

Loan Approval Prediction using Machine Learning

Ms Sharada P. Chavhan¹ and Prof. Dr. N. R. Wankhade²

ME Student, Department of Computer Engineering¹

Head of Department, Department of Computer Engineering²

Late. G. N. Sapkal College of Engineering, Nashik, Maharashtra, India

Abstract: *In today's digital age, financial institutions are facing a significant challenge in managing the increasing volume of loan applications. Traditional loan approval methods, which involve manual evaluation by credit analysts, are time-consuming and prone to errors. To overcome these challenges, machine learning algorithms have emerged as a promising solution for automating the loan approval process. This abstract discusses the use of machine learning algorithms for loan approval prediction. The proposed approach involves collecting and preprocessing a large dataset containing historical loan applications, including various financial and personal attributes. The data is then fed into a machine learning model that predicts the likelihood of loan approval based on the input features. The model is trained using supervised learning techniques such as Random Forest, Logistic Regression, Support Vector Machine, XGboost, Decision Tree, Python. The selected model is then integrated into the lending institution's loan approval process, replacing or augmenting the manual evaluation process. The benefits of this approach include faster processing times, reduced errors, and more consistent decision-making. Additionally, machine learning algorithms can provide insights into the factors that influence loan approval decisions, enabling lending institutions to make more informed decisions and improve their overall loan portfolio management strategies. Machine learning algorithms have the potential to revolutionize the loan approval process by providing a more efficient and accurate alternative to traditional methods. As the volume of loan applications continues to grow, it is essential for lending institutions to adopt these technologies to remain competitive and provide better service to their customers.*

Keywords: Random Forest, Logistic Regression, Support Vector Machine, XGboost, Decision Tree, Python

I. INTRODUCTION

Loan approval prediction using machine learning is a process of developing algorithms and models that can accurately predict whether an individual or a business is likely to be approved for a loan based on various financial and personal factors. This technology has revolutionized the banking and financial industry by providing a faster, more efficient, and more accurate way of assessing loan applications. The traditional loan approval process involves a manual review of an applicant's financial history, credit score, employment status, and other personal information. This process can take several days or even weeks, which can be frustrating for both the applicant and the lender. Moreover, the manual review process can be subjective and prone to errors, leading to inconsistent decisions. Machine learning algorithms, on the other hand, can analyze large volumes of data in real-time and provide an instant decision on whether the loan application is likely to be approved or not. These algorithms use various statistical techniques, such as Random forest, Logistic regression, Support vector machine, XGboost, Decision tree to predict the likelihood of loan repayment based on historical data. The accuracy of these algorithms depends on the quality and quantity of the data used to train them. The more data that is available, the more accurate the predictions will be. Moreover, these algorithms can continuously learn from new data and improve their predictions over time. To improving efficiency and accuracy, machine learning algorithms also help lenders to reduce risk by identifying potential loan defaults early on. By analyzing historical data and predicting which loans are most likely to default, lenders can take proactive measures to mitigate risk and prevent losses. Loan approval prediction using machine learning is a powerful technology that is transforming the banking and

financial industry by providing faster, more efficient, and more accurate loan approval decisions while also helping lenders to reduce risk. As this technology continues to evolve, we can expect even more sophisticated algorithms and models that will further enhance the accuracy and effectiveness of loan approval prediction using machine learning.

