JETIR.ORG ISSN: 2349-5162 | ESTD Year : 2014 | Monthly Issue
JOURNAL OF EMERGING TECHNOLOGIES AND INNOVATIVE RESEARCH (JETIR)
An International Scholarly Open Access, Peer-reviewed, Refereed Journal

# FORECASTING LOAN SUITABILITY WITH MACHINE LEARNING

**Halima Sania Samreen**
E&CE, PESITM Shivamogga, India.

**Prajwal S**
E&CE, PESITM Shivamogga, India.

**Mr. Anil Kumar C**
Asst Professor (Guide)
E&CE, PESITM Shivamogga, India.

**Thanushree GP**
E&CE, PESITM Shivamogga, India.

**Yeshwanth B**
E&CE, PESITM Shivamogga, India.

## ABSTRACT:

A very complicated and fundamental mortgage endorsement system that banks depend on to a great extent for their earnings and profitability is considered the backbone of the banking industry. An advanced statistical approach was used introduced into the process through utilization historical data about past applicants to this system and how they paid off their debt and more efficiency. Three different statistical methods discussed in this paper provide insight into the ways predicting the approval of mortgage applications: Decision Trees and Naïve Bayes. To test these models, we use an openly accessible dataset that contains income, credit history, loan sizes used by potential borrowers among other attributes, with performance being evaluated in forms of accuracy, specificity, prediction accuracy for minority groups and overall efficacy. In relation to our study findings, the boost-based decision tree method outperformed all other approaches generating excellent results on all test data metrics. Finally, we have provided an importance table which identifies some specific features whose presence or absence impact loan approval rate predictions most strongly. This leads us to conclude that boosting based decision tree is a robust predictor of loan approval thereby making it best recommended tools in banking industry position as well as position as a powerful tool for predicting loan approval rates.

**Keywords: Loan approval, Decision tree, Naive bayes, Data sets, Training, Testing, Prediction.**

## INTRODUCTION:

"Examining Credit Approval Systems in the Banking Sector: A Machine Learning Approach. The banking sector is a crucial part of today's economy, facilitating the movement of economic activities and boosting growth. In this sense, credit approval has much influence regarding whether an individual or business can access funds for their financial undertakings. But mortgage approval through conventional means may be protracted, convoluted and error-prone because of human lapse." Machine learning (ML) has become an essential tool used to address several problems faced by mortgage approval systems. This review will evaluate the performance and techniques used in many mortgages' ML-based models that are largely open source packages and evaluation metrics provided with them. We start our investigation by considering the need for green credit analysis methods within this field. To attain this reason, we focus on how ML can ease automatic processing of mortgage certificates using its ability to handle high volumes of data, discover patterns and thus enable wise choices to be made.

In this article, we delve into a multitude of ML algorithms utilized in the mortgage approval process. Specifically, the efficacy of decision trees and naive arrays and highlight the strengths and challenges of each. As we explore their usefulness in various types of datasets and applications, we also emphasize the role of exploratory statistical analysis (EDA) in producing well-informed and skilled mortgage data for ML modeling.

Then, our focus shifts to assessing loan approval systems driven by machine learning. Among the most important evaluation measures we can talk about are accuracy, precision and recall that act as a fundamental evaluation framework for determining the effectiveness of different machine learning models. Additionally, we touch on the critical issue of fairness and bias in ML techniques to ensure they won't exhibit any type of prejudice against particular applicant organizations.

In conclusion, our paper looks into the bright future of ML-based mortgage approval systems. Specifically, shows how these systems can be made better through further advancements in ML techniques such as deep learning and natural language processing. For this reason, while exploring these thrilling possibilities, we consider the challenges involved as well as ethical examination that accompanies ML being used in mortgage approvals. This includes transparency, accountability and responsible AI development which should be at the forefront of every implementation process.

Finally, a comprehensive analysis of how machine-learning-powered mortgage-approval systems work – including their purpose, methods used for realization or assessment targets and what may happen next will be given in this assessment document on loans. Through understanding how ML can achieve or cannot perform results measurement for gauging its efficacy becomes apparent together with addressing equity along with discrimination hence making it easier to apply artificial intelligence in banking loan approval workflows, enhance decision-making, and elevate client contentment.

