FastFunds: A machine learning-driven personal loan approval prediction mobile App

Samuel Akwasi Danso
Isaac Osei Nyantakyi
Dept. of Computer Engineering
Ghana Communication Technology University, Accra, Ghana

Justice Odoom
School of Computer Science and Technology
Southwest University of Science and Technology, Mianyang, China

Emmanuel Attah Kumasi Amponsah
Dept. of Computer Engineering
Ghana Communication Technology University, Accra, Ghana

Diana Danso
Principal Superitendent -Ghana Education Services- Kumasi,Ghana

Keywords
Machine Learning, Loan Approval, KNN, SVM, Mobile App, NFC.

Abstract

The loan approval processes in financial services require precision and efficiency to minimize risks and ensure customer satisfaction. This study addresses the challenge of accurately predicting loan approvals by developing a mobile application that leverages machine learning algorithms. Six machine learning models—Logistic Regression, Decision Tree, Random Forest, Support Vector Machine, K-Nearest Neighbors, and Feed Forward Neural Network—were evaluated using key metrics such as accuracy, precision, recall, specificity, F1-score, geometric mean, and balanced accuracy. Among these, the Feed Forward Neural Network demonstrated the highest accuracy at 94.72%, followed by the Support Vector Machine and KNearest Neighbors, which also showed strong performance across multiple metrics. This research highlights how machine learning strategies can significantly improve the accuracy of loan approval predictions on mobile platforms. The mobile application developed through this work aims to enhance the operational efficiency, accessibility, and user engagement of financial institutions.

Introduction

The identification of risks is impeded by the interconnected financial accounts of individuals, organizations, banks, and governments [6]. Many debtors are failing to repay their loans, which is causing financial organizations to experience significant difficulties. This problem impedes the capacity of these institutions to assist individuals and enterprises and damages their finances. Loan approval is an essential procedure for banks, as they heavily depend on loans. Conversely, it is more challenging to evaluate loan applications and mitigate loan default risks. Machine learning algorithms are capable of effectively managing vast quantities of data intended for loan

approval forecasting. Deborah et al. (2023). The precision and efficacy of loan approval processes can be enhanced by the rapid expansion of machine learning (ML). Machine learning algorithms have the capacity to analyze extensive data sets to identify patterns and insights that may be overlooked by humans. Loan defaults can be accurately predicted through comprehensive borrower data analysis, which enhances decision-making. Risk management and client interaction for financial institutions could be revolutionized by a mobile application for personal loan approval that employs machine learning algorithms.

In [7], the World Bank indicates that approximately 1.7 billion low- and middle-income adults are unable to access formal financial services. This scenario is being transformed by mobile technology, which provides new opportunities for financial inclusion. For example, in Kenya, mobile money platforms like M-Pesa have led to a significant increase in financial inclusion, with 65% of the population using mobile money by 2019, compared to just 2% in 2012.

In recent years, the financial industry has increasingly adopted machine learning (ML) techniques to handle large datasets and make predictive decisions in real-time. However, despite these advancements, many financial institutions still struggle with loan default rates due to inefficiencies in the traditional approval process. Mobile technology presents an opportunity to revolutionize loan processing by enabling faster and more accurate predictions through advanced machine learning algorithms.

Literature review

Leveraging machine learning and deep learning techniques to forecast loan defaults, researchers [1] utilized a dataset from the Lending Club [8-16] and applied several models, including Decision Trees, Random Forest, Logistic Regression, KNearest Neighbors, and a deep learning model (Feedforward Neural Network). The main strength of the study lies in its extensive use of various machine learning and deep learning models, providing a comprehensive analysis of multiple methodologies to identify the most effective approach for forecasting loan defaults. The dataset was preprocessed using efficient techniques, which involved handling missing values and categorical features, both of which are crucial for improving model performance. However, the exclusive use of the Lending Club dataset raises concerns about the generalizability of the model's performance. Its applicability to other datasets or real-world scenarios with different characteristics remains uncertain.

