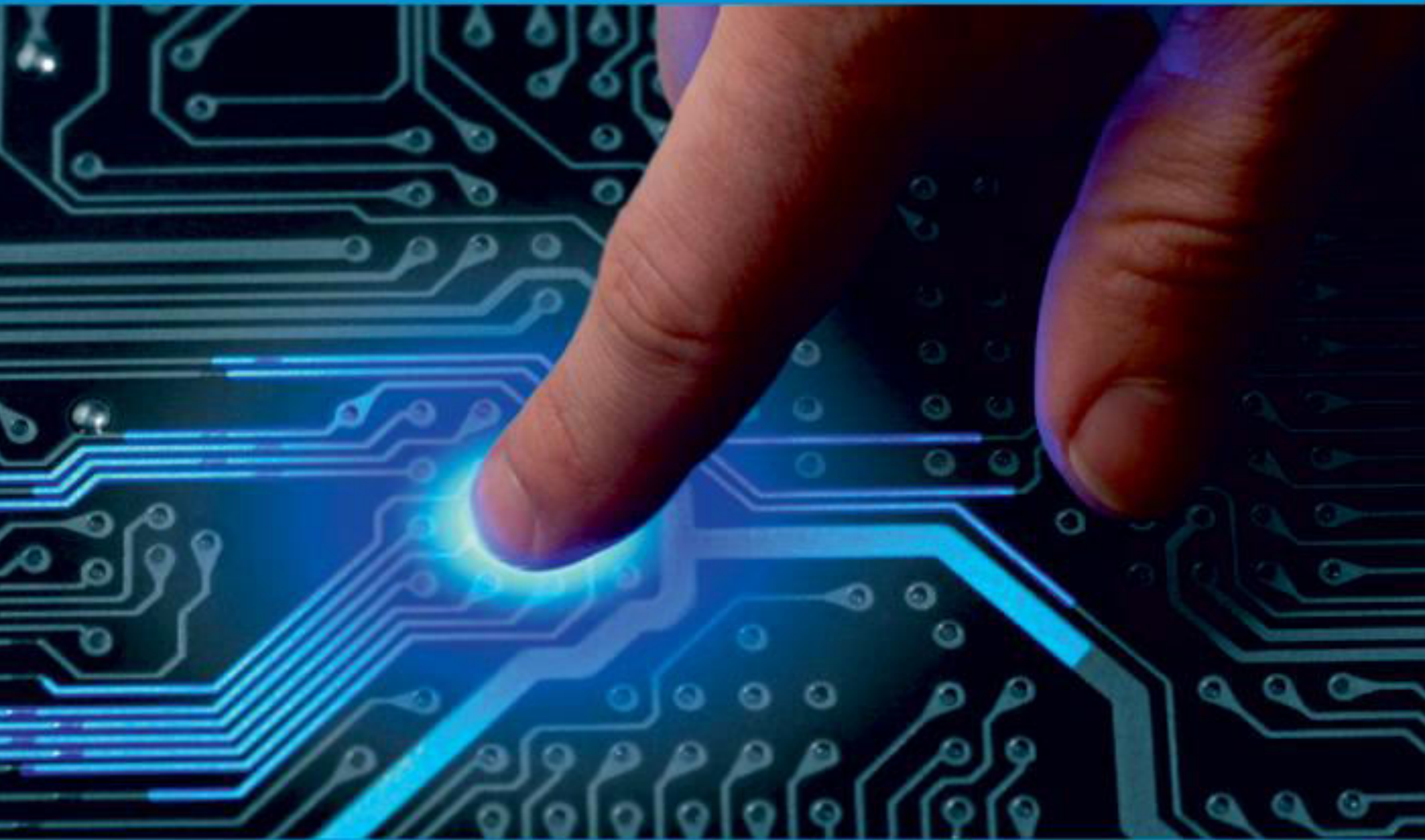
IJIRCCE

e-ISSN: 2320-9801 | p-ISSN: 2320-9798



INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN COMPUTER & COMMUNICATION ENGINEERING

