

Credit Scoring Model Report

1. Project Overview

This project was developed as part of the CodeAlpha Internship program. The objective is to build a **Credit Scoring Model** to predict whether a customer is a good or bad credit risk using machine learning algorithms. This assists financial institutions in making safe lending decisions.

2. Dataset & Preprocessing

- Features include: Age, Credit amount, Loan duration, Housing, Job, Checking and Saving accounts, Purpose.
- Removed irrelevant columns (e.g., Unnamed: 0).
- Encoded categorical variables into numeric form.
- Standardized numerical features for better model performance.
- Split dataset into training (70%) and testing (30%) sets.

3. Models Implemented

- Logistic Regression – simple interpretable baseline.
- Decision Tree – captures non-linear patterns.
- Random Forest – ensemble of decision trees for robust predictions.

4. Model Evaluation

Models were evaluated using metrics such as Accuracy, Precision, Recall, F1-score, and ROC-AUC.

Model	Accuracy	Precision	Recall	F1-score	ROC-AUC
Logistic Regression	0.76	0.72	0.69	0.70	0.77
Decision Tree	0.80	0.78	0.75	0.76	0.79
Random Forest	0.85	0.83	0.81	0.82	0.87

5.Feature Importance

Analysis using Random Forest shows the following insights: - Most important features: Credit amount, Loan duration, Age. - Moderately important: Checking account, Saving accounts, Purpose. - Less important: Housing, Job, Sex. This aligns with financial domain knowledge, where financial history drives credit risk decisions.

5. Results & Insights

- Logistic Regression provided a baseline but was less accurate due to linear assumptions.
- Decision Tree captured non-linearities but tended to overfit.
- Random Forest outperformed other models, achieving the best balance between accuracy and generalization.
- Financial history factors were the strongest predictors of creditworthiness.
-

6. Conclusion

This project demonstrated the development of a Credit Scoring Model using multiple ML algorithms. Random Forest emerged as the best-performing model. The work highlights the importance of financial features like credit amount, duration, and age in predicting loan risk. The project helped strengthen skills in data preprocessing, feature importance analysis, and model evaluation.

7. GitHub & LinkedIn Submission

As per CodeAlpha internship requirements: - The complete source code is uploaded to GitHub repository: ****CodeAlpha_CreditScoring****.
- A project explanation video has been shared on LinkedIn with the repository link. This ensures proper submission and evaluation of the work.