

## Introducción

El problema a resolver durante este proyecto es el de determinar qué variables son más importantes para predecir el porcentaje de la deuda de un conjunto de empresas de diferentes industrias dentro de México, las variables que se consideraron son el porcentaje de variación del precio spot, del tipo de cambio a 360 días, del INPC y del PIB que son variables macro, además de variables individuales de cada empresa como son la variación en los ingresos, en el activo total, en el flujo de operaciones, en el margen neto y en el ROE.

## Metodología

Primero se separaron cada uno de los datos, ya que cada todas las variables spot estaban unidas en la misma tabla, posterior se unieron por empresa ya con las variables macro, de forma posterior se rellenaron valores faltantes, puesto que había tablas con valores nulos, esto se realizó llenando con la mediana de los datos sobre esa misma variable, de forma siguiente se llevó un proceso para obtener el porcentaje variación de las siguientes variables: El ingreso, el activo total y el flujo, después se llevó un proceso de estandarización, el cual consiste en lo siguiente

$$\hat{x}_{(j)} = \frac{x_{(j)} - \bar{x}}{\sigma}$$

donde  $x_{(j)}$  es la medición en el tiempo  $j$  y se transforma en  $\hat{x}_{(j)}$  restándole la media de los datos y dividiéndolos entre la desviación estándar de los mismos, así para cada variable de los datos, de forma siguiente se realizó a observar si estos datos eran independientes y lo que se encontró es que el tipo de cambio estaba altamente correlacionado con el tipo de cambio a 360 y, por otro lado, estaban relacionados el ROE con el margen neto como se puede ver en la siguiente imagen,

