Capstone Project - 3

Company Bankruptcy Prediction

Shantanu Houzwala

Index

Title
❖ Introduction ❖ Problem Statement ❖ Understanding the Data ❖ Exploratory Data Analysis on Features ❖ Models Used ❖ Which model did I choose and why? ❖ Challenges faced ❖ Conclusion

Introduction

- A company faces bankruptcy when they are unable to pay off their debts. The Taiwan Economic Journal for the years 1999 to 2009 has listed the details of company bankruptcy based on the business regulations of the Taiwan Stock Exchange.
- The Taiwan Stock Exchange was established in 1961 and began operating as a stock exchange on 9 February 1962. It is a financial institution located in Taipei, Taiwan. It has over 900 listed companies.
- The data includes a majority of numerical attributes that help understand the possibility of bankruptcy.
- This project aims at analyzing the possibility of whether an organization would face bankruptcy.

Problem Statement

- Prediction of bankruptcy is a phenomenon of increasing interest to firms who stand to lose money because on unpaid debts. Since computers can store huge dataset pertaining to bankruptcy making accurate predictions from them before hand is becoming important.

- The main objective of this project is to use various classification algorithms on bankruptcy dataset to predict bankruptcies with satisfying accuracies long before the actual event.

Understanding the Data

Updated column names and description to make the data easier to understand (Y = Output feature, X = Input features)

Y - Bankrupt?: Class label 1 : Yes , 0: No

X1 - ROA(C) before interest and depreciation before interest: Return On Total Assets(C)

X2 - ROA(A) before interest and % after tax: Return On Total Assets(A)

X3 - ROA(B) before interest and depreciation after tax: Return On Total Assets(B)

X4 - Debt ratio %: Liability/Total Assets

X5 - Liability-Assets Flag: 1 if Total Liability exceeds Total Assets, 0 otherwise

X6 - Net Income Flag: 1 if Net Income is Negative for the last two years, 0 otherwise

