



Loan Application Analysis using Machine Learning

Mrs. N. SUNEETHA

Assistant professor

Computer Science and Engineering

Vignan's Institute of Engineering for Women, Visakhapatnam

Jawaharlal Nehru Technological University Kakinada, Andhra Pradesh, India

N. Swarupa, M.Padmaja, K.Sreeja, N.Ramadevi, K.Sowjanya
B. Tech Students

Computer Science and Engineering

Vignan's Institute of Engineering for Women, Visakhapatnam

Jawaharlal Nehru Technological University Kakinada, Andhra Pradesh, India

Abstract: With the enhancement in the banking sector lots of people are applying for bank loans but the bank has its limited assets which it has to grant to limited people only, so finding out to whom the loan can be granted which will be a safer option for the bank is a typical process. So in this project, try to reduce this risk factor behind selecting the safe person so as to save lots of bank efforts and assets. This is done by mining the Big Data of the previous records of the people to whom the loan was granted before and on the basis of these records/experiences the machine was trained using the machine learning model which gives the most accurate result. The main objective of this project is to predict whether assigning the loan to a particular person will be safe or not. Here Machine learning techniques are used to predict the person who is reliable for a loan, based on the previous record of the person to whom the loan amount is accredited before.

The Four machine learning algorithms like Logistic Regression, Decision Tree, Random Forest and XGBoost are compared and the algorithm with highest accuracy is applied to predict the loan approval of customers.

Keywords: Machine learning, Logistic Regression, Decision Tree, Random Forest, XGBoost, Loan prediction, Python.

1. INTRODUCTION

Distribution of the loans is the core business part of almost every bank. The main portion of the bank's assets directly came from the profit earned from the loans distributed by the banks. The prime objective in the banking environment is to invest their assets in safe hands where it is. Today many banks/financial companies approve loans after a regress process of verification and validation but still there is no surety whether the chosen applicant is the deserving right applicant out of all applicants. Through this system we can predict whether that particular applicant is safe or not and the whole process of validation of features is automated by machine learning technique. The disadvantage of this model is that it emphasizes different weights to each factor but in real life sometimes loans can be approved on the basis of a single strong factor only, which is not possible through this system.

MACHINE LEARNING TECHNIQUES

- Logistic Regression
- Decision Tree
- Random Forest
- XGBoost

LOGISTIC REGRESSION

Logistic regression is an estimation of Logit function. Logit function is simply a log of odds in favour of the event.

DECISION TREE

Decision tree is a type of supervised learning algorithm (having a predefined target variable) that is mostly used in classification problems. In this technique, we split the population or sample into two or more homogeneous sets (or sub-populations) based on the most significant splitter / differentiator in input variables. Decision trees use multiple algorithms to decide to split a node in two or more sub-nodes. The creation of sub-nodes increases the homogeneity of resultant sub-nodes. In other words, we can say that purity of the node increases with respect to the target variable.

RANDOM FOREST

Random Forest is a tree based bootstrapping algorithm wherein certain no. of weak learners (decision trees) is combined to make a powerful prediction model. For every individual learner, a random sample of rows and a few randomly chosen variables are used to build a decision tree model. Final prediction can be a function of all the predictions made by the individual learners.

XGBOOST

XGBoost is an implementation of gradient boosted decision trees designed for speed and performance.

2. PROBLEM STATEMENT

Distribution of the loans is the core business part of almost every bank. The prime objective in the banking environment is to invest their assets in safe hands where it is. Today many banks/financial companies approve loans after a regress process of verification and validation but still there is no surety whether the chosen applicant is the deserving right applicant out of all applicants. Through this system, whether that particular applicant is safe or not is predicted and the whole process of validation of features is automated by machine learning technique.

3. LITERATURE SURVEY

Survey on improving classification of J48 algorithm using bagging, boosting and blending ensemble methods on SONAR dataset using WEKA:

Many researchers have investigated the technique of combining the predictions of multiple classifiers to produce a single classifier. The resulting classifier is more accurate than any individual classifier. This paper investigates the ability of ensemble methods to improve the efficiency of basic J48 machine learning algorithms. Ensemble algorithms like Bagging, Boosting and Blending improved the discrimination between sonar signals bounced off a roughly cylindrical rock in the SONAR dataset. The ranking and standard deviation functionalities provided by the WEKA experimenter helps to determine the effectiveness of a classifier model.

