© 2022 IJNRD | Volume 7, Issue 12 December 2022 | ISSN: 2456-4184 | IJNRD.ORG

IJNRD.ORG ISSN : 2456-4184

INTERNATIONAL JOURNAL OF NOVEL RESEARCH AND DEVELOPMENT (IJNRD) | IJNRD.ORG

An International Open Access, Peer-reviewed, Refereed Journal

Loan Default Prediction using Machine Learning Techniques

Shaistha Tabassum
Dept. of CSE
Dr. Ambedkar Institute of Technology
Bangalore, Karnataka
Shaisthatab97@gmail.com

Dr. Shylaja K R
Professor
Dept. of CSE
Dr. Ambedkar Institute of Technology
Bangalore, Karnataka
shylaja.kr.cs@drait.edu.in

***Abstract*-- Technology has made a lot of improvements, and the banking industry is no exception. Every day, there are more applications to approve loans. While evaluating an application for getting approved, they must take into consideration a few bank policies. The bank is required to determine what is ideal for approval based on a few criteria. To meticulously check out each individual prior recommending them for loan approval is challenging and risky. Based on the cibil score and history of the individual to whom the loan amount was previously accredited, we utilise machine learning in this study to identify who may be trusted for a loan. The main goal of this approach is to forecast whether or not the loan will be sanctioned.**

***Index Terms* - Loan Prediction, KNN, Naïve Bayes, Logistic Regression.**

I. INTRODUCTION

People would prefer to make loans applications online due to the daily data growth driven on by banking firm's modernization. As a typical tool for analyzing data analysis, artificial intelligence (AI) is becoming more and more prominent. People from numerous sectors are already using AI calculations to solve the problems depending on their knowledge of the relevant industries. While granting loans, banks are having a difficult time. The financial managers monitor a large volume of applications daily, which would be challenging and increases the chance of mistakes. The majority of financial institutions make profit through lending, however it could be problematic to choose desirable individuals from the pool of applications. A bank could experience a serious loss due to one oversight. Almost all financial institutions main line of business is loan management. Out of all applicants, one would secure a loan as an outcome of this program. An efficient, fair approach that save the bank time comprises categorizing every applicant during evaluation. The successful completion of every other customer requirements by the authorities concerned favors the customers. With the aid of this approach, software applications can be assigned priority consideration. In order to ascertain a customer's loan condition and formulate strategies, it is possible to use the loan prediction machine learning model. In [1] The effectiveness of data mining approaches for predicting and categorising a client's credit score as outstanding or low was analyzed by the authors. In order to review a loan application and determine whether to approve it or not, banks must perform a critical and essential procedure known as prediction or evaluation of a debt default. This procedure will assist them in identifying loan applicants who may eventually prove to be defaulters. Developing a model that will consider all of an applicant's attributes into consideration and produce an outcome for the application in question has become essential. Data mining techniques have been used to predict the potential defaulters using a dataset comprising information on loan applications, assisting banks in future decision-making. In [3] The objective of their study was to detect and assess the risks involved in a commercial bank loan. To assess the risk associated with loan approval, they also used data mining tools. In order to produce insightful data, it involves the collecting, processing, and analyzing of data from multiple sources. The Authors have employed C4.5 classification algorithm to predict the level of risk involved in granting a loan to a specific person. A Decision Tree Classifier called the C4. 5 algorithm can be used in Data Mining to generate a decision based on a specific dataset (univariate or multivariate predictors).

The greatest thing is that it performs effectively for both applicants and banks. The adding of attributes makes the model significantly advantageous. Attributes for the assessment involve "Gender," "Married," "Eligible dependents," "Education," "Self-Employed," "Applicant