The authors in [2] conducted a study comparing logistic regression and random forest models for predicting loan approvals. Statistical techniques, such as confidence intervals and power analysis, were used, with results showing logistic regression had a higher accuracy (81.30%) than random forest (75.6%). The primary advantage of the study lies in its comparative methodology, which analyses the variations in performance between two widely used algorithms in loan prediction, providing a more comprehensive understanding of algorithmic performance. Utilizing statistical techniques such as confidence intervals, p-values, and power analysis enhances the credibility and scientific rigor of the findings. However, the study seems to have a limited reach, as it solely analyses two models. Integrating a wider range of machine learning or deep learning models could enhance the overall comprehension of the capabilities and constraints of predictive analytics in the field of finance. The study is based on a sample size of " n=10," which may not accurately represent the data. Increased and more varied datasets can enhance the reliable extrapolation of the results. The report does not specify if the models' parameters were altered or adjusted, which can significantly affect their performance.

The authors in [3] aimed to improve financial institution decision-making processes using machine learning-driven loan approval prediction. The study employed preprocessing techniques to cleanse the dataset and utilized various classification algorithms to predict the probability of loan approval. The correctness of the model is assessed by the utilization of KFold cross-validation and baseline modelling. Although the paper presents impressive accuracy rates, it fails to address the performance of these algorithms on different datasets or in varying economic circumstances. The report could provide more details regarding the specific features utilized and the process by which they were derived from raw data. The study primarily examines traditional machine learning models, although more exploration of deep learning methods could yield superior results.

The authors in [4] focused on detecting fraudulent loan applications through six machine learning models: decision tree, KNN, SVM, random forest, AdaBoost, and logistic regression. KNN had the best performance, achieving 83.75% accuracy. The primary advantage of the study is in its evaluation of different machine learning algorithms to determine the most efficient method for identifying fraudulent loan applications. This methodology allows for a comprehensive analysis of the various approaches that can be used to achieve this objective. The paper does not describe how the models were evaluated beyond the initial accuracy test. For example, cross-validation, precision-recall analysis, and ROC curves would provide a more in-depth understanding of each model's performance and reliability. While the study discusses" Analytics Vidhya" data collection, there is no information about the dataset's diversity and representativeness. Details on the magnitude, variety, and completeness of the data would aid in determining the strength of the study's conclusions. Although the study analyses six models, its main focus is on accuracy. Other metrics, such as F1-score, precision, and recall, are also important for a

thorough study, particularly in a problem domain like fraud detection, where false positives and false negatives can be quite costly.

The authors in [5] applied SVM and logistic regression to predict loan status. The study used data from Kaggle, which included client demographics, and aimed to help financial institutions make more informed loan approval decisions. Several studies have explored machine learning algorithms in predicting loan defaults. This paper introduced a robust methodology for managing data, which encompassed preprocessing and feature engineering. These steps are crucial to attain a high level of accuracy in the model. Despite the widespread usage of Logistic Regression and Support Vector Machines, the study lacks any novel methodologies or significant improvements to these established models. This limitation may hinder its potential to advance the field of financial analytics.

Methodology

Our workflow begins with Data Collection, where relevant data is gathered from publicly available sources, such as Kaggle [17-18], which offers a rich repository of datasets for various machine learning tasks. The chosen dataset is often selected based on its relevance to loan approvals, containing key variables such as borrower demographics, financial history, and loan details.

Once the data is collected, it undergoes Data Preprocessing to ensure it is clean, consistent, and usable for machine learning models. This step is crucial, as raw data often contains missing values, inconsistencies, and outliers that can negatively impact model performance. Techniques such as imputation or removal are employed to handle missing values. Additionally, categorical variables are encoded into numerical formats, which makes them compatible with machine learning algorithms. Numerical features, such as income or loan amount, are normalized or scaled to ensure that no single feature disproportionately affects the model's learning process.

Next, the dataset is split into Training and Test sets, typically using an 80/20 split, where 80% of the data is used to train the models and 20% is reserved to test and evaluate the models' performance. Splitting the data ensures that the model is exposed to unseen data during testing, which helps to assess its generalization ability and prevent overfitting. Various machine learning models, including Logistic Regression, Decision Tree, Random Forest, K-Nearest Neighbor (KNN), and Feed-Forward Neural Network (FFNN), are then trained on the training dataset.

Following model training, each model is evaluated using multiple performance metrics such as accuracy, precision, recall, F1-score, specificity, geometric mean and balanced accuracy. Accuracy alone may not always provide the best indication of a model's performance, especially in cases of imbalanced datasets, where metrics like precision (the ability to correctly predict positive cases) and recall (the ability to detect all actual positive cases) are equally important. The F1score offers a balance between precision and recall, making it a key metric for models handling skewed data. Specificity measures the model's ability to correctly identify negative cases. Based on these metrics, the best-performing model is selected. This evaluation process is critical, as different models may excel in various aspects depending on the dataset and evaluation criteria.