## I. LITERATURE REVIEW:

[1] Vishal Singh delves into the application of ML techniques to forecast loan approvals within the banking industry. The authors underscore the utmost significance of accurately predicting an applicant's ability to repay their loan. Moreover, they shed light on the critical role of past data in determining loan approval and how it can affect the bank's financial gains or losses. The study also constructs a machine learning model using a classifier and a support vector machine classifier, both operating on historical data.

Moreover, the authors go through into the numerous benefits of implementing the proposed system. Notably, the system has the potential to remarkably reduce bank losses and expand the number of loans granted. Additionally, the authors stress the significance of various factors – such as loan duration, amount, age, and income – in accurately predicting loan approval. The paper culminates with emphasis on the system's considerable accuracy and the vital role of elements like zip code and credit history in classifying loan applicants.

The article draws upon a references, including "Machine Learning Approach for Cooperative Banks Loan Approval" and "Loan Prediction Using Ensemble Technique," highlighting the authors' dependence on established literature and research in the domains of loan prediction and machine learning.

In this paper, we can gain valuable knowledge on the utilization of ML algorithms in predicting loans, the critical role of historical data, and the potential advantages for banks in enhancing their loan approval procedures.

[2] Gupta and Pant show in their article, "Bank Loan Prediction System using ML," which measures the accuracy of daily projections. But they contend that, in spite of the claims made by others that predictions are nothing more than informed guesses or gut instincts, predictions are nevertheless crucial for predicting future occurrences, whether or not they will occur soon or even years from now. One of the key components of advanced analytics is predictive analytics, which includes a variety of techniques like data mining, statistical analysis, modeling, machine learning, and artificial intelligence.

Adyan Nur Alfiyatin, Hilman Taufiq and Through their work on Regression Analysis for Accurate Forecasting of Real Estate Prices by Particle Swarm Optimization (PSO), associates have made important advances to this subject. The application of supervised learning was also Mohamed Mohadab for predicting rankings of research papers while Kumar Arun et al., created a model that uses support vector machines (SVM) to facilitate the acceptance or rejection of bank loans. and neural networks algorithm from machine learning point of view.

Through an exhaustive review of these noteworthy contributions, Our review of the literature provides insightful information that will guide our future research. Building upon this robust foundation, our objective is to develop a robust and reliable model for accurately forecasting bank loan approvals.

[3] During their presentation, Dr. C. K. Gomathy and Ms. Charulatha highlighted the loan prediction system's productivity and adaptability. They emphasized its tailored capabilities for banking institutions, highlighting its potential for success. The system's effectiveness and seamless integration with different systems were also noted, further emphasizing its practicality. The researchers also acknowledged the current achievements of the system and identified areas for improvement, particularly in regard to safety, reliability, and adaptability, showcasing their commitment to overall performance enhancement. Thus, the decision to incorporate the prediction module into the automated processing system demonstrates a proactive approach towards improving and expanding the system. This strategic thinking highlights the system's evolution and progression.

# II. PROPOSED SYSTEM

## (a) MACHINE LEARNING ALGORITHMS:

Algorithms for ML can be divided into two categories:

Supervised Learning: This is typically an instance of training an algorithm using a labeled dataset where inputs are paired to correct outputs. The learning method among the most widely used algorithms is supervision, which can be found in decision trees, linear regression, Support Vector Machines, and Neural Networks.

Unsupervised Learning: Unsupervised learning is a different approach which work on unlabeled datasets. In this approach, the system needs to recognize the patterns and interrelations by itself, rather than use correct output data. An instance of representatives of this category are K-Means Clustering, Hierarchical Clustering, or PCA (Principal Component Analysis).

For the sake of this study, There are two different machine learning algorithms: employed to facilitate accurate predictions within the dataset: In order to conduct this study, two different ML algorithms are employed to facilitate accurate predictions within the dataset:

(a) Decision Tree: The Decision Tree is a universally applicable algorithm which is efficient for both regression and classification tasks. Consequently, it is its main advantage to adopt the sequence of events mode. The system does this recursive binary greedy to partition the attribute space. By strategically dividing the attribute space into smaller subsets, the decision tree consistently makes optimal decisions with the purpose of minimizing classification errors. Key properties of decision trees include: By strategically dividing the attribute space into smaller subsets, the decision tree consistently makes optimal decisions with the purpose of minimizing classification errors. Key properties of decision trees include:

- Graphical Representation: Decision trees are visually depicted, offering a simple interpretation that closely mimics human decision-making processes.