4. EXISTING SYSTEMS AND DRAWBACKS

Assessing the risk, which is involved in a loan application, is one of the most important concerns of the banks for survival in the highly competitive market and for profitability. These banks receive a number of loan applications from their customers and other people on a daily basis. Not everyone gets approved. Most of the banks use their own credit scoring and risk assessment techniques in order to analyze the loan application and to make decisions on credit approval. In spite of this, there

are many cases happening every year, where people do not repay the loan amounts or default, due to which these financial institutions suffer huge amounts of losses.

5. PROPOSED SYSTEM

In today's world, taking loans from financial institutions has become a very common phenomenon. Everyday a large number of people make applications for loans, for a variety of purposes. But all these applicants are not reliable and everyone cannot be approved. Every year, we read about a number of cases where people do not repay the bulk of the loan amount to the banks due to which they suffer huge losses. The risk associated with making a decision on loan approval is immense. So the idea of this project is to gather loan data from multiple data sources and use data mining algorithms on this data to extract important information and predict if a customer would be able to repay his loan or not. In other words, predict if the customer would be a defaulter or not.

5.1 CONTENT DIAGRAM

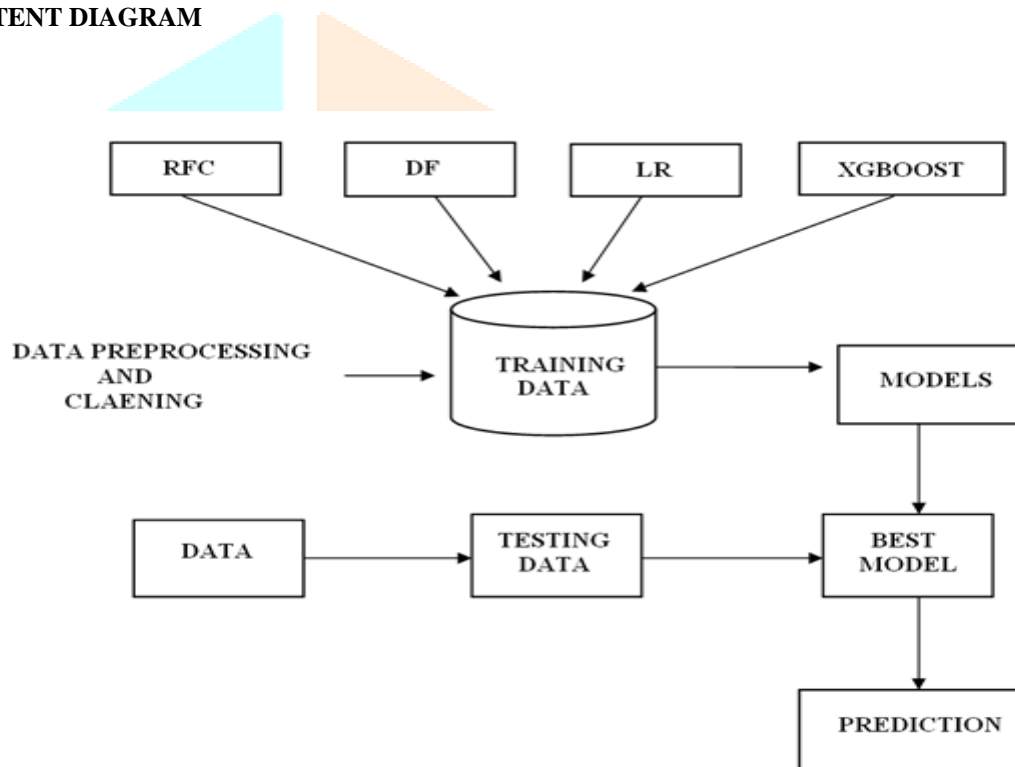


Fig 5.1 Content diagram of proposed system

5.2 FLOW CHART

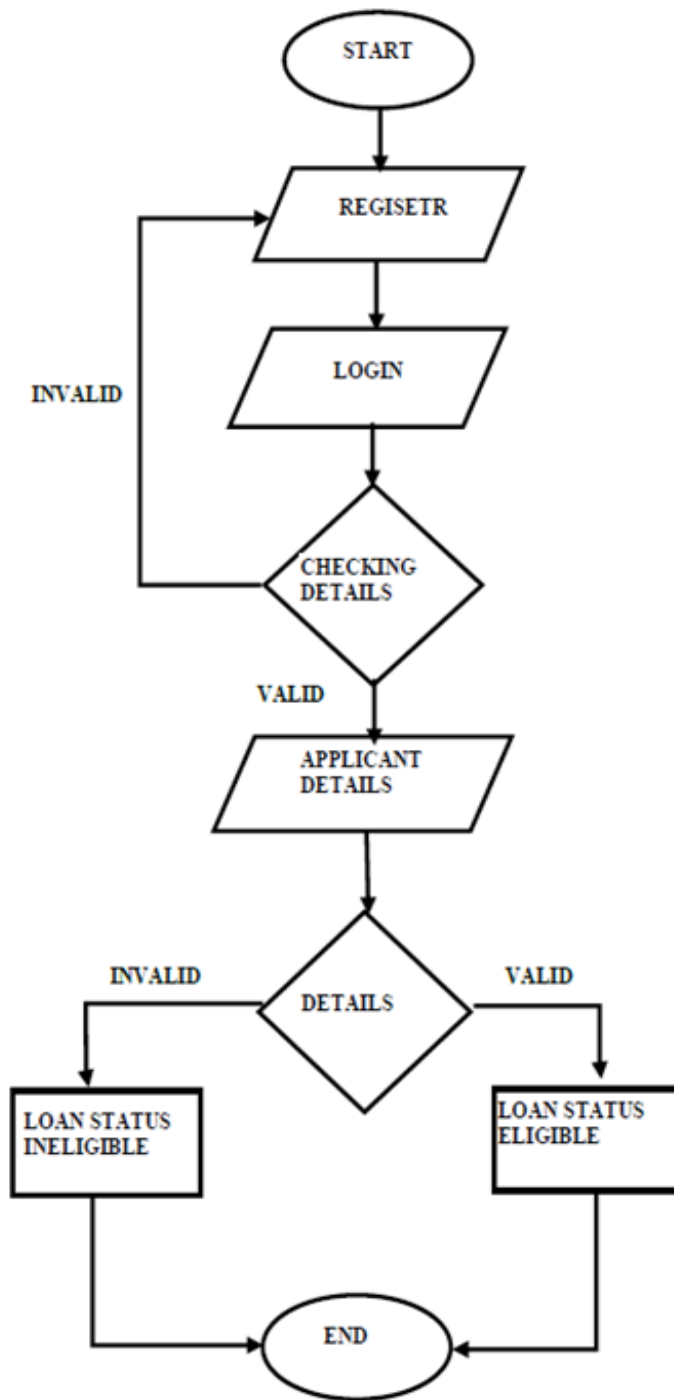


Fig 5.2.1 Flowchart for User

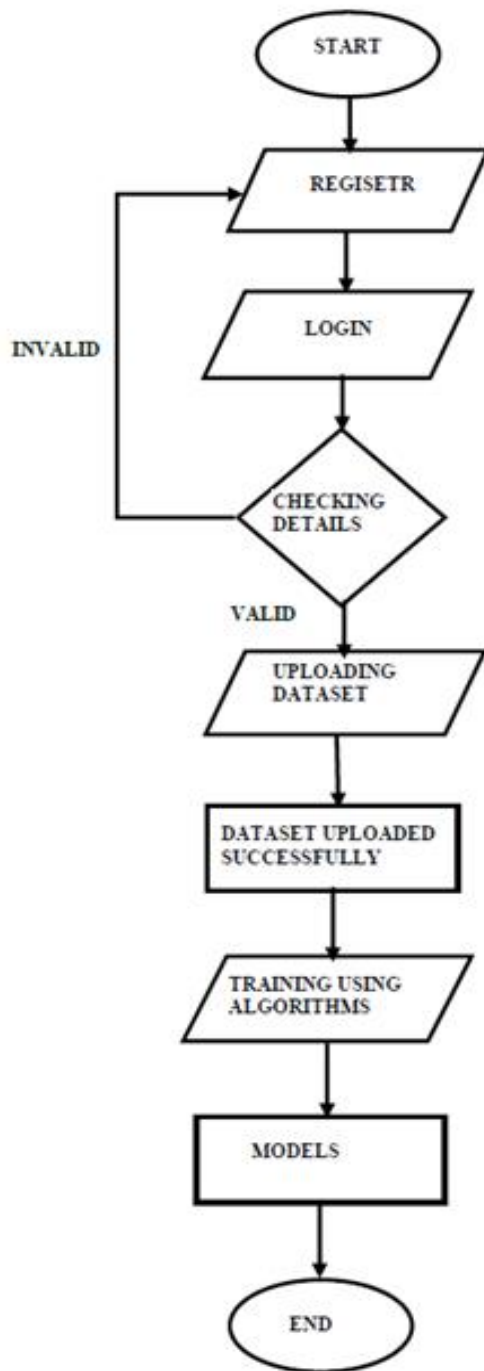


Fig 5.2.2 Flowchart for Admin

6.METHODOLOGY

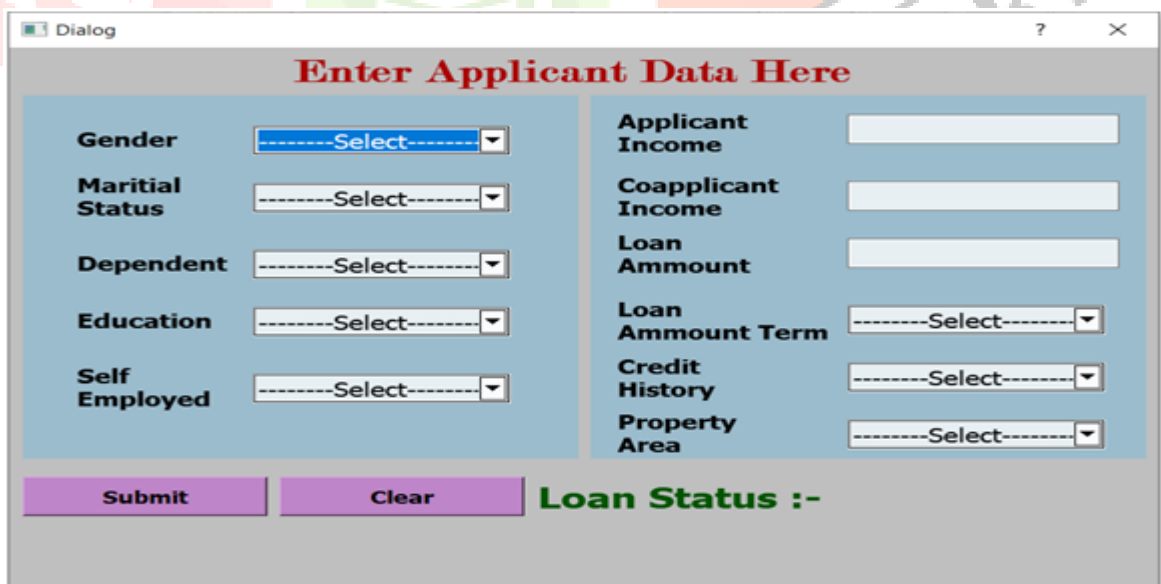
Implementation is the stage of the project when the theoretical design is turned out into a working system. Thus, it can be considered to be the most critical stage in achieving a successful new system and in giving the user confidence that the new system will work effectively. The implementation stage involves careful planning, investigation of the existing system and its constraints on implementation, designing of methods to achieve changeover and changeover methods.

7. RESULTS



The screenshot shows a web application window titled "Dialog" with a wooden background. The main heading is "LOAN APROVAL SYSTEM". Below it, there are input fields for "username" and "Password". A "User" dropdown menu is positioned below the password field. A link "Register here" is provided for new users. A prominent purple "Login" button is at the bottom center.

Fig (a) Login Page



The screenshot shows a web application window titled "Dialog" with a grey background. The heading is "Enter Applicant Data Here". The form is divided into two columns. The left column contains dropdown menus for "Gender", "Marital Status", "Dependent", "Education", and "Self Employed". The right column contains text input fields for "Applicant Income", "Coapplicant Income", "Loan Ammount", "Loan Ammount Term", "Credit History", and "Property Area". At the bottom, there are "Submit" and "Clear" buttons, and a label "Loan Status :-".