Once the optimal model is chosen, it is integrated into a mobile application. For smooth integration, the selected model is often converted into a lightweight format, such as TensorFlow Lite, allowing it to run efficiently on mobile devices. This conversion ensures the model can make predictions quickly, even with limited computational resources available on mobile platforms. The application is designed to handle real-time inputs from users, offering immediate loan approval predictions.

This methodology, as abstracted in Fig. 1, offers a practical and efficient solution for predicting loan approvals using machine learning on a mobile platform. By leveraging advanced algorithms, the solution enhances both the user experience and the operational efficiency of financial institutions, providing quick and reliable assessments of loan applications.

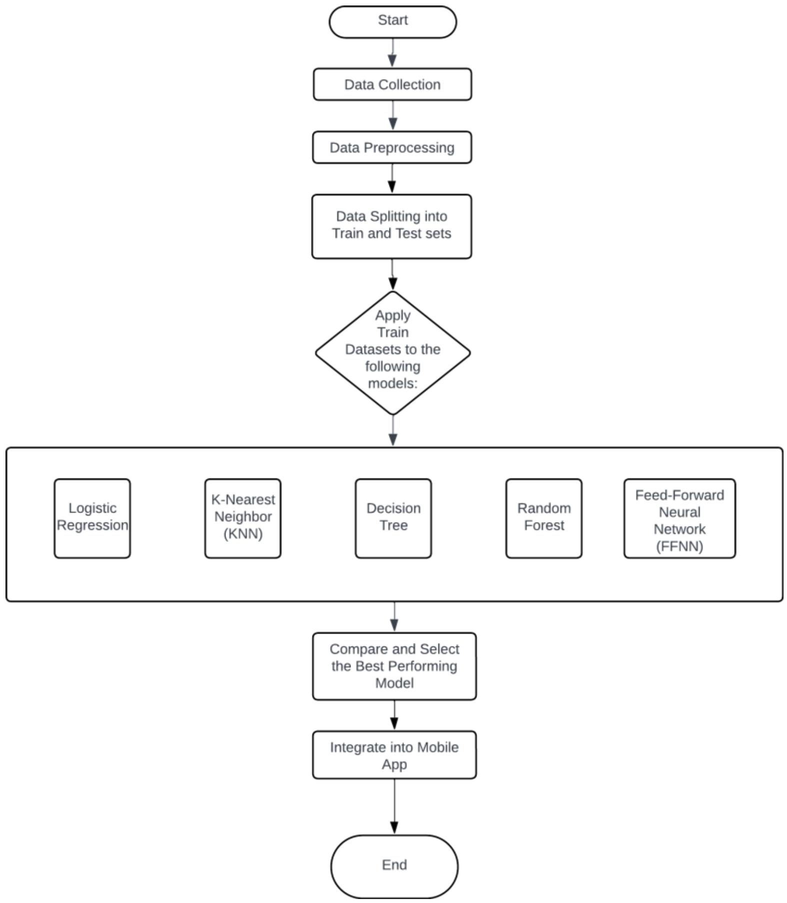


Fig. 1: Proposed Methodology.

A. Machine Learning Models

The following models were implemented and tested on the dataset:

- Logistic Regression: A widely used statistical model for binary classification tasks. It predicts the probability of a binary outcome by applying a logistic function to a linear combination of the input features.
- Decision Tree: A non-parametric model that splits the dataset into subsets based on the most significant features, eventually arriving at a decision by traversing through the branches.
- Random Forest: An ensemble method that builds multiple decision trees and aggregates their results to improve prediction accuracy.
- Support Vector Machine (SVM): This constructs hyperplanes in a multi-dimensional space to classify data points. It is effective in high-dimensional spaces and suitable for cases where the number of features exceeds the number of observations.
- K-Nearest Neighbors (KNN): A simple, nonparametric algorithm that classifies a data point based on the majority class of its k-nearest neighbors.
- Feed Forward Neural Network (FFNN): A deep learning model that consists of multiple layers of neurons, where each layer is fully connected to the next. FFNN can capture complex patterns in the data.

Each model was evaluated based on accuracy, precision, recall, specificity, F1-score, geometric mean and Balanced.

