**Capstone Project Submission Company Bankruptcy Prediction**

|  |
| --- |
| **Summary -** |
| 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 data were collected from the Taiwan Economic Journal for the years 1999 to 2009. Company bankruptcy was defined based on the business regulations of the Taiwan Stock Exchange.  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.  This dataset didn’t have any missing values, this saved some time. Then I moved on to EDA where I plotted the target feature bankruptcy which was heavily imbalance and then plotted positively and negatively correlated features with bankruptcy.  Now comes the Data Preparation where the target feature was imbalanced so I had to implement SMOTE Synthetic Minority Oversampling Technique (SMOTE) which is a statistical technique for increasing the number of cases in your dataset in a balanced way.  Further on scaling data and train test spilt is done to end the Data Preparation.  In this project, total 5 classification models are used which are Logistic Regression, Support Vector Classifier, Decision Trees, Random Forest and XGBoost. After hyperparameter tuning with GridSearch it was found that XGBoost is the optimal model for this project.  To increase the model performance, PCA was also implemented. PCA slightly increased the score but not worth the time took for fitting PCA. One can skip dimension reduction if the gain is tiny. At last, as per the F1 score XGBoost was the best choice for this project.  This was all about my analysis and training that I did, and based on result and my inferences I make the following conclusion:   * There are many attributes that play important roles to decide whether a company will go bankrupt or not. * Net Income Flag plot showed us that most of the companies are running into Losses for the past 2 years. * There are high chances that a company can go Bankrupt if the attributes “Debt Ratio %, Current Liability To Assets, Current Liability To Current Assets” are high. * The best performing model is XGBoost by considering the F1 score which is an ideal metric to choose for a classification model. |
| **Contributor Role** |
| Ritik Vaidande ([vr171k@gmail.com](mailto:vr171k@gmail.com))   * Understanding Data   Understanding different column  Having overview of data   * Exploratory Data Analysis * Target feature analysis * Positively correlated features with bankruptcy * Negatively correlated features with bankruptcy * Visualizations * Bar graphs and Boxplot * Data Preparation * Resampling Data * Modelling * Logistic Regression * Decision Tree * Random Forest * Support Vector Classifier * XGBoost * Conclusions |
| **GitHub Repo and Drive Link** |
| Github Link:-  <https://github.com/vaidande/Company-Bankruptcy-Prediction>  Drive Link:-  https://drive.google.com/drive/folders/1Lzvxtw-hKOsHAhmJG7vx\_xpWN5Gy5Hoy?usp=share\_link |