Fig (b) Applicant Data



Fig (c) Admin Page

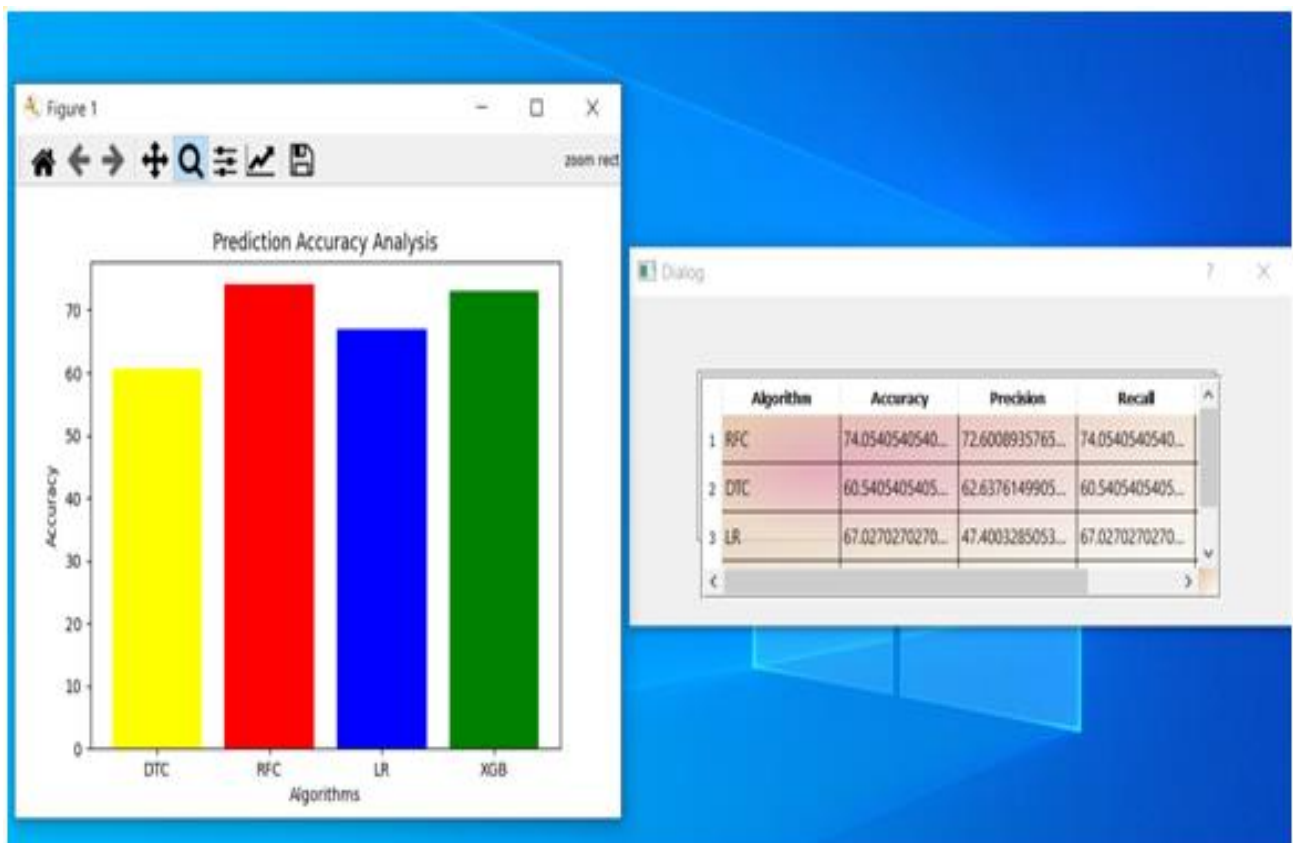


Fig (d) Graphical Representation

CONCLUSION

From a proper analysis of positive points and constraints on the component, it can be safely concluded that the product is a highly efficient component. This application is working properly and meeting all Banker requirements. This component can be easily plugged in many other systems. There have been numbers cases of computer glitches, errors in content and most important weight of features is fixed in automated prediction system, So in the near future the so –called software could be made more secure, reliable and dynamic weight adjustment .In near future this module of prediction can be integrate with the module of automated processing system. The system is trained on old training dataset so future software can be made such that new testing data should also take part in training data after some fixed time.

REFERENCES

- [1]. Rattle data mining tool: available from <http://rattle.togaware.com/rattle-download.html>
- [2]. How to predict Loan Eligibility using Machine Learning Models Build predictive models to automate the process of targeting the right applicants, Mridul Bhandari Sep 15, 2020
- [3]. Aslam U, Aziz H I T, Sohail A and Batcha N K 2019 An empirical study on loan default prediction models Journal of Computational and Theoretical Nanoscience 16 pp 3483–8
- [4]. Breeden J L 2020 A survey of machine learning in credit risk
- [5]. Zhu L, Qiu D, Ergu D, Ying C and Liu K 2019 A study on predicting loan default based on the random forest algorithm The 7th Int. Conf. on Information Technol. and Quantitative Management (ITQM) 162 pp 503–13
- [6]. Challa, M.L., Malepati, V. & Kolusu, S.N.R. Finance Innov (2018) 4: 24. <https://doi.org/10.1186/s40854-018-0107-z>
- [7]. Selvamuthu, D., Kumar, V. & Mishra, A. Finance Innov (2019) 5: 16 <https://doi.org/10.1186/s40854-019-0131-7>.
- [8]. Zhong, X. and Enke, D. Financ Innov (2019) 5 : 24 <https://doi.org/10.1186/s40854-019-0138-0>

