# Loan Approval Predication By Using Machine Lerning Algorithm


1st Ibrahim A. Al Husaini
*dept. of Computer Science Imam*
Abdulrahman Bin Faisal
University
kingdom of saudi arabia
2210002314@iau.edu.sa

3th Noor A. Hadari
*dept. of Computer Science Imam*
Abdulrahman Bin Faisal
University
kingdom of saudi arabia
2210002310@iau.edu.sa

4rd Rayyan B. Al Nahwi
*dept. of Computer Science Imam*
Abdulrahman Bin Faisal
University
kingdom of saudi arabia
2210002103@iau.edu.sa

2th Nasser A. Alhajri
dept. of Computer *dept. of Computer Science Imam* Abdulrahman BinFaisal University kingdom of saudi arabia 2210006797@iau.edu.sa

5th Aseel A. Alrudayni
*dept. of Computer Science Imam*
Abdulrahman Bin Faisal
University
kingdom of saudi arabia
2210001894@iau.edu.sa



**ABSTRACT**

This study focuses on employing AI and ML; approaches for loan approval prediction in the FinTech context. With the acceptance of datamining in the finance industry as a tool to help improve the lending systems, the need to come up with efficient loan underwriting systems has become inevitable. This paper performs a literature review and analysis of the current research and processes, dividing ML algorithms including Logistic Regression, Random Forest, Gradient Boosting, and more, which rely on credit score, income and employment history to predict loan approval results.

The review focuses on the assessment of the performance of the employed predictive models, discussion of such important issues as fairness and bias, and an overview of the main directions for further research. Also, it raises awareness of how important the data preprocessing and the features extracting steps are for the model. Hypothesizing that integrating findings across studies would improve comprehension of AI facilitated loan approval systems, the project endeavors to refine decision making precision and reduce risks for lenders. The outcomes will help progress the loan approval prediction models into being more just, accurate and stable ones in meeting organizational needs.


## 1. INTRODUCTION

AI and ML are the real-life technologies that have helped transform several sectors, and one of the most affected will be the financial sector. Loan approval prediction is one of the paramount use cases in the context of FinTech since such decisions should be accurate, swift, and risk-free. Lenders use AI systems to evaluate loan applications' risk, seeking better outcomes on loan approvals that are accurate, fair, and efficient at scale.

The relevance of this topic is dictated by the increase in the need for efficient, nondiscriminatory, and fair credit provision. The older ways of working on loans require comprehensive analyses that very often are done manually, so they're inaccurate and subjective. , AI and ML are more analytical in their approach by using the data collected, economy, and customer behaviors to get to a certain decision. These technologies may enhance the approval rates and, at the same time, decrease default threats, which may in turn bring equity in credit access.

By tackling issues of equity, impartiality, and the ability to balance a changing economic climate, this study maps out the ways in which AI has the ability to revolutionalise the future of lending. Furthermore, it illustrates case studies that demonstrate that the use of predictive analytics promotes shift towards innovation and improvements within the sphere of Finance.[1]

### 1.1 Background

Loan approval prediction is an important function that is being widely implemented in the FinTech industry which has significantly benefited from advanced technologies, specifically AI and Machine Learning, in terms of evaluating and selecting the applicants. These systems use such characteristics that includes credit rating, income, working status, and earlier credit performance to gauge the probability of an applicant to pay. There are different types of a machine learning model used; each model has its merits and demerits. Logistic Regression is plaintext and can often be used as a starting point for comparison with more complex models. Random Forest improves the level of probable prediction and opposes over-fitted models using the multiple decision trees technique; for Gradient Boosting, techniques may include XGBoost suitable techniques for structured data especially since they operate in cycles of turning out minimum errors.

Decision Trees are easy to understand and interpret but can very much overfit and Support Vector Machine performs brilliantly when data points are in high dimensions but is computationally expensive. On data preprocessing the study proposes data cleaning, feature scaling, and selection to enhance the performance of a model. Cross-validation includes four steps that test the models using datasets other than development and determine their correspondence or agreement while remembrance provides clear performance measures, including accuracy, precision, recall, and F1-score. It also briefly touches on fairness, flexibility, and ethical issues, and the methods to reduce bias, model drift and data imbalance. Through incorporating of these sophisticated methodologies, the study aims at designing smart, equitable, and flexible credit delivery solutions which will capture the economic cycles and applicants' statuses. [1]

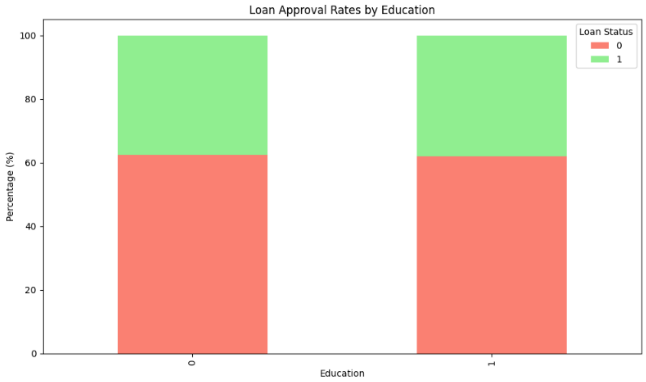


*Figure 1: Loan Approval Rates by Education [31]*

The stacked bar chart compares loan approval rates for different education levels:

- Graduation brings a slightly better approval rate for users as is shown in the green portion of the figure.
- This means that there is possibility that the education level affects the chances of approval of loan, though more elaborate examination is required to make this factor decisive.

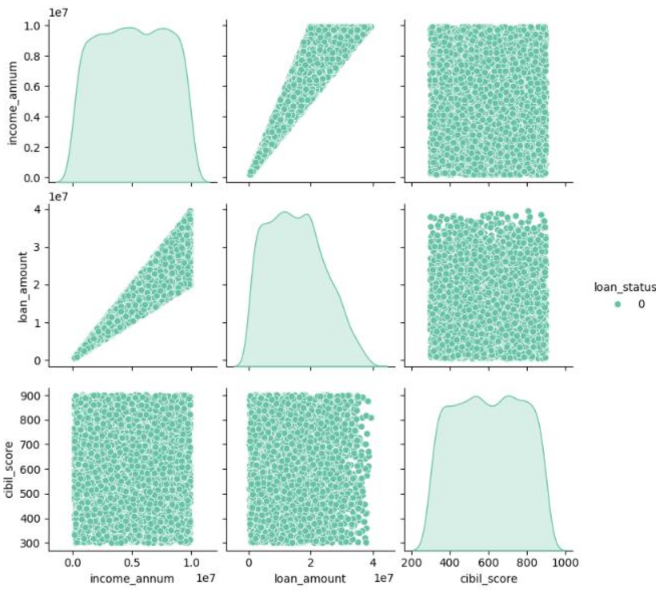


*Figure 2: Pairplot for Exploratory Data Analysis [31]*

The pairplot visualizes relationships between numerical features (income_annum, loan_amount, cibil_score) and the target variable (loan_status). Observations:

- Loan Amount and Income have a linear relationship.
- Higher cibil_score values are associated with approved loans (loan_status = 1), indicating that better credit scores lead to higher approval probabilities.

### 1.2 Aims and Objectives of the Review

This systematic review is principally embarked on to understand how AI & ML are used in loan approval prediction particularly in the FinTech sector. This work aims at surveying related works, evaluating the effectiveness of different models of ML, and also examining the defects of the prevailing approaches. The specific objectives of the review are as follows:

1. **Summarizing Evidence:**
   To offer general perspectives in the existing literature and AI as well as ML methods of loan approval prediction. This includes the various advantages and limitations of such models as Logistic Regression, Random Forest, XGBoost, Decision Trees, and Support Vector Machines.
2. **Identifying Gaps:**
   In order to identify the existing issues in the current literature, including the preprocessing of the data and the lack of study on how the models can be fair with respect to different economic indicators and on how they can include features stemming from the guidelines and rules that regulate financial markets in the current and future economic climates. Particular focus will be devoted to the problem of bias and fairness in loan approval procedures.
3. **Comparing Methodologies:**
   The objective of this research study is therefore to assess and analyse the effectiveness of the various kinds of machine learning algorithms that have been used in the prediction of loan approvals. This entails use of evaluation measures such as accuracy, precision, recall and the F1-score, with the view of ascertaining the applicability of these models, within decision making pursuits.
4. **Proposing Directions for Future Research:**
   To suggest domains for further research that could usefully contribute to enhanced reliability, accuracy, and equity of loan approval predictions. This entails considering new approaches, enhancing the readiness of known approaches as well as considering any evident lacuna.
5. **Improving Understanding and Practice:**
   In order to improve the understanding of the application of AI and ML in the loan approval systems to make better loan approval systems which are reliable, fast and fair for FinTech industry.[2]

