



# BANK LOAN PREDICTION SYSTEM USING MACHINE LEARNING

V Vijay Chandra, S Shaikshavali, T Tarun, U Bharath Teja, P Nanda Kishore

Assistant Professor, Student, Student, Student, Student  
Computer Science and Engineering,  
G Pullaiah College of Engineering and Technology, Kurnool, India

*Abstract* : Credit approval and credit scores are a multi-billion dollar business. Applicants range from students to large companies, from individuals to organizations. It is always a challenge for financial systems to process these applications through scanning. The process of accepting or rejecting an application is time-consuming as there are many variables to consider. Introduce and apply machine learning models using features such as credit history, applicants, credit history, income details and employment information. This project includes background information, sample selection, and more. The design will also be embedded into a web application that will allow users to access data and get real-time predictions about their performance. Applying machine learning models can reduce human bias and slow down processing time. Banks are already considering using AI to determine loan risk and repayment. This project is designed to create a loan agreement using machine learning and put it into a web application through the Flask framework. The goal is to create a powerful model that can identify applicants to predict the likelihood of receiving loan approval.

*IndexTerms* - Machine learning, Bank loan, analytics, AI process, Flask framework.

## I. INTRODUCTION:

- i. Concept
- ii. Data preparation and prioritization
- iii. Data segmentation
- iv. Model
- v. Model deployment

In this section we examine the challenges and opportunities in detail and highlight the need to develop a credit system[2]. Financial reform requires a move away from formal, rules-based models and towards more flexible, data-driven decision-making. Our system aims to not only improve but also transform the process by bringing machine learning to the world of loan approval[3]. We explore the historical context to reveal how technological advances led to this change. Lending is the main business of banks. A significant part of the bank's income comes from direct loans. Although the bank approves the loan after checking and proving the return, it is still not clear whether the selected customer is correct or not. We can predict whether this particular view is safe or not, and the entire approval process is automated through machine reading and writing[4]. Credit forecasts are very useful for retaining both banks and potential customers.

## II. DATA ANALYSIS:

1. Data analysis to estimate the status of consumer loans The main purpose of this report is to analyze the status of consumer loans. The report shows the customer based on certain factors. Classification is done by researching scientific literature [5]. Data analysis is a process of identifying and documenting the core content of training data. Predicting loan approvals using machine learning Machine learning [6] is the result of creating benchmarks based on training models. The model is used to evaluate data to provide accurate results. Here the author uses three algorithms to predict loans.

These are

1. K is closest to
2. Decision Tree
3. Random Forest

The main purpose of this report is to provide instant and accurate results for loan approval to eligible customers. In the financial market, n people will ask for a loan. It is difficult to identify the right customer from data. This method can provide accurate results for n people.

2. Random Forest Model In this report we discuss credit risk and credit analysis. Success in banking largely depends on credit risk analysis. In this report, we use the random forest method to build the model. The reason for using the random forest method is that the random forest method gives better results than K neighbors and decision trees[7].

3. Joint Modeling Research in Loan Payments In this report, the authors used the random forest method to create the model. In this report, the perfect credit model is determined by combining two or more classifications. The combined approach compares two or more models and finds the best model from two or more models to optimize credit, allowing companies to make the right choice when they find a good credit application.

4. Until now, loans were made with pen and paper from various banks. When a large number of customers apply for bank loans, banks take more time to approve their loans.

When the bank approves the loan, it does not guarantee that the selected applicant will be able to repay the loan. Many banks use their own software for loan approval[8]. In the current system, we use data mining algorithms for loan approval; This is an old tool for loan approval. A lot of data is combined to create comprehensive data and different machine learning algorithms are used to produce results. But these technologies have not yet been mined. As a result, large banks are experiencing financial difficulties. We introduced a new credit approval method to solve this problem.

