INTERNSHIP PROJECT

TEAM DETAILS

Team ID : LTVIP2023TMID04034

Team Size : 5

Team Leader : Gousia Irfana

Team Member : Jada Sasi Kanth

Team Member : Rongala Kavya Srija

Team Member : Pachipulusu Radha Vaishnavi

Team Member : Goodupu Venkata Vamsi Krishna

CONTENTS

INTRODUCTION

The objective of this project is to develop an intelligent and automated loan approval system using machine learning techniques. By analyzing a borrower's financial data and credit history, the system will predict the likelihood of loan approval, assisting financial institutions in making informed decisions quickly and accurately. These systems can analyze vast amounts of historical data and uncover hidden patterns to make data-driven decisions.

POURPOSE

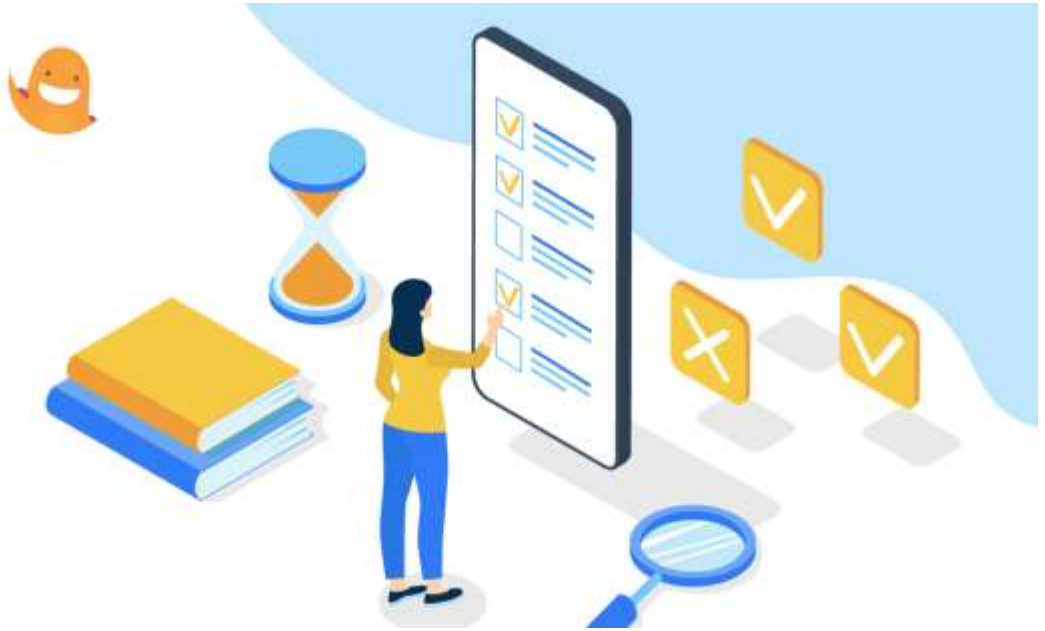
The primary purpose of this project is to streamline the loan approval process and provide a seamless experience to both borrowers and financial institutions.

By employing machine learning algorithms, we aim to achieve the following:

- **Improve Efficiency**
- **Enhance Accuracy**
- **Reduce Bias**
- **Improve Customer Experience**

LITERATURE SURVEY

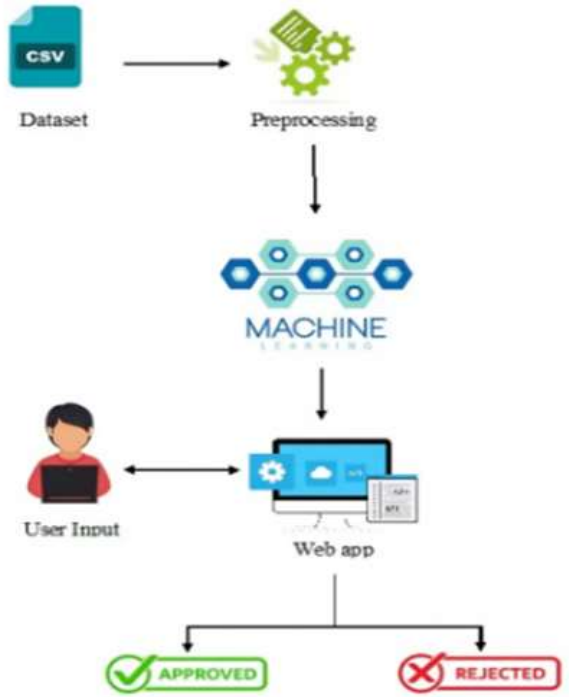
- Predictive Modeling for Loan Approval: A Survey (Author: Gupta S, Year: 2019)
- Machine Learning Approaches for Credit Risk Assessment (Authors: Smith J, Patel R, Year: 2020)
- Explainable AI for Loan Approval Decisions (Authors: Lee C, Kim J, Year: 2021)
- Impact of Feature Engineering on Loan Approval Prediction (Authors: Chen L, Wang H, Year: 2018)
- Addressing Class Imbalance in Loan Approval Prediction (Authors: Zhang Y, Li Q, Year: 2019)



