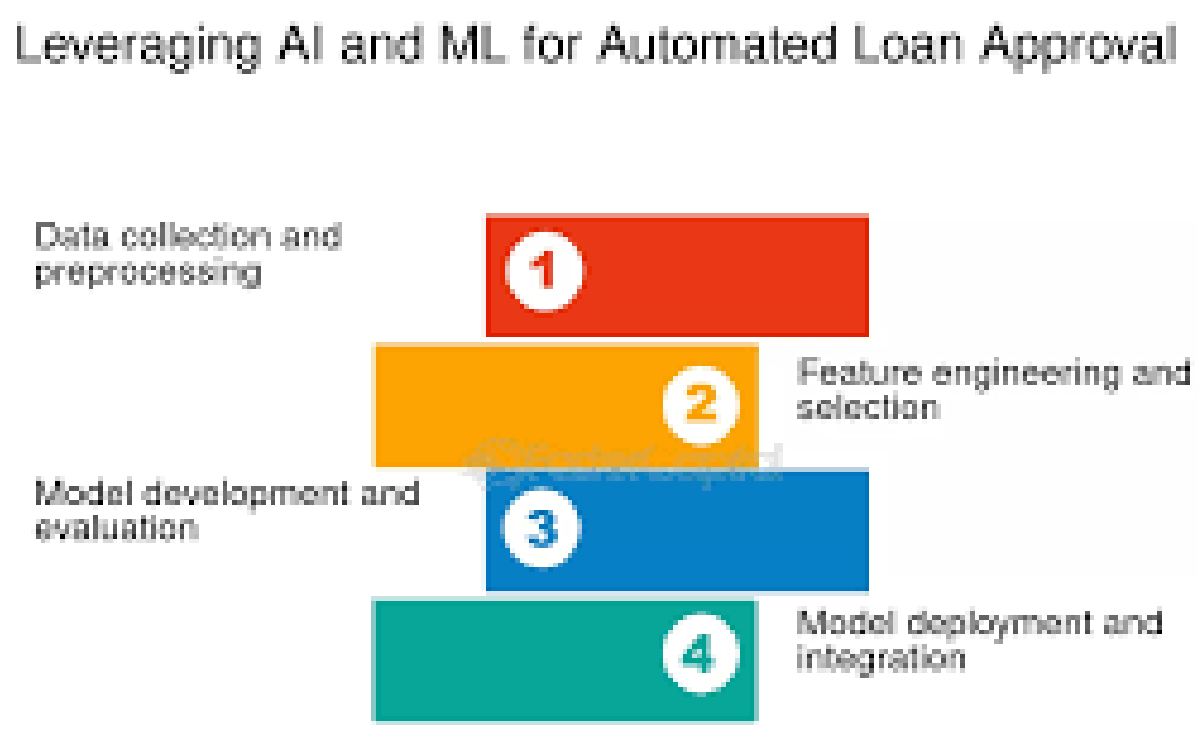
Loan Approval Prediction Model

Introduction



- This project aims to predict loan approval statuses using machine learning models.
- It involves data preprocessing, exploratory data analysis (EDA), model building, and evaluation

Data Source

- Data Used for analysis:

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

The project analyses the past loan datasets by looking at applicants income levels, employment status, loan amount, education and credit history to be decide whether to approve the loan.

These factors have been examined and machine learning applied to be able to automate the loan approval process.

Machine Learning Models

- Dataset was divided into two variables
- X as the features: applicants income levels, employment status, loan amount, education and credit history
- and y as the Loan_Status the target value we want to predict.

Models we will use:

1. Logistic Regression
2. Decision Tree
3. Random Forest

Process of Modelling the Data

1. Importing the model

2. Fitting the model

3. Predicting Loan Status

4. Classification report by Loan Status

5. Overall accuracy

Logistic Regression

Accuracy Score: 0.8216216216216217

Classification Report:

	precision	recall	f1-score	support
0	0.88	0.41	0.56	51
1	0.81	0.98	0.89	134
accuracy			0.82	185
macro avg	0.84	0.69	0.72	185
weighted avg	0.83	0.82	0.80	185

- Loan Approved (Class 1): Precision = 0.83, implying 83% of predicted approvals were correct. Recall = 0.91: The model captured 91% of all actual loan approvals. F1-Score = 0.87: A strong balance of Precision and Recall.
- Loan Rejected (Class 0): Precision = 0.75: implying 75% of predicted rejections were correct. Recall = 0.60: The model only captured 60% of actual loan rejections, indicating room for improvement.
- Overall Accuracy: The model is performing well but has slightly lower recall for predicting loan rejections.

Decision Tree

Decision Tree Accuracy: 0.7837837837837838

Classification Report:

	precision	recall	f1-score	support
0	0.79	0.29	0.43	51
1	0.78	0.97	0.87	134
accuracy			0.78	185
macro avg	0.79	0.63	0.65	185
weighted avg	0.78	0.78	0.75	185

- Class 0 (Rejected Loans): Precision (0.79): 79% of the predicted rejections are correct. Recall (0.29): Only 29% of actual rejections are correctly identified. F1-Score (0.43): Indicates a low balance between precision and recall for this class.
- Class 1 (Approved Loans): Precision (0.78): 78% of the predicted approvals are correct. Recall (0.97): The model captures 97% of actual approvals, which is excellent. F1-Score (0.87): A strong overall performance for this class

Random Forest

```
Random Forest Accuracy: 0.827027027027027

Classification Report:
              precision    recall  f1-score   support

           0       0.83      0.47      0.60        51
           1       0.83      0.96      0.89       134

    accuracy                           0.83       185
   macro avg       0.83      0.72      0.74       185
weighted avg       0.83      0.83      0.81       185
```

- Class 1 (Approved Loans):
- The model performs exceptionally well, with high recall and F1-score. This shows the model can confidently approve loans when the status is "Approved".
- Class 0 (Rejected Loans): The model struggles to identify rejected loans (low recall at 47%). While precision is good, the low recall means many actual rejections are misclassified as approvals. Improvements Over Decision Tree:
- Accuracy improved from 78.3% to 82.7%. Precision, recall, and F1-score for both classes improved slightly.

Summary of Model Results

	Model	Score
2	Random Forest	0.827027
0	Logistic Regression	0.821622
1	Decision Tree	0.783784

```
Final Model Performance on Test Data:
Accuracy: 0.827027027027027

Classification Report:
              precision    recall  f1-score   support

           0       0.83      0.47      0.60        51
           1       0.83      0.96      0.89       134

    accuracy                           0.83       185
   macro avg       0.83      0.72      0.74       185
weighted avg       0.83      0.83      0.81       185
```

- The best model to use is the Random Forest and the same was applied to the test data.
- The model correctly predicts the loan status (approved/rejected) for 82.70% of the test data.
- Class Imbalance: The dataset has 51 rejected loans (minority) and 134 approved loans (majority). This class imbalance causes the model to favor predicting "Approved."“
- There is need to address the class imbalance and explore new features or interactions between existing features to help the model identify rejected loans better.