Income," "Loan Amount," "Loan Amount," "Loan Amount Term," "Credit History," "Property Area," as well as "Loan Status." factors should be considered in order to accurately determine the likelihood of loan default. Therefore, by assessing a customer's potential of defaulting on a loan, Machine Learning Techniques approach makes it simple to identify the proper customers to be targeted for loan providing. The model comes to the conclusion that a bank should evaluate all of a customer's characteristics since they are essential in determining whether to grant credit and in detecting loan defaulters. In order to determine a customer's eligibility for a loan based on their data, banking systems require human processes. Even while manual approaches were frequently effective, they weren't enough when there were many loan applications. It would take a while to make a decision at that point. Bank employees manually review the applicant's qualifications before giving loans to applicants who qualify. The process of reviewing each applicant's data is time-consuming. We used machine learning techniques to build automated loan prediction in order to fix the issue. We'll train the system using data that was previously gathered. for the purpose of evaluating and analysing the process by the machine. Once a qualifying application has been found, the system will inform us of the findings. It serves a purpose, There will be a shorter sanctioning process for loans. The entire process will be computerised to eliminate human error, for those that qualify, a loan will be approved instantly. In [8] The authors proposed a detailed research into data mining technologies and the approach taken by such real-time applications for the detection of loan defaulters. It demonstrates how data mining techniques aid numerous industries in detecting fraud, They developed a conceptual model that loan approval authorities may use to evaluate the reliability of a data mining technology that was being used to estimate the risk attached to a loan application. They also offered demonstrations for several data mining techniques, such as support vector machines, decision trees, and neural networks. Additionally, they used a high average percent hit ratio in combination with the 10-fold cross-validation approach to demonstrate that the forecast was reliable. hence raising the probability that their loans would be repaid on schedule. In [7] when analysing large amounts of data, machine learning methodologies are mostly useful for predicting outcomes. Particularly, machine learning supports data prediction, decision-making, and learning from its own experiences. Employing a Naive Bayesian classifier was recommended by the author. The Navie Bayes algorithm is employed to determine whether to approve a loan for a particular customer. The continuous data associated with each class are distributed according to the Gaussian distribution, according to a typical approach when dealing with continuous values. However the classifier employed in this strategy, which is minimally probabilistic, was developed by applying the Bayes theorem. After carefully examining the Naive Bayes model's advantages and constraints, they came to the convincing conclusion that it is capable of producing outcomes that are superior to those of other models.

The goal of the project is to assist users in loan prediction. The majority of a bank's profit or loss derives from the loans it provides, specifically whether or not its borrowers are maintaining their loan repayments. By identifying potential defaulters on loan, the bank can lower its non-performing assets. This highlights how essential it is to conduct this research. Earlier studies from this era suggest that There are many approaches to look into the problem of preventing loan default. Although producing accurate predictions is essential for maximising profits, it is important to understand the structure of the different methodologies and compare them. Using a predictive analytics method, the problem of predicting loan defaulters is evaluated. Creating a machine learning model is the aim of this project that, using the loan and personal data supplied, can forecast loan default. In order to minimise risk and maximise profit, When deciding whether to provide a loan, the consumer and his financial institution are supposed to utilise the model as a reference tool.

