Capstone Project Submission

Project Summary:

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.

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.

As a first step I performed Exploratory Data Analysis (EDA) on the features of the dataset and found correlations between the target variable i.e. 'Bankrupt' column and other columns.

Based on the findings from EDA, I selected few important features which are positively and negatively correlated with our target variable. This helps in building the predictive classification model.

Since the dataset is highly imbalanced, through "Stratified K Fold Cross-Validation" I have distributed the 80% training set into further training and testing splits.
Since we are dealing with over 50 features, we use "Randomized Search Cross-Validation" as this technique proves to perform better with many features.
After this I performed feature selection with help of "SelectKBest" function which is used to add the most value to the target variable.

Lastly, I did the Comparison of different classification models for final selection of the Model and found that the best performing models for this dataset are “Random Forest Classifier” and “K Nearest Neighbour” on the basis of f-1 score and with 98% accuracy rate.

Name, Email and Tasks Accomplished:

Name: Shantanu Houzwala

Email: shantanu.houzwala10@gmail.com

Tasks Accomplished:

1. Importing Libraries
2. Loading the Dataset
3. EDA on features
4. Data Modeling
5. Comparison of Models
6. Feature selection

Github Link: https://github.com/Shantanuh10/Bankruptcy-Prediction