Consensus Based Bank Loan Prediction Model Using Aggregated Decision Making and Cross Fold Validation Techniques

Ibrahim Hadiza Ndanusa
Department of Computer Science
Federal University of Technology
Minna Niger State, Nigeria
deezah_dans@yahoo.com

Solomon Adelowo Adepoju
Department of Computer Science
Federal University of Technology
Minna Niger State, Nigeria
solo.adepoju@futminna.edu.ng

Adeniyi Oluwaseun Ojerinde
Department of Computer Science
Federal University of Technology
Minna Niger State, Nigeria
o.ojerinde@futminna.edu.ng

Abstract-- Considering the growth of the credit businesses, machine learning models for granting loan permissions with the minimum amount of risk are becoming increasingly popular among banking sectors. Machine Learning based models has proven to be useful in resolving a variety of banking risk prediction issues. ML Predictions are sometimes unfair and biased because they are heavily dependent on randomly selected training data sample for every prediction made. However, this problem can be address by utilizing a cross-validation strategy. Prediction can be improved by combining decisions from different machine learning algorithms (ensemble decision making). The proposed consensus-based prediction model is evaluated using standard performance metrics, and the proposed model achieved an accuracy of 83 percent.

Index Terms: - ML, Machine Learning, **K-NN**, K- Nearest Neighbor **DT**, Decision Tree, **SVM**, Support Vector Machine, **LR**, Logistic Regression.

I. INTRODUCTION

In the event of a global recession, the banking sectors is prepared to respond. Financial institutions are more reliant on loan interest rates acquired from businesses in distress. Such institutions are having difficulty authorizing loans and dealing with current loan defaulters [1]. Banks must take appropriate steps to reduce credit risks in order to reduce costs as much as possible. Customer's assets in banks are likely to weather the current crisis without too much difficulty. It is possible to determine the default status of payment or credit score based on a customer's portfolio [2]. Access to centralized information about bank customers is beneficial to any financial company that obtains and manages consumer data in order to analyze borrowing, purchasing, and repayment patterns [3]

However, predictive machine learning models are thought to be effective at classifying data that has never been seen before into various classes. The predictive models work by learning from labelled observations to estimate the most appropriate category to which a data sample belongs [4]. As a result, such models are widely employed in a variety of industries, including the financial sector [5].

Machine learning (ML) systems are capable of detecting patterns in data and predicting complicated outputs in the face of high uncertainty [6]. In most cases machine learning algorithms parameters has to be fine tune to attain good prediction result (thus, hyperparameter optimization) [7]. Human rely on algorithm precision and accuracy to handle complex tasks in varieties of field, including medicine, finance, and law. In many circumstances, machine learning algorithm can outperform human, especially when working with huge datasets or a large number of input characteristics [8]. Predicting criminal recidivism based on previous conviction of a previous crime, it's an area where machine learning algorithm and expert systems might help human make better judgments [4].

This study primarily tends to address the issue related to biased decision making and improve on existing prediction model using aggregated decision-making approach. However, Tolan and other researchers [4] identifies that precision and prediction tend to be unfair and biased based on a randomly selected training data set for each prediction. Hence, this motivate the adoption of cross fold validation approach to address the identified problem.

Furthermore, this paper aim to develop an improved consensus-based bank loan predication model and approval system using cross-fold validation techniques and aggregated decision-making approach.