II. LITERATURE REVIEW

A. Gahlaut, Tushar, and P. K. Singh [1], The authors investigated the usefulness of data mining methods for forecasting and categorising a customer's credit score as good or bad, Lenders can limit the possibility that they will continue giving loans to borrowers who won't be able to repay them. Predictive modelling methods include Decision Tree, Linear Regression, Support Vector Machine, Neural Network, Adaptive Boosting Model, and Random Forest. The most promising and accurate method for creating a better categorization model is Random Forest. Pidikiti Supriya, Myneedi Pavani, Nagarapu Saisushma, Namburi Vimala Kumari, k Vikash [2], The project is further advanced using feature engineering methodologies and has transitioned to Exploratory Data Analysis (EDA), wherein statistical concepts like the normal distribution, probability density function, etc. are used to analyse the dependent and independent variables. It will be possible to understand the internal dependent and independent variables effectively by evaluating the data of the univariate, bivariate, and multivariate approaches. The logistic regression model is focused on identifying those customers who are eligible for loans, By splitting the probability into two binary outputs, the sigmoid function made it possible. As a result, the Prediction model might be developed. Mrunal Surve, Pooja Thitme, Priya Shinde, Swati Sonawane, and Sandip Pandit. [3], The author's efforts were focused on detecting and analysing the risk connected to a commercial bank loan. To determine the risk involved in lending money, they used data mining tools. It involves gathering, processing, and synthesising data from various sources into insightful knowledge. M. S. Sivasree [4], Algorithms for categorization are frequently used for loan default prediction. Data processing for classification involves creating train and test sets. Utilizing training data, a prediction model is developed and evaluated against a test set. In the initial phase of information gathering, data from datasets of previously approved loans are combined. Exploratory Data Analysis (EDA), which is used in this paper, offers fundamental understanding of any dataset. EDA's primary goal is to identify the key patterns and display them graphically. Bagherpour. [5], The study propose using a clustering approach to increase the accuracy of bank defaulters identified by likelihood. Using the KNN algorithm, which is available in R, the experimental findings were achieved. The organisation builds a prototype model that it may use to decide whether to approve or refuse the loan in the best possible way. Namvar, M. Siami, F. Rabhi, and M. Naderpour [6], The authors made a remark about the issue of class disparity. The problem of the class imbalance is addressed using a variety of strategies. This classification uses a binary system, which causes the output to fall into one of the two possible categories: default or not. Goyal and R. Kaur [7], In order to categorise loan defaulters and swiftly provide the findings, the author suggested using a Naive Bayesian classifier. Calculating the prior and posterior probabilities, it makes the assumption that each input variable is independent. When the inputs have a high level of dimensionality, the naive Bayes classifier is highly suitable. The Naive Bayes classification method is one of the most complex despite its simplicity. The credit-risk management domain is well suited for it. Yu Jin and Yudan Zhu [8], They compared various data mining methods, such as support vector machines, decision trees, and neural networks in order to use data mining to estimate the risk related to loan application proceedings. Additionally, they demonstrated the forecast's accuracy using a high average hit ratio and the 10-fold cross-validation method. To assess the quality, the cumulative lift curve was evaluated. The SVM approach produced the best outcomes. L. Puro, J. E. Teich, H. Wallenius, J. Wallenius [9], The literature has written extensively about the elements that determine the risk of loan default. Puro and Jeffrey Teich's research suggests that borrowers with higher credit ratings are more likely to have their loan applications approved and experience fewer loan defaults. M. C. Tsai, S. P. Lin, C. C. Cheng, Y. P. Lin [10], The prediction of loan defaults can be represented as a standard binary-class classification problem that can be resolved by a number of machine learning approaches, such as neural networks, logistic regression, and discriminant analysis.

III. PROBLEM STATEMENT

Banks and home finance companies are active in lending in every region of the world, providing various kinds of loans, including mortgages, personal loans, business loans, and others. Rural, semi-urban, and urban areas all have these businesses. These enterprises confirm whether a customer is eligible for a loan after they apply for one. The study provides a way to use machine learning technologies to automate this approach. Meanwhile, the customer will submit a loan application online. The information about the applicant's gender, marital status, qualification, dependents, annual income, loan amount, and credit history is collected through this application. The bank will focus on these customers by automating this method by employing machine learning, and The customer categories that qualify for loan amounts will first be determined by the algorithm. The algorithm will first identify those groups of customers who qualify for loan amounts so that the bank can concentrate on these clients. Every financial institution must deal with the extremely common and critical circumstance of loan prediction when carrying out lending operations. Automating the loan approval procedure can lower operational costs simultaneously improving customer service responsiveness. Significant operational cost savings and an increase in customer satisfaction have been reported. To decrease the risk of loan default, the bank must determine which loans from customers to approve and which to reject. The benefits, however, can only be attained if the bank has a reliable model to make this prediction.

IV. METHODOLOGY

System architecture is a conceptual representation of a system's structure, behaviour, and other elements. A system architecture is a representation of a system that is structured in such a way that it is simple to explain about its structures and actions. System architectural planning determines the WebApp's overall hypermedia framework. The architecture design is influenced by the declared navigation philosophy, the stated goals for the WebApp, the information to be displayed, the users, and other factors. The architecture of a web application describes how it is set up to handle user interaction, carry out internal processing activities, facilitate navigation, and present content. The utilisation of the application's design platform is considered when defining WebApp architecture. Prior to the classification phase, our model required detailed data analysis as well as the application of several preprocessing techniques as shown in figure 1. Once the data has been updated, three distinct classification algorithm models will be applied to forecast loan default, and the outcomes of the models will be compared with those of the same model's previous predictions before data preparation.