PROPOSED SOLUTION

The "Predicting Personal Loan Approval Using Machine Learning" project proposes the following key steps and methodologies:

- **Data collection**
- **Data cleaning and preparation**
- **Exploratory Data Analysis (EDA)**
- **Model building**
- **Model evaluation and hyperparameter tuning**
- **Model deployment**
- **Performance testing and transparency**
- **Report creation**



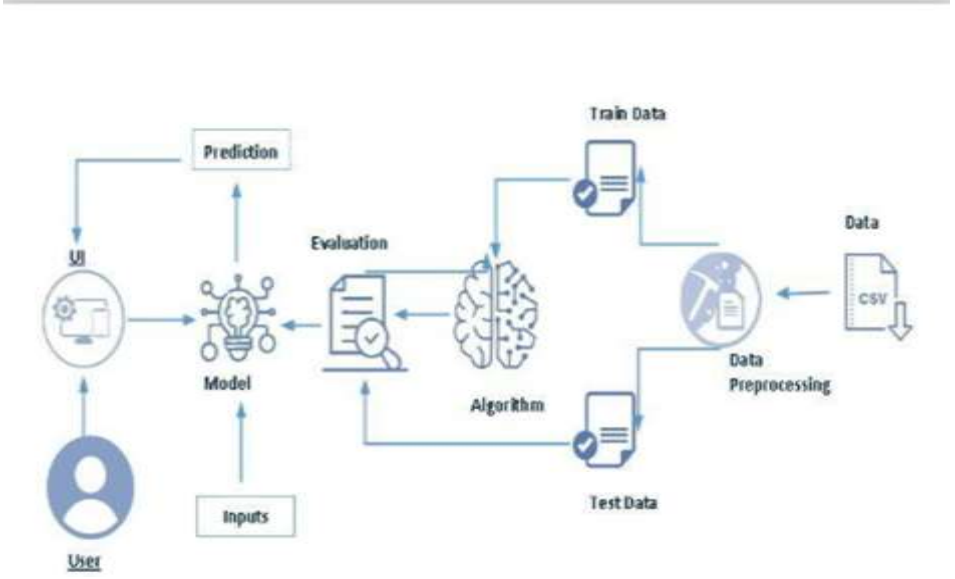
THEORETICAL ANALYSIS

- Project flow Diagram
- Dataset
- Data Preparation
- Data Cleaning
- Feature Engineering
- Exploratory Data Analysis
- Model Building
- Testing the Model
- Performance Testing & Hyper parameter Tuning
- Save the Best Model
- Integrate with Web Framework

Project Flow Diagram

- At the core of the system is the machine learning model, which takes input data from the borrower. The model then processes the data and generates a prediction regarding the likelihood of loan approval. This prediction, along with its confidence score, is conveyed back to the user through the web-based interface.

- The web interface, built using Python-Flask, provides a user-friendly platform for borrowers to enter their details and interact with the system. The integration of IBM Machine Learning enables seamless model deployment and real-time prediction capabilities.
- The project's architecture aims to strike a balance between functionality and simplicity. The streamlined process ensures quick responses to loan applicants while maintaining the accuracy and reliability of the loan approval predictions



Data Preprocessing

Dataset:

The dataset used in this project was obtained from Kaggle.

Data Preparation:

