.556F_4 + 0.570F_5 + 5.316F_6 + 0.505F_{10} + 0.159F_{12} + 5.806F_{15}$$

$$P(empresa\ en\ bancarrota) = \frac{e^z}{1 + e^z}$$

Muestra 2

- *Factor 6* (estimador = 6.856): Por cada incremento en una unidad del factor apalancamiento, la probabilidad de que se produzca bancarrota frente a que no lo haga aumenta en un 685.6%.

- *Factor 1* (estimador = 0.097):): Por cada incremento en una unidad del factor rentabilidad, la probabilidad de que se produzca bancarrota frente a que no lo haga se reduce en un 90.3%.

$$z = 0.097F_1 + 2.472F_2 + 6.856F_6 + 0.130F_{12}$$

$$P(empresa\ en\ bancarrota) = \frac{e^z}{1 + e^z}$$

Muestra 3

- *Factor 6* (estimador = 7.230): Por cada incremento en una unidad del factor apalancamiento, la probabilidad de que se produzca bancarrota frente a que no lo haga aumenta en un 723%.

- *Factor 1 (*estimador = 0.065): Por cada incremento en una unidad del factor rentabilidad, la probabilidad de que se produzca bancarrota frente a que no lo haga se reduce en un 93.5%.

$$z = 0.065F_1 + 2.810F_2 + 0.677F_4 + 7.230F_6 + 0.134F_{12}$$
$$P(empresa\ en\ bancarrota) = \frac{e^z}{1 + e^z}$$

Muestra 4

- *Factor 6* (estimador = 6.856): Por cada incremento en una unidad del factor apalancamiento, la probabilidad de que se produzca bancarrota frente a que no lo haga aumenta en un 685.6%.

- *Factor 1* (estimador = 0.035): Por cada incremento en una unidad del factor rentabilidad, la probabilidad de que se produzca bancarrota frente a que no lo haga se reduce en un 96.5%.

$$z = 0.035F_1 + 1.857F_2 + 0.498F_5 + 2.583F_6$$
$$P(empresa\ en\ bancarrota) = \frac{e^z}{1 + e^z}$$

Todas las muestras coincidieron en que el ***factor1*** **(rentabilidad)** era la variable que más influía negativamente a la detección de empresas en bancarrota. Es decir, **existía una menor probabilidad de detectar empresas insolventes si su valor de rentabilidad era alto**.

Por otra parte, 3 de las 4 muestras indicaron que el ***factor6*** **(apalancamiento)** era la variable que más influía positivamente en la detección de empresas en bancarrota. Es decir, **existía una mayor probabilidad de detectar empresas en bancarrota si su grado de apalancamiento era elevado.**

Otro factor que apareció en el 75% de las muestras fue el ***factor12*** **(solvencia).** Este factor influía negativamente a la detección de empresas en bancarrota. Es decir, **existía una menor probabilidad de detectar empresas insolventes si su solvencia económica era elevada.**

6.4. Resultados de algoritmo KNN

El criterio de elección del mejor parámetro k se decidió en función de la sensibilidad media y la tasa de error media para el fichero de entrenamiento por el método de validación cruzada y para el fichero de validación.

La sensibilidad media es el resultado de la sensibilidad promedio de las 4 muestras. La tasa de error media es también el resultado de la tasa de error promedio de las 4 muestras.

Se buscó aquel parámetro que tenga una mayor sensibilidad media y una menor tasa de error media. Una sensibilidad elevada indica una mejor clasificación de las empresas en bancarrota,. mientras que, una menor tasa de error media significa una mejor clasificación global del modelo.

Dado el intervalo [1,10], se tomaron solo los valores impares como posibles valores de k debido a los posibles empates en las votaciones al hacer la clasificación. Tampoco, se aumentó el valor de k más allá de 10 debido al costo computacional.

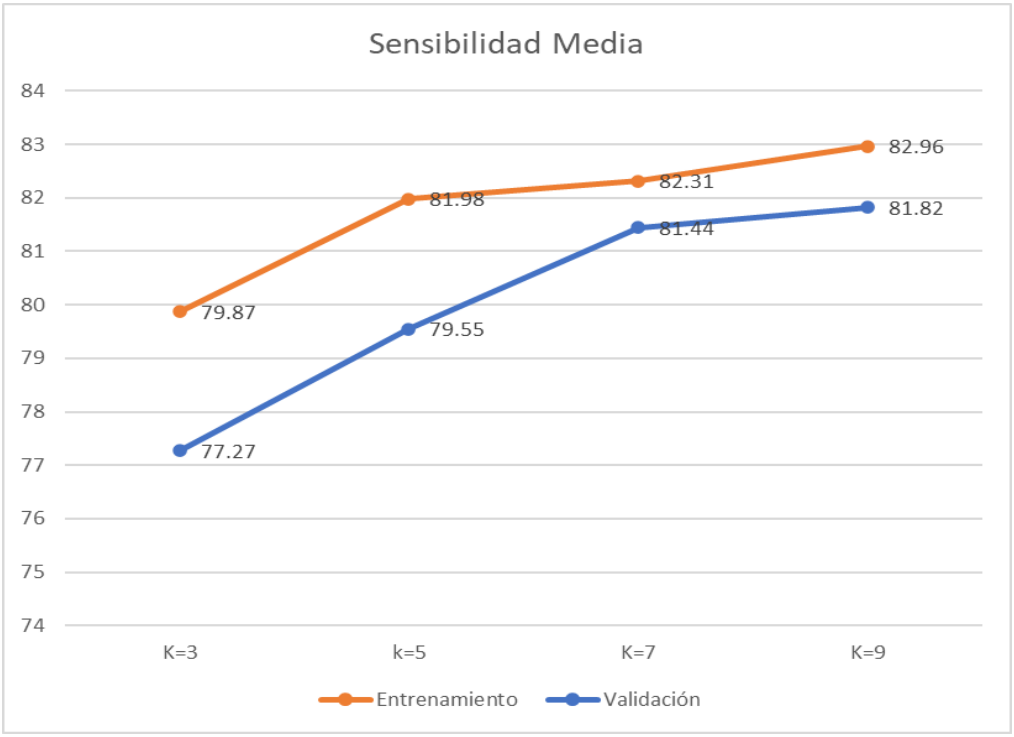


Figura 8. Sensibilidad combinada para el conjunto de entrenamiento y de validación.

La sensibilidad más alta se consiguió en el parametro k=9. En la clasificación del fichero del entrenamiento por validación cruzada el algoritmo knn logró detectar en media, un 82.96% de las empresas en bancarrota, mientras que en el conjunto de validación, se logró detectar en media un 81,82% de las empresas en bancarrota.

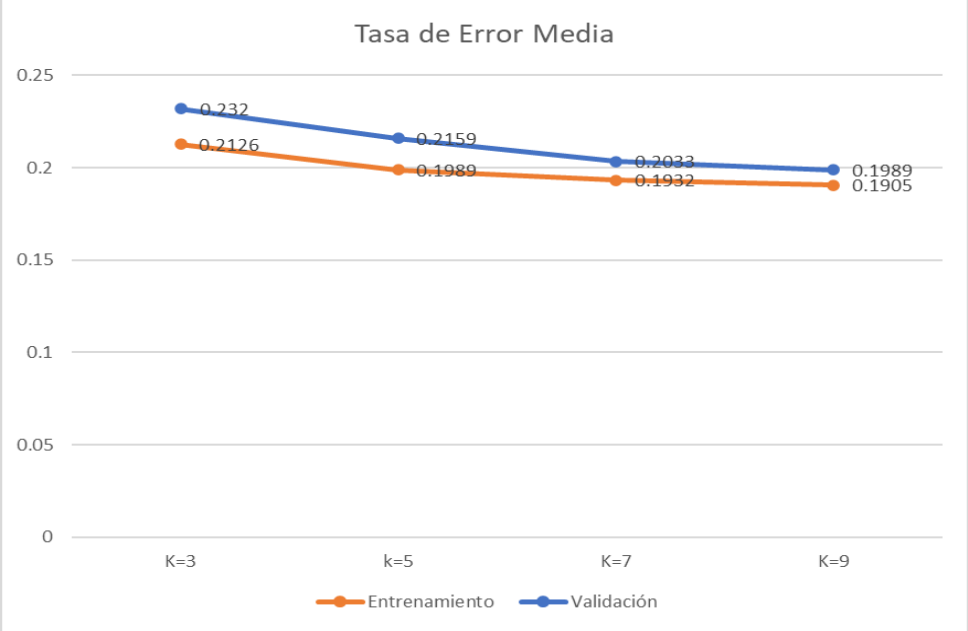


Figura 9. Tasa de error combinada para el conjunto de entrenamiento y validación.

Las tasas son tan solo el cociente entre las predicciones incorrectas y el total de predicciones. En términos medios**, la tasa de error más baja se alcanzó también en el parámetro k=9.**

En la clasificación del fichero de entrenamiento por validación cruzada, el algoritmo logró una tasa de error media de 0.1905, mientras que, en el fichero de validación, una tasa de error media de 0.1989.

Era de esperar que las observaciones del fichero de entrenamiento las clasificará mejor que las de validación dado que estas observaciones construyeron el algoritmo de clasificación.

Una vez fijado el parámetro, es necesario ver si las distancias entre los centroides de las categorías de la variable *Bankrupt_* son significativas o de lo contrario, no lo son. Un centroide es un vector de tamaño n variables que contiene la media de las observaciones de ese conglomerado.

Una mayor distancia indica un mayor poder discriminatorio de la prueba. De esta forma, si existe una mayor cercanía entre los centroides habrá un mayor solapamiento de funciones de densidad de las categorías y habrá una mayor tasa de error.

Estadísticos F, NDF=20, DDF=287 para la distancia cuadrada para Bankrupt_		
De Bankrupt_	0	1
0	0	13.20187
1	13.20187	0

Estadísticos F, NDF=20, DDF=287 para la distancia cuadrada para Bankrupt_		
De Bankrupt_	0	1
0	0	12.48818
1	12.48818	0

Estadísticos F, NDF=20, DDF=287 para la distancia cuadrada para Bankrupt_		
De Bankrupt_	0	1
0	0	14.12669
1	14.12669	0

Estadísticos F, NDF=20, DDF=287 para la distancia cuadrada para Bankrupt_		
De Bankrupt_	0	1
0	0	12.93999
1	12.93999	0

Tablas 14.1, 14.2, 14.3 y 14.4. Estadístico F para la significancia de la distancia cuadrada entre los centroides.

Los grados de libertad del estadístico F fueron, para el numerador 20 grados puesto que la base de datos constaba de 21 variables y la del denominador 287 grados ya que se analizaron 308 observaciones (308-21=287).

El contraste de la F sirvió para evaluar la significatividad de la distancia entre los centroides. Se contrastó si la distancia cuadrática es igual a 0. Las hipótesis del contraste fueron las siguientes:

$$H_0: D^2 = 0 \qquad H_1: D^2 \neq 0$$

El cálculo del p-valor para todas las muestras se calculó de la siguiente forma:

$$P\text{-}valor = P\{F_{20,287} > f\} =< 0.0001$$

Por lo tanto, se comprobó que, para las 4 muestras, existe evidencias estadísticas como para rechazar la distancia cuadrática nula.

Validación

Se tomaron 2 tipos de matriz de confusión para evaluar la capacidad predictiva: Validación cruzada y método de espera.

El resultado de cada matriz de confusión se tomó en base a las 4 muestras de entrenamiento por lo que los resultados se expresaron en términos medios.

Validación cruzada

La sensibilidad media que se alcanzó por este método fue de un 82.95%. Esto quiere decir que, **en media, el 82.95% de las empresas que estaban en bancarrota fueron detectados a través del modelo.**

La especificidad media que se obtuvo fue de un 79.38%. Dicho de otra forma, **en media, el 79.38% de las empresas que no estaban en bancarrota fueron detectados a través del modelo.**

Se obtuvo una tasa de error media para la categoría mayoritaria de 20.62%, mientras que para la minoritaria fue de un 17.05%.

Número de observaciones y porcentaje clasificado en Bankrupt			
De Bankrupt	0	1	Total
0	122.25	31.75	154
	79.38	20.62	100
1	26.25	127.75	154
	17.05	82.95	100
Total	148.5	159.5	308
	48.22	51.79	100
Anteriores	0.5	0.5	

Estimaciones de cuenta de error para Bankrupt			
	0	1	Total
Tasa	0.2062	0.1705	0.1884
Anteriores	0.5	0.5	

Tablas 15.1 y 15.2. Matriz de confusión y tasa de error combinada por el método de validación cruzada.

Método de espera (*hold-out method*)

Este método evalúa la clasificación modelo sobre el conjunto de validación.

Siguiendo este método, se consiguió un 81.82% de sensibilidad media. Dicho de otra forma, **el modelo consiguió detectar en media, el 81.82% de todas las empresas en bancarrota del fichero de validación.**

En cuanto a la especificidad, el modelo logró detecta en media un 78.4% de las empresas. Es decir**, de todas las empresas que tenían solvencia económica del fichero de validación, se clasificó en media, un 21.6% como empresas en bancarrota.**

Se obtuvo una tasa de error media para la categoría de interés del 18.18%, mientras que, para la otra categoría, un 21.6%.

Esta técnica de validación ha sido la que proporcionó los resultados más bajos de los 3 efectuados. Esto es debido a que el método de espera es efectivo cuando se tienen suficientes observaciones, de lo contrario, habrá desajustes en cuanto información.

Número de observaciones y porcentaje clasificado en Bankrupt_			
De Bankrupt_	0	1	Total
0	1551.5	427.5	154
	78.4	21.6	100
1	12	54	154
	18.18	81.82	100
Total	1563.5	481.5	308
	76.46	23.54	100
Anteriores	0.5	0.5	

Estimaciones de cuenta de error para Bankrupt_			
	0	1	Total
Tasa	0.216	0.1818	0.1989
Anteriores	0.5	0.5	

Tablas 16.1 y 16.2. Matriz de confusión y tasa de error combinada entre las 4 muestras por el método de espera.

La técnica de validación cruzada es la más eficaz cuando se tienen escasos registros. Es eficaz también cuando se usan ficheros con grandes registros, pero a mayor cantidad, mayor costo computacional. Cuando esto sucede, es recomendable usar el método de espera.