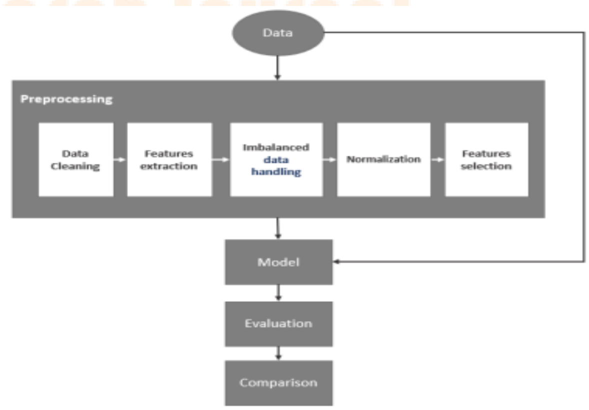

Figure 1: Approach

Supervised Learning

The bulk of real-world applications of machine learning make use of supervised learning. With the use of supervised learning, you may identify the function that transforms your input (x) variables into your output (Y) variables. Y = f (X)

In order to anticipate output variables (Y) given fresh input data, it is important to estimate the mapping function as correctly as feasible (x). The reason it is called supervised learning is because an algorithm that learns from the training dataset is comparable to a teacher who directs the learning process. As a result of our knowledge of the correct responses, The algorithm iteratively predicts using the training data and receives feedback from the teacher. Once the algorithm delivers satisfactory results, learning is complete. Problems with regression and classification fall under the heading of supervised learning difficulties. A classification challenge arises when a category is the output variable, such as "loan approved" or "loan not approved" or "loan defaulter".

Modules: Data Pre-processing, Train the Model & Model Analysis

Data Preprocessing

Data preprocessing is an important step in the machine learning process since the quality of the data and the knowledge that can be drawn from it have a direct impact on how well our model can learn. This makes preprocessing the data before submitting it to the model very essential. Unclean data can be transformed into clean data sets via a procedure called data preparation. In other words, if data is collected from different sources, it is done so in a way that prevents analysis. Therefore, a few procedures are used to convert the data into a minimal, clean data collection. Prior to performing iterative analysis, this procedure is employed. Data preparation is the term for the procedure as shown in figure 1, It contains: Data Cleaning, Feature Extraction, Imbalanced data handling, Normalization and Features selection.

Correcting or deleting inaccurate, damaged, improperly formatted, duplicate, or insufficient information from a dataset is known as data cleaning. The removal of irrelevant values, data duplication, correction of structural errors, handling of missing data, filtering out of data outliers, and data validation are some of the most fundamental techniques for cleaning data in data mining. Since we can't get decent results from unclean data, regardless of how sophisticated our ML algorithm is, data cleansing is essential. The methods and procedures used to clean the data will vary depending on the dataset. Normalization is a popular technique used in the process of data preparation for machine learning. The objective of normalization is to transform the numerical values in the dataset's columns to a common scale without losing any information or distorting the value ranges. Normalization is to transform features to be on a similar scale. This improves the model's performance and training stability. Unevenly distributed observations for the target class are referred to as "imbalanced data" in datasets. Properly identifying the minority classes is essential in specific cases like fraud detection or disease prediction. Therefore, the model should give the minority class equal weight or importance, rather than being biased to only detect the majority class. choosing proper evaluation metric, SMOTE (Synthetic Minority Oversampling Technique), BalancedBaggingClassifier, K-fold cross-validation, and resampling the training set are a few ways to handle imbalanced data. A process called feature extraction converts any data, including text and images, into numerical features that machine learning algorithms can understand. On the other hand, a machine learning technique called feature selection is used on these (numerical) features. Since unstructured data from the real world exists, data preparation is required. Real-world data typically includes:

Train Model

Model training is the next step after model construction. We succeeded in creating models that use our data. Create train and test datasets from the dataset. Using the training dataset, we will finally build and train the model.

Model Analysis

The algorithms Logistic Regression, Naïve Bayes, and KNN, which have a model accuracy and visualisation component, are used in this research.

Model Selection

Model selection is a process of evaluating a finalized machine learning model for a particular training dataset of a Loan customer from a collection of competing machine learning models.