- No Assumption on Attribute Distribution: Decision trees make no assumptions about the distribution of attributes or predictors, making them adaptable to both numerical and categorical variables.

This study utilizes a classification tree to accurately forecast the loan status of applicants, taking into account various characteristics. The classification tree relies on a democratic approach, assigning each instance to the most typical seen loan status among similar cases in its particular region. Key elements that may impact these determinations include employment status, credit history, loan amount, and other relevant factors.

**(b) Naive bayes:**

Naive Bayes, a widely used probabilistic classification algorithm based on Bayes' theorem and known for its simplicity and robustness, represents one of the tools in the arsenal of a Data Scientist. It has excellent computational efficiency built into it which assumes independence of features. Therefore, it is good for tasks like text classification including spam detection and sentiment analysis. It is adaptable, for instance, its different types of naive bayes, including Bernoulli for binary data, Multinomial for discrete data, and Gaussian for continuous data which is applicable in a variety of different applications. Although the method is simple and its effectiveness is unquestionable, it generally delivers successful results, especially when the independence assumption is sufficiently valid. Its superior computational efficiency enables it for use across datasets of different sizes. Nevertheless, it faces obstacles including sensitivity to nonrelated characteristics and the limitations of the independence assumption. Thus, researchers have started considering more variations of the technique and mixing it with other methods which lead to the improvement of its operation.

Concerning the default-predicting loan, Naive Bayes operates under the presumption that characteristics are independent of conditions given the class label. Despite its apparently trivial implementation, it is efficient and computationally friendly. Applying the Naive Bayes for prediction of loan defaults would imply the usage of historical data with labeled instances (like defaulted or not defaulted) and relevant features.

## (b) DATASET DESCRIPTIONS AND PRE-PROCESSING:

This paper consists of the data sets from Kaggle.com that have person of different age groups and genders. The dataset has thirteen attributes namely Loan_ID, Gender, Status of Marriage, Dependents,Education, Self-employed, Applicant_Income,Co-applicant_Income, Loan_Amount, Loan_Term, Credit_History, Property_Area, and Loan_Status.as shown below.

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 614 entries, 0 to 613
Data columns (total 13 columns):
 #   Column             Non-Null Count  Dtype
---  ------             --------------  -----
 0   Loan_ID            614 non-null    object
 1   Gender             601 non-null    object
 2   Married            611 non-null    object
 3   Dependents         599 non-null    object
 4   Education          614 non-null    object
 5   Self_Employed      582 non-null    object
 6   ApplicantIncome    614 non-null    int64
 7   CoapplicantIncome  614 non-null    float64
 8   LoanAmount         592 non-null    float64
 9   Loan_Amount_Term   600 non-null    float64
 10  Credit_History     564 non-null    float64
 11  Property_Area      614 non-null    object
 12  Loan_Status        614 non-null    object
dtypes: float64(4), int64(1), object(8)
memory usage: 62.5+ KB
```

The collection is the most critical component to be applied in ML training on the basis of historical loan application records to predict loan approvals. The precision of the system's predictions is significantly in connection with the quality and the dimension of the dataset. Luckily, the dataset readily obtained and accessed through the provided Kaggle link: https://www.kaggle.com/datasets/devzohaib/loan-elfilibilty-prediction.

It is possible to obtain a more precise training and a better performance among the machine learning techniques employed in the modeling of loan eligibility by using this dataset. The availability of a detailed and well-structured database is vital in the process of designing accurate and successful predictive models for loan approval decision making.

In consideration of the shaping of loan approval prediction, the optimal plan is the one that will involve the perfect process of data pre-processing in ensuring the creation of good models. This involves several important steps: - Humanize the given sentence.

Data cleaning: Incomplete values are ascribed using imputation or simply marked as missing, and outliers are detected and processed to avoid skewness in results. When this process is done and the resulting dataset now has 0 NA (missing) values.

