

Loan Approval and Credit Risk Classification Analysis

Group 24 - Danny Lynn, Sanjit Saji, Yelena Merriam, Zecheng Wang

Why Analyze Loan Approval?

- Financial institutions rely on efficient and fair loan approval processes.
- Understanding key predictors helps reduce risk and improve decision-making.
- This project explores a dataset of 45,000 loan applications to predict approval outcomes.

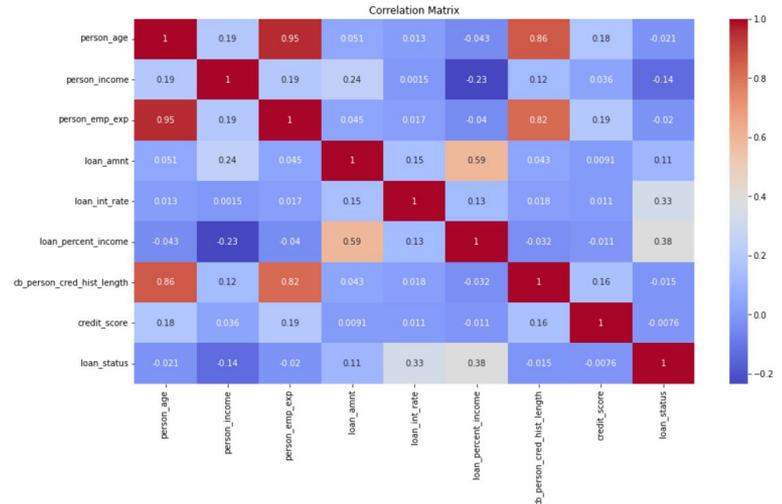
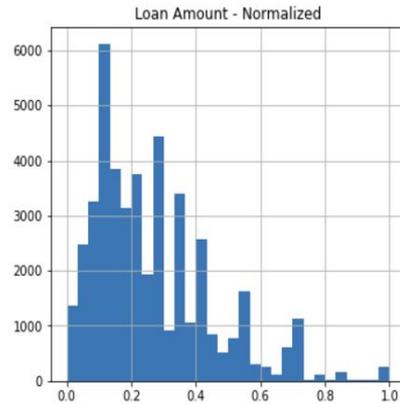
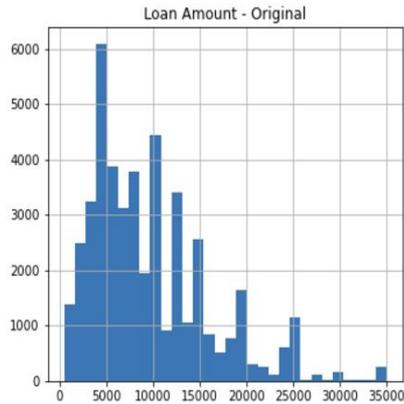


Dataset Highlights

- Size: 45,000 records, 14 variables.
- Key Variables:
 - Demographics: Age, Gender, Education, Income.
 - Loan Details: Amount, Interest Rate, Intent.
 - Financial Indicators: Credit Score, Loan Defaults.
 - Target: Loan Status (1 = Approved, 0 = Rejected).

Methodology - 1 (Data Preparation and Analysis)

- Cleaned data by handling outliers, normalizing numerical features, and encoding categorical variables.
- Conducted Exploratory Data Analysis (EDA) to uncover patterns like the relationship between income and loan approval.