Volume 10, Issue 7, July 2022

INTERNATIONAL STANDARD SERIAL NUMBER INDIA

Impact Factor: 8.165

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 8.165 |

|| Volume 10, Issue 7, July 2022 ||

| DOI: 10.15680/IJIRCCE.2022.1007018 |

Accelerating Loan Compliance System Using Machine Learning Algorithm

Ritik Koul, Rameez Ali Mohommed, Vipul Gupta, Shatrunjay Pratap, Byre Gowda

Dept. of ISE, SIR MVIT, India

Dept. of ISE, SIR MVIT, India

Dept. of ISE, SIR MVIT, India

Dept. of ISE, SIR MVIT, India

Asst. Professor, Dept. of ISE, SIR MVIT, India

ABSTRACT: Loans are the major source from where the banks make their profit. Although many people apply for a loan. Choosing a real applicant to pay off a loan is difficult. Because it is done manually, there can be a lot of confusion in selecting real applicants. Therefore, we are developing a loan prediction system using machine learning so that the system automatically selects suitable candidates. This is useful for both bank employees and applicants. The loan term is greatly shortened. This article uses some machine learning algorithms such as decision trees,support vector machine,random forest classifier,logic regression,K-nearest neighbour to predict loan data.

I. INTRODUCTION

Distribution of the loans is the most critical part of any financial institution like bank. The main portion of the bank's assets directly come from the profit earned from the loans distributed by the banks. The prime objective in banking environment is to invest their assets efficiently and at large scale in form of loans. Today many banks/financial companies approves loan after a rigorouss 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 appropriate for loan approval or not and the whole process of validation of features is automated by machine learning techniques.

Loan Prediction is very helpful for the employees of banks as well as for the applicant too. The goal of this project is to provide a fast, immediate and easy way to select qualified candidates. This can provide special benefits to the bank. The loan prediction system can automatically calculate the weight of each function involved in loan processing, and in the new test data, the same function given the associated weight is processed.

II. LITERATURE REVIEW

Sarwesh Site, Dr. Sadhna K. Mishra[1] proposed the method in which two or more classifiers are combined together build a ensemble model for better prediction. They

he used bagging and strengthening techniques and then used them random forest technique. The classifiers process is k improve data performance and provide better efficiency. In this work, the authors describe various ensemble techniques for binary classification and also for multi-class classification. A new technique that is described by the authors for the ens emble is the COB which gives effective classification performance, but also compromised by noise and outlier classification data.Finally, they concluded that the algorithm is file-based improves the results for the training dataset.

A.R.Ghatge, P.P.Halkarnikar[2] develops artificial a neural network model for bank credit risk prediction.A feedback propagation neural network isused to predict credit default. They also compare results with the bank's manual calculations in 2004, 2005 and 2006. The results are better and higher performance compared to the bank's manual calculations.

Alaraj, M., Abbod, M. [3] introduce a credit risk model which are based on homogeneous and heterogeneous classifiers. An ensemble model based on three classifiers which are logistic artificial neural network, logistic regression and vector machine support. The results show that the set of heterogeneous classifiers has improved performance and accuracy compared to a set of homogeneous classifiers.

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 8.165 |

|| Volume 10, Issue 7, July 2022 ||

| DOI: 10.15680/IJIRCCE.2022.1007018 |

M. Yaghini, T. Zhiyan and M. Fallahi [4] reports a a model that is based on feedforward neural networks identify bad customers in the bank. They use three different strategies such as fast, dynamic and multiple strategy. To avoid redrawing the model, cross validation is done on the model. Evaluate the proposed model, the result of the neural network is compared with a

common prediction methods namely decision tree and logistic regression. The results indicate that three layers neural network based on backpropagation learning the fast strategy algorithm has higher accuracy.

III. PROBLEM STATEMENT

Financial companies, banks handle different types of loans, such as student loans, shop loans, home loans,personal loans, etc. All of which are part of the national loan types. All businesses and banks are located in towns, towns and cities. After customer applies for a loan, this bank/company wants to verify the details of customer as to whether the applicant is eligible for a loan. The main purpose of the system is to get loans with or without approvals based on the trained machine learning model.

