Bank Loan Classification

- By Nalin Malla

Abstract

This project aims to develop a predictive machine learning model to determine the acceptance of personal loans based on historical customer data. Utilizing a dataset comprising 15 features, I performed comprehensive data preprocessing to ensure data quality, followed by exploratory data analysis (EDA) to uncover underlying patterns and relationships. Feature engineering was conducted to enhance model performance by quantifying categorical variables and selecting the most relevant features. Several machine learning algorithms were evaluated, with the final model chosen based on its accuracy and efficiency.

In addition to the predictive model, I designed and implemented a chatbot interface that allows users to interact with the model in real-time. The chatbot guides users through a conversational flow, collecting necessary information and providing instant loan acceptance predictions. Error handling mechanisms were integrated to ensure a seamless user experience even in cases of invalid inputs.

The combination of a robust predictive model and an intuitive chatbot interface demonstrated that the relationship postulated during the EDA were in fact correct. This in turned revealed that the predictive model was working properly. This project is a simple example which underscores the potential for machine learning applications in streamlining financial services and enhancing customer engagement.

Table of Contents

Introduction

In the competitive landscape of financial services, banks and lending institutions continually seek innovative methods to enhance decision-making processes and customer engagement. One critical area of focus is the prediction of loan acceptance, which can significantly impact a bank's portfolio and customer satisfaction. Accurate predictions enable banks to tailor their services to individual needs, optimize marketing strategies, and manage risk effectively.

This project addresses the challenge of predicting personal loan acceptance using machine learning techniques. By analyzing historical customer data, I aim to identify patterns and factors that influence loan approval decisions. The dataset provided contains a range of features, including demographic information, financial behavior, and banking activity, which serve as inputs for the predictive model.

Moreover, recognizing the importance of accessibility and user-friendliness in financial technology applications, I have developed a chatbot interface that interacts with users in a conversational manner. This chatbot serves as a medium for users to inquire about their loan acceptance likelihood without navigating complex interfaces or understanding technical jargon.

The integration of a machine learning model within a chatbot framework represents an intersection of data science and user experience design. It is a simple example how technology can be leveraged to provide personalized financial insights while maintaining an engaging and straightforward user interaction.

In the following sections, I will detail the methodologies employed in data preprocessing, exploratory data analysis, feature engineering, model training, and chatbot development. I will also discuss my key findings and insights gleaned from the analysis.

Methodology

The methodology of this project is structured into several key components: data preprocessing, exploratory data analysis (EDA), feature engineering, model training, and chatbot development. Each component plays a vital role in the overall success of the loan acceptance prediction system.

1. **Data Preprocessing:**
   The initial step involved profiling the data to understand it's structure and properties. Then I cleaned the dataset to ensure accurate model training. This process included handling missing values by imputation or removal, correcting data types for consistency, and addressing outliers that could skew the results. The goal was to create a clean and reliable dataset for analysis.

2. **Exploratory Data Analysis (EDA):**
   EDA was conducted to gain insights into the dataset's characteristics and uncover any underlying patterns. This included univariate analysis to understand individual feature distributions, bivariate analysis to explore relationships with the target variable, and correlation analysis to identify multicollinearity among features. Some

3. **Feature Engineering:**
   In this phase, I relabeled the values in categorical variables with numerical values in order to improve model performance. This process also involved discarding irrelevant features which were identified during the EDA to reduce the noise.

4. **Model Training & Evaluation:**
   Multiple machine learning algorithms like, KNN, Decision Tree, SVM and Logistic Regression were considered for this project. Hyperparameter tuning was performed using grid search to optimize the model's performance. I evaluated models based on their training & test accuracy and the result of their subsequent confusion matrix. The final model was chosen based on its ability to make accurate prediction based on unseen data and its efficiency.

5. **Chatbot Development:**
   The chatbot was designed with a focus on user experience. A conversational flow was created that would guide users through the process of obtaining a loan acceptance prediction. The trained machine learning model was integrated into the chatbot's backend to provide real-time predictions based on user input.

6. **Testing & Validation:**
   The Chatbot was tested with different sets of inputs and when I inputted best values for key attributes like. Income, Home Ownership, CCAvg, Education, Mortgage & CD Account and worst values for other irrelevant attributes, based on the Exploratory Data Analysis done before, it is predicated that loan application will be approved. This demonstrated that the relationship discovered between various factors were correct and the trained model was working accurately.

