ISSN (Online): 2455-3662
EPRA International Journal of Multidisciplinary Research (IJMR) - Peer Reviewed Journal
Volume: 10| Issue: 5| May 2024|| Journal DOI: 10.36713/epra2013 || SJIF Impact Factor 2024: 8.402 || ISI Value: 1.188

PREDICTIVE MODELING FOR LOAN APPROVAL: A MACHINE LEARNING APPROACH

Valmiki Sarath Kumar[1], K. Vijayalakshmi[2]
[1]*Schoolof Computer Science and Applications, REVA University, Bangalore, Karnataka, India*
[2]*School of Computer Science and Applications, REVA University, Bangalore, Karnataka, India*

Article DOI: https://doi.org/10.36713/epra17042
DOI No: 10.36713/epra17042

ABSTRACT

Exploring machine learning approaches to enhance the effectiveness and precision of procedures related to bank loan approval. This investigation encompasses various methods such as logistic regression, decision trees, linear regression, as well as GaussianNB, Random Forest, and SVM. Utilizing a substantial dataset containing past loan applications and diverse applicant attributes like demographics, credit scores, income levels, and employment histories. The research endeavors to evaluate the recall, accuracy, precision, and F1-score metrics of various algorithms. Additionally, it investigates the interpretability and transparency of machine learning models to offer further insight into the variables affecting decisions on loan acceptance. The study emphasizes the efficacy of logistic regression, which outperformed SVM (77%), GaussianNB (78%), random forests (78%), and decision trees (69%), achieving the highest accuracy of 80% in loan approval. By implementing this model, we can enhance ML-driven loan approval processes within the banking industry, thereby elevating decision-making standards and enhancing consumer satisfaction.

KEYWORDS— *Machine Learning Algorithms, Loan Approval, LogisticRegression, DecisionTree, Linear Regression, GaussianNB, RandomForest, SupportVectorMachine (SVM), Decision-Making,*

I. INTRODUCTION

The loan approval process holds significance for both lenders and applicants. Traditional evaluation methods, while somewhat effective in specific scenarios, often fall short in meeting the rapid and accurate demands of the contemporary market. This is where predictive modeling powered by machine learning stands out as an innovative solution. The foremost priority for any bank is to guarantee the security of its assets [1]. In the digital era, the banking industry heavily relies on advanced technologies. Debt is widely regarded as the primary service provided by financial institutions and a significant revenue stream for companies. The assessment of credit risk emerges as a critical factor demanding meticulous attention [2]. Daily a considerable number of individuals apply for personal loans with many receiving them from different institutions. Interest-free loans constitute the primary revenue stream for banks [3]. Machine learning (ML), a subset of artificial intelligence (AI) investigates statistical models and techniques enabling computers to learn from data, thereby making predictions or assessments without explicit programming. This process is often termed as teaching computers to perform tasks more autonomously relies on data or expertise. It encompasses analyzing data using statistical models and algorithms to recognize patterns and draw conclusions or predictions. Machine learning methods are essential for constructing models that reliably categorize loan applicants according to their ability to repay loans [4].

Data mining techniques prove invaluable in this domain involving the extraction of valuable insights from vast and unorganized datasets. These insights aid decision-making processes by executing crucial operations. In the realm of loan approval prediction numerous conventional data mining techniques exist alongside machine learning algorithms designed to automate the prediction of loan statuses. These algorithms effectively and quickly determine loan eligibility. Among them are decision trees, logistic regression, random forests, and gradient boosting. This study aims to mitigate the risk linked to loan approval by predicting the loan status through analysis of various loan attributes or characteristics. The Loan Prediction System assigns weight to each attribute involved in loan processing automatically. These weights are then applied to newly acquired test data during processing [5]. Loan prediction offers significant advantages for banks, employees, and applicants alike. The primary goal of this work is to offer a practical and effective approach for selecting the optimal course of action. With the loan prediction system's ability to automatically assess the importance of each parameter, users can make optimal decisions. We can schedule appointments to review customer status and ascertain their eligibility. The system efficiently prioritizes specific applications for assessment. The target audience for this study report is the governing authority of banks or financial institutions. It's important to note that the process remains confidential with no ability for any involved parties to alter it.

