**Capstone Project Submission**

**Instructions:**

i) Please fill in all the required information.

ii) Avoid grammatical errors.

|  |
| --- |
| **Team Member’s Name, Email and Contribution:** |
| 1. Name: - Kanika Bhardwaj, Email ID: -bhardwajk496@gmail.com |
| **Please paste the Gather Repo link.** |
| Github Link:- https://github.com/KanikaBhardwaj1/Company-bankruptcy-prediction/blob/main/Copy\_of\_COMPANY\_BANKRUPTCY\_PREDICTION.ipynb |
| **Please write a short summary of your Capstone project and its components. Describe the problem statement, your approaches and your conclusions. (200-400 words)** |
| Data science can be summarized into five steps: capture, maintain process, analyze, and communicate. Bankruptcy prediction is an important problem in finance, since successful predictions would allow stakeholders to take early actions to limit their economic losses. In recent years many studies have explored the application of machine learning models to bankruptcy prediction with financial ratios as predictors. This study extends this research by applying machine learning techniques to a quarterly data set covering financial ratios for a large sample of the Taiwan Economic Journal for the years 1999 to 2009. We find that tree-based ensemble methods, especially XGBoost, can achieve a high degree of accuracy in out-of-sample bankruptcy prediction. We next apply our best model, using XGBoost, to the problem of predicting the overall bankruptcy.  Hereafter we applied Linear Regression, Decision Trees and Random Forest Algorithms. Visualizations showed us how our model’s predictions were close to Test Data. It was evident that XGBoost were performing well. When we did analysis on model evaluation result we observed that decision tree model were not a good performers. As, XGBoost was providing us good accuracy, we can say that it’s a model to be opted for. We got good f1\_ score for XGBoost i.e. close to 96%. When we did analysis on different types of model we got good accuracy for XGBoost i.e. close to 96.1%.  The performance of the LDA and NB models is affected by the undersampling rate. This fact shows that in a specific range, finding the undersampling rate which can improve the performance of the model is a feasible method of data preprocessing. It is feasible to improve the model performance in future work by increasing the undersampling rate while keeping the optimal information and speed. |