II. OBJECTIVE & SCOPE OF PROPOSED SYSTEM

1. The objective of this project is to develop a machine learning model that can predict whether a loan application will be approved or not based on various features and attributes of the applicant. The model will be trained on historical loan application data, and its accuracy will be evaluated using various performance metrics. The ultimate goal is to provide a reliable and efficient tool for lending institutions to make informed decisions about loan approvals, reducing the risk of default and improving overall financial performance.
2. To develop a predictive model that accurately predicts whether an applicant will be approved for a loan or not based on their financial and personal information.
3. To identify the most important factors that influence loan approval decisions, such as credit score, income, employment history, and debt-to-income ratio.
4. To provide insights into the loan approval process and help financial institutions make more informed decisions about lending.
5. To improve the efficiency and accuracy of loan processing by automating the decision-making process, reducing the need for manual review and approval by loan officers.
6. To provide a better user experience for loan applicants by providing them with quicker and more accurate feedback on their loan applications, reducing wait times and uncertainty.
7. Use techniques such as pruning, early stopping, or model compression to reduce the size and complexity of the model while maintaining accuracy. This can improve its efficiency in a production environment with limited resources such as memory or CPU power.
8. Use techniques such as regularization, dropout, or ensemble methods to improve the robustness of the model in the face of new and unseen data with different distributions or correlations than those seen during training. This can prevent overfitting and improve its generalizability in a production environment with diverse customer profiles and loan types.
9. Use techniques such as online learning, active learning, or reinforcement learning to continuously update and refine the model based on new data and feedback from customers or loan officers. This can improve its accuracy and relevance over time while minimizing risks and costs associated with loan defaults or frauds.

III. FEATURES OF PROJECT

1. User behavior patterns
2. Data analysis
3. Credit scoring
4. Risk assessment
5. Automation
6. Fraud detection
7. Improved accuracy
8. Personalized loan offers
9. Real-time decision making
10. Regulatory compliance
11. Improved customer experience

IV. Literature Review

In this research work prediction model uses several attributes of the applicant which also include non-financial attributes, obtained highly reliable model when compared to the ones which include only financial attributes. For our dataset the best accuracy is achieved for the XDBoost model which is 0.82 and clearly see in results that XGBoost is best in all metrics and conclude it is best model for taken dataset. To build an ML application which can reduce the

time required to approve a loan using ML based prediction model to approve the loan with minimal human intervention by filtering huge number of applications and forward very few applications for human verification. In this work used several popular machine learning models are Random Forest tree, Logistic Regression, Decision tree, Support Vector machine and XGB. The results demonstrated that XGB has provided higher accuracy in the presenting assessment.[1]

In this paper, Machine Learning (ML) algorithms are used to extract patterns from a common loan approved dataset and retrieve patterns in forecasting future loan defaulters. Customers' past data, such as their age, income, loan amount, and tenure of work, will be used to conduct the analysis. To determine the maximum relevant features, i.e. the factors that have the most impact on the prediction outcome, various ML algorithms such as Random Forest, Support Vector Machine, K-Nearest Neighbor and Logistic Regression, were used. These mentioned algorithms are evaluated with the standard metrics and compared with each other. The random forest algorithm achieves better accuracy.[2]

In this loan prediction system based on machine learning is developed, in which the system will automatically identify the qualified candidates. This is beneficial to both the bank personnel and the applicant. The loan approval process will be greatly shortened. The loan data is predicted by using the hybrid model of Naïve Bayes (NB) and Decision Tree (DT) algorithms. First, the dataset is given to the three classification algorithms – Support Vector Machine (SVM), NB and DT Algorithms and the prediction is done with these three algorithms. The accuracy of each of these three is used to assess performance. The creation of the hybrid model increases accuracy. The dataset is given to NB for training and the prediction of NB is given to DT Algorithm for training. Test data are sent to the model for prediction after training. The model is evaluated, and the performance is measured in terms of different metrics form sklearn metrics. This prediction of loan range is useful for bank staff to give the loan amount accordingly. The NB algorithm checks for equality and independence of all the features in the dataset. In DT algorithm, the tree is constructed based on the information gain value. The attribute with high information gain value is placed as the root node and also the other nodes are constructed based on information gain value. The proposed hybrid model predicts - yes or no, and based on the prediction, whether the loan is to be sanctioned or denied for the applicant is specified.[3]

