***ERURJ* 2025, 4, 1, 2271-2289**
Review Article

A Data Driven Model for Predicting Loan Approval Using Machine Learning Approaches

Shady Abdelhady [1], Reham Abdallah [2]Anton Magdy Fawzy [3], Mazen Alaa Mohamed [4], Yasser Salaheldin [5, *]

[1,2] *Assistant Lecturer, Business Technology Department, Faculty of Management, Economics and Business Technology, Egyptian Russian University, Cairo, Egypt.*
[3,4] *Under-graduate student, Fourth level, Business Technology Department, Faculty of Management, Economics and Business Technology, Egyptian Russian University, Cairo, Egypt.*
[5] *AI Assistant Professor, Artificial Intelligence Department, Faculty of Computers and Information, Sadat Academy for Management Science, Cairo, Egypt.*

*Corresponding author(s): Yasser Salaheldin Kamel, E-mail: yasser-salah@eru.edu.eg

Received 29th July 2024 Revised 9th December 2024 Accepted 30th January 2025

DOI:10.21608/erurj.2025.291674.1149

ABSTRACT

Predicting loan approval is an essential task for banks and financial institutions, as it entails assessing the risk and profitability of lending money to potential borrowers. Loan approval processes in financial institutions are often complex and time-consuming, relying on manual assessments that can be biased and inconsistent. These difficulties make it more difficult to properly evaluate risks and guarantee ethical lending practices. To address these issues, this research proposes an approach rooted in machine learning to anticipate the approval status of loans for applicants based on their personal and financial characteristics. The research depends on a dataset consisting of 5000 loan applicants with 14 attributes, acquired from Kaggle. We assess the model using three classification techniques such as decision tree, k-nearest neighbors (KNN), and support vector machine (SVM). Metrics like accuracy, precision, recall, and F1-score were used to assess the performance of the models. The SVM model demonstrates an accuracy of 96%, while the decision tree achieves 92% and KNN attains 86%. According to these findings, we conclude that using the SVM model is a reliable and effective method for predicting loan approval status.

Keywords: Loan Approval, Machine Learning, Decision Tree, Neural Network, Classification

1. Introduction

Credit scoring systems are essential for lending institutions as they aim to determine the probability default for clients and adhere to a low-loss principle for sustainability. These systems support credit applications, manage the risk's credit, and significantly influence non-performing loans, which can ultimately result in bankruptcy, financial disasters, and impact environmental sustainability. (Végh et al., 2023) (Kannan, et al., 2023).

Over the past decade, there has been a shift from credit scoring models based on credit officers or experts to more technologically advanced methods. This shift has driven the demand for automated credit decision-making systems that can minimize lost opportunities and credit losses, thereby reducing potential risks for lending institutions. Consequently, automated credit scoring has gained immense importance, particularly due to the rise of financial services that operate without human participation. A significant event is the recent creation of South Korea's first internet-only banking firm (Prabaljeet, et al., 2023). Essentially, the utilization of technology and automation to streamline operations and decrease costs in contemporary lending institutions underscores the need for developing accurate credit scoring models.

Traditional approaches for evaluating credit risk typically rely on credit scores and reports to estimate the probability of potential losses. However, these reports frequently do not support comprehensive information about the borrower's creditworthiness. Several indicators, including financial factors, demographic data, and consumer behavior such as transaction history and spending patterns, significantly influence credit risk assessments. To manage the complexity of evaluating multiple factors and offer more thorough and scalable assessments (Chang, Park, 2018).

While constructing an efficient model to assess clients' creditworthiness is challenging, Machine Learning (ML) now plays a crucial role in credit scoring applications. ML algorithms can analyze huge datasets and determine significant patterns, leading to more accurate predictions and enhanced credit-scoring models (AC, et al., 2023). (Prabaljeet, et al,2023).

Banks and other financial organizations can significantly enhance their lending operations by utilizing machine learning techniques to develop models that predict loan approval. By accurately assessing creditworthiness and predicting loan approval outcomes, the model can help mitigate risks, reduce default rates, and enhance the overall efficiency of the lending method.

This paper's main objective is to create a prediction model that can accurately determine the loan approval status of applicants using their personal and financial attributes. By leveraging machine learning algorithms, we aim to automate and optimize the loan approval process for banks and financial institutions. The specific goals of this study include:

– Collecting a comprehensive dataset of loan applicants with relevant personal and financial attributes.
– Applying machine learning algorithms to train and then test predictive models for loan approval.
– - Evaluating the models' performance using appropriate metrics such as F1-score, accuracy, precision, and recall.
– Comparing how well various machine learning techniques including decision trees and neural networks predict loan approval status.
– Providing insights and recommendations based on the findings to improve loan approval procedures and reduce hazards for lending institutions.

The paper is organized as follows: In the introduction part, the importance of creating a machine learning-based predictive model for loan approval is outlined, along with the challenges faced by financial organizations and banks. The section also establishes the objectives and significance of the study. The following literature review section offers a thorough examination of existing research in the field. The methodology section describes the framework and approach used in developing the predictive model, detailing the steps taken during training and testing. Additionally, this section provides information about the dataset used. The results section presents the findings, analyzing and interpreting them to offer insights into the effectiveness of the predictive model for predicting loan approval. The conclusion section highlights the contributions of the predictive model in improving the loan approval process and discusses the study's limitations, as well as suggesting future research areas. This structure ensures a coherent flow of information, progressing from the introduction to the literature analysis and methods, results, and conclusion, providing a comprehensive understanding of the study's objectives, methods, and outcomes.