X7 - Borrowing dependency: Cost of Interest-bearing Debt

X8 - Current Liability to Assets

X9 - Current Liabilities/Equity

X10 - Current Liability to Assets

X11 - Current Liability to Current Assets

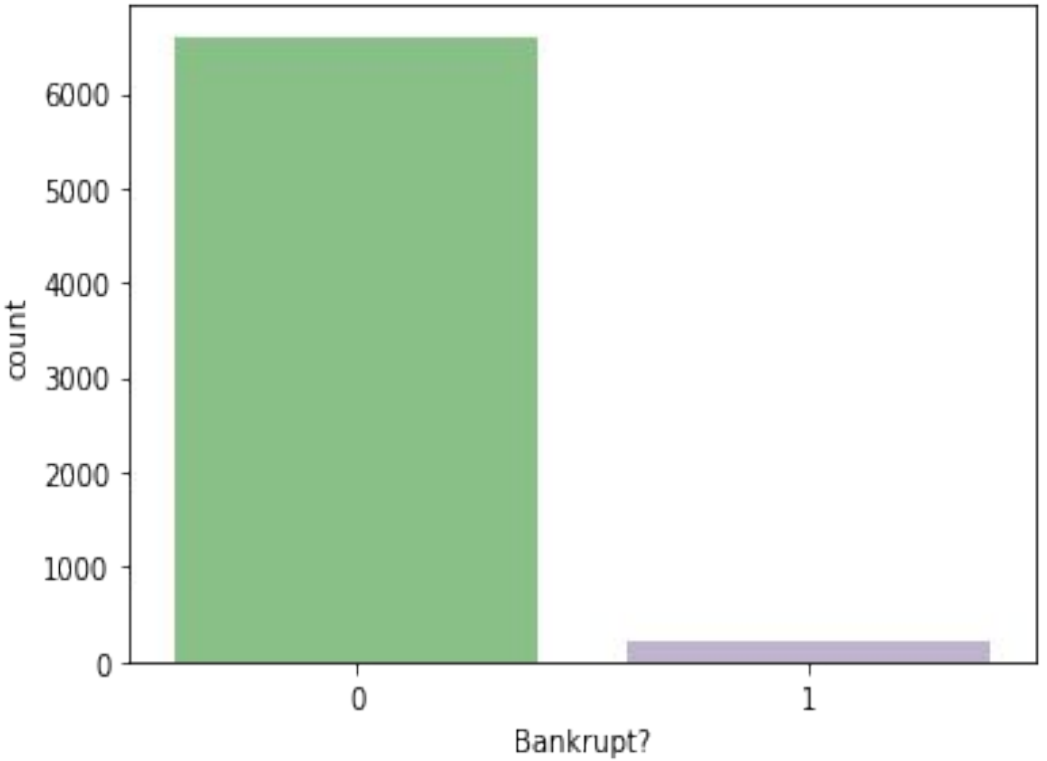
X12 - Liability to Equity

Exploratory Data Analysis (EDA) on Features

Finding number of Bankrupt Companies

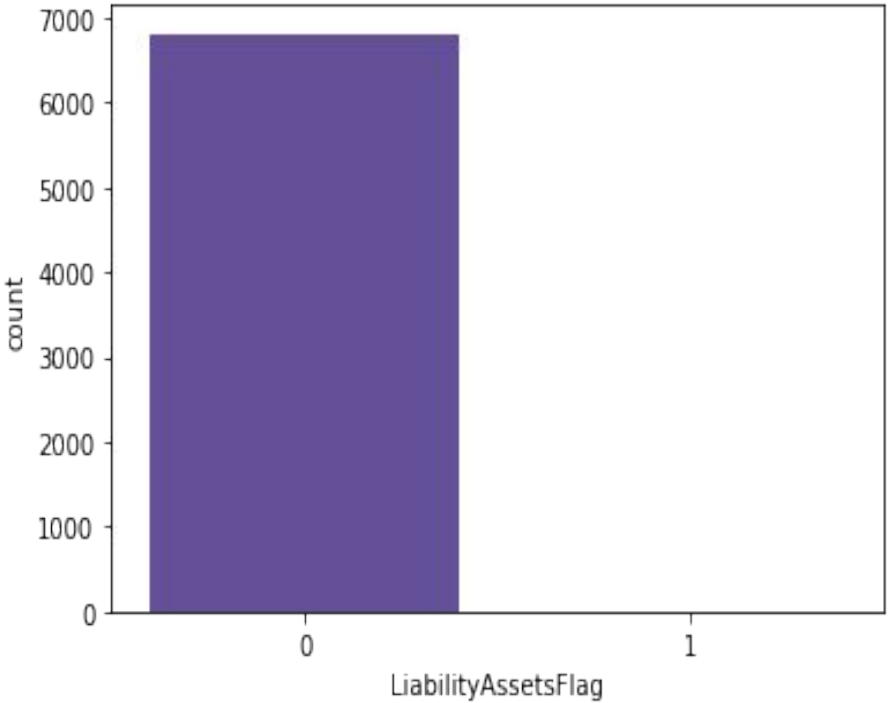
- **The number of organizations that have gone bankrupt in 10 years between 1999 – 2000 is few.**

- **In the chart Bankrupt is denoted by Class label where**
  1 : Yes
  0: No



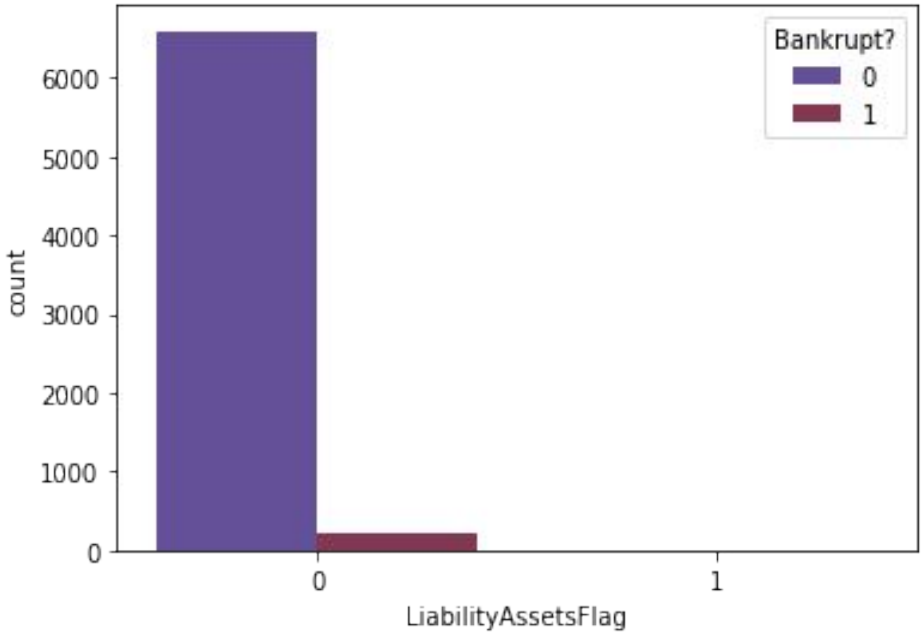
Counting total number of liability-assets flag

- The "Liability-Assets" flag denotes the status of an organization, where if the total liability exceeds total assets, the flagged value will be 1, else the value is 0.
- A majority number of times, organizations/company's assets are more than their liabilities.
- This is a good sign as this reduces the possibility of bankruptcy of a company.



Finding correlation between bankrupt and liability-assets flag features.

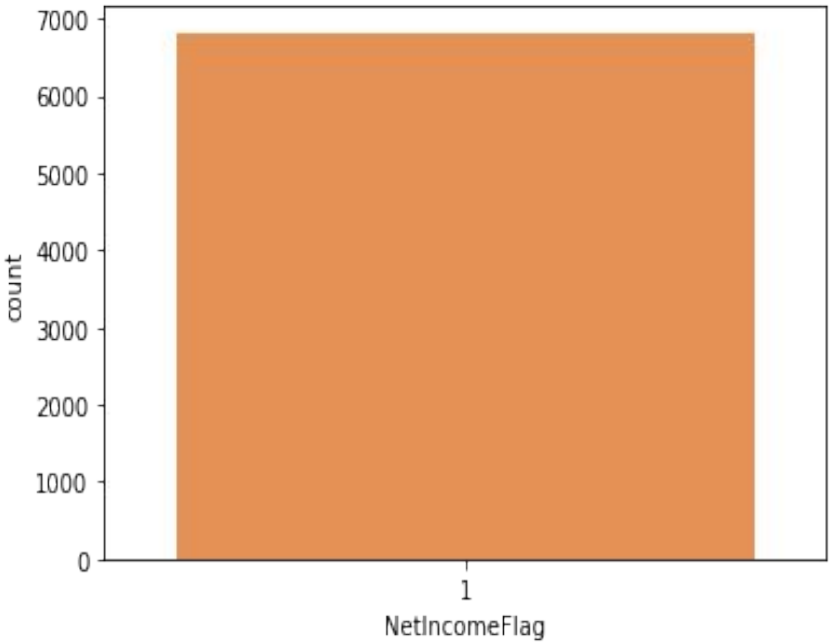
- A small portion of organizations suffers bankruptcy, although possessing more assets than their liabilities.
- Several companies possess many assets, which is always a good sign for an organization.
- However, an organization cannot guarantee not being bankrupt, although owning several assets.



