Corporation bankruptcy is some kind of legal way of handling those businesses which are unable to repay their outstanding debts. The organization’s solvency is calculated using the probability value of the scenario where the company is not able to pay its debt. The main frameworks of corporate bankruptcy are: (1) the smaller the failure probability, larger the reservoir and net liquid asset operation flow; (2) greater the failure probability, greater the debt amount and operations fund expands. Business insolvency can occur two ways: Receivership and Liquidity. Few financial ratios that influence business insolvencies are: -1) Profit Scenario 2) Costs Incurred 3) Capital turnover 4) Liquidity Ratio 5) Asset 6) Capital 7) Growth Rate 8) Tobin’s Q.

In 2008, Lehman Brothers filed for the biggest bankruptcy event in the history of America with more than 600 billion in debts. Just before becoming bankrupt, it was the fourth-largest U.S. investment bank. This firm went bankrupt due to its huge investment to the mortgage and real estate markets of America. Following the trend of most of the IB banks, they were mostly dependent on short-term markets in order to make profit in billions each day. Ultimately, it was a failure in securing funding that caused this sudden bankruptcy. Currently due to this ongoing global coronavirus pandemic, all leading experts predicted that the economic depression, we are going to face in coming few years, will be much higher on scale than what we faced in 2008. In this study, different machine learning techniques are employed to predict bankruptcy. Further based on the performance of the classifiers, the best model is chosen for development of a probabilistic decision modelling in Python programming language. The modelling can be utilized by stock holders and investors to predict the performance of a company based on their financial history.

This paper uses a sample dataset of US Corporate firms comprising thirteen financial ratios as feature variables and after data preprocessing, an analysis based on nine different machine learning techniques (Logistic Regression, KNN, SVM, Naive Bayes, Decision Tree, Random Forest, AdaBoost, XgBoost, CatBoost) on the dataset is done. I have used Accuracy Scores, ROC-AUC Curve, Confusion Matrix, Precision, Recall as well as F Score and Cumulative Gain Chart. The best models came out to be Random Forest, XgBoost, AdaBoost, CatBoost. After applying an Ensemble Voting Method on top 5 algorithms, it votes for Random Forest and the boosting algorithms to be the best two predictors on both training and testing data of bankruptcy cases, thus reducing over fitting problem. Final model is able to correctly predict all 8033 bankrupt firms correctly and 17208 non-bankrupt firms correctly out of 17217 in the test dataset. Only 9 corporations have been misclassified as bankrupt when they have no chance of going bankrupt. Furthermore, this model allocates significance to discrete components related to assets, liquidities.

The results suggest if we compare it with the literature review then our study indicates due to advancement of modern machine learning and artificial intelligence-based technologies, much better accuracy in bankruptcy prediction results can be achieved with gradient boosting algorithms like XgBoost, Ada Boost, Cat Boost alongside traditional random forest algorithm. Earlier the techniques like principal component analysis or oversampling to achieve more accurate results were not common, but in our research work, we have introduced the application of these modern techniques to handle dimensionality and data imbalance problem with ease. The predictive model helps to produce high test accuracy in predicting bankruptcy. In future to be continued with the research, I want to create multi-year bankruptcy prediction model which, can predict potential bankruptcies several years prior in advance. Also, we want to see if there is any underlined relation between what the firm’s social media activities and odds of bankruptcy using Natural Language Processing based sentiment analysis. Additionally, I want to apply recent deep learning algorithms to predict bankruptcy.

REFERENCES

[9] Cat Boost vs XG Boost: Mar 14, 2018