2. Literature Review

Dansana, Debabrata, et al. (2024). This paper provides a detailed comparative analysis, demonstrating how selected machine learning techniques can differentiate between regular and high-risk loan applicants unlikely to repay their debts. Testing results reveal that XGBoost achieved 84% accuracy on the initial dataset, exceeding KNN (83%), as well as gradient boost (83%). Regarding the additional dataset, with an accuracy of 85%, random forest is the most accurate method, followed by decision trees and KNN at 83%. In addition to accuracy, the precision, recall, and overall performance of these algorithms were evaluated. A confusion matrix analysis was conducted, yielding numerical results that highlight the superior performance of XGBoost and random forest in the classification tasks of the first dataset, and XGBoost and decision tree in the second dataset. These results offer significant explanations. for researchers and practitioners in selecting models to improve their classification efforts in terms of accuracy and precision.

Muhammad, Iqbal, et al.2024. This Loan Approval Prediction dataset has 4,269 items and 14 characteristics. Based on a comprehensive analysis of the 14 variables, the most crucial factor is the loan amount, whereas the bank asset worth has the least influence. Different Decision Tree ensemble models were also used to examine the dataset. Considering the results, the accuracy was 0.9974, the precision was 0.9969, the AUC was 0.9998, and the F1 Score was 0.9966. the comparison analysis shows that the XGBoost model performs better than the others.

Yang, Chunyu, 2024. This paper presented the suggested models for the issue and provided instructions on how to apply the chosen model. LR, DT, RF, SVM, Ada Boost, and neural networks are the six potential models that have been chosen. When comparing the interpretability and accuracy of these models, logistic regression performed the best. Test accuracy rates are highest for logistic regression (0.76) and second highest for SVM (0.73). Neural networks have an accuracy rating of 0.63, making them the least accurate model.

Uphade, et al.2024. The research aims to develop a prediction model that may be utilized to precisely identify possible defaulters. To evaluate the likelihood of a loan default in situations with balanced and unbalanced data, research uses a wide range of machine learning techniques, such as LR, DT, RF, k-NN, SVM, XG Boost, AdaBoost, and Gradient Boosting Machines. The study's results show that RF algorithm performed the best on imbalanced datasets, with a remarkable accuracy score of 0.91. SVM and logistic regression showed similar effectiveness, with corresponding accuracy ratings of 0.91 and 0.90. With a perfect accuracy score of 1.00 in balanced datasets, the RF model unexpectedly performed better than other models.

Debabrata Dansana, et al. (2024). This study aims to assess whether loans should be approved for certain people or entities. An evaluation of performance and identification of possible borrowers for loan approval was done using the Random Forest Regressor (RFR) model. The findings suggest that banks must not solely focus on wealthy customers but should also examine more important client attributes that are important in determining credit approval and loan default prediction.

Nallakaruppan, M. K., et al.2024. The proposed project aims to create a driving application providing justifications for loan denials. This application incorporates a Random Forest with an XAI (Explainable Artificial Intelligence) framework to elucidate the reasons behind both the approval or denial of a loan application. The Random Forest-based method achieved the highest accuracy, sensitivity, and specificity, with scores of 0.998, 0.998, and 0.997 respectively. Additionally, the LIME and SHAPLEY Translators are used to offer insights using both regional and worldwide surrogate models that explain the influence of various factors on the features.

Végh et al. (2023). The paper also evaluates two deep learning algorithms, the Long Short-Term Memory Network (LSTM) and the Deep Neural Network (DNN). for analyzing loan approval. The dataset was initially made available on Kaggle in July of 2023 and consists of 4269 instances with Nine category and three numerical characteristics, including the target variable. The findings indicate that the Deep Neural Network (DNN) outscored all other machine learning and deep learning algorithms in tests, with the best F1-score, accuracy, precision, and recall. Both validation and test data showed that the best accuracy, exceeding 98%, was achieved using neural networks and ensemble classification models. One of the limitations of the study is that it did not address the problems of the data before entering it into the model, nor did it clarify the importance of the features that affect the dependent variable.

Ndayisenga and Theoneste (2023). In this study, the researchers aimed to enhance commercial banks' ability to predict borrower behavior. They developed and tested various frameworks with Bank of Kigali data, the largest financial institution in Rwanda by assets and customer base (Mpaka, 2019). The dataset included over 58,096 observations and 29 variables related to loan repayment. A training set (70%) and a test set (30%) were created from the data. The team employed ensemble methods, integrating various machine learning methodologies to determine the optimal strategy for predicting bank loan defaults. Their analysis showed that Gradient Boosting yielded the best results in predicting loan defaults, followed by XGBoost, while other models like decision trees, random forest, and logistic regression performed less effectively.

Kannan, MK Jayanthi.et al. (2023). This study describes a loan approval prediction system that uses a variety of machine learning techniques to calculate the probability of loan approval based on a borrower's credit history. The study assesses several algorithms, including LR, DT, RF, KNN, ANN, Naive Bayes, Adaboost, and Voting Classifier, for their effectiveness in loan prediction. The findings indicate strong performance, with accuracy rates of 86% for Logistic Regression, 74% for Decision Tree, 86% for Support Vector Machine (RBF), and 86% for Naive Bayes. However, the study had some significant shortcomings. First, it does not mention data preprocessing issues or efforts to improve data quality before feeding it into the models. Second, it lacks a detailed analysis or explanation of the significance of the features impacting the dependent variable.