IV. THE REFLECTIVE PROCESS

Step 1: We explore the input datasets.
Step 2: After exploring the data we do univariate analysis on the variables.
Step 3: Then we perform bivariate analysis on the variables.
Step 4: After step3 we treat the missing values present in the dataset.
Step 5: We perform preprocessing like Feature Engineering, Outliers Removal, Feature Scaling, and Transformation.
Step 6: We put the refined dataset into the machine learning models and performed the following.
6.1: Logistic Regression
6.2: KNN Classifier
6.3: Decision Tree Classifier
6.4: Random Forest
6.5: Support Vector Machine (SVM)
Step 7: We evaluate the results of the models and provide the best result and machine learning model to the user.

Logistic Regression:
This type of statistical model (also known as the logit model) is commonly used for classification and predictive analytics. Logistic regression estimates the probability of an event such as: B. Voting or non-voting based on a particular set of independent variable data. The result is a probability, so the dependent variable is limited between 0 and 1. In logistic regression, logit transformations are applied to probabilities. H. The ratio of the likelihood of success to the likelihood of failure.

KNN Classifier:
The K-nearest neighbors (KNN) algorithm is a type of supervised ML algorithm that can be used for both classification and regression prediction problems. However, it is mainly used for classification of predictive problems in industry. The following two properties would well define KNN −

Lazy Learning Algorithm − KNN is a lazy learning algorithm because it does not have a dedicated training phase and uses all data for training during classification.

Non-parametric learning algorithm − KNN is also a non-parametric learning algorithm because it does not assume anything about the underlying data.

Decision Tree Classifier:
A decision tree is a supervised learning technique that can be used for both classification and regression problems, but is mostly preferred for solving classification problems. It is a tree-structured classifier where internal nodes represent features of the dataset, branches represent decision rules, and each leaf node represents an outcome.
There are two nodes in a decision tree, which are the decision node and the leaf node. Decision nodes are used to make decisions and have multiple branches, while Leaf nodes are the output of these decisions and contain no further branches.

Random Forest:
As a supervised learning technique, Random Forest is a popular machine learning algorithm.
Both classification and regression problems can be solved with it in machine learning. This method relies on ensemble learning to solve complex problems and improve performance by combining multiple classifiers.