Counting total number of Net Income Flag.

- The "Net Income" flag denotes the status of an organization's income in the last two years, where if the net income is negative for the past two years, the flagged value will be 1, else the value is 0.

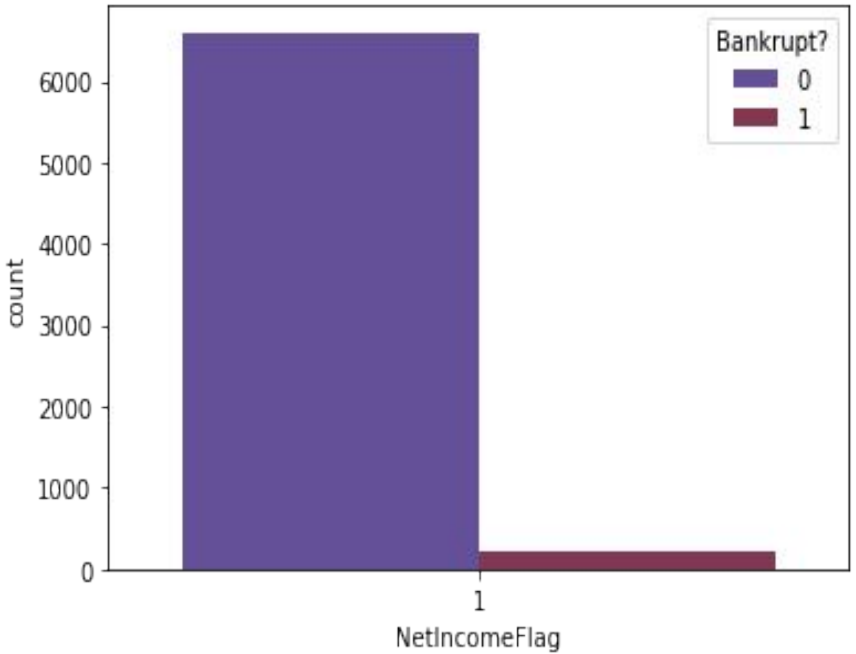
- We observe that all the records have been exhibiting a loss for the past two years.



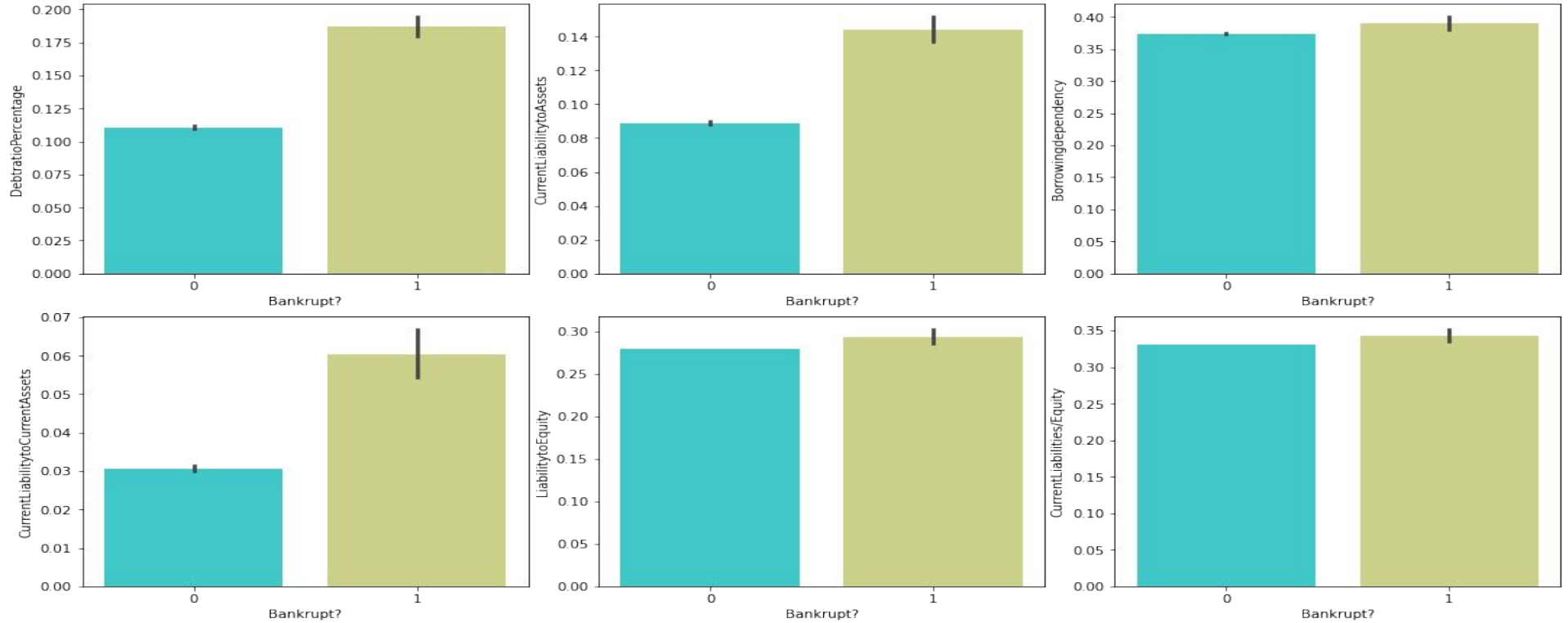
Finding correlation between bankrupt and Net Income Flag features

