

Estimación del Rating en base a ciertas métricas

Rubén Vicente Arévalo^{a*}

^aEstadística y Métodos Matemáticos para la Investigación:
Introducción a las Técnicas Estadísticas para la Investigación.

*ruben1.6180@gmail.com

Resumen

La calificación crediticia (o Rating) de una compañía es uno de los principales indicadores de la solvencia y el riesgo crediticio de empresas que los inversores analizan al considerar invertir en una empresa, pudiendo ser un factor decisivo en algunas decisiones de inversión. Sin embargo, no todas las compañías poseen una calificación crediticia, ya que obtenerla implica que las empresas deben pagar a las agencias de calificación crediticia para ser evaluadas. Ante esta situación, el presente trabajo propone el desarrollo de un modelo que permita estimar la calificación crediticia de cualquier empresa a partir de sus estados financieros y sector, aplicando el cálculo de ciertas métricas financieras específicas.

Palabras Clave: Rating, regresión multinomial, ratios financieros, machine learning, random forest.

1. Introducción

Las agencias de calificación crediticias, son entidades cruciales en los mercados financieros, encargadas de evaluar el riesgo crediticio de diversas entidades, como empresas individuales, acciones, bonos gubernamentales, corporativos o municipales, y otros valores financieros. Estas calificaciones sirven para indicar la probabilidad de que el emisor de un valor incumpla con sus obligaciones de pago, clasificando las inversiones en “grado de inversión” (menor riesgo) y “especulativas” (mayor riesgo), lo cual informa a los inversores sobre la calidad y el riesgo asociado con cada inversión. En nuestro caso nos centraremos en empresas individuales, ya que es donde más datos se disponen y el objetivo de nuestro trabajo.

Las tres principales agencias proveedoras de Rating son Moody’s, Standard and Poor’s (S&P) y Fitch, las cuales dominan el mercado, controlando alrededor del 95 % de las calificaciones. Cada una tiene su propia escala de calificación y modificadores (1,2,3, o “+”, “-”) para indicar la fuerza relativa dentro de las categorías de calificación.

Las calificaciones van desde AAA, que indica la calidad más alta, hasta D, que señala incumplimiento (Default). Un aspecto notable es que las calificaciones no suelen expresarse en términos de probabilidades numéricas de incumplimiento, sino más bien a través de descripciones de la fortaleza financiera del deudor. En la Tabla 1 se muestran todas las calificaciones posibles para cada agencia ¹. Nuestra agencia de referencia y a la cual hemos obtenido acceso es la agencia S&P.

El proceso de calificación implica un análisis detallado por parte de analistas y un comité de calificación, que considera la información financiera, pronósticos de gestión, y datos macroeconómicos, entre otros. Nosotros solo dispondremos de la información financiera de las compañías ya que los pronósticos no se suele tener acceso y son datos más subjetivos, al igual que las previsiones macroeconómicas de cada compañía.

En resumen, estas agencias ofrecen una perspectiva esencial sobre el riesgo crediticio asociado a las inversiones, aunque su papel ha sido cuestionado y ahora están sujetas a una mayor supervisión y regulación para proteger la integridad de los mercados financieros.

Otro de los motivos por lo que es de interés obtener el Rating de las compañías es porque, a la hora de valorar una compañía, uno de los métodos que se suele usar es el método de valoración por múltiplos. Este método consiste en la obtención de diferentes múltiplos de compañías comparables a la que se quiere valorar para poder aplicar el múltiplo medio a la misma. Uno de los problemas principales es la obtención de este grupo de compañías comparables las cuales tienes que ser lo más parecidas a tu compañía objetivo. Entre las variables que se buscan se encuentra la calidad crediticia, es decir, el Rating. Sin embargo la mayoría de las compañías no disponen de Rating y por tanto tener una estimación de su Rating sería de gran utilidad en este proceso.

¹Estos son los posibles Ratings para largo plazo, pero también existen calificaciones a corto plazo.

Moody's	S&P	Fitch	Descripción
Aaa	AAA	AAA	Prime
Aa1 Aa2 Aa3	AA+ AA AA-	AA+ AA AA-	Alto grado
A1 A2 A3	A+ A A-	A+ A A-	Grado alto medio
Baa1 Baa2 Baa3	BBB+ BBB BBB-	BBB+ BBB BBB-	Grado bajo medio
Ba1 Ba2 Ba3	BB+ BB BB-	BB+ BB BB-	Sin grado de inversión especulativo
B1 B2 B3	B+ B B-	B+ B B-	Altamente especulativo
Caa1 Caa2 Caa3	CCC+ CCC CCC-	CCC	Extremadamente especulativo
Ca	CC C		Pocas esperanzas de pago
C - -	D	DDD DD D	Impago

Cuadro 1: Calificación de las principales agencias crediticias

Otros escenarios donde es necesario saber el Rating de una compañía es en la toma de decisiones a la hora invertir, ya que uno de los criterios suele ser que la compañía tenga un Rating superior a un cierto umbral.

Analizando la bibliografía disponible observamos que existen estudios similares que se han llevado a cabo para obtener las probabilidad de quiebra de las empresas cotizadas y no cotizadas en el mercado de Estados Unidos [1]. Otros de los intereses que puede tener el conocer el Rating de una compañía es para motivos contables, ya que bajo las NIIF hay muchas ocasiones en las que se necesita estimar la calidad crediticia de una contraparte [2].

Pero, los artículos más recientes y numerosos se han focalizado en la predicción de los Ratings mediante técnicas de Machine Learning (ML) [5], ya que se puede usar una gran cantidad de datos y proporcionan mejores resultados que los métodos estadísticos tradicionales. En este respecto el artículo realizado por Parisa, Ionut y Rupak [4] aplica diferentes métodos, tanto de ML como de Redes Neuronales y árboles de decisión. Otros estudios no solo incluyen métricas financieras si no también la capitalización de la compañía, ya que es una variable que económicamente se podría esperar que tuviera impacto en el riesgo de la compañía [3].

Nuestro análisis ha seguido un diseño muestral ya que hemos obtenido los datos del mercado y dado que tenemos la capacidad de sacar todos los datos disponibles, no hemos aplicado ninguna selección muestral.

Para realizar un análisis financiero detallado y estimar el Rating de una empresa, es crucial tener su información financiera y comprender la naturaleza de cada dato. A continuación, se explica brevemente cada uno de los elementos de información financiera que se han utilizado:

- **Company ID:** Identificador único de la empresa en la base de datos. Sirve para organizar y referenciar de manera única cada empresa dentro del estudio.
- **Rating:** Calificación crediticia de la empresa, otorgada por la agencia S&P. Refleja la capacidad y la probabilidad de la empresa de cumplir con sus obligaciones financieras.
- **Sector:** Clasificación de la industria a la que pertenece la empresa. Proporciona contexto sobre el mercado y la economía en la que opera la empresa.
- **Current Assets:** Activos corrientes. Incluyen recursos económicos que se espera sean convertidos en efectivo o utilizados en la producción de ingresos dentro de un año o ciclo operativo, lo que incluye efectivo, equivalentes de efectivo, cuentas por cobrar e inventarios.

- Cash and equivalents: Efectivo y equivalentes de efectivo. Representa los recursos líquidos disponibles de inmediato para uso de la empresa, incluyendo el efectivo en mano y los activos que pueden convertirse rápidamente en efectivo.
- Receivables: Cuentas por cobrar. Son montos adeudados a la empresa por clientes que han recibido bienes o servicios pero aún no han pagado.
- Total Assets: Activos totales. Representan todos los recursos económicos controlados por la empresa.
- Current Liabilities: Pasivos corrientes. Obligaciones financieras que deben ser pagadas dentro de un año o ciclo operativo, como cuentas por pagar, deudas a corto plazo y otras obligaciones a corto plazo.
- Long Term Debt: Deuda a largo plazo. Compromisos financieros que tienen un vencimiento mayor a un año, reflejando préstamos y otras obligaciones financieras que la empresa debe pagar en el futuro.
- Total Debt: Deuda total. Incluye la suma de todas las obligaciones financieras de la empresa, tanto a corto como a largo plazo.
- Total Equity: Patrimonio neto total. Medida del capital propio de los accionistas en la empresa y representa el valor residual de los activos de la empresa una vez que se han pagado todas sus obligaciones.
- Gastos financieros netos: Gastos financieros netos. Reflejan el costo total de la deuda de una empresa, incluyendo intereses pagados menos los intereses recibidos, ajustados por cualquier otra forma de ingreso o gasto financiero.
- EBITDA: Ganancias antes de intereses, impuestos, depreciación y amortización. Es un indicador de la rentabilidad operativa de una empresa, excluyendo los efectos de decisiones de financiamiento e impuestos, así como el desgaste y obsolescencia de los activos.