Viswanatha V, Ramachandra A.C,et al .(2023). To ascertain the probability of accepting individual loan requests, this study proposes the combination of ensemble learning approaches with ML models. Four different algorithms were utilized: RF, NB, DT, and KNN. Through this approach, the research achieved an improved accuracy rate of 83.73%, with the Naive Bayes algorithm performing as the most effective.

Uddin, Nazim, et al.2023. This comprehensive study involves preprocessing the data, using SMOTE to balance the data effectively, and deploying multiple machine learning models, including Decision Trees, Support Vector Machines, K-Nearest Neighbors, Gaussian Naive Bayes, AdaBoost, Gradient Boosting, Logistic Regression, and Advanced Deep Learning Models like Recurrent Neural Networks, Deep Neural Networks, and Long Short-Term Memory Models. With an impressive accuracy of 87.26%, the results show that the voting-based ensemble model performs better than other state-of-the-art methods as well as individual ML models like Extra Trees.

Nalawade, Shubham, et al.2022. They compared the accuracy of different machine learning algorithms. We got a percentage of accuracy ranging from 75-85% but the best accuracy we got was from Logistic Regression i.e., 88.70%. The system includes a user interface web application where the user can enter the details required for the model to predict. The drawback of this model is that it takes into consideration many attributes but in real life sometimes the loan application can also be approved on a single strong attribute, which will not be possible using this system.

Nureni et al (2022). The models were trained using eight distinct algorithms: K-means, RF, DT, LR, SVM, NB, and KNN. The findings showed that the frameworks yielded a range of outputs. Based on the findings from both datasets, Random Forest and Naive Bayes came in second and third, respectively, with accuracy levels of 82.16% and 77.34%, and Logistic Regression at 83.24% and 78.13%.

3. Materials and proposed model

3.1 aim of research

The objective of the research is to improve loan approval processes and reduce the threat of default and financial loss. This will be achieved by creating a dependable predictive model that aids lenders in making well-informed decisions and provides them with a reliable tool for evaluating loan applications. The model will leverage historical loan data, encompassing customer information, financial attributes, and loan-specific features, to achieve its objectives.
This research carries significant weight in reducing the necessity of human intervention in the decision-making procedure with regards to the approval of loans.

3.2 Materials

In this study, the materials used for defining data sources regarding loan approval are crucial for developing a comprehensive and accurate predictive model. The primary data sources include historical loan data obtained from financial institutions, encompassing a wide range of attributes and variables relevant to loan approval. These data sources comprise customer information such as demographics, credit history, employment details, income, and financial statements.
Additionally, loan-specific features, such as loan amount, interest rate, loan purpose, loan term, and collateral information, are considered. By leveraging these materials and data sources, The research seeks to identify the various factors influencing loan approval decisions and accurately predict the risk associated with loan applications.

Table 1 data description

No.	Variables	Description
1	ID	Customer ID
2	Age	Customer's age in completed years
3	Experience	#years of professional experience
4	Income	Annual income of the customer ($000)
5	Zip Code	Home Address ZIP code.
6	Family	Family size of the customer
7	CCAvg	Avg. spending on credit cards per month ($000)
8	Education	Education Level. 1: Undergrad; 2: Graduate; 3: Advanced/Professional
9	Mortgage	Value of house mortgage if any. ($000)
10	Personal Loan	Did this customer accept the personal loan offered in the last campaign?
11	Securities Account	Does the customer have a securities account with the bank?
12	CD Account	Does the customer have a certificate of deposit (CD) account with the bank?
13	Online	Does the customer use Internet banking facilities?
14	CreditCard	Does the customer use a credit card issued by UniversalBank?

3.3 Proposed model

The machine learning approach proposed in Figure 2 is designed to address the uncertainty in postgraduate admission. The framework consists of five distinct steps: (1) Data exploration phase, (2) Data preprocessing phase to prepare the data for applying the EDM models, (3) Data splitting into training and testing sets, (4) Training of ML models, and (5) Evaluation phase to assess the accuracy of the constructed model. Each of these steps will be thoroughly discussed and explained below.