- **Very few of the organizations that have had negative income in the past two years suffer from bankruptcy.**

- **Many organizations that have suffered losses for the past two years have stabilized their business, thus avoiding bankruptcy**

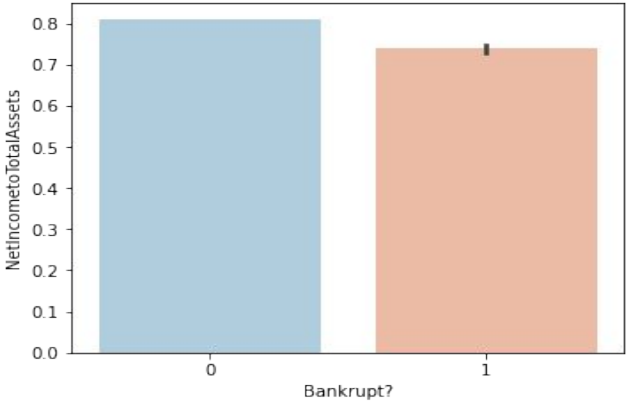


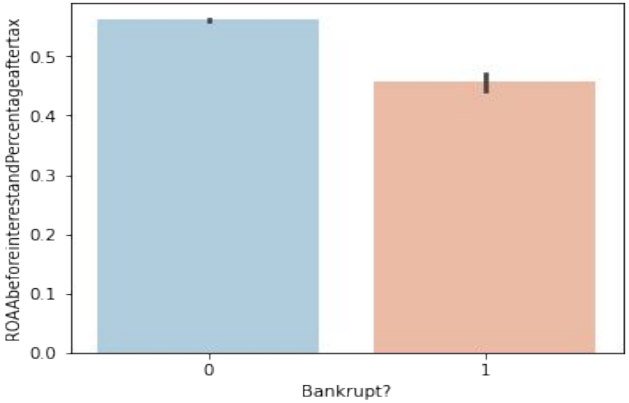
Analyzing top six positively correlated attributes

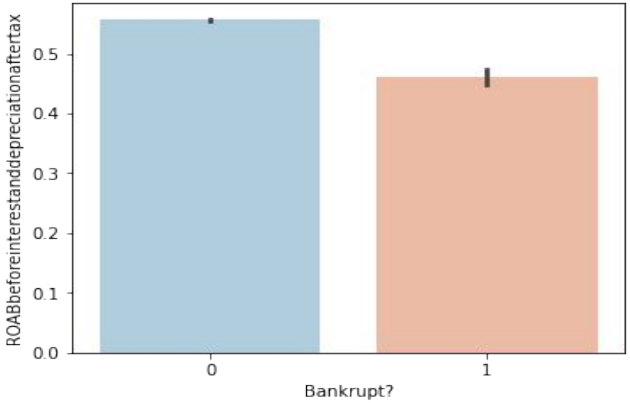


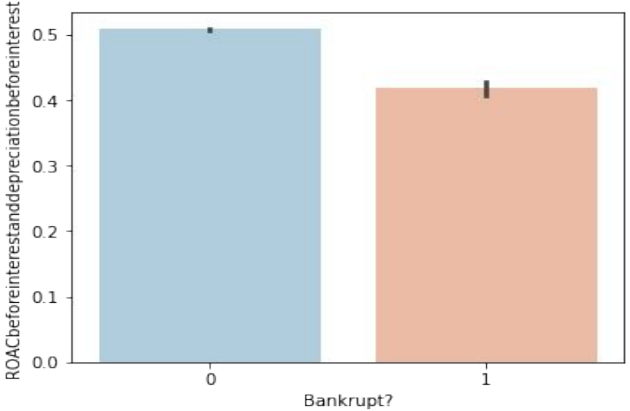
- **We see that three attributes - "Debt Ratio %, Current Liability To Assets, Current Liability To Current Assets" are commonly high in bankrupt organizations.**

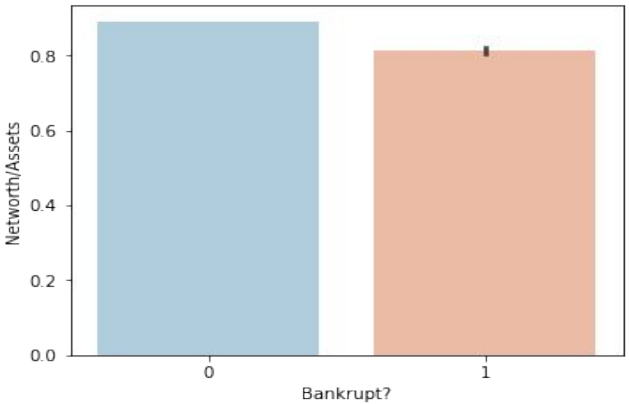
Analyzing top six negatively correlated attributes