De las 2 técnicas analizadas, **es más importante tener en cuenta los resultados del método de validación cruzada ya que esta técnica consigue eliminar el sobreajuste**. Por lo cual, ofrece resultados más fiables y precisos.

6.5. Resultados de bosques aleatorios

Primeramente, se determinó los hiperparámetros más convenientes antes de continuar con el algoritmo. **Los valores que se modificaron fueron el número de variables** (VARS_TO_TRY) y el **número de árboles** (MAXTREES).

- Número de variables: Según la regla del pulgar y por redondeo hacia arriba, en cada iteración **se seleccionaron aleatoriamente 5 variables** (VARS_TO_TRY $= \sqrt{20} \approx 4.47 \approx 5$).

- Número de árboles: Para las 4 muestras, inicialmente se ensayó con 300 árboles. La elección del número de árboles se decidió en función de la tasa de clasificación errónea de las observaciones fuera de bolsa (OOB). La figura 10 muestran estas tasas para las 4 muestras.

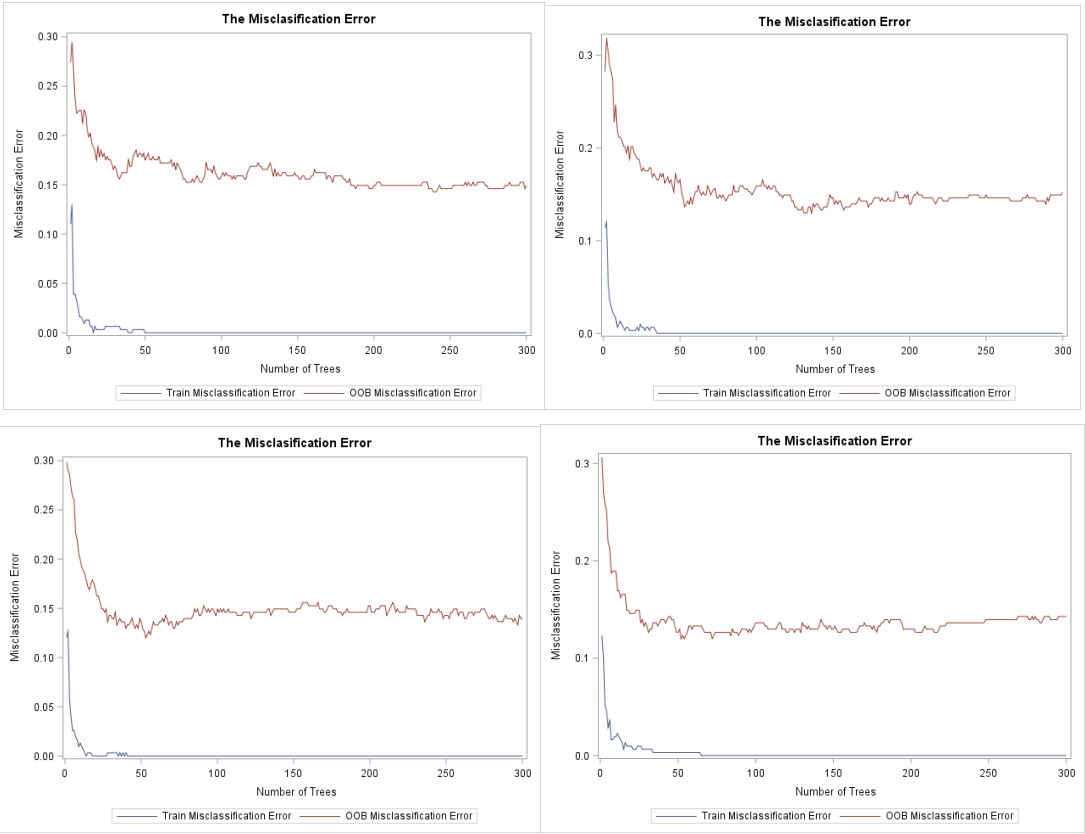


Figura 10.1, 10.2 10.3 y 10.4. Tasa de error en función del número de árboles.

A partir de los 200 árboles, la tasa error de clasificación de las observaciones OOB se mantuvieron estables. Otros casos, como en la muestra 2 y 4 comenzaron a ascender (empeorar).

La tasa de error para el fichero de entrenamiento se mantuvo cercano a 0 a partir de los 50 árboles. El **número final de árboles fijado fue de 200 árboles** y el criterio de división el método de Gini.

Estadísticos de ajuste base	
Estadístico	**Valor**
Average Square Error	0.25
Misclassification Rate	0.5
Log Loss	0.693

Tabla 17. Estadísticos de ajuste base.

Los estadísticos de ajuste base son los estadísticos calculados antes de la realización del algoritmo. La tasa de clasificación errónea tomó valor 0.5 ya que es la proporción de observaciones de empresas que no cayeron en bancarrota (*Bankrupt_* = 0)

Estadísticos de ajuste							
Número de árboles	**Número de hojas**	**Error cuadrado de la media (Train)**	**Error cuadrado de la media (OOB)**	**Tasa de clasificaciones incorrectas (Train)**	**Tasa de clasificaciones incorrectas (OOB)**	**Pérdida Log (Train)**	**Pérdida Log (OOB)**
1	**33**	0.1104	0.274	0.11039	0.274	2.542	6.314
2	**64**	0.0674	0.281	0.12987	0.295	0.516	6.162
.	.	.	.	.	.	.	.
199	**6183**	0.0197	0.12	0	0.146	0.12	0.396
200	**6216**	0.0197	0.12	0	0.149	0.12	0.396
Muestra 2							
1	**32**	0.1136	0.282	0.11364	0.282	2.617	6.499
2	**68**	0.0787	0.306	0.12013	0.319	0.81	6.816
.	.	.	.	.	.	.	.
199	**6354**	0.0202	0.124	0	0.149	0.124	0.404
200	**6390**	0.0202	0.124	0	0.14	0.124	0.404
Muestra 3							
1	**32**	0.1201	0.298	0.12013	0.298	2.766	6.871
2	**65**	0.0666	0.277	0.12662	0.29	0.448	5.952
.	.	.	.	.	.	.	.
199	**5942**	0.0177	0.11	0	0.153	0.113	0.363
200	**5969**	0.0177	0.11	0	0.153	0.113	0.364
Muestra 4							
1	**31**	0.1234	0.306	0.12338	0.306	2.841	7.056
2	**69**	0.0706	0.272	0.10065	0.271	0.656	5.939
.	.	.	.	.	.	.	.
199	**6018**	0.0193	0.114	0	0.13	0.118	0.448
200	**6051**	0.0193	0.114	0	0.13	0.118	0.448

Tabla 18. Estadísticos de ajuste para las 4 muestras.

Por simplicidad, se muestran los estadísticos de ajuste para los 2 primeros y últimos árboles.
En la primera y segunda columna se indica el número de árboles y de hojas (nodos terminales). Se muestra en forma acumulada ya que el resultado es el promedio de los anteriores arboles más el actual.

En las siguientes dos columnas se indica el error cuadrado de la media (*ASE*) para el conjunto de entrenamiento y para las observaciones que no entraron en el modelo (*OOB*) respectivamente.
El ASE es la suma de errores al cuadrado (SCE) dividido entre en número de observaciones.

La quinta y sexta columna indican el ratio de mala clasificación para el conjunto de entrenamiento y para las observaciones que no entraron en el modelo (*OOB)* respectivamente.
El ratio de mala clasificación es la proporción de observaciones mal clasificadas

Las últimas dos columnas indican la función de perdida $L(p_i) = -\log(p_i)$, donde p_i es la probabilidad de clasificación correcta. Por lo tanto, interesa un valor de *L(p)* cercano a 0.

Como es un proceso cumulativo, las estadísticas de ajuste va disminuyendo cada vez que se añaden más arboles de decisión al bosque aleatorio.

Para las 4 muestras, los valores en los árboles 199 y 200 son muy similares. Esto se debe a que, llegado un punto, los errores se estabilizan e incrementar el número de árboles solo ocasionan un mayor costo computacional e incluso en algunos casos empeoran la tasa de mala clasificación como pasó en la muestra 2 y 4.

Importancia de la variable de reducción de pérdida					
Variable	Número de reglas	Gini	Gini OOB	Margen	Margen OOB
Factor1	475	0.107677	0.07851	0.215354	0.18403
Factor2	422	0.078715	0.04735	0.15743	0.12565
Factor10	307	0.036213	0.00736	0.072426	0.04273
Factor6	336	0.037576	0.00531	0.075152	0.04214
Factor3	187	0.016689	-0.00224	0.033379	0.01364

Importancia de la variable de reducción de pérdida					
Variable	Número de reglas	Gini	Gini OOB	Margen	Margen OOB
Factor1	522	0.111016	0.07722	0.222032	0.18709
Factor2	364	0.059121	0.03478	0.118242	0.09354
Factor6	295	0.039128	0.0092	0.078256	0.04666
Factor10	332	0.037216	0.00508	0.074432	0.04215
Factor3	273	0.028651	0.00192	0.057303	0.02996

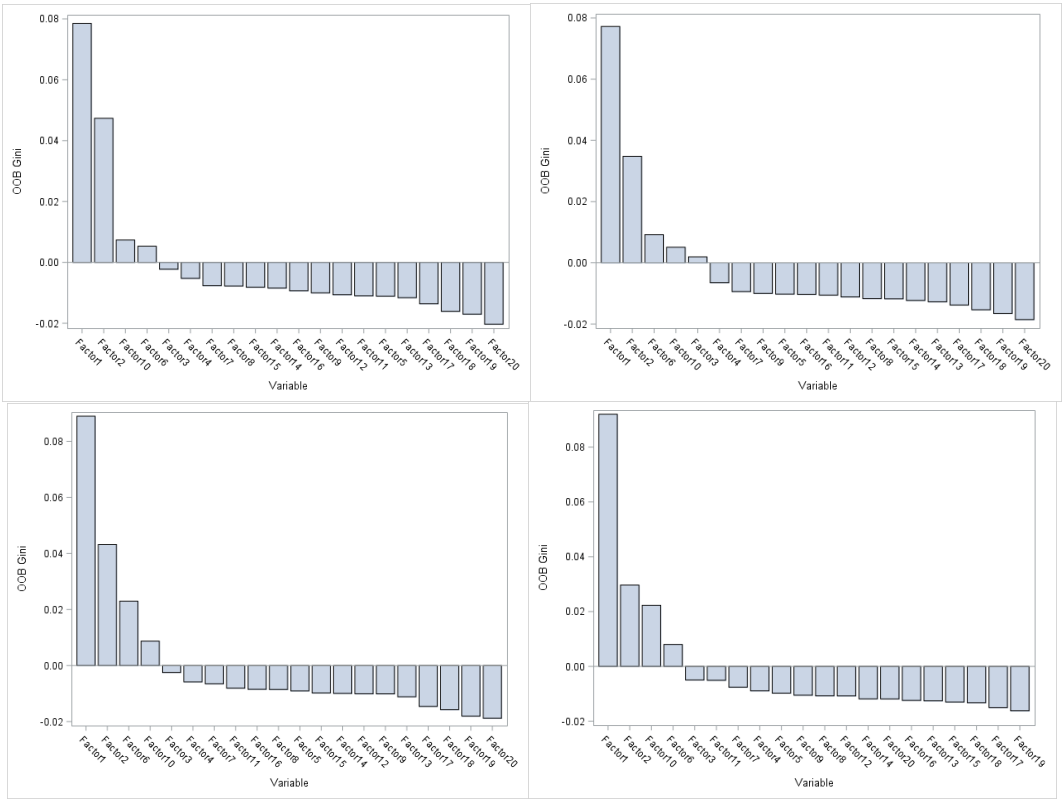
Importancia de la variable de reducción de pérdida					
Variable	Número de reglas	Gini	Gini OOB	Margen	Margen OOB
Factor1	495	0.120531	0.08898	0.241063	0.2074
Factor2	336	0.066613	0.0432	0.133226	0.10942
Factor6	316	0.047164	0.02297	0.094329	0.06862
Factor10	288	0.033058	0.00872	0.066116	0.04126
Factor3	142	0.012649	-0.00246	0.025297	0.01067

Importancia de la variable de reducción de pérdida					
Variable	Número de reglas	Gini	Gini OOB	Margen	Margen OOB
Factor1	485	0.12968	0.09198	0.25936	0.22177
Factor2	357	0.061376	0.02968	0.122752	0.08949
Factor10	322	0.045516	0.02229	0.091032	0.06734
Factor6	296	0.033181	0.00799	0.066362	0.04183
Factor3	186	0.012884	-0.00495	0.025769	0.00717

Tablas 19.1, 19.2, 19.3 y 19.4. Predominancia de las variables en el modelo Random Forest.

Las tablas 22 muestran la importancia de la variable de reducción de perdida. Están ordenadas según el índice de Gini de las observaciones OOB. Por simplicidad, solo se muestran las 5 más notorias.

El resultado es el mismo para las 4 muestras. Las variables más destacadas fueron el **factor 1, 2, 6, 10 y 3.** Estos factores como se mencionó anteriormente representan la **rentabilidad**, las **obligaciones**, el **riesgo** o **apalancamiento**, el **colchón financiero** de la empresa y las **tasas de interés**.



Figuras 20.1, 20.2, 20.3 y 20.4. Importancia de variables según Gini sobre las observaciones OOB.

Para finalizar, se validaron los modelos *Random Forest* a través de la matriz de confusión combinada. Su elaboración siguió la misma estructura que los anteriores. Fue la combinación de las 4 muestras.

El modelo consiguió un 85.98% de sensibilidad media. Es decir, **de todas las empresas insolventes del conjunto de validación, el modelo consiguió detectar en media, el 85.98% de ellas.**

En cuanto a la especificidad, el modelo logró detecta en media un 82.82% de las empresas. Es decir, **de todas las empresas que tenían solvencia económica del conjunto de validación, se clasificó en media, un 17.18% de ellas como empresas en bancarrota.**

Tabla de Bankrupt_ por I_Bankrupt_				
		I_Bankrupt_(Into: Bankrupt_)		
		0	1	Total
Bankrupt_				
0	Frecuencia	1639	340	1979
	Pct fila	82.82	17.18	
1	Frecuencia	9.25	56.75	66
	Pct fila	14.02	85.98	
Total	Frecuencia	1648.25	396.75	2045

Tabla 20. Matriz de confusión combinada.