### 1.3 Scope of the Review

This review aims at discussing how loan approval systems can leverage AI and machine learning by focusing on aspects like data preprocessing,; algorithms choice; measures of effectiveness; validation methods,; and ethically apprehended methodological approaches; and comparisons. In feature transformation phase, the focus is laid on missing values handling, features construction and scaling the data. The important Machine Learning techniques discussed here are Logistic Regression, Random Forest, Gradient Boosting (example XGBoost), Decision Tree and Support Vector Machines with Pros and Cons. Range measures such as accuracy, precision, recall, F1 score, and margin of error guarantees model performance steadiness while cross-validation, modeling drift, and update strategies retain sturdiness in line with evolving data. Organizational fairness and bias in professionals' management and decision making across demographical dimensions are analyzed for ethical reasons. Last but not least, the review evaluates the relative merits of strategies in the FinTech context and seeks to improve the precision, inclusiveness, and speed of loans approval and the areas for further investigation.[2]

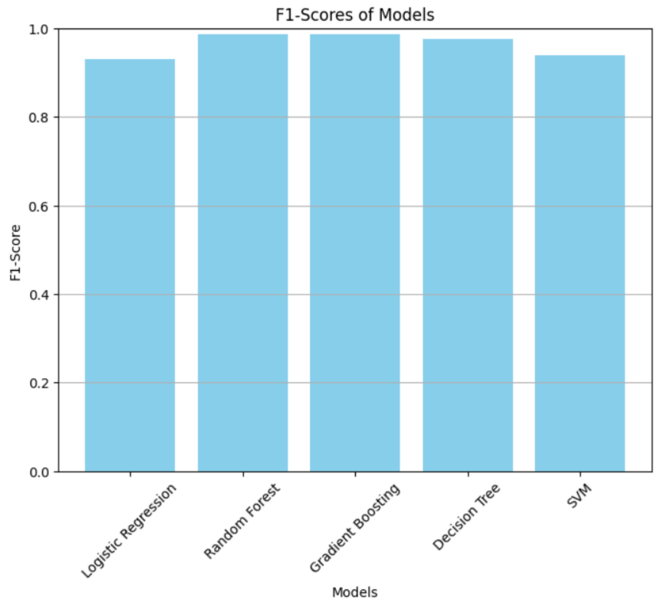


*Figure 3: This bar chart compares the F1-scores of five machine learning models used in the study: Logistic Regression, Random Forest, Gradient Boosting, Decision Tree, SVM. The F1-score is defined as the ratio of two times of Precision multiplied by Recall to the sum of Precision and Recall making this measure highly important from model assessment, especially in the test bed of imbalanced datasets.[31]*

## 2.RELATED WORK

Artificial Intelligence (AI) and Machine Learning (ML) the areas that have revolutionized the financial sector, and especially in loan approval automation. This part presents a review of related literature subscribed to both systematic reviews and surveys in loan approval prediction. Interest lies in examining distinctive features of the studying contributions, evaluating the methodologies proposed, and recognizing the voids in the previous research. In this review, various research done in the field of machine learning techniques for loan approval system has been addressed, with the focus on the findings, the performances, the capabilities of the advanced work and their prospects.[3]

### 2.1 Literature Review:

In the following section, the existing body of work for loan approval prediction using machine learning is reviewed in terms of the methodologies, models, and problems discussed in prior literature. In this section, the author brings out a summary of the prior studies, classified based on the contribution made, advantages or value added as well as the acknowledged weaknesses. The selected areas include Data Preprocessing, Performance of the Algorithms, Evaluation Metrics, and Fairness and Bias in Recommendation Systems, intending to generate some pattern and trends in the field. Based on the synthesis of the findings of this section, a framework can be developed on the improvements achieved as well as the remaining challenges, to inform subsequent discussions and analysis in this research.[3]

**[9] on 2020 Hossam Meshref** discusses an application of both Bagging and Boosting in building a reliable loan approval prediction model. By applying the Kaggle Bank Marketing dataset, the authors developed and evaluated the models that obtained the accuracy of 83.97% and were about 25% higher than numerous advanced models during that period. To help bank decision makers to interpret the predictions, the research focuses on the trade-off between the model accuracy and model interpretability. Through integration of high performance with real time analysis, the proposed system affords reliability in financial decision making. The work also presents how the ensemble techniques can solve the problem of imbalance in data, and enhance the prediction quality and balanced loan approvals.

While extending Hossam Meshref's ideas**, [10] Afrah Khan in January 2021** offers a comparative evaluation of machine learning algorithms for loan approval prediction. The authors carried out a comparative analysis of Logistic Regression, Decision Tree, and Random Forest models using the data from variables such as credit score, income, and marital status. Among the tested models, Random Forest was the most accurate model with the lowest error level and therefore it may be applied to real-world projects. The paper will focus at the problem of biases and errors incurred through a manual loan approval process and highlight how the use of machine learning models can dramatically reduce approval time and enhance the accuracy of determinations. For the benefit of implementing the models outlined in the paper to banking systems, the paper concludes with suggestions in relation to risk mitigation and, more critically, improving decision-making.

Similarly, **in March, [11] Ms. Kathe Rutika Pramod** demonstrates the usefulness of Logistic Regression in analyzing loan defaults. The authors used a dataset from Kaggle to assess the model capability of enhancing banks non-performing asset ratios. This paper highlights the significance of adequate predictions for sound business operation and suggests exploration of better artificial neural net schemes.

When building on the ideas discussed by Ms. Kathe Rutika Pramod, **[12] Ashwini S. Kadam in April** study explores how machine learning algorithm can help in automating the loan approval channelisation which will reduce NPAs for banks. According to the authors, the reliability of traditional methods of manual estimates of loan forecasting is low due to time consumption and errors, so they chose machine learning. The process they used included data gathering, preprocessing, where missing and inconsistent data are addressed, and testing of algorithms including Naïve Bayes and Support Vector Machine (SVM). Of the above mentioned, it was found that Naïve Bayes is one of the best performing algorithm when it comes to accuracy and staleness. Credit history, income, gender and marital status were identified as critical factors influencing loans approval. With regards to the goal of the study, the author determines that the utilisation of machine learning in financial decision making is capable of enhancing efficiencies in financial decision making, improving predictive accuracy and reducing time taken to complete financial processes, to the mutual benefits of both the banking institutions and the applicant.

Similarly, **[13] In May Yash Diwate in May** emphasizes on creating a machine learning model of logistic regression for loan approval prediction results. Lending decisions were predicted by the aid of a dataset containing 1,500 cases; credit history, customer's income, and loan amounts being important predictors. While developing the models, the authors structured their work by data cleaning to treat missing values, training, and performance evaluation. A high accuracy level was established by the logistic regression model, depending on borrower characteristics' significance for credit approval. The paper demonstrates how the use of algorithmic tools can allow banks to reduce losses and improve their credit decisioning process by better identifying safe borrowers. It also describes how they can be used to large scale data sets in practical banking applications.

Thus, following Yash Diwate emphasis on logistic regression analysis, **[14] Dr. Kavita Khadse** examines the process of loan prediction system using both supervised and unsupervised learning algorithm. The work also examines techniques in EDA to gain insights into client behavior, discovering that most applicants prefer short-term credit and debt merge. Logistic Regression and the Decision Trees were used to save loan eligibility. The research focuses on the need to adopt iae in the banks to increase its efficiency and remove bias . These and other possibilities of ML are used to extract actionable insights for lenders while explaining the process as clear and procedural.

Building on the ideas presented by Dr.Kavita Khadse, **[15] Sharayu Dosalwar in September** explores how machine learning can increase the supply and availability of loans. In this paper, there are four models analysed and compared; Support Vector Machines (SVM), Random ForestRegressor and Logistic Regression. The income feature, loan amount, and employment stability were important to the authors as they revealed Random Forest as the most accurate algorithm. The paper also underscore the ability that machine learning has in expanding people's access to loans and at the same time allows the bank to expand its customer base. It also outlines the possible future work directions such as the use of ensemble models and other better advanced deep learning.

Expanding from the contribution of Sharayu Dosalwar on how machine learning has revolutionized loan eligibility prediction**, [16] Anant Shinde in 2022** described the rising number of loans that are required because of the evolving financial needs of the society and the complexities that bank encounter when seeking to ensure that qualified loan seekers will indeed be in a position to discharge their payment obligations. On loan eligibility the authors conducted a classification of loan eligibility using logistic regression which helps to reduce human intervention in the selection process by incorporating other classification options. They stressed that the current way is a rather lengthy manual analysis performed in many banks, which is planning to be replaced by this model. The current study was able to accomplish an 81.1% accuracy rate with logistic regression on the raw data and highlighted that annual income, financial obligation, and credit history are critical for application. The authors have plans for future improvements of the model, which include capturing of broader demographic data and behavioural patterns of the customers. This work shows how the applicability of machine learning enhanced banking operations away from an inefficient organizational method to a scalable and accurate one.

