

Universidad Cenfotec

Data Analytics /Big Data (XTOL)

Capstone Project

“Predicción de Bancarrota Empresas Polacas”

Estudiante:

Deily Rojas Mejias

Abril 2020

## Objetivo

Diseñar e implementar un modelo de inteligencia de negocios efectivo por medio de machine learning, que permita determinar si una empresa polaca se expone en declararse en bancarrota según su comportamiento. El modelo por desarrollar tomara como base la información financiera de las empresas correspondiente al año 2007.

## Administración de la información

Para el análisis de la información se cuenta con 64 columnas de información correspondiente a métricas financieras orientadas en su Rentabilidad, Liquidez y Solvencia. Esto para 7027 empresas polacas. En el dataset publicado se cuenta con información de los últimos 5 años, sin embargo para focalizar esta investigación, se analiza únicamente el año 2007.

En el análisis de la información, y fue necesario reforzar el conocimiento de estados financieros y sus diferentes métricas sobre las compañías con la finalidad de comprender contablemente que significan aspectos como rentabilidad o liquidez.

Por lo tanto, de los datos en estudio las métricas fueron clasificadas en 3 categorías (Rentabilidad, Liquidez y Solvencia). Las cuales nos permitirán, generar comportamiento que nos indiquen si esta en riesgo de bancarrota la compañía.

### Rentabilidad

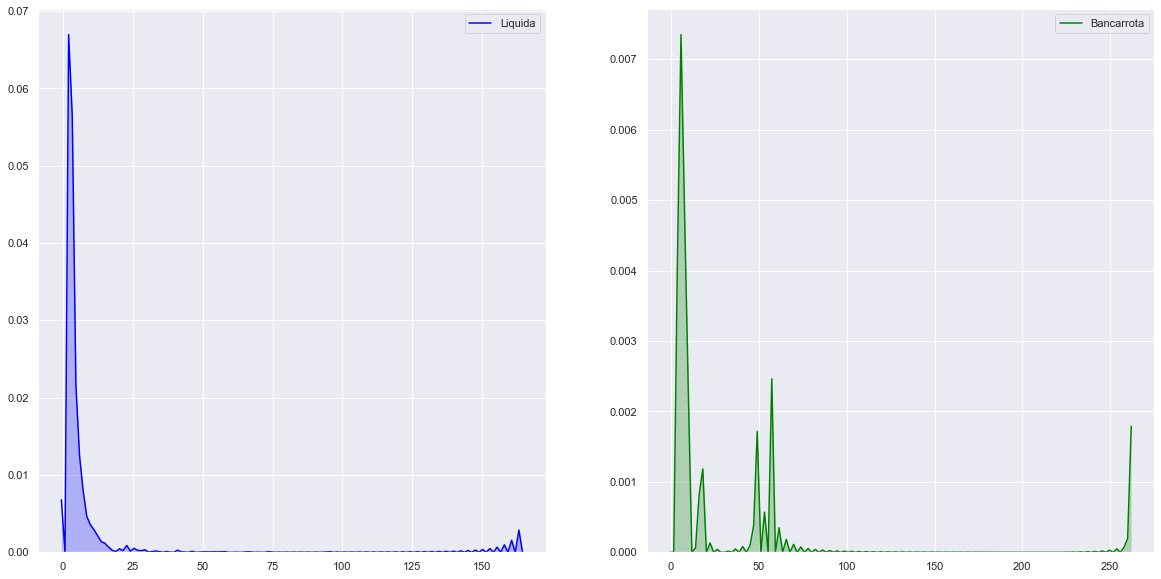
#### Rentabilidad de los activos: Utilidad de operación/activo total