B. Model Training

- Data Splitting: The dataset (X train, y train) is used to train the machine learning models. The dataset comprises a set of features (X _train) and their related labels or targets (y train). The features encompass data such as the borrower's credit score, income level, loan amount, and other pertinent financial attributes. The target variable, y train, usually represents the result, such as whether the loan was repaid successfully or defaulted. The testing dataset, consisting of features (X test) and labels (y test), replicates the structure of the training dataset. This allows for evaluating the model's accuracy, precision, and recall using new and unseen data.

- Data Scaling: Prior to training, it is crucial to standardize the features in the training dataset (X train). Scaling is an essential preprocessing step in numerous machine learning algorithms, especially those that are influenced by the magnitude of input features, such as SVMs, knearest neighbors, and neural networks. Scaling was important because it ensured equal contribution of all features to the result and prevented features with larger ranges from overpowering the training process. Additionally, it accelerated the convergence of machine learning algorithms when the features have a comparable scale and are distributed close to a normal distribution.
- Scaling Method: The StandardScaler function from the preprocessing module of Scikit-learn was utilized. The scaler was first used on the training data (X train) to obtain the scaling parameters (such as mean and standard deviation in the case of StandardScaler). Later, it was used to convert both the training and testing datasets. To minimize overfitting and erroneous performance ratings, it was crucial to just apply the scaler to the training data and not transfer any information from the test set.
- Training and Evaluation: After scaling the data appropriately, different models were trained using the training dataset. The performance of these models was evaluated on a scaled testing dataset using problem-specific metrics such as recall, precision, accuracy, and F1 score for classification tasks.

C. Mobile App Design

The mobile app, developed using Flutter, is designed to enhance the operational efficiency, accessibility, and customer engagement of financial institutions by providing real-time predictions of loan defaults on the go. Users can input borrower information, receive accurate loan default predictions, manage past predictions, and obtain immediate feedback with detailed explanations and suggestions, as outlined in Fig. 2. Additionally, the app integrates NFC technology to streamline
the fund receiving process, improving transaction convenience for users.

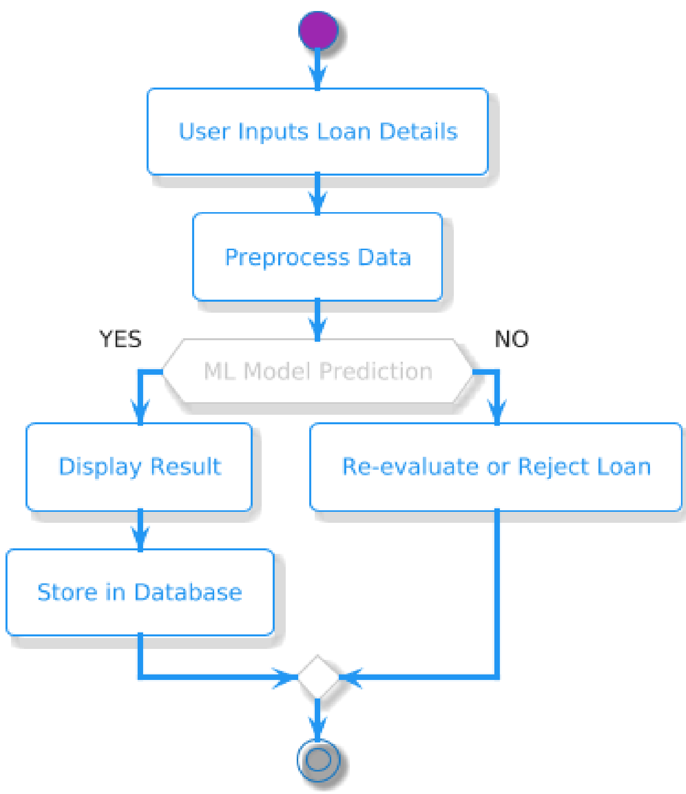


Fig. 2: Flowchart of Loan Request

Fig. 2 illustrates the step-by-step workflow of the loan prediction process within the app. The process begins with the user inputting loan details. This input is then subject to Data Preprocessing, which includes crucial validation checks to ensure data accuracy and completeness. Following preprocessing, the data is fed into a machine learning model for prediction. Based on the prediction outcome, the app either displays the result and stores it in the database for future reference or prompts a re-evaluation or rejection of the loan application if the prediction results are unsatisfactory. This flowchart provides a visual understanding of how the application manages data to ensure accurate loan predictions.