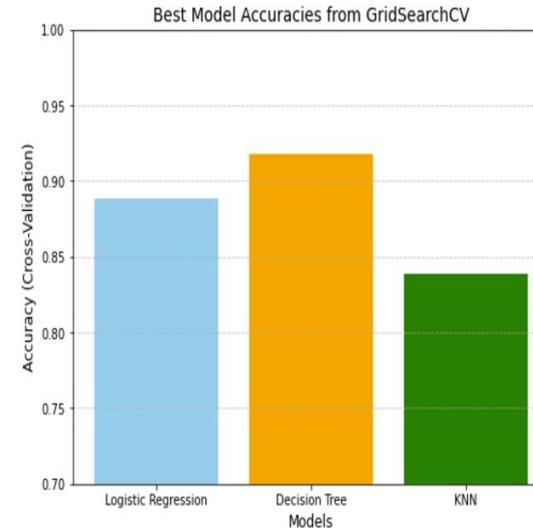
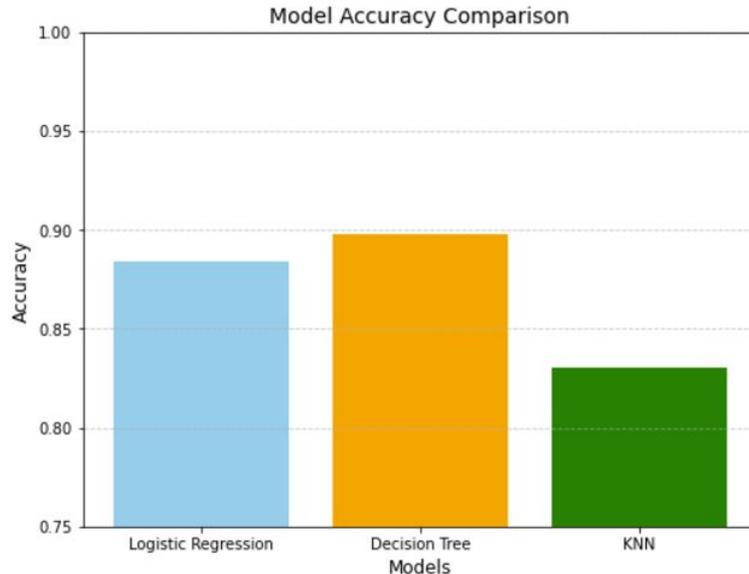


Methodology - 2 (Modeling and Evaluation)

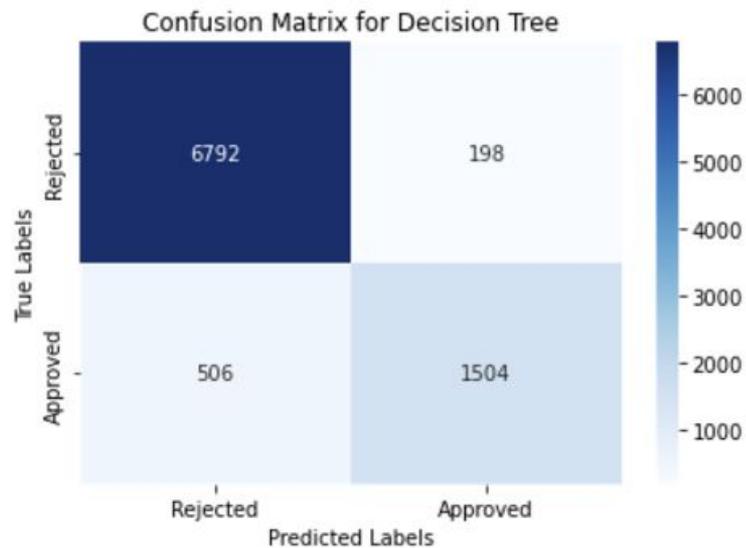
- The models were evaluated on the test set using accuracy.
- Decision Tree performed best, followed by Logistic Regression and KNN.

Logistic Regression Accuracy: 0.88
Decision Tree Accuracy: 0.90
KNN Accuracy: 0.83

Best Logistic Regression: `LogisticRegression(C=0.01, max_iter=5000, random_state=42)` with accuracy: 0.89
Best Decision Tree: `DecisionTreeClassifier(max_depth=10, min_samples_leaf=2, random_state=42)` with accuracy: 0.92
Best KNN: `KNeighborsClassifier(metric='manhattan', n_neighbors=9)` with accuracy: 0.84



Conclusion



Evaluation Metrics for Best Decision Tree:

F1 Score: 0.81

Recall: 0.75

Classification Report:

	precision	recall	f1-score	support
Rejected	0.93	0.97	0.95	6990
Approved	0.88	0.75	0.81	2010
accuracy			0.92	9000
macro avg	0.91	0.86	0.88	9000
weighted avg	0.92	0.92	0.92	9000