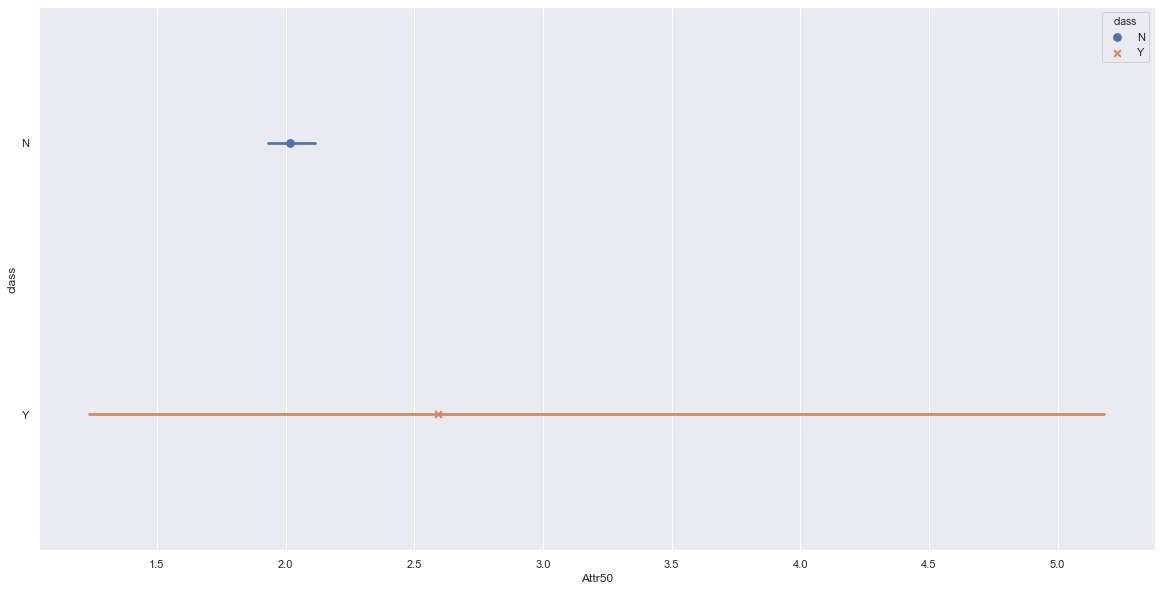
### C:\Users\drojasm\AppData\Local\Microsoft\Windows\INetCache\Content.MSO\54AE8F3C.tmp

Validando la métrica de rentabilidad basada en sus activos, permite denotar que las empresas que se encuentran en bancarrota, no generaban las utilidades necesarias y se estancan en cifras cercanas a 0, comparándolas con las liquidas son variables entre ganancias y pérdidas.

### Liquidez

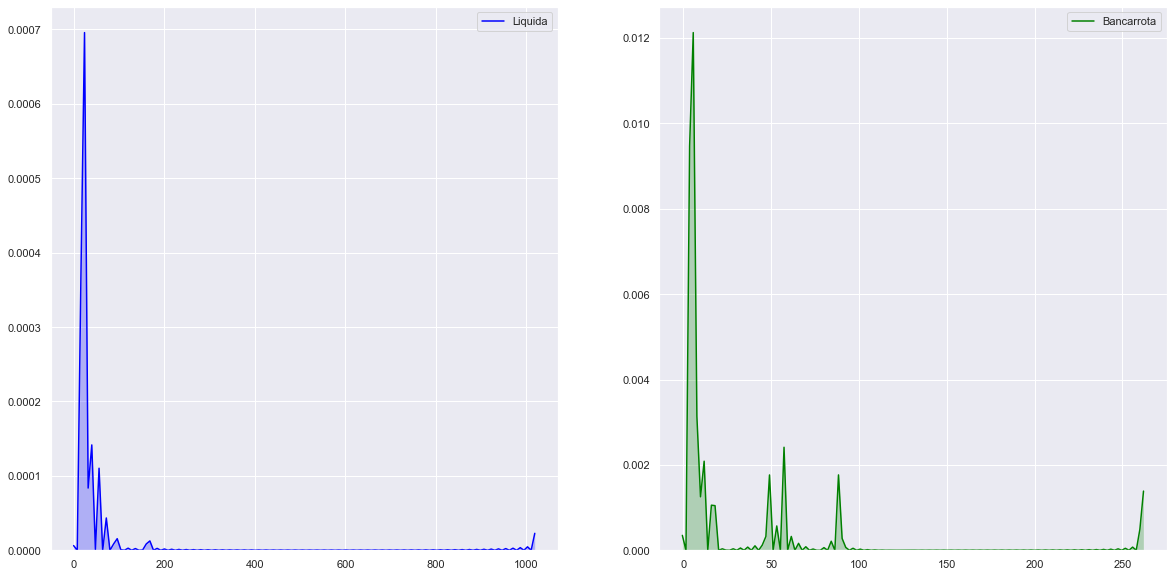
#### Radio de Liquidez - Activo circulante/pasivo circulante