ISSN (Online): 2455-3662
EPRA International Journal of Multidisciplinary Research (IJMR) - Peer Reviewed Journal
Volume: 10| Issue: 5| May 2024|| Journal DOI: 10.36713/epra2013 || SJIF Impact Factor 2024: 8.402 || ISI Value: 1.188

II.LITERATURE REVIEW

To determine the likelihood of approval for a bank loan Vandana Sharma et al [6] investigates the feasibility of employing ML techniques for credit risk assessment amidst persistent worries regarding loan defaults. It scrutinizes existing methodologies and underscores the significance of precise credit evaluation, particularly given the rising prominence of peer-to-peer lending platforms. To enhance predictive efficacy, the proposed model incorporates logistic regression, feature engineering, and data refinement. Nevertheless, it acknowledges various potential limitations such as issues concerning data integrity, logistic regression's limitations in detecting complex relationships, and the model's dependence on specific datasets and credit scoring mechanisms.

Mohammad Ahmad Sheikh et al [7] study report underscores the importance of accurate predictions for maximizing profits in the banking industry through the application of logistic regression. It leverages Kaggle data and focuses on customer attribute beyond mere account validation. The proposed methodology mitigates default risks and enhances operational efficiency and customer contentment by automating loan approval procedures. Apart from feature engineering and preprocessing missing variable imputation is employed. Logistic regression is selected as the model due to its commendable accuracy of 0.811 on the test dataset and its effectiveness in classification endeavours. Nancy Deborah R et al [8] delves into the potential uses of machine learning methods, such as Support Vector Classifiers (SVM), in predicting loan approval status. SVM exhibited an 83% accuracy rate in this context. The research underscores the importance of employing dependable prediction models when addressing the obstacles banks encounter in evaluating loan applications and mitigating credit risk. It underscores the significance of considering various factors to achieve precise forecasts of loan status, including data quality, hyperparameter tuning, and potential biases in both data and models. Furthermore, it illustrates how SVM serves as a valuable tool that can be enhanced over time by incorporating new features and data sources to enhance prediction accuracy and adaptability to evolving market dynamics.

Sk. Sharmila et al [9] study delves into utilizing a Decision Tree Classifier for bank loan approval, presenting a fresh perspective on loan prediction. It underscores the growing inclination towards employing machine learning models in loan assessment and the demand for more effective decision-making algorithms. The research advocates for the Decision Tree method citing its clarity and effectiveness, while scrutinizing conventional techniques such as linear regression and Gaussian Naive Bayes. Following training on past loan data, the suggested model exhibits significant promise in expediting loan approval procedures, boasting an impressive accuracy rate of 95% and minimal loss at 0.09%. Anshika Gupta et al [10] utilizes machine learning methods to tackle the increasing need for streamlined loan approval procedures within the banking sector. It focuses on forecasting the likelihood of loan approval through supervised learning techniques, specifically random forest, and logistic regression algorithms, leveraging application data. The report underscores the indispensability of artificial intelligence (AI) in expediting banking operations and mitigating risks linked to manual assessments. Through feature engineering and data preprocessing, the study constructs a dataset encompassing credit history, income, and educational background. It underscores the significance of predictive analytics in augmenting decision-making processes and curbing fraudulent activities, advocating for the advancement and integration of such systems in forthcoming endeavors.

Praveen Tumuluru et al [11] Analyzes the application of machine learning methods for predicting loan defaults, a critical issue affecting the overall financial stability and success of banks. The research evaluates various techniques including Random Forest, Support Vector Machine, K-Nearest Neighbor, and Logistic Regression to evaluate the risk involved in loan approval decisions. Its central premise is that automating the process reduces bank risk and losses. Random Forest emerges as the top performer, achieving an accuracy of 81% in predicting loan approval, suggesting its potential for future loan forecasting. The research recommends exploring additional machine learning approaches to enhance prediction accuracy further. Krishna Mridha et al [12] delves into the utilization of machine learning algorithms by financial institutions to streamline the loan approval process, thereby mitigating risks associated with customer behavior. It examines various classification techniques, such as logistic regression, and assesses the accuracy of each model using data sourced from Kaggle. Future investigations will focus on crafting a hybrid model that integrates deep learning methodologies to enhance prediction accuracy while emphasizing crucial features related to loan repayment capacity.