Use-case diagram for Model

Use case for model analysis: Model based analysis as shown in figure 2, is a technique for analysis that employs models to carry out the analysis as well as to record and present the findings. The user in this instance makes use of KNN, Naïve Bayes, and Logistic Regression Algorithms. Pre-processing data use case illustration: When we pre-process data, we make changes to it before we submit it to the algorithm. A technique for turning unclean data into clean data sets is data preparation. Use case illustration for the train model The user split the data first. Then Train the data that has been splited, test the data as the data is divided, and finally, fit the data as illustrated in figure 3.

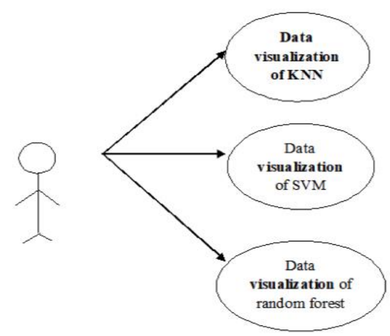

Figure 2: use-case for model analysis

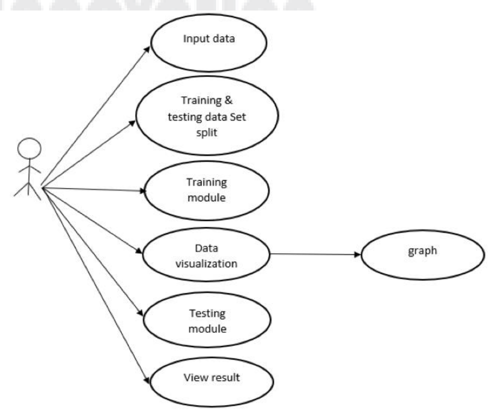

Figure 3: User

V. MODELS

Each of these models has various benefits and challenges, including the statistical noise that predictive error brings into the data, the inadequacies of the sample data and the limits imposed by each model type. The chosen model fulfills the requirements and limitations of the project's stakeholders (Bank and Customers. As a result, the model is selected as the prediction of loan acceptance is a specific type of classification problem. We have used three different models, KNN, Logistic Regression and Naïve Bayes Algorithm for loan default prediction.

1. K-Nearest (KNN) Neighbor

One of the simplest machine learning algorithms, K-Nearest Neighbor, applies the supervised approach. The KNN technique places the new case in the category that most closely resembles the currently available categories on the assumption that the new instance and the data are related to the cases that are already available. The KNN method categorises fresh input based on similarity while saving all previously collected data. The KNN technique can then be used to efficiently classify newly generated data into the appropriate category. The classification problems are where the KNN technique is most frequently used, despite the fact that it may be used to solve regression and classification problems. Due to the non-parametric nature of the KNN technique, no presumptions are made regarding the underlying information. It is commonly referred as lazy learner algorithm since it only intervenes on the training dataset when classification is necessary and preserves the training dataset rather than learning from it immediately. When new data is received, the KNN algorithm categorises it into a category that closely resembles the new data and then simply stores the knowledge learned during the training phase.

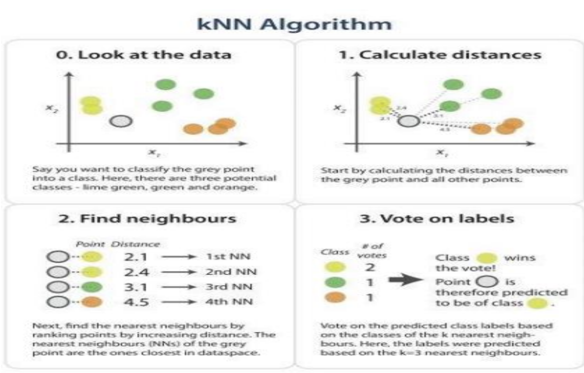

Figure 4: KNN Algorithm

2. Logistic Regression:

A classification technique called logistic regression is used to interpret the data and explain the correlation between a dependent binary variable and one or more independent variables. A method for classifying the data is logistic regression. It is used to predict a binary outcome by utilising a number of independent variables. (1/0, Yes/No, True/False). Machine learning algorithms can learn to recognise patterns that are indicative of loan default by analysing data from prior loan applications. The loan amount and terms for new loan applicants can then be automatically determined using these patterns.
The logistic regression method is one supervised learning strategy. It is used to calculate or predict the probability of a binary (yes/no) occurrence. Using machine learning to assess a person's likelihood of loan eligibility is one example of logistic regression. The topic of foreseeing loan defaulters is studied using the Logistic regression model, a crucial technique in predictive analytics. Data collection is done for research and forecasting. Different performance metrics have been derived using logistic regression models that have been run. By assigning observations to a certain set of classes, the statistical method of logistic regression achieves categorical categorization. It can be used to categorise data points in binary form. An output can be classified categorically into any of the two classes (1 or 0). The logistic sigmoid function is used to modify the output of the logistic regression to return a probability value.

