With Recession around the corner and more layoffs in the market. I was researching on what metrics the companies/organizations go bankrupt.

Good bankruptcy prediction is important for any financial institution to make lending decisions to firms. Some of the co-corporate terminologies which will help identify bankruptcy predictions are **Net Income to Total Assets, Interest Coverage Ratio, Cash Flow to Liability, Retained Earnings to Total Assets, Total Asset Growth Rate, and Operating Profit Rate.**

The most important questions I wanted to answer are **“Relationship between Bankruptcy event and Net income to total assets”**

**“Relationship between Bankruptcy event to Total Assets****,” “Relationship between Bankruptcy event to Total Assets Growth Rate”** I will add more use cases as I start working on the data set further.

Many Organizations usually though indicate negative revenue, there could be a chance the overall assets valuation can go up, which could be a good indicator for liquidity financed to be guaranteed, we will have to see from further data analysis how this relation works out.

There are a few other parameters like Current interest payments, Gross returns before interest payment, and Final Gross after interest payment all the parameters are considered to qualify a firm eligible for financial loan.

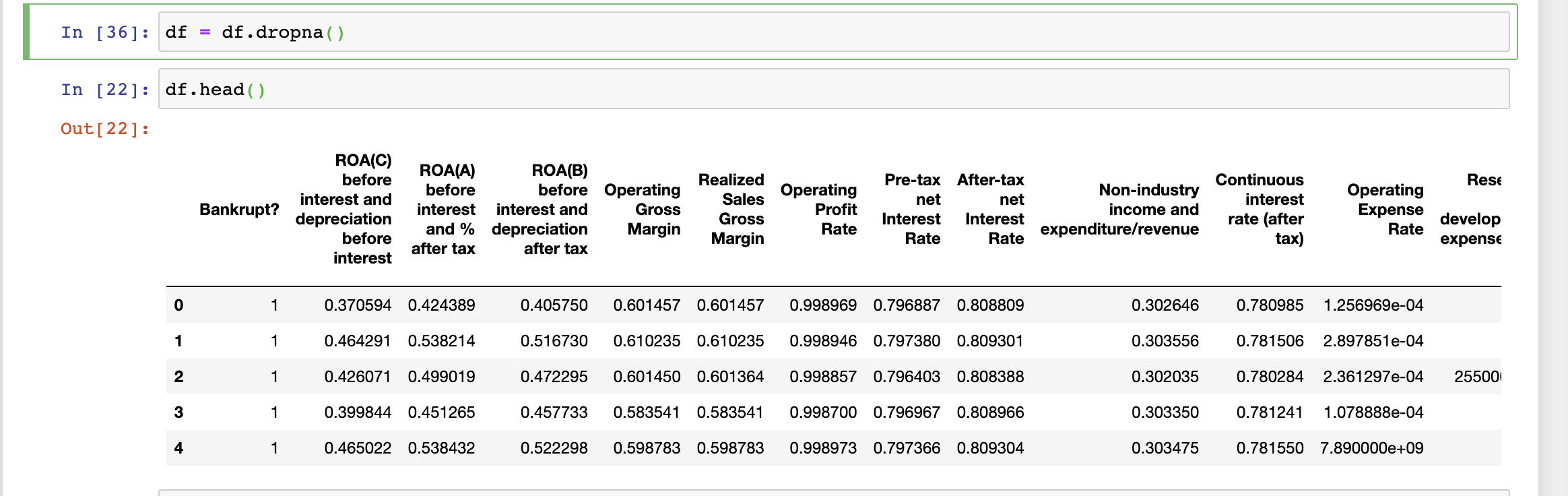
I am taking the data set from Kaggle: <https://www.kaggle.com/code/jiunkailee/company-bankruptcy-prediction-dt-rf-knn-nn/data>.

I should be able to answer the initial questions of **“Relationship between Bankruptcy event and Net income to total assets,” “Relationship between Bankruptcy event to Retained earnings to Total Asset,” and “Relationship between Bankruptcy event to Total Assets Growth Rate.”**

The initial visualization charts **Correlation matrix's, histograms and density plots** are used to observe the strength of relationships between bankruptcy attribute and other attributes.