Estos datos los hemos obtenido para 10.437 empresas y para los años comprendidos entre 2014 y 2023, obteniendo una total de 82.837 casos. A continuación se muestran tres ejemplos:

Rating	Sector	Current Assets	Cash and equi.	Rec.	Total Assets	Current Liab.	LT Debt	Total Debt	Total Equity	Gastos fin.	EBITDA
AAA	Industrials	131	40	79	948	375	140	298	365	(10)	132
AAA	Industrials	166	10	135	1.993	583	232	530	640	(14)	168
AAA	Health Care	40.948	15.191	8.503	96.321	18.638	9.689	13.216	53.758	(296)	16.813

Cuadro 2: Ejemplos de las métricas en millones de Euros de tres compañías

Cada uno de estos datos proporciona información sobre la empresa, pero el problema es que todos estos ratios no son comparables entre empresas, y por tanto hay que calcular métricas que faciliten la comparación entre ellas. Estas métricas se explicarán en la Sección 2.1 ².

A continuación detallamos la estructura del artículo: En la Sección 2 haremos todos los cálculos y análisis estadísticos previos a la realización de los modelos, los cuales se estudiarán y analizarán en la Sección 3. En esta Sección 3 aplicaremos diferentes modelos tanto de clasificación como de regresión o de Machine Learning (ML), para intentar obtener el mejor modelo posible.

El código, modelos y resultados han sido generados con Python 3 y están disponibles en: https://github.com/RubenPhy/rating_analysis.

²Cabe destacar que solo estamos usando las compañías que disponen de Rating, así que nuestra muestra estará sesgada por estas compañías. Aunque haría falta realizar un análisis más detallado, teniendo un número tan grande de datos y una distribución no-normal, hemos observado mediante el test estadístico de Kruskal-Wallis que es una opción para comparar las medianas entre los grupos, y es el equivalente no paramétrico de la ANOVA de un solo factor y se utiliza para determinar si hay diferencias estadísticamente significativas entre tres o más grupos independientes. Observamos que la media de los ratios financieros calculados en la Sección 2.1 son estadísticamente diferentes. Lo cual puede deberse a varios factores pero el principal es que suelen ser empresas más pequeñas.

2. Análisis de los datos

2.1. Obtención de los datos

Antes de poder hacer los modelos que estimen el Rating de las compañías tenemos que obtener una muestra lo mejor posibles. Para ello, en esta sección preparemos y analizaremos los datos que más tarde usaremos en los modelos. Para empezar, de los 82.837 compañías mencionadas anteriormente, una vez eliminamos todas las que tienen, datos vacíos, o errores en cualquiera de los campos, obtenemos un total de 52.592 casos.

Respecto al tipo de variables de nuestros datos, la mayoría de las variables son numéricas, salvo el Rating y el sector. A continuación explicamos con más detalle cada una de ellas:

1. Métricas financieras: estas variables son todas variables numéricas.
2. Sector: es una variables categórica nominal, es decir tiene dos o más categorías pero no tienen un orden intrínseco.
3. Rating: es la variable dependiente y es una variable categórica ordinal, es decir tienen un orden intrínseco.

El siguiente paso es calcular las métricas que utilizaremos para nuestros modelos. Estas métricas estandarizan los datos entre compañías y son las siguientes:

- Net Debt to EBITDA: Este ratio compara la deuda neta (deuda total menos efectivo y equivalentes de efectivo) con el EBITDA (Ganancias antes de intereses, impuestos, depreciación y amortización). Mide la capacidad de una empresa para pagar su deuda utilizando sus ganancias operativas antes de gastos no monetarios e intereses. Por tanto, un valor menor indica que la empresa puede pagar su deuda más rápidamente.

$$\text{Net Debt to EBITDA} = \frac{\text{Total Debt} - \text{Cash and equivalents}}{\text{EBITDA}}$$

El problema de este ratio es que:

- El EBITDA puede estar cerca de 0, lo cual hace que el ratios esten entorno $+\infty$, $-\infty$
- La Deuda neta puede ser 0 o negativa ya que puedes tener más caja que deuda.

Por tanto, para los 4 casos posibles, tenemos que:

- $EBITDA > 0$ y $Net Debt > 0$: Hemos limitando este ratio a 15.
 - $EBITDA > 0$ y $Net Debt < 0$: Hemos usando un 0 ya que esté sería el mejor escenarios y por tanto le asignamos el mejor valor posible.
 - $EBITDA < 0$ y $Net Debt < 0$: Le hemos asignado el mayor valor posible.
 - $EBITDA < 0$ y $Net Debt > 0$: Este es un caso raro, así que eliminaremos estos casos.
- Debt to Assets: Muestra la proporción de activos de la empresa que es financiada por deuda. Un ratio más alto indica una mayor proporción de financiamiento mediante deuda, lo que puede sugerir un mayor riesgo financiero.

$$\text{Debt to Assets} = \frac{\text{Total Debt}}{\text{Total Equity} + \text{Total Debt}}$$

- Current Ratio: Compara los activos corrientes con los pasivos corrientes. Este ratio mide la capacidad de una empresa para pagar sus obligaciones a corto plazo con sus activos a corto plazo. Un ratio superior a 1 sugiere que la empresa tiene suficientes activos corrientes para cubrir sus pasivos a corto plazo.

$$\text{Current Ratio} = \frac{\text{Current Assets}}{\text{Current Liabilities}}$$

- Quick Ratio: Similar al Current Ratio, pero excluye inventarios de los activos corrientes. Esto se debe a que los inventarios a menudo son menos líquidos. Este ratio proporciona una medida más estricta de la liquidez a corto plazo de una empresa.

$$\text{Quick Ratio} = \frac{\text{Cash and equivalents} + \text{Receivables}}{\text{Current Liabilities}}$$

- **Cash Ratio:** Aún más estricto que el Quick Ratio, este ratio compara solo el efectivo y equivalentes de efectivo con los pasivos corrientes. Ofrece una visión de la capacidad de la empresa para pagar sus deudas a corto plazo utilizando únicamente su efectivo disponible y activos fácilmente convertibles en efectivo.

$$\text{Cash ratio} = \frac{\text{Cash and equivalents}}{\text{Current Liabilities}}$$

- **Long Term Debt to Equity:** Este ratio compara la deuda a largo plazo con el patrimonio neto de la empresa. Proporciona una indicación de cómo la empresa está financiando sus operaciones (a través de deuda versus financiamiento de los accionistas) y el nivel de riesgo financiero a largo plazo.

$$\text{Long Term Debt to Equity} = \frac{\text{Long Term Debt}}{\text{Total Equity}}$$

- **Debt to Equity:** Muestra la proporción entre el total de deuda (a corto y largo plazo) y el patrimonio neto. Este ratio ayuda a entender cómo una empresa se financia y puede indicar si está utilizando demasiada deuda.

$$\text{Debt to Equity} = \frac{\text{Total Debt}}{\text{Total Equity}}$$

- **Debt Ratio:** Es la proporción del total de activos financiados por deuda. Un ratio mayor sugiere una mayor dependencia de la financiación con deuda, lo cual puede ser un indicador de riesgo financiero.

$$\text{Debt ratio} = \frac{\text{Total Assets}}{\text{Total Debt}}$$

- **Financial Leverage:** También conocido como el ratio de apalancamiento financiero, este ratio mide el efecto del endeudamiento en la rentabilidad de la empresa. Un mayor apalancamiento significa que la empresa está utilizando más deuda para financiar sus operaciones, lo cual puede aumentar la rentabilidad pero también el riesgo.

$$\text{Financial Leverage} = \frac{\text{Total Assets}}{\text{Total Equity}}$$

Una vez tenemos las métricas, eliminaremos los datos extremos ³ ya que nos distorsionaran la muestra y no nos permitirá entrenar bien los modelos. Una vez tenemos todos los datos podemos empezar a analizar los. Lo primero que haremos será ver la distribución de cada métrica por Rating. Como se puede ver en la Figura 1, si que hay un diferencia de distribuciones para los diferentes Ratings.