Figure 5: Logistic Regression model

3. Naive Bayes algorithm

To handle classification problems, the Bayes theorem-based supervised learning method known as the Naive Bayes algorithm is used. Text classification using a large training set is its main use. The Naive Bayes algorithm is among the most sophisticated methods for categorising data. It is ideal for the field of credit risk managers. A group of classification methods referred to as Naive Bayes classifiers are constructed using the Bayes Theorem. They are utilised for tasks like prediction and anomaly detection in many different domains. A graph with nodes and directed linkages represents the variables, where each node stands in for a distinct variable and the arcs represent the connections between them. Even though there is no information available in fraud detection about the Naive Bayes, Using the same reasoning, it is possible to determine the set of factors that lead to frauds.
To categorise loan defaulters, a Naive Bayesian classifier can be used, and the results can be obtained rapidly. Calculating the prior and posterior probabilities, it makes the assumption that each input variable is independent. The Naive Bayes algorithm is employed to determine whether or not to approve a loan for a specific customer. This algorithm can be used to make predictions in real time because it is efficient and quick. This method is frequently employed for multi-class predictions. This method makes it simple to ascertain the probabilities of many target classes.

The supervised machine learning method known as the Naive Bayes algorithm is based on the Bayes theorem. If previous information is available, the Bayes theorem is used to calculate the likelihood of a hypothesis. Conditional probabilities are a factor. The accuracy, avoidance of false alarms, and concentration on fraud in bank credit default were not sufficiently addressed by the current fraud prediction algorithms employed in bank credit administration. In an effort to identify credits that could go into default and anticipate potential frauds in the transaction based on data from previous transactions, this work employed the Nave Bayes approach to predict fraud in credit default.

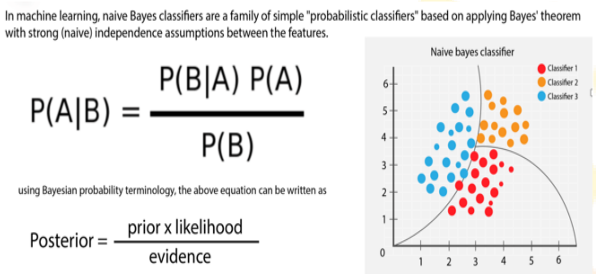

Figure 6: Naïve bayes classifier

VI. RESULT ANALYSIS

The outcome of this whole project with output is described below using three algorithms as mentioned

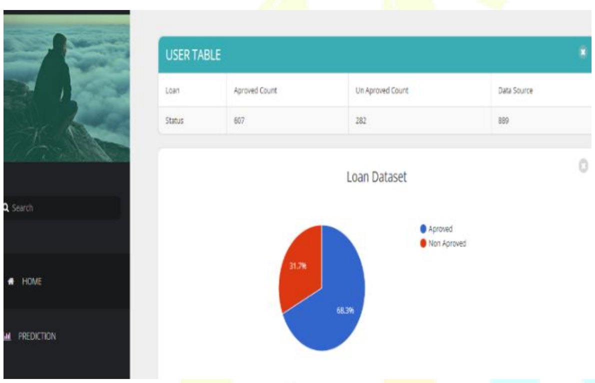

Figure 7: Home – User Table

The figure 7 above demonstrates the Home page, includes a variety of data sources, and the status of the applicant's loan (loan approved or rejected) counts.