```
Loan_ID              0
Gender               0
Married              0
Dependents           0
Education            0
Self_Employed        0
ApplicantIncome      0
CoapplicantIncome    0
LoanAmount           0
Loan_Amount_Term     0
Credit_History       0
Property_Area        0
Loan_Status          0
LoanAmount_log       0
dtype: int64
```

Feature scaling: To make sure all numeric features are scaled the same, data standardization is done to stop the model from being dominated by one characteristic.

Feature coding: Categorical variables are represented as numerical by employing encoding methods such as one-shot coding and label coding and thus in the modeling process are more suitable.

## (c) METHODOLOGY:

Data Collection: Present the historical loan data which include both approved a In recent years, the power of language has emerged as a crucial concern in the world. Language serves is a channel via which individuals can communicate, preserve culture, and express themselves creatively. However, it also carries the risk of marginalizing certain communities, exerting social pressure, and even influencing societal attitudes and values. The relevant feature should be taken into consideration. They will be the number of credit score, income, loan amount, employment history and others.

Data Pre-processing: Depending about the data's values collected, use appropriate (appropriately) cleaning methods to deal with the missing values, transform the non-numeric data into numerical formats as required.

Data Splitting: Separate the pre-processed into two distinct subgroups of the dataset as follows— training set and validation set. The distinctive use of the training set is to train the machine learning models while the test set acts as an independent data set for the model's evaluation.

Building a Decision Tree: Trained a model of decision trees using the training data. The algorithm known as the decision tree structurally partitions based on the associated features, thereby creates a tree structure that is used in order to predictions.

Naive Bayes Model: Use the same data set you utilized to train the Naïve Bayes classifier. Unlike other classification algorithms, Naive Bayes based on probabilities is the best at making decisions.

Evaluate Models: Evaluate the model of decision trees and Naive Bayes model using designed test information to obtain a complete evaluation. Therefore, the execution of models is measured by the assessment that is employed for identifying the best model and is suited for loan approval predictions.

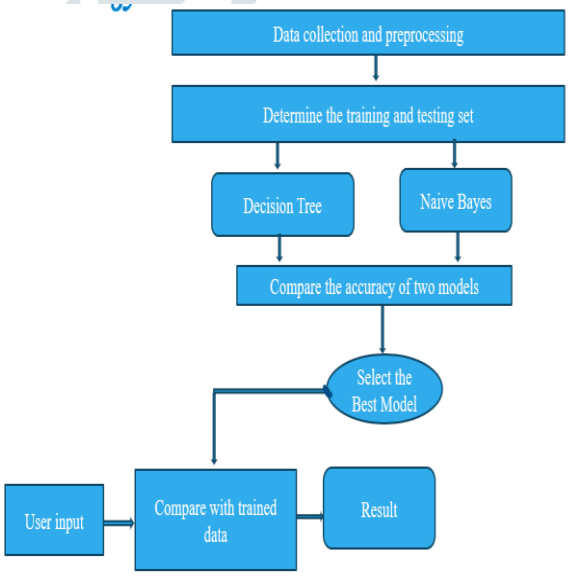


**Figure:** Block Diagram of the loan default Predication

## (d) EVALUATION METRICS:

The Loan Application Prediction Scoring Matrix consists of four key steps to accurately assess the execution of a trained model: The Loan Application Prediction Scoring Matrix consists of four key steps to accurately assess the performance of a trained model:

1. Data Split: Initially, the model is evaluated, the data-set is divided into two distinct subsets - testing set and training set. The training set consists 80% of the information and the testing set 20% of the information. The model's results evaluation can then be confidently done because of the unseen data being considered.

