**Artificial Intelligence and Machine**

**Learning**

Project Report

Semester-IV (Batch-2022)

Loan Approval Prediction

Submitted By: - Supervised By: - Shubham Singhal

|  |  |  |
| --- | --- | --- |
| Gautam Kumar | 2210991577 | G11 |
| Gaurav Anand | 2210991570 | G11 |
| Himanshu | 2210991663 | G11 |
| Nidhish | 2210990606 | G11 |

**Department of Computer Science and Engineering**

**Chitkara University Institute of Engineering & Technology, Chitkara University, Punjab**

**Table of Contents:**

|  |  |  |
| --- | --- | --- |
| **Sr. no.** | **Content** | **Page No.** |
| **1.** | **Introduction :- Background, Objectives, Significance** | 3-6 |
| **2.** | **Problem Definition and Requirements :- Problem Statement, Software Requirements, Hardware Requirements, Data Sets** | 7-8 |
| **3.** | **Proposed Design / Methodology :- Schematic Diagram, File Structure, Algorithms Used** | 9-12 |
| **4.** | **Results :- Screenshots, Metrics** | 13-15 |
| **5.** | **References** | 16 |

**Introduction :-**

**Overview**

**In the financial sector, assessing the risk associated with loan applications is a critical task for banks and lending institutions. The decision to approve or reject a loan application is influenced by multiple factors, including the applicant's credit history, income, employment status, and other demographic information. Traditional methods of loan evaluation often rely heavily on manual processes and the subjective judgment of loan officers, which can be time-consuming and prone to errors or biases. To enhance the efficiency and accuracy of loan approval decisions, many institutions are turning to machine learning (ML) and artificial intelligence (AI) solutions.**

**Objective:-**

The primary objective of this project is to develop a predictive model that can automatically assess and determine the likelihood of a loan application's approval. By leveraging historical loan data, the model aims to provide a more objective and data-driven approach to loan approval, thereby minimizing the risk for the lender and improving the customer experience by speeding up the decision-making process.

Significance

The implementation of an AI/ML-based loan approval system offers several significant benefits:

Improved Accuracy: Machine learning models can analyze vast amounts of data and identify patterns that may not be evident to human evaluators, leading to more accurate predictions.

Efficiency: Automating the loan approval process reduces the time and resources required, enabling faster decision-making and processing times.

Consistency: AI/ML models apply the same criteria to every application, ensuring consistent evaluations and reducing biases that might affect human judgment.

Scalability: Automated systems can handle large volumes of applications simultaneously, making it easier to scale operations as the number of applicants grows.

**Problem Definition and Requirements**

**Problem Statement :-**

Develop a machine learning model to predict whether a loan application will be approved or rejected by a financial institution based on the applicant's information and loan details. This predictive model will assist banks in streamlining the loan approval process, minimizing human bias, and enhancing decision-making efficiency.

Key Features (Inputs):

Applicant Information:

Gender (e.g., Male, Female)

Marital Status (e.g., Married, Single)

Number of Dependents

Education (e.g., Graduate, Not Graduate)

Employment Status (e.g., Employed, Self-Employed)

Applicant Income

Co-applicant Income

**Software Requirements :-**

1.1 Purpose

The purpose of this document is to define the requirements for the Loan Approval Prediction System, which aims to automate the process of predicting loan approvals based on applicants' data using machine learning algorithms.

**The system will:**

**Collect and preprocess data from applicants.**

**Train machine learning models to predict loan approval outcomes.**

**Provide an interface for loan officers to input applicant data and get predictions.**

**Offer insights and explanations for predictions.**

Definitions, Acronyms, and Abbreviations

AI: Artificial Intelligence

ML: Machine Learning

SRS: Software Requirements Specification

API: Application Programming Interface

UI: User Interface

Hardware Requirement :-

To develop and deploy a loan approval prediction model using AI/ML, the hardware requirements will depend on several factors, including the size of the dataset, the complexity of the model, and whether you're planning to train the model locally or on a cloud platform. Here’s a general guide to help you determine the necessary hardware requirements

Development and Training Phase

Laptop/Desktop for Small-scale Development:

CPU: Intel i5/i7 or AMD Ryzen 5/7

RAM: 16 GB minimum (32 GB preferred for larger datasets)

Storage: SSD with at least 500 GB

GPU: NVIDIA GTX 1050 Ti or better (if planning to use GPU acceleration with libraries like TensorFlow or PyTorch)

Data Sets : -

For an AI/ML project focused on predicting loan approval, you can utilize several well-known datasets that are publicly available. These datasets contain various attributes related to loan applications, which can be used to train and evaluate machine learning models. Here are some recommended datasets:

1. UCI Machine Learning Repository:

2. Kaggle: Home Credit Default Risk

3. LendingClub Loan Data

4. UCI Machine Learning Repository: Credit Approval Data Set

5. Kaggle:

**Proposed Design / Methodology :-**

**Schematic Diagram:-**

**┌───────────────────────────┐**

**│ Data Collection │**

**│ (Customer Information, │**

**│ Credit History, Income, │**

**│ Employment Status, etc.)│**

**└────────────┬──────────────┘**

**│**

**▼**

**┌───────────────────────────┐**

**│ Data Preprocessing │**

**│ (Cleaning, Handling │**

**│ Missing Values, │**

**│ Encoding Categorical │**

**│ Variables, Scaling) │**

**└────────────┬──────────────┘**

**│**

**▼**

**┌───────────────────────────┐**

**│ Feature Engineering │**

**│ (Creating New Features, │**

**│ Feature Selection) │**

**└────────────┬──────────────┘**

**│**

**▼**

**┌───────────────────────────┐**

**│ Data Splitting │**

**│ (Training Set, │**

**│ Validation Set, │**

**│ Test Set) │**

**└────────────┬──────────────┘**

**│**

**▼**

**┌───────────────────────────┐**

**│ Model Selection │**

**│ (Choosing Algorithms: │**

**│ Logistic Regression, │**

**│ Decision Trees, │**

**│ Random Forest, etc.) │**

**└────────────┬──────────────┘**

**│**

**▼**

**┌───────────────────────────┐**

**│ Model Training │**

**│ (Training Models on │**

**│ Training Data) │**

**└────────────┬──────────────┘**

**│**

**▼**

**┌───────────────────────────┐**

**│ Model Evaluation │**

**│ (Using Validation Set, │**

**│ Metrics: Accuracy, │**

**│ Precision, Recall, │**

**│ F1 Score, AUC-ROC) │**

**└────────────┬──────────────┘**

**│**

**▼**

**┌───────────────────────────┐**

**│ Hyperparameter Tuning │**

**│ (Optimizing Model │**

**│ Parameters: Grid Search,│**

**│ Random Search, etc.) │**

**└────────────┬──────────────┘**

**│**

**▼**

**┌───────────────────────────┐**

**│ Final Model Evaluation │**

**│ (Using Test Set, Final │**

**│ Metrics Calculation) │**

**└────────────┬──────────────┘**

**│**

**▼**

**┌───────────────────────────┐**

**│ Model Deployment │**

**│ (Deploying Model to │**

**│ Production, Creating │**

**│ APIs, Integration) │**

**└────────────┬──────────────┘**

**│**

**▼**

**┌───────────────────────────┐**

**│ Monitoring and │**

**│ Maintenance │**

**│ (Tracking Model │**

**│ Performance, Updating │**

**│ Model, Handling Drift) │**

**└───────────────────────────┘**

**File Structure:-**

loan\_approval\_prediction/

│

├── data/

│ ├── raw/

│ │ ├── tra

│ │ ├── test.csv

│ │ └── additional\_data.csv

│ ├── processed/

│ │ ├── train\_processed.csv

│ │ ├── test\_processed.csv

│ │ └── feature\_engineered.csv

│ └── external/

│ ├── external\_data\_1.csv

│ └── external\_data\_2.csv

│

├── notebooks/

│ ├── 01\_data\_exploration.ipynb

│ ├── 02\_data\_cleaning.ipynb

│ ├── 03\_feature\_engineering.ipynb

│ ├── 04\_model\_building.ipynb

│ ├── 05\_model\_evaluation.ipynb

│ └── 06\_model\_deployment.ipynb

│

├── scripts/

│ ├── data\_preprocessing.py

│ ├── feature\_engineering.py

│ ├── model\_training.py

│ ├── model\_evaluation.py

│ ├── model\_inference.py

│ └── utils.py

│

├── models/

│ ├── random\_forest/

│ │ ├── rf\_model.pkl

│ │ └── rf\_metrics.json

│ ├── xgboost/

│ │ ├── xgb\_model.pkl

│ │ └── xgb\_metrics.json

│ └── logistic\_regression/

│ ├── lr\_model.pkl

│ └── lr\_metrics.json

│

├── reports/

│ ├── figures/

│ │ ├── correlation\_matrix.png

│ │ ├── feature\_importance.png

│ │ ├── model\_performance.png

│ └── final\_report.pdf

│

├── tests/

│ ├── test\_data\_preprocessing.py

│ ├── test\_feature\_engineering.py

│ ├── test\_model\_training.py

│ ├── test\_model\_evaluation.py

│ └── test\_utils.py

│

├── requirements.txt

├── README.md

├── setup.py

└── .gitignore

**Algorithms Used :-**

To develop a loan approval prediction algorithm using AI/ML, we can outline a comprehensive process that includes data collection, data preprocessing, feature engineering, model selection, training, evaluation, and deployment. Here's a detailed overview of each step:

1. Data Collection

2. Data Preprocessing

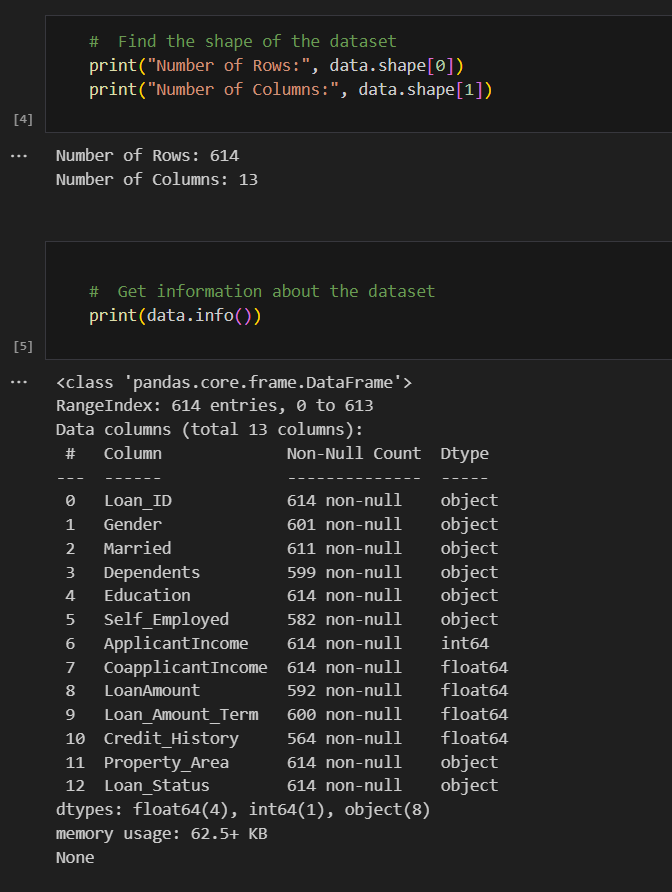
3. Feature Engineering

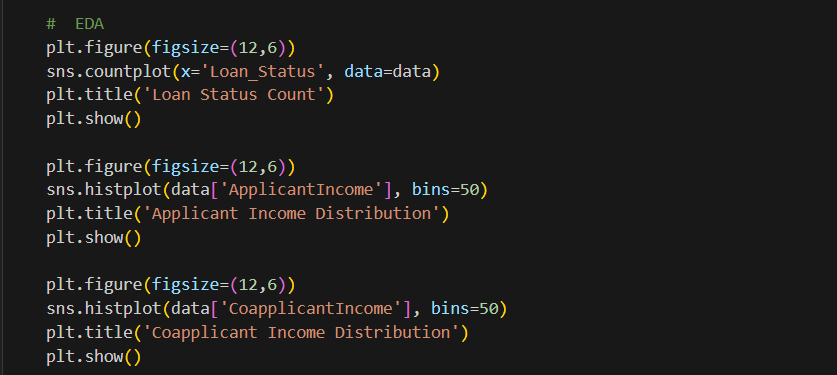
4. Model Selection

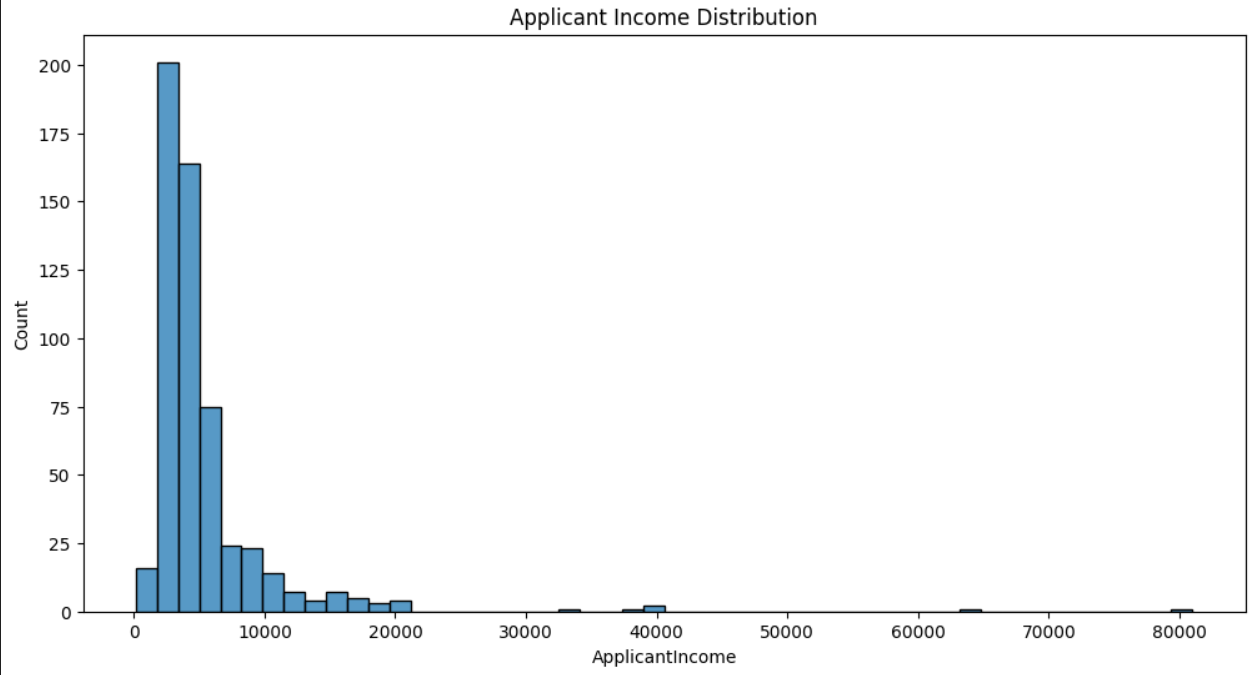
5. Model Training