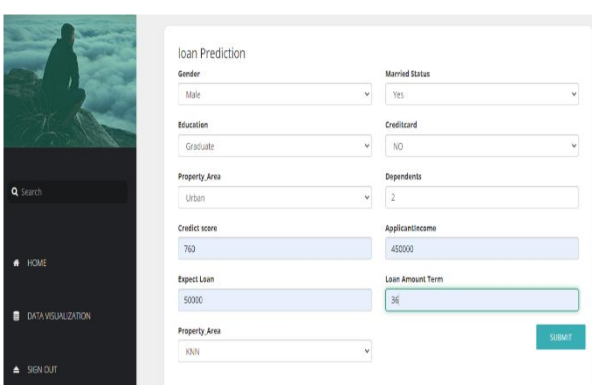

Figure 8: Data Visualization

The above figure 8 demonstrates the Data Visualization, if the applicant needs to apply, they can fill out all the necessary bank information and check their eligibility for a loan. These businesses determine a customer's loan eligibility after they submit an application. The research presents a mechanism for automating this method using machine learning technology. The customer will then complete a loan application online. The information on this form includes the applicant's details like gender, marital status, qualification, number of dependents, annual income, loan amount, and credit history. By automating this process with machine learning, the bank will concentrate on these customers. The algorithm will first identify the customer categories that are qualified for loan amounts.

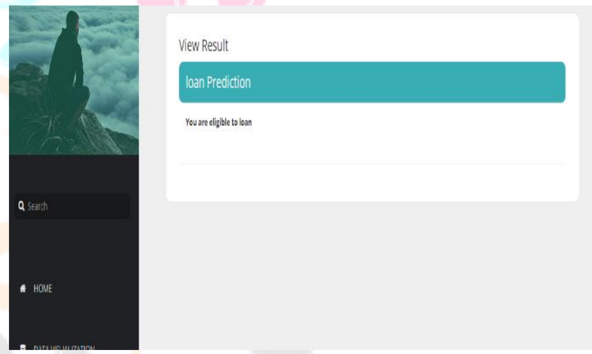

Figure 9: Result

The above figure 9 demonstrates the Output, Predicting the Eligibility Status of a Loan applicant. The project's goal is to build a machine learning model, that can forecast loan default using the loan and personal data provided. The consumer and his financial institution should use the model as a guide when considering whether to grant a loan in order to reduce risk and maximise profit. Automating the loan approval procedure can improve customer satisfaction while reducing operational costs. Also an increase in customer satisfaction and significant operational cost savings have been observed.

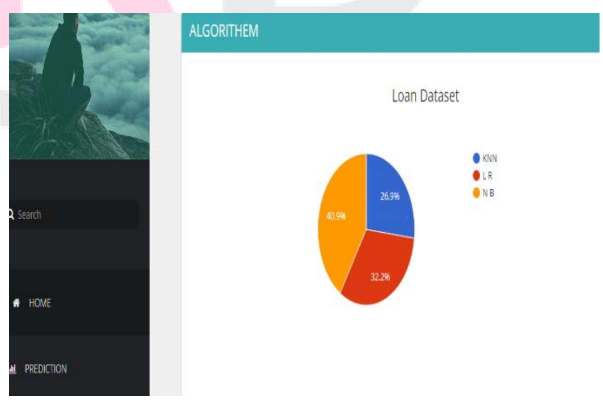

Figure 10: Accuracy

The figure 10 above demonstrates comparison between the three algorithms used to forecast loan status. Model evaluation is a method for evaluating a model's performance in consideration of certain limitations. It is important to consider that the model cannot be overfit or dropped upon when being evaluated. There are numerous techniques available to assess the model's performance. Calculations are performed to evaluate the accuracy of each algorithm.

CONCLUSION

In order to decrease human interference and boost productivity, the rapidly expanding IT sector of today needs to develop new technology and upgrade existing technology. Anyone looking to apply for a loan or the banking system will use this model. It will be very beneficial for managing banks. It is abundantly obvious from the data analysis that it lessens all fraud committed at the time of loan acceptance. Everyone values their time highly, Consequently, by doing this, both the bank's wait time and the applicant's wait time will be decreased. Although it seems that it won't work in some uncommon circumstances where a single parameter is sufficient for the choosing, it is quite reliable and effective in some circumstances. This prediction module can be developed further and added to in the future. The software could be changed in the future to allow it to accept both new testing data and training data and predict as needed. Currently, the system is trained using historical training data.