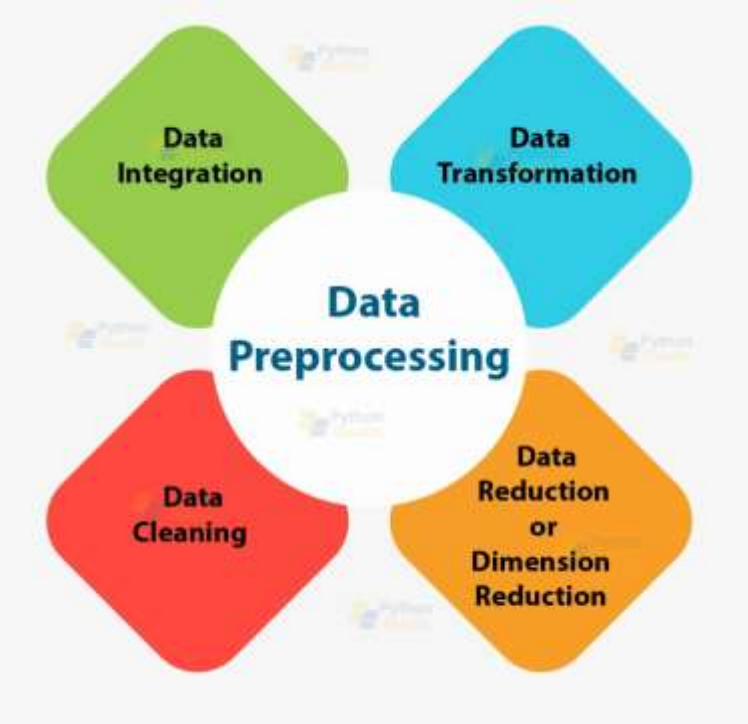
Data preparation is a critical step in machine learning. It involves processing the raw data

- **Data Cleaning:**
- **Handling Missing Values**
- **Dealing with Duplicate Records**

Feature Engineering:

Feature engineering involves creating new features or transforming existing features to enhance the model's predictive power.

- **Debt-to-Income Ratio:**
- **Loan-to-Income Ratio**

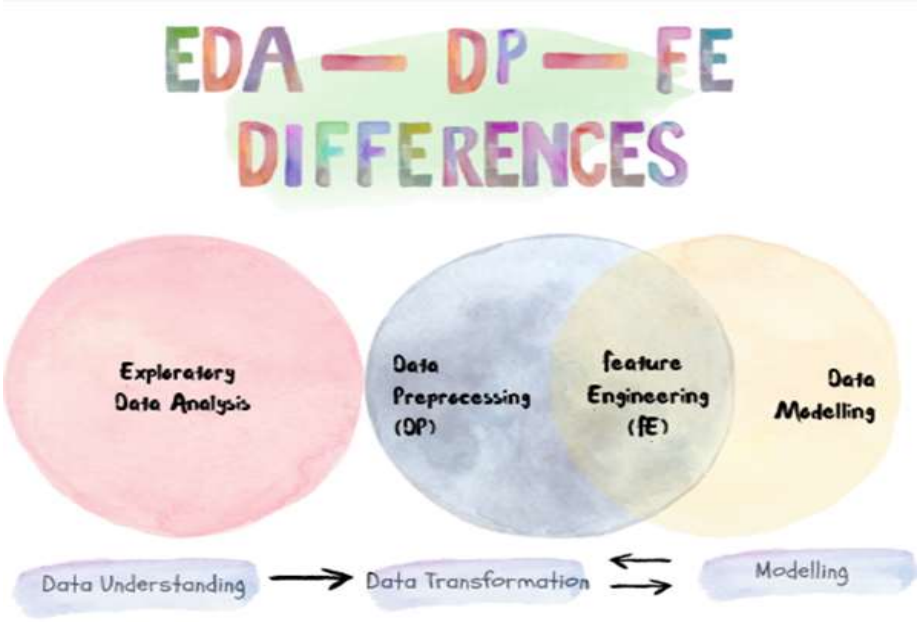


Exploratory Data Analysis

- During EDA, various statistical measures and visualizations are used to gain insights into the data distribution, relationships between features, and potential correlations between features and loan approval status.

Some key findings from the EDA process include:

- **Distribution of Credit Scores:** Analyzing the distribution of credit scores among loan applicants provides insights into the creditworthiness of the borrowers.
- **Income vs. Loan Amount:** Exploring the relationship between income and loan amount reveals trends in loan applications based on borrowers' financial capabilities.

