Proceedings of the Second International Conference on Artificial Intelligence and Smart Energy (ICAIS-2022)
IEEE Xplore Part Number: CFP22OAB-ART; ISBN: 978-1-6654-0052-7

2022 Second International Conference on Artificial Intelligence and Smart Energy (ICAIS) | 978-1-6654-0052-7/22/$31.00 ©2022 IEEE | DOI: 10.1109/ICAIS53314.2022.9742800

Comparative Analysis of Customer Loan Approval Prediction using Machine Learning Algorithms

Praveen Tumuluru[1]
Department of CSE
Koneru Lakshmaiah Education Foundation, Vaddeswaram AP, INDIA
praveenluru@gmail.com

Lakshmi Ramani Burra[2]
Department of CSE
PVP Siddhartha Institute of technology, Kanuru, Vijayawada, AP,INDIA,
ramanimythili@gmail.com

M.Loukya[1]
Department of CSE
Koneru Lakshmaiah Education Foundation Vaddeswaram AP, INDIA

S.Bhavana[1]
Department of CSE
Koneru Lakshmaiah Education Foundation, Vaddeswaram AP, INDIA

CH.M.H.SaiBaba[3]
Department of CSE
Koneru Lakshmaiah Education Foundation , Vaddeswaram AP, INDIA

N Sunanda[4]
Department of CSE
Koneru Lakshmaiah Education Foundation Vaddeswaram AP, INDIA

Corresponding author: praveenluru@gmail.com

Abstract – In today's increasingly competitive market, estimating the risk involved in a loan application is one of the most crucial challenges for banks' survival and profitability. The banks receive many loan applications from their customers and other individuals daily. Not every applicant is accepted. Most banks employ their credit scoring and risk assessment procedures to examine loan applications and make credit approval decisions. Despite this, many incidents of people failing to repay loans or defaulting on them occur every year, causing financial institutions to lose a significant amount of money. In this study, Machine Learning (ML) algorithms are used to extract patterns from a common loan-approved dataset and retrieve patterns in forecasting future loan defaulters. Customers' past data, such as their age, income, loan amount, and tenure of work, will be used to conduct the analysis. To determine the maximum relevant features, i.e. the factors that have the most impact on the prediction outcome, various ML algorithms such as Random Forest, Support Vector Machine, K-Nearest Neighbor and Logistic Regression, were used. These mentioned algorithms are evaluated with the standard metrics and compared with each other. The random forest algorithm achieves better accuracy.

Keywords— Random Forest, Machine Learning, K-Nearest Neighbor, Support Vector Machine, Logistic Regression.

I.INTRODUCTION

Nowadays, public depend on the bank loans to cover their unavoidable requirements. In recent years, the volume of loan applications has increased dramatically. Sanction of a loan is all the time loaded with peril[7,8]. Banking managements are very anxious about their clients' ability to reimburse their debts. Loan endorsement decisions are not always easy, even after captivating several measures and extensively scrutinizing the loan application data. This course of action must be automated in order for loan approval to be less hazardous and for banks to big lose less money in as a result.

The Artificial intelligence (AI) is a speedily embryonic technology that is utilized to answer a variety of real-world challenges perfectly[9, 10]. The Machine Learning (ML) is an (AI) Artificial intelligence techniques for improving prediction systems. A fundamental machine learning model is shown in Figure-1. The model created by the educating ML algorithm is used to build the prediction. Before making prediction-related decisions, the Machine Learning methods be able to examine the section of test data[11,12]. The problem of loan acceptance in the banking sector was solved using machine learning algorithms in this study.

II. LITERATURE REVIEW

Since the global financial crisis, risk management in banks has become increasingly important in guiding bank decision-making. The sanctioning of loans to potential individuals is a big part of risk management. However, because machine learning algorithms are black boxes, many loan providers vary the outcome with the more comprehensive systematic literature reviews that focus on the application of powerful machine learning in banking risk administration.

The authors proposed a ML based Loan Prediction using the Decision Tree and Random Forest algorithms. The major objective is to find the scenery, credibility and background of the client applying for numerous loans. The technique investigative data analysis is used to transaction with the problem of rejecting or approving the loan request or the loan prediction. Based on this, it primarily focuses on determining regardless the loan giving to a meticulous person or an organization shall be permitted or not. [1]

The primary goal is to determine whether issuing a loan to a specific individual is safe or not. The instigators provide various machine learning models to lower the risk factor associated with finding a safe individual and save a

significant amount of bank time and resources[2]. The author explored an ML algorithm for loan sanctioning process prediction. The goal is to reduce the risk of finding a suitable person to return the loan on time, allowing the bank to keep its nonperforming assets on hold. This can be accomplished by feeding the bank's preceding records of customers who have obtained loans into a well-trained machine learning models that can produce an exact result. [3]

The authors proposed a logistic regression method for loan approval prediction with a sample set for loan approval applications. [4]. The authors implemented a machine learning approach to predict credit retrieval. Credit retrieval is a perilous matter for the banking sector, and forecasting is difficult. Different machine learning approaches were used to predict credit retrieval, and the gradient expansion algorithms (GBM) outperformed the other machine learning approaches. [5]. The authors devised a method that automatically amasses data for an aspirant and calculates his credit score. To collect the user information, this model uses the social media. [6].