Mencionar que para un primer análisis hemos agrupado los Rating por una categoría más amplia que la oficial, Tabla 1. Empleando la misma aproximación, también hemos podido observar que esta misma diferencia de distribuciones en las métricas se puede observar para cada Sector, indicándonos que es una variable relevante a tener en cuenta.

Para poder emplear los diferentes modelos de regresión y clasificación, nuestras dos variables categóricas las hemos transformado en variables numéricas asignado un numero a cada sector y para el caso del Rating asignado un 0 a la categoría AAA, un 1 a la AA+ y así sucesivamente.

Otro dato importante a observar es como se distribuyen las compañías por Rating. Ya que como se observa en la Figura 2 siguen una distribución normal, con la mayoría de empresas clasificadas entorno al Ratings BBB. Esto hace que tengamos que hacer un *oversample* de los datos menos observados para que nuestro modelo no de más peso a los Ratings más frecuentes y estime siempre esta categoría.

Este *oversample* lo hemos realizado tanto sobre las categorización de las agencias y como sobre nuestra categorización más general.

2.2. Análisis previos

Antes de diseñar nuestro modelo de regresión hay que validar algunas hipótesis sobre nuestros datos, ya que el modelo de regresión no resulta valido para todos los tipos de datos. Para que las conclusiones del modelo sean válidas, los datos que manejamos deben cumplir los siguientes supuestos:

³Para no hacer este artículo muy extenso no vamos a entrar en detalles de como hemos ajustado cada una de las métricas, pero hay que tener en cuenta que cada una es diferentes y que por tanto no es únicamente eliminar el percentil 5%. Ya que, para algunas de las métricas hay que asegurarse que no sean 0. Todo estos ajustes se pueden observar en los Notebooks 2 y 3 del Github https://github.com/RubenPhy/rating_analisis

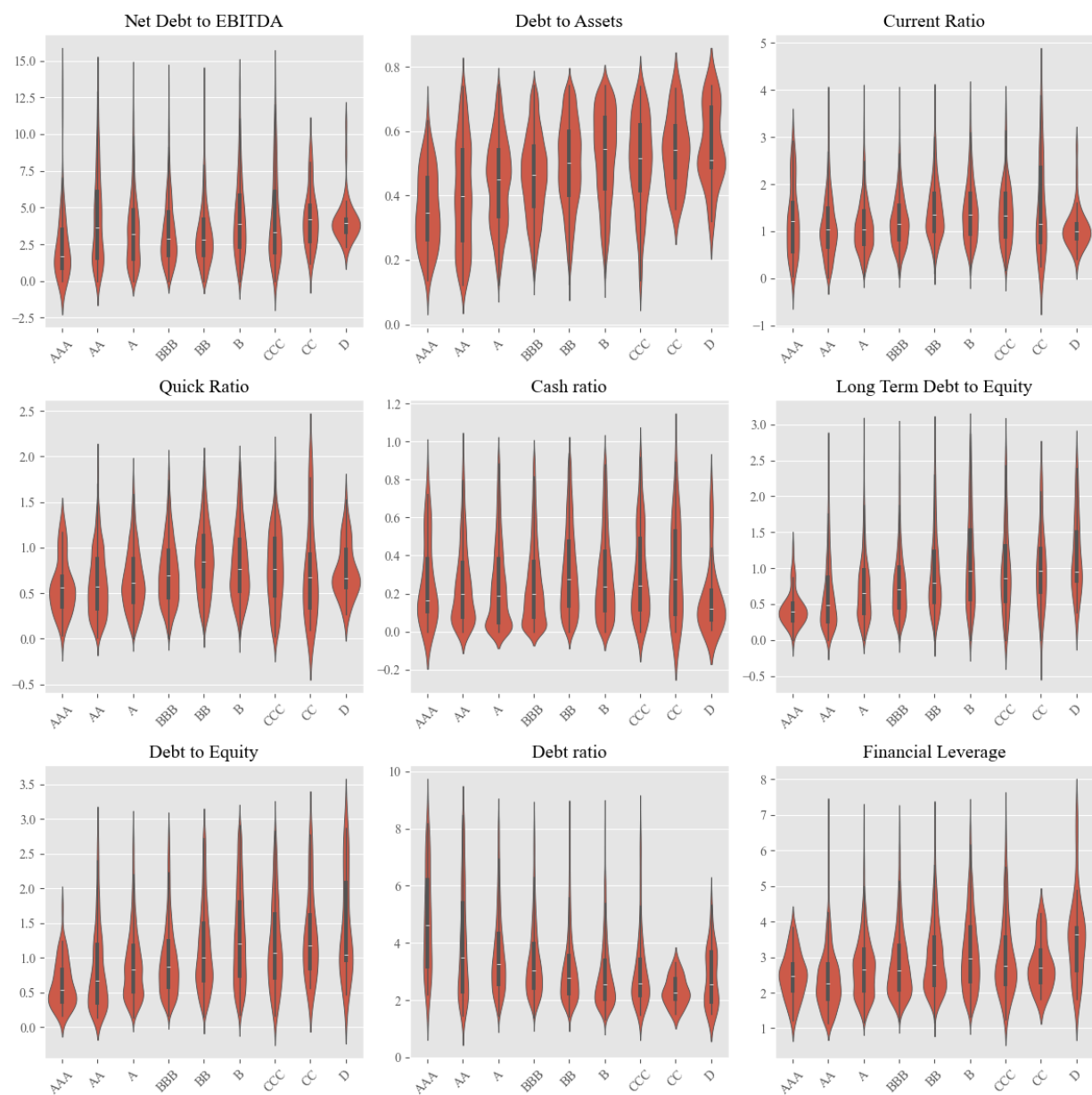


Figura 1: Distribución de las métricas por Rating

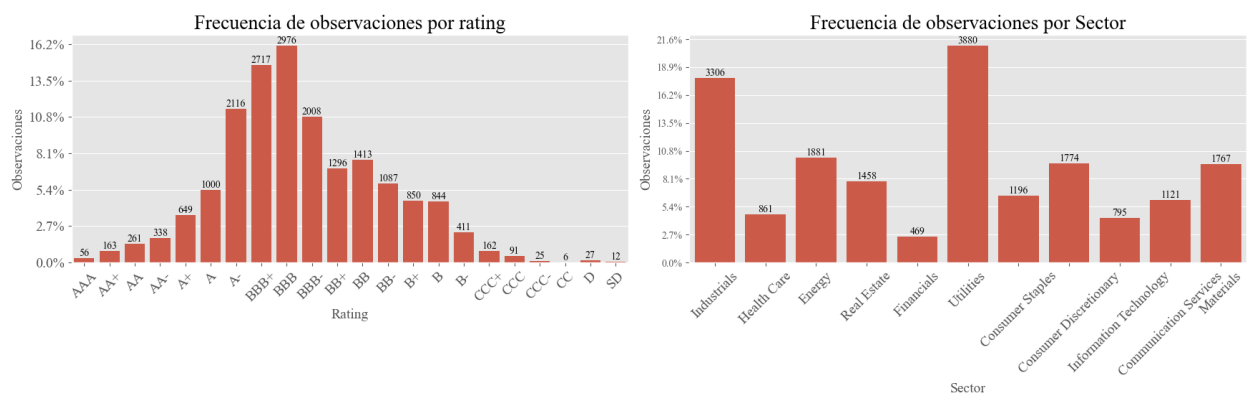


Figura 2: Distribución de los Ratings y los sectores

1. Sólo la variable dependiente Y se trata como aleatoria.
2. Las observaciones de $Y = (y_1, y_2, \dots, y_n)$ han de ser independientes. Esta hipótesis se cumple ya que las compañías son en su gran mayoría independientes de otras.
3. Para cada individuo de la población se cumple que $y_i = b_0 + b_1 x_i + u_i$ siendo u_i una perturbación aleatoria conocida también como residuo.
4. Condición de homoscedasticidad, que supone que $Var[Y]$ es constante respecto de X . Como observaremos más adelante esta es una condición que nuestras variables no cumplen.
5. La variable Y sigue una distribución normal. Como hemos observado en la Figure 8, nuestra variable dependiente sigue una distribución normal.