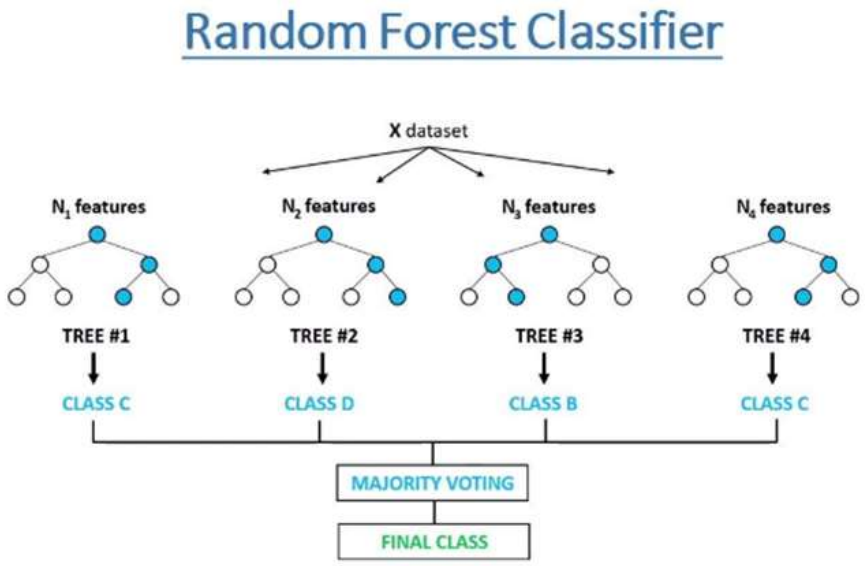


Model Building

Training the Model in Multiple Algorithms:

In this project, three popular machine learning algorithms are selected for training:

- **Logistic Regression**
- **Decision Trees**
- **Random Forests**

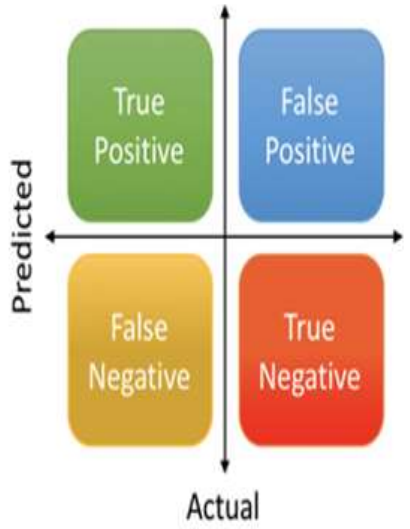


Testing The Model

$$\text{Precision} = \frac{\text{True Positive}}{\text{Actual Results}} \quad \text{or} \quad \frac{\text{True Positive}}{\text{True Positive + False Positive}}$$

$$\text{Recall} = \frac{\text{True Positive}}{\text{Predicted Results}} \quad \text{or} \quad \frac{\text{True Positive}}{\text{True Positive + False Negative}}$$

$$\text{Accuracy} = \frac{\text{True Positive + True Negative}}{\text{Total}}$$



Performance Testing & Hyper parameter Tuning

- Testing Model with Multiple Evaluation Metrics
- Comparing Model Accuracy Before & After Applying Hyperparameter Tuning

****DecisionTreeClassifier****
Confusion matrix
[[90 20]
 [35 94]]
Classification report

	precision	recall	f1-score	support
0	0.72	0.82	0.77	110
1	0.82	0.73	0.77	129
accuracy			0.77	239
macro avg	0.77	0.77	0.77	239
weighted avg	0.78	0.77	0.77	239

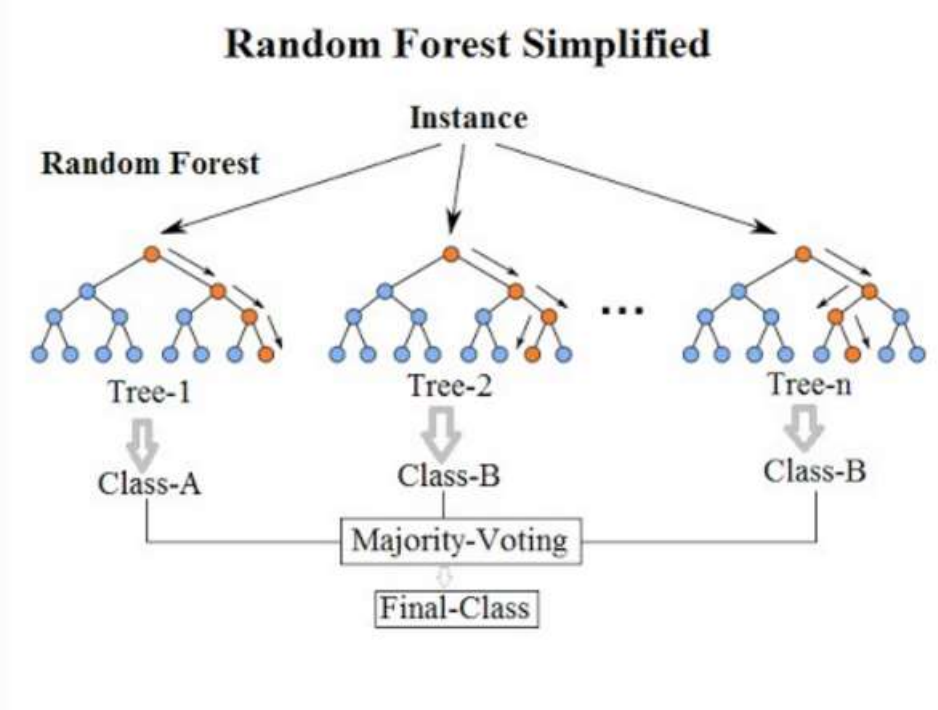
RandomForestClassifier
Confusion matrix
[[88 22]
 [19 110]]
Classification report

	precision	recall	f1-score	support
0	0.82	0.80	0.81	110
1	0.83	0.85	0.84	129
accuracy			0.83	239
macro avg	0.83	0.83	0.83	239
weighted avg	0.83	0.83	0.83	239