7. CONCLUSIONES

7.1. Conclusión de objetivo general

1. La detección de empresas insolventes se realizó de forma satisfactoria por regresión logística, algoritmo KNN y bosques aleatorios. Para cada técnica estadísitca, se cumplieron todas las hipotesis prefijadas. Además, **los 3 modelos alcanzaron una sensibilidad media superior al 81%.**

7.2. Conclusión de objetivos específicos

2. El método que más se ajustó a la base de datos fue la regresión logística. Fijando una sensibilidad media mínima del 90% entre las 4 muestras (Punto de corte = 0.38), se alcanzó una sensibilidad media del 88.64% en el fichero de validación.

La inferior precisión de los otros 2 algoritmos se debe en parte a que *PROC DISCRIM* y *PROC HPFOREST* carecen de la opción *ctable* por lo que no se pudo variar el punto de corte distinto a 0.5. Como consecuencia, la sensibilidad no podia variar.

3. De los 20 factores analizados en la regresión logística, los más frecuentes fueron tan solo 4. **Los *factores 1, 2 y 6* aparecieron en todos los modelos construidos**. Mientras que el *factor 12*, apareció en el 75% de ellos.

Los *factores 1, 2, 6* estan representados respectivamente por las variables **rentabilidad**, **deuda** y **riesgo**. El *factor 12* esta representado por la **solvencia** de la empresa.

4. De las 3 variables usadas en el modelo de Fitzpatrick (1932) y Beaver (1936) que eran ratios de liquidez deuda y rotación, **una variable se encontraba en el modelo de regresión logística final** representada en forma de factor. La variable encontrada **fue ratio de deuda** y se representó en el ***factor 2***.

De las 5 variables usadas el el modelo Z-Score de Altman (1968) que eran rentabilidad, liquidez, solvencia, apalancamiento y actividad financiera, **3 se encuentraban en el modelo final** representadas en forma de factor. Las variables **rentabilidad**, **apalancamiento** y **solvencia** fueron representadas por los ***factores 1, 6 y 12*** respectivamente.

8. REFERENCIAS BIBLIOGRÁFICAS

Joffrey L. Leevy et al, 2018. *A survey on addressing high-class imbalance in big data.*

Dorian Pyle, 1999. *Data preparation for data mining.*

Apuntes:

- Juan M. Marín. *Regresión logística.*
- Luis C. Rioja et al. *Regresión Logística: Fundamentos y aplicación a la investigación sociológica.*
- Marcelo Chávez, 2017. *Introducción a los métodos multivariantes.*
- *Luis Bolaños, 2020. Análisis factorial.*
- *William Raseman, 2018. Multivariate Distances: Mahalanobis vs. Euclidean.*
- Zijie Zhu y Mengtian Zhang, 2020. *K-Nearest Neighbors(KNN) Classification with Different Distance Metrics.*

Irini Mavrou, 2015. *Exploratory factor analysis: conceptual and methodological issues*.

N Ayuni e I Sari, 2018. *Analysis of factors that influencing the interest of Bali State Polytechnic's students in entrepreneurship.*

Hair Jr et al, 2010. Multivariate Data Analysis.

Barbara Tabachnick y Linda Fidell, 2001. *Using multivariate statistics.*

M. Collins et al, 2002. *A generalization of principal components analysis to the exponential family.*

Le Agrawal y Yogesh Maheshwari, 2019. *Efficacy of industry factors for corporate default prediction.*

Daniel Ogachi et al, 2020. *Corporate Bankruptcy Prediction Model, a Special Focus on Listed Companies in Kenya.*

Magda Gabriela et al, 2014. *Modelo de predicción de quiebra en micro y pequeñas empresas (MiPyMEs).*

Gergely Fejér-Király, 2015. *Bankruptcy Prediction: A Survey on Evolution, Critiques, and Solutions.*

Rafael Becerra et al, 2020. *Deep Recurrent Convolutional Neural Network for Bankruptcy Prediction: A Case of the Restaurant Industry*.

Amir F. Atiya, 2001. *Bankruptcy Prediction for Credit Risk Using Neural Networks: A Survey and New Results*.

Marcus D. Odom y Ramesh Sharda, 1990. *A Neural Network Model for Bankruptcy Prediction*.

James Gearheart, 2020. *End-to-End Data Science with SAS: A Hands-On Programming Guide.*