Así, en función de la independencia, normalidad de los datos y el tamaño de la muestra caben 4 posibilidades:

- Datos independientes que siguen una distribución Normal o el tamaño de ambas muestras es ≥ 30 . Se utiliza la prueba T de Student para datos independientes. Previamente se debe realizar una comparación de varianzas mediante el contraste de hipótesis adecuado.
- Datos independientes que no siguen una distribución Normal y el tamaño de las muestras es < 30 . No puede utilizarse la prueba T de Student y se deben usar pruebas no paramétricas.
- Datos apareados que siguen una distribución Normal o el tamaño de ambas muestras es ≥ 30 . Se utiliza la prueba T de Student para datos apareados.
- Datos apareados que no siguen una distribución Normal y tamaño de muestra es < 30 . Deben utilizarse pruebas no paramétricas, en este caso la prueba T de Wilcoxon.

Antes de realizar los contrastes de hipótesis vamos explicar de manera general en que consisten los contrastes de hipótesis. Cuando se comparan dos poblaciones mediante muestras aleatorias, nuestro objetivo es conocer si los parámetros de interés (medias, proporciones o varianzas) son diferentes en las poblaciones. Para ello realizamos contrastes de hipótesis que consisten en someter a prueba la hipótesis nula H_0 (ausencia de diferencias). Los contrastes habitualmente son bilaterales, es decir:

- Hipótesis nula (H_0): No hay diferencia entre los grupos.
- Hipótesis alternativa (H_1): Existe una diferencia significativa entre los grupos.

La probabilidad de que una diferencia observada entre 2 parámetros (medias, proporciones o varianzas) se deba realmente al azar se denomina grado de significación o p -valor. Cuando rechazamos una hipótesis nula podemos equivocarnos, ese error denominado error de primera especie debe ser menor del 5%. Por tanto, solo aceptaremos una diferencia como significativa cuando la probabilidad de que se deba simplemente al azar sea menor del 5%. Por tanto, para aceptar una diferencia observada como real deberá cumplirse que $p < 5\%$.

Pero para poder realizar estos Test antes tendremos que analizar si nuestros datos siguen una distribución Normal. La distribución Normal es una distribución continua de probabilidad. Es la distribución de probabilidad más utilizada debido a que es la que siguen muchas variables en la naturaleza y a que es la base de la inferencia estadística. Muchas pruebas estadísticas requieren como condición de aplicación que los datos no sean diferentes de una Normal. Por ello tienen gran importancia las pruebas de comprobación de normalidad. Existen pruebas gráficas y otras numéricas. Las más utilizadas son:

1. Representar los datos en papel probabilístico. Se trata de un gráfico de dispersión que compara los valores observados con los que esperaríamos encontrar si la distribución fuera Normal. Si los puntos siguen una recta diagonal los datos siguen una distribución Normal. Como se puede observar en la Figura 3, los datos no muestran una distribución normal, salvo en el caso de la variable dependiente.
2. Representar un histograma y comprobar la forma de campana. Esto se puede observar en la Figura 4, donde además de la variable independiente se observa que la métrica Debt to Asset también sigue una variable prácticamente normal.