" Loan Approval Prediction based on Machine Learning Approach" Author- Kumar Arun, Garg Ishan, Kaur Sanmeet Year- 2018 The main objective of this paper is to predict whether assigning the loan to particular person will be safe or not. This paper is divided into four sections (i) Data Collection (ii) Comparison of machine learning models on collected data (iii) Training of system on most promising model (iv) Testing.[1]

"Loan Prediction using machine learning model" Year- 2019 whether or not it will be safe to allocate the loan to a particular person. This paper has the following sections (i) Collection of Data, (ii) Data Cleaning and (iii) Performance Evaluation. Experimental tests found that the Naïve Bayes model has better performance than other models in terms of loan forecasting. With the enhancement in the banking sector lots of people are applying for bank loans but the bank has its limited assets which it has to grant to limited people only, so finding out to whom the loan can be granted which will be a safer option for the bank is a typical process. So in this project we try to reduce this risk factor behind selecting the safe person so as to save lots of bank efforts and assets. This is done by mining the Big Data of the previous records of the people to whom the loan was granted before and on the basis of these records/experiences the machine was trained using the machine learning model which give the most accurate result.[2]

"Loan Prediction using Decision Tree and Random Forest" Author- Kshitiz Gautam, Arun Pratap Singh, Keshav Tyagi, Mr. Suresh Kumar Year-2020. In India the number of people or organization applying for loan gets increased every year. The bank have to put in a lot of work to analyse or predict whether the customer can pay back the loan amount or not (defaulter or non-defaulter) in the given time. The aim of this paper is to find the nature or background or credibility of client that is applying for the loan. We use exploratory data analysis technique to deal with problem of approving or rejecting the loan request or in short loan prediction. The main focus of this paper is to determine whether the loan given to a particular person or an organization shall be approved or not.[6]

Xin Li, Xianzhong Long, Guozi Sun, Geng Yang, and Huakang Li This paper mainly introduces the main application of LSTM-SVM model in user loan risk prediction, and elaborates the current economic background, traditional risk forecasting method. On this basis, the prediction methodology based on LSTM method and SVM method is proposed, and the prediction results are compared with the traditional algorithm, and the feasibility of the model is confirm. However, the LSTM-SVM method proposed in this paper actually has few limits and needs to be improved in future research [7].

Kumar Arun and Garg Ishan, in their research paper tested a total of six different machine learning approaches, including neural networks, support vector machines, random forests, decision trees, linear models, and Adaboost. There are four sections to this study. (i) Gathering of data (ii) Model evaluation using ML on the collected information (iii) System training using the most feasible model (iv) After the system has been trained on the most promising model, it is put to the test. R programming language was used to create this system. They didn't represent the data results for easier comprehension and comparison, but this problem can be solved by offering data visualization in the form of graphs or other matrix forms.[8]

Authors Initially said, the information was cleansed. The next step was exploratory data analysis and feature engineering. They had done visualization through graphs. For loan prediction, four models are used. Decision Tree, Naive Bayes, Support Vector Machines, and Logistic Regression methods are the four methods. They determined confidently showing the Naive Bayes model is very capable of delivering superior results to other models after thoroughly studying positive attributes and constraints.[9]

Authors said a set of data was obtained from the banking sector. The data set is in the ARFF (Attribute-Relation File Format) format, which Weka understands. They used exploratory data analysis to solve the challenge of granting or rejecting loan requests, as well as short-term loan projection. In their research, they did an exploratory data analysis. For prediction, two machine learning classification models are used Decision Tree and Random Forest. In their analysis, they chose the random forest method. [10]

The author, Vaidya, Ashlesha uses logistic regression as a machine learning tool in paper and shows how predictive approaches can be used in real world loan approval problems. His paper uses a statistical model (Logistic Regression) to predict whether the loan should be approved or not for a set of records of an applicant. Logistic regression can even work with power terms and nonlinear effect. Some limitations of this model are that it requires independent variables for estimation and a large sample is required for parameter estimation. [11]