**Preparing the Data:**

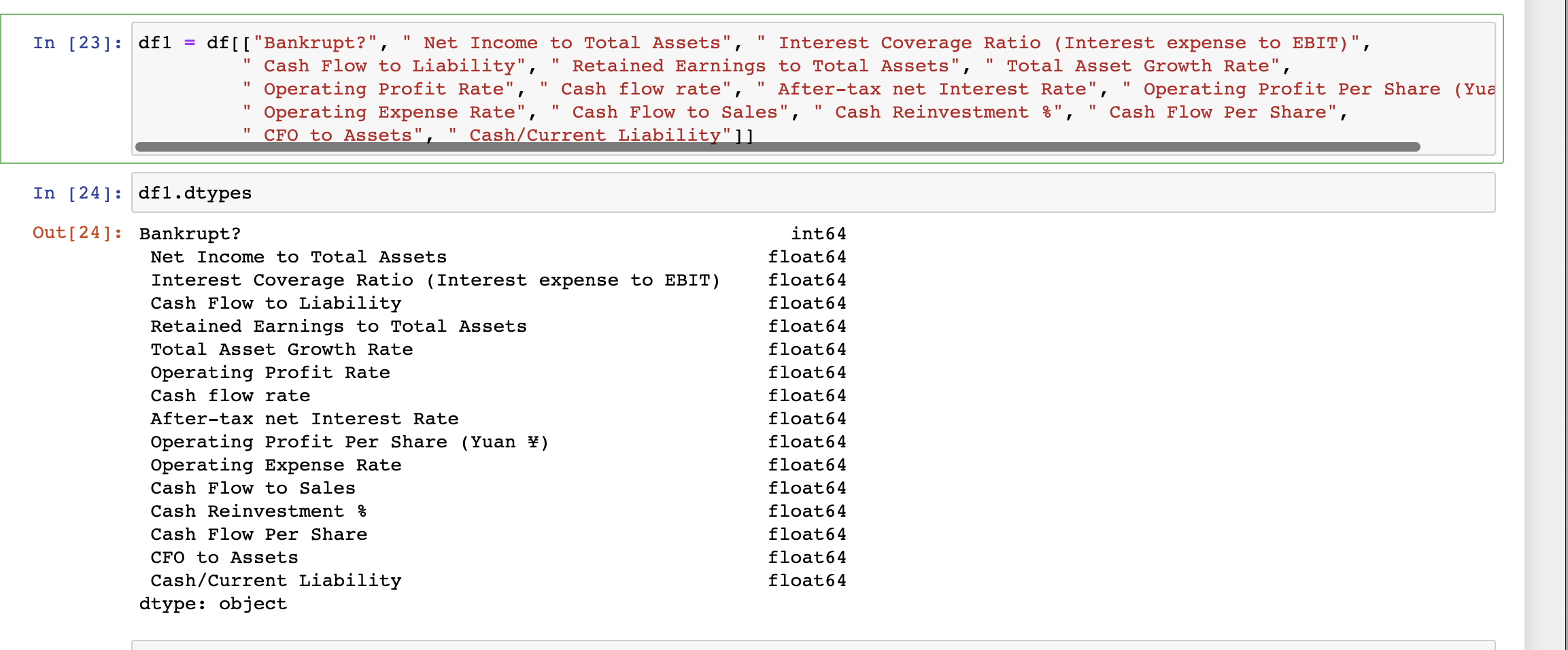
The Data is prepared by starting with removing Not Available attributes in DataFrame, split the train and test data sets, smote the variables to correct imbalanced data sets, Normalize the train data set.



**Adjustment of Initial Data Frame:**

The initial Data Frame is adjusted to include only required key attributes for analysis which are

