PROJECT ON CORPORATE BANKRUPTCY PREDICTION

RUSHIRAJ JAGTAP

INTRODUCTION

- We need predictive tools to better predict such events to prevent such catastrophic occurrence from happening again.
- The purpose of the bankruptcy prediction is to assess the financial condition of a company.
- Its future perspectives within the context of long-term operation on the market.
- when applied to real world problems of corporate accounting exclusively focused towards predicting corporate bankruptcy.

DATASET AND FEATURES

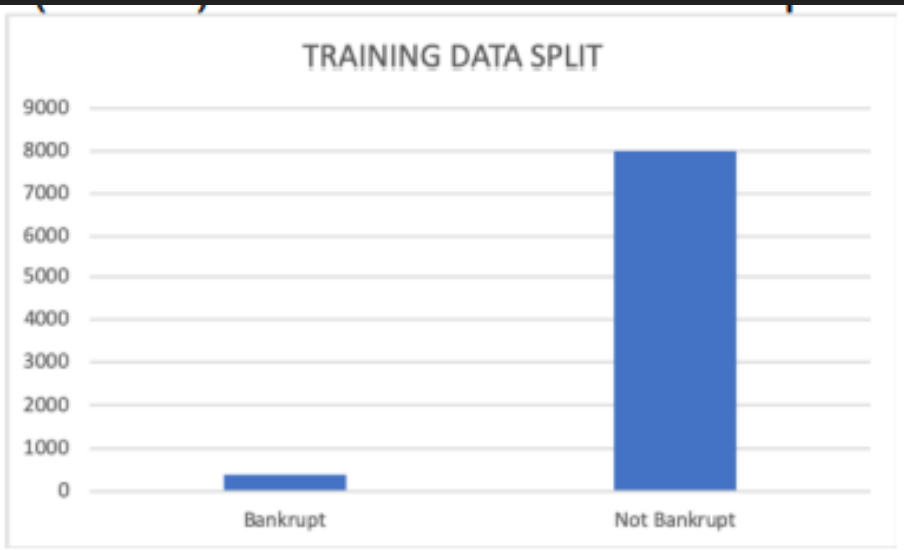

Figure 1 Training Data Split

To evaluate the performance and accuracy of various techniques we are relying on dataset which is hosted on UCI Machine Learning Repository.

The data that is available is in classified in 5 cases depending on forecasting period.

We choose and split the data randomly into 80% for training and hold out 20% data for testing.

MODEL IMPLEMENTATION

- Logistic Regression-

  Logistic regression is a linear model of classification. We use L2 regularization to avoid overfitting the data and introducing high variance in our model.

- Support Vector Machine-

  A good separation is achieved by the hyper-plane. the larger the margin the lower the generalization error of the classifier.

- Neural Network-

  Adam Gradient descent though performs better, did not perform well for us. Our model had 5 neurons in 2 hidden layers. We simply used logistic function with Square Error loss function.

- Naïve Bayes-

  For instance Gaussian based model was resulting in negative score no matter how we fine tune it.

RESULTS

Based on our observation SVM performed better than any other model in terms of maximum accuracy and minimum loss.vAccuracy is 94.66%

SVM to be performing better for this scale of data and features was expected as we learned in class since it uses vectors aka subset of data to operate. Same can be said for performance for Logistic Regression and Neural Network.

Classifier	Test Accuracy	Test Log Loss	Train Accuracy	Train Time
LogisticRegression	0.946692	1.841192	0.952392	1.01
SVM	0.947168	1.824752	1.000000	2.7
NeuralNetwork	0.947168	1.824752	0.954297	0.8
NaiveBayes	0.755831	8.433488	0.765413	0.006
ExtremeGradientBoosting	0.526416	16.357391	1.000000	1.8
LightGradientBoostingMachine	0.708710	10.061028	1.000000	0.7

CONCLUSION

- The project evaluated the approaches for the problem of predicting the bankruptcy basing on the financial factors.
- To solve the stated classification problem, we applied various models.
- The results gained by the SVM, Logistic and Neural Network were significantly better than the results gained by Boosting.