5. According to the author, the prediction process begins with data maintenance and processing, data set test analysis and modeling without cost change, and continues with the model measuring and evaluating the data. Logistic regression models were applied. The highest accuracy achieved for the first data set was 0.811. Compare models based on performance measures such as sensitivity and specificity[9]. When the results were analyzed, the following conclusions were reached. However, other characteristics of consumers that play an important role in credit decisions and predicting delays also need to be evaluated. The company does not appear to take into account some other characteristics, such as gender and marital history. The credit prediction system helps businesses make decisions to allow or deny customers' credit requests. This helps the commercial bank to establish a good distribution system. This means that if the customer can repay the minimum amount, his system can prevent future risks. Implemented and tested for name along with other methods (using Weka tool) that are better than the general data search model. The authors propose to develop a credit model that predicts loan applicants based on the customer's creditworthiness or standard. The proposed model showed a score of 75.08 when classifying loan applicants using R-Package. Lenders may use this interpretation when making credit decisions for loan applications. A comparative study of different iteration levels was also conducted. The replicated function is a base 30 ANN model that provides a higher level of optimization than other scenarios. This model can be used to prevent large losses in banks. Predicting Android apps using machine learning classification models. This model is available in the open source software R. The application works well and meets the needs of every bank. The disadvantage of this model is that it gives different weights to each element, but in practice it is possible to determine credit based on a single strength, which cannot be provided with this system. This product can be easily connected to many other systems. There are bad computer situations and incorrect content and worst features are fixed by clarifying using the machine and soon the name of the software will be safer, smarter, more reliable and more. Risk assessment and prediction is an important task for the banking industry in determining the suitability of good and lazy loan applicants.

Risk assessment is carried out in primary and secondary education to increase the accuracy of risk assessment. Use data mining to extract customer data and select relevant features. Track policy forecast for each loan type based on predefined criteria. Both the applicant's approval and rejection are considered "qualified" and are considered "invalid." Corresponding experimental results show that the proposed method has better prediction accuracy and takes less time than existing methods. The main purpose of this design is to predict which customers will get the loan back because lenders need to see the problem of borrowers not being able to repay the threat. A study of three models shows that logistic regression with ratings relative to other models, random forests and decision trees[10]. People with bad credit probably can't get this because they may choose not to pay. In many cases, applicants with high net worth can qualify for forgiveness that will allow them to repay the loan. Some views on sex and marriage seem beyond the point.

### III.IMPORTANT PROCESS:

Bank employees verify applicants' information and provide loans to eligible applicants. It takes a lot of time to review each applicant's details. Artificial neural network model for predicting bank credit risk. Feedforward backpropagation neural networks are used to predict loan defaults. A method of combining two or more distributions to create a composite model for better prediction. They use bagging and boosting techniques and then use random forest techniques. The action of the classifier is to improve the performance of the data, providing better performance[11]. In this work, the authors describe various integration methods for binary and multi-class classification. The new integration method described by the authors is COB, which provides a good classification function, but also suffers from noise and poor classification.

- The traditional loan approval process is challenged by strict regulations and manual evaluation procedures.

The disadvantages of these systems are:

- Risk of error
- Long decision-making time
- Problems adapting to an adverse financial environment
- The limitations of relying on historical rules and subjective evaluation are obvious.
- These challenges lay the groundwork for our new machine learning-driven solutions
- Analyzing the details of each applicant takes a lot of time and effort. Human error may occur in verifying all details. Has the ability to distribute loans to unqualified applicants.

### IV.APPROVAL PROCESS:

Credit approval is a software used for loan approval in commercial banks. We use machine learning algorithms in this application. Machine learning is the process of creating symmetric patterns based on existing data; The standard used to evaluate new data[12]. The system contains training data and testing data. Training data is used to build the model. The model is applied to the test data set to obtain the desired results. We use an integrated system to create models. The Random Forest algorithm uses this combination and creates models based on existing training data.

The application process is a beacon of innovation in the fintech industry.

- Leverages advanced machine learning algorithms to radically increase the accuracy and efficiency of the loan approval process.
- Core principles include:
  - Predictive models
  - Advanced data analytics
  - Pattern recognition
- The system is designed to be responsive and intelligent
- For dynamic business, we focus on changing demand and changing customer profile.