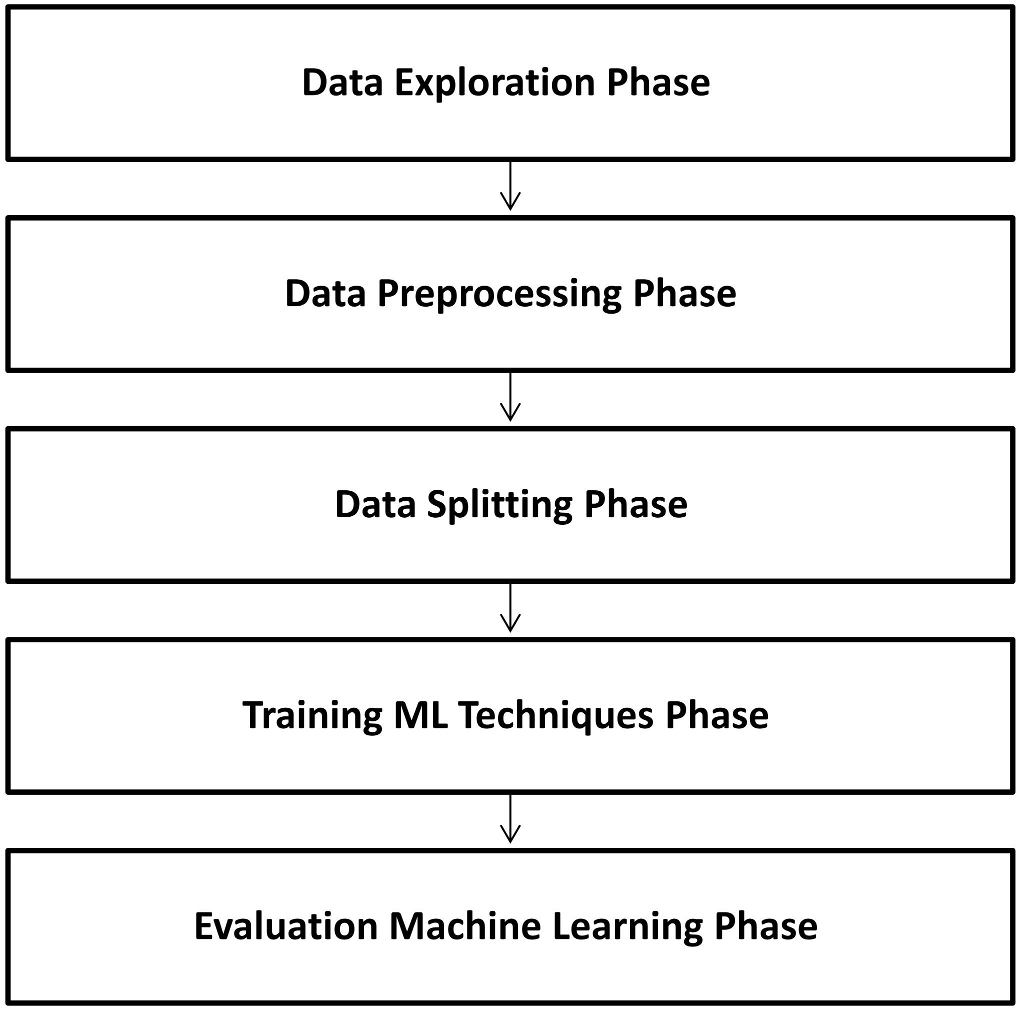


Figure 1 machine learning pipeline

3.3.1 Data exploration phase

In the data exploration phase of predicting loan approval, the objective is to thoroughly analyze the loan dataset and gain insights into its characteristics. This phase is vital for comprehending the data before developing a predictive model using machine learning techniques. During this phase, the loan dataset is obtained and examined for any problems with data quality, such as missing values or outliers. Descriptive statistics are calculated to summarize the numerical variables, while the frequency of different categories is determined for the categorical variables.

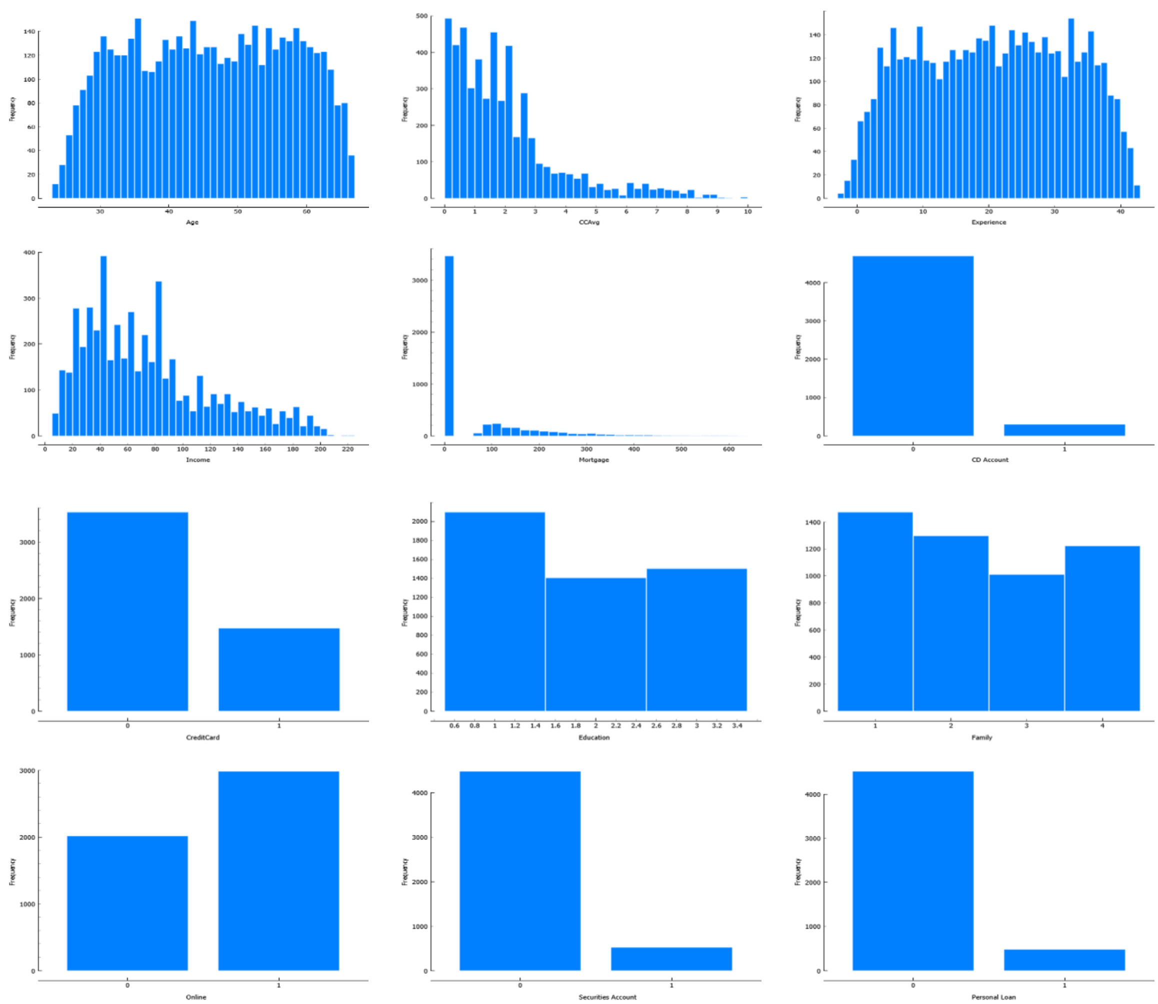


Figure 2 distribution of variables (cont.)