Trishita Saha et al [13] delves into the utilization of machine learning algorithms by financial institutions to streamline the loan approval process, thereby mitigating risks associated with customer behavior. It examines various classification techniques, such as logistic regression, and assesses the accuracy of each model using data sourced from Kaggle. Future investigations will focus on crafting a hybrid model that integrates deep learning methodologies to enhance prediction accuracy while emphasizing crucial features related to loan repayment capacity. Ch. Naveen kumar et al [14] the importance of loan management within the banking sector, particularly examining the impact of defaults on bank profitability. It underscores the vital nature of accurately predicting loan defaulters and emphasizes the growing role of machine learning techniques in addressing this challenge. Additionally, it underscores the necessity for innovative approaches to enhance the precision of loan eligibility prediction, alongside outlining several existing industry practices. Gaurav Parmar et al [15] the specific study article, previous research on data rebalancing methods aimed at mitigating bias in sensitive categories such as age, gender,

ISSN (Online): 2455-3662
EPRA International Journal of Multidisciplinary Research (IJMR) - Peer Reviewed Journal
Volume: 10| Issue: 5| May 2024|| Journal DOI: 10.36713/epra2013 || SJIF Impact Factor 2024: 8.402 || ISI Value: 1.188

and race would certainly be referenced. This review would scrutinize the limitations of existing techniques and underscore the risks associated with both overfitting and underfitting machine learning models.

III. METHODOLOGY

The loan approval prediction strategy utilizes statistical methods and historical data to forecast the outcome of loan approval. Algorithms that we are going to use in the paper is Logistic Regression, Linear Regression, Decision Tree, SVM, Random Forest and GaussionNB. All these algorithms are trained on historical data to learn the relationships between the different factors that affect loan approval. The loan approval prediction methodology can be a valuable tool forbanks. Banks can enhance their evaluation of loan applications and lower the chances of loan defaults by implementing this measure.

A. Model Planning

The first step involves importing the necessary libraries as shown in “fig.1” for machine learning tasks. These packages encompass NumPy for numerical computations, pandas for data processing, and scikit-learn for machine learning techniques. The next step is the data cleaning process which involves addressing concerns and ensuring data quality by performing dataset cleaning. Data often contains errors, omissions, and inconsistencies that need to be resolved. Following this the subsequent stage is label encoding aims to convert categorical variables into numerical values making them compatible with models for processing. Our objective is to encode the categorical variables within the dataset. The next stage divides the dataset into two categories: training and testing. The training set is utilized to train the model while the testing set is used to assess its performance on unseen data. The training set contains 0.65% of the dataset, whereas the testing set contains 0.35%. The training data is then used to develop various machine learning models which include linear regression, logistic regression, decision trees, random forests, support vector machines, or Naive Bayes. After the training phase, these models are assessed on a testing set to determine their effectiveness with new data.

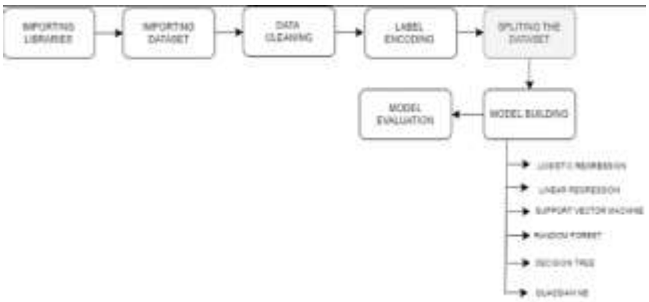

Fig. 1. Model planning of the models

B. Dataset

For all the machine learning models we need to build models by using the dataset divide them into train and test data. We need to train the model by using train data and then applying the model on the test data. In this Dataset we have 12 columns and 614 customer records. All these records are in the csv format. We can use panda’s library for reading the file and by using python we can apply all the algorithms. In that dataset Status is the target variable and the other variables- “Gender”, “Married”, “Dependents”, “Education”, “Self-Employed”, “Applicant Income”, “Co-applicant Income”, “Loan Amount”, “Term”, “Credit History”, “Area”.

TABLE I. DATASET DESCRIPTION VARIABLES AND TYPES.