2.Model Training: -The model employed to forecast the loan application is thereafter instructed on the training set competently, employing powerful algorithms for example Naive Bayes and Decision Trees so as to improve the level of thoroughness.

3. Prediction: - After the model is trained, is used in the test set to ascertain the precision of the predictions based on how the model would perform on new data.

4. Confusion Matrix: - Once the Scoring Matrix is completed, a confusion matrix is constructed, using the matrix to offer invaluable insights on the model's performance. True Positives (TP): Approved loans correctly predicted

True Negatives (TN): Rejected loans correctly predicted

False Positives (FP): Rejected loans wrongly assumed to be approved

False Negatives (FN): Approved loans wrongly assumed to be rejected

Example of a Confusion Matrix:

| Predicted Approved | Predicted Rejected | |
|---|---|---|
| Actually Approved | TP | FN |
| Actually Rejected | FP | TN |

Validation: Ensure the model's robustness and generalizability by validating it on additional datasets or conducting cross-validation tests.

## III. EXPERIMENTAL RESULTS:

For our analysis, we use an unbalanced data set to evaluate two different supervised learning models represented in the imbalanced data set. Accuracy will be compared to decide the best approach, that is, the most effective one. The given figure showcases the data that has been employed in instruction these models. The practiced data contains attributes like Loan_ID, Gender, Marital status, Dependents, Education, Self Employed, Applicant Income, Co-Applicant Income, Loan Amount, Loan Term, Credit History, Property Area, and Loan Status.

| Loan_ID | Gender | Married | Dependen | Education | Self_Empl | ApplicantIn | CoapplicaI | LoanAmou | Loan_Amo | Credit_His | Property_A | Loan_Status |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| LP001002 | Male | No | 0 | Graduate | No | 5849 | 0 | | 360 | 1 | Urban | Y |
| LP001003 | Male | Yes | 1 | Graduate | No | 4583 | 1508 | 128 | 360 | 1 | Rural | N |
| LP001005 | Male | Yes | 0 | Graduate | Yes | 3000 | 0 | 66 | 360 | 1 | Urban | Y |
| LP001006 | Male | Yes | 0 | Not Gradu | No | 2583 | 2358 | 120 | 360 | 1 | Urban | Y |
| LP001008 | Male | No | 0 | Graduate | No | 6000 | 0 | 141 | 360 | 1 | Urban | Y |
| LP001011 | Male | Yes | 2 | Graduate | Yes | 5417 | 4196 | 267 | 360 | 1 | Urban | Y |
| LP001013 | Male | Yes | 0 | Not Gradu | No | 2333 | 1516 | 95 | 360 | 1 | Urban | Y |
| LP001014 | Male | Yes | 3+ | Graduate | No | 3036 | 2504 | 158 | 360 | 0 | Semiurban | N |
| LP001018 | Male | Yes | 2 | Graduate | No | 4006 | 1526 | 168 | 360 | 1 | Urban | Y |
| LP001020 | Male | Yes | 1 | Graduate | No | 12841 | 10968 | 349 | 360 | 1 | Semiurban | N |
| LP001024 | Male | Yes | 2 | Graduate | No | 3200 | 700 | 70 | 360 | 1 | Urban | Y |
| LP001027 | Male | Yes | 2 | Graduate | | 2500 | 1840 | 109 | 360 | 1 | Urban | Y |
| LP001028 | Male | Yes | 2 | Graduate | No | 3073 | 8106 | 200 | 360 | 1 | Urban | Y |
| LP001029 | Male | No | 0 | Graduate | No | 1853 | 2840 | 114 | 360 | 1 | Rural | N |
| LP001030 | Male | Yes | 2 | Graduate | No | 1299 | 1086 | 17 | 120 | 1 | Urban | Y |