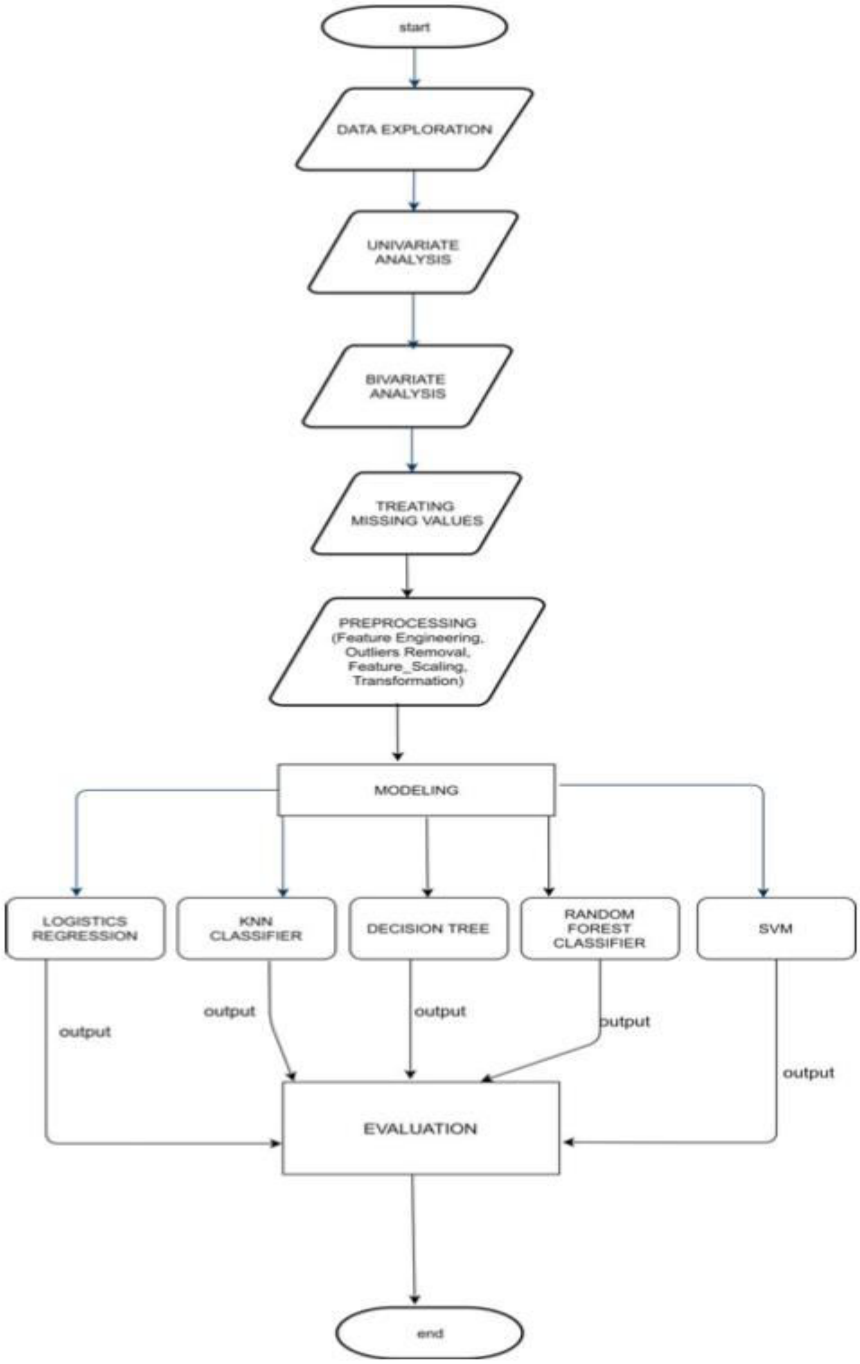
| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 8.165 |

|| Volume 10, Issue 7, July 2022 ||

| DOI: 10.15680/IJIRCCE.2022.1007018 |

As the name suggests, "Random Forest is a classifier that contains a number of decision trees on different subsets of a given data set and takes the average to improve the predictive accuracy of that data set." Instead of relying on a single decision tree, a random forest takes a prediction from each tree and predicts and predicts the final output based on the majority of votes.

Support Vector Machine(SVM):
Support Vector Machine or SVM is one of the most popular supervised learning algorithms used for both classification and regression problems. However, it is primarily used for classification problems in machine learning.
By using the SVM algorithm, we create the best decision boundary from the n-dimensional space so that we can easily assign new data points to the correct category in the future. This best decision boundary is called the hyperplane.
SVM selects extreme points/vectors that help in creating the hyperplane.



| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 8.165 |

|| Volume 10, Issue 7, July 2022 ||

| DOI: 10.15680/IJIRCCE.2022.1007018 |

V. PROPOSED SYSTEM

This loan approval system benefits both the bank and its customers. Banks can use our model to embed our model in a user interface that clients can use. This helps the bank avoid unnecessary processing of applications that do not meet the minimum criteria based on previous experience. Meanwhile, it will benefit customers who apply for a loan to the bank that is most likely to be approved for a loan. Thus, the system can save time, resources and effort for both the bank and its customers.

In machine learning, we use semi-automated gaining knowledge of the data to determine if it is a loan will or will not be approved. The classification could be a supervised learning in which the response is categorical, i.e its unit area values in a finite unordered set. Easy matter classification is used, scikit learn. Praim primacy of this the company does not have to maintain a ground team verify and verify customer records. They can easily check whether the loan must be approved by this prediction or not Model.

In this article, we try to develop the user interface flexibly meaning graphical concepts connected through a browser interface.Our goal is to implement a machine learning model in such a way that it is possible to classify,with the highest degree of accuracy possible, master card fraud by a dataset collected from Kaggle. after an initial survey of knowledge, we tend to know that we could implement a random forest model for the best accuracy reports.