The dataset for predicting loan approval consists of 5000 rows and 14 columns. It comprises 5 categorical features and 9 numeric features. The dataset is clean, with no missing data, which eliminates the need for imputation or data cleaning in this regard. Based on the data exploration,

Name	Distribution	Mean	Mode	Median	Dispersion	Min.	Max.	Missing
Age		45.34	35	45	0.25	23	67	0 (0 %)
Experience		20.10	32	20	0.57	-3	43	0 (0 %)
Income		73.77	44	64	0.62	8	224	0 (0 %)
ZIP Code		93152.50	94720	93437	0.02	9307	96651	0 (0 %)
Family		2.40	1	2	0.48	1	4	0 (0 %)
CCAvg		1.93794	0.3	1.5	0.901723	0.00	10	0 (0 %)
Education		1.88	1	2	0.45	1	3	0 (0 %)
Mortgage		56.50	0	0	1.80	0	635	0 (0 %)
Personal Loan			0		0.316			0 (0 %)
Securities Account			0		0.335			0 (0 %)
CD Account			0		0.228			0 (0 %)
Online			1		0.674			0 (0 %)
CreditCard			0		0.606			0 (0 %)

Figure 3 decriptive statistics of variables

we noticed that the experience variable contains negative values (52 value), which are meaningless noise in the data. These negative values might be a result of data inconsistencies or errors in the available dataset. The presence of negative values in this variable is unusual and can have a negative impact on data analysis and building a loan approval prediction model. It is important to handle these values carefully during the data exploration phase. We can take actions such as excluding the negative values or replacing them with missing values (e.g., using the absolute value of those negative values) so that they do not affect our analysis and modeling. It is necessary to document this finding and take appropriate steps to deal with these unusual values in subsequent stages of the modeling process.

3.3.2 Data preprocessing phase

During the data preprocessing phase, multiple actions are taken to transform and prepare the dataset for machine learning. A key step involves converting numerical representations for categorical parameters. One-hot encoding is a popular technique used for this purpose. It involves creating binary columns corresponding to each category within a categorical variable. Each column represents a specific category, and a value of 1 is assigned if the observation belongs to that category, while 0 is assigned otherwise.

Additionally, numerical features may require discretization to convert continuous values into discrete intervals or categories. This process helps simplify the data and can be achieved using techniques like binning or quantile-based discretization.

Furthermore, the experience variable contains noise or invalid values. To handle this, noisy data points can be removed from the dataset to ensure data integrity. then estimate values in the experience variable, imputation methods can be utilized to estimate and fill in the missing values. By performing these data preprocessing steps, the dataset is transformed into a suitable format, categorical variables are encoded numerically, noisy data is removed, and missing values are imputed, making the dataset ready for further analysis and model development.

3.3.4 Data splitting phase

In this study, we employed the k-fold method (with k=10) to partition the data. The k-fold cross-validation approach involves the random division of the dataset into two parts: the Training dataset and the Testing dataset The dataset is divided into 'k' sub-samples using this technique; k-1 sub-samples are used for training and one sub-sample is kept away for testing. The procedure of k-fold cross-validation is carried out k times, with the training and testing datasets being adjusted accordingly (Yadav,2016). As depicted in Figure 5, this methodology was implemented.

3.3.5 Training ML techniques phase

The training phase of classification machine learning algorithms plays a pivotal role in constructing models capable of accurately classifying data points into predefined categories. This phase involves the utilization of labeled training data to enable the algorithm to learn and generalize patterns and relationships within the data. In our research, we employed a combination

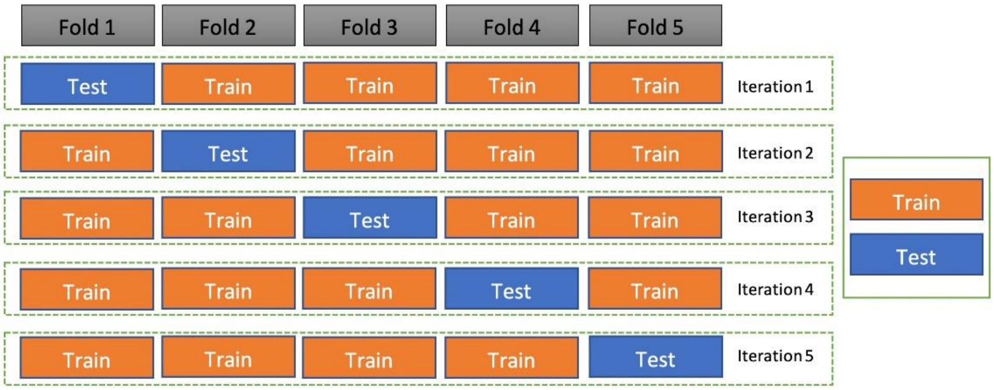


Figure 4 k-fold valudation method

of classification machine learning algorithms, including Decision Trees (DT), K-Nearest Neighbors (KNN), Support Vector Machines (SVM), and Logistic Regression (LR). Each algorithm offers unique characteristics and advantages, allowing us to explore different approaches to classification tasks and evaluate their performance.