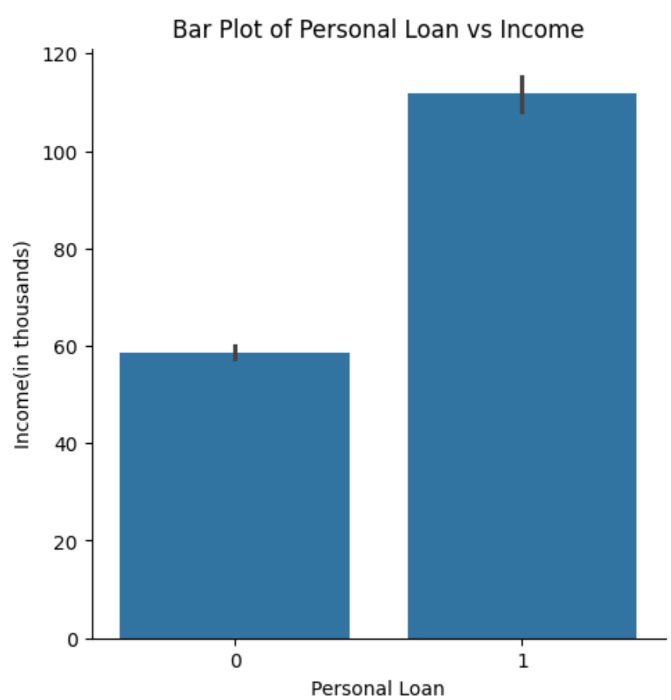
Insights drawn from EDA

I drew the following insights from exploratory data analysis:

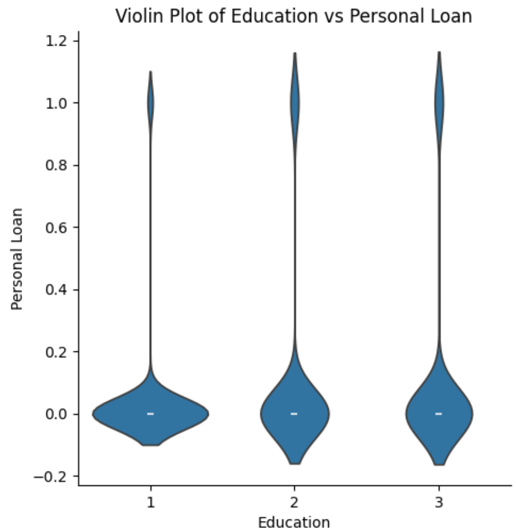
Key Influencers of Personal Loan:

There are two key factors which heavily influence approval of personal loan. They are:

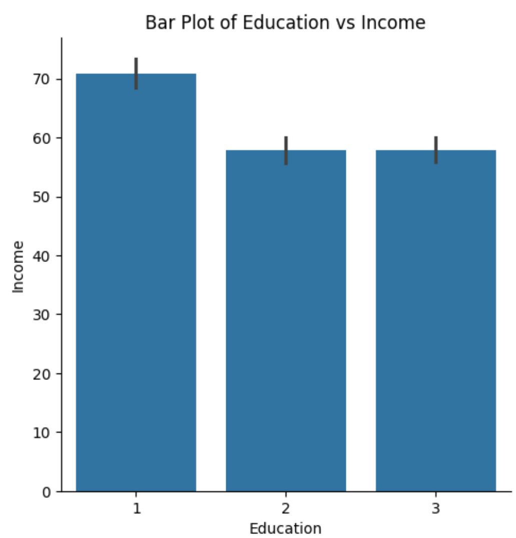
1. **Income:**
   Income has a positive relation with approval of Personal loan. In average the people who have been approved for Personal loans have almost twice the monthly Income of those who aren't approved. This is shown below:



2. **Education:**
   Level of Education also seems to be positively correlated to Personal loan. This is shown in the Violin Plot given below.



This plot shows that people who have Masters or Professional Degrees got more personal loans compared to those who had Bachelor's degree, even though people with bachelor's degree were the ones who applied for more loans.

