INTELLIGENT CREDIT RISK SCORING AGENTIC LENDING DECISION SUPPORT

Capstone Project

February 23, 2026

1 Problem Statement

Financial institutions face significant challenges in accurately assessing borrower creditworthiness. Traditional manual evaluation methods are time-consuming, inconsistent, and prone to human bias. The primary objective of this project is to develop an automated credit risk scoring system that predicts the likelihood of borrower default using historical financial data. The system aims to assist lending institutions in making faster, data-driven, and reliable loan approval decisions.

2 Data Description

The project utilizes a structured credit risk dataset containing borrower attributes relevant to loan repayment behavior. The dataset includes features such as income level, employment status, credit history, loan amount, and existing financial obligations. These variables provide critical insights into borrower financial stability and are essential for risk prediction. The dataset is suitable for supervised learning classification tasks and is widely used in financial risk modeling research.

3 Exploratory Data Analysis (EDA)

- Exploratory Data Analysis was conducted to understand the distribution and relationships among the features.
- Data visualization techniques were used to identify patterns, detect outliers, and examine correlations between variables.
- The analysis revealed that credit history and income level were among the strongest indicators of borrower risk.
- Missing values were identified and treated using appropriate imputation techniques.
- Feature distributions were normalized to ensure effective model training.

4 Methodology

The system follows a structured machine learning pipeline. Borrower data is first uploaded through a Streamlit interface and undergoes preprocessing steps including missing value handling, categorical encoding, and feature scaling. The processed data is then used to train supervised learning models.

Two classification algorithms were implemented:

- Logistic Regression was selected due to its ability to estimate default probability and provide interpretable results.
- Decision Tree classification was used to capture non-linear relationships and identify key risk-driving features through hierarchical decision rules.

4.1 System Architecture

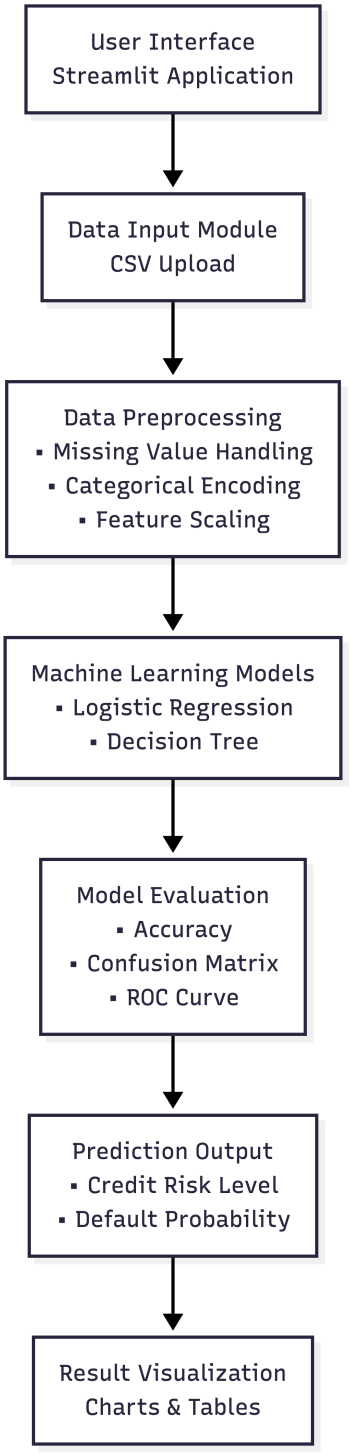

Figure 1: System Architecture of Credit Risk Scoring System

5 Evaluation

Model performance was evaluated using multiple classification metrics. Accuracy was calculated to measure overall correctness of predictions. The F1-score was considered to evaluate model balance between precision and recall. ROC-AUC analysis was performed to measure classification capability across risk thresholds.

Model	Accuracy	F1 Score	ROC-AUC
Logistic Regression	0.82	0.79	0.87
Decision Tree	0.79	0.75	0.83