VARIABLE NAME	DESCRIPTION	TYPE
Gender	Male / Female	Character
Married	Applicant married (Y/N)	Character
Dependents	Number of dependents	Integer
Education	Educated/ Not Education	String
Self Employed	Self Employed(Y/N)	Character
Application Income	Applicant income	Integer
Co-Applicant Income	Coapplicant income	Integer
Loan Amount	Loan amount in thousands	Integer
Term	Term of loan in months	Integer
Credit History	Credit history of the applicant	Integer
Area	Urban/Semi Urban/Rural	String
Status	Loan Approved(Y/N)	Character

From the above Table I, we can know the features of the variables which are present in the data set. The next step is to handle the nullvalues. This is done by different methods, but we have enough data, so we are removing the null values. All the records of the applicants are in the form of categorical and numerical data. The dataset has missing values and null values. Before the analysis we need to normalize the dataset by removing the null values.

C. Data preprocessing

Before removing the null values, the value count of every variable that present in the data set is shown in Table II.

TABLE II. DATASET WITH NULL VALUES

Gender	601
Married	611
Dependents	599
Education	614
Self Employed	582
Application Income	614
Co-Applicant Income	614
Loan Amount	614
Term	600
Credit History	564
Area	614
Status	614
Dtype: int64	

We can’t handle imbalance data and we can’t apply machine learning algorithms also to the imbalance records, so we need to normalize this data by removing the null values. After removing the null values, the value count of every variable is shown in Table III.

ISSN (Online): 2455-3662
EPRA International Journal of Multidisciplinary Research (IJMR) - Peer Reviewed Journal
Volume: 10| Issue: 5| May 2024|| Journal DOI: 10.36713/epra2013 || SJIF Impact Factor 2024: 8.402 || ISI Value: 1.188

TABLE III. DATASET WITHOUT NULL VALUES

Gender	499
Married	499
Dependents	499
Education	499
Self Employed	499
Application Income	499
Co-Applicant Income	499
Loan Amount	499
Term	499
Credit History	499
Area	499
Status	499
Dtype: int64	

Normalizing the dataset is done. The dataset contains both training and testing segments. By using the train data, we need to train the model and test the model by using the test data.

As shown in the "fig.2" y-axis depicts the count of dependents, while the x-axis indicates the number of dependents. Individuals without dependents exhibited the highest loan approval rate, hovering around 92%. The approval rate notably declines for applicants with one dependent, dropping to approximately 66%. Candidates with two or three dependents consistently experience a lower acceptance rate (approximately 55%). For applicants with more than three dependents, the loan approval rate is the lowest, standing at around 28%.

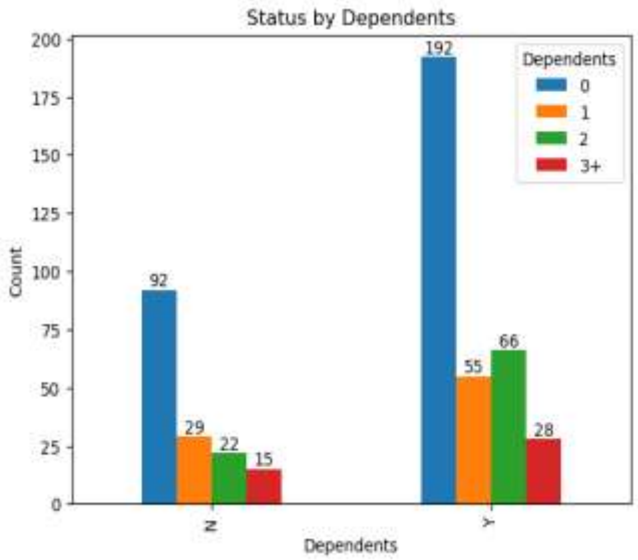

Fig. 2. Status by Dependents

As shown in "fig.3" the graph illustrates self-employment status, with the x-axis depicting two options: "Yes" and "No". It indicates a decline in self-employment, with fewer individuals opting for it. The blue bars represent the non-self-employed, with a maximum count of 24. The orange line denotes the self-employed, peaking at 134 individuals. Model planning of the models.