III. PROPOSED METHODOLOGY

The algorithms proposed for this work is Random Forest, Logistic Regression, Support Vector Machine (SVM), (K-NN) K- Nearest Neighbor.

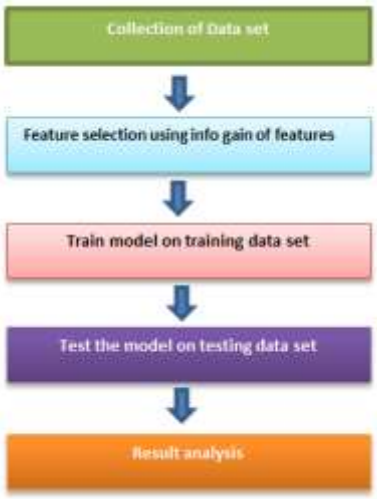

Figure.1: Basic Model of Machine Learning

Data-set description:

The data set is extensive because there are a lot of customers who apply for the loan. This data set contains various parameters such as income, education, loan amount etc. and is shown in Table.1. The model is trained, and then the customer details entered as applicants act as test data. Using the training and testing dataset to fit with different algorithms and find out the results and predict if the customer is eligible for the loan or not

Machine Learning Algorithms:

A. Random forest:

Random Forest is an ML algorithm widely used and is a part of the supervised learning technique[13]. The machine learning algorithm is used for both editing and retrieving issues. It is built on the integrated learning principle, which uses many disciplines to solve complicated problems and increase model performance.

It subdivides the tree containing several decision trees for a set of different databases and measures to increase the prediction data accuracy. Rather than depending on a single tree for decision, a random forest takes a forecast from each tree and supports it with numerous predictable votes to anticipate the eventual conclusion and is depicted in Figure.2. The random forest method produces higher accuracy and less overfitting when the number of trees is large. This is collected from kaggle repository.

Table.1: Dataset Description

Variable Name	Description	Type
Loan_ID	Unique Loan ID	Integer
Self_Employed	Self Employed(Y/N)	Character
Maritial_Status	Applicant Married(Y/N)	Character
Education_Qualification	Graduate/Undergraduate	String
Gender	Male/Female	Character
Dependents	Number of Dependents	Integer
Applicant_Income	ApplicantIncome	Integer
CO_Aplicant_Income	Co-applicant Income	Integer
Loan_Amount	Amount of Loan in thousands	Integer
Loan_Amount_Term	Loan Term in months	Integer
Credit_History	Guidelines of Credit History	Integer

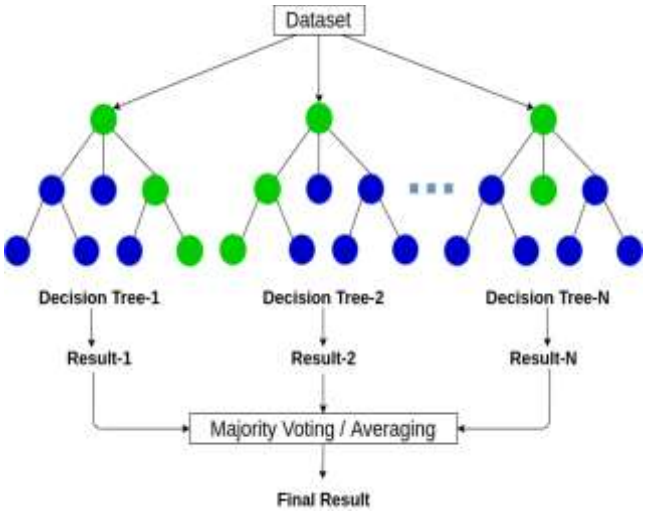

Figure.2: Random Forest Model

B. Support Vector Machine:

The SVM is a machine readable learning technique that can split, retrieve, and identify data. Drawing a straight line amid two classes is how the SVM segment line works[14,15,16]. All data points on one lateral of the line will be recorded as single category, while all topics on the supplementary side of the line will be written as subsequently another category and are shown in Figure.3.