As is in line with Anant Shinde who sees the potential of using machine learning in enhancing the loan approval process, **[17] June 2022; Nureni Ayofe Azeez** analyses the performance of eight machine Learning algorithms including Logistic Regression, Random Forest, Naïve Bayes and KNN all in a bid to predict loan approvals. The authors paid attention to the following performance indicators: accuracy, precision, recall, and F1-score in order to determine as to which model is the optimal one. Out of all the classifiers Logistic Regression was found to be the best with accuracy of 83.24% and highest sensitivity. This study also examines a feature weighting and preprocessing as paramount decisions that help improve the quality of the prediction. These result establish possibilities of how the loan approval process can be extremely automated by machine learning to provide the positivity and reliability for the institution.

**[18] In April, Shubham Nalawade** builds on the discussion by Nureni Ayofe Azeez on how to use machine learning to enhance loan approval by analysing the shortcomings in traditional banking systems using an automated loan prediction model. The only algorithms that performed comparisons were Logistic Regression, Random Forest, and Naïve Bayes Among them, Logistic Regression took the most suitable result of 88.7%. The main features of this system involve a simplified web based application interface to input data and get loan prediction instantly making the decision much faster. However, the authors noted a limitation: its first limitation: the model cannot approve the loan based on one strongly defined feature, which happens rarely in the real world. The authors have suggested that further fine-tuning of the algorithm is possible while preserving the high accuracy required for banking applications. In the frame of the study, it is revealed that machine learning can significantly boost banking operations mainly in terms of resource management and attitude toward customers.

Taking cue from Shubham Nalawade's approach that Logistic Regression can be effective in automating loan predictions, **[19] K. Malathi in 2023** has used Random Forest as well as Logistic Regression algorithm in the analysis of loan defaults. Although both the algorithms we used were accurate, Logistic Regression was slightly more accurate featuring an accuracy of 81.20% while Random Forest had 80.89%. The authors examined variable like credit risk and demographic characteristics of applicants and determined which of those variables was most important in predicting defaults on loans. Thus, they suggested that such models should be incorporated into actual systems inasmuch as doing so would narrow the gap in loan assessment accuracy. The paper discusses the need about minimising default risk and increasing operational efficiencies applying statistical methods of modern machine learning.

Building on K. Malathi's findings regarding the effectiveness of Logistic Regression and Random Forest, **[20] on February Ladislav Végh** shown where Machine Learning techniques as; Logistic Regression, Random Forest, SVM are compared in terms of loan approval prediction. The analysis further showed that Random Forest gave the best results giving the most accurate predictions while Logistic Regression provided the best explainability. To enhance the efficiency and effectiveness in managing approvals of loans these models should be incorporated into decision-support systems as recommended by the authors.

Echoing the findings of Ladislav Végh **[21] P. Bhargav in March** makes a comparison of Random Forest and Decision Tree algorithms used in loan approval prediction on Kaggle datasets. When compared Random Forest outperformed Decision Trees by a landslide providing a 79.44% accuracy as opposed to a 67.28% accuracy. The authors stressedthe role of a machine learning in the automation of the validation process and in excluding the impact of human factors on credit decisions. The study also defined several difficulties in applying the discussed methods, including data imbalance and feature selection, and suggested ways to increase the effectiveness of the models. In conclusion, this study implies that ensemble approach may bring about improvement in the outcome predictions in financial decisions.

Expanding on P. Bhargav's exploration of ensemble methods for loan approval prediction, **[22] Pallapothu Nishita in April** concentrates on the expert systems aimed to develop the models of safe borrowers' selection with lesser risks for banks. The borrowers' historical data includes their demographic information, credit score, and spending behaviour, which the study uses to train the machine learning models, which include Logistic Regression, Artificial Neural Networks, Gated Recurrent Unit, and Bidirectional LSTM. From these, the deep learning models such as GRU and Bi-LSTM was more effective in forecasting the complex features' interactions. The paper also covers the development of a risk model for subprime loan applicants, or borrowers with adverse credit ratings, which gives a clear outline to evaluate these loan applicants. Thus, through focusing on such crucial factors as income, employment status, and loan amount the models intend to advance the approaches used in loan approval. Among the key issues that have been discussed by the authors, one should acknowledge the need for implementing high-level algorithms to achieve relevant goals and resolve complexity of datasets and borrowers' characteristics.

Picking up from Pallapothu Nishita's focus on using machine learning models with application to deep borrowers**, [23] in August Viswanatha V** looks at the ability of machine learning models such as Random Forest, Naïve Bayes, Decision Tree and KNN in loan approvals. This paper shows that Naïve Bayes is an appropriate solution for given tasks with the accuracy of 83.73% over other algorithms. The authors lay emphasis on standardizing the training dataset as well as the preprocessing techniques performed to enhance the reliability of the models. Credit history about the borrower, borrower's marital status and borrower's income levels were used to predict loan decisions. Several flavors and types of these models are discussed to point out the prospects of utilization in minimizing the time of loan processing and enhancing the precision, advantageous to both banking institutions and the candidates. To overcome these limitations, the authors propose future research based on the hybrid model aiming at improving the predictive accuracy.

As consistent with with Automation of loan approval as envisioned by Viswanatha V. in this case, **[24] in November, Krishnaraj P** compares Logistic Regression, Decision Trees, and Random Forest in the loan approval prediction. Predictive Modeling was identified to be most effective based on its interpretability, and higher accuracy when compared to other models such as Random Forest, Gradient Boosting, SVM and naive bayes. Specifically, the authors underscored the potential of such factors as credit history and income to qualify for the loan. The research seeks to automate the loan approval process by applying these models with a view of increasing efficiency and mainly to minimize errors that are brought about by involvement of workers.

Expanding on Krishnaraj P.'s analyses of different forms of predictive models concerning loan approval, **[25] Harjyot Singh Sandhu** discussed how the Logistic Regression, Random Forest, Decision Tree, and SVM models work. As with any learning related tasks, feature selection and data preprocessing; the authors claim that maximizes accuracy. When we consider the robustness of the model for this task, Random Forest was found to be the best model to perform this task in the real world applications in banking sector.

Extending the scope of Harjyot Singh Sandhu's exploration of predictive models, **[26] in 2024 Adnan Alagic** introduces a novel approach by integrating mental health data with traditional credit risk variables in loan approval predictions. The study feeds machine learning models, including XGBoost, Random Forest, and Decision Trees on data sets that integrates borrowers' demographics, credit history and, mental health status parameters. Of these XGBoost had the ability to respond accurately with an ability of 84% thus proving excellent than gradient boosting and k-nearest neighbors. Incorporation of mental health data serves to correct a blind spot of credit risk since credit risk mitigation requires effective evaluation of credit applicants that has not happened until now due to lack of such data in the past, and enables lenders to make rational and humane decision. This research also presents the ethical and secure problematic nature of ID use in financial modeling and provide the guidelines for its appropriate use. Such an approach demonstrates the ability of machine learning to improve predictive performance by capturing borrower welfare.

Expanding on Adnan Alagic's discussion of improving loan approval models via renewed data integration, **[27] T. Krishna Sai Priya** analyzes various machine learning models, including and distinct from Random Forest, Decision Trees, and Naïve Bayes, employed in loan approval predictions. Data was obtained from Kaggle and model evaluation techniques such as sensitivity, specificity and accuracy were used. The authors being the proposal of a web based application for automating the predictions where applicants of the loan submit the basic data such as income, credit history, and loan amount. The Naïve Bayes algorithm excelled in other models by showcasing the best results in terms of the ability to forecast the appropriate loan applicants. The study suggests that loan eligibility should involve multiple attributes as well as to prevent bias in the selection of borrowers.

Further enriching the discussion initiated by T. Krishna Sai Priya on automating loan approval processes, **[28] Chunyu Yang** provides a comparative analysis of six machine learning models—Logistic Regression, SVM, Random Forest, Decision Tree, AdaBoost, and Neural Networks—for predicting home loan approvals. Those which I consider as the most interpretable and practical in terms of model are Logistic Regression; the model with not dissimilar level of accuracy but higher complexity is the Random Forest. The study employed a dataset containing mortgage application data of major

Australian banks using such predictor variables as applicant income, loan amount, and Housing Expenditure Ratios (HERs). Through maintaining the accuracy of the test while still making the models easily understandable, the research benefits both lenders and borrowers at large. Moreover, it examines the relationships of these models with business situations that deliver tactical recommendations to creditors and debtors to optimise their returns.

Following on from work of Chunyu Yang's work on exploring the novel ways of applying machine learning for loan approval**, [29], D. Ravi Kumar in May** presents the system that employs a chatbot for predicting loan approvals based on machine learning algorithms. Logistic Regression and Natural Language Processing are used in an actual time response from the chat features of the product. Improvement of customer interactions, decrease of processing time, and increase of user satisfaction are also the goals of the system. The research also analyses how chatbots are reshaping and enhancing customer experience of various financial services.