Random forest because it was a good candidate for binary classification. The sklearn Python library was used for the implementation project, we used Kaggle datasets for credit card fraud detection,using pandas to dataframe for class ==0 for no fraud and class==1 for fraud,matplotlib for plotting fraud and non-fraud data, train_test_split for data extraction (split arrays or matrices into random train and test subsets) and used Logistic Regression machine learning algorithm for fraud detection and printing score prediction by algorithm. Confusion at last matrix was plotted as true and predicted. The main part of this post uses sklearn preprocessing the MinMax method is scalar, i.e. it helps to normalize the data. Model selection using cross-validation, split train/test, kfold, GridSearchCV.

VI. CONCLUSIONS

From a proper analysis of the positive points and limitations of the element, it can be safely concluded that the product is a highly productive element. This use works properly and fulfills all requirements of the Banker. This member can be freely connected to many other systems. There have been mathematical cases of computer glitches, content violations, and the most important feature weight is fixed in the automated prophetic system, so the so-called software could be safer, more trustworthy, and more dynamic in the near future. In the near future, this prophecy module may be integrated with the automated processing system module. The system is trained on the old training data set in the future software can be made so that the new test date should also participate in the training data after a certain correction period.

REFERENCES

[1]Rattle data mining tool: available from http://rattle.togaware.com/rattle-download.html

[2]S.S. Keerthi and E.G. Gilbert. Convergence of a generalizeSMO algorithm for SVM classifier design. Machine Learning, Springer, 46(1):351–360, 2002.

[3]K. Manikandan, S. Shanmugan, R. Ashish kumar, Vadde Venkat Harish, T. Jansi Rani, T. Nishanthi, Trans membranous fetal movement and pressure sensing. Materials Today: Proceedings, Available online 13 May 2020. https://doi.org/10.1016/j.matpr.2020.04.497

[4]Zoran Ereiz. Predicting Default Loan using Machine Learning (OptiML), https://www.researchgate.net/publication/338948751_Predicting_Default_Loans_Using_Machin e_Learning_OptiML

[5]Premalatha V., Vineesha K. and Srinivasarao M., "International Journal of Scientific and Technology Research", 2020, 9(1), pp. 1005-1008.

[6]J.R. Quinlan. Induction of decision trees. MachinelearningSpringer, 1(1):81–106, 1086.

[7]M. A. Sheikh, A. K. Goel and T. Kumar, "An Approach for Prediction of Loan Approval using Machine Learning Algorithm," 2020 International Conference on Electronics and Sustainable Communication Systems (ICESC), 2020, pp. 490-494, doi: 10.1109/ICESC48915.2020.9155614.

[8]A. Gupta, V. Pant, S. Kumar and P. K. Bansal, "Bank Loan Prediction System using Machine Learning," 2020 9th International Conference System Modeling and Advancement in Research Trends (SMART), 2020, pp. 423-426, doi: 10.1109/SMART50582.2020.9336801.

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 8.165 |

|| Volume 10, Issue 7, July 2022 ||

| DOI: 10.15680/IJIRCCE.2022.1007018 |

[9]Pidikiti Supriya, Myneedi Pavani, Nagarapu Saisushma, Namburi Vimala Kumari, kVikash,“Loan Prediction by using Machine Learning Models”, InternationalJournalofEngineering andTechniques.Volume 5 Issue 2, Mar-Apr 2019
[10]Aboobyda Jafar Hamid and Tarig Mohammed Ahmed, ―Developing Prediction Model of Loan Risk in Banks using Data Mining‖, Machine Learning and Applications: An International Journal (MLAIJ), Vol.3, No.1, pp. 1-9, March 2016.

INNO SPACE
SJIF Scientific Journal Impact Factor
Impact Factor: 8.165

INTERNATIONAL STANDARD SERIAL NUMBER INDIA

निस्केयर
NISCAIR

INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN COMPUTER & COMMUNICATION ENGINEERING