Valencia Delfa y Vicente Hernanz, 2005. *Análisis multivariante I.*

9. ANEXO

9.1. Código SAS

```
libname libreria "F:\final";


data datos;
set libreria.datos;
run;

/*Información sobre el conjunto de datos*/
proc contents data=datos;run;

/*información sobre la variable de interés*/
proc freq data=datos;
tables bankrupt_;
run;

/*Correlación de variables*/
proc corr data=datos plots(MAXPOINTS=NONE)=all;
run;



/**************QUITAR VARIABLES CORRELADAS******************/
proc glm data=datos;
     model Bankrupt_= _ROA_C__before_interest_and_depr--
_After_tax_net_Interest_Rate _Continuous_interest_rate__after--
_Debt_ratio__
_Long_term_fund_suitability_rati--_Cash_Current_Liability
_Operating_Funds_to_Liability--_Fixed_Assets_to_Assets
_Equity_to_Long_term_Liability--_No_credit_Interval
_Net_Income_to_Stockholder_s_Equ--_Interest_Coverage_Ratio__Intere
_Equity_to_Liability;
run;quit;

/*Matriz singular*/
data datos_singular;
obs_id=_n_;
set datos;
keep obs_id bankrupt_ _ROA_C__before_interest_and_depr--
_After_tax_net_Interest_Rate _Continuous_interest_rate__after--
_Debt_ratio__
_Long_term_fund_suitability_rati--_Cash_Current_Liability
_Operating_Funds_to_Liability--_Fixed_Assets_to_Assets
_Equity_to_Long_term_Liability--_Current_Liability_to_Current_As
_Net_Income_to_Total_Assets--_No_credit_Interval
_Net_Income_to_Stockholder_s_Equ--_Interest_Coverage_Ratio__Intere
_Equity_to_Liability;
run;



/*Poner nombre de variables en el log*/
proc contents data=datos_singular out=nombres; run;
data _null_;
   set nombres;
   put name;
run;
```

```
/************ESTANDARIZACIÓN DE LOS DATOS*******************/

PROC STANDARD DATA=datos_singular MEAN=0 STD=1 OUT=datos_st;
var _ROA_C__before_interest_and_depr--_Equity_to_Liability;
RUN;

data libreria.datos_st;
set datos_st;
run;



/******************ANÁLISIS FACTORIAL************************/
PROC FACTOR data=datos_st /*COV*/ SIMPLE /*solucion*/ MSA /*KMO y MSO*/
METHOD=PRINCIPAL PRIORS=ONE ROTATE=QUARTIMAX;
  VAR VAR22 _Accounts_Receivable_Turnover _After_tax_Net_Profit_Growth_Rat
_After_tax_net_Interest_Rate _Allocation_rate_per_person
_Average_Collection_Days _Borrowing_dependency _CFO_to_Assets
_Cash_Current_Liability _Cash_Flow_Per_Share
_Cash_Flow_to_Equity _Cash_Flow_to_Liability _Cash_Flow_to_Sales
_Cash_Flow_to_Total_Assets _Cash_Reinvestment__ _Cash_Total_Assets
_Cash_Turnover_Rate _Cash_flow_rate _Contingent_liabilities_Net_wort
_Continuous_Net_Profit_Growth_Ra
_Continuous_interest_rate__after _Current_Asset_Turnover_Rate
_Current_Assets_Total_Assets _Current_Liabilities_Equity
_Current_Liabilities_Liability _Current_Liability_to_Current_As
_Current_Ratio _Debt_ratio__ _Degree_of_Financial_Leverage__D
_Equity_to_Liability _Equity_to_Long_term_Liability
_Fixed_Assets_Turnover_Frequency _Fixed_Assets_to_Assets
_Interest_Coverage_Ratio__Intere _Interest_Expense_Ratio
_Interest_bearing_debt_interest _Inventory_Current_Liability
_Inventory_Turnover_Rate__times_
_Inventory_Working_Capital _Inventory_and_accounts_receivab
_Liability_to_Equity _Long_term_Liability_to_Current
_Long_term_fund_suitability_rati _Net_Income_to_Stockholder_s_Equ
_Net_Income_to_Total_Assets
_Net_Value_Growth_Rate _Net_Value_Per_Share__A_ _Net_Value_Per_Share__B_
_Net_Value_Per_Share__C_ _Net_Worth_Turnover_Rate__times_
_Net_profit_before_tax_Paid_in_c _No_credit_Interval
_Operating_Expense_Rate _Operating_Funds_to_Liability
_Operating_Gross_Margin _Operating_Profit_Growth_Rate
_Operating_Profit_Per_Share__Yua _Operating_Profit_Rate
_Operating_profit_Paid_in_capita _Operating_profit_per_person
_Per_Share_Net_profit_before_tax _Persistent_EPS_in_the_Last_Four
_Pre_tax_net_Interest_Rate _Quick_Asset_Turnover_Rate
_Quick_Assets_Current_Liability _Quick_Assets_Total_Assets _Quick_Ratio
_ROA_A__before_interest_and___af _ROA_B__before_interest_and_depr
_ROA_C__before_interest_and_depr _Realized_Sales_Gross_Margin
_Realized_Sales_Gross_Profit_Gro _Regular_Net_Profit_Growth_Rate
_Research_and_development_expens _Retained_Earnings_to_Total_Asse
_Revenue_per_person _Tax_rate__A_ _Total_Asset_Growth_Rate
_Total_Asset_Return_Growth_Rate _Total_Asset_Turnover
_Total_assets_to_GNP_price _Total_debt_Total_net_worth
_Total_expense_Assets _Total_income_Total_expense _Working_Capital_Equity
_Working_Capital_to_Total_Assets _Working_capitcal_Turnover_Rate;
RUN;
```

```
/*se quita _Revenue_per_person 0.1232*/
PROC FACTOR data=datos_st /*COV*/ SIMPLE /*solucion*/ MSA /*KMO y MSO*/
METHOD=PRINCIPAL PRIORS=ONE ROTATE=QUARTIMAX;
  VAR VAR22 _Accounts_Receivable_Turnover _After_tax_Net_Profit_Growth_Rat
_After_tax_net_Interest_Rate _Allocation_rate_per_person
_Average_Collection_Days _Borrowing_dependency _CFO_to_Assets
_Cash_Current_Liability _Cash_Flow_Per_Share
_Cash_Flow_to_Equity _Cash_Flow_to_Liability _Cash_Flow_to_Sales
_Cash_Flow_to_Total_Assets _Cash_Reinvestment__ _Cash_Total_Assets
_Cash_Turnover_Rate _Cash_flow_rate _Contingent_liabilities_Net_wort
_Continuous_Net_Profit_Growth_Ra
_Continuous_interest_rate__after _Current_Asset_Turnover_Rate
_Current_Assets_Total_Assets _Current_Liabilities_Equity
_Current_Liabilities_Liability _Current_Liability_to_Current_As
_Current_Ratio _Debt_ratio__ _Degree_of_Financial_Leverage__D
_Equity_to_Liability _Equity_to_Long_term_Liability
_Fixed_Assets_Turnover_Frequency _Fixed_Assets_to_Assets
_Interest_Coverage_Ratio__Intere _Interest_Expense_Ratio
_Interest_bearing_debt_interest _Inventory_Current_Liability
_Inventory_Turnover_Rate__times_
_Inventory_Working_Capital _Inventory_and_accounts_receivab
_Liability_to_Equity _Long_term_Liability_to_Current
_Long_term_fund_suitability_rati _Net_Income_to_Stockholder_s_Equ
_Net_Income_to_Total_Assets
_Net_Value_Growth_Rate _Net_Value_Per_Share__A_ _Net_Value_Per_Share__B_
_Net_Value_Per_Share__C_ _Net_Worth_Turnover_Rate__times_
_Net_profit_before_tax_Paid_in_c _No_credit_Interval
_Operating_Expense_Rate _Operating_Funds_to_Liability
_Operating_Gross_Margin _Operating_Profit_Growth_Rate
_Operating_Profit_Per_Share__Yua _Operating_Profit_Rate
_Operating_profit_Paid_in_capita _Operating_profit_per_person
_Per_Share_Net_profit_before_tax _Persistent_EPS_in_the_Last_Four
_Pre_tax_net_Interest_Rate _Quick_Asset_Turnover_Rate
_Quick_Assets_Current_Liability _Quick_Assets_Total_Assets _Quick_Ratio
_ROA_A__before_interest_and___af _ROA_B__before_interest_and_depr
_ROA_C__before_interest_and_depr _Realized_Sales_Gross_Margin
_Realized_Sales_Gross_Profit_Gro _Regular_Net_Profit_Growth_Rate
_Research_and_development_expens _Retained_Earnings_to_Total_Asse
/*_Revenue_per_person*/ _Tax_rate__A_ _Total_Asset_Growth_Rate
_Total_Asset_Return_Growth_Rate _Total_Asset_Turnover
_Total_assets_to_GNP_price _Total_debt_Total_net_worth
_Total_expense_Assets _Total_income_Total_expense _Working_Capital_Equity
_Working_Capital_to_Total_Assets _Working_capitcal_Turnover_Rate;
RUN;


/*se quita VAR22 0.1389*/
PROC FACTOR data=datos_st /*COV*/ SIMPLE /*solucion*/ MSA /*KMO y MSO*/
METHOD=PRINCIPAL PRIORS=ONE ROTATE=QUARTIMAX;
  VAR /*VAR22*/ _Accounts_Receivable_Turnover
_After_tax_Net_Profit_Growth_Rat _After_tax_net_Interest_Rate
_Allocation_rate_per_person _Average_Collection_Days _Borrowing_dependency
_CFO_to_Assets _Cash_Current_Liability _Cash_Flow_Per_Share
_Cash_Flow_to_Equity _Cash_Flow_to_Liability _Cash_Flow_to_Sales
_Cash_Flow_to_Total_Assets _Cash_Reinvestment__ _Cash_Total_Assets
_Cash_Turnover_Rate _Cash_flow_rate _Contingent_liabilities_Net_wort
_Continuous_Net_Profit_Growth_Ra
_Continuous_interest_rate__after _Current_Asset_Turnover_Rate
_Current_Assets_Total_Assets _Current_Liabilities_Equity
_Current_Liabilities_Liability _Current_Liability_to_Current_As
_Current_Ratio _Debt_ratio__ _Degree_of_Financial_Leverage__D
```

```
_Equity_to_Liability _Equity_to_Long_term_Liability
_Fixed_Assets_Turnover_Frequency _Fixed_Assets_to_Assets
_Interest_Coverage_Ratio__Intere _Interest_Expense_Ratio
_Interest_bearing_debt_interest _Inventory_Current_Liability
_Inventory_Turnover_Rate__times_
_Inventory_Working_Capital _Inventory_and_accounts_receivab
_Liability_to_Equity _Long_term_Liability_to_Current
_Long_term_fund_suitability_rati _Net_Income_to_Stockholder_s_Equ
_Net_Income_to_Total_Assets
_Net_Value_Growth_Rate _Net_Value_Per_Share__A_ _Net_Value_Per_Share__B_
_Net_Value_Per_Share__C_ _Net_Worth_Turnover_Rate__times_
_Net_profit_before_tax_Paid_in_c _No_credit_Interval
_Operating_Expense_Rate _Operating_Funds_to_Liability
_Operating_Gross_Margin _Operating_Profit_Growth_Rate
_Operating_Profit_Per_Share__Yua _Operating_Profit_Rate
_Operating_profit_Paid_in_capita _Operating_profit_per_person
_Per_Share_Net_profit_before_tax _Persistent_EPS_in_the_Last_Four
_Pre_tax_net_Interest_Rate _Quick_Asset_Turnover_Rate
_Quick_Assets_Current_Liability _Quick_Assets_Total_Assets _Quick_Ratio
_ROA_A__before_interest_and___af _ROA_B__before_interest_and_depr
_ROA_C__before_interest_and_depr _Realized_Sales_Gross_Margin
_Realized_Sales_Gross_Profit_Gro _Regular_Net_Profit_Growth_Rate
_Research_and_development_expens _Retained_Earnings_to_Total_Asse
/*_Revenue_per_person*/ _Tax_rate__A_ _Total_Asset_Growth_Rate
_Total_Asset_Return_Growth_Rate _Total_Asset_Turnover
_Total_assets_to_GNP_price _Total_debt_Total_net_worth
_Total_expense_Assets _Total_income_Total_expense _Working_Capital_Equity
_Working_Capital_to_Total_Assets _Working_capitcal_Turnover_Rate;
RUN;

/*se quita _Accounts_Receivable_Turnover 0.26*/
PROC FACTOR data=datos_st /*COV*/ SIMPLE /*solucion*/ MSA /*KMO y MSO*/
METHOD=PRINCIPAL PRIORS=ONE ROTATE=QUARTIMAX;
  VAR /*VAR22*/ /*_Accounts_Receivable_Turnover*/
_After_tax_Net_Profit_Growth_Rat _After_tax_net_Interest_Rate
_Allocation_rate_per_person _Average_Collection_Days _Borrowing_dependency
_CFO_to_Assets _Cash_Current_Liability _Cash_Flow_Per_Share
_Cash_Flow_to_Equity _Cash_Flow_to_Liability _Cash_Flow_to_Sales
_Cash_Flow_to_Total_Assets _Cash_Reinvestment__ _Cash_Total_Assets
_Cash_Turnover_Rate _Cash_flow_rate _Contingent_liabilities_Net_wort
_Continuous_Net_Profit_Growth_Ra
_Continuous_interest_rate__after _Current_Asset_Turnover_Rate
_Current_Assets_Total_Assets _Current_Liabilities_Equity
_Current_Liabilities_Liability _Current_Liability_to_Current_As
_Current_Ratio _Debt_ratio__ _Degree_of_Financial_Leverage__D
_Equity_to_Liability _Equity_to_Long_term_Liability
_Fixed_Assets_Turnover_Frequency _Fixed_Assets_to_Assets
_Interest_Coverage_Ratio__Intere _Interest_Expense_Ratio
_Interest_bearing_debt_interest _Inventory_Current_Liability
_Inventory_Turnover_Rate__times_
_Inventory_Working_Capital _Inventory_and_accounts_receivab
_Liability_to_Equity _Long_term_Liability_to_Current
_Long_term_fund_suitability_rati _Net_Income_to_Stockholder_s_Equ
_Net_Income_to_Total_Assets
_Net_Value_Growth_Rate _Net_Value_Per_Share__A_ _Net_Value_Per_Share__B_
_Net_Value_Per_Share__C_ _Net_Worth_Turnover_Rate__times_
_Net_profit_before_tax_Paid_in_c _No_credit_Interval
_Operating_Expense_Rate _Operating_Funds_to_Liability
_Operating_Gross_Margin _Operating_Profit_Growth_Rate
_Operating_Profit_Per_Share__Yua _Operating_Profit_Rate
```

```
_Operating_profit_Paid_in_capita _Operating_profit_per_person
_Per_Share_Net_profit_before_tax _Persistent_EPS_in_the_Last_Four
_Pre_tax_net_Interest_Rate _Quick_Asset_Turnover_Rate
_Quick_Assets_Current_Liability _Quick_Assets_Total_Assets _Quick_Ratio
_ROA_A__before_interest_and___af _ROA_B__before_interest_and_depr
_ROA_C__before_interest_and_depr _Realized_Sales_Gross_Margin
_Realized_Sales_Gross_Profit_Gro _Regular_Net_Profit_Growth_Rate
_Research_and_development_expens _Retained_Earnings_to_Total_Asse
/*_Revenue_per_person*/ _Tax_rate__A_ _Total_Asset_Growth_Rate
_Total_Asset_Return_Growth_Rate _Total_Asset_Turnover
_Total_assets_to_GNP_price _Total_debt_Total_net_worth
_Total_expense_Assets _Total_income_Total_expense _Working_Capital_Equity
_Working_Capital_to_Total_Assets _Working_capitcal_Turnover_Rate;
RUN;

/*se quita _Current_Ratio 0.2736*/
PROC FACTOR data=datos_st /*COV*/ SIMPLE /*solucion*/ MSA /*KMO y MSO*/
METHOD=PRINCIPAL PRIORS=ONE ROTATE=QUARTIMAX;
  VAR /*VAR22*/ /*_Accounts_Receivable_Turnover*/
_After_tax_Net_Profit_Growth_Rat _After_tax_net_Interest_Rate
_Allocation_rate_per_person _Average_Collection_Days _Borrowing_dependency