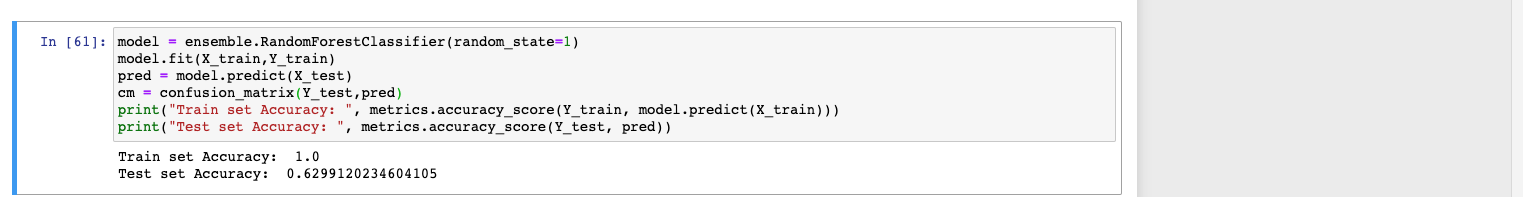
Bankrupt? int64  
 Net Income to Total Assets float64  
 Interest Coverage Ratio (Interest expense to EBIT) float64  
 Cash Flow to Liability float64  
 Retained Earnings to Total Assets float64  
 Total Asset Growth Rate float64  
 Operating Profit Rate float64  
 Cash flow rate float64  
 After-tax net Interest Rate float64  
 Operating Profit Per Share (Yuan ¥) float64  
 Operating Expense Rate float64  
 Cash Flow to Sales float64  
 Cash Reinvestment % float64  
 Cash Flow Per Share float64  
 CFO to Assets float64  
 Cash/Current Liability float64



**Models used for Evaluation:**

The planned models used for Evaluation will be Random Forest, Decision Tree Classifier and K Neighbors Classifier to check on Test and Train data sets.

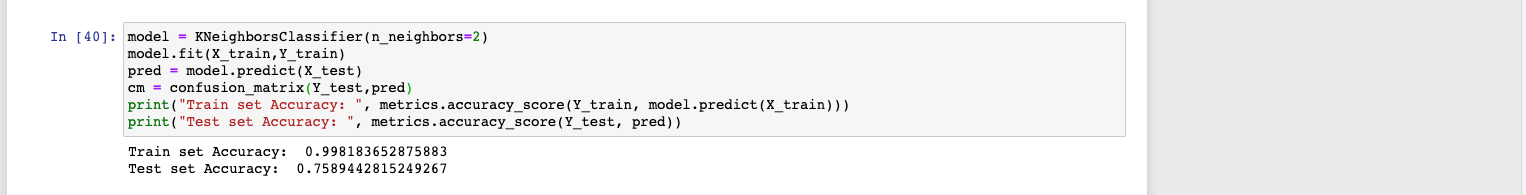
Below are the model evaluations of Random Forest, Decision Tree and KNN classifier on test and train data sets



The Random Forest Classifier has

Train set Accuracy: 1.0

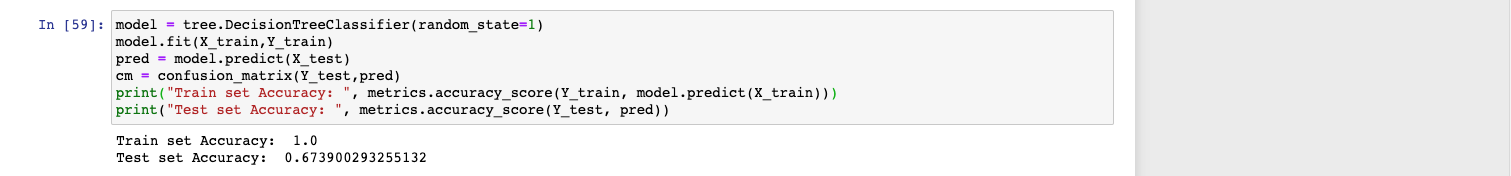
Test set Accuracy: 0.6



The KNeighbors Classifiers have

Train set Accuracy: 0.99

Test set Accuracy: 0.75



The DecisionTree Classifier has

Train set Accuracy: 1.0

Test set Accuracy: 0.67

From the comparison of the models, K neighbor classifier has the most accuracy with a K value of 2, test set accuracy up to 75%