C. K-Nearest Neighbor:

The K-NN algorithm perceives resemblances among new instances/data and existing cases and allots every recent instance to a category that most closely look like current categories, as shown in Figure 4. The K-NN algorithm retains current data and categorizes novel data points established on their comparisons[17,18]. This means, that the KNN technique can quickly sort data into a meaningful segment regardless of where it comes from. Figure.3 Support Vector Machine Model.

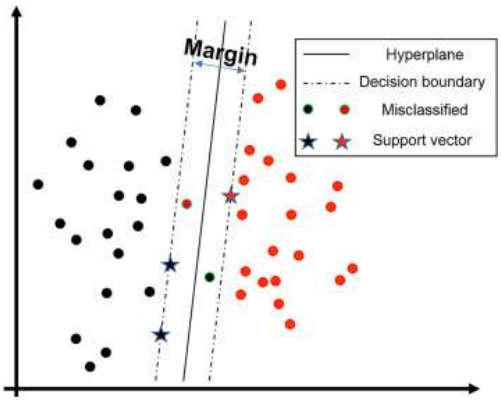

Figure.3: Support Vector Machine Model

Although the K-NN technique will be used for editing and undoing, it's most typically used to solve editing problems. This method is non-parametric, so it does not make any assumptions about the underlying data.

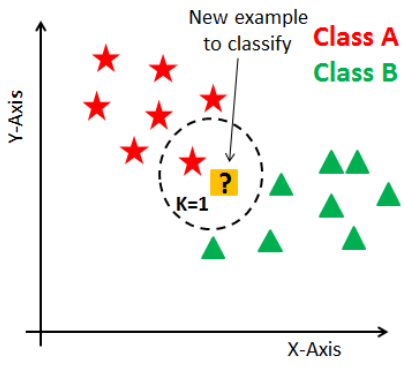

Figure.4: K-Nearest Neighbor Model

D. Logistic regression:

The likelihood of a finite number of outcomes, often two, is modelled using logistic regression, comparable to linear regression.

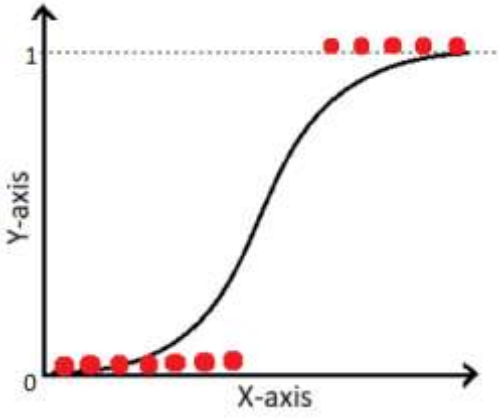

Figure.5: Logistic Regression Model

When modelling probabilities of outcomes, logistic regression is preferred over linear regression for various reasons[19]. A logistic equation is designed so that the output values can only be between 0 and 1which is shown in Figure.5.

IV. RESULTS AND DISCUSSIONS

To forecast loan approvals for loan requests, three machine learning algorithms are applied to the test-data. The data is used for training is 70% and testing 30%. Several machine learning methods' accuracy is calculated and compared. The test dataset accuracy for the proposed ML models is shown in Table.2 and in Figure.6.

Table.2: Test dataset Accuracy

Algorithm	Accuracy in(%)
SVM	73.2
K-Nearest Neighbor	68
Random forest	81
Logistic regression	77

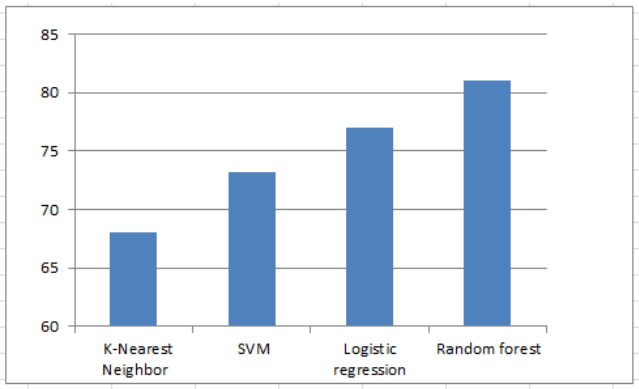

Figure.6: Test Accuracy of ML algorithms

V. CONCLUSION

This study proposes the use of machine learning algorithms to forecast loan acceptance. Customers' loan approval status is predicted using three machine learning algorithms. SVM has 73.2% accuracy, K-NN has 68% accuracy, Random Forest has 81% accuracy, and Logistic Regression has 77% accuracy. For loan prediction, the Random Forest algorithm has the uppermost accuracy. In the future, a thorough investigation of alternative machine learning methods for loan endorsement forecast.