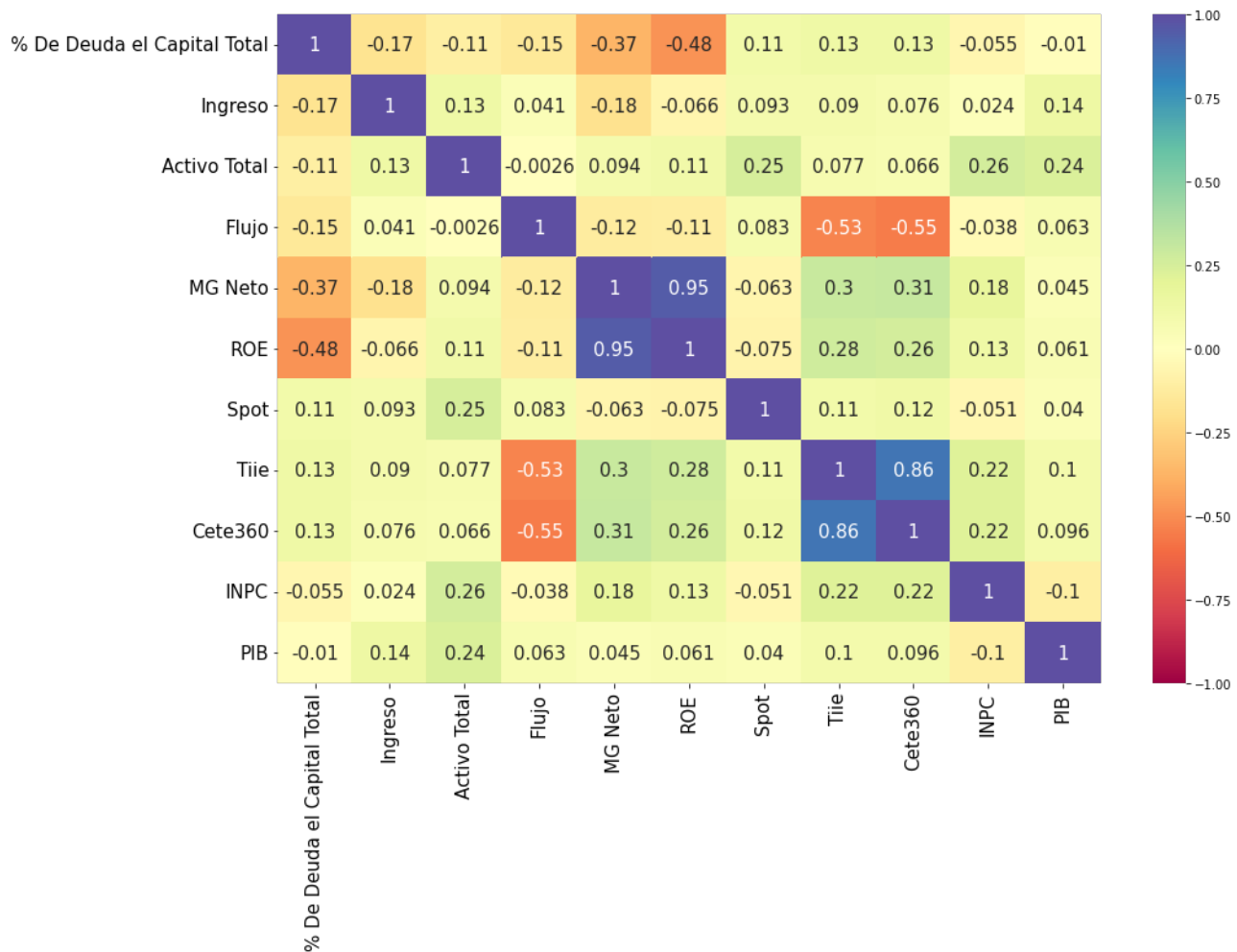


Figura 1: Correlación de las variables

así que se decidió eliminar el tiie así como el ROE, por último se realizaron 3 tipos de regresiones que son la Elastic Net, la HUBER y la RIDGE y la de tipo Bayesiana. Posteriormente, se realizaron las predicciones y se tomaron los mejores modelos para cada empresa, la forma en que evaluamos esto es usando el error cuadrático medio, posteriormente se tomaron los pesos que arrojaron estos modelos y vemos cuáles eran los que influían mejor.

Antes de presentar los resultados vamos a hablar un poco de los supuestos así como de la parte matemática de estas regresiones, lo bueno de estas regresiones es que sus supuestos son diferentes a los de una regresión tradicional, ya que estos se relajan, por ejemplo, estos no piden la homoscedasticidad o la normalidad de los errores, de hecho solo piden la independencia de las variables, que se logró eliminando tanto el Tiie como el ROE.

Ahora vamos a hablar de los ajustes de los modelos.

Primero empecemos con **Elastic Net** la cual busca minimizar con respecto de  $w$  el cual es el vector de pesos para la siguiente función:

$$\frac{1}{2n} \|Xw - y\|_2^2 + \alpha \rho \|w\|_1 + \frac{\alpha(1 - \rho)}{2} \|w\|_2^2$$

donde  $n$  es la cantidad de elementos para ajustar la regresión,  $X$  los datos que tenemos,  $y$  los pesos a estimar,  $\alpha$  y  $\rho$  son valores arbitrarios que nosotros asignamos, por último  $\| - \|_1$  y  $\| - \|_2$  son la norma l1

y  $l_2$  respectivamente.

Para la regresión **Huber** tenemos que se busca minimizar la siguiente función con respecto de  $w$  y  $\sigma$ :

$$\sum_{i=1}^n \left( \sigma + H_{\epsilon} \left( \frac{X_i w - y_i}{\sigma} \right) \sigma \right) + \alpha \|w\|_2^2$$

Figura 2:

donde

$$H_{\epsilon}(z) = \begin{cases} z^2, & \text{if } |z| < \epsilon, \\ 2\epsilon|z| - \epsilon^2, & \text{otherwise} \end{cases}$$

Figura 3:

Para la regresión de tipo **Ridge** se busca minimizar la siguiente función con respecto de  $w$ :

$$\|Xw - y\|_2^2 + \alpha \|w\|_2^2$$

Figura 4:

Por último la regresión **Bayesina** estima la siguiente probabilidad y con base a eso estima:

$$p(y|X, w, \alpha) = \mathcal{N}(y|Xw, \alpha)$$

Figura 5:

## Resultados

Los resultados de los errores cuadráticos medios para cada modelo y para cada empresa se pueden ver en la siguiente tabla.

	Elastic Net	Bayesiana	HUBER	Ridge
ALFAA	179.244317	183.567303	16427.799902	249.527514
ALSEA	825.515633	783.830721	987684.928023	1197.464147
AMXL	371.514787	400.574276	65426.743867	735.958080
ASURB	218.873081	224.730452	36809.222835	1208.785905
BIMBOA	203.135942	199.011977	263948.705422	136.458349
GCC	37.900758	37.907771	28338.079151	89.104267
CEMEXCPO	42.375800	48.872506	13465.103625	702.619776
ELEKTRA	194.340044	194.485752	5198.598937	297.063641
FEMSAUBD	105.702529	99.892882	36096.970100	53.433778
GCARSOA1	149.053763	144.882799	5733.724861	108.894730
GMEXICOB	57.736880	2455.192870	4522.152983	485.542173
GRUMAB	75.463812	71.766184	19626.589180	380.847140
KIMBERA	715.290179	706.712270	570004.006880	552.896363
LIVEPOLC-1	41.433363	46.245194	41090.249643	51.061981
ORBIA	325.024665	312.174747	11012.651337	280.034539
PE&OLES	66.822058	60.588418	16521.611609	277.093793
TLEVISACPO	93.503059	814.332361	46803.004483	377.973460
WALMEX	123.080205	102.471570	219351.863251	119.561978