Save The Best Model

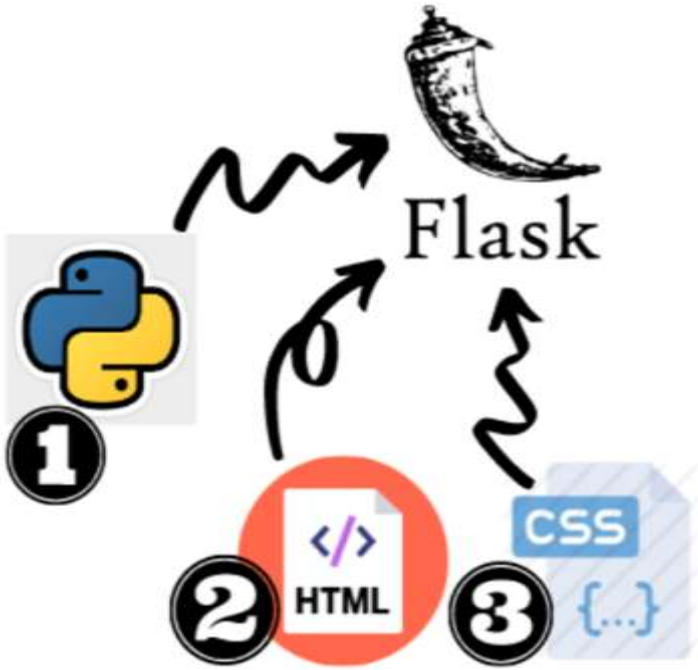
- Based on the evaluation results, the Random Forests model with hyperparameter tuning was selected as the best-performing model. This model was saved for deployment in the production environment.

- The choice of the Random Forests model is driven by its superior accuracy and robustness in predicting loan approvals. The ensemble nature of the Random Forests algorithm reduces the risk of overfitting and enhances the model's ability to generalize to unseen data



Integrate with Web Framework

- To provide a user-friendly interface for borrowers, the best model was integrated with the Python-Flask web framework. Flask enables the creation of a responsive and interactive web application that connects borrowers with the machine learning model.
- The prediction, along with a confidence score, is conveyed back to the user through the web interface.
- The use of Flask simplifies the model deployment process and provides borrowers with a seamless and efficient loan approval experience.