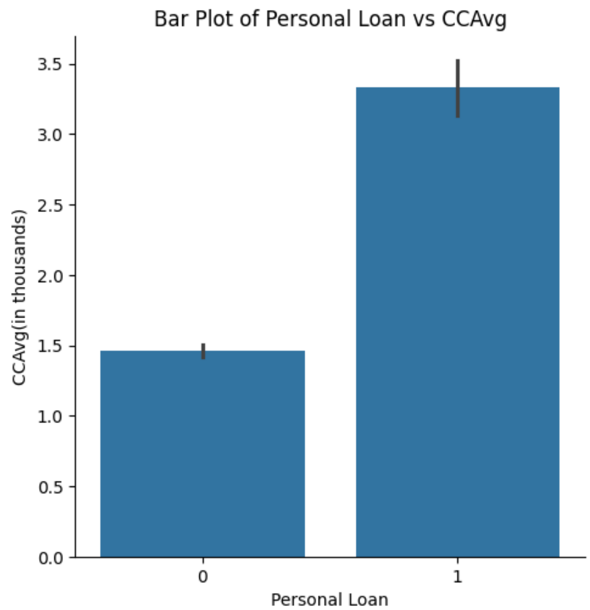


This bar plot shows that people who have Bachelor's Degree earn more in average compared to those with Masters and Professional Degree. But, despite this they have a lower chance of getting Personal Loans. This means that **Education is factor independent of Income which is essential for predicting acceptance of Personal Loan**.

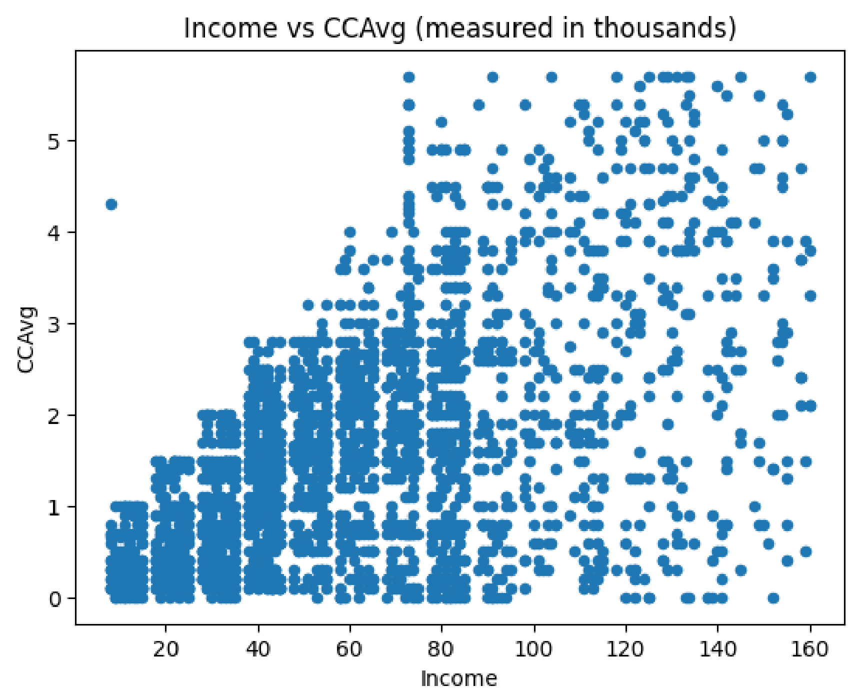
Other significant influencers of Personal Loan:

1. **Monthly credit card expense (CCAvg):**

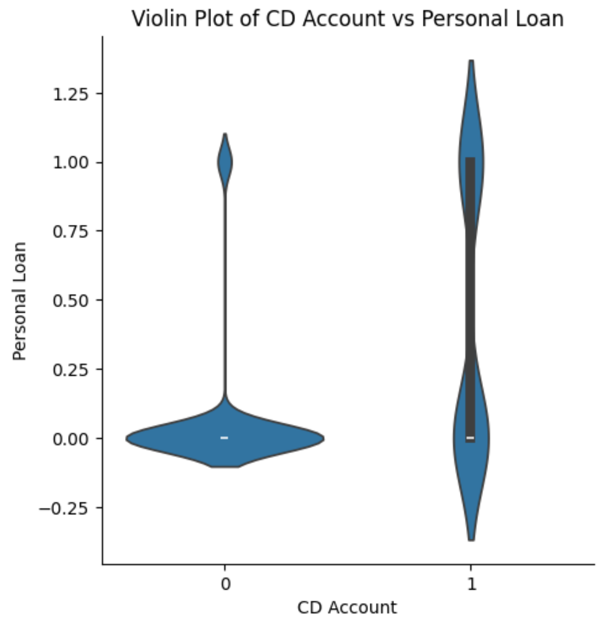
   CCAvg seems to have a positive relation with Personal Loan.