_CFO_to_Assets _Cash_Current_Liability _Cash_Flow_Per_Share
_Cash_Flow_to_Equity _Cash_Flow_to_Liability _Cash_Flow_to_Sales
_Cash_Flow_to_Total_Assets _Cash_Reinvestment__ _Cash_Total_Assets
_Cash_Turnover_Rate _Cash_flow_rate _Contingent_liabilities_Net_wort
_Continuous_Net_Profit_Growth_Ra
_Continuous_interest_rate__after _Current_Asset_Turnover_Rate
_Current_Assets_Total_Assets _Current_Liabilities_Equity
_Current_Liabilities_Liability _Current_Liability_to_Current_As
/*_Current_Ratio*/ _Debt_ratio__ _Degree_of_Financial_Leverage__D
_Equity_to_Liability _Equity_to_Long_term_Liability
_Fixed_Assets_Turnover_Frequency _Fixed_Assets_to_Assets
_Interest_Coverage_Ratio__Intere _Interest_Expense_Ratio
_Interest_bearing_debt_interest _Inventory_Current_Liability
_Inventory_Turnover_Rate__times_
_Inventory_Working_Capital _Inventory_and_accounts_receivab
_Liability_to_Equity _Long_term_Liability_to_Current
_Long_term_fund_suitability_rati _Net_Income_to_Stockholder_s_Equ
_Net_Income_to_Total_Assets
_Net_Value_Growth_Rate _Net_Value_Per_Share__A_ _Net_Value_Per_Share__B_
_Net_Value_Per_Share__C_ _Net_Worth_Turnover_Rate__times_
_Net_profit_before_tax_Paid_in_c _No_credit_Interval
_Operating_Expense_Rate _Operating_Funds_to_Liability
_Operating_Gross_Margin _Operating_Profit_Growth_Rate
_Operating_Profit_Per_Share__Yua _Operating_Profit_Rate
_Operating_profit_Paid_in_capita _Operating_profit_per_person
_Per_Share_Net_profit_before_tax _Persistent_EPS_in_the_Last_Four
_Pre_tax_net_Interest_Rate _Quick_Asset_Turnover_Rate
_Quick_Assets_Current_Liability _Quick_Assets_Total_Assets _Quick_Ratio
_ROA_A__before_interest_and___af _ROA_B__before_interest_and_depr
_ROA_C__before_interest_and_depr _Realized_Sales_Gross_Margin
_Realized_Sales_Gross_Profit_Gro _Regular_Net_Profit_Growth_Rate
_Research_and_development_expens _Retained_Earnings_to_Total_Asse
/*_Revenue_per_person*/ _Tax_rate__A_ _Total_Asset_Growth_Rate
_Total_Asset_Return_Growth_Rate _Total_Asset_Turnover
_Total_assets_to_GNP_price _Total_debt_Total_net_worth
_Total_expense_Assets _Total_income_Total_expense _Working_Capital_Equity
_Working_Capital_to_Total_Assets _Working_capitcal_Turnover_Rate;
RUN;
```

```
/*se quita _Total_Asset_Return_Growth_Rate 0.2884*/
PROC FACTOR data=datos_st /*COV*/ SIMPLE /*solucion*/ MSA /*KMO y MSO*/
METHOD=PRINCIPAL PRIORS=ONE ROTATE=QUARTIMAX;
  VAR /*VAR22*/ /*_Accounts_Receivable_Turnover*/
_After_tax_Net_Profit_Growth_Rat _After_tax_net_Interest_Rate
_Allocation_rate_per_person _Average_Collection_Days _Borrowing_dependency
_CFO_to_Assets _Cash_Current_Liability _Cash_Flow_Per_Share
_Cash_Flow_to_Equity _Cash_Flow_to_Liability _Cash_Flow_to_Sales
_Cash_Flow_to_Total_Assets _Cash_Reinvestment__ _Cash_Total_Assets
_Cash_Turnover_Rate _Cash_flow_rate _Contingent_liabilities_Net_wort
_Continuous_Net_Profit_Growth_Ra
_Continuous_interest_rate__after _Current_Asset_Turnover_Rate
_Current_Assets_Total_Assets _Current_Liabilities_Equity
_Current_Liabilities_Liability _Current_Liability_to_Current_As
/*_Current_Ratio*/ _Debt_ratio__ _Degree_of_Financial_Leverage__D
_Equity_to_Liability _Equity_to_Long_term_Liability
_Fixed_Assets_Turnover_Frequency _Fixed_Assets_to_Assets
_Interest_Coverage_Ratio__Intere _Interest_Expense_Ratio
_Interest_bearing_debt_interest _Inventory_Current_Liability
_Inventory_Turnover_Rate__times_
_Inventory_Working_Capital _Inventory_and_accounts_receivab
_Liability_to_Equity _Long_term_Liability_to_Current
_Long_term_fund_suitability_rati _Net_Income_to_Stockholder_s_Equ
_Net_Income_to_Total_Assets
_Net_Value_Growth_Rate _Net_Value_Per_Share__A_ _Net_Value_Per_Share__B_
_Net_Value_Per_Share__C_ _Net_Worth_Turnover_Rate__times_
_Net_profit_before_tax_Paid_in_c _No_credit_Interval
_Operating_Expense_Rate _Operating_Funds_to_Liability
_Operating_Gross_Margin _Operating_Profit_Growth_Rate
_Operating_Profit_Per_Share__Yua _Operating_Profit_Rate
_Operating_profit_Paid_in_capita _Operating_profit_per_person
_Per_Share_Net_profit_before_tax _Persistent_EPS_in_the_Last_Four
_Pre_tax_net_Interest_Rate _Quick_Asset_Turnover_Rate
_Quick_Assets_Current_Liability _Quick_Assets_Total_Assets _Quick_Ratio
_ROA_A__before_interest_and___af _ROA_B__before_interest_and_depr
_ROA_C__before_interest_and_depr _Realized_Sales_Gross_Margin
_Realized_Sales_Gross_Profit_Gro _Regular_Net_Profit_Growth_Rate
_Research_and_development_expens _Retained_Earnings_to_Total_Asse
/*_Revenue_per_person*/ _Tax_rate__A_ _Total_Asset_Growth_Rate
/*_Total_Asset_Return_Growth_Rate*/ _Total_Asset_Turnover
_Total_assets_to_GNP_price _Total_debt_Total_net_worth
_Total_expense_Assets _Total_income_Total_expense _Working_Capital_Equity
_Working_Capital_to_Total_Assets _Working_capitcal_Turnover_Rate;
RUN;


/*se quita _Inventory_Working_Capital 0.316*/
PROC FACTOR data=datos_st /*COV*/ SIMPLE /*solucion*/ MSA /*KMO y MSO*/
METHOD=PRINCIPAL PRIORS=ONE ROTATE=QUARTIMAX;
  VAR /*VAR22*/ /*_Accounts_Receivable_Turnover*/
_After_tax_Net_Profit_Growth_Rat _After_tax_net_Interest_Rate
_Allocation_rate_per_person _Average_Collection_Days _Borrowing_dependency
_CFO_to_Assets _Cash_Current_Liability _Cash_Flow_Per_Share
_Cash_Flow_to_Equity _Cash_Flow_to_Liability _Cash_Flow_to_Sales
_Cash_Flow_to_Total_Assets _Cash_Reinvestment__ _Cash_Total_Assets
_Cash_Turnover_Rate _Cash_flow_rate _Contingent_liabilities_Net_wort
_Continuous_Net_Profit_Growth_Ra
_Continuous_interest_rate__after _Current_Asset_Turnover_Rate
_Current_Assets_Total_Assets _Current_Liabilities_Equity
_Current_Liabilities_Liability _Current_Liability_to_Current_As
/*_Current_Ratio*/ _Debt_ratio__ _Degree_of_Financial_Leverage__D
```

```
_Equity_to_Liability _Equity_to_Long_term_Liability
_Fixed_Assets_Turnover_Frequency _Fixed_Assets_to_Assets
_Interest_Coverage_Ratio__Intere _Interest_Expense_Ratio
_Interest_bearing_debt_interest _Inventory_Current_Liability
_Inventory_Turnover_Rate__times_
/*_Inventory_Working_Capital*/ _Inventory_and_accounts_receivab
_Liability_to_Equity _Long_term_Liability_to_Current
_Long_term_fund_suitability_rati _Net_Income_to_Stockholder_s_Equ
_Net_Income_to_Total_Assets
_Net_Value_Growth_Rate _Net_Value_Per_Share__A_ _Net_Value_Per_Share__B_
_Net_Value_Per_Share__C_ _Net_Worth_Turnover_Rate__times_
_Net_profit_before_tax_Paid_in_c _No_credit_Interval
_Operating_Expense_Rate _Operating_Funds_to_Liability
_Operating_Gross_Margin _Operating_Profit_Growth_Rate
_Operating_Profit_Per_Share__Yua _Operating_Profit_Rate
_Operating_profit_Paid_in_capita _Operating_profit_per_person
_Per_Share_Net_profit_before_tax _Persistent_EPS_in_the_Last_Four
_Pre_tax_net_Interest_Rate _Quick_Asset_Turnover_Rate
_Quick_Assets_Current_Liability _Quick_Assets_Total_Assets _Quick_Ratio
_ROA_A__before_interest_and___af _ROA_B__before_interest_and_depr
_ROA_C__before_interest_and_depr _Realized_Sales_Gross_Margin
_Realized_Sales_Gross_Profit_Gro _Regular_Net_Profit_Growth_Rate
_Research_and_development_expens _Retained_Earnings_to_Total_Asse
/*_Revenue_per_person*/ _Tax_rate__A_ _Total_Asset_Growth_Rate
/*_Total_Asset_Return_Growth_Rate*/ _Total_Asset_Turnover
_Total_assets_to_GNP_price _Total_debt_Total_net_worth
_Total_expense_Assets _Total_income_Total_expense _Working_Capital_Equity
_Working_Capital_to_Total_Assets _Working_capitcal_Turnover_Rate;
RUN;

/*se quita _Net_Value_Growth_Rate 0.318*/
PROC FACTOR data=datos_st /*COV*/ SIMPLE /*solucion*/ MSA /*KMO y MSO*/
METHOD=PRINCIPAL PRIORS=ONE ROTATE=QUARTIMAX;
  VAR /*VAR22*/ /*_Accounts_Receivable_Turnover*/
_After_tax_Net_Profit_Growth_Rat _After_tax_net_Interest_Rate
_Allocation_rate_per_person _Average_Collection_Days _Borrowing_dependency
_CFO_to_Assets _Cash_Current_Liability _Cash_Flow_Per_Share
_Cash_Flow_to_Equity _Cash_Flow_to_Liability _Cash_Flow_to_Sales
_Cash_Flow_to_Total_Assets _Cash_Reinvestment__ _Cash_Total_Assets
_Cash_Turnover_Rate _Cash_flow_rate _Contingent_liabilities_Net_wort
_Continuous_Net_Profit_Growth_Ra
_Continuous_interest_rate__after _Current_Asset_Turnover_Rate
_Current_Assets_Total_Assets _Current_Liabilities_Equity
_Current_Liabilities_Liability _Current_Liability_to_Current_As
/*_Current_Ratio*/ _Debt_ratio__ _Degree_of_Financial_Leverage__D
_Equity_to_Liability _Equity_to_Long_term_Liability
_Fixed_Assets_Turnover_Frequency _Fixed_Assets_to_Assets
_Interest_Coverage_Ratio__Intere _Interest_Expense_Ratio
_Interest_bearing_debt_interest _Inventory_Current_Liability
_Inventory_Turnover_Rate__times_
/*_Inventory_Working_Capital*/ _Inventory_and_accounts_receivab
_Liability_to_Equity _Long_term_Liability_to_Current
_Long_term_fund_suitability_rati _Net_Income_to_Stockholder_s_Equ
_Net_Income_to_Total_Assets
/*_Net_Value_Growth_Rate*/ _Net_Value_Per_Share__A_
_Net_Value_Per_Share__B_ _Net_Value_Per_Share__C_
_Net_Worth_Turnover_Rate__times_ _Net_profit_before_tax_Paid_in_c
_No_credit_Interval _Operating_Expense_Rate _Operating_Funds_to_Liability
_Operating_Gross_Margin _Operating_Profit_Growth_Rate
_Operating_Profit_Per_Share__Yua _Operating_Profit_Rate
```

```
_Operating_profit_Paid_in_capita _Operating_profit_per_person
_Per_Share_Net_profit_before_tax _Persistent_EPS_in_the_Last_Four
_Pre_tax_net_Interest_Rate _Quick_Asset_Turnover_Rate
_Quick_Assets_Current_Liability _Quick_Assets_Total_Assets _Quick_Ratio
_ROA_A__before_interest_and___af _ROA_B__before_interest_and_depr
_ROA_C__before_interest_and_depr _Realized_Sales_Gross_Margin
_Realized_Sales_Gross_Profit_Gro _Regular_Net_Profit_Growth_Rate
_Research_and_development_expens _Retained_Earnings_to_Total_Asse
/*_Revenue_per_person*/ _Tax_rate__A_ _Total_Asset_Growth_Rate
/*_Total_Asset_Return_Growth_Rate*/ _Total_Asset_Turnover
_Total_assets_to_GNP_price _Total_debt_Total_net_worth
_Total_expense_Assets _Total_income_Total_expense _Working_Capital_Equity
_Working_Capital_to_Total_Assets _Working_capitcal_Turnover_Rate;
RUN;

/*se quita _Quick_Assets_Current_Liability 0.3105*/
PROC FACTOR data=datos_st /*COV*/ SIMPLE /*solucion*/ MSA /*KMO y MSO*/
METHOD=PRINCIPAL PRIORS=ONE ROTATE=QUARTIMAX;
  VAR /*VAR22*/ /*_Accounts_Receivable_Turnover*/
_After_tax_Net_Profit_Growth_Rat _After_tax_net_Interest_Rate
_Allocation_rate_per_person _Average_Collection_Days _Borrowing_dependency
_CFO_to_Assets _Cash_Current_Liability _Cash_Flow_Per_Share
_Cash_Flow_to_Equity _Cash_Flow_to_Liability _Cash_Flow_to_Sales
_Cash_Flow_to_Total_Assets _Cash_Reinvestment__ _Cash_Total_Assets
_Cash_Turnover_Rate _Cash_flow_rate _Contingent_liabilities_Net_wort
_Continuous_Net_Profit_Growth_Ra
_Continuous_interest_rate__after _Current_Asset_Turnover_Rate
_Current_Assets_Total_Assets _Current_Liabilities_Equity
_Current_Liabilities_Liability _Current_Liability_to_Current_As
/*_Current_Ratio*/ _Debt_ratio__ _Degree_of_Financial_Leverage__D
_Equity_to_Liability _Equity_to_Long_term_Liability
_Fixed_Assets_Turnover_Frequency _Fixed_Assets_to_Assets
_Interest_Coverage_Ratio__Intere _Interest_Expense_Ratio
_Interest_bearing_debt_interest _Inventory_Current_Liability
_Inventory_Turnover_Rate__times_
/*_Inventory_Working_Capital*/ _Inventory_and_accounts_receivab
_Liability_to_Equity _Long_term_Liability_to_Current
_Long_term_fund_suitability_rati _Net_Income_to_Stockholder_s_Equ
_Net_Income_to_Total_Assets
/*_Net_Value_Growth_Rate*/ _Net_Value_Per_Share__A_
_Net_Value_Per_Share__B_ _Net_Value_Per_Share__C_
_Net_Worth_Turnover_Rate__times_ _Net_profit_before_tax_Paid_in_c
_No_credit_Interval _Operating_Expense_Rate _Operating_Funds_to_Liability
_Operating_Gross_Margin _Operating_Profit_Growth_Rate
_Operating_Profit_Per_Share__Yua _Operating_Profit_Rate
_Operating_profit_Paid_in_capita _Operating_profit_per_person
_Per_Share_Net_profit_before_tax _Persistent_EPS_in_the_Last_Four
_Pre_tax_net_Interest_Rate _Quick_Asset_Turnover_Rate
/*_Quick_Assets_Current_Liability*/ _Quick_Assets_Total_Assets _Quick_Ratio
_ROA_A__before_interest_and___af _ROA_B__before_interest_and_depr
_ROA_C__before_interest_and_depr _Realized_Sales_Gross_Margin
_Realized_Sales_Gross_Profit_Gro _Regular_Net_Profit_Growth_Rate
_Research_and_development_expens _Retained_Earnings_to_Total_Asse
/*_Revenue_per_person*/ _Tax_rate__A_ _Total_Asset_Growth_Rate
/*_Total_Asset_Return_Growth_Rate*/ _Total_Asset_Turnover
_Total_assets_to_GNP_price _Total_debt_Total_net_worth
