**Capstone Project Submission**

**Instructions:**

i) Please fill in all the required information.

ii) Avoid grammatical errors.

|  |
| --- |
| **Team Member’s Name, Email and Contribution:** |
| 1. Turab Ahmed ([thurab455@gmail.com](mailto:thurab455@gmail.com))   The followings are contributed on the Company Bankruptcy Prediction,   * Performed EDA. * Handled Class Imbalance. * Checked Multicollinearity. * Trained 80% and Tested 20% with the data. * Applied various models to the training and testing data to predict Bankruptcy. * Applied Grid Cross Validation to figure out the learning rate and to improve the Recall score for the model |
| **Please paste the GitHub Repo link.** |
| Turab’s Github Link: - [GitHub Thurab/Company\_Bankruptcy\_prediction\_Classification\_usecase](https://github.com/Thurab/Company_Bankruptcy_prediction_Classification_usecase) |
| **Please write a short summary of your Capstone project and its components. Describe the problem statement, your approaches and your conclusions. (200-400 words)** |
| **Problem Statement: -**  Estimating the risk of corporate bankruptcies is of large importance to creditors and investors. There are large indirect and direct costs associated with financial distress and Corporate bankruptcies can have serious effects both locally and globally on employees, investors, customers, suppliers, and their financiers are all affected when a company disappears in some cases a corporate bankruptcy can cause an entire industry to suffer. We will be predicting whether a company may go for a bankruptcy or not by using different machine learning models  **Approach:-**  The followings are the approach on the Company Bankruptcy Prediction,   * Performed EDA. * Handled Class Imbalance. * Checked Multicollinearity. * Trained 80% and Tested 20% with the data. * Applied various models to the training and testing data to predict Bankruptcy. * Applied Grid Cross Validation to figure out the learning rate and to improve the Recall score for the model |
| **Conclusions:-**   * Applied **Logistic Regression** with Maximum iteration=1000 and default regression strength(C=1.0), the model able to give Recall **88.37%.** * Next, tried with **Decision Trees** model with ‘Gini’ criterion, the model able to give Recall **51.16%**. * Next, applied **K Nearest Neighbor** with Euclidean and Manhattan distance, the model able to produce a recall of **44.18%**, which is still lower than Logistic regression * Next, applied **Support Vector Machines**, the model able to produce the recall score of **79.06%.** * Next, applied with **Random Forest Classifier**, the model able to produces a recall score of **62.79%** only. * Next, tried with **XGBoost (Extreme Boost)** Classification, the model able to produce a recall score of **69.76%** only. * Next, tried with **Artificial Neural Network** with 76 input layers, 76 hidden layers and 1 output layer, the model able to produces a recall score of **34.88%**. * By comparing all the models, **Logistic Regression** model gives us very good recall score of **88.37%** * We applied Grid Search Cross Validation to Logistic Regression with **cv=5** for further improvement on the accuracy, grid search cv gives us the regression score of C=10 to get good result. This produced a recall score of **88%** * With that we have built a model which predicts Company Bankruptcy prediction with **88%** of Recall using Logistic Regression. |