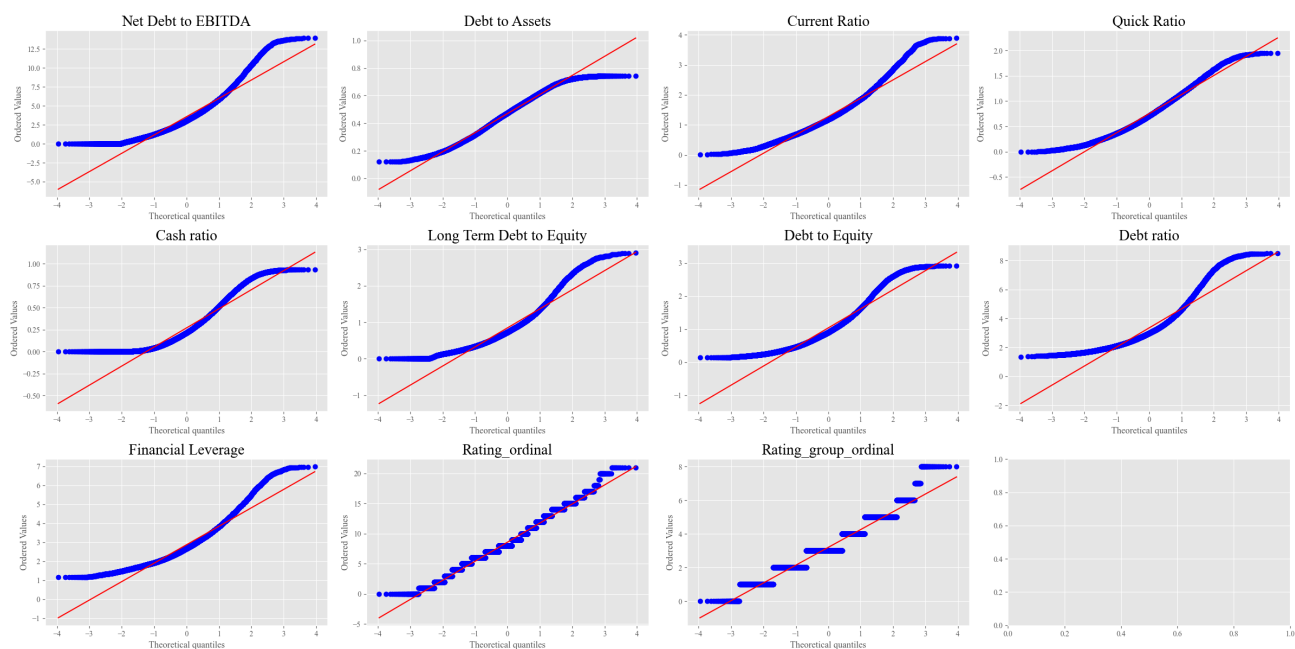


Figura 3: QQ plot de todas la variables

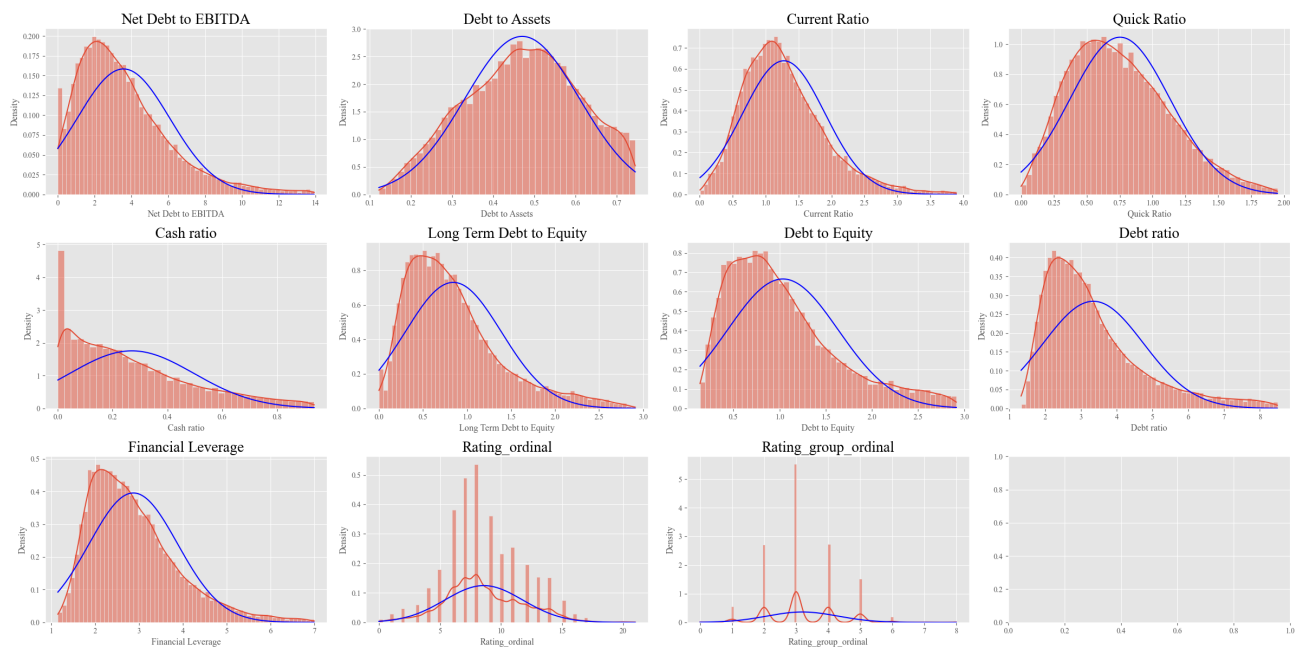


Figura 4: Histograma de todas las variables

Métrica	Coefficiente de asimetría	Coefficiente de curtosis	Simetría	Mesocúrtica
Net Debt to EBITDA	1.20	1.67	No es simétrica	No es mesocúrtica
Debt to Assets	-0.12	-0.70	No es simétrica	No es mesocúrtica
Current Ratio	0.92	1.17	No es simétrica	No es mesocúrtica
Quick Ratio	0.53	-0.16	No es simétrica	No es mesocúrtica
Cash ratio	0.90	0.05	No es simétrica	No es mesocúrtica
Long Term Debt to Equity	1.14	1.17	No es simétrica	No es mesocúrtica
Debt to Equity	0.99	0.50	No es simétrica	No es mesocúrtica
Debt ratio	1.34	1.59	No es simétrica	No es mesocúrtica
Financial Leverage	1.09	1.29	No es simétrica	No es mesocúrtica
Rating	0.32	0.16	No es simétrica	No es mesocúrtica
Rating agrupado	0.38	0.56	No es simétrica	No es mesocúrtica

Cuadro 3: Comprobación de distribución simétrica y mesocúrtica

3. Comprobar que la distribución es simétrica y mesocúrtica. Se puede hacer comprobando que los coeficientes de asimetría (CA) y de apuntamiento o curtosis (CC) no son significativamente diferentes de 0 mediante sus intervalos de confianza del 95 %, o bien comprobando que los coeficientes tipificados de asimetría y curtosis están comprendidos entre ± 2 . Este análisis se puede comprobar en la Tabla 3
4. Utilizar pruebas de bondad de ajuste como la prueba de Kolmogorov-Smirnov (K-S). Se basan en comparar la distribución acumulada de los valores de la muestra observada con la distribución acumulada que se obtendría en el supuesto que siguiera una distribución Normal con la misma media y varianza. Si en la prueba de K-S se obtienen valores de $p < 5\%$, la distribución difiere significativamente de una Normal. Dado que tenemos una serie de datos mayor a 5000, el librería de Python *scipy* recomienda usar el Test de Lilliefors, el cual es un tipo de test de Kolmogorov-Smirnov. Realizando este test para nuestras variables independientes tampoco obtenemos una distribución normal de los datos, con un nivel de significatividad del 1 %.

Nuestro siguiente paso es comprobar la homocedasticidad de los datos. Dado que los datos ya hemos comprobado que no se comportan de manera normal usaremos el test de Levene. El test de Levene pone a prueba la hipótesis nula de que todas las muestras proceden de poblaciones con varianzas iguales y es una alternativa en el caso de que hay desviaciones significativas de la normalidad.

Cuando obtienes p-valores menores a 0.05 en la prueba de Levene, indica que las varianzas entre tus grupos no son iguales, es decir, tus datos violan la suposición de homocedasticidad. Esto afecta el uso de pruebas paramétricas que asumen igualdad de varianzas entre los grupos, como el t-test para muestras independientes y la ANOVA. Aplicar transformaciones a tus datos puede ayudar a estabilizar la varianza. Algunas transformaciones que hemos intentado para poder ajustar los datos son:

1. Logarítmica: $\log x$ es útil para datos con varianzas que aumentan con la media.
2. Raíz Cuadrada: \sqrt{x} para datos de conteo o tasas.
3. Box-Cox: La transformación Box-Cox puede ser aplicada a datos positivos y busca una transformación que estabilice la varianza y aproxime los datos a una distribución normal.

Después de probar estas tres técnicas no hemos conseguido mejorar la la homocedasticidad.

Una vez vemos que no son normales y no tienen varianzas homogéneas debemos considerar el uso de métodos no paramétricos para nuestros análisis, como la prueba de Kruskal-Wallis para comparar medianas en lugar de medias. Realizando este análisis para las 9 métricas, obtenemos que hay diferencias estadísticamente significativas entre las medias de los Ratings, con un nivel de significatividad del 5 %.

Una vez que has determinado que existen diferencias significativas entre los grupos mediante la prueba de Kruskal-Wallis, el siguiente paso es realizar un análisis post-hoc para identificar específicamente entre qué pares de grupos existen esas diferencias. La Prueba de Dunn es una opción popular para comparaciones múltiples después de una Prueba de Kruskal-Wallis, ya que está diseñada para datos no paramétricos. La Figura 5, nos dice que para los cuadrantes donde el p-value es proximo a 0 (es decir blanco) las medias de las muestras son diferentes. Es decir, las medias de los datos con Rating peores (entorno al CC) son significativamente diferentes del resto