FUTURE ENHANCEMENT

For future study, the model should be further enhanced by developing an even better dataset, possibly by considering only specific and significant features or categories and experimenting with statistical requirements in order to determine the optimal probability level. Perhaps a hybrid model that first builds an SVM model using training data can be tested, according to the study. Only the data from the first stage that was correctly predicted would be used as the dataset for the neural network model that would thereafter be developed. On this basis, accuracy might be improved while the number of false Positives and true negatives is further reduced.

REFERENCES

1. A. Gahlaut, Tushar, and P. K. Singh, "Prediction analysis of risky credit using Data mining classification models," 2017 8th International Conference on Computing, Communication and Networking Technologies (ICCCNT), 2017.
2. Pidikiti Supriya, Myneedi Pavani, Nagarapu Saisushma, Namburi Vimala Kumari, k Vikash,"Loan Prediction by using Machine Learning Models", International Journal of Engineering and Techniques.Volume 5 Issue 2, Mar-Apr 2019.
3. Mrunal Surve, Pooja Thitme, Priya Shinde, Swati Sonawane, and Sandip Pandit. "Data mining techniques to analyze risk giving loan(bank)" Internation Journal Of Advance Research And Innovative Ideas In Education Volume 2 Issue 1 2016 Page 485-490.
4. M. S. Sivasree, "Loan Credibility Prediction System Based on Decision Tree Algorithm," Int. J. Eng. Res. Technol., 2015.
5. Bagherpour, "Predicting mortgage loan default with machine learning methods," Univ. California / Riverside, 2017.
6. Namvar, M. Siami, F. Rabhi, and M. Naderpour, "Credit risk prediction in an imbalanced social lending environment," arXiv Prepr.arXiv1805.00801, 2018.
7. Goyal and R. Kaur, "Loan Prediction Using Ensemble Technique.," Int. J. Adv. Res. Comput. Commun. Eng., vol. 5, no. 3, pp. 523–526, 2016.
8. Yu Jin and Yudan Zhu, " A data-driven approach to predict default risk of loan for online Peer-to-Peer(P2P) lending," School of Information, Zhejiang University of Finance and Economics, 310018 Hangzhou, China.
9. L. Puro, J. E. Teich, H. Wallenius, J. Wallenius, "Borrower Decision Aid for people-to-people lending," Decision Support Systems, Vol. 49, no.1, pp. 52-60, April. 2010, 10.1016/j.dss.2009.12.009.
10. M. C. Tsai, S. P. Lin, C. C. Cheng, Y. P. Lin, "The consumer loan default predicting model – An application of DEA–DA and neural network," Expert Systems with Applications, Vol. 36, No. 9, pp. 11682–11690, 10.1016/j.eswa.2009.03.009.
11. A. Bachmann, A. Becker, D. Buerckner, M. Hilker, F. Kock, M. Lehmann, P. Tiburtius, and B. Funk, ''Online peer-to-peer lending—A literature review,'' J. Internet Banking Commerce, vol. 16, no. 2, pp. 1–19, Aug. 2011.
12. R. Emekter, Y. Tu, B. Jirasakuldech, and M. Lu, ''Evaluating credit risk and loan performance in online peer-to-peer (P2P) lending,'' Appl. Econ., vol. 47, no. 1, pp. 54–70, Jan. 2015.
13. Y. Jin and Y. Zhu, ''A data-driven approach to predict default risk of loan for online peer-to-peer (P2P) lending,'' in Proc. 5th Int. Conf. Commun. Syst. Netw. Technol., Apr. 2015, pp. 609–613.
14. A. Byanjankar, M. Heikkila, and J. Mezei, ''Predicting credit risk in peerto-peer lending: A neural network approach,'' in Proc. IEEE Symp. Ser. Comput. Intell., Dec. 2015, pp. 719–725.
15. Y. Xia, C. Liu, and N. Liu, ''Cost-sensitive boosted tree for loan evaluation in peer-to-peer lending,'' Electron. Commerce Res. Appl., vol. 24, pp. 30–49, Jul. 2017.
16. Nikhil Madane, Siddharth Nanda,"Loan Prediction using Decision tree", Journal of the Gujrat Research History,Volume 21 Issue 14s, December 2019.
17. C. Serrano-Cinca and B. Gutiérrez-Nieto, "The use of profit scoring as an alternative to credit scoring systems in peer-to-peer (P2P) lending," Decis. Support Syst., vol. 89, pp. 113–122, Sep. 2016.