_Total_expense_Assets _Total_income_Total_expense _Working_Capital_Equity
_Working_Capital_to_Total_Assets _Working_capitcal_Turnover_Rate;
RUN;
```

```
/*se quita _Current_Liabilities_Liability 0.330*/
PROC FACTOR data=datos_st /*COV*/ SIMPLE /*solucion*/ MSA /*KMO y MSO*/
METHOD=PRINCIPAL PRIORS=ONE ROTATE=QUARTIMAX;
  VAR /*VAR22*/ /*_Accounts_Receivable_Turnover*/
_After_tax_Net_Profit_Growth_Rat _After_tax_net_Interest_Rate
_Allocation_rate_per_person _Average_Collection_Days _Borrowing_dependency
_CFO_to_Assets _Cash_Current_Liability _Cash_Flow_Per_Share
_Cash_Flow_to_Equity _Cash_Flow_to_Liability _Cash_Flow_to_Sales
_Cash_Flow_to_Total_Assets _Cash_Reinvestment__ _Cash_Total_Assets
_Cash_Turnover_Rate _Cash_flow_rate _Contingent_liabilities_Net_wort
_Continuous_Net_Profit_Growth_Ra
_Continuous_interest_rate__after _Current_Asset_Turnover_Rate
_Current_Assets_Total_Assets _Current_Liabilities_Equity
/*_Current_Liabilities_Liability*/ _Current_Liability_to_Current_As
/*_Current_Ratio*/ _Debt_ratio__ _Degree_of_Financial_Leverage__D
_Equity_to_Liability _Equity_to_Long_term_Liability
_Fixed_Assets_Turnover_Frequency _Fixed_Assets_to_Assets
_Interest_Coverage_Ratio__Intere _Interest_Expense_Ratio
_Interest_bearing_debt_interest _Inventory_Current_Liability
_Inventory_Turnover_Rate__times_
/*_Inventory_Working_Capital*/ _Inventory_and_accounts_receivab
_Liability_to_Equity _Long_term_Liability_to_Current
_Long_term_fund_suitability_rati _Net_Income_to_Stockholder_s_Equ
_Net_Income_to_Total_Assets
/*_Net_Value_Growth_Rate*/ _Net_Value_Per_Share__A_
_Net_Value_Per_Share__B_ _Net_Value_Per_Share__C_
_Net_Worth_Turnover_Rate__times_ _Net_profit_before_tax_Paid_in_c
_No_credit_Interval _Operating_Expense_Rate _Operating_Funds_to_Liability
_Operating_Gross_Margin _Operating_Profit_Growth_Rate
_Operating_Profit_Per_Share__Yua _Operating_Profit_Rate
_Operating_profit_Paid_in_capita _Operating_profit_per_person
_Per_Share_Net_profit_before_tax _Persistent_EPS_in_the_Last_Four
_Pre_tax_net_Interest_Rate _Quick_Asset_Turnover_Rate
/*_Quick_Assets_Current_Liability*/ _Quick_Assets_Total_Assets _Quick_Ratio
_ROA_A__before_interest_and___af _ROA_B__before_interest_and_depr
_ROA_C__before_interest_and_depr _Realized_Sales_Gross_Margin
_Realized_Sales_Gross_Profit_Gro _Regular_Net_Profit_Growth_Rate
_Research_and_development_expens _Retained_Earnings_to_Total_Asse
/*_Revenue_per_person*/ _Tax_rate__A_ _Total_Asset_Growth_Rate
/*_Total_Asset_Return_Growth_Rate*/ _Total_Asset_Turnover
_Total_assets_to_GNP_price _Total_debt_Total_net_worth
_Total_expense_Assets _Total_income_Total_expense _Working_Capital_Equity
_Working_Capital_to_Total_Assets _Working_capitcal_Turnover_Rate;
RUN;

/*se quita _Long_term_Liability_to_Current 0.3224*/
PROC FACTOR data=datos_st /*COV*/ SIMPLE /*solucion*/ MSA /*KMO y MSO*/
METHOD=PRINCIPAL PRIORS=ONE ROTATE=QUARTIMAX;
  VAR /*VAR22*/ /*_Accounts_Receivable_Turnover*/
_After_tax_Net_Profit_Growth_Rat _After_tax_net_Interest_Rate
_Allocation_rate_per_person _Average_Collection_Days _Borrowing_dependency
_CFO_to_Assets _Cash_Current_Liability _Cash_Flow_Per_Share
_Cash_Flow_to_Equity _Cash_Flow_to_Liability _Cash_Flow_to_Sales
_Cash_Flow_to_Total_Assets _Cash_Reinvestment__ _Cash_Total_Assets
_Cash_Turnover_Rate _Cash_flow_rate _Contingent_liabilities_Net_wort
_Continuous_Net_Profit_Growth_Ra
_Continuous_interest_rate__after _Current_Asset_Turnover_Rate
_Current_Assets_Total_Assets _Current_Liabilities_Equity
/*_Current_Liabilities_Liability*/ _Current_Liability_to_Current_As
/*_Current_Ratio*/ _Debt_ratio__ _Degree_of_Financial_Leverage__D
```

```
_Equity_to_Liability _Equity_to_Long_term_Liability
_Fixed_Assets_Turnover_Frequency _Fixed_Assets_to_Assets
_Interest_Coverage_Ratio__Intere _Interest_Expense_Ratio
_Interest_bearing_debt_interest _Inventory_Current_Liability
_Inventory_Turnover_Rate__times_
/*_Inventory_Working_Capital*/ _Inventory_and_accounts_receivab
_Liability_to_Equity /*_Long_term_Liability_to_Current*/
_Long_term_fund_suitability_rati _Net_Income_to_Stockholder_s_Equ
_Net_Income_to_Total_Assets
/*_Net_Value_Growth_Rate*/ _Net_Value_Per_Share__A_
_Net_Value_Per_Share__B_ _Net_Value_Per_Share__C_
_Net_Worth_Turnover_Rate__times_ _Net_profit_before_tax_Paid_in_c
_No_credit_Interval _Operating_Expense_Rate _Operating_Funds_to_Liability
_Operating_Gross_Margin _Operating_Profit_Growth_Rate
_Operating_Profit_Per_Share__Yua _Operating_Profit_Rate
_Operating_profit_Paid_in_capita _Operating_profit_per_person
_Per_Share_Net_profit_before_tax _Persistent_EPS_in_the_Last_Four
_Pre_tax_net_Interest_Rate _Quick_Asset_Turnover_Rate
/*_Quick_Assets_Current_Liability*/ _Quick_Assets_Total_Assets _Quick_Ratio
_ROA_A__before_interest_and___af _ROA_B__before_interest_and_depr
_ROA_C__before_interest_and_depr _Realized_Sales_Gross_Margin
_Realized_Sales_Gross_Profit_Gro _Regular_Net_Profit_Growth_Rate
_Research_and_development_expens _Retained_Earnings_to_Total_Asse
/*_Revenue_per_person*/ _Tax_rate__A_ _Total_Asset_Growth_Rate
/*_Total_Asset_Return_Growth_Rate*/ _Total_Asset_Turnover
_Total_assets_to_GNP_price _Total_debt_Total_net_worth
_Total_expense_Assets _Total_income_Total_expense _Working_Capital_Equity
_Working_Capital_to_Total_Assets _Working_capitcal_Turnover_Rate;
RUN;


/*se quita _Working_capitcal_Turnover_Rate 0.36*/
PROC FACTOR data=datos_st /*COV*/ SIMPLE /*solucion*/ MSA /*KMO y MSO*/
METHOD=PRINCIPAL PRIORS=ONE ROTATE=QUARTIMAX;
  VAR /*VAR22*/ /*_Accounts_Receivable_Turnover*/
_After_tax_Net_Profit_Growth_Rat _After_tax_net_Interest_Rate
_Allocation_rate_per_person _Average_Collection_Days _Borrowing_dependency
_CFO_to_Assets _Cash_Current_Liability _Cash_Flow_Per_Share
_Cash_Flow_to_Equity _Cash_Flow_to_Liability _Cash_Flow_to_Sales
_Cash_Flow_to_Total_Assets _Cash_Reinvestment__ _Cash_Total_Assets
_Cash_Turnover_Rate _Cash_flow_rate _Contingent_liabilities_Net_wort
_Continuous_Net_Profit_Growth_Ra
_Continuous_interest_rate__after _Current_Asset_Turnover_Rate
_Current_Assets_Total_Assets _Current_Liabilities_Equity
/*_Current_Liabilities_Liability*/ _Current_Liability_to_Current_As
/*_Current_Ratio*/ _Debt_ratio__ _Degree_of_Financial_Leverage__D
_Equity_to_Liability _Equity_to_Long_term_Liability
_Fixed_Assets_Turnover_Frequency _Fixed_Assets_to_Assets
_Interest_Coverage_Ratio__Intere _Interest_Expense_Ratio
_Interest_bearing_debt_interest _Inventory_Current_Liability
_Inventory_Turnover_Rate__times_
/*_Inventory_Working_Capital*/ _Inventory_and_accounts_receivab
_Liability_to_Equity /*_Long_term_Liability_to_Current*/
_Long_term_fund_suitability_rati _Net_Income_to_Stockholder_s_Equ
_Net_Income_to_Total_Assets
/*_Net_Value_Growth_Rate*/ _Net_Value_Per_Share__A_
_Net_Value_Per_Share__B_ _Net_Value_Per_Share__C_
_Net_Worth_Turnover_Rate__times_ _Net_profit_before_tax_Paid_in_c
_No_credit_Interval _Operating_Expense_Rate _Operating_Funds_to_Liability
_Operating_Gross_Margin _Operating_Profit_Growth_Rate
_Operating_Profit_Per_Share__Yua _Operating_Profit_Rate
```

```
_Operating_profit_Paid_in_capita _Operating_profit_per_person
_Per_Share_Net_profit_before_tax _Persistent_EPS_in_the_Last_Four
_Pre_tax_net_Interest_Rate _Quick_Asset_Turnover_Rate
/*_Quick_Assets_Current_Liability*/ _Quick_Assets_Total_Assets _Quick_Ratio
_ROA_A__before_interest_and___af _ROA_B__before_interest_and_depr
_ROA_C__before_interest_and_depr _Realized_Sales_Gross_Margin
_Realized_Sales_Gross_Profit_Gro _Regular_Net_Profit_Growth_Rate
_Research_and_development_expens _Retained_Earnings_to_Total_Asse
/*_Revenue_per_person*/ _Tax_rate__A_ _Total_Asset_Growth_Rate
/*_Total_Asset_Return_Growth_Rate*/ _Total_Asset_Turnover
_Total_assets_to_GNP_price _Total_debt_Total_net_worth
_Total_expense_Assets _Total_income_Total_expense _Working_Capital_Equity
_Working_Capital_to_Total_Assets /*_Working_capitcal_Turnover_Rate*/;
RUN;

/*se quita _Cash_Flow_to_Sales 0.13*/
PROC FACTOR data=datos_st /*COV*/ SIMPLE /*solucion*/ MSA /*KMO y MSO*/
METHOD=PRINCIPAL PRIORS=ONE ROTATE=QUARTIMAX;
  VAR /*VAR22*/ /*_Accounts_Receivable_Turnover*/
_After_tax_Net_Profit_Growth_Rat _After_tax_net_Interest_Rate
_Allocation_rate_per_person _Average_Collection_Days _Borrowing_dependency
_CFO_to_Assets _Cash_Current_Liability _Cash_Flow_Per_Share
_Cash_Flow_to_Equity _Cash_Flow_to_Liability /*_Cash_Flow_to_Sales*/
_Cash_Flow_to_Total_Assets _Cash_Reinvestment__ _Cash_Total_Assets
_Cash_Turnover_Rate _Cash_flow_rate _Contingent_liabilities_Net_wort
_Continuous_Net_Profit_Growth_Ra
_Continuous_interest_rate__after _Current_Asset_Turnover_Rate
_Current_Assets_Total_Assets _Current_Liabilities_Equity
/*_Current_Liabilities_Liability*/ _Current_Liability_to_Current_As
/*_Current_Ratio*/ _Debt_ratio__ _Degree_of_Financial_Leverage__D
_Equity_to_Liability _Equity_to_Long_term_Liability
_Fixed_Assets_Turnover_Frequency _Fixed_Assets_to_Assets
_Interest_Coverage_Ratio__Intere _Interest_Expense_Ratio
_Interest_bearing_debt_interest _Inventory_Current_Liability
_Inventory_Turnover_Rate__times_
/*_Inventory_Working_Capital*/ _Inventory_and_accounts_receivab
_Liability_to_Equity /*_Long_term_Liability_to_Current*/
_Long_term_fund_suitability_rati _Net_Income_to_Stockholder_s_Equ
_Net_Income_to_Total_Assets
/*_Net_Value_Growth_Rate*/ _Net_Value_Per_Share__A_
_Net_Value_Per_Share__B_ _Net_Value_Per_Share__C_
_Net_Worth_Turnover_Rate__times_ _Net_profit_before_tax_Paid_in_c
_No_credit_Interval _Operating_Expense_Rate _Operating_Funds_to_Liability
_Operating_Gross_Margin _Operating_Profit_Growth_Rate
_Operating_Profit_Per_Share__Yua _Operating_Profit_Rate
_Operating_profit_Paid_in_capita _Operating_profit_per_person
_Per_Share_Net_profit_before_tax _Persistent_EPS_in_the_Last_Four
_Pre_tax_net_Interest_Rate _Quick_Asset_Turnover_Rate
/*_Quick_Assets_Current_Liability*/ _Quick_Assets_Total_Assets _Quick_Ratio
_ROA_A__before_interest_and___af _ROA_B__before_interest_and_depr
_ROA_C__before_interest_and_depr _Realized_Sales_Gross_Margin
_Realized_Sales_Gross_Profit_Gro _Regular_Net_Profit_Growth_Rate
_Research_and_development_expens _Retained_Earnings_to_Total_Asse
/*_Revenue_per_person*/ _Tax_rate__A_ _Total_Asset_Growth_Rate
/*_Total_Asset_Return_Growth_Rate*/ _Total_Asset_Turnover
_Total_assets_to_GNP_price _Total_debt_Total_net_worth
_Total_expense_Assets _Total_income_Total_expense _Working_Capital_Equity
_Working_Capital_to_Total_Assets /*_Working_capitcal_Turnover_Rate*/;
RUN;
```

```
/*se quita _Inventory_Current_Liability 0.41*/
PROC FACTOR data=datos_st /*COV*/ SIMPLE /*solucion*/ MSA /*KMO y MSO*/
METHOD=PRINCIPAL PRIORS=ONE ROTATE=QUARTIMAX;
  VAR /*VAR22*/ /*_Accounts_Receivable_Turnover*/
_After_tax_Net_Profit_Growth_Rat _After_tax_net_Interest_Rate
_Allocation_rate_per_person _Average_Collection_Days _Borrowing_dependency
_CFO_to_Assets _Cash_Current_Liability _Cash_Flow_Per_Share
_Cash_Flow_to_Equity _Cash_Flow_to_Liability /*_Cash_Flow_to_Sales*/
_Cash_Flow_to_Total_Assets _Cash_Reinvestment__ _Cash_Total_Assets
_Cash_Turnover_Rate _Cash_flow_rate _Contingent_liabilities_Net_wort
_Continuous_Net_Profit_Growth_Ra
_Continuous_interest_rate__after _Current_Asset_Turnover_Rate
_Current_Assets_Total_Assets _Current_Liabilities_Equity
/*_Current_Liabilities_Liability*/ _Current_Liability_to_Current_As
/*_Current_Ratio*/ _Debt_ratio__ _Degree_of_Financial_Leverage__D
_Equity_to_Liability _Equity_to_Long_term_Liability
_Fixed_Assets_Turnover_Frequency _Fixed_Assets_to_Assets
_Interest_Coverage_Ratio__Intere _Interest_Expense_Ratio
_Interest_bearing_debt_interest /*_Inventory_Current_Liability*/
_Inventory_Turnover_Rate__times_
/*_Inventory_Working_Capital*/ _Inventory_and_accounts_receivab
_Liability_to_Equity /*_Long_term_Liability_to_Current*/
_Long_term_fund_suitability_rati _Net_Income_to_Stockholder_s_Equ
_Net_Income_to_Total_Assets
/*_Net_Value_Growth_Rate*/ _Net_Value_Per_Share__A_
_Net_Value_Per_Share__B_ _Net_Value_Per_Share__C_
_Net_Worth_Turnover_Rate__times_ _Net_profit_before_tax_Paid_in_c
_No_credit_Interval _Operating_Expense_Rate _Operating_Funds_to_Liability
_Operating_Gross_Margin _Operating_Profit_Growth_Rate
_Operating_Profit_Per_Share__Yua _Operating_Profit_Rate
_Operating_profit_Paid_in_capita _Operating_profit_per_person
