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.

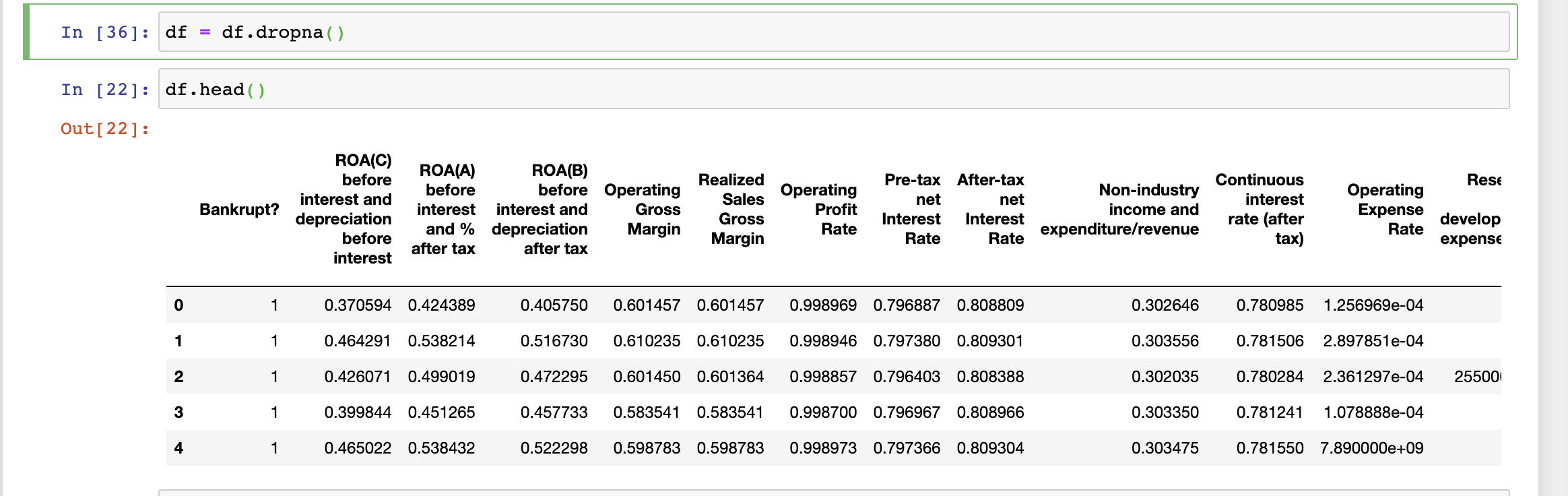
I am taking the data set from Kaggle: <https://www.kaggle.com/code/jiunkailee/company-bankruptcy-prediction-dt-rf-knn-nn/data>.

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 Total Assets,” 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:**

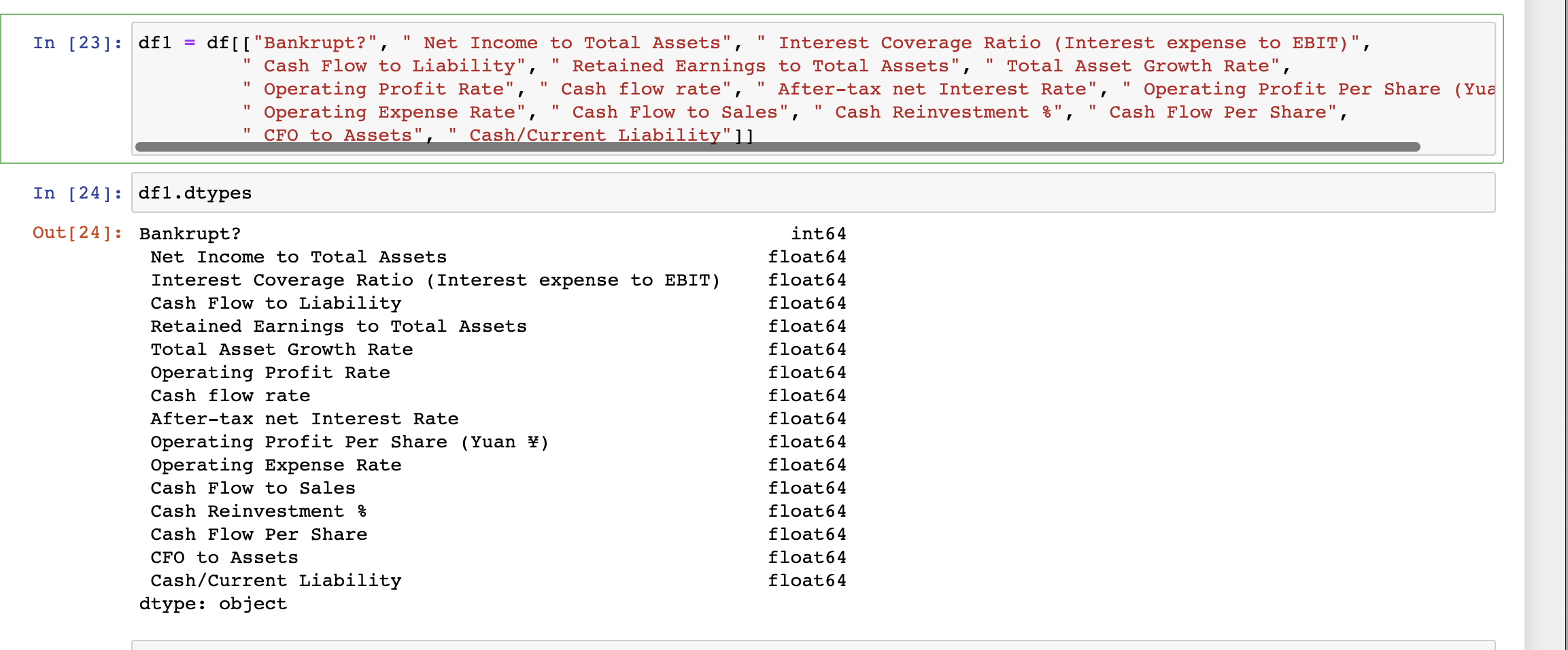
The Data is prepared by starting with removing Not Available attributes in DataFrame, split the train and test data sets, smote the variables to correct imbalanced data sets, Normalize the train data set.