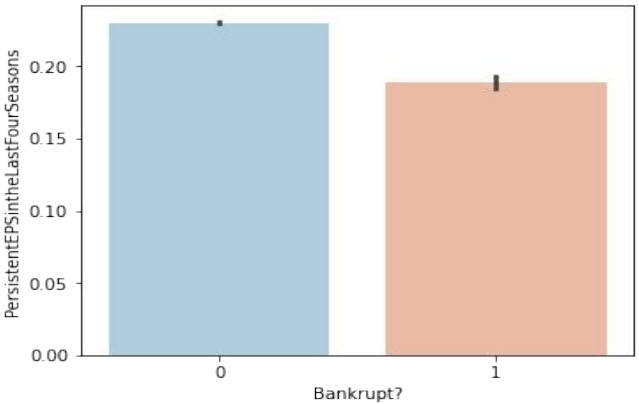






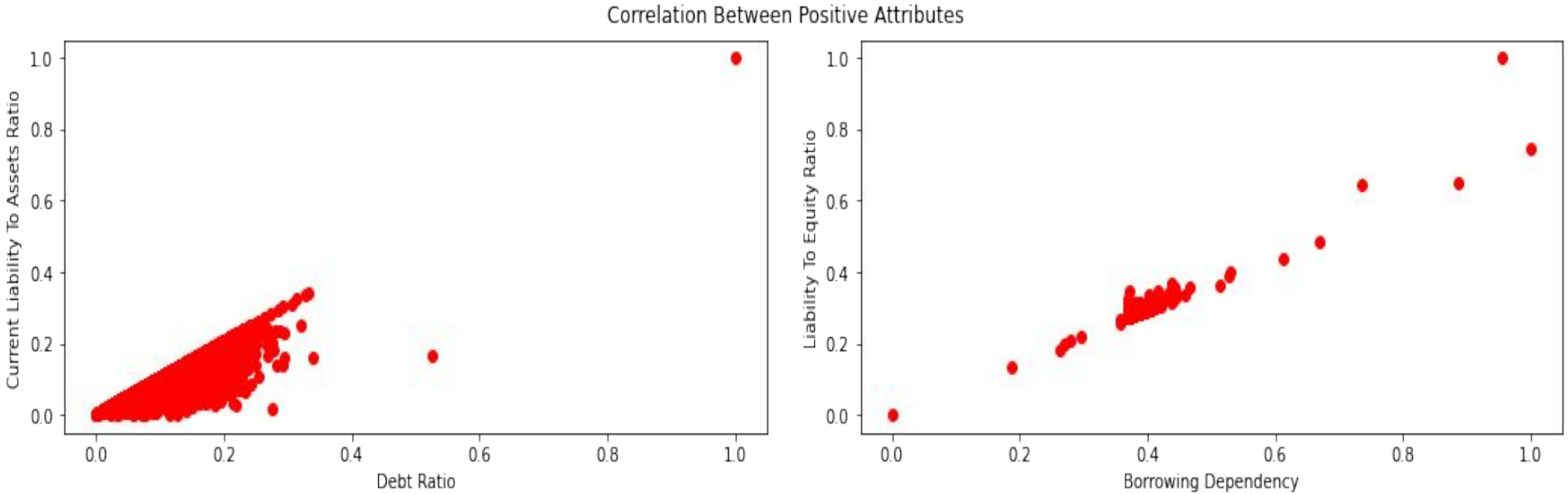






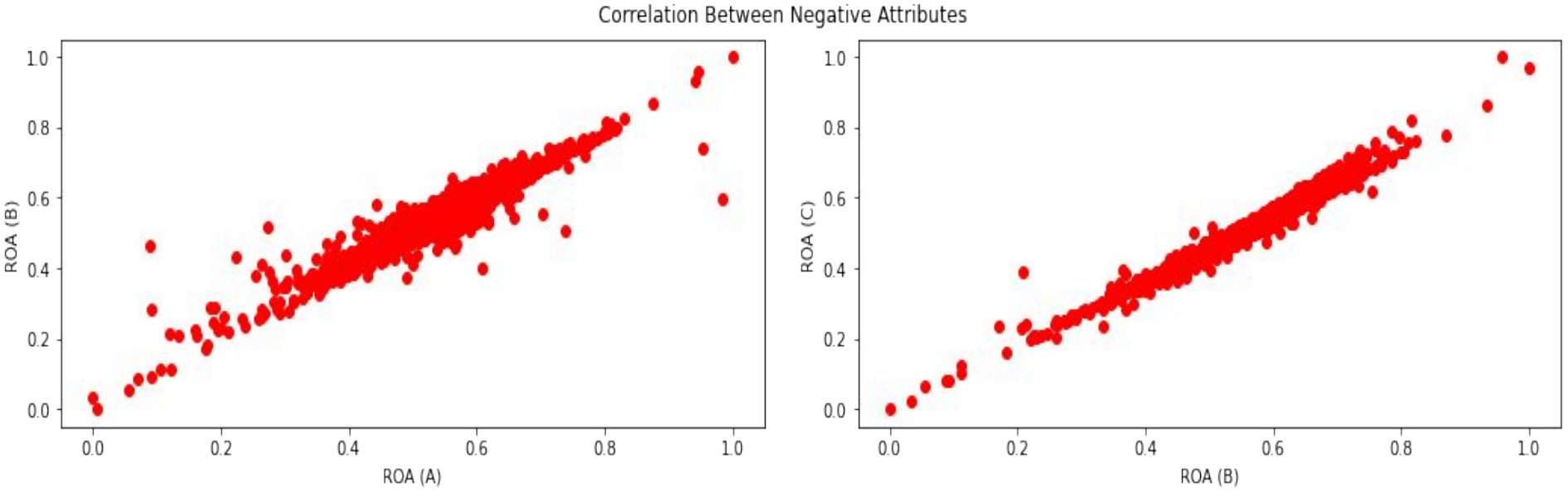
- **These attributes show us that the more the assets and earning of a company, the less likely is the organization to be bankrupt.**