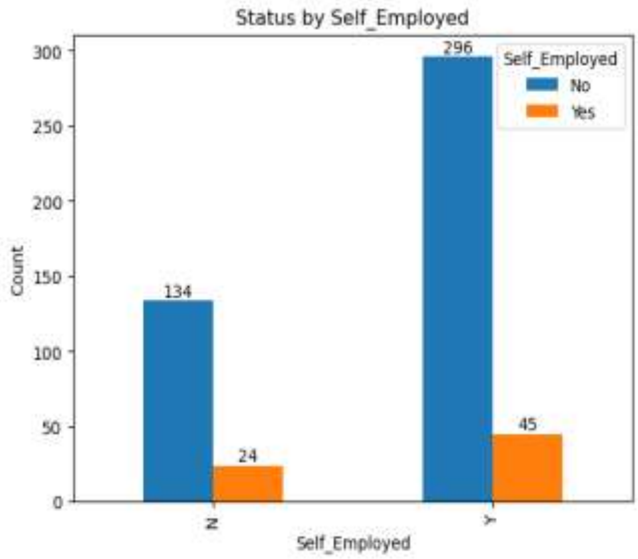

Fig. 3. Status by Self employment

The heatmap correlations between each variable in the dataset are shown in "Fig.4" below. So that we can easily do further analysis.

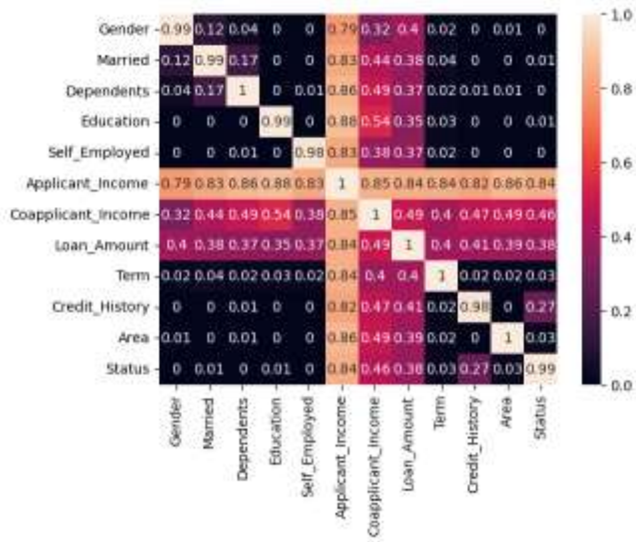

Fig. 4. Representing the correlation between attributes using the heatmap

D. Logistic Regression

Supervised machine learning methods like logistic regression are utilized to address binary classification tasks by evaluating the likelihood of an event, outcome, or observation. Equation 1 is the mathematical representation of logistic regression. This results in a dichotomous output, typically represented as true/false, 0/1, yes/no. Logistic regression examines the association between independent variables to categorize data into various groups. Commonly employed in predictive modelling, this approach employs mathematical techniques to ascertain the probability of an event belonging to a specific category.

$$f(x) = \frac{1}{1+e^{-x}} \quad (1)$$

ISSN (Online): 2455-3662
EPRA International Journal of Multidisciplinary Research (IJMR) - Peer Reviewed Journal
Volume: 10| Issue: 5| May 2024|| Journal DOI: 10.36713/epra2013 || SJIF Impact Factor 2024: 8.402 || ISI Value: 1.188

E. Decision Tree

Decision trees serve as a supervised learning technique commonly employed for classification tasks, but they can also address regression problems. Structurally resembling a tree, they feature leaf nodes representing individual outcomes, decision nodes for branching decision rules, and root nodes for dataset attributes. Comprising two types of nodes Decision and Leaf decision trees utilize decision nodes for making choices and leaf nodes to display decision outcomes. Dataset attributes guide the testing and decision-making process. Graphical representation of decision tree is shown in "Fig .5".

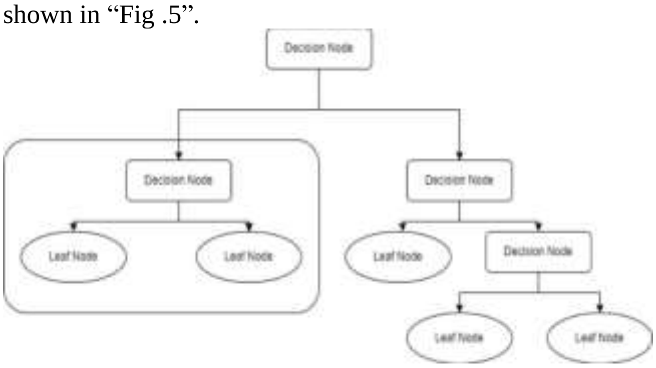

Fig. 5. Decision Tree

F. Linear Regression

Linear regression, a statistical technique commonly employed in machine learning and data science predictive analysis, establishes a direct relationship between an independent variable, which remains constant amidst changing factors, and a dependent variable, whose value is influenced by changes in the independent variable. In essence, this method utilizes mathematical modelling to predict the values of continuous or numerical variables such as age, income, sales, and product price, employing supervised learning techniques to forecast outcomes.

$$y = \beta_0 + \beta_1 X + \beta_2 X + \dots\dots\dots + \beta_n X \quad (2)$$

Where in a linear regression model Equation 2, Y represents the dependent variable, and the independent variables are represented by X1, X2, ..., Xp. The intercept is symbolized by β0, and the slopes by β1, β2, ..., βn.