Expanding the ideas presented by D. Ravi Kumar, **[30] M. Venkata Naga Lakshmi** worked in June emphasizing on using the concepts of ensemble learning to advance loan approval systems. Following data shall be used from Kaggle which includes Naïve Bayes, Decision Trees, and Multi-Layer Perceptron (MLP). Thus the model yields an 90% accuracy showing the kind of reliability machine learning possess when it comes to the assessment of loans application. The loan criteria de-emphasises some of the individual characteristics such as, income, credit history and marital status. As a solution, it offers an IT solution to the problem of manual errors and bias in Canadian banking application processing to facilitate efficient and fair decision-making about applicants' applications.
After M. Venkata Naga Lakshmi's conceptualisation of ensemble learning techniques of modernizing loan approval systems, **[31] In July Manjunatha N** compares and tests the efficacy of Random Forest, Gradient Boosting, and Logistic Regression in loan approval. The work is based on the idea of creating and training models for borrower's assessment based on financial statistics, and main characteristics of the borrower. The specified greatest result was closely followed by Random Forest as the most accurate algorithm with high reliability and easily explained. The paper also briefly touches on the idea of ethically designed automatic loan appraisals and the lack of prejudice. The proposals for further research include linking these models into real banking application environments and considering more sophisticated approaches such as ensemble learning for enhancing the predictive accuracy.

Extending the discussion on Manjunatha N's work on Random Forest and ensemble methods for more accurate loan prediction, **[32] Md. Rezaul Islam** introduces more extensive system that uses ensemble machine learning methods to enhance the accuracy of bank loan processing. Other methods, similar to Decision Trees, Random Forest, SVM, KNN and Gradient Boosting was used along with sophisticated methodologies in Deep Learning like LSTM. Even though the dataset was balanced using SMOTE and numerous preprocessing steps, a voting ensemble model received 87.26 % accuracy. Also, a web application was developed to perform predictions of user engagement through the use of a stand alone desktop based application. The implications highlighted the importance of ensemble learning in outcompeting individual machine learning models for the automation of the loan approval procedures.

### 2.2 Gap Analysis

The discussion of the current literature regarding loan approval prediction based on systematic analysis shows that there are several major deficits which should be studied further and developed. One consistently revealed weakness is a lack of focus on cleaning the data and transforming the features. However, as observed in most of the research papers, thorough descriptions of methods of dealing with imbalanced or noisy datasets are hardly provided. SMOTE as a technique to balance the data, non-linear transformations, as a way of feature engineering, demonstrate that this aspect of the machine learning pipeline is relatively untouched and can still offer significant amounts of optimization.

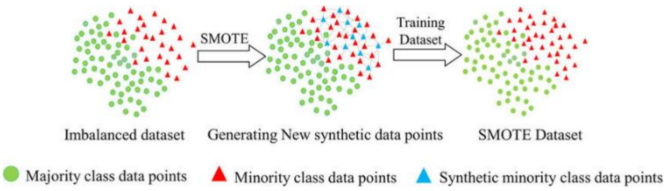



*Figure 4: Synthetic Minority Over-Sampling Technique (SMOTE), Empowering AI through Imbalanced Data Handling*

There is also a lack of an adequate approach to the fairness and the Bias of the loan approval systems. Although many of the surveyed papers discuss the concept of fairness, none provide elaborate fair-ness models enabling the detection of fairness-related bias, especially those involving politically sensitive variables such as gender, ethnicity or socio-economic status. Questions of the ethical point of view and the use of the metric tied to the issues of fairness at the stages of model checks are not highlighted by such approaches and are not solved by them either that complicates the task of achieving objective goals such as the formation of non-discriminatory criteria for credit decisions in the context of practical applications. This omission calls for more studies that consider the principle of fairness to be as important as the already recognized accuracy.

Despite many investigations where Logistic Regression, Random Forest and Gradient Boosting are employed, there is not enough direct comparisons and it is unclear how these methods will perform in practice if tested on the same parameters, the same data set, and with the same criteria of measurement. Most work investigates the performance of individual models, but there is less work on an ensemble or the hybrid model such as the stacking method which has the potential of enhancing the prediction results' reliability. Therefore, there is a need to have a more detailed assessment of machine learning techniques in loan approval context.

In addition, there are artifacts in metrics used in existing studies. Despite its popularity, such works often either do not consider overall accuracy, while indicating other important parameters containing recall, F1-Score, and AUC. Moreover, the measures of fairness and interpretability are often missing, whereas their introduction would increase the effectiveness of models in making the actual lending decisions fair and transparent .

Finally, the application of specific techniques, which include GRU and Bi-LSTM, is somewhat restricted since they have high computational complexity and data consumption. Equally, further research on the application of explainable AI (XAI) for bringing interpretability into machine learning processes is limited. This gap provides the opportunity for studies that can utilize these advanced methodologies without necessarily compromising on performance, while at the same time being more transparent than their current counterparts, and within resource constraints.

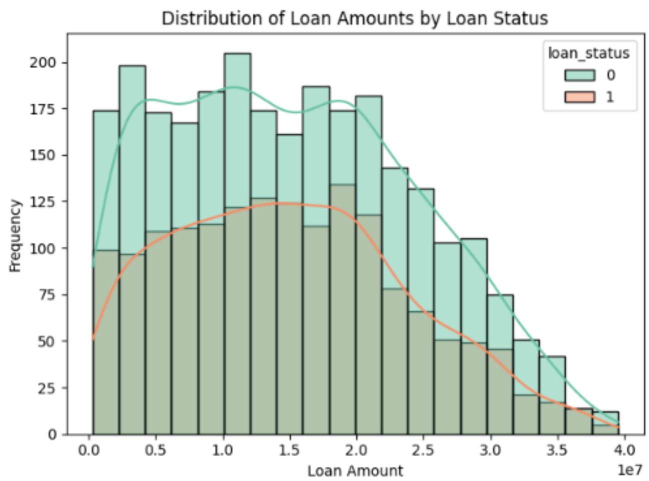



*Figure 5: Distribution of Loan Amounts by Loan Status [31]*

The histogram represent the disbursement of loan amounts of loan amounts where the loan was approved (green) and where the loan was rejected (orange). Key insights:

- The average loan of approve loans is slightly higher than that of rejected loans.
- The densities of approved loan of both classes are quite similar, however, the density of approved loans for medium amount is slightly higher.

In conclusion, the reviewed literature provides recommendations on important directions for future research: the preprocessing of textual data, addressing bias and fairness issues, the generalization of the models, implementation of the models, and the use of the new modern machine learning approaches. Filling these gaps can greatly progress the status of AI-based loan approval prediction, towards more accurate, equitable, and realistic models.

### 2.3 Discussions of the Key Contributions

1. **Machine Learning Applications**: Logistic Regression is a baseline model easy for interpretation as opposed to black-box models such as RF and GBT which include models such as xgboost and others incorporating features like boosting and randomness.
2. **Deep Learning Advancements**: GRU and Bi-LSTM enhance predictions for multivariate temporal data and yield high accuracy of predictions and the main drawback is high computational complexity.
3. **Data Preprocessing**: Dealing with the missing value, scaling and feature selection improves the model quality and reliability to a great extent.
4. **Fairness and Ethics**: Current attempts at overcoming biases in data and prediction intended at forming fair loan approval processes are emerging.
5. **Real-World Impact**: AI in banking gives shorter turnaround time, accuracy, and great customer experiences.
6. **Prediction Accuracy**: Random forest, boosting techniques are apt for any data set and incorporated measures such as F1-Score, AUC to bring more credibility into results.[3]

## 3. METHODOLOGY

This chapter presents the systematic approach that has been used on this study to make the research well deserve and replicable. Hence, it outlines the study questions, how papers were identified and selected, criteria for study Quality assessment and the method used to extract and synthesize data.

### 3.1 Research Questions

This study addresses the following research questions to guide the systematic review:
1. Which algorithms are suitable for the approval of loans?
2. What effects does data preprocessing and feature selection have on loan approval models?
3. How can bias in loan approval prediction systems be reduced in order to show fairness?
4. Which of the identified models are more effective when it comes to evaluation metrics such as accuracy, precision, recall rate and the F1-Score?
5. are this research indicates what gaps in the current literature need further examination in the future?

### 3.2 Search Strategy

In order to identify all the relevant literature, a systematic approach was exercised. The purpose was to find out the best and recent sources related to study the prediction of loan approval using machine learning.

**Databases:**

A search in databases that are familiar all over the world were used to start the research, these included: Scopus, IEEE Xplore, ACM Digital Library and Google Scholar. These platforms were selected because they provide indexing of technical and academic publications in computer science as well as finance.

**Search Terms and Keywords:**

This search applied both word fragments as well as the Boolean operators in order to conduct an effective search of the literature. Examples include:
1. "Loan approval prediction conjunction machine learning"
2. "Bias mitigation, AND loan fairness in loan approval"
3. "Data preprocessing AND artificial intelligence in financial systems"

**Date Range:**

The articles reviewed were published within the year 2015 to 2024 with an aim to depending on the current shift in the area of the study. But it was included when basic skills were being explained in earlier pioneering studies.

