MD Shibli Mollah

Report summarizing the findings and insights

Loan Approval Prediction Analysis

1. Data Overview

The dataset includes the following features:

"income": The applicant's income.

"credit_score": The applicant's credit score.

"loan_amount": The amount of the loan requested.

"loan_term": The duration of the loan.

"employment_status": The employment status of the applicant (e.g., employed, self-employed).

"loan_approved": The target variable indicating whether the loan was approved (1 for yes, 0 for no).

First Few Rows:

This section provides a glimpse of the data, showcasing different applicant profiles:

income	credit_score	loan_amount	loan_term	employment_status	loan_approved
5849	475	230	120	employed	1
4583	731	128	360	self-employed	0
3000	478	66	360	employed	1
2583	382	120	360	employed	1
6000	654	141	360	employed	1

2. Model Evaluation

KNN Model Metrics:

"Accuracy": 50.41%

"Precision": 50.00%

"Recall"   : 49.18%

"F1 Score" : 49.59%

"Confusion Matrix":

  [[32 30]

   [31 30]]

This indicates the model made 32 true negatives, 30 false positives, 31 false negatives, and 30 true positives.

Decision Tree Model Metrics:

"Accuracy": 47.97%
"Precision": 47.95%
"Recall"   : 57.38%
"F1 Score" : 52.24%
"Confusion Matrix":
   [[24 38]
    [26 35]]

This shows the model made 24 true negatives, 38 false positives, 26 false negatives, and 35 true positives.

3. Feature Importances from Decision Tree:
The importance of each feature in predicting loan approval, ranked from most to least important:

Feature	Importance
income	0.602899
credit_score	0.143495
loan_amount	0.136009
employment_status	0.072445
loan_term	0.045152

Insights:

- The income feature is the most significant predictor of loan approval, indicating that applicants with higher incomes are more likely to have their loans approved.
- Credit score and loan amount also play considerable roles, while employment status and loan term contribute less to the decision-making process.

4. Recommendations Based on Findings
1. "Model Performance":

- Both models have relatively low accuracy, suggesting that the current feature set may not sufficiently capture the factors influencing loan approval. Consider exploring additional features or enhancing data quality (e.g., handling missing values or outliers).
- KNN performed slightly better than the Decision Tree in terms of accuracy and F1 score, but both models have room for improvement.

2. "Feature Engineering":

- Investigate other potential features that could enhance model performance, such as:
    - Previous loan history (if available).
    - Debt-to-income ratio.
    - Other financial metrics or demographics.

3. "Data Balance":

Check the balance of the target variable. If the data is imbalanced (e.g., a lot more approvals than rejections), consider techniques like oversampling the minority class or undersampling the majority class.

4. "Model Tuning":

Experiment with hyperparameter tuning for both KNN (e.g., varying n_neighbors) and Decision Trees (e.g., max_depth, min_samples_split) to see if performance improves.

5. "Deployment Considerations":

If the goal is to deploy a model, ensure continuous monitoring and re-evaluation against new data, as the financial landscape can change.