G. Random Forest

The Random Forest algorithm stands as one of the most effective machine learning techniques for tree-based learning. During the training phase, it generates multiple Decision Trees as shown in "Fig. 6". To mitigate overfitting and foster diversity, every tree is built by randomly selecting features and data samples. This approach fosters greater variation among the trees, leading to improved prediction performance. In regression tasks, the algorithm computes the average output of all trees, while in classification problems, it aggregates the average votes from each tree. By integrating insights from multiple trees, Random Forest facilitates robust and precise decision-making, ensuring reliable outcomes [16].

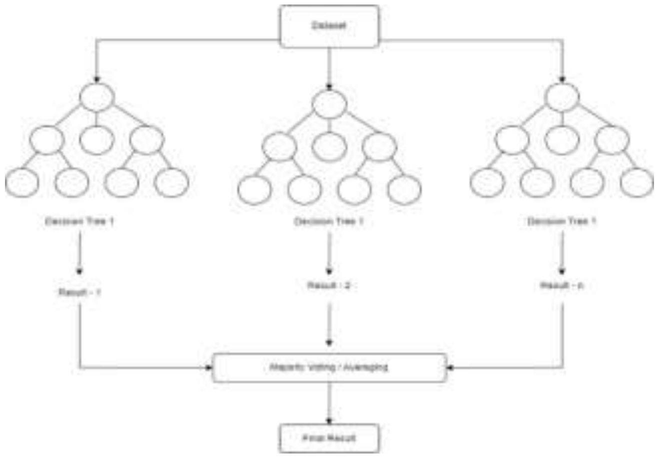

Fig. 6. Random Forest

H. Navie Bayes model GaussionNB

Gaussian Naive Bayes applies the Naive Bayes algorithm specifically to data that follows a normal (Gaussian) distribution. In this approach, it assumes that the probability of each x_i within y_k aligns with the Gaussian Distribution, expressed in Equation 3.

$$e^{-\frac{(x-\mu)^2}{2\sigma^2}} \quad 1/(\sigma\sqrt{2\pi}) = P(x\_i/y) \qquad (3)$$

This approach calculates the posterior probability for each class and then assigns the data point to the class with the highest likelihood to classify new data points.

H. Support Vector Machine (SVM)

Support vector machines (SVMs) play a vital role in various machine learning tasks like regression, linear or nonlinear classification, and outlier detection. They are indispensable in applications such as text and image classification, facial and handwriting recognition, anomaly detection, gene expression analysis, and spam detection. SVMs excel in handling high-dimensional data and capturing nonlinear relationships, making them versatile and robust. These methods aim to identify the hyperplane in the feature space that optimally separates classes, thus reducing the distance between them effectively.

IV. MODEL EVALUATION

The model assesses an applicant's risk of defaulting on a loan by considering various factors, such as their historical loan information. Some key metrics used to evaluate how effectively the model categorizes loan applicants into accepted or denied groups include recall, accuracy, precision, and F1-score.

A. Confusion Matrix

A confusion matrix is a useful tool in machine learning for assessing the performance of a classification model by providing insights into its accuracy. As shown in the "fig.7" It helps differentiate between true positives, true negatives, false positives, and false negatives, offering a clear evaluation of the model's effectiveness.