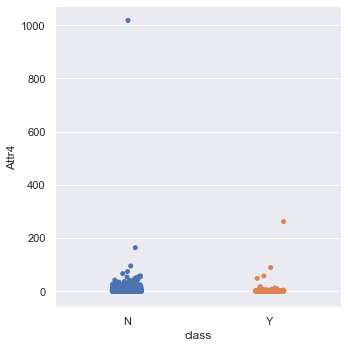




En el indicador de radio de liquidez, indica que si el comercio posee una métrica cercana a 0 posee más liquidez que las alejadas a 0, en particular del caso en análisis permite destacar que las empresas en bancarrota indican variabilidad en su liquidez mientras que las restantes se mantienen estables.

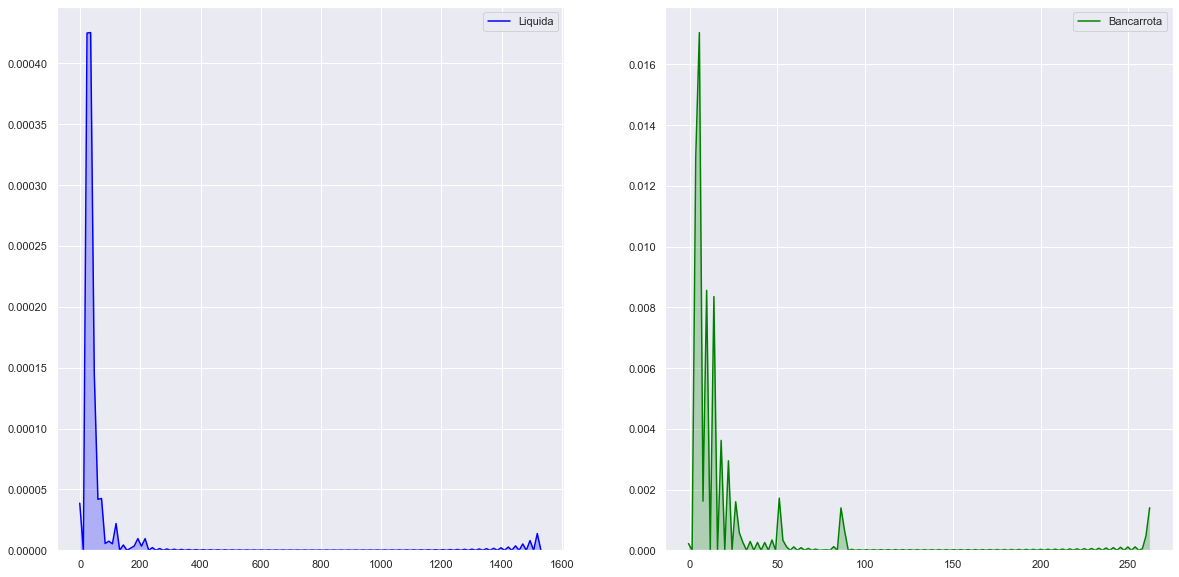
### Radio de Liquidez - Corto Plazo

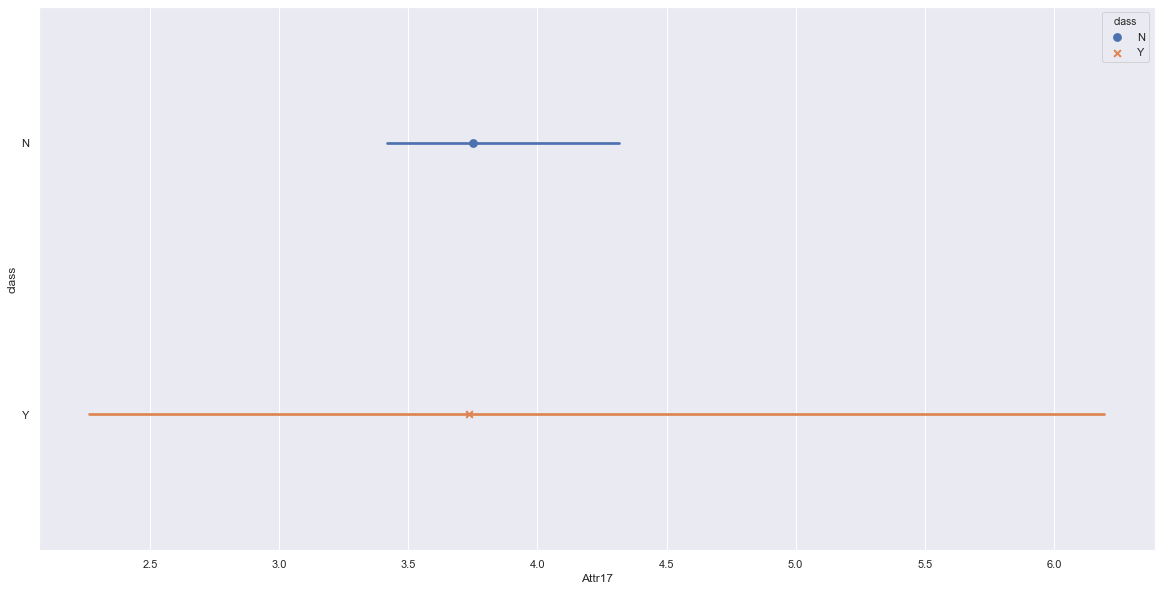




Manteniendo la métrica pero en este caso el contexto aplica a numéricamente en corto plazo, el comportamiento se mantiene, con mayor liquidez en empresas liquidas.