Finding correlation between positive attributes



- **There is a positive relation between attributes that have a high correlation with the target attribute.**

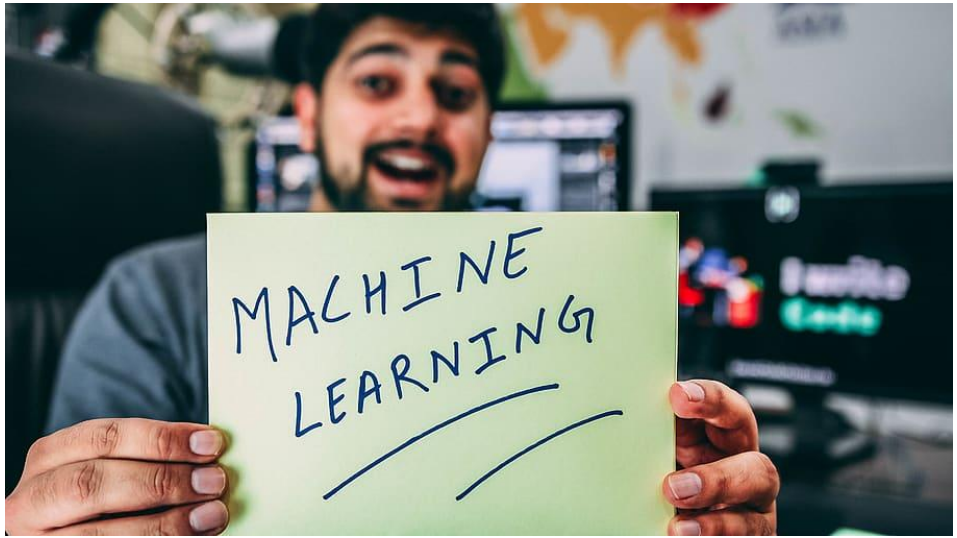
Finding correlation between negative attributes



- **There is a positive relation between attributes that have a low correlation with the target attribute.**

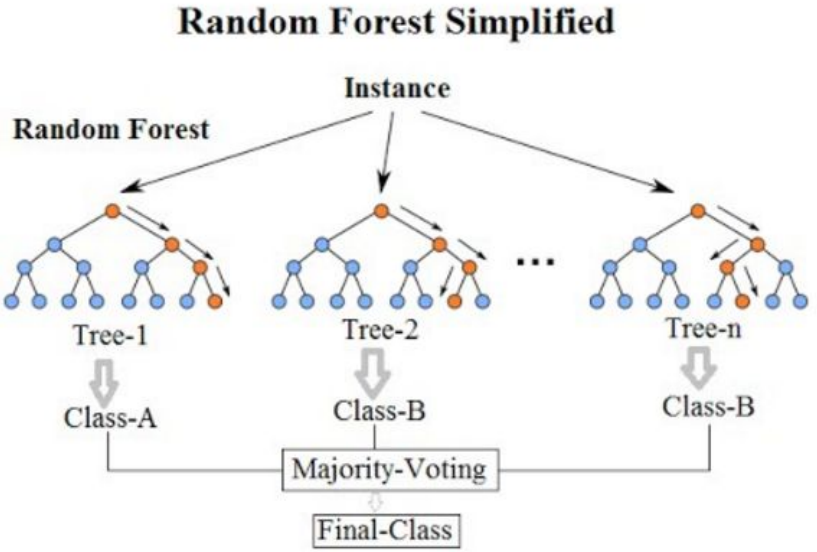
Models Used

1. Random Forest Classifier
2. K Nearest Neighbour
3. Support Vector Classifier
4. Logistic Regressor
5. Decision Tree Classifier



