Application of Ensemble Models in Credit Scoring Models

Anjali Chopra¹
Priyanka Bhilare¹

Abstract:
Loan default is a serious problem in banking industries. Banking systems have strong processes in place for identification of customers with poor credit risk scores, however most of the credit scoring models need to be constantly updated with newer variables and statistical techniques for improved accuracy. While totally eliminating default is almost impossible, loan risk teams however minimize the rate of default thereby protecting banks from the adverse effects of loan default. Credit scoring models have used logistic regression and linear discriminant analysis for identification of potential defaulters. Newer and contemporary machine learning techniques have the ability to outperform classic old age techniques. This paper aims to conduct empirical analysis on publically available bank loan dataset to study banking loan default using decision tree as the base learner and comparing it with ensemble tree learning techniques like bagging, boosting and random forests. The results of the empirical analysis suggest that the gradient boosting model outperforms the base decision tree learner indicating that ensemble model works better than individual models. The study recommends that the risk team should adopt newer contemporary techniques to achieve better accuracy resulting in effective loan recovery strategies.

Keywords: Credit Scoring Model, probability of default, ensemble methods, accuracy and precision

¹K.J Somaiya Institute of Management Studies and Research, Mumbai, Maharashtra, India

Corresponding Author: Anjali Chopra, K.J Somaiya Institute of Management Studies and Research, Mumbai, Maharashtra, India
E-mail: anjali.chopra@somaiya.edu