### Radio de Liquidez - Total

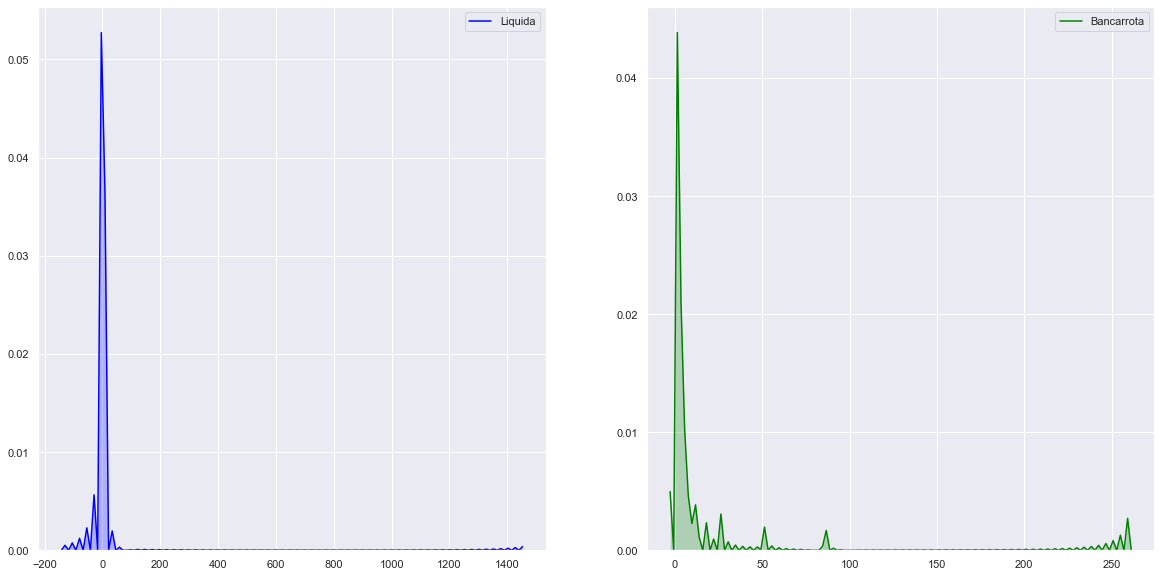


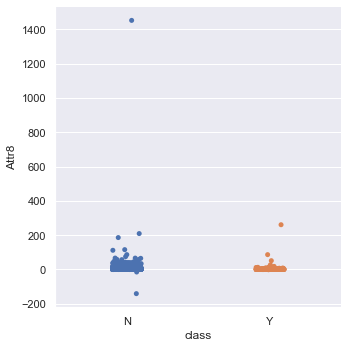


Finalmente, misma métrica, pero analizada al total de los activos y pasivos de la compañías, nos muestra que las empresas liquidas poseen más distribución positiva en relación a los contextos anteriores. Pero al compararlo con las empresas en bancarrota se denota una estabilidad importante en su liquidez.

### Solvencia

#### Razón de deuda - Capital contable/Pasivo total





Con respecto a la solvencia, se observa que las empresas liquidas poseen en promedio más activos disponibles para enfrentar sus deudas, que las empresas en bancarrota.

## Conclusiones

Recabando los indicadores en las distintas categorías, podemos indicar que el comportamiento de una empresa antes de quedar en quiebra es debido a erróneos manejos en su rentabilidad, liquidez y Solvencia. Estos aspectos se podrían profundizar aún más con un experto en finanzas y permitir así interactuar con otras metricas importantes para las organizaciones.

## Evaluación del Modelo

Luego de explorada la información, se llevó a cabo un proceso de Ciencia de datos con la finalidad de predecir con los datos suministrados si una empresa estará en bancarrota, modelando los escenarios por medio de 3 diferentes algoritmos (Random Forest, Support Vector Machine y ADOBoost).

Luego de desarrollado el modelo, por métricas de la librería sklearn metrics nos permitirá determinar el score del modelo en su aplicación. Y si es confiable para determinar si el cliente es elegible para el préstamo o no, según sus características. Además, por medio de Matrices de Confusión se evaluará la eficacia del modelo contra el conjunto de prueba para el modelo con mejores métricas.

|  |  |
| --- | --- |
| Modelo | Score |
| Random Forest | 0.99 |
| Support Vector Machine | 0.96 |
| AdaBoost | 0.95 |

Bajo esta perspectiva el rendimiento de Random Forest es ligeramente superior en su aplicación por lo que será el modelo elegido.

Al aplicar y predecir el set de certificación métricas de efectivas medidas por Rsquared y RMSE

|  |  |  |
| --- | --- | --- |
| Modelo | Accuracy | Kappa |
| Random Forest | 0.97 | 0.44 |

Las cuales son métricas positivas, permitiendo con una exactitud del 97% en la predicción de que una empresa este en bancarrota. Recordando los comentarios realizados al inicio en el cual se limitaba la investigación a el primero año, seria idóneo en la búsqueda de métricas de calidad incluir más información de años anteriores que permitan comparar los escenarios en periodos anteriores y validar así los comportamientos. Ya que con métricas del modelo del 0,97 se podría estar prediciendo en un contexto de overfitting.