Loan Approval Prediction - Detailed Report

This report presents the implementation and results of a Machine Learning project for Loan Approval Prediction. The dataset contains information about loan applicants, their financial details, and whether their loan applications were approved or rejected. The workflow includes: 1. Data loading and cleaning. 2. Handling categorical and numerical features. 3. Splitting the dataset into training and testing sets. 4. Training Logistic Regression and Decision Tree classifiers. 5. Evaluating model performance using confusion matrices and classification reports.

Dataset Information
- Total Rows: 4,269
- Total Columns: 13
- Target Column: *loan_status*
- Missing values: None

The dataset includes both categorical features (e.g., education, self_employed) and numeric features (e.g., income_annum, cibil_score, loan_amount).

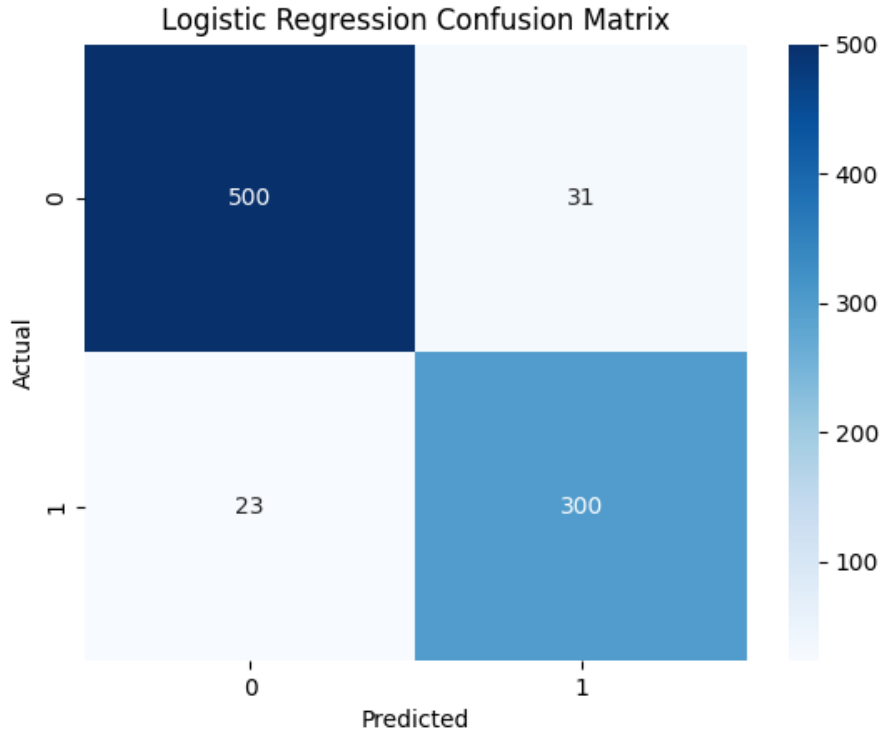
Data Preprocessing
- All column names and string values were stripped of unnecessary spaces.
- Numeric features were standardized using *StandardScaler*.
- Categorical features were one-hot encoded using *OneHotEncoder*.
- A *ColumnTransformer* combined both numeric and categorical transformations.

Model Training
Two models were trained and evaluated:
1. **Logistic Regression** with balanced class weights.
2. **Decision Tree Classifier** with balanced class weights.

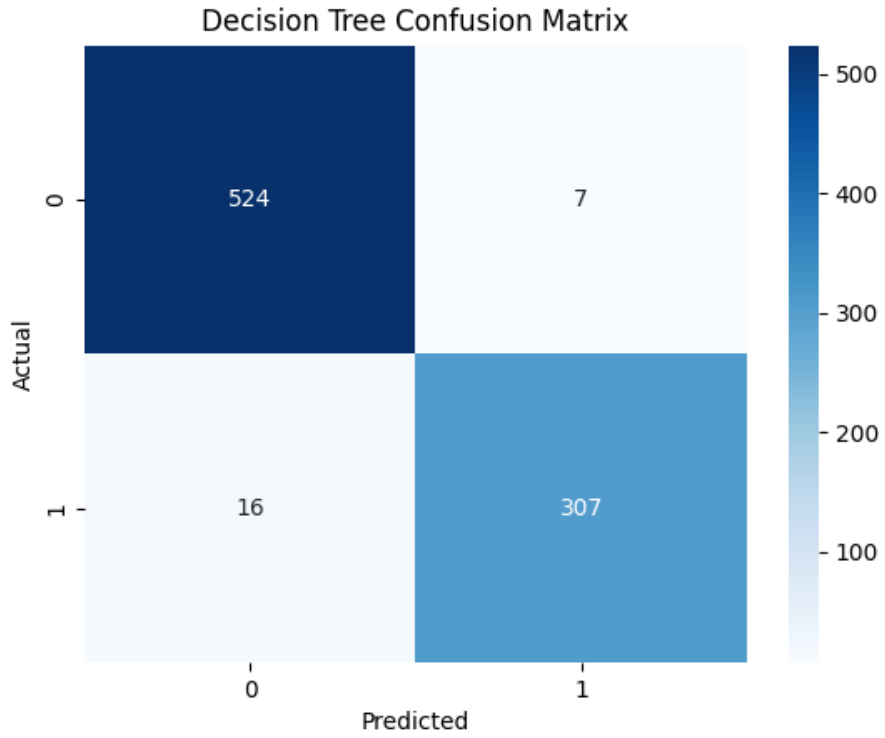
Both models were trained using 80% of the dataset and tested on 20% (stratified split).

Results and Evaluation
The models were evaluated using confusion matrices and classification reports. Below are the confusion matrices obtained from the experiments.

Logistic Regression Confusion Matrix



Decision Tree Confusion Matrix



Model Performance Metrics

Logistic Regression:
- Precision (Approved): 0.96
- Recall (Approved): 0.94
- F1-score (Approved): 0.95
- Precision (Rejected): 0.91
- Recall (Rejected): 0.93
- F1-score (Rejected): 0.92
- Overall Accuracy: 94%

Decision Tree:
- Precision (Approved): 0.97
- Recall (Approved): 0.99
- F1-score (Approved): 0.98
- Precision (Rejected): 0.98
- Recall (Rejected): 0.95
- F1-score (Rejected): 0.96
- Overall Accuracy: 97%

Conclusion

Both Logistic Regression and Decision Tree performed well on the dataset. While Logistic Regression provided strong balanced results, the Decision Tree achieved slightly higher accuracy and recall for approved loans. Decision Trees may overfit on certain datasets, so Logistic Regression could generalize better in real-world scenarios with new data. This project demonstrates how machine learning can support financial institutions in making efficient and data-driven loan approval decisions.