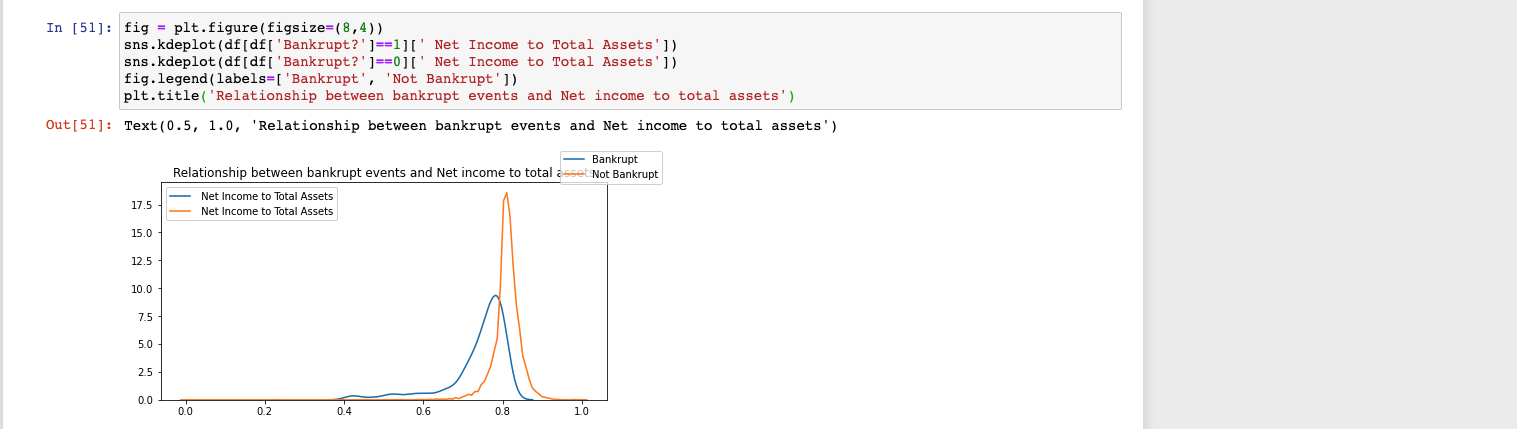
**Results Finalization:**

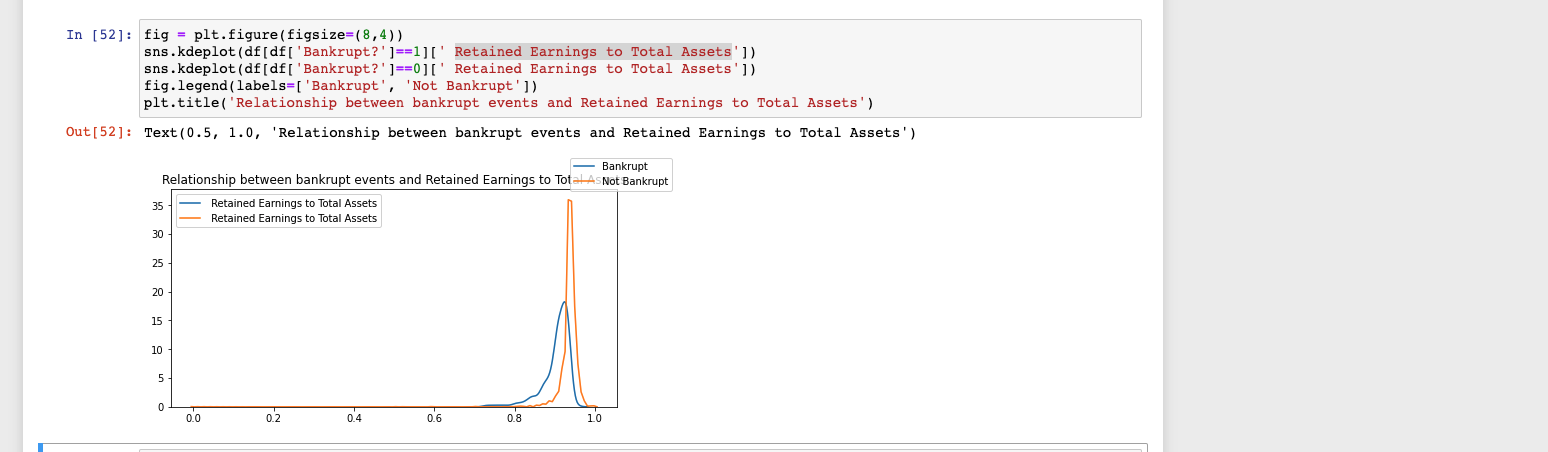
The feature importance of variables from the model KNN according to the order of priority of variables are