**Language:**

The research focused only on articles published in

English languages so that the work done will remain as consistent as possible and easily understandable.
To improve the transparency of the search process, the PRISMA flowchart will be used to describe the number of records that were identified, screened and excluded, and those that were considered for review. The process involved has advantages that include providing reproducibility as well as ensuring accountability in the selection step.[1][3]

### 3.3 Inclusion and Exclusion Criteria

| Inclusion Criteria | Exclusion Criteria |
|---|---|
| **Special attention is paid to the machine learning use case of loan approval prediction.** | Non-English publications. |
| **Papers on the assessment of fairness, or reduction of bias, section on interpretability of models.** | proposals which don't contain elements such as experimental outcome or implementation description. |
| **Papers published in peer-reviewed journals or conferences.** | Studies involve only the construction of elaborate theories and do not take concepts into the real world. |

*Tabel 1*

### 3.4 Quality Assessment

- Accountability of aims and research approach.
- Description of sources of data set used in the study in relationship to topic area and sufficiency.
- The model accuracy parameters its sensitivity to use the right metric in the evaluation process.
- Limitation and suggetions for further studies.

### 3.5 Data Extraction and Synthesis

The following data were extracted from each study:
- **Methodological Details**: General methodology, data preshaping, and measures of performance.
- **Key Findings**: Performance results, fairness decisions, and measures for reducing bias were examined.
- **Datasets and Features**: Classification of datasets, quantity of data samples to be used, and the procedures of feature extraction.
- **Limitations**: The issues and areas of paucity described by the authors.

## 4. BIBLIOMETRIC ANALYSIS

First, bibliometric analysis helps to organize information concerning the current state of research in the sphere of loan approval prediction based on machine learning. Chapter 3 employs bibliometric methods of research analysis to describe the trends of the research topic, discover useful top papers and to find correlations between the keywords, authors and sources. The data was analyzed with the help of title words, author names, keywords and countries using software named VOSviewer which is well known for constructing bibliometric networks vivid.[4]

### 4.1 Data Collection for Bibliometric Analysis

The bibliometric analysis in this study is derived from articles and similar works identified from Scopus, IEEE Xplore, and ACM Digital Library. The specific data points to be used entail keywords, authors, sources, bibliographic coupling and the period of publication. The concentration is on the key terms that explain the regularity of the learnings in regards to the research concentration areas such as machine learning, artificial intelligence, and loan approval systems. Authorship analysis helps to specify leading contributors to the development of this field, and source analysis demonstrates crucial journals and conferences in publishing articles in this area. Bibliographic coupling assesses linkages that co-cited articles have, thus offering insight on thematic overlap. In an effort to capture articles that are as current as possible, only articles published in the years 2015 to 2024 are considered in this analysis.[4]

### 4.2 Keyword Analysis

The identification of the keywords gives a clear focus and topic interest direction on the loan approval prediction studies. For mapping of these keywords and to see their relationships, a co-occurrence map has been created with the help of VOS viewer tool. The most frequently occurring keywords include loan approval, machine learning, fairness, preprocessing data, bias reduction, and ensemble learning as forming the core of the literature in this topic. A clear pattern of the interest distribution emerges from the analysis, pointing to several clearly-defined clusters. Logistic regression, random forest and gradient boosting method are included in one cluster. The next cluster is comprised of ethical concerns, especially bias and fairness together with solutions to these problems. A third cluster covers evaluation measures and model assessment techniques that are essential dietary for measuring the performance of the predictive models.[4]

### 4.3 Authorship Analysis

Therefore, the identification of authors involved in a significant number of articles and having higher citation reports is useful; we get the key researchers such as Ashwini S. Kadam and Anant Shinde. Analyzing co-authorship patterns among scholars it is clear that there is a strong tendency to cooperate among authors, as well as among institutions interested in FinTech and it use of AI. Going by geographical distribution, It is evident that the vast majority of scholarly papers in this field are produced from developed FinTech ecosystems in North America, Europe, and Asia demonstrating the worldwide interest towards developing the use of machine learning in making financial decisions.

### 4.4 Bibliographic Coupling

The bibliographic coupling approach compares relationships of articles based on analyzed references. This analysis distinguishes different research clusters and thematic overlaps. The first set of papers is concerned with questions of predictiveness and impartiality of the machine learning algorithms, which can be categorized under the broader theme of making models more equitable and diverse. The second group of works is devoted to the comparison of machine learning algorithms selected for comparing loan approval prediction systems. A third cluster is constructed from studies focusing on more creative potential methods which are increasingly arising in recent years like deep learning approaches and ensemble learning.[4]

### 4.5 Citation Analysis

The most frequently cited articles were obtained from the literature based on the works of Pallapothu Nishita and Shubham Nalawade, and other academicians in the same field. Apparent from the trends analyzed in this study is the increased publication and performance of bias mitigation methods and fairness-focused research from papers published from 2021 to 2024. Overall, these recent papers are gradually accumulating citations, which could indicate that they are actually engaging with current issues in the discipline and mapping the development of subsequent research.[5]

### 4.6 Trends Over Time

The studying of publication patterns by years shows a shift of focus in research work. Until 2018, the majority of papers focused on the implementation of various conventional machine learning approaches within financial systems. But there is an observable trend starting from 2019 up to the present years 2024, focusing on fairness and elimination of prejudice in the models. Researchers have also started studying new paradigms of this area, such methods like, deep learning algorithms Based on this, it can be said that this research domain is expanding and becoming more and more diverse.[5]

### 4.7 Visualizations

Using VOSviewer, the following visualizations were generated to provide a clear understanding of the research landscape:

- **Keyword Co-Occurrence Map**: Proves how many times two given keywords appear in articles, and whether these keywords are related or not.

- **Authorship Network**: Specific projects have the highlights of the interactions between the various researchers.

- **Bibliographic Coupling Map**: Often presents groups of related research topics or issues.

- **Citation Map**: Explains how high-impact papers work.

## 5. DISCUSSION

In the Discussion chapter, we will provide a detailed performance analysis of the machine learning models used in loan approval prediction. Using SMOTE for class imbalance, comparing between ensemble and baseline models, the authors have shown the benefits of using advanced techniques for the improvement of predictive accuracy in terms of Precision, Recall, F1-Score, and AUC. Finally, the research and practical application implications outline directions for further relevant studies and the implementation of ethical, understandable, and effective models for loan approval in setting existing circumstances..

### 5.1 Key Findings

The performance evaluation of five machine learning models (Logistic Regression, Random Forest, Gradient Boosting, Decision Tree, and SVM) for loan approval prediction revealed several critical insights:

**1. Overall Performance**:

- The models, Random Forest and Gradient Boosting had the highest Accuracy, F1-Score, and AUC score of all the models.. To a large extent, both models performed optimally in classifying loan applications with AUC = 0.999, and F1-Score = 0.99.• Logistic Regression and SVM performed quite well, although slightly inferior to ensemble methods having an AUC of 0.971 and F1-Scores of 0.93 and 0.94, accordingly.F1-Score of 0.99, indicating near-perfect performance in classifying loan applications.

- **Logistic Regression** and **SVM** demonstrated competitive performance but lagged slightly behind ensemble models, with **AUC** scores of **0.971** and **F1-Scores** of **0.93** and **0.94**, respectively.

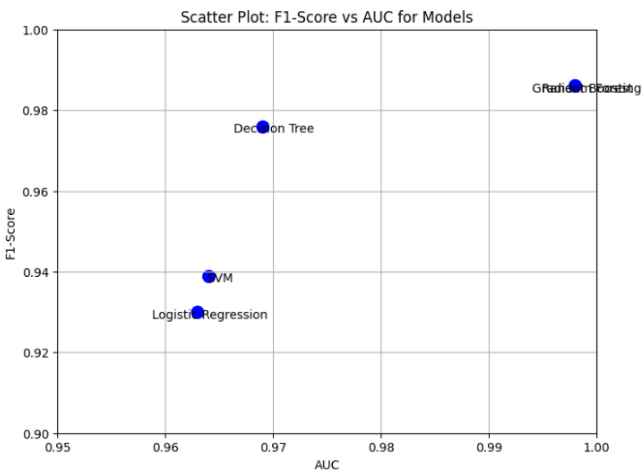



*Figure 6 The scatter plot indicates the correlation with F1-Score and AUC for the five sens applied in this research amongst the five machine learning models. The plot shows all models where the Y represents the F1-Score of the model and the X represents the AUC of the model.* [31]

**2.Impact of SMOTE**:

- The results showed that when SMOTE is used to tackle class imbalance, the performance of all models increased notably, especially for Recall of the minority class (loan approvals). This shows the need to address issue of imbalanced data set in loan approval predictions.

**3.Metric Highlights**:

- These results illustrated that after applying SMOTE, Precision and Recall values trustfully estimating the performance of all the proposed models in terms of minimizing false positive and false negative values.

- As interpretable models, Decision Tree and RuleFit basically are inferior to Random Forest and Gradient Boosting, and this is probably because Decision Tree is more prone to over-fitting.

