Loan Eligibility Classification Using Logistic Regression

Paul Law Lik Pao
Faculty of Computing
Universiti Malaysia Pahang Al-Sultan Abdullah
Pekan, Malaysia.
paullaw9904@gmail.com

Mohd Arfian Ismail*
[1]Faculty of Computing
Universiti Malaysia Pahang Al-Sultan Abdullah
Pekan, Pahang, Malaysia.
[2]Pusat Kecemerlangan Kecerdasan Buatan & Sains Data
Universiti Malaysia Pahang Al-Sultan Abdullah
Gambang, Kuantan, Pahang, Malaysia.
arfian@ump.edu.my

Abstract—**Machine learning is becoming increasingly vital in various domains, including loan eligibility classification, d ue to its ability to analyze large amounts of data, develop predictive models, adapt to new information, and automate processes. This research paper presents a study on loan eligibility classification using a machine learning approach by comparing the performance of three Machine Learning algorithms which were Logistic Regression, Random Forest, and Decision Tree. This research was conducted using Python and Jupyter Notebook for data analysis and model development. The models were then evaluated on the testing set using evaluation metrics such as Accuracy, Precision, Recall, And F1-Score. The performance of the models was compared to identify the most effective algorithm for loan eligibility classification. A mong t he t hree M L a pproach, t he LR model appears to be the most effective at classify loan eligibility, with the 82% accuracy score, 82% recall score, 81% precision score and 79% F1 score.**

Keywords—Machine Learning, Classification, Loan, Predictive model, Logistic Regression, Python, Streamlit

I. INTRODUCTION

One of the oldest industries in this world is banking. This industry started when merchants made grain loans to farmers and traders transported products between towns or cities. This barter system has been in existence for centuries around the world until the invention of money and start the banking activities by using the money. The first m odern b ank i n the history was founded in Siena, where located at the Italian Renaissance in 15th century. The bank is called Banca dei Pashi di Siena [1].

In the modern era, one of the main challenges in purchasing properties or necessary items is having sufficient f unds. To overcome this challenge, people often apply for loans from banks, enabling them to acquire the needed cash once their loan application is approved. Consequently, there has been a yearly increase in loan applications in Malaysia, particularly due to inflation a nd r ising p roperty a nd c ar p rices [2]. Loans serve as the primary revenue source for banks in Malaysia, as they generate significant p rofits th rough th e in terest charged on loans.

As been stated, loans occasionally become necessary for those needed the money for business purpose. By considering the topic, the machine learning method involved in the loan application process to speed up the loan approval progress automatically.

The purpose of this research is to figure out the best ML method in loan classification and develop an interactive interface which allow user to predict the loan application eligibility by using Streamlit.

II. MATERIAL AND METHODS

A. Comparison of existing ML methods

The ML classifies into three primary categories which are supervised ML, unsupervised ML, and semi-supervised learning [3]. This study focusses in classification categories of supervised ML and research will be conducted. Logistic Regression, Random Forest and Decision Tree are three existing classification method in ML that have been selected. The reasons choose these three ML classification methods are ease to use in solving the loan eligibility classification which is binary classification issues. These classification ML methods able to predict the discrete value into classes and it is easy to interpret as these algorithms can smoothly handle qualitative target variables [4]. The comparison of the existing ML methods is shown in Table I.

TABLE I
COMPARISON OF MACHINE LEARNING METHODS

Characteristics	LR	DT	RF
Classification Accuracy	Lower	Lower	Higher
Training Speed	Fast	Fast	Slow
Complexity	Simple	Simple	Moderate
Prediction Speed	Fast	Fast	Fast
Ease of Implementation	Easy	Middle	Hard

B. Methodology

The classification of loan eligibility using LR was conducted in this study. It includes the project requirement which discuss the step of LR works, mathematical formulation used and research framework where it covers literature review, data collection, data pre-processing and applying algorithm and