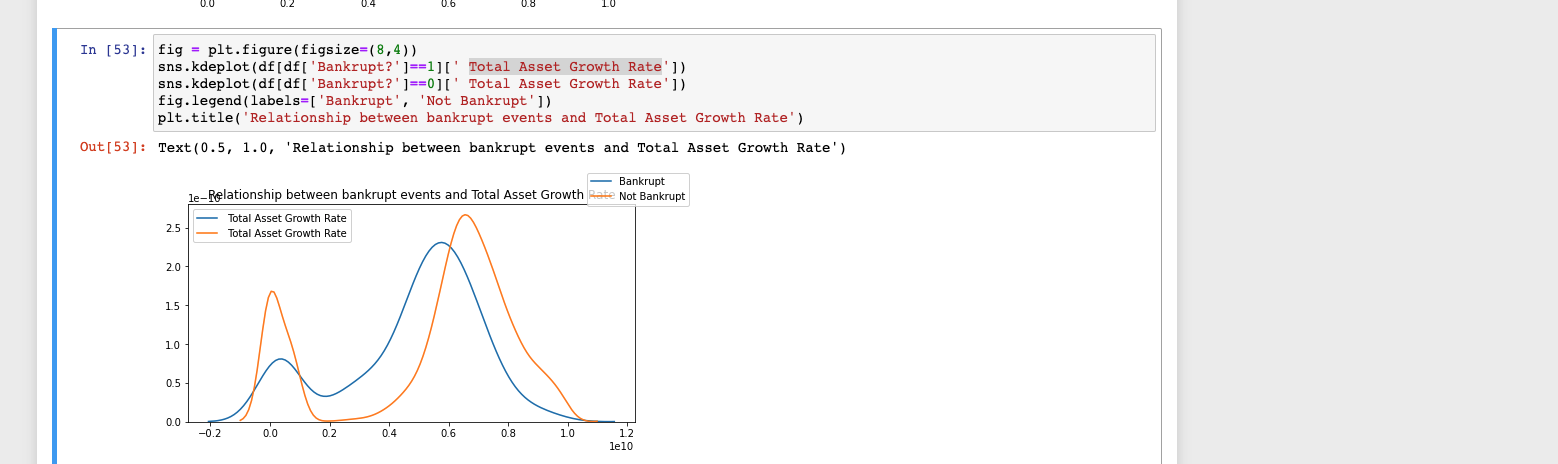
**Net Income to Total Assets, Retained Earnings to Total Assets, After-tax net Interest Rate, Interest Coverage Ratio (Interest expense to EBIT) , Operating Profit Rate and Cash Flow to Liability have direct impact on Bankruptcy event and should be considered while making decision choice of sanctioning loans**

**After the models are trained and tested, I analyzed the Feature importance of variables to understand their significance with bankruptcy events**

**Some features Net Income to Total Assets, Retained Earnings to Total Assets and Total Asset Growth to understand their importance in determining the bankruptcy event**







**Net Income to Total Assets growth will decrease bankruptcy chance**

**Retained earnings to Total Assests will decrease bankruptcy chance**

**After Tax net interest rate has direct impact on bankruptcy event.**

**There is no straightforward relationship between Bankrupt and Total Assests growth rate**

**There is no straightforward relationship between Bankrupt and Total Assests growth rate**

**Model Deployment:**

**Once the Model is tuned further as per business requirements, I should be able to deploy it to cloud infrastructure as microservice-based applications servicing user inputs and generating predictions accordingly.**

**Risks:**

I feel good the Data Set used from Kaggle is good, with results to be clear and precise currently, I do not. See any risks.

**Reference Link:**

https://www.sciencedirect.com/science/article/abs/pii/S0377221716000412

Kaggle Data Sources:

<https://www.kaggle.com/code/jiunkailee/company-bankruptcy-prediction-dt-rf-knn-nn/data>

Additional Data Source:

<https://www.kaggle.com/code/edacebeci/predict-bankruptcy/data>

https://scholarworks.lib.csusb.edu/cgi/viewcontent.cgi?article=5256&context=etd-project