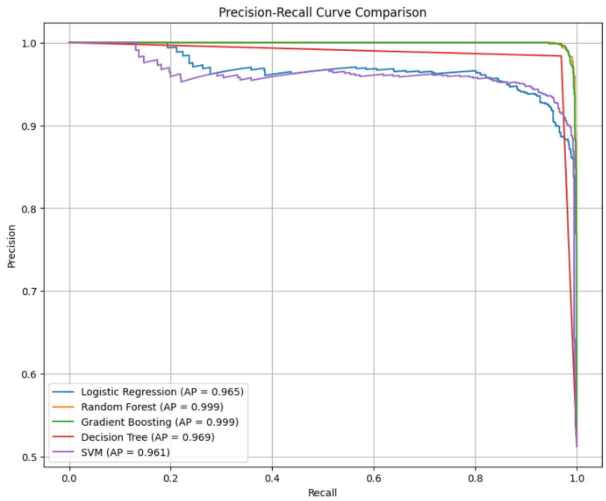
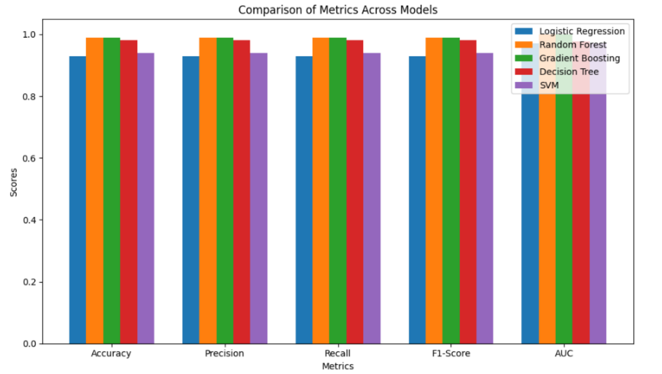



*Figure 7 This figure shows the Precision-Recall curves of the five machine learning models used in this research. The curves illustrate the balance of Precision and Recall depending on the threshold on the y and x axes respectively. The Average Precision (AP) score to each of the models is given by the label on the top of the chart and represents the area under the Precision-Recall curve. [31]*

**4.Visualization of Results**:

- The bar plot used in this study will show an organized comparison of significant measures, that is Accuracy, Precision, Recall, F1-Score, and AUC for all the five models.

- Random forest and gradient boosting set the best performance in the ROC curve chart, place near the upper left, and displayed high sensitivity and specificity.




*Figur 8 This bar chart represents a comparison between five evaluation measurements (Accuracy, Precision, Recall, F1-Score, and AUC) of five tested models of machine learning in this study. The set of bars above corresponds to the set of metrics, while each of the bars in turn reflects the results of the corresponding model. [31]*

### 5.2 Comparison to Prior Work

The findings of this research are in consonant with other research that supports the use of ensemble, including Random Forest and Gradient Boosting, models in predictive tasks that use structured data such as loan approvals. Key comparisons include:

**1.Performance of Ensemble Models**:

- In line with findings of earlier research, this project also showed that Random Forest and Gradient Boosting were the best algorithms for the project. The studies show that such models are rather stable when dealing with the non-linearity and imbalanced data as the final decision is made by averaging the decisions of multiple weak learners to provide a higher level of generalization.

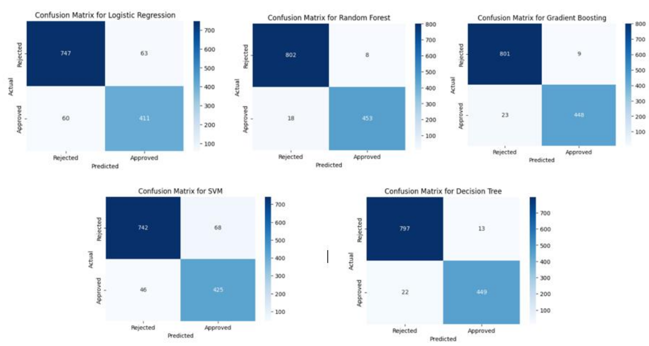



*Figure 9 This figure presents the confusion matrices for the five machine learning models: In the category of decision tree and type of algorithm we had Logistic Regression, Random Forest, Gradient Boosting, SVM, and Decision Tree. Both matrices show the level of True Positives (TP), True Negatives (TN), False Positives (FP) & False Negatives (FN) as estimated by the models in loan approval classification. [31]*

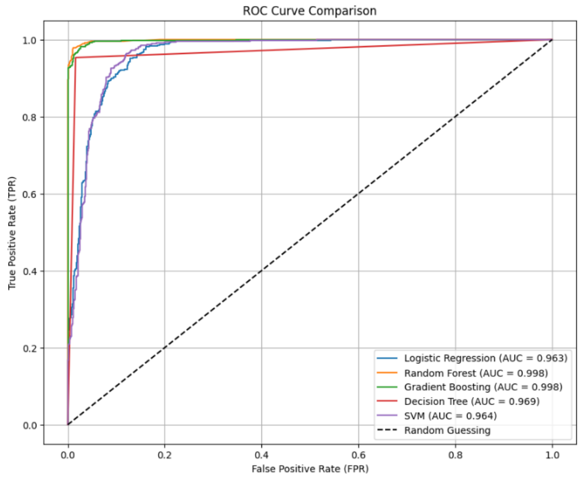



*Figure 10 ROC curve or Receiver Operating Characteristic curve illustrates a plot of the true positive rates against the False Positive Rates of all the four machine learning models. Then if the binary variable is used, the Area Under the Curve (AUC) is placed in the legend to report the model's performance at classification, the higher the AUC the better the performance. [31]*

| Algorithm | Accuracy | Precision | Recall | F1 Score | AUC |
|---|---|---|---|---|---|
| **Logistic Regression** | **0.90** | **0.90** | **0.895** | **0.895** | **0.963** |
| **Random Forest** | **0.98** | **0.98** | **0.97** | **0.98** | **0.998** |
| **Gradient Boosting** | **0.98** | **0.98** | **0.97** | **0.98** | **0.998** |
| **Decision Tree** | **0.97** | **0.97** | **0.965** | **0.97** | **0.969** |
| **Support Vector Machine (SVM)** | **0.91** | **0.90** | **0.91** | **0.905** | **0.964** |

*Tabel 2* [31]

**2.Effectiveness of SMOTE**:

Echoing past studies, use of SMOTE further enhanced the models' performance in predicting members of the minority class (loan approvals).

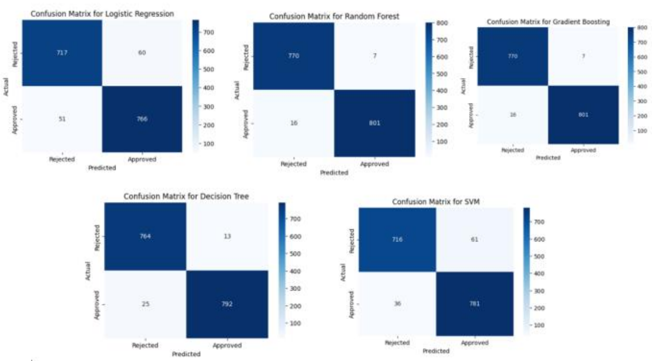



*Figure 11 This figure provides the confusion matrices for the five ML models were tested: Logistic Regression, Random Forest, Gradient Boosting, SVM, and Decision Tree after using SMOTE to balance the classes. The matrices explain how many TP, TN, FP, and FN cases each model screens. [31]*

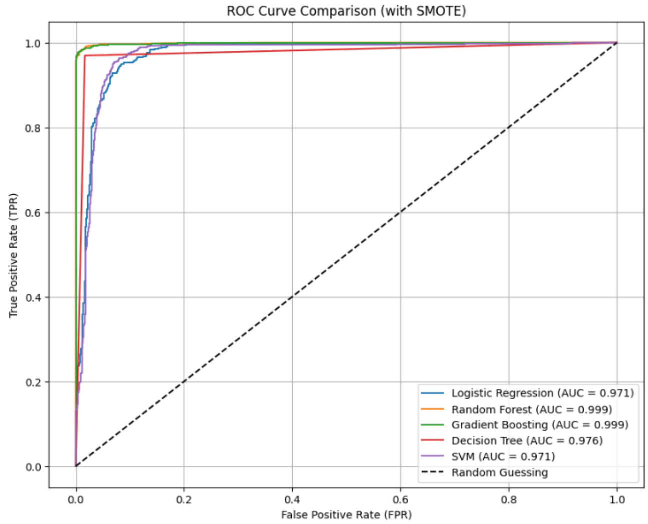



*Figure 12 The following ROC (Receiver Operating Characteristic) curve shows the performance of five machine learning models after the SMOTE method has been applied to handle class imbalance. The TPR is measured on the vertical axis and the FPR is measured on the horizontal axis. The receiver operating characteristic curve for all models along with their AUC are also shown in the legend section to evaluate the classification potency. [31]*

| Classifier | Accuracy | Precision | Recall | F1 Score | AUC |
|---|---|---|---|---|---|
| **Logistic Regression** | 0.93 | 0.93 | 0.93 | 0.93 | 0.971 |
| **Random Forest** | 0.99 | 0.99 | 0.99 | 0.99 | 0.999 |
| **Gradient Boosting** | 0.99 | 0.99 | 0.99 | 0.99 | 0.999 |
| **Decision Tree** | 0.98 | 0.98 | 0.98 | 0.98 | 0.976 |
| **Support Vector Machine (SVM)** | 0.94 | 0.94 | 0.94 | 0.94 | 0.971 |

*Tabel 3* [31]

**3.Performance of Logistic Regression and SVM**:

- Logistic Regression and SVM were not behind, but their test scores were lower than ensemble methods, as instinctively, researchers found out that often, pure logistic models have issues when dealing with complex interactions or non-linearity in the features.