Results and analysis

A. Model Results

The results from the evaluation of the models are summarized in Table I. Among all models, the Feed Forward Neural Network (FFNN) achieved the highest accuracy of 94.72%, followed closely by the Support Vector Machine. The Decision Tree algorithm, while interpretable, performed less effectively due to its tendency to overfit the training data. As shown in Fig. 3, the accuracy of the FFNN was evaluated across 6 epochs, displaying slight fluctuations before stabilizing around the highest accuracy value.

TABLE I: Model Evaluation Results

Model	Acc (%)	Pre	Recall	Spe	F1	Geo	IBA
LR	94.07	0.94	0.93	0.95	0.93	0.94	0.88
DT	81.99	0.85	0.86	0.78	0.86	0.81	0.67
RF	84.83	0.87	0.87	0.82	0.87	0.85	0.72
SVM	94.07	0.94	0.93	0.95	0.93	0.94	0.88
KNN	92.70	0.93	0.92	0.94	0.92	0.93	0.86
FFNN	94.72	0.94	0.93	0.96	0.93	0.95	0.89

The superior performance of the Feed Forward Neural Network (FFNN) and Logistic Regression is evident in metrics such as accuracy, specificity, and F1-score. FFNN achieved the highest specificity (0.96) and a strong F1-score (0.93), highlighting its effectiveness in correctly identifying true positives and negatives, while minimizing false positives. Logistic Regression also performed well, with a specificity of 0.95 and an accuracy of 94.07%, demonstrating a balanced performance across key metrics. These results indicate FFNN's potential for mobile integration, particularly when optimized for efficiency using TensorFlow Lite. Additionally, Support Vector Machine (SVM) delivered competitive performance with faster prediction times, making it ideal for computationally efficient applications.

The FFNN's accuracy across different epochs, as depicted in Fig. 3, illustrates the learning behavior of the model during training, where the accuracy dipped in the early stages but improved significantly in subsequent epochs.

B. Model Integration

The TensorFlow Lite model file (.tflite) was integrated into the mobile app's codebase. The integration involved using the TensorFlow Lite interpreter, which is responsible for running the model on the mobile device. This interpreter was embedded into the app, allowing the best performing model (FFNN) to make predictions directly on the device.

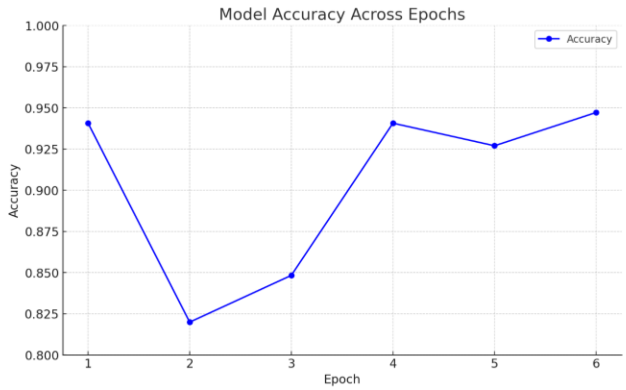


Fig. 3: FFNN Accuracy Graph.

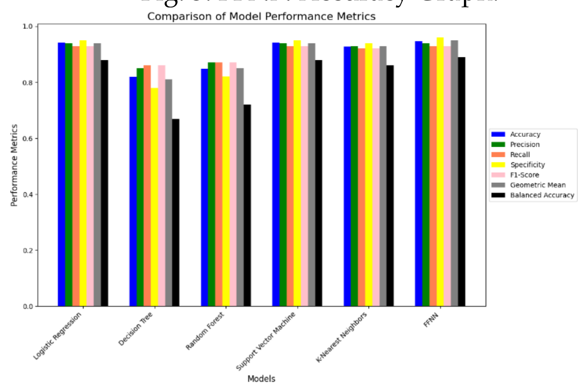


Fig. 4: Performance indicators of various models

The app is designed to collect user inputs relevant to loan approval (such as income, loan term etc.) through the user interface. These inputs are preprocessed to match the format required by the FFNN model. Once the user inputs are collected, they are passed to the TensorFlow Lite interpreter, which feeds them into the FFNN model for prediction. The model processes the input data and returns a prediction, indicating whether the loan is likely to be approved or not.