_Per_Share_Net_profit_before_tax _Persistent_EPS_in_the_Last_Four
_Pre_tax_net_Interest_Rate _Quick_Asset_Turnover_Rate
/*_Quick_Assets_Current_Liability*/ _Quick_Assets_Total_Assets _Quick_Ratio
_ROA_A__before_interest_and___af _ROA_B__before_interest_and_depr
_ROA_C__before_interest_and_depr _Realized_Sales_Gross_Margin
_Realized_Sales_Gross_Profit_Gro _Regular_Net_Profit_Growth_Rate
_Research_and_development_expens _Retained_Earnings_to_Total_Asse
/*_Revenue_per_person*/ _Tax_rate__A_ _Total_Asset_Growth_Rate
/*_Total_Asset_Return_Growth_Rate*/ _Total_Asset_Turnover
_Total_assets_to_GNP_price _Total_debt_Total_net_worth
_Total_expense_Assets _Total_income_Total_expense _Working_Capital_Equity
_Working_Capital_to_Total_Assets /*_Working_capitcal_Turnover_Rate*/;
RUN;


/*se quita _Interest_Coverage_Ratio__Intere 0.46*/
PROC FACTOR data=datos_st /*COV*/ SIMPLE /*solucion*/ MSA /*KMO y MSO*/
METHOD=PRINCIPAL PRIORS=ONE ROTATE=QUARTIMAX;
  VAR /*VAR22*/ /*_Accounts_Receivable_Turnover*/
_After_tax_Net_Profit_Growth_Rat _After_tax_net_Interest_Rate
_Allocation_rate_per_person _Average_Collection_Days _Borrowing_dependency
_CFO_to_Assets _Cash_Current_Liability _Cash_Flow_Per_Share
_Cash_Flow_to_Equity _Cash_Flow_to_Liability /*_Cash_Flow_to_Sales*/
_Cash_Flow_to_Total_Assets _Cash_Reinvestment__ _Cash_Total_Assets
_Cash_Turnover_Rate _Cash_flow_rate _Contingent_liabilities_Net_wort
_Continuous_Net_Profit_Growth_Ra
_Continuous_interest_rate__after _Current_Asset_Turnover_Rate
_Current_Assets_Total_Assets _Current_Liabilities_Equity
/*_Current_Liabilities_Liability*/ _Current_Liability_to_Current_As
/*_Current_Ratio*/ _Debt_ratio__ _Degree_of_Financial_Leverage__D
```

```
_Equity_to_Liability _Equity_to_Long_term_Liability
_Fixed_Assets_Turnover_Frequency _Fixed_Assets_to_Assets
/*_Interest_Coverage_Ratio__Intere*/ _Interest_Expense_Ratio
_Interest_bearing_debt_interest /*_Inventory_Current_Liability*/
_Inventory_Turnover_Rate__times_
/*_Inventory_Working_Capital*/ _Inventory_and_accounts_receivab
_Liability_to_Equity /*_Long_term_Liability_to_Current*/
_Long_term_fund_suitability_rati _Net_Income_to_Stockholder_s_Equ
_Net_Income_to_Total_Assets
/*_Net_Value_Growth_Rate*/ _Net_Value_Per_Share__A_
_Net_Value_Per_Share__B_ _Net_Value_Per_Share__C_
_Net_Worth_Turnover_Rate__times_ _Net_profit_before_tax_Paid_in_c
_No_credit_Interval _Operating_Expense_Rate _Operating_Funds_to_Liability
_Operating_Gross_Margin _Operating_Profit_Growth_Rate
_Operating_Profit_Per_Share__Yua _Operating_Profit_Rate
_Operating_profit_Paid_in_capita _Operating_profit_per_person
_Per_Share_Net_profit_before_tax _Persistent_EPS_in_the_Last_Four
_Pre_tax_net_Interest_Rate _Quick_Asset_Turnover_Rate
/*_Quick_Assets_Current_Liability*/ _Quick_Assets_Total_Assets _Quick_Ratio
_ROA_A__before_interest_and___af _ROA_B__before_interest_and_depr
_ROA_C__before_interest_and_depr _Realized_Sales_Gross_Margin
_Realized_Sales_Gross_Profit_Gro _Regular_Net_Profit_Growth_Rate
_Research_and_development_expens _Retained_Earnings_to_Total_Asse
/*_Revenue_per_person*/ _Tax_rate__A_ _Total_Asset_Growth_Rate
/*_Total_Asset_Return_Growth_Rate*/ _Total_Asset_Turnover
_Total_assets_to_GNP_price _Total_debt_Total_net_worth
_Total_expense_Assets _Total_income_Total_expense _Working_Capital_Equity
_Working_Capital_to_Total_Assets /*_Working_capitcal_Turnover_Rate*/;
RUN;

/*se quita _Average_Collection_Days 0.48*/
PROC FACTOR data=datos_st /*COV*/ SIMPLE /*solucion*/ MSA /*KMO y MSO*/
METHOD=PRINCIPAL PRIORS=ONE ROTATE=QUARTIMAX;
  VAR /*VAR22*/ /*_Accounts_Receivable_Turnover*/
_After_tax_Net_Profit_Growth_Rat _After_tax_net_Interest_Rate
_Allocation_rate_per_person /*_Average_Collection_Days*/
_Borrowing_dependency _CFO_to_Assets _Cash_Current_Liability
_Cash_Flow_Per_Share
_Cash_Flow_to_Equity _Cash_Flow_to_Liability /*_Cash_Flow_to_Sales*/
_Cash_Flow_to_Total_Assets _Cash_Reinvestment__ _Cash_Total_Assets
_Cash_Turnover_Rate _Cash_flow_rate _Contingent_liabilities_Net_wort
_Continuous_Net_Profit_Growth_Ra
_Continuous_interest_rate__after _Current_Asset_Turnover_Rate
_Current_Assets_Total_Assets _Current_Liabilities_Equity
/*_Current_Liabilities_Liability*/ _Current_Liability_to_Current_As
/*_Current_Ratio*/ _Debt_ratio__ _Degree_of_Financial_Leverage__D
_Equity_to_Liability _Equity_to_Long_term_Liability
_Fixed_Assets_Turnover_Frequency _Fixed_Assets_to_Assets
/*_Interest_Coverage_Ratio__Intere*/ _Interest_Expense_Ratio
_Interest_bearing_debt_interest /*_Inventory_Current_Liability*/
_Inventory_Turnover_Rate__times_
/*_Inventory_Working_Capital*/ _Inventory_and_accounts_receivab
_Liability_to_Equity /*_Long_term_Liability_to_Current*/
_Long_term_fund_suitability_rati _Net_Income_to_Stockholder_s_Equ
_Net_Income_to_Total_Assets
/*_Net_Value_Growth_Rate*/ _Net_Value_Per_Share__A_
_Net_Value_Per_Share__B_ _Net_Value_Per_Share__C_
_Net_Worth_Turnover_Rate__times_ _Net_profit_before_tax_Paid_in_c
_No_credit_Interval _Operating_Expense_Rate _Operating_Funds_to_Liability
_Operating_Gross_Margin _Operating_Profit_Growth_Rate
_Operating_Profit_Per_Share__Yua _Operating_Profit_Rate
```

```
_Operating_profit_Paid_in_capita _Operating_profit_per_person
_Per_Share_Net_profit_before_tax _Persistent_EPS_in_the_Last_Four
_Pre_tax_net_Interest_Rate _Quick_Asset_Turnover_Rate
/*_Quick_Assets_Current_Liability*/ _Quick_Assets_Total_Assets _Quick_Ratio
_ROA_A__before_interest_and___af _ROA_B__before_interest_and_depr
_ROA_C__before_interest_and_depr _Realized_Sales_Gross_Margin
_Realized_Sales_Gross_Profit_Gro _Regular_Net_Profit_Growth_Rate
_Research_and_development_expens _Retained_Earnings_to_Total_Asse
/*_Revenue_per_person*/ _Tax_rate__A_ _Total_Asset_Growth_Rate
/*_Total_Asset_Return_Growth_Rate*/ _Total_Asset_Turnover
_Total_assets_to_GNP_price _Total_debt_Total_net_worth
_Total_expense_Assets _Total_income_Total_expense _Working_Capital_Equity
_Working_Capital_to_Total_Assets /*_Working_capitcal_Turnover_Rate*/;
RUN;

/*se quita Total_income_Total_expense 0.481*/
PROC FACTOR data=datos_st /*COV*/ SIMPLE /*solucion*/ MSA /*KMO y MSO*/
METHOD=PRINCIPAL PRIORS=ONE ROTATE=QUARTIMAX;
  VAR /*VAR22*/ /*_Accounts_Receivable_Turnover*/
_After_tax_Net_Profit_Growth_Rat _After_tax_net_Interest_Rate
_Allocation_rate_per_person /*_Average_Collection_Days*/
_Borrowing_dependency _CFO_to_Assets _Cash_Current_Liability
_Cash_Flow_Per_Share
_Cash_Flow_to_Equity _Cash_Flow_to_Liability /*_Cash_Flow_to_Sales*/
_Cash_Flow_to_Total_Assets _Cash_Reinvestment__ _Cash_Total_Assets
_Cash_Turnover_Rate _Cash_flow_rate _Contingent_liabilities_Net_wort
_Continuous_Net_Profit_Growth_Ra
_Continuous_interest_rate__after _Current_Asset_Turnover_Rate
_Current_Assets_Total_Assets _Current_Liabilities_Equity
/*_Current_Liabilities_Liability*/ _Current_Liability_to_Current_As
/*_Current_Ratio*/ _Debt_ratio__ _Degree_of_Financial_Leverage__D
_Equity_to_Liability _Equity_to_Long_term_Liability
_Fixed_Assets_Turnover_Frequency _Fixed_Assets_to_Assets
/*_Interest_Coverage_Ratio__Intere*/ _Interest_Expense_Ratio
_Interest_bearing_debt_interest /*_Inventory_Current_Liability*/
_Inventory_Turnover_Rate__times_
/*_Inventory_Working_Capital*/ _Inventory_and_accounts_receivab
_Liability_to_Equity /*_Long_term_Liability_to_Current*/
_Long_term_fund_suitability_rati _Net_Income_to_Stockholder_s_Equ
_Net_Income_to_Total_Assets
/*_Net_Value_Growth_Rate*/ _Net_Value_Per_Share__A_
_Net_Value_Per_Share__B_ _Net_Value_Per_Share__C_
_Net_Worth_Turnover_Rate__times_ _Net_profit_before_tax_Paid_in_c
_No_credit_Interval _Operating_Expense_Rate _Operating_Funds_to_Liability
_Operating_Gross_Margin _Operating_Profit_Growth_Rate
_Operating_Profit_Per_Share__Yua _Operating_Profit_Rate
_Operating_profit_Paid_in_capita _Operating_profit_per_person
_Per_Share_Net_profit_before_tax _Persistent_EPS_in_the_Last_Four
_Pre_tax_net_Interest_Rate _Quick_Asset_Turnover_Rate
/*_Quick_Assets_Current_Liability*/ _Quick_Assets_Total_Assets _Quick_Ratio
_ROA_A__before_interest_and___af _ROA_B__before_interest_and_depr
_ROA_C__before_interest_and_depr _Realized_Sales_Gross_Margin
_Realized_Sales_Gross_Profit_Gro _Regular_Net_Profit_Growth_Rate
_Research_and_development_expens _Retained_Earnings_to_Total_Asse
/*_Revenue_per_person*/ _Tax_rate__A_ _Total_Asset_Growth_Rate
/*_Total_Asset_Return_Growth_Rate*/ _Total_Asset_Turnover
_Total_assets_to_GNP_price _Total_debt_Total_net_worth
_Total_expense_Assets /*_Total_income_Total_expense*/
_Working_Capital_Equity _Working_Capital_to_Total_Assets
/*_Working_capitcal_Turnover_Rate*/;
RUN;
```

```
/***************ÚLTIMO MODELO DE ANÁLISIS FACTORIAL*****************/
/*Número de factores N=20*/
/*Médida de adecuación KMO=0.8030*/
PROC FACTOR data=libreria.datos_st /*COV*/ SIMPLE /*solucion*/ MSA /*KMO y
MSO*/  METHOD=PRINCIPAL PRIORS=ONE ROTATE=QUARTIMAX N=20 /*para ver cuantos
factores nos vamos a quedar*/ REORDER  /*Matriz residual para ver bondad de
ajuste*/
  OUT=TABLA1 OUTSTAT=TABLA2 /*las tablas out sirven para hacer las
representaciones*/;
  VAR /*VAR22*/ /*_Accounts_Receivable_Turnover*/
_After_tax_Net_Profit_Growth_Rat _After_tax_net_Interest_Rate
_Allocation_rate_per_person /*_Average_Collection_Days*/
_Borrowing_dependency _CFO_to_Assets _Cash_Current_Liability
_Cash_Flow_Per_Share
_Cash_Flow_to_Equity _Cash_Flow_to_Liability /*_Cash_Flow_to_Sales*/
_Cash_Flow_to_Total_Assets _Cash_Reinvestment__ _Cash_Total_Assets
_Cash_Turnover_Rate _Cash_flow_rate _Contingent_liabilities_Net_wort
_Continuous_Net_Profit_Growth_Ra
_Continuous_interest_rate__after _Current_Asset_Turnover_Rate
_Current_Assets_Total_Assets _Current_Liabilities_Equity
/*_Current_Liabilities_Liability*/ _Current_Liability_to_Current_As
/*_Current_Ratio*/ _Debt_ratio__ _Degree_of_Financial_Leverage__D
_Equity_to_Liability _Equity_to_Long_term_Liability
_Fixed_Assets_Turnover_Frequency _Fixed_Assets_to_Assets
/*_Interest_Coverage_Ratio__Intere*/ _Interest_Expense_Ratio
_Interest_bearing_debt_interest /*_Inventory_Current_Liability*/
_Inventory_Turnover_Rate__times_
/*_Inventory_Working_Capital*/ _Inventory_and_accounts_receivab
_Liability_to_Equity /*_Long_term_Liability_to_Current*/
_Long_term_fund_suitability_rati _Net_Income_to_Stockholder_s_Equ
_Net_Income_to_Total_Assets
/*_Net_Value_Growth_Rate*/ _Net_Value_Per_Share__A_
_Net_Value_Per_Share__B_ _Net_Value_Per_Share__C_
_Net_Worth_Turnover_Rate__times_ _Net_profit_before_tax_Paid_in_c
_No_credit_Interval _Operating_Expense_Rate _Operating_Funds_to_Liability
_Operating_Gross_Margin _Operating_Profit_Growth_Rate
_Operating_Profit_Per_Share__Yua _Operating_Profit_Rate
_Operating_profit_Paid_in_capita _Operating_profit_per_person
_Per_Share_Net_profit_before_tax _Persistent_EPS_in_the_Last_Four
_Pre_tax_net_Interest_Rate _Quick_Asset_Turnover_Rate
/*_Quick_Assets_Current_Liability*/ _Quick_Assets_Total_Assets _Quick_Ratio
_ROA_A__before_interest_and___af _ROA_B__before_interest_and_depr
_ROA_C__before_interest_and_depr _Realized_Sales_Gross_Margin
_Realized_Sales_Gross_Profit_Gro _Regular_Net_Profit_Growth_Rate
_Research_and_development_expens _Retained_Earnings_to_Total_Asse
/*_Revenue_per_person*/ _Tax_rate__A_ _Total_Asset_Growth_Rate
/*_Total_Asset_Return_Growth_Rate*/ _Total_Asset_Turnover
_Total_assets_to_GNP_price _Total_debt_Total_net_worth
_Total_expense_Assets /*_Total_income_Total_expense*/
_Working_Capital_Equity _Working_Capital_to_Total_Assets
/*_Working_capitcal_Turnover_Rate*/;
RUN;


/***************REPRESENTACIÓN CORRELACIÓN*******************/
data datos_factor_representacion;
set tabla1;
keep  bankrupt_ Factor1--Factor20;
run;
```

```
ods graphics /height=1600 width=1900 imagemap;

proc template;
   define statgraph corrHeatmap;
    dynamic _Title;
     begingraph;
          entrytitle _Title;
        rangeattrmap name='map';
        /* select a series of colors that represent a "diverging"  */
        /* range of values: stronger on the ends, weaker in middle */
        /* Get ideas from http://colorbrewer.org                   */
        range -1 - 1 / rangecolormodel=(cxD8B365 cxF5F5F5 cx5AB4AC);
        endrangeattrmap;
        rangeattrvar var=r attrvar=r attrmap='map';
        layout overlay /
          xaxisopts=(display=(line ticks tickvalues))
          yaxisopts=(display=(line ticks tickvalues));