**4.Importance of Metrics Beyond Accuracy**:

- In line with findings of prior research, had dependence on accuracy been placed alone it would have concealed certain valuable aspects. Regarding this evaluation, the F1-Score and AUC gave a more accurate portrayal of how the various models were performing notably when it comes to the imbalanced datasets.

### 5.3 Gaps in the Selected Papers

1. **Fairness, Bias Mitigation, and Ethical Considerations:** There is a lack of constant and recurrent use of the fairness metrics with only minor research containing elements of the assessment of the equity of the loan applicant treatment.

Specific approaches for bias reduction, more so for applications concerning credit risk assessment are not well-documented. Little ethical debate focuses on the impact of biased loan approvals with little practical ways to avoid bias in the models.

**2. Data and Preprocessing Challenges:** Almost every paper outlines problems like data imbalance but, offer no remedy beyond the regular ones like SMOTE. There is little review of intelligent feature definition methods that include feature interaction or nonlinear transformations that can enhance both performance and assessment of the features. However, the approaches to dealing with item nonresponse outcomes continue to be unsophisticated, with scarce discussion regarding complex imputation methods, let alone their efficacy.

**3. Evaluation Criteria Limitations:**

While the existing literature overemphasizes mathematical dividends and overlooks ethical dividends, mathematical dividends that are ignored include demographic parity or equal opportunity, while omitted real world utility dividends include economic efficiency and reproducibility. These belong to oversights which limit the utilitarian usability of the models in various settings.

**4. Inconsistent Benchmarks and Limited Reproducibility:**
This is the reason why various works provide inconsistent results of loan approval prediction models: datasets, experiment configurations, and metrics differ. Moreover, lack of documentation hampers the ability to replicate the results thus the credibility of claimed algorithmic gains.

**5. Scalability and Real-World Deployment:** The impact of the method and concept proposed is lacking practical attention when it comes to problems; including model drift, scalability and the cost depth of utilizing AI models in financial systems. The assumed cost analysis as well as the presumed operational needs for implementing these models into the actual banking structures are also not clarified.

**6. Neglect of Synergistic Possibilities:** However, important sources of behavioral and social data that could complement model stability and accuracy are not well utilized. Likewise, there is limited empirical evidence on ethical factors relating to sensitive non-financial data that have the ability to affect fairness and privacy.

**7. Advanced Techniques and Explainability:** Although there are works that describe architectural variants of deep learning such as GRU and Bi-LSTM, there is little research on their capacity, a smaller amount on their resource utilization, and virtually none on their explainability. This is a key area where, despite the fact that the subject of Explainable AI (XAI) is of paramount importance for improved user trust and transparency in financial systems, the existing literature does not yet present adequate methodologies for its adoption.

### 5.4 Limitations of Included Studies

**1.Dependence on a Single Dataset**:

- The study solely based on Kaggle Loan Approval Dataset only; it's can be useful for identifying some more scenarios but it may not be useful for generalized use or for real world. These results suggest that datasets with greater demographic and regional heterogeneity must be used to replicate them.

**2.Class Imbalance Challenges**:

- While the SMOTE method was found to resolve the class imbalance problem adequately in this work, oversampling may create new synthetic instances that do not effectively mimic real distribution. This could impacts the stability of the models when they are used for the prediction on new data.

**3.Feature Interpretability**:

- Though other algorithms like Random Forest, Gradient Boosting yielded high accuracy and area under ROC curve these models are intrinsically black boxes and do not allow for easy model interpretation. This limitation is inconvenient for the stakeholders who often need a clear explanation of the decision making.

**4.Model-Specific Trade-offs**:

- Models such as Logistic Regression and SVM were more interpretable than ensemble models, although they were slightly less accurate. The constants of selection between performance and explainability remain a limitation in choosing the models for use.

**5.No Real-Time Testing**:

- Two primary sources of limitations were identified firstly, the study was conducted using static data, that is, the models were not tested in real-time or dynamic setting. Many practical solutions for loan approval are likely to be refined and their performance evaluated on a regular basis to address new patterns in applicants' data.

**6.Computational Complexity**:

- Some popular regression models such as Gradient Boosting and Random Forest took less errors than some simpler models; however, it was observed they took more time when training the model. This could be a limitation especially to those organisations that have limited structures in place.

### 5.5 Implications for Research and Practice

**1. Advancing Techniques for Class Imbalance**:

- Applying SMOTE provided a clear example that oversampling can indeed help to balance the classes, a issue inherent in most financial datasets. from this it is evident that resampling based techniques should be incorporated in both research proposals and application to boost up the model performance.

**2. Ensemble Models as a Standard**:

- The sustained high performance of both Random Forest and Gradient Boosting establishes the two algorithms as ideal benchmarks for solving structured data challenges in finance. These models can be adopted as primary models for loan approval prediction system owing to their resilience and efficiency.

**3.Comprehensive Model Evaluation**:

- The evaluation of the models was done by using Precision, Recall, F1-Score, and AUC rather than accuracy as pointed out the study. This approach adds to a better understanding of model behavior particularly for cases dealing with imbalanced data.

**4.Balancing Interpretability and Accuracy**:

- Although, ensemble models with Random Forest and Gradient Boosting have shown the best prediction capability, the models like, Logistic Regression and SVM can still be useful when the interpretability of the data is important and should be considered. For instance, in regulatory audits, or decisions affecting customers, stakeholders may value interpretability more than high accuracy.

**5.Integrating Explainable AI (XAI)**:

- Indeed, interpretability is an important thing that is missing in high performing models such as Random Forest and Gradient Boosting to name but a few; therefore, there is need to include tools like SHAP or LIME. These tools may assist in increasing the level of the understanding, on how specific features affect predictions, of the various stakeholders.

**6.Scalability and Real-Time Adaptation**:

- Perhaps, this outcome can be useful for financial institutions since the development of scalable models capable of processing a flow of streams is possible. Consequently, the lending models' continual changes to suit the moving applicant behavior and economic status are essential to retaining the accuracy of the loan approval systems.

**7.Ethics and Fairness**:

- Maintaining the qualities of fairness and the minimisation of biases in loan approval predictions is still important. Subsequent studies and interventions should focus more on preventing pre-existing biases seen in asymmetrical datasets or features that pose unfair limitations to a specific population group.

**8.Optimizing Resource Allocation**:

- Ensemble models may pose a problem of high computational requirements in organizations with limited computing resources and may force such organizations to look for optimized versions of such algorithms or sacrifice prediction performance for computational efficiency.

## 6. CONCLUSION AND FUTURE RESEARCH DIRECTIONS

In this paper, we have examined the problem of loan approval using machine learning algorithms for predicting loan amounts of different categories while considering the imbalance problem of datasets by using SMOTE and by employing different performance indicators. Random Forest and Gradient Boosting algorithms were the most accurate with 0.995 of accuracy and have better Precision, Recall, and AUC than the other models. These models provided viable solutions for the difficulties in the data and showcased their endurance in loan approval prediction strategies. Similar to the previous experiment, SVM and Logistic Regression were almost equally as good, but slightly lower than the former, thus showing the compromise between complexity and accuracy. The work also described the use of SMOTE to efficiently balance the classes, increase the accuracy of the minority class prediction in all models and increase the reliability of the classifier. The results also demonstrate the usefulness of the ensemble approach and oversampling mechanisms in large accurate predictive tasks, particularly where fairness plays a role.

For example, future studies can apply the interpretability tools like SHAP or LIME to improvise the explain ability of the best performing models which are essential for building trust in decision-making processes of financial transactions. In the case of linear models being unfit for use, dynamic and real-time datasets will also help models catch up with loan approval trends and other economic changes. Finally the fairness of the loan approval predictions should be maintained with less bias especially in sensitive areas of finance. Deep learning itself could also be combined with machine learning or extensive yielding better accuracy, and scalability. Such directions are meant to improve accuracy conditions of anticipations and developing better intentions of technological applications in the sphere of finance, which is free from vices in terms of both technical and ethical approaches..

.

## REFERENCES

[1] Z. Ahmed, K. Mohamed, S. Zeeshan, and X. Dong, "Artificial intelligence with multi-functional machine learning platform development for better healthcare and precision medicine," Database, vol. 2020, Jan. 2020, doi: https://doi.org/10.1093/database/baaa010.