C. Mobile App Results

Fig. 5 illustrates the home screen of the mobile application" FastFunds Capital." Upon logging in, the user is welcomed by name, along with a summary of their account details. The main features accessible from the home screen include Transactions, Transfer, Request Loan, and Loan History. These options provide quick access to essential financial functions, such as viewing transaction history, transferring funds, applying for loans, and checking past loan requests. The balance section is displayed in a secure manner, ensuring the user's financial data is protected, with an option to hide or display the balance. This intuitive layout enhances user experience by offering a simple yet comprehensive dashboard for managing loans and other financial operations.

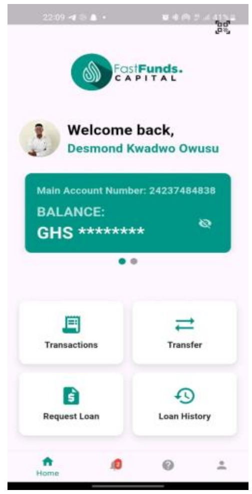


Fig. 5: Main Page of Mobile App

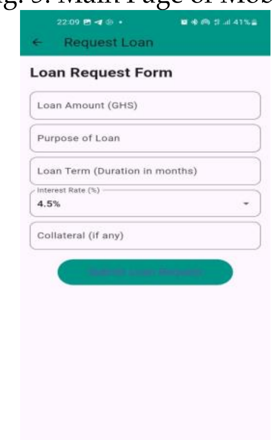


Fig. 6: Loan Request Form

Fig. 6 presents the Loan Request Form screen of the mobile application. This interface allows users to request a loan by filling in critical details such as the loan amount, purpose of the loan, loan term (in months), interest rate, and any collateral if applicable. The form is designed to be simple and user-friendly, ensuring that all necessary information is captured before submission. Once the required fields are completed, the user can submit the loan request by tapping the" Submit Loan Request" button.

Fig. 7 shows the Loan Request Result pop-up that appears after a user submits a loan application. Upon successful submission, the app evaluates the loan request using the integrated machine learning model. In this instance, the result indicates that the user qualifies for the loan, and they will receive an approval message shortly. This real-time feedback enhances the user experience by providing immediate and clear communication regarding the status of their loan request. Users are reassured about their loan qualification before receiving further confirmation through the app's notification system.

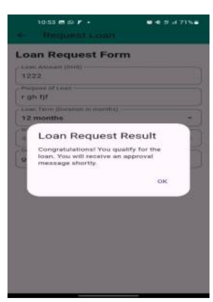
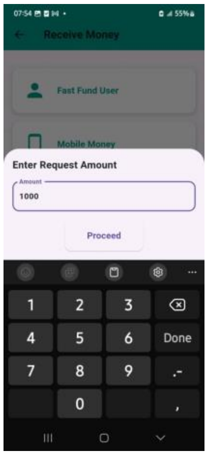
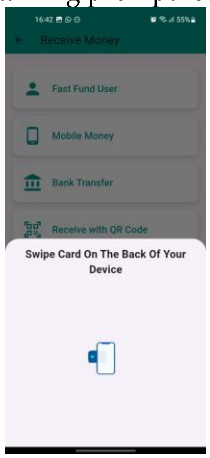


Fig. 7: Loan Prediction Results

D. NFC Integration

The integration of NFC technology into the mobile application is specifically engineered to provide a smooth and effective experience for users that require monetary transactions. By enabling customers to begin contactless transactions with minimal input, it streamlines the fund receiving process. The interface facilitates users in entering the appropriate amount, scanning their NFC card, securely validating the transaction with a PIN, and obtaining prompt feedback as shown in Fig.8 and Fig. 9.