| LP001032 | Male | No | 0 | Graduate | No | 4950 | 0 | 125 | 360 | 1 | Urban | Y |
| LP001034 | Male | No | 1 | Not Gradu | No | 3596 | 0 | 100 | 240 | | Urban | Y |
| LP001036 | Female | No | 0 | Graduate | No | 3510 | 0 | 76 | 360 | 0 | Urban | N |
| LP001038 | Male | Yes | 0 | Not Gradu | No | 4887 | 0 | 133 | 360 | 1 | Rural | N |
| LP001041 | Male | Yes | 0 | Graduate | | 2600 | 3500 | 115 | | 1 | Urban | Y |
| LP001043 | Male | Yes | 0 | Not Gradu | No | 7660 | 0 | 104 | 360 | 0 | Urban | N |
| LP001046 | Male | Yes | 1 | Graduate | No | 5955 | 5625 | 315 | 360 | 1 | Urban | Y |

**Fig:** Trained data

| Loan_ID | Gender | Married | Dependen | Education | Self_Empl | ApplicantIn | CoapplicaI | LoanAmou | Loan_Amo | Credit_His | Property_Area |
|---|---|---|---|---|---|---|---|---|---|---|---|
| LP001015 | Male | Yes | 0 | Graduate | No | 5720 | 0 | 110 | 360 | 1 | Urban |
| LP001022 | Male | Yes | 1 | Graduate | No | 3076 | 1500 | 126 | 360 | 1 | Urban |
| LP001031 | Male | Yes | 2 | Graduate | No | 5000 | 1800 | 208 | 360 | 1 | Urban |
| LP001035 | Male | Yes | 2 | Graduate | No | 2340 | 2546 | 100 | 360 | | Urban |
| LP001051 | Male | No | 0 | Not Gradu | No | 3276 | 0 | 78 | 360 | 1 | Urban |
| LP001054 | Male | Yes | 0 | Not Gradu | Yes | 2165 | 3422 | 152 | 360 | 1 | Urban |
| LP001055 | Female | No | 1 | Not Gradu | No | 2226 | 0 | 59 | 360 | 1 | Semiurban |
| LP001056 | Male | Yes | 2 | Not Gradu | No | 3881 | 0 | 147 | 360 | 0 | Rural |
| LP001059 | Male | Yes | 2 | Graduate | | 13633 | 0 | 280 | 240 | 1 | Urban |
| LP001067 | Male | No | 0 | Not Gradu | No | 2400 | 2400 | 123 | 360 | 1 | Semiurban |
| LP001078 | Male | No | 0 | Not Gradu | No | 3091 | 0 | 90 | 360 | 1 | Urban |
| LP001082 | Male | Yes | 1 | Graduate | | 2185 | 1516 | 162 | 360 | 1 | Semiurban |
| LP001083 | Male | No | 3+ | Graduate | No | 4166 | 0 | 40 | 180 | | Urban |
| LP001094 | Male | Yes | 2 | Graduate | | 12173 | 0 | 166 | 360 | 0 | Semiurban |
| LP001096 | Female | No | 0 | Graduate | No | 4666 | 0 | 124 | 360 | 1 | Semiurban |
| LP001099 | Male | No | 1 | Graduate | No | 5667 | 0 | 131 | 360 | 1 | Urban |
| LP001105 | Male | Yes | 2 | Graduate | No | 4583 | 2916 | 200 | 360 | 1 | Urban |
| LP001107 | Male | Yes | 3+ | Graduate | No | 3786 | 333 | 126 | 360 | 1 | Semiurban |
| LP001108 | Male | Yes | 0 | Graduate | No | 9226 | 7916 | 300 | 360 | 1 | Urban |
| LP001115 | Male | No | 0 | Graduate | No | 1300 | 3470 | 100 | 180 | 1 | Semiurban |
| LP001121 | Male | Yes | 1 | Not Gradu | No | 1888 | 1620 | 48 | 360 | 1 | Urban |
| LP001124 | Female | No | 3+ | Not Gradu | No | 2083 | 0 | 28 | 180 | 1 | Urban |

**Fig:** Test data

The given picture shows a test dataset that shares some attributes with the training dataset. In this scenario, we adopted an 80:20 ratio for instruction and assessment the model. When comparing the two models that are crucial to consider other measurements like precision, recall, and F1 score. The metrics may differ according to the particular needs of your loan application prediction task. We found because the precision of the decision tree model was around 70.73% while the naive Bayesian model had an accuracy of about 83%. The following images will show these details