ISSN (Online): 2455-3662
EPRA International Journal of Multidisciplinary Research (IJMR) - Peer Reviewed Journal
Volume: 10| Issue: 5| May 2024|| Journal DOI: 10.36713/epra2013 || SJIF Impact Factor 2024: 8.402 || ISI Value: 1.188

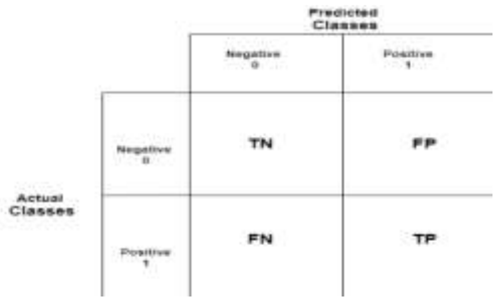

Fig. 7. Confusion Mtrix

B. Accuracy

The accuracy of the model has been evaluated using predetermined measures. While a balanced class model exhibits outstanding accuracy, an imbalanced class model shows significant inaccuracies. The mathematical representation of accuracy is represented as Equation 4.

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN} \tag{4}$$

C. Precision:

The forecast accuracy level of an optimistic model is determined by Equation 5. By dividing the total number of correctly predicted positive outcomes by the total number of positive outcomes incorrectly forecasted.

$$Precision = \frac{TP}{TP+FP} \tag{5}$$

D. Recall:

Recall value, also known as sensitivity, measures the proportion of correctly identified positive events compared to all actual positive cases. In Formula 6, the denominator, (TP + FN), signifies the total number of positive instances. This metric enhances identification accuracy and influences the probability of the model overlooking positive occurrences.

$$Recall(Sensitivity) = \frac{TP}{TP+FN} \tag{6}$$

E. F1 Score:

The F1 Score serves as a key indicator of a model's optimal performance level, being a harmonic mean of recall and accuracy. Attaining the highest F1 Score necessitates high values for both recall and accuracy. Any declines in either recall or accuracy can lead to a notable decrease in the final F1 score. Equation 7 for the F1 Score employs the sum of recall and accuracy as its numerator. Consequently, a model that effectively predicts positive events and avoids underestimating positives by predicting negatives can achieve a high F1 score (accuracy + recall).

$$F1 - Score = \frac{2TP}{2TP+FP+FN} \tag{7}$$

V. RESULTS

Below Table IV displays the accuracies achieved by the decision tree, random forest, logistic regression, support vector machine, and Gaussian Naive Bayes models.

TABLE IV. COMPARING THE PERFORMANCE OF EACH MODEL

Algorithms	Accuracy
Logistic Regression	80%
Decision Tree	69%
Random Forest	78%
GaussianNB	78%
SVM	77%

TABLE V. LINEAR REGRESSION MODEL METRICS RESULTS

MSE	0.154
RMSE	0.392
Train Set Score	0.27
Test Set Score	0.32

Among various machine learning techniques evaluated, logistic regression exhibited the highest accuracy at 80%, surpassing SVM (77%), GaussianNB (78%), random forests (78%), and decision trees (69%). Notably, linear regression displayed inferior performance based on metrics like MSE, RMSE, and scores. Particularly in predicting loan approvals, logistic regression emerged as the most accurate model.

VI. CONCLUSION

In the loan approval project, several machine learning models wereevaluated, and logistic regression emerged as the most accurate model with an accuracy of 80%. The logistic regression model demonstrated superior performance compared to the SVM, Decision Tree, Random Forest, GaussianNB, and Linear Regression models in accurately predicting loan approval. There are several reasons why the logistic regression model has achieved higher accuracy. Originally, logistic regression was frequently and consistently employed to address binary classification problems. It serves as an effective tool for determining whether to accept or reject something, as it is specifically designed to estimate the probability of an event occurring. Logistic regression tends to exhibit lower overfitting compared to more complex models such as decision trees and random forests. When a model learns the training dataset too well, it is said to be overfitting and has poor generalization to new data. Logistic regression is efficient and somewhat understandable compared to other models like support vector machines (SVM) or random forests. If the relationship between the input variables and the loan approval decision is relatively linear, logistic regression can effectively capture this pattern and make accurate predictions. In conclusion, the loan approval project found that logistic regression outperformed other models, achieving the highest accuracy with 80%.From the proper view of analysis this system can be used perfectly for detection of clients who are eligible for approval of loan. In the future, there may be opportunities to expand upon this research further, leading to improved software upgrades that enhance correctness, security, and reliability.