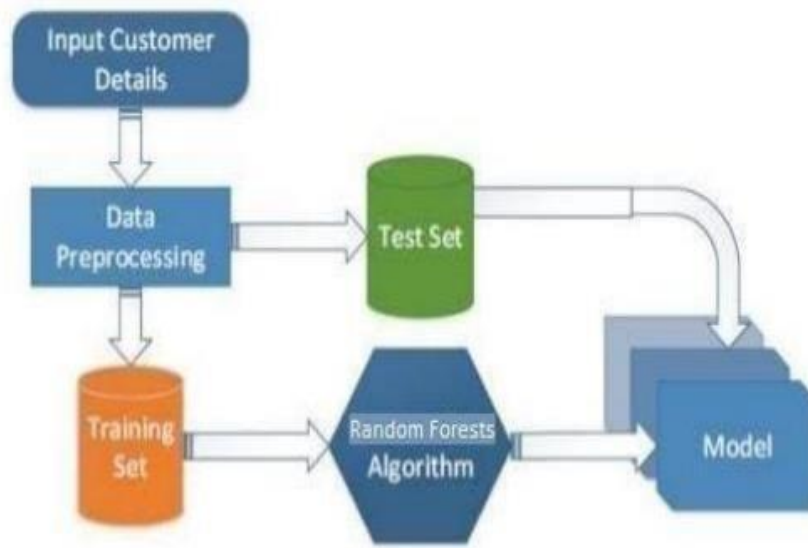


Fig.1. System Architecture

V. MODULES :

1. DATA COLLECTION:

Collections in software are collections of data consisting of lists, sets, tuples, and dictionaries. They have different features depending on demand and purpose. A list is declared in square brackets, can be modified, stores return values, and can be accessed using indexes[13]. The archive takes center stage, explaining the careful processes involved in obtaining, verifying and organizing historical credit data. More in-depth investigation of data, quality checks, and strategies for dealing with missing or inaccurate data. The main source of strong data for the next stage is approximate.

It is estimated during the data collection, training and testing process to predict customers' credit defaults. Usually the 8020 ratio is used to separate the training process and the testing process. The data model created using decision tree is used for the training process and depends on the accuracy of the test set to perform the prediction test.

Attribute Name	Category
Loan_ID	Qualitative
Gender	Categorical
Married	Categorical
Dependents	Qualitative
Education	Categorical
Self_Employed	Categorical
ApplicantIncome	Qualitative
CoapplicantIncome	Qualitative
LoanAmount	Qualitative
Loan_Amount_Term	Qualitative
Credit_History	Qualitative
Property_Area	Categorical

Table1. Dataset attributes

1.1. Preprocessing:

1.2. Archived data may contain missing values that may cause conflicts. In order to get better results, the data must first be processed to improve the results of the algorithm. We must remove outliers and variables must be replaced. We use the chart function to solve these problems. Preliminary data for credit prediction using machine learning. Data preprocessing includes tag coding, missing values

handling, field selection, standardization, etc. Contains. We need to do all these steps on our dataset, even though it is a relatively clean and structured dataset.

### 1.3. Feature Engineering:

1.4. Four feature engineering that is the cornerstone of machine learning in detail. Beyond selection, this chapter discusses the transformation process used for raw materials to provide relevant and useful properties. Techniques such as dimensionality reduction,

1.5. parameterization and retrieval are examined in detail and it is shown how these steps increase the predictive power of the model.

### 1.6. Model training:

1.7. The model training module is published and the selection of machine learning algorithms is discussed. A detailed description of the training process, including hyperparameter tuning and cross-validation techniques, provides insight into the search for the best-performing model. Real-world examples and case studies show that the flexibility of the system varies[14]. Now we need to train the model on the training data and perform fortune telling on the test data. We can divide the training data into two parts: training and validation. We can train the model in this training section and use it to make predictions for the witness section[15]. In this way we can verify our prophecy because we have the correct prophecy of the proverbs (we have no test data)

### 1.8. Evaluation of the model:

1.9. Critique of the model leads to Elaboration evaluation, which provides an in-depth look at the various metrics used to evaluate performance. Accuracy, precision, and recall, as well as model interpretation and identification of bias, are also investigated[16]. The purpose of this section is to provide users with information on how to measure and adjust their models. Choose appropriate machine learning algorithms for credit prediction, such as logistic regression, decision trees, random forests, support vector machines or neural networks. Consider using a combination of techniques to combine multiple models to improve the accuracy of predictions.

### 1.10. User Interface:

1.11. The user interface module is designed to act as a bridge between the algorithms and the end user. Analysis of user design ensures that the connection not only provides predictions, but also provides insights into the features that influence the predictions. Visual communication and customer-friendly features are at the center of this search.

## 2. ALGORITHM APPLICATION:

2.1. The algorithm process provides in-depth study of the chosen model, providing continuous and practical explanations of the principles, advantages and applications of each algorithm.