Requirements

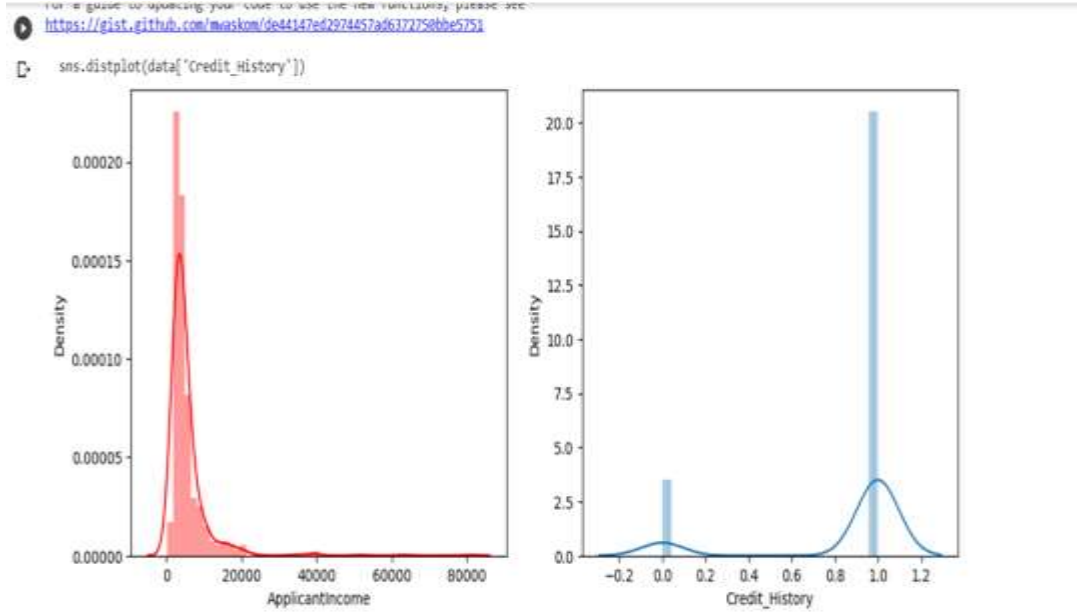
Hardware Requirements

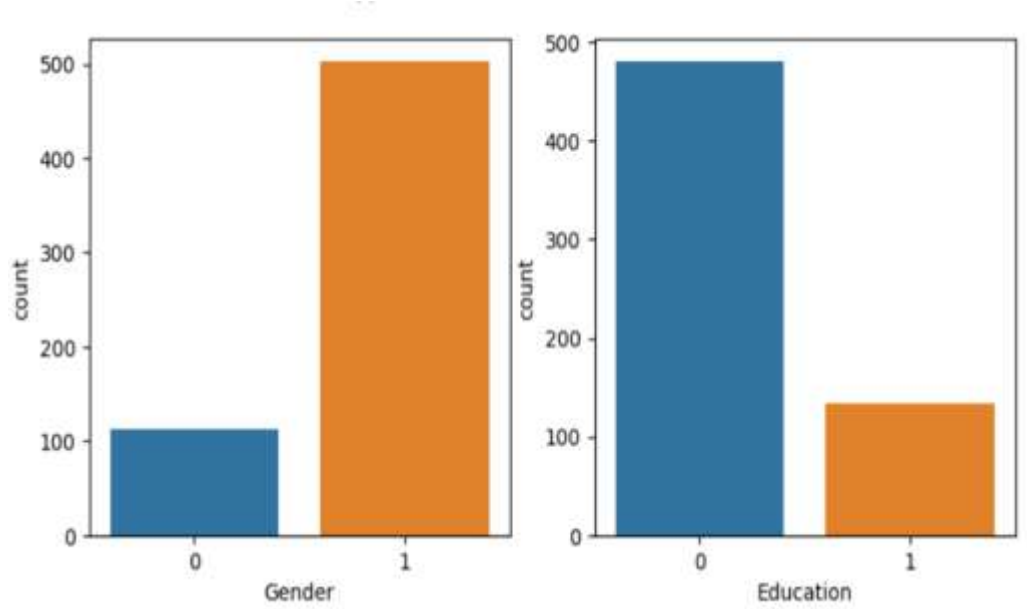
A computer with sufficient processing power and memory to handle data processing and model training efficiently. The hardware must be capable of running the required software components smoothly. The availability of a stable internet connection is essential for web-based interactions with the loan approval system