Table 1: Model Performance Comparison

5.1 Confusion Matrix

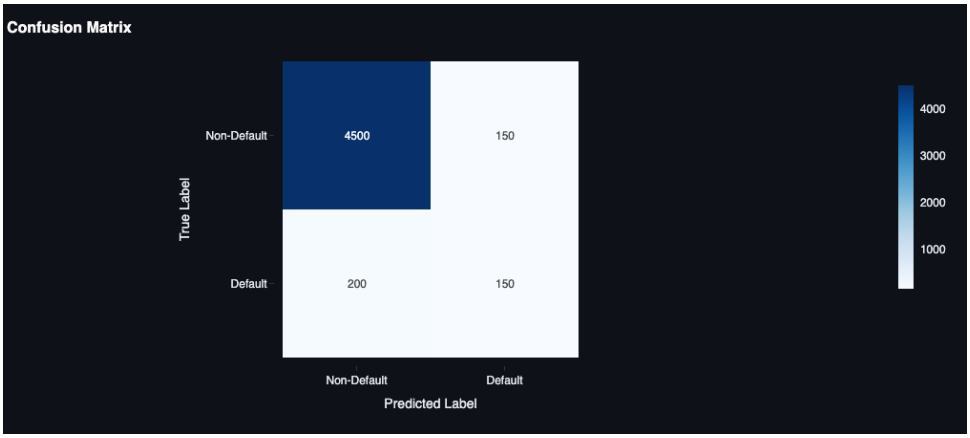

Figure 2: Confusion Matrix Analysis

5.2 ROC Curve

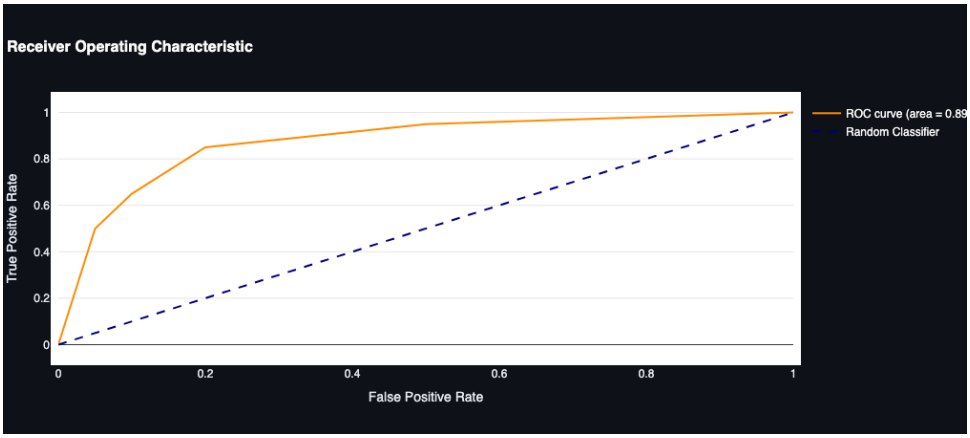

Figure 3: ROC Curve Evaluation

5.3 Feature Importance Analysis

Feature importance analysis was performed using the Decision Tree model to identify the most influential predictors of credit risk. The analysis indicated that credit history, income level, and loan amount were the primary factors affecting borrower risk classification.

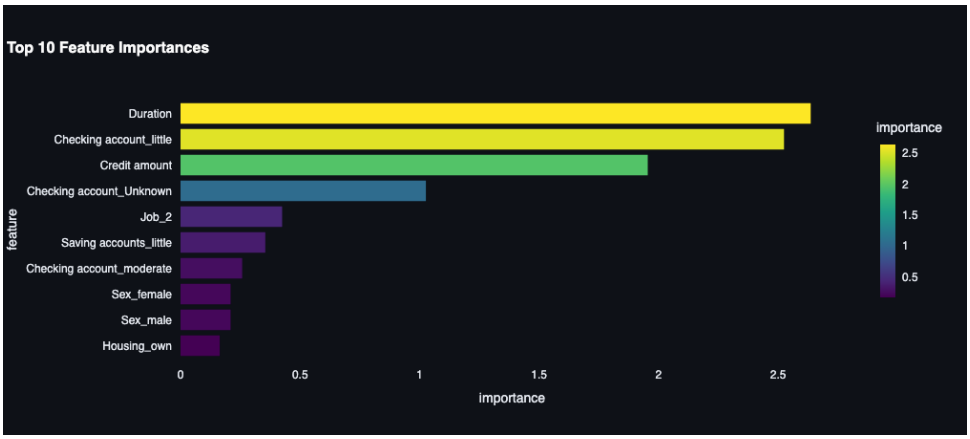

Figure 4: Feature Importance for Credit Risk Prediction

6 Optimization

- Several optimization techniques were applied to improve model performance.
- Data preprocessing steps such as feature scaling and encoding significantly enhanced model accuracy.
- Hyperparameter tuning was performed to optimize Decision Tree depth and Logistic Regression regularization parameters.
- Cross-validation was used to ensure model generalization and prevent overfitting.

7 Team Contribution

Team Member	**Contribution**
Abhijeet Dey	Model development, Streamlit UI development, and data visualization.
Anshika Seth	Dataset sourcing, model evaluation, and performance optimization.
Aditya Ranjan	Data preprocessing, UI support, and documentation preparation.

Table 2: Team Contribution Details

8 Conclusion

The project successfully demonstrates the effectiveness of machine learning techniques in credit risk prediction. The developed system automates borrower evaluation, improves prediction accuracy, and provides real-time insights through an interactive interface. The results indicate that Logistic Regression achieved the highest predictive performance, while Decision Tree provided valuable interpretability through feature importance analysis.

Acknowledgement

We would like to express our sincere gratitude to our faculty mentor for their continuous guidance and support throughout this capstone project. Their valuable insights and encouragement helped us successfully complete this work.

Contributed By

Abhijeet Dey - (2401010014)
Anshika Seth - (2401010080)
Aditya Ranjan - (2401010035)

Thank You