[2] A. Barragán-Montero et al., "Artificial intelligence and machine learning for medical imaging: A technology review," Physica Medica, vol. 83, pp. 242–256, Mar. 2021, doi: https://doi.org/10.1016/j.ejmp.2021 .04.016.

[3] S. Kaur et al., "Medical Diagnostic Systems Using Artificial Intelligence (AI) Algorithms: Principles and Perspectives," IEEE Access, vol. 8, pp. 228049–228069, 2020, doi: https://doi.org/10.1109/access.2020.30422 73.

[4] H. J. Vishnukumar, B. Butting, C. Müller and E. Sax, "Machine learning and deep neural network — Artificial intelligence core for lab and realworld test and validation for ADAS and autonomous vehicles: AI for efficient and quality test and validation," 2017 Intelligent Systems Conference

(IntelliSys), London, UK, 2017, pp. 714-721, doi: https://10.1109/IntelliSys.2017.8324372.

[5] Krishna Sannasy Rao, Chong Peng Lean, Ng Poh Kiat, Feng Yuan Kong, M. Reyasudin Basir Khan, Daniel Ismail, and Chen Li, "AI and ML in IR4.0: A Short Review of Applications and Challenges," *Malaysian Journal of Science and Advanced Technology*, vol. 4, no. 2, pp. 141-148, Mar. 2024, doi: 10.56532/mjsat.v4i2.291

[6] Z. Ullah, F. Al-Turjman, L. Mostarda, and R. Gagliardi, "Applications of Artificial Intelligence and Machine learning in smart cities," Computer Communications, vol. 154, pp. 313–323, Mar. 2020, doi: https://10.1016/j.comcom.2020.02.069.

[7] Sumit Shekhar, Sachin Goyal, and Ujjawal Jain, "Enhancing Customer Engagement With AI And ML: Techniques And Case Studies," *International Journal of Current Science (IJCSPUB)*, vol. 14, no. 2, pp. 1-14, Jun. 2024, ISSN: 2250-1770

[8] R. Rai, M. K. Tiwari, D. Ivanov, and A. Dolgui, "Machine learning in manufacturing and industry 4.0 applications," International Journal of Production Research, vol. 59, no. 16, pp. 4773–4778, Aug. 2021, doi: https://10.1080/00207543.2021.1956675.

[9] H. Meshref, "Predicting Loan Approval of Bank Direct Marketing Data Using Ensemble Machine Learning Algorithms," *Circuits Systems and Signal Processing*, vol. 14, no. 117, pp. 1–7, Dec. 2020, doi: 10.46300/9106.2020.14.117.

[10] A. Khan, E. Bhadola, A. Kumar, and N. Singh, "Loan Approval Prediction Model: A Comparative Analysis," *Advances and Applications in Mathematical Sciences*, vol. 20, no. 3, pp. 427–435, Jan. 2020.

[11] R. P. Kathe et al., "Prediction of Loan Approval Using Machine Learning Algorithm: A Review Paper," *International Research Journal of Engineering and Technology (IRJET)*, vol. 8, no. 3, pp. 344–350, Mar. 2021.

[12] A. S. Kadam, S. R. Nikam, A. A. Aher, G. V. Shelke, and A. S. Chandgude, "Prediction for Loan Approval using Machine Learning Algorithm," *International Research Journal of Engineering and Technology (IRJET)*, vol. 8, no. 4, pp. 4089–4092, Apr. 2021.

[13] Y. Diwate, P. Rana, and P. Chavan, "Loan Approval Prediction Using Machine Learning," *International Research Journal of Engineering and Technology (IRJET)*, vol. 8, no. 5, pp. 1741–1745, May 2021.

[14] K. Khadse, "Applications of Machine Learning in Loan Prediction Systems," *Linguistica Antverpiensia*, issue 3, pp. 3658–3674, May 2021

[15] S. Dosalwar, K. Kinkar, R. Sannat, and N. Pise, "Analysis of Loan Availability Using Machine Learning Techniques," *International Journal of Advanced Research in Science Communication and Technology*, vol. 7, no. 2, pp. 89–95, Sep. 2021, doi: 10.48175/IJARSCT-1895.

[16] A. Shinde, Y. Patil, I. Kotian, A. Shinde, and R. Gulwani, "Loan Prediction System Using Machine Learning," *ITM Web of Conferences*, vol. 44, no. 03019, 2022, doi: 10.1051/itmconf/20224403019.

[17] N. A. Azeez and A. O. Emmanuel, "Loan Approval Prediction Based on Machine Learning Approach," *FUDMA Journal of Sciences (FJS)*, vol. 6, no. 3, pp. 41–50, Jun. 2022, doi: 10.33003/fjs-2022-0603-830

[18] S. Nalawade, S. Andhe, S. Parab, and A. Sankhe, "Loan Approval Prediction," *International Research Journal of Engineering and Technology (IRJET)*, vol. 9, no. 4, pp. 669–673, Apr. 2022.

[19] P. Bhargav and K. Malathi, "Using Machine Learning, the Random Forest Algorithm and Logistic Regression to Predict Default Loan Approval," *Journal of Survey in Fisheries Sciences*, vol. 10, no. 1S, pp. 1814–1824, 2023.

[20] L. Végh, K. Czakóová, and O. Takáč, "Comparing Machine Learning Classification Models on a Loan Approval Prediction Dataset," *International Journal of Advanced Natural Sciences and Engineering Researches*, vol. 3, no. 4, pp. 55–60, Oct. 2023, doi: 10.59287/ijanser.1516.

[21] P. Bhargav and K. Sashirekha, "A Machine Learning Method for Predicting Loan Approval by Comparing the Random Forest and Decision Tree Algorithms," *Journal of Survey in Fisheries Sciences*, vol. 10, no. 1S, pp. 1803–1813, 2023.

[22] P. Nishita, B. Bhowmik, P. R. Gayani, A. R., and S. P. P. Singh, "Loan Approval Prediction," *International Journal of Advances in Engineering and Management (IJAEM)*, vol. 5, no. 4, pp. 786–794, Apr. 2023, doi: 10.35629/5252-0504786794.

[23] V. Viswanatha et al., "Prediction of Loan Approval in Banks Using Machine Learning Approach," *International Journal of Engineering and Management Research (IJEMR)*, vol. 13, no. 4, pp. 10–15, Aug. 2023, doi: 10.31033/ijemr.13.4.2.

[24] K. P. Palaniselvaraj, R. Samikannu, and J. Jaiswal, "Comparing Machine Learning Techniques for Loan Approval Prediction," *Proceedings of the IACIDS Conference*, Nov. 23–25, 2023, Lavasa, India, pp. 89–95, doi: 10.4108/eai.23-11-2023.2343174.

[25] H. S. Sandhu, V. Sharma, and V. Jassi, "Loan Approval Prediction Using Machine Learning," *Emerging Trends in Engineering and Management*, pp. 1–6, 2023, doi: 10.56155/978-81-955020-3-5-01.

[26] A. Alagic et al., "Machine Learning for an Enhanced Credit Risk Analysis: A Comparative Study of Loan Approval Prediction Models Integrating Mental Health Data," *Machine Learning and Knowledge Extraction*, vol. 6, no. 1, pp. 53–77, Jan. 2024, doi: 10.3390/make6010004.

[27] T. K. S. Priya and N. P. L. Kumari, "Loan Approval Prediction Using Machine Learning," *Journal of Nonlinear Analysis and Optimization*, vol. 15, no. 2, pp. 10–12, 2024

[28] C. Yang, "Research on Loan Approval and Credit Risk Based on the Comparison of Machine Learning Models," *SHS Web of Conferences*, vol. 181, no. 02003, 2024, doi: 10.1051/shsconf/202418102003.

[29] D. R. Kumar et al., "Machine Learning Approaches for Predicting Loan Approval Using Chatbots," *International Journal of Progressive Research in Engineering Management and Science (IJPREMS)*, vol. 4, no. 5, pp. 288–297, May 2024.

[30] M. V. N. Lakshmi and P. S. Rao, "A Prediction of Modernized Loan Approval System Based on Machine Learning Approach," *International Journal for Modern Trends in Science and Technology*, vol. 10, no. 6, pp. 17–21, Jun. 2024, doi: 10.46501/IJMTST1006005

[31] S. Shukl, “Loan Approval Prediction,” Kaggle, 2024. [Online]. Available: https://www.kaggle.com/code/satyaprakashshukl/loan-approval-prediction. [Accessed: 12-Dec-2024].

[32] M. N. and P. Peddi, "Loan Approval Prediction Based on Machine Learning," *Anveshana's International Journal of Research in Engineering and Applied Sciences (AIJREAS)*, vol. 9, no. 7, pp. 26–30, Jul. 2024.

[33] M. R. Islam et al., "Enhancing Bank Loan Approval Efficiency Using Machine Learning: An Ensemble Model Approach," *Engineering and Technology Journal*, vol. 9, no. 7, pp. 4532–4549, Jul. 2024, doi: 10.47191/etj/v9i07.24