The research and work done by Arora, Nisha and Pankaj Deep Kaur aimed at forecasting whether an applicant can be a loan defaulter or not. It uses Bolasso to select most relevant attributes based on their robustness and then applied to classification algorithms like Random Forest, SVM, Naive Bayes and KNearest Neighbours (KNN) to test how accurately they can predict the results. It is concluded that Bolasso enabled Random Forest algorithm (BS-RF) provides the best results in credit risk evaluation and gives better accuracy by using optimised feature selection methods. [12]

In paper authored by Yang, Baoan, et al., the use of artificial neural networks in an early warning system for predicting loan risk is discussed wherein it covers the early warning signals for deteriorating financial situations. The ability of an applicant to repay the loan is determined to be the most relevant aspect in the financial analysis. The early warning system in this paper uses artificial neural network that is utilizing the traditional early warning concepts. This system based on ANN proves to be a very effective decision tool and early warning system for banks and other commercial lending organizations. [13]

V. REPRESENTATION OF THE METHODOLOGY

Collect relevant data for loan applicants, including attributes such as age, income, credit score, employment history, loan amount, etc. This data can be obtained from various sources such as banks, credit bureaus, or online loan applications. Clean the collected data by removing any inconsistencies, duplicates, or missing values. Perform feature engineering, which involves transforming and selecting relevant features from the dataset. This can include techniques like one-hot encoding, normalization, or handling missing values. Split the dataset into training and testing sets. The training set will be used to train the machine learning model, while the testing set will be used to evaluate the model's performance. Choose an appropriate machine learning algorithm for loan approval prediction. Popular algorithms for this task include Random forest, Logistic regression, Support vector machine, Xgboost, Decision tree. Select the algorithm that best suits the problem and dataset. Train the selected machine learning model using the training dataset. The model learns from the input features and the corresponding loan approval labels in this step. The training process involves adjusting the model's parameters to minimize the prediction error. Evaluate the trained model's performance using the testing dataset. Common evaluation metrics for loan approval prediction include accuracy, precision, recall, and F1 score. These metrics provide insights into how well the model predicts loan approvals Fine-tune the model's hyper parameters to improve its performance. Hyper parameters are configuration settings of the machine learning

algorithm that are not learned during the training process. Tuning them optimizes the model's performance on the validation set. Once the model is trained and tuned, use it to predict the loan approval probability for new loan applicants. Apply the model to unseen data using the learned parameters and evaluate the loan approval probability. This involves integrating the model into the business workflow or application, where it can be used to automatically predict loan approvals for new applicants.