- **Decision Tree**

A decision tree is a framework that uses a tree-like structure to progressively split the dataset into smaller subsets. This division is guided by the characteristic that offers the best informative value for the specific task being addressed. Inside nodes of the tree represent decisions that depend on specific features, whereas the leaf nodes correspond to classes that align with the most appropriate target values. In certain instances, leaf nodes may also hold probability vectors, which indicate the likelihood of various target attribute values. Decision trees prove to be particularly advantageous in scenarios where the correlation between the input and output variable is intricate and nonlinear, such as credit risk predictions [Javed,2015]. Their easily interpretable structure enables stakeholders to visualize and comprehend the decision-making process, making them useful for explanations. To increase decision trees' robustness and performance, they can also be combined with other ML strategies, such as ensemble methods. It is important to note that decision-makers often favor simpler decision trees because they are easier to understand. Research indicates that the complexity of a decision tree significantly affects its accuracy. [Breiman,1984]. Hence, finding the right balance between complexity and accuracy is crucial when utilizing decision trees in practical applications.

- **KNN**

One popular classification method in several domains, including image recognition, speech recognition, and natural language processing, is the K-nearest neighbors (KNN) algorithm [Jena, 2021]. Classifying a new data point according to the classes of its K-nearest neighbors in the training dataset is the basic idea behind the K-nearest neighbors (KNN) technique. KNN operates on the fundamental premise that objects in close proximity display comparable prediction values. As a result, KNN calculates the distance between each point in the training dataset and a new data point using metrics based on distance, such as the Manhattan, Euclidean, and Chebyshev distances.

- **SVM**

In machine learning, Support Vector Machines (SVM) are a reliable classification algorithm that are frequently utilized. The goal is to find the best hyperplane in the feature space that maximally divides data points belonging to various classes. With the use of the kernel technique, SVMs are skilled at handling datasets that are both linearly and non-linearly separable. The capacity of SVM to handle complicated decision boundaries is a major benefit in classification problems. SVM can convert the data into a higher-dimensional space where it becomes linearly separable by using different kernel functions, such as polynomial kernels or the radial basis function (RBF). Because of this characteristic, SVM is useful in situations where a simple linear boundary cannot simply divide the data into relationships and non-linear patterns (Li, 2006; et al., 2008).

3.3.6 Evaluation machine learning phase

A research paper's evaluation section is essential for determining a model or algorithm's efficacy and performance. It entails a careful examination and contrast of the model's predictions with reference or ground truth data. In this paper, we employ the confusion matrix, a commonly utilized evaluation method. A thorough and comprehensive depiction of a model's predictions is provided by the confusion matrix. The true positives, true negatives, false positives, and false negatives are shown in tabular form. The count or percentage of accurate and inaccurate predictions the model made for each class is represented by each element in the matrix. Accuracy, precision, and recall are just a few of the performance indicators that may be obtained by examining the confusion matrix. These measurements provide insightful information about how predictive.

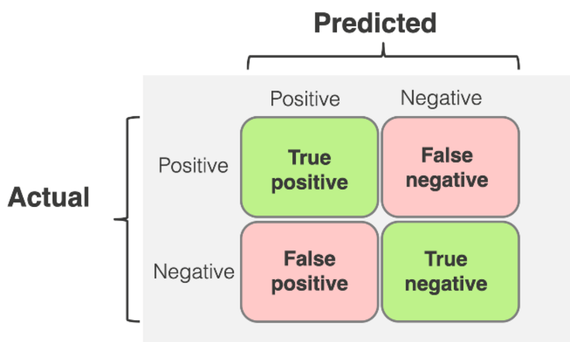
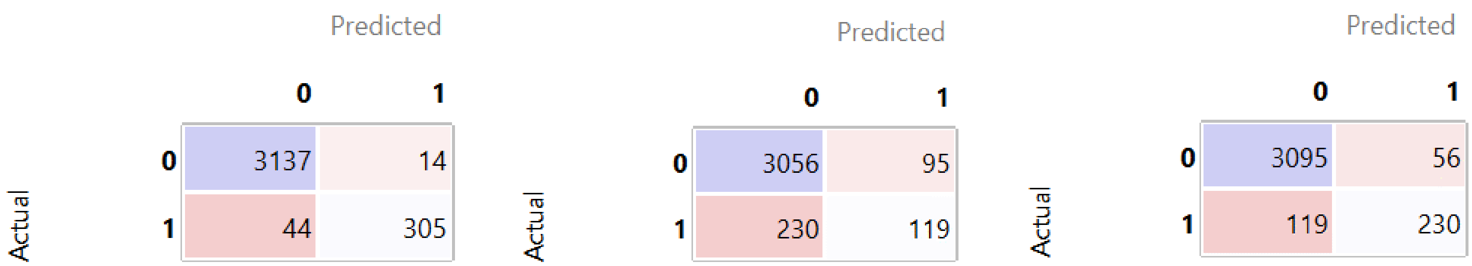


Figure 5 confusion matrix



(a) Decision tree (b) KNN (c) SVM

Figure 6 confusion matrix of ml algorithms

The outcomes' discussion of the machine learning models, namely Support Vector Machines (SVM), k-nearest Neighbors (KNN), and Decision Trees (DT), reveals interesting findings. The SVM model revealed an accuracy of 96%, compared to the other two models. The KNN model's accuracy was 86%, while the Decision Tree (DT) model's accuracy was 92%. The SVM model successfully distinguished between the classes and produced correct predictions, as evidenced by its high accuracy. SVMs are recognized for their capability to capture non-linear correlations in the data and manage complex decision boundaries. This adaptability likely played a significant role in its superior performance for this task.