### 2.2. Decision tree

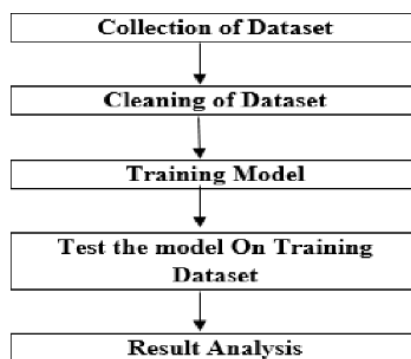
2.3. Many important aspects of machine learning - decision tree. Recursive methods of building trees, deciding on splits, and dealing with overfitting are discussed[17].

2.4. Real-world comparisons and scenarios demonstrate the flawed nature of the algorithm.

### 2.5. Random Forest

2.6. Random forest algorithm follows the rules of the decision tree. The difference between them is that the decision tree algorithm decides only one key to give the output, while the random forest algorithm compares many decision trees and gives satisfactory results for most of the decision trees. Random Forest is a learning method that attracts a lot of interest[18]. This section explains the

concept of packaging and aggregating multiple decision trees. The benefits of random forests in reducing optimization and increasing the accuracy of predictions are explained with examples.

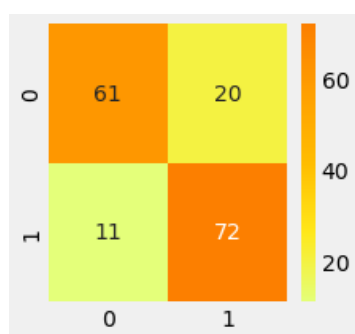


2.7. Fig.2. Random Forest algorithm

2.8. When we load the training data, the system creates the model using the random forest algorithm and then when we load the test data using the model, the system provides the required output. The output consists of two groups labeled yes/no. Yes means the customer has fulfilled the loan agreement, No means the customer has not met the credit agreement. In this way the system provides the required output.

## 2.9. Gradient Boosting

2.10. Gradient boosting actually “boosts” many weak predictive models into strong models in the form of a collection of weak models. Here a poor predictive model can be any model that is slightly better than the prediction. We need to find a good way to “connect” weak models to create a strong model. The mystery of gradient boosting algorithms, including XGBoost, has been solved. One way to combine poor learners into good models is to conduct research on the effectiveness of algorithms in handling relationships in data.



2.11.

## 2.12. Fig.3. Gradient Boosting algorithm

2.13. Many aspects regarding flexibility and efficiency. Gradient boosting is an integrated machine learning technique that creates strong predictive models by combining multiple weak models. It uses gradient descent as a method to reduce errors during training.

## VI.PURPOSE OF THE MODEL:

### 1.1. Credit approval prediction:

2. The main function of the model is to predict credit approval results. Use historical data and machine learning algorithms to make the right decisions. It includes many things like Cibil loan includes income, employment, loan maturity, loan amount, assets and loan amount. This data is often used in machine learning and data analytics to create models and algorithms that predict the likelihood of receiving loan approval based on certain characteristics.

### 2.1. Risk Factor Analysis:

3. Provides a better understanding of the various risks associated with a loan application. Identify and describe risks to improve decision making. The purpose of RFA is to identify and understand the underlying principles that ultimately drive the behavior of the plan's high-level strategy, cost, and performance measures.

### 3.1. Behavioral Analysis:

4. Look at the patterns and behaviors that customers exhibit regarding their loan applications.

5. Enable organizations to gain in-depth insight into customer preferences and trends.

### 5.1. Business analysis:

6. Discuss the ability to analyze business and strategic needs. Enable financial institutions to monitor market changes and adjust their strategies accordingly.

### 6.1. Guided Decision Making:

7. Compare real-life situations where organizations can use insights to make better decisions.

8. Show the value of decisions made in line with business objectives

## VII.CONCLUSION:

Thanks to the policy analysis method, this standard can be effectively used to check customer compliance. For credit approval. The software works perfectly and meets all banking regulations. The system can be easily installed on any operating system. The system will become even more usable in the coming days as the technology is moved online. The system is safe and reliable. Since we use the random forest algorithm, the system provides very good returns. It does not matter if there are many customers applying for credit. The system accepts N files. customers. In the future we can add more algorithms to this system to get more results. In short, the "Bank Loan Prediction System Using Machine Learning" is a revolution in the financial sector. The system leverages advanced machine learning algorithms to provide unparalleled accuracy and efficiency in the loan approval process. Its adaptability to dynamic business conditions and insightful data analysis make it a strategic asset for financial institutions. During the work, we encouraged the development of the work by solving problems with problem-solving strategies. The impact of the exchange system lies in its ability to provide advice, accelerate decision-making, and address the complexities of the changing financial environment. As stakeholders evaluate this technological revolution, the system has become a beacon for data-driven decision-making in finance. It heralds a future where predictive models and model validation will redefine the way financial institutions operate. This decision highlights significant achievements and invites people to embrace a future where smart machines improve the financial services ecosystem.