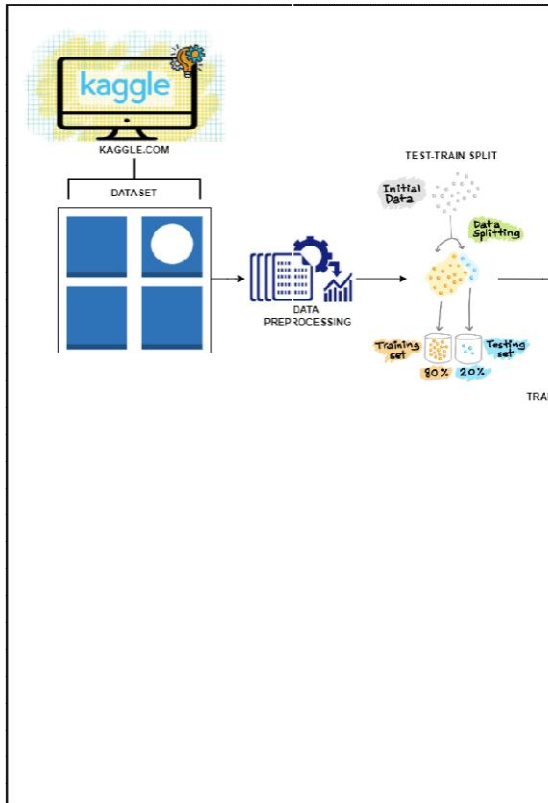


Fig : Representation Of The Methodology

VI. PROGRAMMING ARCHITECTURE

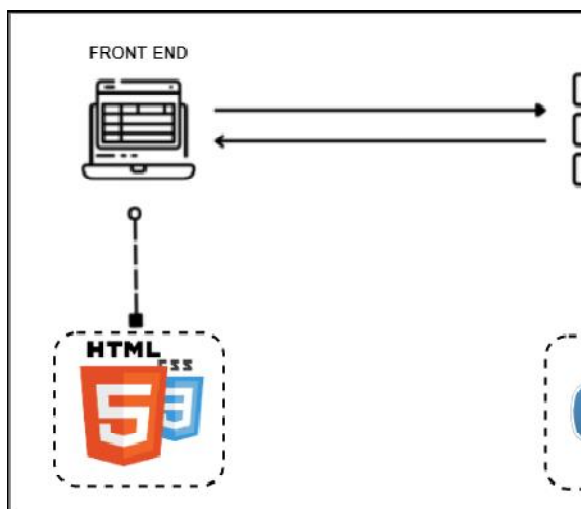


Figure: Programming Architecture

VII. ADVANTAGES

1. **Improved Accuracy:** Machine learning algorithms can learn complex patterns and relationships in the data that may not be easily detectable by traditional statistical methods. This can lead to more accurate predictions of loan approval, reducing the risk of default and improving the overall profitability of the lending institution.
2. **Faster Processing:** Machine learning models can process large volumes of data quickly and efficiently, allowing lenders to make decisions in real-time or near-real-time. This can enable faster loan approvals, which can be a competitive advantage in a crowded marketplace.
3. **Better Risk Management:** By considering a wide range of factors and variables, machine learning models can provide a more holistic view of the borrower's creditworthiness and potential risk of default. This can help lenders make more informed decisions about loan terms, interest rates, and collateral requirements, minimizing their exposure to credit losses.
4. **Enhanced Customer Experience:** By providing personalized loan offers and recommendations based on the borrower's unique circumstances and preferences, machine learning models can improve the overall customer experience and satisfaction. This can lead to higher customer loyalty and repeat business, as well as positive word-of-mouth referrals.
5. **Greater Transparency:** By providing clear explanations of how the model makes its predictions based on the input features, machine learning models can promote greater transparency and fairness in the lending process. This can help build trust and confidence with stakeholders, regulators, and customers, as well as reduce the risk of legal or reputational issues.
6. **Lower Costs:** By automating many of the manual processes involved in loan approval and underwriting, machine learning models can reduce the costs associated with these activities, such as staffing, training, and compliance. This can improve the overall profitability of the lending institution by freeing up resources for other strategic initiatives or investments.

VIII. APPLICATION AREAS

1. Consumer loans
2. Mortgage loans
3. Business loans
4. Peer-to-peer lending
5. Fraud detection
6. Credit risk assessment
7. Consumer Lending
8. Small Business Lending
9. Mortgage Lending
10. Student Loan Lending
11. Commercial Real Estate Lending
12. Credit Card Lending

IX. HARDWARE REQUIREMENTS

1. CPU Quad Core (not counting hyper-threading) 2.4Ghz, Intel VT or AMDV (Intel i3 or better)
2. Memory 4 GB
3. The ability to install more memory is desirable. Disk 512 GB SSD or better
4. Graphics Accelerated, Gaming Support Nvidia is preferred over AMD 1920 by 1080 resolution is recommended (at least on an external port) At least 1280 by 1024 resolution
5. HDMI output recommended (perhaps with an adapter)
6. Mouse An external mouse (USB or Bluetooth) is desirable.
7. USB USB 3.0 desirable for an external disk Other USB ports may be needed for: mouse, printer, mic-in, and headphones-out, depending on how these are connected.
8. External monitor A 23 or larger HDMI monitor is recommended, with reasonable resolution.