ISSN (Online): 2455-3662
EPRA International Journal of Multidisciplinary Research (IJMR) - Peer Reviewed Journal
Volume: 10| Issue: 5| May 2024|| Journal DOI: 10.36713/epra2013 || SJIF Impact Factor 2024: 8.402 || ISI Value: 1.188

VII. REFERENCES

1. *Suliman Mohamed Fati, "Machine Learning-Based Prediction Model for Loan Status Approval",2021.*
2. *Malika Becha, Olfa Dridi, Oumayma Riabi, Yasmine Benmessaoud, "Use of Machine Learning Techniques in Financial Forecasting", 2020.*
3. *C. Prasanth, R. Praveen Kumar, A. Rangesh, N. Sasmitha, Dhiyanesh B," Intelligent Loan Eligibility and Approval System based on Random Forest Algorithm using Machine Learning",2023.*
4. *Hamid, A.J., and Ahmed, T.M. "Developing prediction model of loan risk in banks using data mining", 2016.*
5. *Kumar Arun, Garg Ishan, Kaur Sanmeet," Loan Approval Prediction based on Machine Learning Approach", 2016.*
6. *Vandana Sharma, Amit Singh, Ashendra Kumar Saxena and Vineet Saxena" A Logistic Regression Based Credit Risk Assessment Using WoE Bining and Enhanced Feature Engineering Approach ANOVA and Chi-Square",2023.*
7. *Mohammad Ahmad Sheikh, Amit Kumar Goel, Tapas Kumar," An Approach for Prediction of Loan Approval using Machine Learning Algorithm",2020.*
8. *Nancy Deborah R, Alwyn Rajiv S, Vinora A, Manjula Devi C, Mohammed Arif S, Mohammed Arif G S," An Efficient Loan Approval Status Prediction Using Machine Learning",2023.*
9. *Sk. Sharmila, P V S Sandhya, P. Suhani kousar, P. Anuradha, S. Deekshitha," Bank Loan Approval using Machine Learning", 2024.*
10. *Anshika Gupta, Vinay Pant , Sudhanshu Kumar and Pravesh Kumar Bansal," Bank Loan Prediction System using Machine Learning",2020.*
11. *Praveen Tumuluru, Lakshmi Ramani Burra, M.Loukya, S.Bhavana, CH.M.H.SaiBaba, N Sunanda," Comparative Analysis of Customer Loan Approval Prediction using Machine Learning Algorithms",2022.*
12. *Krishna Mridha, Dipayan Barua, Meghla Monir Shorna, Hasan Nouman Nouman, Md Hasanul Kabir, Ajay Vikram Singh, "Credit Approval Decision using Machine Learning Algorithms",2022.*
13. *Trishita Saha, Saroj Kumar Biswas, Saptarsi Sanyal, Neeta Verma, Biswajit Purkayastha, "Credit Risk Prediction using Extra Tree Ensembling Technique with Genetic Algorithm", 2023.*
14. *Ch. Naveen kumar, D. keerthana , M Kavitha , M Kalyani, "Customer Loan Eligibility Prediction using Machine Learning Algorithms in Banking Sector",2022.*
15. *Gaurav Parmar, Rimi Gupta, Tejas Bhatt, G.J. Sahani, Brijeshkumar Y. Panchal, Hiren Patel, "Data Re-Balancing using Fuzzy Clustering and SMOT Mechanism",2023.*
16. *Kusumlata Bhatt, Pranchal Sharma, Megha Verma, Dr Kadambri Agarwal, "Loan Status Prediction in the Banking Sector using Machine Learning",2023.*
17. *Kumar Arun et al., 2021 proposed Loan Approval Prediction based on Machine Learning Approach, OSR Journal of Computer Engineering (IOSR-JCE).*
18. *Ramachandra H V et al., 2021 proposed Design and Simulation of Loan Approval Prediction Model using AWS Platform, International Conference on Emerging Smart Computing and Informatics (ESCI).*
19. *Ms. Kathe Rutika Pramod1, Ms. Dapse Punam Laxman2, Ms. Panhale Sakshi Dattatray3, Ms. Avhad Pooja Prakash4, Mr. Ghorpade Dinesh B.5, 1-4 Information Technology Engineering SVIT, Nashik Maharashtra, India. 5 Information Technology (Assistant Proffessor) SVIT, Nashik Maharashtra, India an Approach for Prediction of Loan Approval Using Machine Learning Algorithm 2021.*
20. *Mahankali Gopinatha K. Srinivas Shankar Maheep and R. Sethuraman, 2021 proposed Customer Loan Approval Prediction Using Logistic Regression, Smart Intelligent Computing and Communication Technology.*