REFERENCES

[1] Arun, Kumar, Garg Ishan, and Kaur Sanmeet. "Loan approval prediction based on machine learning approach." IOSR J. Comput. Eng 18.3 (2016): 18-21.

[2] Kumar, R., et al. "Prediction of loan approval using machine learning." Int J Adv Sci Technol 28 (2019): 455-460.

[3] Sheikh, Mohammad Ahmad, Amit Kumar Goel, and Tapas Kumar. "An Approach for Prediction of Loan Approval using Machine Learning Algorithm." 2020 International Conference on Electronics and Sustainable Communication Systems (ICESC). IEEE, 2020.

[4] Lopes, Rogerio Gomes, Marcelo Ladeira, and Rommel Novaes Carvalho. "Use of machine learning techniques in the prediction of credit recovery." Advances in Science, Technology and Engineering Systems Journal 2.3 (2017): 1432-1442.

[5] Lohokare, Jay, Reshul Dani, and Sumedh Sontakke. "Automated data collection for credit score calculation based on financial transactions and social media." 2017 International Conference on Emerging Trends & Innovation in ICT (ICEI). IEEE, 2017.,

[6] Hrushikesava, Raju Dr S., et al. "Tourism Enhancer App: User-Friendliness of a Map with Relevant Features." IOP Conference Series: Materials Science and Engineering. Vol. 981. No. 2. IOP Publishing, 2020.

[7] Tumuluru, Praveen, et al. "A Review of Machine Learning Techniques for Breast Cancer Diagnosis in Medical Applications." 2019 Third International conference on I-SMAC (IoT in Social, Mobile, Analytics and Cloud)(I-SMAC). IEEE, 2019.

[8] Ramani, B. Lakshmi, and Praveen Tumuluru. "Deep learning and fuzzy rule-based hybrid fusion model for data classification." International Journal of Recent Technology and Engineering 8.2 (2019): 3205-3213.

[9] Tumuluru, Praveen, and Bhramaramba Ravi. "Chronological grasshopper optimization algorithm-based gene selection and cancer classification." Journal of Advanced Research in Dynamical & Control Systems 10.3 (2018).

[10] Tumuluru, Praveen, and Bhramaramba Ravi. "GOA-based DBN: Grasshopper optimization algorithm-based deep belief neural networks for cancer classification." International Journal of Applied Engineering Research 12.24 (2017): 14218-14231.

[11] Jampani, Varun, Martin Kiefel, and Peter V. Gehler. "Learning sparse high dimensional filters: Image filtering, dense crfs and bilateral neural networks." Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2016.

[12] Praveen, T. "Shortest path Enhancement using improved Bellman Ford Algorithm in PPI Journal of Engineering and Applied Sciences." Medwell Journals [2017].

[13] Tumuluru, Praveen, Bhramaramba Ravi, and Sujatha Ch. "A survey on identification of protein complexes in protein–protein interaction data: Methods and evaluation." Computational Intelligence Techniques for Comparative Genomics. Springer, Singapore, 2015. 57-72.

[14] B. Venkateswarlu, V.Viswanath Shenoi, Praveen Tumuluru, "CAViaR-WS-based HAN: conditional autoregressive value at risk-water sailfish-based hierarchical attention network for emotion classification in COVID-19 text review data" Social Network Analysis and Mining [2022], https://doi.org/10.1007/s13278-021-00843-y.

[15] Chen, Joy Iong-Zong, and Kong-Long Lai. "Deep Convolution Neural Network Model for Credit-Card Fraud Detection and Alert." Journal of Artificial Intelligence 3, no. 02 (2021): 101-112.

[16] Tripathi, Milan. "Sentiment Analysis of Nepali COVID19 Tweets Using NB, SVM AND LSTM." Journal of Artificial Intelligence 3, no. 03 (2021): 151-168.

[17] Kottursamy, Kottilingam. "A review on finding efficient approach to detect customer emotion analysis using deep learning analysis." Journal of Trends in Computer Science and Smart Technology 3, no. 2 (2021): 95-113.

[18] Anand, C. "Comparison of Stock Price Prediction Models using Pre-trained Neural Networks." Journal of Ubiquitous Computing and Communication Technologies (UCCT) 3, no. 02 (2021): 122-134.

[19] Andi, Hari Krishnan. "An Accurate Bitcoin Price Prediction using logistic regression with LSTM Machine Learning model." Journal of Soft Computing Paradigm 3, no. 3 (2021): 205-217.