9. Laptop or Desktop Windows 11 or macOS 12.4 or above. Linux is also acceptable if a mainstream distribution (e.g. Ubuntu).

X. SOFTWARE REQUIREMENTS

1. Operating System: Windows XP and later versions
2. Front End: HTML, CSS
3. Programming Language: Python
4. Dataset: Kaggle.com
5. Domain: Machine Learning
6. Algorithm: Random forest, Logistic regression, Support vector machine, Xgboost, Decision tree.

XI. TEST DATA REQUIREMENTS

Unit Testing

Unit testing concentrates verification on the smallest element of the program – the module. Using the detailed design description important control paths are tested to establish errors within the bounds of the module. In this system each sub module is tested individually as per the unit testing such as campaign, lead, contact etc are tested individually. Their input field validations are test

Integration testing

Once all the individual units have been tested there is a need to test how they were put together to ensure no data is lost across interface, one module does not have an adverse impact on another and a function is not performed correctly. After unit testing each and every sub module is tested with integrating each other.

XII. SYSTEM TESTING FOR THE CURRENT SYSTEM

In this level of testing we are testing the system as a whole after integrating all the main modules of the project. We are testing whether system is giving correct output or not. All the modules were integrated and the flow of information among different modules was checked. It was also checked that whether the flow of data is as per the requirements or not. It was also checked that whether any particular module is non-functioning or not i.e. once the integration is over each and every module is functioning in its entirety or not.

1. Functional testing: this involves testing the functionality of the system to ensure that it meets the required specifications and performs as expected. This includes testing the churn prediction accuracy, input data handling, and output interpretation.
2. Performance testing: this involves testing the system's performance under different load conditions to ensure that it can handle the expected workload and respond within acceptable time limits. This includes testing the system's scalability, resource utilization, and response time.
3. Security testing: this involves testing the system's security features to ensure that it can protect sensitive customer data from unauthorized access, theft, or misuse. This includes testing the system's authentication, authorization, and encryption mechanisms.
4. Compatibility testing: this involves testing the system's compatibility with different operating systems, databases, and hardware configurations to ensure that it can operate in a variety of environments.
5. Usability testing: this involves testing the system's user interface and user experience to ensure that it is intuitive, easy to use, and meets the needs of the end-users.
6. Regression testing: this involves testing the system's functionality after making changes or updates to ensure that the changes have not introduced any unintended side effects or regressions.
7. Acceptance testing: this involves testing the system's functionality from the perspective of the end-users to ensure that it meets their requirements and expectations. This includes testing the system's accuracy, reliability, and ease of use.
8. Recovery testing: this involves testing the system's ability to recover from failures, errors, or disasters to ensure that it can continue operating and providing service to the end-users.

9. Stress testing: this involves testing the system's performance under extreme load conditions to ensure that it can handle unexpected or catastrophic events.
10. Exploratory testing: this involves testing the system's functionality and behavior in unanticipated or unexpected scenarios to ensure that it can handle unexpected situations and provide accurate and reliable results. In this level of testing we tested the following: -
 - Whether all the forms are properly working or not.
 - Whether all the forms are properly linked or not.
 - Whether all the images are properly displayed or not.
 - Whether data retrieval is proper or not

XIII. CONCLUSION

In conclusion, the use of machine learning algorithms for loan approval prediction is a promising solution for financial institutions facing the challenge of managing increasing volumes of loan applications. By collecting and preprocessing historical loan application data, training machine learning models using supervised learning techniques, and evaluating their performance using various metrics, lending institutions can automate the loan approval process, resulting in faster processing times, reduced errors, and more consistent decision-making. Additionally, machine learning algorithms can provide insights into the factors that influence loan approval decisions, enabling lending institutions to make more informed decisions and improve their overall loan portfolio management strategies. As the volume of loan applications continues to grow, it is essential for lending institutions to adopt these technologies to remain competitive and provide better service to their customers.