```
In [48]: from sklearn import metrics
         print('The accuracy of decision tree is: ', metrics.accuracy_score(y_pred,y_test))

         The accuracy of decision tree is:  0.7073170731707317
```

**Fig**: Accuracy of Decision Tree

```
In [52]: print('The accuracy of Naive Bayes is: ', metrics.accuracy_score(y_pred,y_test))

         The accuracy of Naive Bayes is:  0.8292682926829268
```

**Fig**: Accuracy of Naïve Bayes

We have decided to choose the Naive Bayes model with 83% accuracy as your primary model for predicting the loan application, you can proceed as follows:

Validation: Make sure the accuracy that is obtained as 83% is invariable across various validation sets. This validation process is vital for the confirmation among the model's reliability and its capability to perform efficiently across diverse datasets.

Feature Importance: Research calculated importance from implementation of Naive Bayes model. Analysis of feature importance can supply valuable clues as to which factors are most important and offer critical interpretations into the determining factors of loan approval.

The issue is detected using the test information applying it to the model, which is produced from the trained data through the use of Naïve Bayes Algorithm. The ultimate loan status, therefore, includes a binary response (1 or 0), reflecting whether a particular customer is approved or denied a loan. The score of one (1) demonstrates that the client is eligible for loan approval but a status of zero (0) implies that the applicant is not in the position to acquire a loan. This binary classification does help in interpreting the model's prediction, giving the certainty that the customer is eligible for the loan if it is classified as such or otherwise.

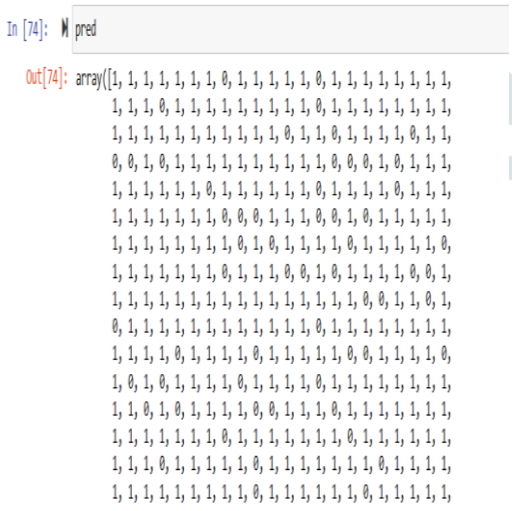


**Fig :** Predicated result

## IV. CONCLUSION

This article on the effectiveness of utilizing the modern loan-approval systems that are substituted for smart machine learning techniques is highlighted in this text. These articles argue that, by consulting significant factors such as loan term, amount applied for, applicant's age, their income and credit history, the banking sector can practice evidence-based and informed decision-making in its assessment process. They learn from historical data, tend to be formulated predictively and thus could revolutionize the banking sector. Out of all the approaches used, tree-structured methodologies, mainly those employing boosting techniques and decision trees, have demonstrated the greatest efficacy and relevance regarding power to generalize and interpretability. Credit history remains a key factor in creditworthiness, but the trend of making decisions with the use of overly simplistic trees has its own drawbacks. Bank credit models are acknowledged by many as imperfect representations of the complex processes of bank credit appraisal. With the implementation of our bank loan forecasting setup that is predicted on machine learning, we have found a solid and dependable tool for exactly predicting loan approvals. This collaborative strategy effectively minimizes the dangers of fraud, accelerates the process and, thus, saves a lot of time for both financial entities and borrowers. The future potential of breakthroughs and the need to input new data for testing is acknowledged, but these studies help us in understanding the statistical learning techniques as they did already in the past. Ultimately, this helps to facilitate process automation, improve the credit approval process, and reduce the risks in the constantly changing environment of banking.

## V. REFERENCES :

[1] Vishal Singh and Ayushman Yadav, "Prediction of Modernized Loan Approval System Based on Machine Learning Approach" IEEE, 2021.
[2] Mohammad J. Hamayel and Mohammad More, "Improving personal loan delivery methods in banks using machine learning methods and approaches in Palestine", IEEE, 2021.
[3] Mohamed Alaradi and Sawsan Hilal, "Tree Methods for Credit Approval", IEEE 2020.
[4] Anshika Gupta and Vinay Pant, "A Bank Loan Forecasting System Using Machine Learning", IEEE 2020.
[5] G. Arutjothi and C. Senthamarai, "Loan condition prediction in commercial bank using machine learning classifier", IEEE 2017.
[6] Deepak Ishwar Gouda, Ashok Kumar, Anil Manjunatha Madivala, Dilip Kumar, "Loan Approval Prediction Based On Machine Learning", International Research Journal of Engineering and Technology IRJET], 2021.
[7] Dr. C. K. Gomathy, Ms. Charulatha, Mr. Akash and Ms. Sowjanya, "Loan Forecasting Using Machine Learning", International Research Journal of Engineering and Technology (IRJET), Volume: 08 Issue: 10 | October 2021.
[8] P. L. Srinivasa Murthy, G. Soma Shekar, P. Rohith, and G. Vishnu Vardhan Reddy, "Loan Approval Prediction System Using Machine Learning" IEEE, vol. 4(1), January–June 2020.

[9] N.Guna sree, M.Divya, N.Reshma, Mr. K.Rajendra Prasad, "Loan Approval Prediction Using Machine Learning Algorithm- Decision tree" IEEE, Vol-11 Issue-01 – 2021.

[10] Sachin Magar, N.S.Nikam, Nilesh Taksale and Suprem Hajare "Loan Eligibility Prediction Using Machine Learning Algorithms", Journal of Emerging Technologies and Innovative Research (JETIR), Volume 9, Number 8, August 2022.