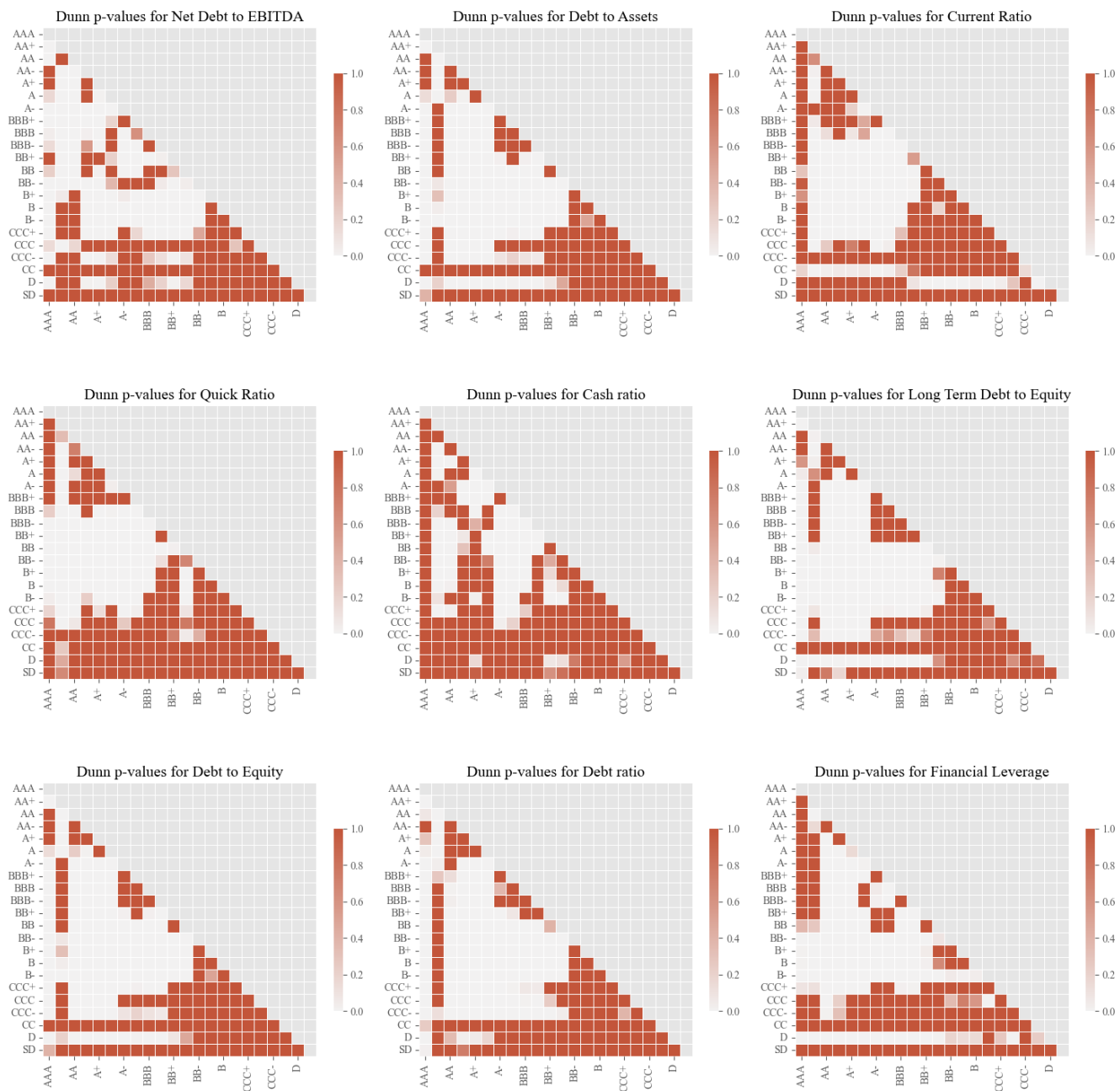


Figura 5: P-Values obtenidos de la prueba de Dunn

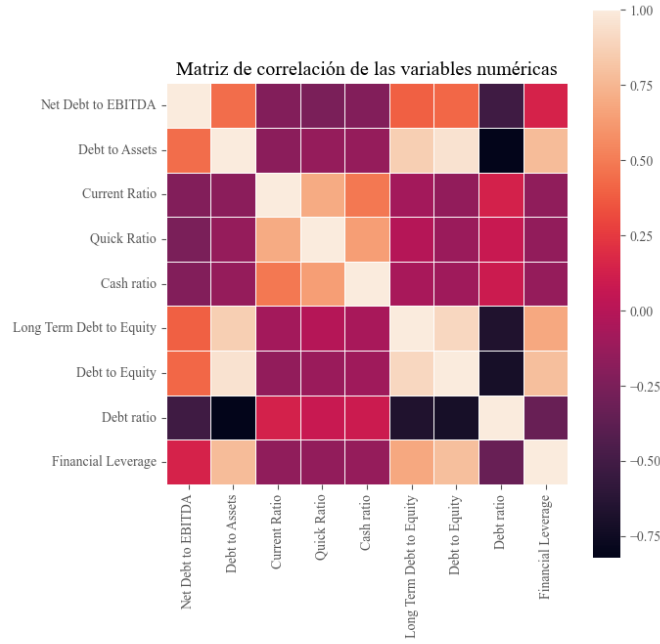


Figura 6: Matriz de correlación

Otra de las hipótesis a comprobar antes de realizar los modelos es ver si existe correlación entre las variables independientes. Sin embargo, cuando las variables no siguen una distribución normal, y quieres evaluar la significancia de las correlaciones, puedes utilizar métodos de correlación no paramétricos junto con pruebas de significancia adecuadas. Dos de las medidas de correlación no paramétricas más comunes son la correlación de Spearman y la correlación de Kendall. Estas medidas no asumen normalidad y son adecuadas para datos ordinales, rangos, o distribuciones no normales.

Para terminar de analizar los datos veremos la colinialidad de una, muestra aleatorio de 1,000 puntos de los datos. Como se observa en Figura 7, algunas de métricas siguen cierta correlación entre ellas, por tanto, a la hora de realizar los modelos, se podría prescindir de estas variables.

3. Modelos

Antes de realizar los test estadísticos de las variables dependientes analizaremos las frecuencias de los datos para cada Rating. Como observamos en la Figura 8, si que se puede observar para la mayoría de las métricas diferencias en las frecuencias de los datos para cada Rating. Esto nos ayudará a obtener mejores predicciones con nuestros modelos.

En este apartado probaremos los modelos de regresión y clasificación que hemos considerado más adecuados. Desde los modelos más básicos de clasificación hasta algoritmos de ML como el Random Forest.

3.1. Modelos de clasificación Lineal

El modelo más sencillo de clasificación consiste en, dada la métrica i , el Rating estimado sería el Rating el cual tenga su media más cercana, es decir, para un valor x y un conjunto de medias $M = \{m_1, m_2, \dots, m_n\}$, calculamos la diferencia absoluta entre x y cada m_i en M , generando un conjunto D de diferencias:

$$D = \{|x - m_1|, |x - m_2|, \dots, |x - m_n|\}$$

Buscamos el índice k tal que D_k es el mínimo en D , donde $D_k = |x - m_k|$ es la diferencia mínima absoluta entre x y las medias en M . Matemáticamente, esto se puede expresar como:

$$k = \underset{i}{\operatorname{argmin}} |x - m_i|$$



Figura 7: Análisis de colinialidad

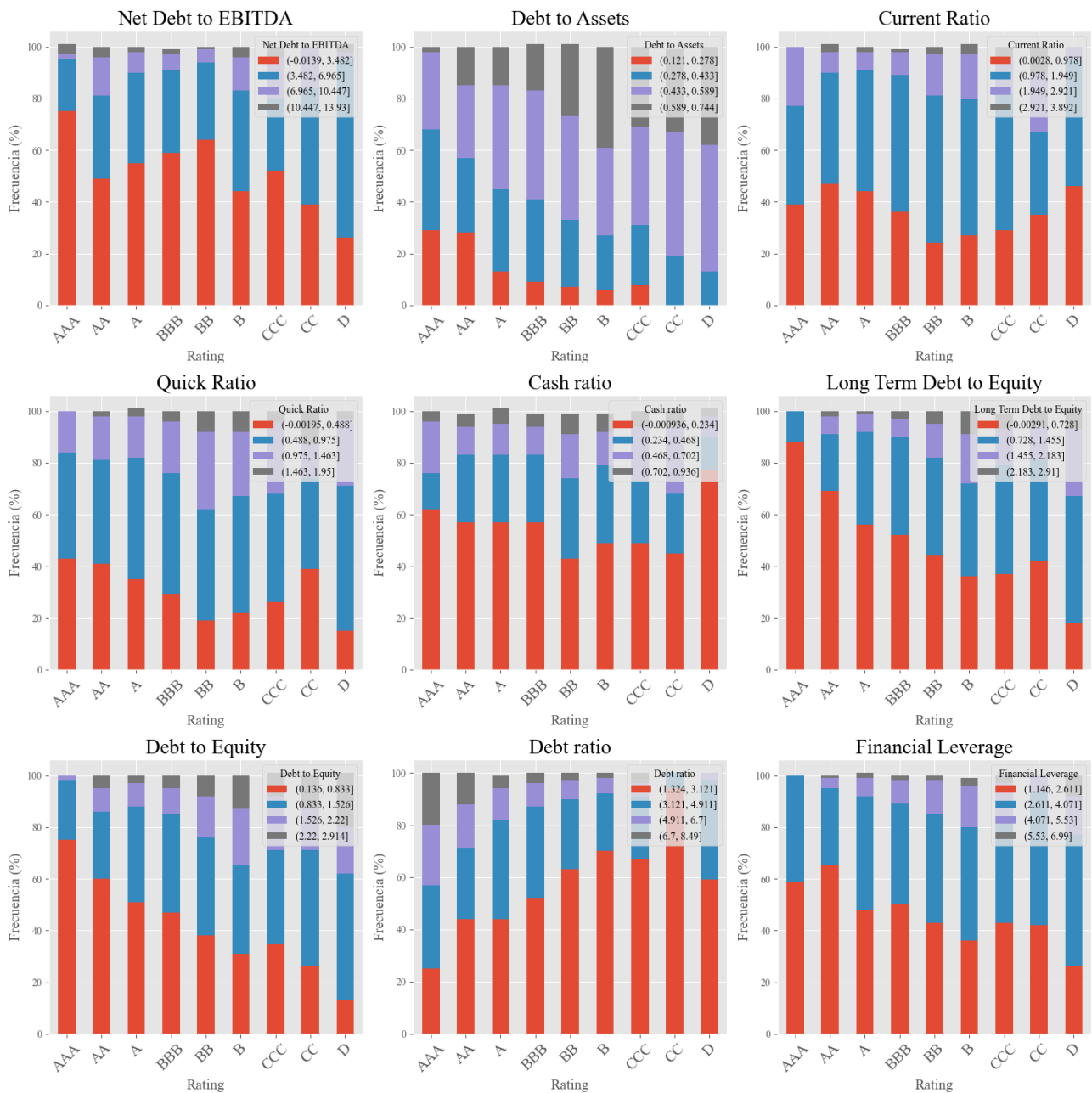


Figura 8: Frecuencia de las métricas por Rating

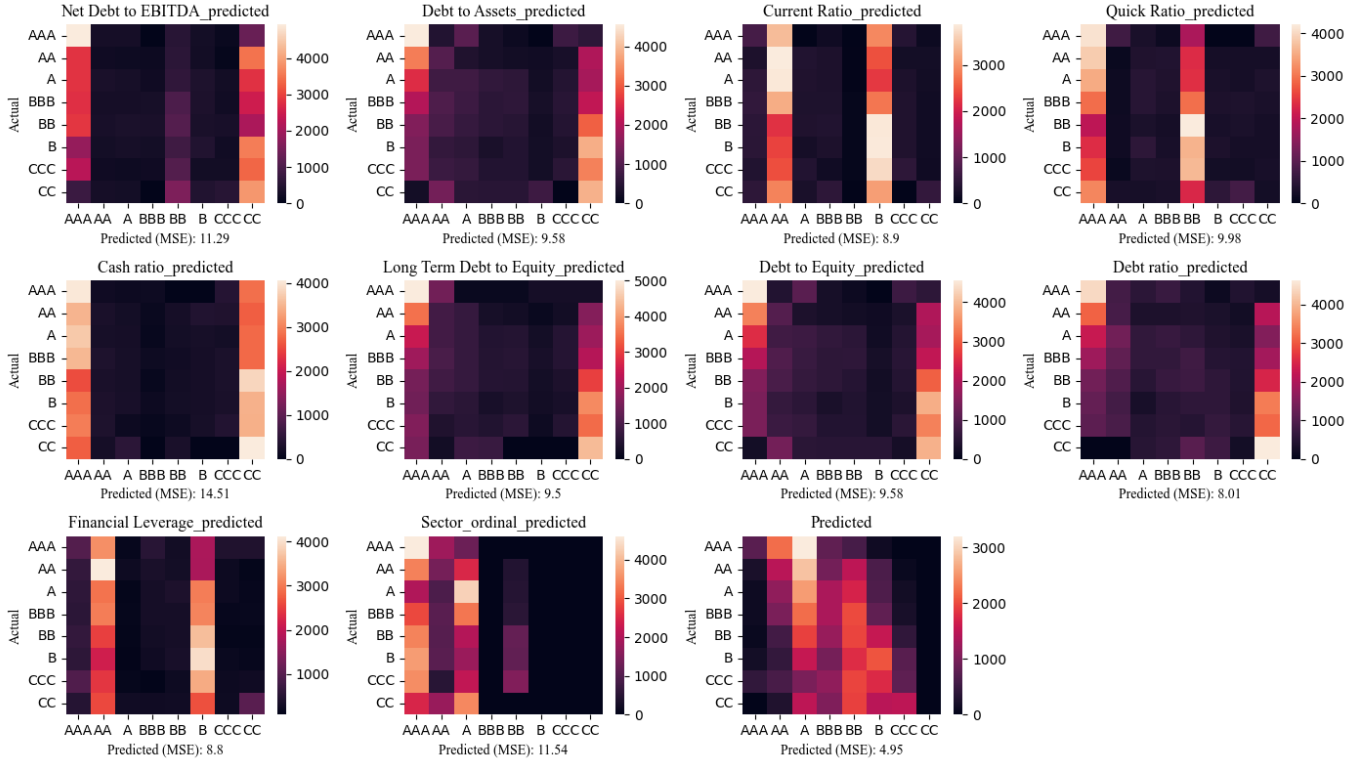


Figura 9: Matriz de confusión para cada una de las métricas y la media.

Donde argmin devuelve el índice i del elemento de M que minimiza la diferencia absoluta con x , esencialmente encontrando la media más cercana a x según la diferencia absoluta.

Este análisis lo realizamos para una de las 9 métricas y luego de manera conjunta, es decir, haciendo la media de los resultados de las 9 clasificaciones. Para observar los resultados obtenidos, hemos hecho una matriz de confusión para cada caso. Como se puede observar en la Figura 9, los resultados cuando lo vemos métrica a métrica no son buenos, pero con la media de los 9 resultados, el Error Cuadrático Medio o Mean Square Error (MSE) mejoró considerablemente.

3.2. Modelos de regresión

Para estimar nuestro Rating partiendo de los datos métricas financieras hemos utilizado tres tipos de modelos:

1. **Regresión Lineal:** La regresión lineal es uno de los métodos más simples y ampliamente utilizados para modelar la relación entre una variable dependiente y una o más variables independientes. El objetivo es encontrar una en el hiperplano que mejor se ajuste a los datos, minimizando la suma de las diferencias cuadradas entre los valores observados y los valores predichos por el modelo. La forma general de la regresión lineal para una variable dependiente y y n variables independientes (x_1, x_2, \dots, x_n) es:

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_n x_n + \epsilon$$

Donde β_0 es el término de intercepción, β_1, \dots, β_n son los coeficientes que representan el peso de cada variable independiente en la predicción de y , y ϵ es el término de error. Como se puede observar en la Figura 10 con este modeo obtenemos los peores resultados.

2. **Regresión Polinómica de grado 3:** La regresión polinómica es una forma de regresión lineal en la que la relación entre la variable independiente x y la variable dependiente y se modela como un polinomio de grado n . En este caso las características de entrada son elevadas a potencias hasta el grado 3 (incluyendo combinaciones de características), lo cual permite modelar relaciones no lineales entre las variables. Por ejemplo, para una única

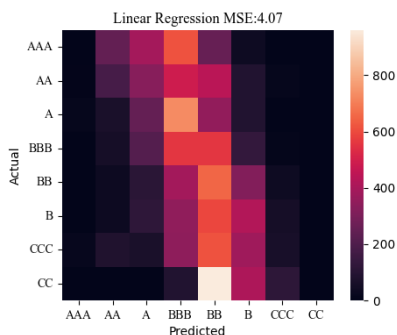


Figura 10: Regresión Lineal

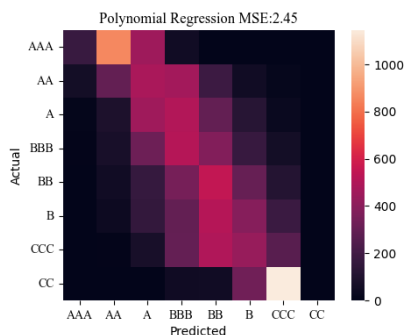


Figura 11: Regresión Polinómica

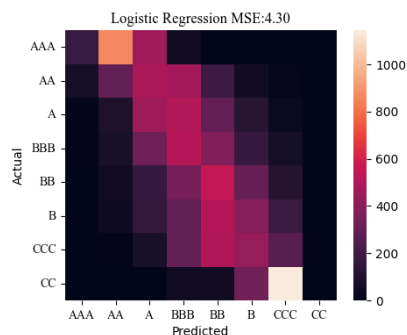


Figura 12: Regresión Logística

Tipo de regresión	Aciertos exactos	Acierto (± 1)
Regresión Lineal	17,25 %	49,61 %
Regresión Logística	18,33 %	51,67 %
Regresión Polinomial	21,32 %	66,86 %

Cuadro 4: Precisión de las estimaciones

variable independiente x , el modelo polinómico de tercer grado sería:

$$y = \beta_0 + \beta_1 x + \beta_2 x^2 + \beta_3 x^3 + \epsilon$$

En este modelos el MSE observado en la Figura 11 se reducen considerablemente.