Software Requirements

- Python
- IBM Machine Learning
- Flask
- HTML, CSS

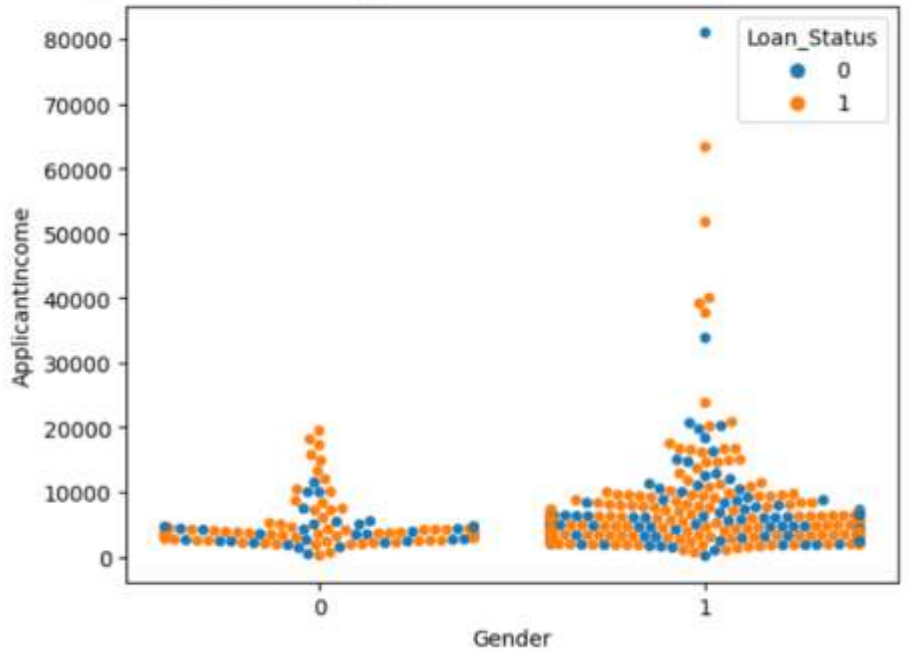
RESULTS

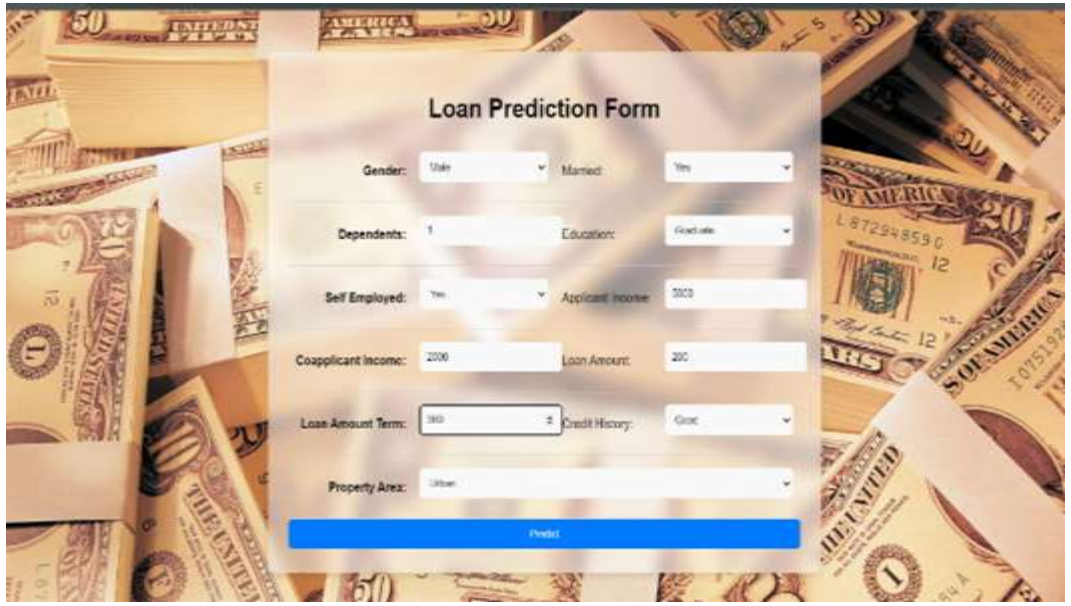

https://gist.github.com/mwaskom/de44147ed2974457ad6372750bbe5751
sns.distplot(data['Credit_History'])
Density
ApplicantIncome
Credit_History


count
Gender
Education

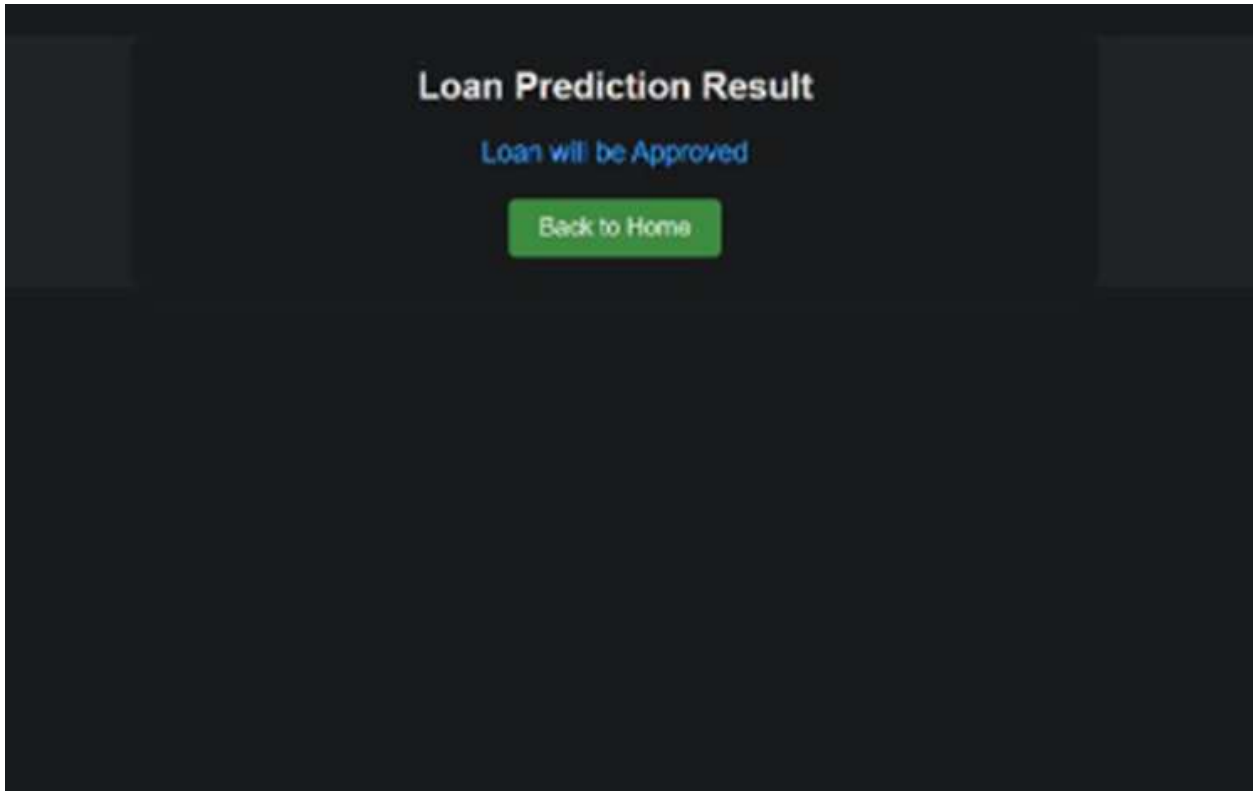
RESULTS

```
/usr/local/lib/python3.10/dist-packages/seaborn/categorical.py:3544: UserWarning: 6
  warnings.warn(msg, UserWarning)
```