**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 Logistic Regression, Decision Tree Classifier and K Neighbors Classifier to check on Test and Train data sets.

I will compare the results of each model and see which model provides me with more accurate results and select one model for final analysis.

**Results Finalization:**

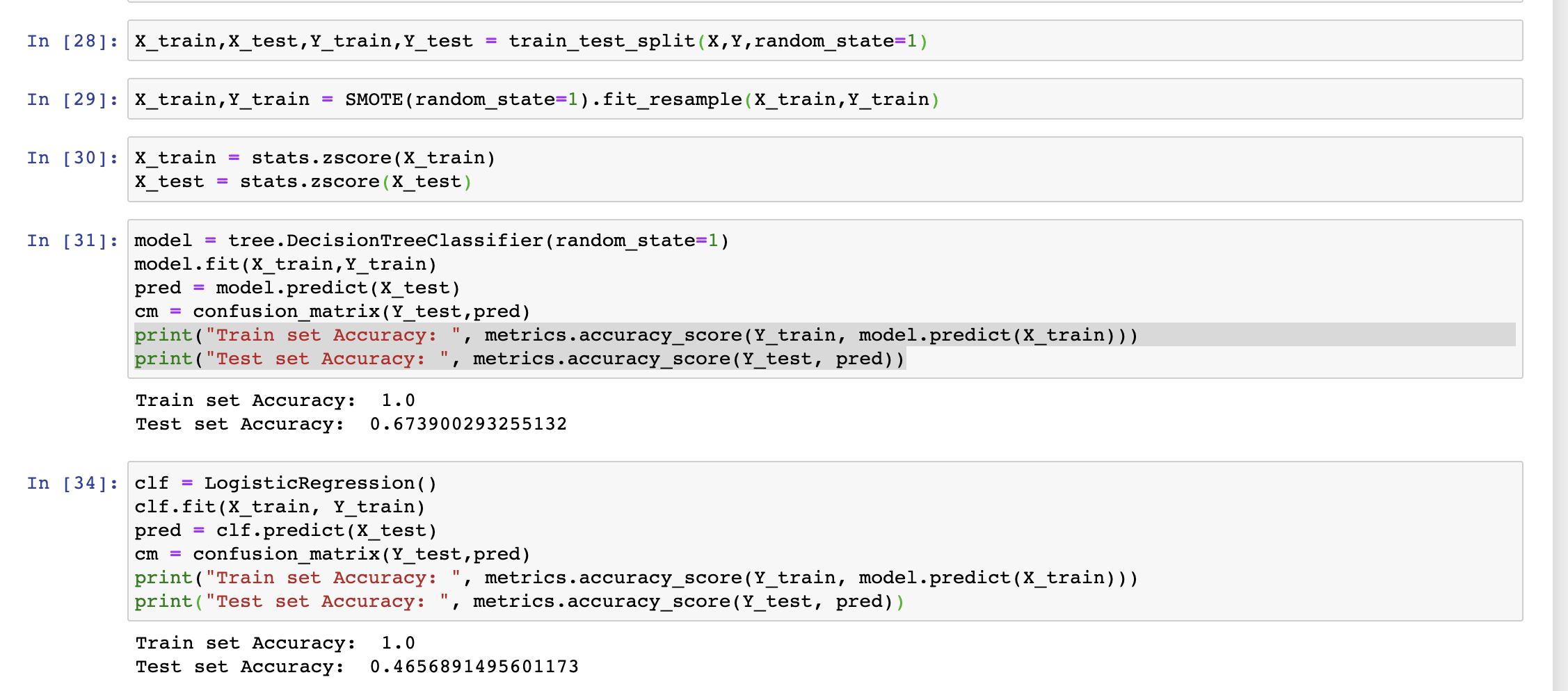
**Split the data to train and test and calculating the accuracy of the test data set.**

I evaluated the models using Logistic Regression, Decision Tree Classifier and K NeighborsClassifier

For Logistic Regression, test set accuracy came up to 0.465 around 46% accuracy

For Decision Tree Classifier test set accuracy came up to 0.673 which is around 67%

For K neighborsClassifier() test set accuracy came up to 0.998 around 99%.



From the above analysis, I find K Neighbors Classifiers as more accurate than other models. I am in the final process of plotting the analysis between different attributes of the data set Current Liability to Operating Expenses, Cash Flow Rate to Operating Profit Rate. I will add up few more use scenarions to explain the results by final project week submission.

But from the initial analysis it is revealed the following observations:

Current Liability will increase if Operating profit decreases which increases bankruptcy rate.

Cash Flow to sale increases the Current Liability will decrease decreasing Bankruptcy Rate.

**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>