## VIII.REFERENCES:

- [1] M. Sheikh, A. Goel, T. Kumar, "An Approach for Prediction of Loan Approval using Machine Learning Algorithm," International Conference on Electronics and Sustainable Communication Systems (ICESC), (2020).
- [2] S. M S, R. Sunny T, "Loan Credibility Prediction System Based on Decision Tree Algorithm," International Journal of Engineering Research & Technology (IJERT) Vol. 4 Issue 09, (2015).
- [3] A. Kumar, I. Garg and S. Kaur, "Loan Approval Prediction based on Machine Learning Approach," IOSR Journal of Computer Engineering, (2016).
- [4] Dr K. Kavitha, "Clustering Loan Applicants based on Risk Percentage using K-Means Clustering Techniques," IJARCSSE - Volume 6, Issue 2, (2016).
- [5] P. Dutta, "A STUDY ON MACHINE LEARNING ALGORITHM FOR ENHANCEMENT OF LOAN PREDICTION", International Research Journal of Modernization in Engineering Technology and Science, (2021).
- [6] G. Arutjothi, Dr C. Senthamarai, "Prediction of Loan Status in Commercial Bank using Machine Learning Classifier," Proceedings of the International Conference on Intelligent Sustainable Systems, (2017).

- [7] P. Supriya, M. Pavani, N. Saisushma, N. Kumari and K. Vikas, "Loan Prediction by using Machine Learning Models," International Journal of Engineering and Techniques, (2019).
- [8] R. Salvi, R. Ghule, T. Sanadi, M. Bhajibhakare, "HOME LOAN DATA ANALYSIS AND VISUALIZATION," International Journal of Creative Research Thoughts (IJCRT), (2021).
- [9] B. Srinivasan, N. Gnanasambandam, S. Zhao, R. Minhas, "Domain-specific adaptation of a partial least squares regression model for loan defaults prediction," 11th IEEE International Conference on Data Mining Workshops, (2011).
- [10] M. V. Reddy, Dr B. Kavitha, "Neural Networks for Prediction of Loan Default Using Attribute Relevance Analysis," International Conference on Signal Acquisition and Processing, (2010).
- [11] G. Chornous, I. Nikolskyi, "Business-Oriented Feature Selection for Hybrid Classification Model of Credit Scoring," IEEE Second International Conference on Data Stream Mining & Processing August (2018).
- [12] K. Hanumantha Rao., G. Srinivas., A. Damodhar., M. Vikas Krishna., Implementation of Anomaly Detection Technique Using Machine Learning Algorithms: International Journal of Computer Science and Telecommunications, Vol. 2, Issue 3, 2011.
- [13] S.S. Keerthi., E.G. Gilbert., Convergence of a generalize SMO algorithm for SVM classifier design, Machine Learning, Springer, Vol. 4, Issue 1, pp. 351-360, 2002.
- [14] Ekta Gandotra., Divya Bansal., Sanjeev Sofat., Malware Analysis and Classification: A Survey, Journal of Information Security, Vol. 05, Issue 02, pp. 56-64, 2014.
- [15] Rattle data mining tool, <http://rattle.togaware.com/rattle-download.html>.
- [16] Aafer Y., Du W., Yin H., Droid APIMiner: Mining API-Level Features for Robust Malware Detection in Android, Security and privacy in Communication Networks, Springer, pp 86-103, 2013.
- [17] J. R. Quinlan., Induction of Decision Tree, Machine Learning, Vol. 1, No. 1. pp. 81-106.
- [18] Andy Liaw., Matthew Wiener., Classification and Regression by random Forest, Vol. 2, Issue 3, pp. 9-22, 2002.

When the bank approves the loan, it does not guarantee that the selected applicant will be able to repay the loan. Many banks use their own software for loan approval[8]. In the current system, we use data mining algorithms for loan approval; This is an old tool for loan approval. A lot of data is combined to create comprehensive data and different machine learning algorithms are used to produce results. But these technologies have not yet been mined. As a result, large banks are experiencing financial difficulties. We introduced a new credit approval method to solve this problem.