However, with an accuracy of 86%, the KNN model also demonstrated remarkable performance. KNN is an easy-to-understand algorithm that groups data points according to how near they are

Model	AUC	CA	F1	Prec	Recall	MCC
Tree (1)	0.924	0.983	0.983	0.983	0.983	0.905
kNN (1)	0.860	0.907	0.897	0.893	0.907	0.389
SVM (1)	0.964	0.950	0.948	0.947	0.950	0.701

Figure 7 machine learning algorithms performance.

to the k nearest neighbors. While it may not have achieved the same level of accuracy as SVM, its performance is still noteworthy. The DT model performed well in the classification task, as seen by its 92% accuracy rate. DT are highly valued for their interpretability and the ability to capture complex decision-making processes. The model's accuracy suggests that it successfully learned the underlying patterns in the data and made accurate predictions.

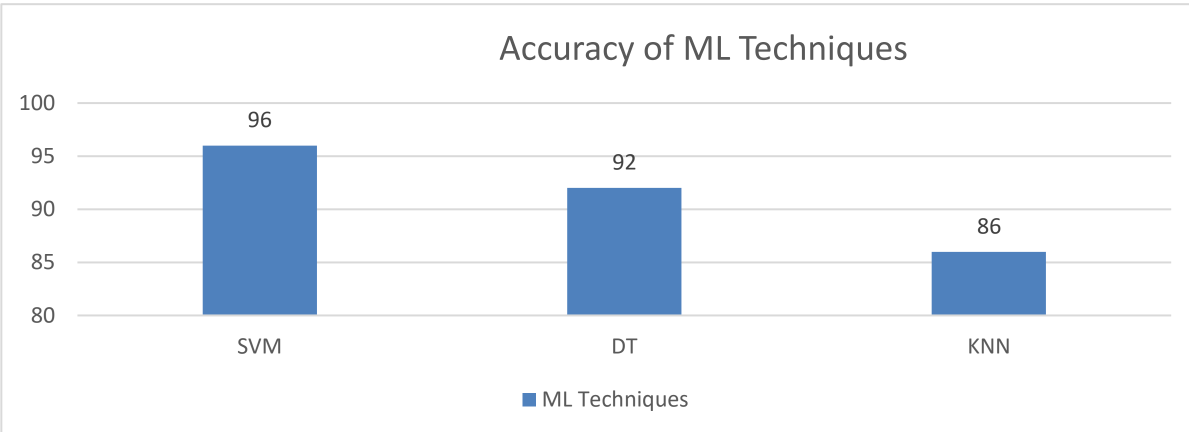


Figure 8 Comparsion between ML techniques

4. Conclusion

This study addresses important issues including bias, inefficiency, and inconsistency in conventional human processes by demonstrating the performance of machine learning models in forecasting loan approval statuses. Three classification methods decision tree, k-nearest neighbors (KNN), and support vector machine (SVM) were applied and assessed using a dataset of 5000 applicants which had 14 attributes. The experimental results highlight the superior performance of the SVM model, achieving an accuracy of 96%, surpassing the decision tree (92%) and KNN (86%) models. The SVM model's higher accuracy, combined with its precision, recall, and F1-score metrics, indicates its reliability and robustness for practical applications. These results demonstrate that machine learning can significantly improve the efficiency and fairness of loan approval processes, allowing financial institutions to make more data-driven, unbiased decisions. Overall, the adoption of such predictive models offers financial institutions a valuable tool to streamline their workflows, minimize risks, and enhance customer satisfaction. Future research can focus on expanding the dataset, incorporating additional features, and exploring advanced machine learning techniques to further improve predictive accuracy and scalability.

5. References:

[1] AC, Ramachandra, and Vishwas KN. "Prediction of Loan Approval in Banks Using Machine Learning Approach." (2023).

[2] architsharma01-loan-approval-prediction-dataset, https://www.kaggle.com/datasets/architsharma01/loan-approval-prediction-dataset

[3] Bhagat, Abhishek. Predicting Loan Defaults using Machine Learning Techniques. Diss. California State University, Northridge, 2018.

[4] Breiman, L., et al. "Construction of trees from a learning sample." Classification and Regression Trees, 1st ed.; Taylor & Francis Group: Abingdon, UK (1984): 21-23.

[5] Chang, Hyejung, and Mikyung Park. "A smart e-form for effective business communication in the financial industry." Business Communication Research and Practice 1.2 (2018): 95-101.

[6] Dansana, Debabrata, et al. "Analyzing the impact of loan features on bank loan prediction using Random Forest algorithm." Engineering Reports 6.2 (2024): e12707.

[7] Javed, K.; Hamid, F. A comparative study of decision tree algorithms for nonlinear and complex relationships between input features and output variables. Int. J. Adv. Res. Comput. Sci. Softw. Eng. 2015, 29, 65–74.

[8] Jena, Bhagyalaxmi, Anita Mohanty, and Subrat Kumar Mohanty. "Gender recognition of speech signal using knn and svm." (2020).

[9] Kannan, MK Jayanthi, and A. R. Nithej. "ML Based Loan Approval Prediction System A Novel Approach."

[10] Karthikeyan, S. M., and Pushpa Ravikumar. "A Comparative Analysis of Feature Selection for Loan Prediction Model." International Journal of Computer Applications 975: 8887.

[11] Kennedy, Kenneth. "Credit scoring using machine learning." (2013).