REFERENCES

- [1]. ChintamAnusha,,Rajendra Kumar G,” An Approach to Loan Approval prediction Using Boosting Ensemble Learning”,2023 Third International Conference on Advances in Electrical, Computing, Communication and Sustainable Technologies (ICAECT) | 978-1-6654-9400-7/23/\$31.00 ©2023 IEEE | DOI: 10.1109/ICAECT57570.2023.10118093,
- [2]. Praveen Tumuluru, Lakshmi Ramani Burra,” Comparative Analysis of Customer Loan Approval Prediction using Machine Learning Algorithms”, 2022 Second International Conference on Artificial Intelligence and Smart Energy (ICAIS) | 978-1-6654-0052-7/22/\$31.00 ©2022 IEEE | DOI: 10.1109/ICAIS53314.2022.9742800, IEEE Xplore Part Number: CFP22OAB-ART; ISBN: 978-1-6654-0052-7
- [3]. Kavitha M N, Saranya S S, Dhinesh E,” Hybrid ML Classifier for Loan Prediction System”, 2023 International Conference on Sustainable Computing and Data Communication Systems (ICSCDS) | 978-1-6654-9199-0/23/\$31.00 ©2023 IEEE | DOI: 10.1109/ICSCDS56580.2023.10104831, IEEE Xplore Part Number: CFP23AZ5-ART; ISBN: 978-1-6654-9199-0
- [4]. Kumar Arun, Garg Ishan, Kaur Sanmeet, —Loan Approval Prediction based on Machine Learning Approachl, IOSR Journal of Computer Engineering (IOSR-JCE), Vol. 18, Issue 3, pp. 79-81, Ver. I (May-Jun. 2016).
- [5]. Pidikiti Supriya, Myneedi Pavani, NagarapuSaisushma, Namburi Vimala Kumari, kVikash,“Loan Prediction by using Machine Learning Models”, InternationalJournalofEngineeringandTechniques.Volume 5 Issue 2, Mar-Apr 2019
- [6]. Pratap Singh, Keshav Tyagi,”Loan Prediction using Decision tree and Random Forest ”, Journal of the Gujrat Research History, Volume 21 Issue 14s, December 2020.
- [7]. Xin Li, Xianzhong Long, Guozi Sun, Geng Yang, and Huakang Li “Overdue Prediction of Bank Loans Based on LSTM-SVM”Jiangsu Key Lab of Big Data and Security and Intelligent Processing Nanjing University of Posts and Telecommunications, Nanjing, 210023, China.
- [8]. Kumar Arun, Garg Ishan, Kaur Sanmeet- Loan Approval Prediction based on Machine Learning Approach- IOSR Journal of Computer Engineering, p-ISSN: 2278-8727 PP 18-21.

- [9]. E. Chandra Blessie, R. Rekha- Exploring the Machine Learning Algorithm for Prediction the Loan Sanctioning Process- (IJITEE) ISSN: 2278-3075, Volume-9 Issue-1, November 2019.
- [10]. Kshitiz Gautam, Arun Pratap Singh, Keshav Tyagi - Loan Prediction using Decision Tree and Random Forest- (IRJET) Volume: 07 Issue: 08 | Aug 2020.
- [11]. Vaidya and Ashlesha, Predictive and probabilistic approach using logistic regression: Application to prediction of loan approval, 2017 8th International Conference on Computing, Communication and Networking Technologies (ICCCNT). IEEE, 2017.
- [12]. Arora, Nisha and Pankaj Deep Kaur, A Bolasso based consistent feature selection enabled random forest classification algorithm: An application to credit risk assessment, Applied Soft Computing 86 (2020), 105936.
- [13]. Yang, Baoan, et al, An early warning system for loan risk assessment using artificial neural networks, Knowledge-Based Systems 14.5-6 (2001), 303-306.