Figura 6: Resultados de los MSE (Mean Squarred Error)

Posteriormente para cada compañía se tomo el mejor modelo y se observaron los pesos que se ajustaron, también como hay pesos positivos y negativos podemos ver que hay factores que hacen que la empresa se endeude más o se endeude menos, también podemos ver con estos pesos cuales son los factores que influyen más y cuales los que influyen menos. Aquí algunas gráficas que muestran esto para ALSEA. Las gráficas para las demás empresas se pueden ver en el nootebook anexo. Un ejemplo de lo anterior se ve aquí por ejemplo tenemos el conjunto de pesos 1, 2, 3, -4, -5 donde cada peso está dado para las variables  $x_1, x_2, x_3, x_4, x_5$  dados los pesos calculamos el influencia total de la siguiente forma: para  $x_1 = 1$ , entonces su influencia es  $\frac{1}{1+2+3+4+5} = \frac{1}{15}$ , así para todas, ahora para calcular su influencia «Positiva» se toman los valores positivos y se realiza el mismo calculo para  $x_1$  el resultado es  $\frac{1}{1+2+3} = \frac{1}{6}$ , ahora para calcular la influencia «Negativa» el calculo se hace de la siguiente forma  $\frac{|-4|}{|-4|+|-5|} = \frac{4}{9}$ .

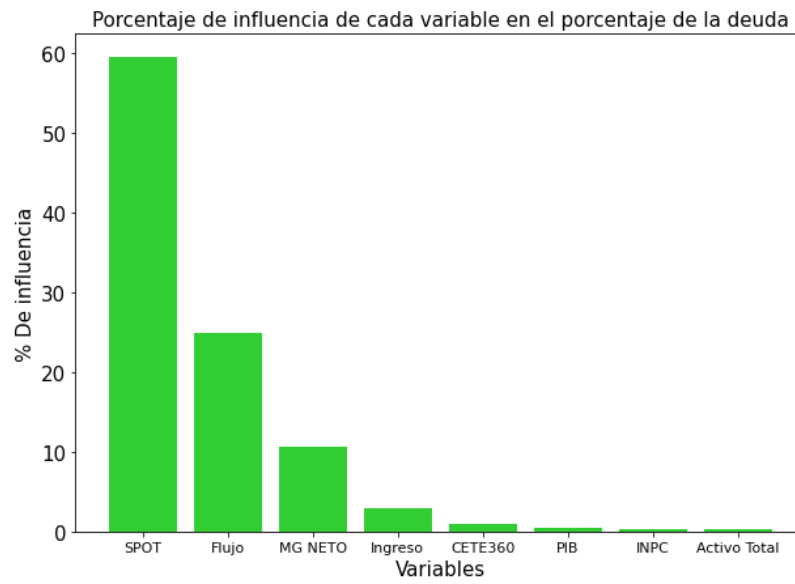


Figura 7:

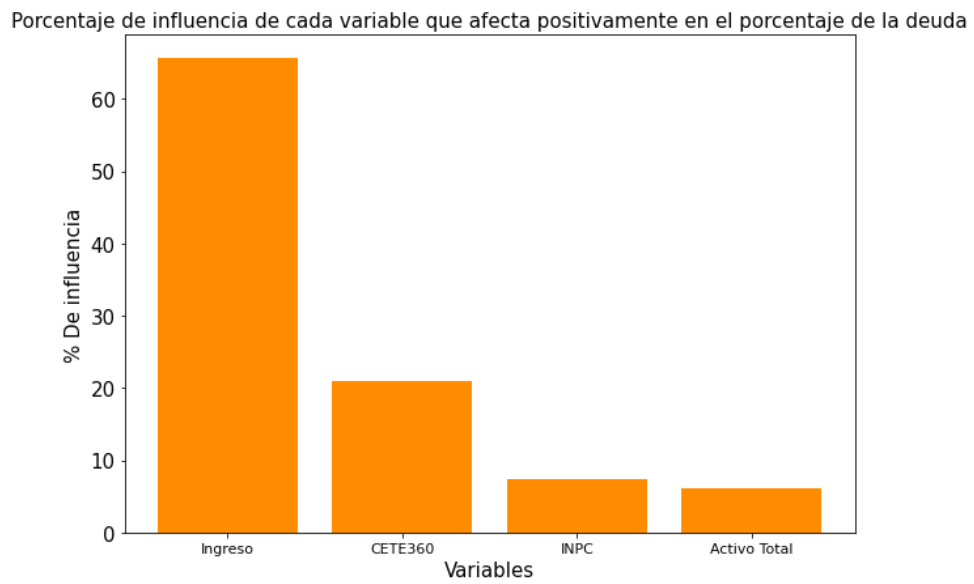


Figura 8:

Porcentaje de influencia de cada variable que afecta negativamente en el porcentaje de la deuda

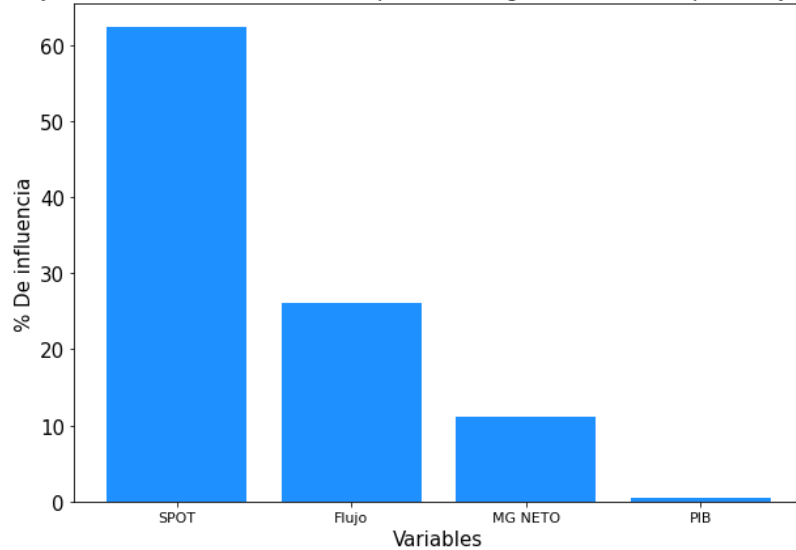


Figura 9:

Para la empresa **ALFAA** las variables de mayor importancia fueron: el **flujo** y el **Margen neto** en ese orden.

Para la empresa **ALSEA** fueron el precio **Spot**, el **flujo** y el **Margen Neto** en ese orden.

Para **AXML** fueron el **El flujo**, el **activo total**, el **margen neto**, el **ingreso** y el **precio spot** en ese orden.

Para **ASURB** fueron el **Margen neto**, el **precio spot** y el **flujo** en ese orden.

Para **BIMBOA** fueron **El Margen neto**, el **ingreso** y el **flujo**

Para **GCC** fue solo **El flujo**

Para **CEMEXCPO** fueron **El Margen Neto**, el **precio Spot** y el **Flujo**.

Para **ELEKTRA** fueron el **Ingreso** y el **flujo**.

Para **FEMSAUBD** fueron el **EL Margen Neto**, el **Ingreso**, el **Flujo**, el **Activo total** y el **precio Spot**

Para **GCARSO1** fueron el **El Ingreso**, el **Margen Neto**, el **Activo total**, el **precio Spot** y el **PIB**

Para **GMEXICOB** fueron el **El flujo** y el **Margen neto**

Para **GRUMAB** fueron el **El flujo** y el **Margen neto**, el **precio spot** y el **ingreso**

Para **KIMBERA** fueron el **EL Margen Neto**, el **Ingreso**, el **Flujo**, el **PIB** y el **CETE360**

Para **LIVEPOLC-1** fueron el **El flujo** y **precio spot**

Para **ORBIA** fueron el **El Activo Total**, el **ingreso** y el **margen neto**

Para **PE&OLES** fueron el **El flujo, el precio spot y el Margen Neto**

Para **TLEVISACPO** fueron el **El flujo, el margen Neto y el precio spot**

Para **WALMEX** fueron el **El precio spot y el flujo**

## **Conclusiones**

Podemos observar que en la mayoría de las empresas las variables que posee mayor importancia son las individuales de cada una, a excepción del precio spot que llega a aparecer en la mayoría de estas, ahora dentro de las variables individuales las que parecen tener mayor importancia son El flujo, el ingreso y el Margen neto que son las que aparecen en la mayoría de los grupos.