[12] Khadse, Kavita. "Applications of machine learning in loan prediction systems." Linguistica Antverpiensia 3 (2020): 3658-3674. with Data Analysis" JCR. 2021: 372-382

[13] Kumar, Ashwani, Raman Dugyala, and Pronaya Bhattacharya. "Prediction of Loan Scoring Strategies Using Deep Learning Algorithm for Banking System." Innovations in Information and Communication Technologies (IICT-2020) Proceedings of International Conference on ICRIHE-2020, Delhi, India: IICT-2020. Cham: Springer International Publishing, 2021.

[14] Li, Sheng-Tun, Weissor Shiue, and Meng-Huah Huang. "The evaluation of consumer loans using support vector machines." Expert Systems with Applications 30.4 (2006): 772-782.

[15] Luthfi, Emha Taufiq, and Ferry Wahyu Wibowo. "Loan Payment Prediction Using Adaptive Neuro Fuzzy Inference System." International Journal of Simulation: Systems, Science & Technology (2017).

[16] Mamun, M. A., Afia Farjana, and Muntasir Mamun. "Predicting Bank Loan Eligibility Using Machine Learning Models and Comparison Analysis." Proceedings of the 7th North American International Conference on Industrial Engineering and Operations Management, Orlando, FL, USA. 2022.

[17] Manglani, Richa, and Anuja Bokhare. "Logistic Regression Model for Loan Prediction: A Machine Learning Approach." 2021 Emerging Trends in Industry 4.0 (ETI 4.0). IEEE, 2021.

[18] Muhammad, Iqbal, et al. "Performance Analysis of Ensemble Learning and Feature Selection Methods in Loan Approval Prediction at Banks." Journal of Artificial Intelligence and Engineering Applications (JAIEA) 3.2 (2024): 557-564.

[19] Nallakaruppan, M. K., et al. "An Explainable AI framework for credit evaluation and analysis." Applied Soft Computing 153 (2024): 111307.

[20] Ndayisenga, Theoneste. Bank loan approval prediction using machine learning techniques. Diss. 2021

[21] Nishita, Pallapothu, et al. "Loan Approval Prediction."

[22] Nureni, Azeez A., and O. E. Adekola. "LOAN APPROVAL PREDICTION BASED ON MACHINE LEARNING APPROACH." FUDMA JOURNAL OF SCIENCES 6.3 (2022): 41-50.

[23] Nalawade, Shubham, et al. "Loan Approval Prediction." International Research Journal of Engineering and Technology 9.04 (2022): 669-673.

[24] Patel, Harshita, et al. "A New Fuzzy Adaptive Algorithm to Classify Imbalanced Data." Computers, Materials & Continua 70.1 (2022).

[25] Patlannagari Hasitha Reddy, Nusrath Mohammad, "Machine Learning for Loan Prediction Dataset with Data Analysis" JCR. 2021: 372-382

[26] Rahman, Arifa Tur, Md Rakib Hasan Purno, and Samiya Akter Mim. "Prediction of the Approval of Bank Loans Using Various Machine Learning Algorithms." 2023 IEEE World Conference on Applied Intelligence and Computing (AIC). IEEE, 2023.

[27] Rath, Golak Bihari, Debasish Das, and BiswaRanjan Acharya. "Modern approach for loan sanctioning in banks using machine learning." Advances in Machine Learning and Computational Intelligence: Proceedings of ICMLCI 2019. Springer Singapore, 2021.

[28] Saini, Prabaljeet Singh, Atush Bhatnagar, and Lekha Rani. "Loan Approval Prediction using Machine Learning: A Comparative Analysis of Classification Algorithms." 2023 3rd International Conference on Advance Computing and Innovative Technologies in Engineering (ICACITE). IEEE, 2023.

[29] Sarkar, A. "Machine learning techniques for recognizing the loan eligibility." International Research Journal of Modernization in Engineering Technology and Science 3.12 (2021): 1135-1142.

[30] Sharma, Hitesh K., et al. "Machine Learning based model for Loan Amount Prediction and Distribution.

[31] Sivasree, M. S., and T. Rekha Sunny. "Loan credibility prediction system based on decision tree algorithm." Int. J. Eng. Res. Technol 4.09 (2015): 825.

[32] Uphade, D. B., A. A. Muley, and S. V. Chalwadi. "Identification of Most Preferable Machine Learning Technique for Prediction of Bank Loan Defaulters." Indian Journal of Science and Technology 17.4 (2024): 343-351.

[33] Uddin, Nazim, et al. "An ensemble machine learning based bank loan approval predictions system with a smart application." International Journal of Cognitive Computing in Engineering 4 (2023): 327-339.

[34] Vashist, Shail Vatsal, and Aman Sharma. "Loan Approval Predictor using Data Science and Machine Learning." (2020).

[35] Végh, Ladislav, Krisztina Czakóová, and Ondrej Takáč. "Comparing Machine Learning Classification Models on a Loan Approval Prediction Dataset." (2023).

[36] Yadav, Sanjay, and Sanyam Shukla. "Analysis of k-fold cross-validation over hold-out validation on colossal datasets for quality classification." 2016 IEEE 6th International conference on advanced computing (IACC). IEEE, 2016.

[37] Yang, Chunyu. "Research on loan approval and credit risk based on the comparison of Machine learning models." SHS Web of Conferences. Vol. 181. EDP Sciences, 2024.

[38] Yasaswini, Polana, et al. "ANALYSIS AND FORECASTING OF BANK LOAN APPROVAL DATA USING MACHINE LEARNING ALGORITHMS.