RESULTS



ADVANTAGES

- **Enhanced Efficiency**: The automation of loan approval processes reduces the time and resources required for manual evaluations, leading to faster response times and increased operational efficiency for financial institutions.
- **Improved Accuracy**: Machine learning algorithms can process vast amounts of historical loan data, leading to more accurate and data-driven predictions. The model's ability to learn from historical data contributes to higher precision and recall rates.
- **Objective Decision-Making**: By removing human judgment from the approval process, the system ensures fairness and objectivity in loan decisions. This reduces the potential for biases in loan approvals and promotes equitable treatment of all borrowers.
- **Real-Time Analysis**: The integration of a web-based interface allows borrowers to receive instant loan approval predictions. This real-time analysis enhances the customer experience and expedites loan processing.

DISADVANTAGES

While the proposed solution offers numerous benefits, it is essential to acknowledge potential limitations:

- **Dependency on Historical Data:** The model's performance relies on the quality and relevance of historical loan data. As market conditions and customer behaviors evolve, the model may require periodic updates and retraining to maintain accuracy.

- **Model Uncertainty:** Although the machine learning model provides a confidence score for its predictions, there may be instances where the model's certainty is low due to ambiguous or sparse data points. In such cases, manual review or additional verification may still be necessary.

- **Ethical Considerations:** While the model aims to be unbiased, there is always a risk of perpetuating or amplifying existing biases present in the historical data. Ensuring fairness in machine learning models requires careful examination and mitigation of biases throughout the development process.

Addressing these limitations requires ongoing monitoring, feedback, and continuous improvement of the loan approval prediction system.

APPLICATIONS

- **Credit Counseling:** Loan applicants with lower chances of loan approval can benefit from credit counseling services. By understanding the factors affecting their creditworthiness, individuals can take steps to improve their credit scores and financial standing.
- **Personal Finance Management:** Individuals can use the loan approval prediction model as part of their personal finance management. By estimating loan approval chances, they can plan their finances more effectively and avoid unnecessary loan applications.
- **Credit Score Improvement:** Loan applicants can learn from the model's insights to identify areas for credit score improvement. By understanding the impact of various features on loan approval, individuals can work on enhancing their credit profiles.
- **Consumer Protection:** The project's model can assist in safeguarding consumers by ensuring fair and consistent loan approval decisions based on objective criteria. This helps protect individuals from potential bias or discrimination in loan approvals.

APPLICATIONS

Credit Agency Information

Public Financial Statements

Bank & Trade References

Financial Stress Prediction

CONCLUSION

In conclusion, the "Predicting Personal Loan Approval Using Machine Learning" project successfully achieved its objectives of developing an automated loan approval system. By leveraging machine learning algorithms, the project provides an efficient and accurate solution for financial institutions to assess loan applications objectively. The integration of a user-friendly web interface allows borrowers to check their loan eligibility conveniently. The project's ability to improve the loan approval process offers potential benefits to both financial institutions and loan applicants.

The successful implementation of the project demonstrates the transformative power of machine learning and artificial intelligence in the financial sector. As technology continues to evolve, it is imperative for financial institutions to embrace innovation and adopt automated systems that enhance efficiency, accuracy, and customer experiences.