          heatmapparm x = x y = y colorresponse = r / xbinaxis=false
ybinaxis=false
             colormodel=THREECOLORRAMP name = "heatmap" display=all;
          continuouslegend "heatmap" /
            orient = vertical location = outside title="Correlación de
Pearson";
        endlayout;
      endgraph;
   end;
run;

%prepCorrData(in=datos_factor_representacion,out=datos_fa_r);

proc sgrender data=datos_fa_r template=corrHeatmap;
    dynamic _title="Matriz de Correlaciones después del AF";
run;




/****************DIVIDIR TRAIN-TEST (HOLD-OUT METHOD)****************/
data datos_factor;
set tabla1;
keep obs_id bankrupt_ Factor1--Factor20;
run;

proc sort data = datos_factor out=ordenado;
by bankrupt_;
run;

/*División*/
proc surveyselect data = ordenado out = sample method = srs samprate = 0.7
seed = 291195  outall;
strata bankrupt_;
run;

data input test;
set sample;
drop selected selectionprob SamplingWeight ;
if selected = 1 then output input;
 else output test;
run;
```

```
/*Comprobación*/
ods graphics on;
proc freq data=input;
tables bankrupt_ / nofreq nocum plots=FreqPlot(scale=Percent) out=Freq1Out;
/* No:5280 obs y Si:176 obs*/
run;

proc freq data=test;
tables bankrupt_ / nofreq nocum plots=FreqPlot(scale=Percent) out=Freq1Out;
/* No:1319 obs y Si:44 obs*/
run;



/*Se guarda en la libreria*/
/*************************************************************************/
data libreria.input;
set input;
run;

data libreria.test;
set test;
run;

data input;
set libreria.input;
run;

data test;
set libreria.test;
run;


/*******************TÉCNICAS DE MUESTREO*********************/
proc surveyselect data=input(where=(bankrupt_=1)) out=valor_1  method=srs
n=154 ;
run;


/*muestra 1*/
PROC SURVEYSELECT DATA=input(where=(bankrupt_=0))
METHOD=URS N=154 SEED=25 outhits OUT=muestra1 (DROP=NumberHits ExpectedHits
SamplingWeight);
STRATA bankrupt_;
RUN;

data libreria.muestra1;
set valor_1 muestra1;
run;


/*muestra 2*/
PROC SURVEYSELECT DATA=input(where=(bankrupt_=0))
METHOD=URS N=154 SEED=44 outhits OUT=muestra2 (DROP=NumberHits ExpectedHits
SamplingWeight);
STRATA bankrupt_;
RUN;

data libreria.muestra2;
set valor_1 muestra2;
run;
```

```
/*muestra 3*/
PROC SURVEYSELECT DATA=input(where=(bankrupt_=0))
METHOD=URS N=154 SEED=31 outhits OUT=muestra3 (DROP=NumberHits ExpectedHits
SamplingWeight);
STRATA bankrupt_;
RUN;

data libreria.muestra3;
set valor_1 muestra3;
run;


/*muestra 4*/
PROC SURVEYSELECT DATA=input(where=(bankrupt_=0))
METHOD=URS N=154 SEED=36 outhits OUT=muestra4 (DROP=NumberHits ExpectedHits
SamplingWeight);
STRATA bankrupt_;
RUN;

data libreria.muestra4;
set valor_1 muestra4;
run;



/*******************MUESTRA 1************************/
/*REGRESION LOGISTICA 1*/
ods graphics on;
proc logistic data=libreria.muestra1 plots(only  maxpoints=none)=roc;
  model bankrupt_(event = "1") =factor1--factor20/selection=stepwise expb
rsquare stb lackfit details ctable  outroc=troc ;
  output out=preds predprobs=individuals;
  score data=libreria.test out=valpred outroc=vroc;
  store bankrupt_Model / label='Estudio de bancarrota';
  roc; roccontrast;
run;

/*Cambiar punto de corte para fichero de validación*/
proc plm restore=bankrupt_Model;
   score data=libreria.test out=NewScore predicted / ilink; /* ILINK
devuelve las probabilidades de clasificación */
run;

data ScoreCutpoint;
cutpoint = 0.38;
set NewScore;
if Predicted > cutpoint then
   Pred_bankrupt_ = 1;
else Pred_bankrupt_ = 0;
run;
proc freq data=ScoreCutpoint;
  table bankrupt_*Pred_bankrupt_ / nopercent nocol out=CellCounts;
run;

 data CellCounts;
  set CellCounts;
  Match=0;
  if bankrupt_=Pred_bankrupt_  then Match=1;
run;
```

```
proc means data=CellCounts mean;
   freq count;
   var Match;
run;



/*KNN 1*/
proc discrim data=libreria.muestra1 method=npar k=9 listerr
crosslisterr crossvalidate distance testdata=libreria.test
testout=toscore_out;
class bankrupt_;
var factor1--factor20;
priors proportional;
run;



/*RANDOM FOREST 1*/
proc hpforest data=libreria.muestra1 vars_to_try=5 maxtrees=200;
target bankrupt_/level=binary;
input factor1--factor20 /level=interval;
save file = "E:\salidas portatil\bosque aleatorio\model_fit.bin";
ods output fitstatistics = fitstats(rename=(Ntrees=Trees));
ods output VariableImportance = Variable_Importance;
ods output Baseline = Baseline;
run;

proc hp4score data=libreria.test;
score file= "E:\salidas portatil\bosque aleatorio\model_fit.bin" OUT =
rf_scored
run;


proc freq data=rf_scored;
   table bankrupt_*I_bankrupt_ /nopercent nocol out=CellCounts;
run;

 data CellCounts;
   set CellCounts;
   Match=0;
   if bankrupt_=I_bankrupt_ then Match=1;
run;
proc means data=CellCounts mean;
   freq count;
   var Match;
run;



/*****************************MUESTRA 2*****************************/
/*REGRESION LOGISTICA 2*/
ods graphics on;
proc logistic data=libreria.muestra2 plots(only  maxpoints=none)=roc;
   model bankrupt_(event = "1") =factor1--factor20/selection=stepwise expb
rsquare stb lackfit details ctable outroc=troc ;
   output out=preds predprobs=individuals;
   score data=libreria.test out=valpred outroc=vroc;
   store bankrupt_Model / label='Estudio de bancarrota';
   roc; roccontrast;
run;
```

```
/*Cambiar punto de corte para fichero de validación*/
proc plm restore=bankrupt_Model;
   score data=libreria.test out=NewScore predicted / ilink; /* ILINK
devuelve las probabilidades de clasificación */
run;

data ScoreCutpoint;
cutpoint = 0.38;
set NewScore;
if Predicted > cutpoint then
   Pred_bankrupt_ = 1;
else Pred_bankrupt_ = 0;
run;
proc freq data=ScoreCutpoint;
  table bankrupt_*Pred_bankrupt_ / nopercent nocol out=CellCounts;
run;

 data CellCounts;
  set CellCounts;
  Match=0;
  if bankrupt_=Pred_bankrupt_  then Match=1;
run;
proc means data=CellCounts mean;
  freq count;
  var Match;
run;



/*KNN 2*/
proc discrim data=libreria.muestra2 method=npar k=9 listerr
crosslisterr crossvalidate distance testdata=libreria.test
testout=toscore_out;
class bankrupt_;
var factor1--factor20;
priors proportional;
run;



/*RANDOM FOREST 2*/
proc hpforest data=libreria.muestra2 vars_to_try=5 maxtrees=200;
target bankrupt_/level=binary;
input factor1--factor20 /level=interval;
save file = "E:\salidas portatil\bosque aleatorio\model_fit.bin";
ods output fitstatistics = fitstats(rename=(Ntrees=Trees));
ods output VariableImportance = Variable_Importance;
ods output Baseline = Baseline;
run;

proc hp4score data=libreria.test;
score file= "E:\salidas portatil\bosque aleatorio\model_fit.bin" OUT =
rf_scored
run;


proc freq data=rf_scored;
  table bankrupt_*I_bankrupt_ /nopercent nocol out=CellCounts;
run;
```

```
 data CellCounts;
   set CellCounts;
   Match=0;
   if bankrupt_=I_bankrupt_ then Match=1;
run;
proc means data=CellCounts mean;
   freq count;
   var Match;
run;



/*****************************MUESTRA 3***********************************/
/*REGRESION LOGISTICA 3*/
ods graphics on;
proc logistic data=libreria.muestra3 plots(only  maxpoints=none)=roc;
   model bankrupt_(event = "1") =factor1--factor20/selection=stepwise expb
rsquare stb lackfit details ctable outroc=troc ;
   output out=preds predprobs=individuals;
   score data=libreria.test out=valpred outroc=vroc;
   store bankrupt_Model / label='Estudio de bancarrota';
   roc; roccontrast;
run;

/*Cambiar punto de corte para fichero de validación*/
proc plm restore=bankrupt_Model;
    score data=libreria.test out=NewScore predicted / ilink; /* ILINK
devuelve las probabilidades de clasificación */
run;

data ScoreCutpoint;
cutpoint = 0.38;
set NewScore;
if Predicted > cutpoint then
    Pred_bankrupt_ = 1;
else Pred_bankrupt_ = 0;
run;
proc freq data=ScoreCutpoint;
   table bankrupt_*Pred_bankrupt_ / nopercent nocol out=CellCounts;
run;

 data CellCounts;
   set CellCounts;
   Match=0;
   if bankrupt_=Pred_bankrupt_  then Match=1;
run;
proc means data=CellCounts mean;
   freq count;
   var Match;
run;


/*KNN 3*/
proc discrim data=libreria.muestra3 method=npar k=9 listerr
crosslisterr crossvalidate distance testdata=libreria.test
testout=toscore_out;
class bankrupt_;
var factor1--factor20;
priors proportional;
run;
```

```
/*RANDOM FOREST 3*/
proc hpforest data=libreria.muestra3 vars_to_try=5 maxtrees=200;
target bankrupt_/level=binary;
input factor1--factor20 /level=interval;
save file = "E:\salidas portatil\bosque aleatorio\model_fit.bin";
ods output fitstatistics = fitstats(rename=(Ntrees=Trees));
ods output VariableImportance = Variable_Importance;
ods output Baseline = Baseline;
run;

proc hp4score data=libreria.test;
score file= "E:\salidas portatil\bosque aleatorio\model_fit.bin" OUT =
rf_scored
run;


proc freq data=rf_scored;
   table bankrupt_*I_bankrupt_ /nopercent nocol out=CellCounts;
run;

 data CellCounts;
   set CellCounts;
   Match=0;
   if bankrupt_=I_bankrupt_ then Match=1;
run;
proc means data=CellCounts mean;
   freq count;
   var Match;
run;



/*****************************MUESTRA 4*********************************/
/*REGRESION LOGISTICA 4*/
ods graphics on;
proc logistic data=libreria.muestra4 plots(only  maxpoints=none)=roc;
   model bankrupt_(event = "1") =factor1--factor20/selection=stepwise expb
rsquare stb lackfit details ctable outroc=troc ;
   output out=preds predprobs=individuals;
   score data=libreria.test out=valpred outroc=vroc;
   store bankrupt_Model / label='Estudio de bancarrota';
   roc; roccontrast;
run;

/*Cambiar punto de corte para fichero de validación*/
proc plm restore=bankrupt_Model;
    score data=libreria.test out=NewScore predicted / ilink; /* ILINK
devuelve las probabilidades de clasificación */
run;

data ScoreCutpoint;
cutpoint = 0.38;
set NewScore;
if Predicted > cutpoint then
    Pred_bankrupt_ = 1;
else Pred_bankrupt_ = 0;
run;
proc freq data=ScoreCutpoint;
   table bankrupt_*Pred_bankrupt_ / nopercent nocol out=CellCounts;
run;
```

```
 data CellCounts;
   set CellCounts;
   Match=0;
   if bankrupt_=Pred_bankrupt_  then Match=1;
run;
proc means data=CellCounts mean;
   freq count;
   var Match;
run;



/*KNN 4*/
proc discrim data=libreria.muestra4 method=npar k=9 listerr
crosslisterr crossvalidate distance testdata=libreria.test
testout=toscore_out;
class bankrupt_;
var factor1--factor20;
priors proportional;
run;



/*RANDOM FOREST 4*/
proc hpforest data=libreria.muestra4 vars_to_try=5 maxtrees=200;
target bankrupt_/level=binary;
input factor1--factor20 /level=interval;
save file = "E:\salidas portatil\bosque aleatorio\model_fit.bin";
ods output fitstatistics = fitstats(rename=(Ntrees=Trees));
ods output VariableImportance = Variable_Importance;
ods output Baseline = Baseline;
run;

proc hp4score data=libreria.test;
score file= "E:\salidas portatil\bosque aleatorio\model_fit.bin" OUT =
rf_scored
run;


proc freq data=rf_scored;
   table bankrupt_*I_bankrupt_ /nopercent nocol out=CellCounts;
run;

 data CellCounts;
   set CellCounts;
   Match=0;
   if bankrupt_=I_bankrupt_ then Match=1;
run;
proc means data=CellCounts mean;
   freq count;
   var Match;
run;
```

```
/**********************GRÁFICOS RANDOM FOREST************************/
/*Determinar el número óptimo de árboles*/
data fitstats;
set fitstats;
label Trees = "Number of Trees";
run;

title "The Misclasification Error";
proc sgplot data = fitstats;
 series x=Trees y=MiscAll/legendlabel='Train Misclassification Error';
 series x=Trees y=MiscOOB/legendlabel='OOB Misclassification Error';
 yaxis label='Misclassification Error';
run;


/*Importancia de Gini según observaciones OOB*/
title "Feature Importance Gini";
 proc sgplot data = Variable_Importance;
 vbar Variable /response=GiniOOB  groupdisplay = cluster
categoryorder=respdesc;
 run;



/**********************MACRO PROC CORR*********************************/

%macro prepCorrData(in=,out=);
  /* Run corr matrix for input data, all numeric vars */
  proc corr data=&in. noprint
    pearson
    outp=work._tmpCorr
    vardef=df
  ;
  run;

  /* prep data for heat map */
data &out.;
  keep x y r;
  set work._tmpCorr(where=(_TYPE_="CORR"));
  array v{*} _numeric_;
  x = _NAME_;
  do i = dim(v) to 1 by -1;
    y = vname(v(i));
    r = v(i);
    /* creates a lower triangular matrix */
    if (i<_n_) then
      r=.;
    output;
  end;
run;

proc datasets lib=work nolist nowarn;
  delete _tmpcorr;
quit;
%mend;
```