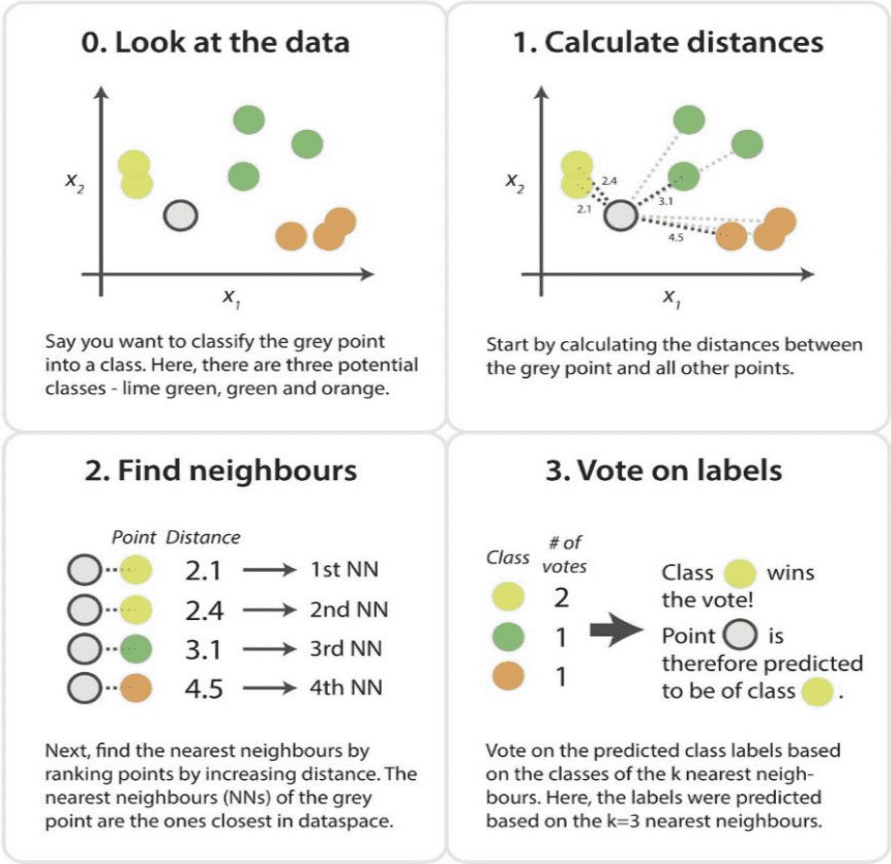
Random Forest Classifier

Model Score	Precision	Recall	F1 score	ROC-AUC score
98.9%	0.81	0.92	0.86	0.96



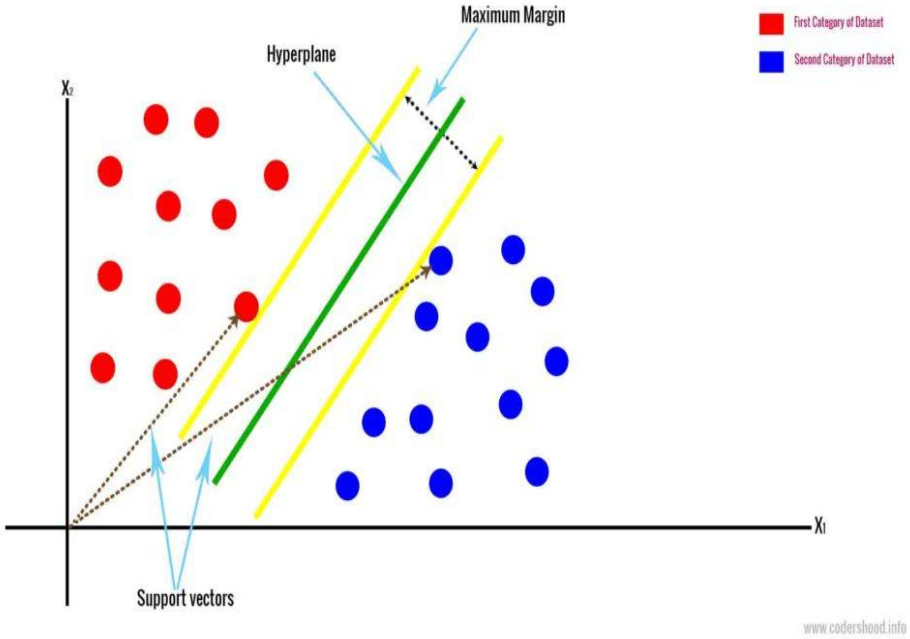
K Nearest Neighbour

Model Score	Precision	Recall	F1 score	ROC-AUC Score
98.02%	0.67	0.94	0.78	0.96



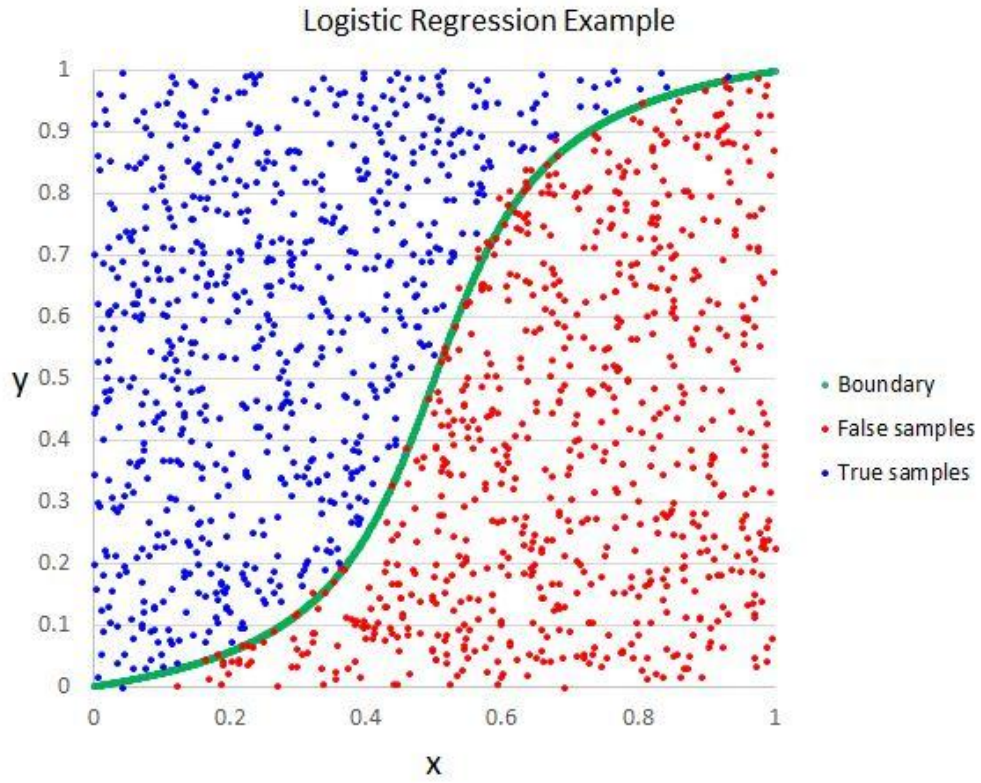
Support Vector Classifier

Model Score	Precision	Recall	F1 score	ROC-AUC Score
94.43%	0.40	0.94	0.56	0.94



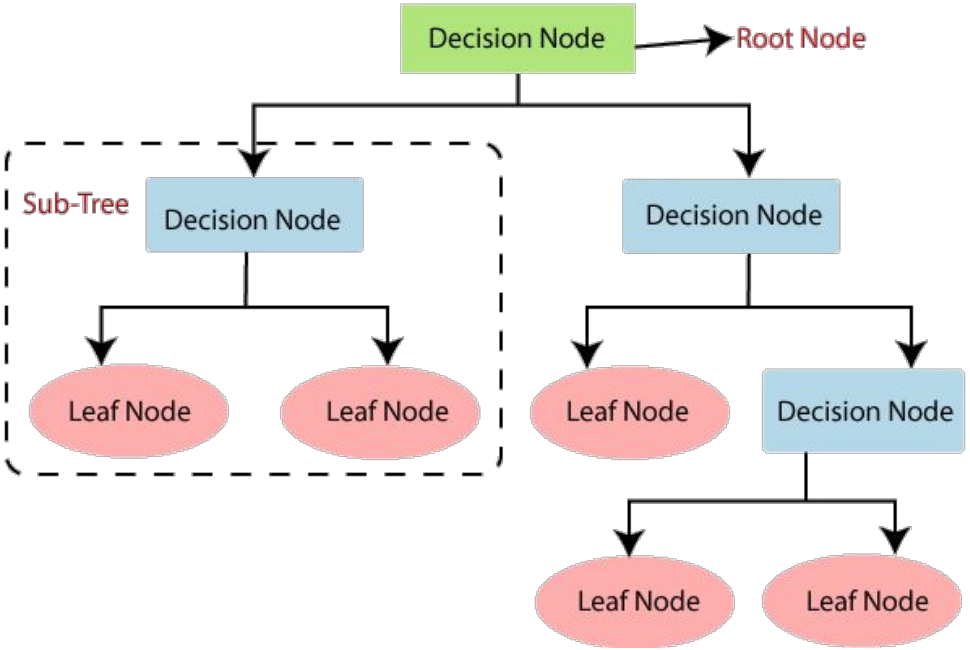
Logistic Regressor

Model Score	Precision	Recall	F1 score	ROC-AUC Score
88.86%	0.23	0.82	0.36	0.86



Decision Tree Classifier

Model Score	Precision	Recall	F1 score	ROC-AUC score
84.68%	0.18	0.90	0.31	0.87



Which Model did I choose and Why?

- **Out of all the models “Random Forest Classifier” and “K Nearest Neighbour” are the best performers. We choose F-1 Score as the deciding factor.**

- **F1-Score is an evaluation metric, that is used to express the performance of the machine learning model (or classifier). It gives the combined information about the precision and recall of a model. This means a high F1-score indicates a high value for both recall and precision.**

- **Generally, F1-score is used when we need to compare two or more machine learning algorithms for the same data. Thus, we opt for the algorithm whose f1 score is higher.**

Challenges Faced

- Dealing with imbalanced dataset the right way
- Finding the best attributes to work with through feature selection
- Choosing right model to get best results.
- Understanding the financial concepts related to bankruptcy

Conclusion

- An organization cannot guarantee not being bankrupt, although owning several assets.
- Most of the organizations in the dataset are running into losses for the past two years as their net income poses to be negative.
- An increase in the values of the attributes “Debt Ratio %, Current Liability To Assets, Current Liability To Current Assets” causes an organization to suffer heavy losses, thus resulting in bankruptcy.
- There seems to be a relation between attributes that have a high correlation with the target attribute and a low correlation with the target attribute.
- Random Forest Classifier and K Nearest Neighbour Algorithms gives the highest F-1 score and accuracy of 98%.
- Thus the main objective of this project, to analyze the dataset and building a predictive model to predict bankruptcy of companies is successfully achieved.

Thank You