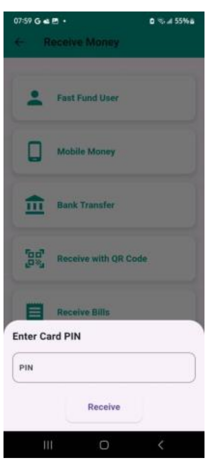


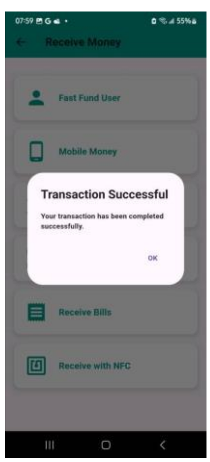
(a) Amount form (b) Swipe card

Fig. 8

Conclusion

This paper demonstrated the effectiveness of using machine learning models for predicting personal loan approvals on a mobile platform. The Feed Forward Neural Network outperformed other models in terms of accuracy and specificity. The integration of NFC technology for secure transactions adds an additional layer of trust and usability for users. The models were evaluated on a fixed dataset, which may not sufficiently capture the breadth of real-world scenarios, thus limiting the generalizability of the results. Future work can focus on improving the diversity of datasets, exploring hybrid models that merge the advantages of different machine learning techniques to enhance overall prediction precision and reliability, and optimizing resource consumption for mobile environments.





(a) Enter Pin (b) Successful Transaction

Fig. 9

References

A. Lakshmanarao, et al., "Loan Default Prediction Using Machine Learning Techniques and Deep Learning ANN Model," *2023 Annual International Conference on Intelligent Systems*, 2023.

R. Vivek, R. Mahaveerakannan, "Analyze the Lack of Accuracy in Loan Prediction using Logistic Regression and Random Forest," *Proceedings of IEEE ICONSTEM*, 2023.

R. Priscilla, T. Siva, M. Karthi, K. Vijayakumar, and Rajin Gangadharan, "Baseline modeling for early prediction of loan approval system," in *Proceedings of the International Conference on Artificial Intelligence and Knowledge Discovery in Concurrent Engineering, ICECONF 2023*, Institute of Electrical and Electronics Engineers Inc., 2023. doi: https://doi.org/10.1109/ICECONF57129.2023.10083650.

N. Hasan, Tanvir Anzum, T. Hasan, and N. Jahan, "Machine learning algorithm to predict fraudulent loan requests," in 2021 *12th International Conference on Computing Communication and Networking Technologies, ICCCNT 2021*, Institute of Electrical and Electronics Engineers Inc., 2021. doi: https://doi.org/10.1109/ICCCNT51525.2021.9579517.

K. Yadav and S. Singh, "Loan status prediction using SVM and logistic regression," in 2023 *14th International Conference on Computing Communication and Networking Technologies, ICCCNT 2023*, Institute of Electrical and Electronics Engineers Inc., 2023. doi: https://doi.org/10.1109/ICCCNT56998.2023.10307473.

World Development Report 2022: Finance for an Equitable Recovery, Februrary 2022, doi: https://doi.org/10.1596/978-1-4648-1730-4.

World Bank, "Fintech and financial inclusion," Private Sector Development Blog, Oct. 26, 2021. [Online]. Available: https://blogs.worldbank.org/psd/fintech-and-financial-inclusion.

https://www.openintro.org/data/index.php?

Lai, L. (2020, August). Loan default prediction with machine learning techniques. In 2020 International Conference on Computer Communication and Network Security (CCNS) (pp. 5-9). IEEE.

Xu, J., Lu, Z., & Xie, Y. (2021). Loan default prediction of Chinese P2P market: a machine learning methodology. Scientific Reports, 11(1), 18759.

Aslam, U., Tariq Aziz, H. I., Sohail, A., & Batcha, N. K. (2019). An empirical study on loan default prediction models. Journal of Computational and Theoretical Nanoscience, 16(8), 3483-3488.

Tiwari, A. K. (2018). Machine learning application in loan default prediction. JournalNX, 4(05), 1-5.

Sheikh, M. A., Goel, A. K., & Kumar, T. (2020, July). An approach for prediction of loan approval using machine learning algorithm. In 2020 international conference on electronics and sustainable communication systems (ICESC) (pp. 490-494). IEEE.

Madaan, M., Kumar, A., Keshri, C., Jain, R., & Nagrath, P. (2021). Loan default prediction using decision trees and random forest: A comparative study. In IOP conference series: materials science and engineering (Vol. 1022, No. 1, p. 012042). IOP Publishing.

Granström, D., & Abrahamsson, J. (2019). Loan default prediction using supervised machine learning algorithms.

Krasovytskyi, D., & Stavytskyy, A. (2024). Predicting Mortgage Loan Defaults Using Machine Learning Techniques. Ekonomika, 103(2), 140-160.

https://www.kaggle.com/datasets/nikhil1e9/loan-default

Banachewicz, K., & Massaron, L. (2022). The Kaggle Book: Data analysis and machine learning for competitive data science. Packt Publishing Ltd.