   But this is partly because of the positively curve-linear relation between Income and CCAvg.

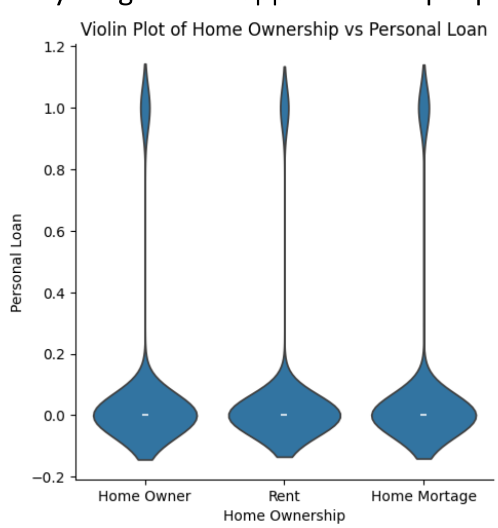


2. **CD Account:**
Having a CD Account is also a positive indicator for getting your personal loan approved.

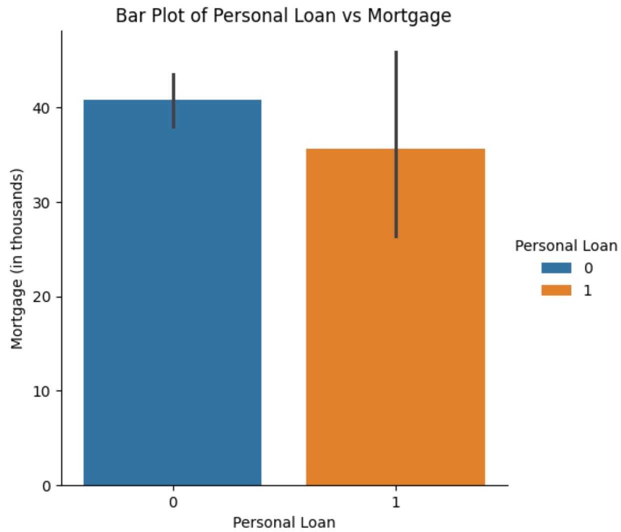


This violin plot shows that most people don't have an CD Account and only a few among them have been granted loan, whereas for the minority who do have a CD Account, about almost 1/4th of them have been granted loan.

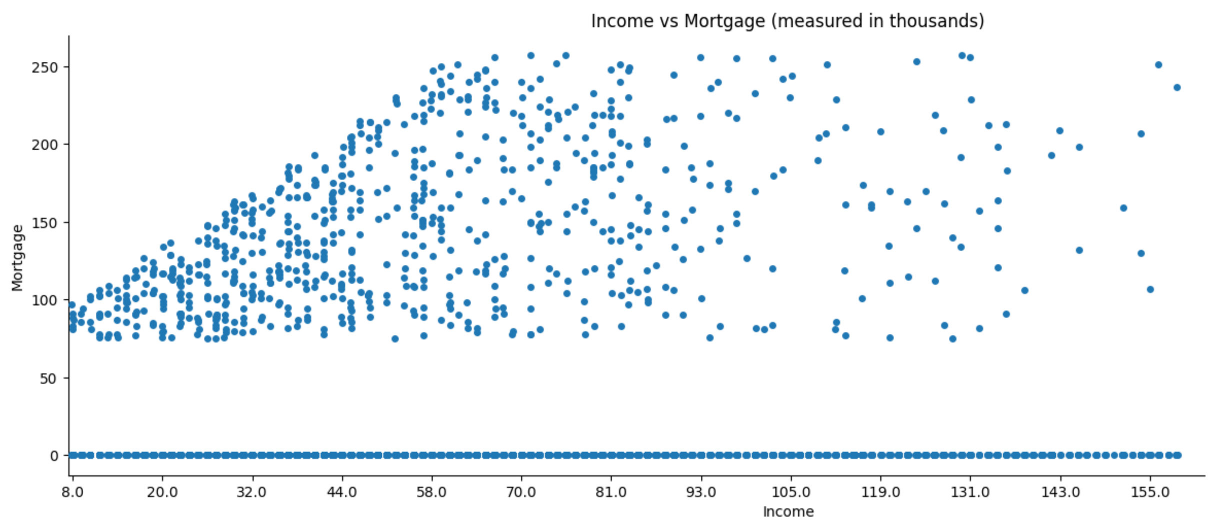
3. **Home Ownership:**
We can see from the violin plot given below that people who rent their home are the least likely to get loan approval and people who have home mortgage are the most likely.