3. **Modelos Logístico:** La regresión logística ordinal es útil cuando la variable dependiente es ordinal, es decir, cuando las categorías tienen un orden natural pero las distancias entre categorías no son necesariamente iguales. El modelo LogisticAT asume que hay un único conjunto de coeficientes (similar a la regresión logística multinomial) pero difiere en que los umbrales para decidir entre categorías consecutivas no son equidistantes; estos se aprenden automáticamente de los datos. El enfoque busca minimizar una función de pérdida que penaliza la clasificación incorrecta de las observaciones, con el objetivo de establecer umbrales que mejor separen las categorías en orden.

3.3. Modelos de ML

Por último hemos creado un modelo Random Forest (Bosque Aleatorio) de ML. Este modelos modelo de Machine Learning es un método de ensamble que opera construyendo múltiples árboles de decisión durante el entrenamiento y produciendo la clase que es la predicción media de los árboles individuales. Este método se basa en combinar las predicciones de varios arboles de decisión de aprendizaje automático para hacer predicciones más precisas que cualquiera de los modelos individuales. A continuación explicamos algunos de los puntos más importantes del Borque Aleatorio.

1. **Árboles de decisión:** Son la unidad básica de un Random Forest. Un árbol de decisión es un modelo que hace predicciones siguiendo una serie de reglas de decisión binarias (sí/no) basadas en las características de los datos. Aunque poderosos, individualmente, los árboles de decisión pueden ser propensos al sobreajuste.
2. **Ensamblaje (Bootstrap Aggregating o Bagging):** Para crear un Random Forest, primero se genera una serie de árboles de decisión. Cada árbol se entrena con una muestra aleatoria de los datos de entrenamiento, seleccionada con reemplazo, conocida como bootstrap sample. Esto ayuda a asegurar que cada árbol sea diferente y reduzca el riesgo de sobreajuste.
3. **Aleatorización:** Además de la aleatorización introducida por el bootstrap sampling, Random Forest introduce más variabilidad en los árboles de decisión al limitar el conjunto de características que se pueden considerar para dividir en cada nodo de los árboles. Esto significa que, en lugar de buscar la mejor característica al dividir un nodo, el algoritmo busca la mejor característica entre un subconjunto aleatorio de características. Esto hace que los árboles sean más diversos y mejora el rendimiento del modelo general.

4. **Agregación:** Después de entrenar, las predicciones de todos los árboles individuales se combinan para formar la predicción final. Para problemas de clasificación, esto suele significar tomar la moda (la clase más frecuente) de las predicciones de todos los árboles. Para regresión, se toma el promedio de las predicciones.

Dividiendo nuestros datos en dos conjuntos de Train y Test hemos entrenado el modelo de Random Forest con 100 Árboles de decisión, obteniendo un acierto en el 93 % de los casos. Sin embargo, en la sección 4 analizaremos este modelo con más detalle.

4. Resultados

Como hemos mencionado anteriormente, al haber realizado un *oversample* de nuestra muestra, cuando hacemos las estimaciones con datos diferentes a los que el modelo ha usado para entrenar, no podemos asegurar que sean datos diferentes, ya que para algunas categorías, la muestra original era varios ordenes de magnitud menor que para la categoría más frecuente, como se observa en la Figura 2.

Por ello hemos sacado todas las métricas para los principales índices del mundo para el año 2024, los cuales son datos que nunca han pasado por el modelo. De esta manera obtenemos un buen indicador de la precisión de cada uno de los modelos. Aplicando los mismos filtros y la misma enumeración para las variables categóricas que en nuestro datos originales, llegamos a una muestra de 753 empresas. En la Tabla 4 observamos que el mejor modelo para las 3 medidas de error que hemos calculado es el modelo de Random Forest. Por tanto, hemos realizado las estimaciones con el modelo de Random Forest y representado la matriz de confusión para el modelo de Random Forest en la Figura 13. La cual muestra unos resultados bastante buenos, con un acierto en el 61 % de los casos y con un acierto aproximado con ± 2 Ratings en el 78 %.

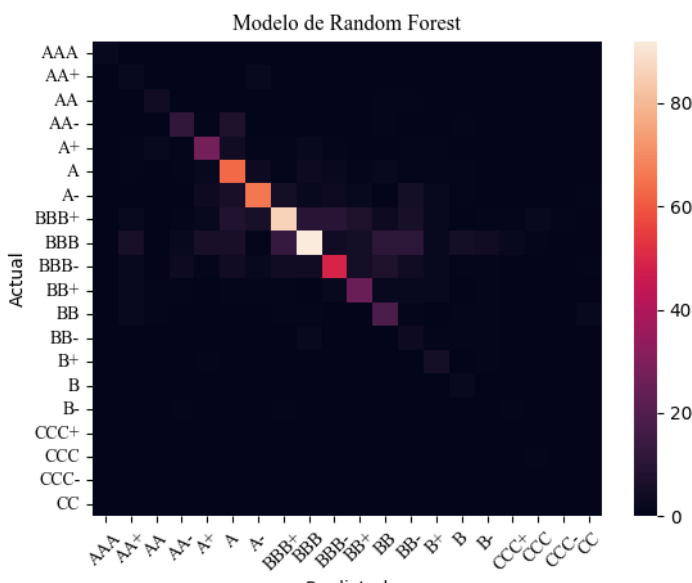


Figura 13: Matriz de confusión para los nuevos datos y el modelo de Random Forest.

5. Conclusiones

Después de analizar los datos y realizar diferentes tipos de modelos, podemos concluir que mediante los modelos de regresión y clasificación “clásica” obtenemos niveles de precisión en nuestras estimaciones aceptables, pero que esta precisión aumenta considerablemente con el modelo de ML de Random Forest. Este trabajo, podría dar pie a una serie de trabajos que analizaran con más profundidad algunos aspectos que no se han analizado en este trabajo pero podrían mejorar las estimaciones del Rating para algunas compañías. Entre estos trabajos se encuentran:

1. Se podría realizar este mismo análisis añadiendo otras métricas financieras, tales como:

- Coste de la financiación.
 - Capitalización de mercado.
2. Intentar normalizar los datos para ver si los modelos de regresión reducen su error.
 3. Estudiar si hay diferencia entre las métricas para cada año, ya que nosotros hemos usado datos de diferentes años y esta podría ser una variable a tener en cuenta.
 4. Ver si hay diferencias estadísticas entre las métricas por países y si esta podría ser otra variable independiente.
 5. Analizar si nuestro error de predicción varía por países o por sector.
 6. Aplicar otras técnicas de ML.
 7. Estudiar algoritmos de clasificación no supervisados, para poder clasifique todas las empresas sin Rating.
 8. Buscar versiones de los modelos de ML más sencilla para poder exportar estos modelos a aplicaciones como *Excel*, de manera que sea más accesible para otros usuarios.

Referencias

- [1] D. Creal, R. Gramacy y R. Tsay. Market-Based Credit Ratings. En: *SSRN Electronic Journal* (ene. de 2012). <https://doi.org/10.2139/ssrn.2310260>.
- [2] D. Delgado-Vaquero y J. Morales-Díaz. Estimating a Credit Rating for Accounting Purposes: A Quantitative Approach. En: *Estudios de Economía Aplicada* **36** (mayo de 2018). <https://doi.org/10.25115/eea.v36i2.2539>.
- [3] K. Galil, A. Hauptman y R. Rosenboim. Prediction of Corporate Credit Ratings with Machine Learning: Simple Interpretative Models. En: *Finance Research Letters* **58** (oct. de 2023), pág. 104648. <https://doi.org/10.1016/j.frl.2023.104648>.
- [4] P. Golbayani, I. Florescu y R. Chatterjee. A comparative study of forecasting Corporate Credit Ratings using Neural Networks, Support Vector Machines, and Decision Trees. En: (jul. de 2020). <https://doi.org/10.1016/j.najef.2020.101251>.
- [5] M. Wallis, K. Kumar y A. Gepp. Credit Rating Forecasting Using Machine Learning Techniques. En: *Managerial Perspectives on Intelligent Big Data Analytics* (ene. de 2019), págs. 180-198. <https://doi.org/10.4018/978-1-5225-7277-0.ch010>.