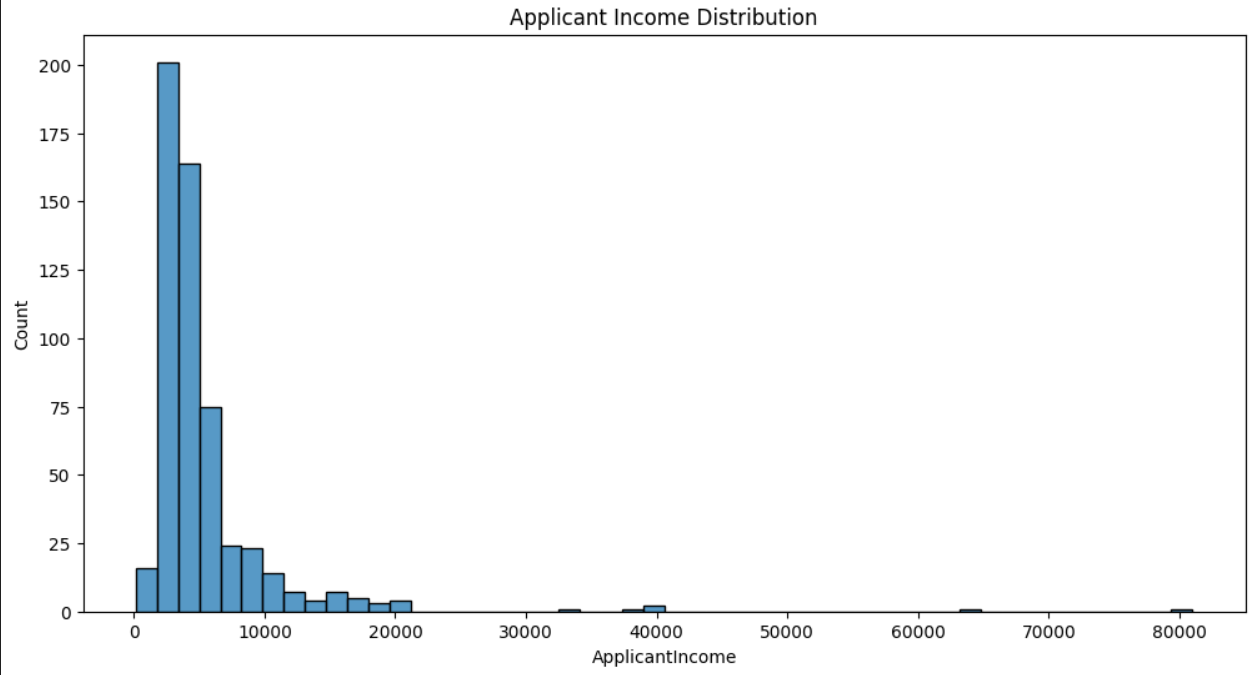
Results:-

Screenshots:-









= **References:-**

<https://www.kaggle.com/datasets/altruistdelhite04/loan-prediction-problem-dataset>

https://youtu.be/p3-7qW\_t5bw?si=5LO8HU3s-ziwG56L