4. **Mortgage:**
   Mortgage seems to have a negative relation with personal loan as people who have been approved for personal loans have less mortgage in average. This is shown in the following figure:



But this too seems to be a result of its curve-linear relation with Income.



In the scatter plot we can see that starting from income of 85000 the number of people who have mortgage starts to decline.

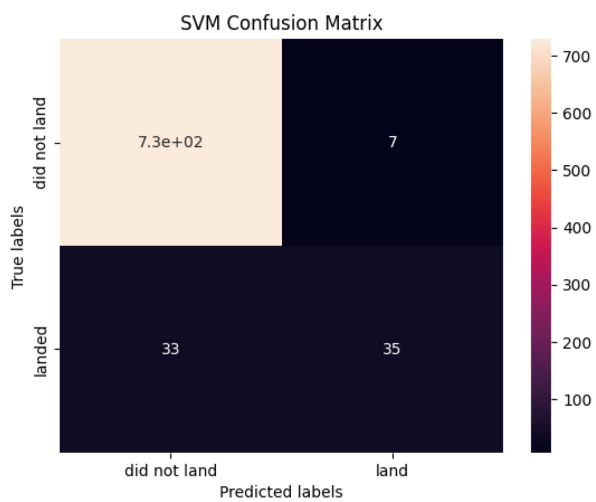
Comparison of Machine Learning Models

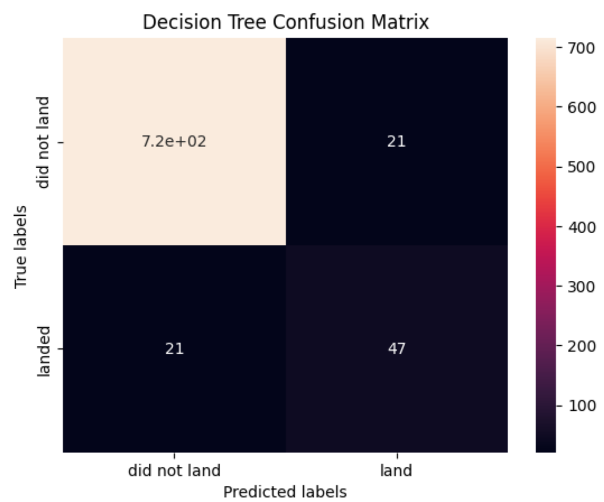
Test Accuracy

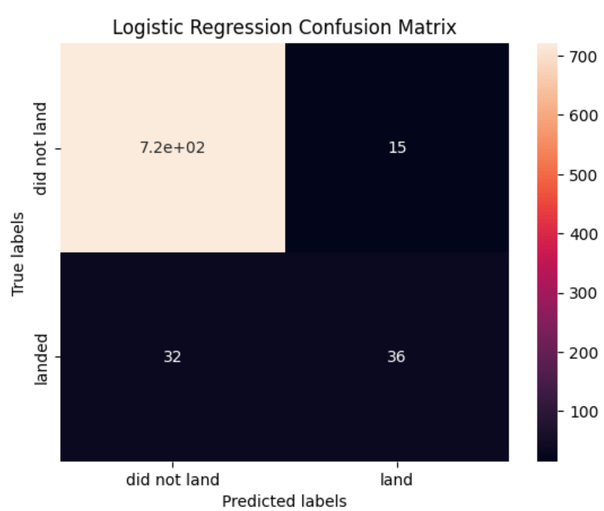
Rank	Model	Test Performance
1	SVM	0.950249
2	Decision Tree	0.947761
3	Logistic Regression	0.941542
4	K Near Neighbor	0.921642

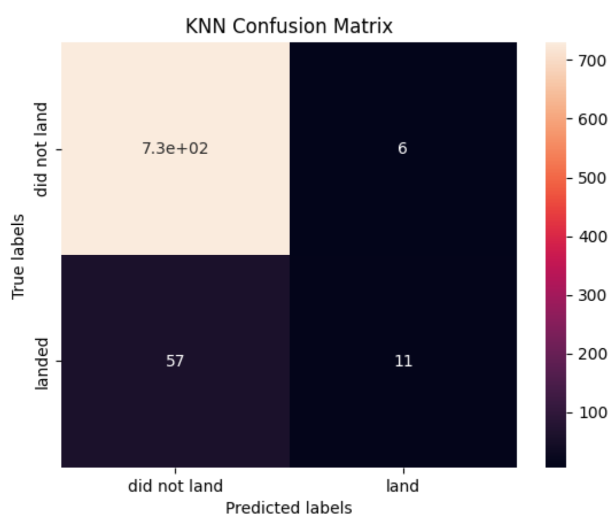
We can clearly see that SVM is the most accurate model with test accuracy of 95.02%.

Confusion Matrix









1. **KNN**: False negatives seem to be a big problem for this KNN model.

2. **Decision Tree**: Here, problem of false negative is substantially lower compared to KNN but the problem of false positive seems to have risen.

3. **SVM**: This model has by far the best accuracy score of 95.02% but is a little worse compared to the Decision Tree model as the true positive is a little lower. It also took an exceptionally long time to train this model i.e. about 30 minutes, this is with optimization for best performance.

4. **Logistic Regression**: This model has the highest false positives but is still better than the KNN model due to its lower false negative and higher true positives.

Best Model Choice

While the SVM model has the best accuracy but as it took a very long time to train, i.e. about 30 min in my case, and has less true positives compared to the second-best model i.e. Decision Tree model with 94.77% accuracy. **I have decided that the Decision Tree model is better overall**.

Chatbot Testing

```
PS X:\Data-Science\Projects\Bank-Lone-Classification> python .\chat_bot.py
Welcome to the Loan Acceptance Prediction Chatbot!
Please enter your name: Nalin
Hello Nalin, let me get some other information in order to predict if your loan application will be accepted or not.
Please enter your age (only whole numbers): 0
Use the following number in place of your gender,
0 for Female
1 for Male
2 for Others
Please enter your gender: 0
Please enter your experience i.e. number of years (only whole numbers): 0
Please enter your monthly income: 120000
Use the following number in place of your home ownership status,
0 for Home Mortgage
1 for Home Owner
2 for Rent
Please enter your home ownership status: 0
Please enter the number of members in your family: 0
Please enter your monthly average credit card expense: 3400
Use the following number in place of your level of education,
1 for Bachelors Degree
2 for Masters Degree
3 for Advanced/Professional Degree
Please enter your education level: 3
Please enter your total mortgage: 30000
If you have a security account enter 1, else enter 0.
Do you have a security account? : 0
If you have a certificate of deposit account enter 1, else enter 0.
Do you have a certificate of deposit account? : 1
If you have a internet banking enter 1, else enter 0.
Do you have a internet banking? : 0
If you have a credit card enter 1, else enter 0.
Do you have a credit card? : 0
[0, 0, 0, 120.0, 0, 0, 3.4, 3, 30.0, 0, 1, 0, 0]
Congratulations! Your loan application is predicted to be accepted.
```

The Chatbot was tested with different sets of inputs and when I inputted best values for key attributes like. Income, Home Ownership, CCAvg, Education, Mortgage & CD Account and worst values for other irrelevant attributes, based on the Exploratory Data Analysis done before, it is predicated that loan application will be approved. This demonstrated that **the relationship discovered between various factors were correct and the trained model was working accurately**.

Conclusion

In conclusion, this project demonstrates a useful application of machine learning for financial decision-making. This project has given me many insights and helped develop my data science skill.

Project Outcomes

The project achieved several significant outcomes:

1. **Discovery of key factors**:
   The relationships which were discovered during the Exploratory Data Analysis were proven to be true when the model was tested via the chatbot. This allowed us to know the factors which determine personal loan approval.

2. **Predictive Model Success**:
   The machine learning model attained an accuracy rate of 94.77%, indicating a high level of precision in predicting loan acceptance.

3. **Model Integration**:
   The seamless integration of the predictive model into the chatbot interface allowed for real-time predictions, enhancing user engagement and satisfaction.

Discussion

The project's outcomes affirm the efficacy of machine learning in enhancing financial decision-making processes. The predictive model's high accuracy demonstrates its potential as a reliable tool for loan approval predictions. The chatbot's user-friendly design further extends this technology's reach to a broader audience.

While promising, the project acknowledges the limitations inherent in data-driven models, such as potential biases and the need for continuous data updates to maintain accuracy. Future work will focus on refining the model and expanding the chatbot's capabilities to ensure it remains a valuable asset in the evolving landscape of financial services.