

CENG 562 Project Proposal

1. Project Title: Interpretable Model Comparison for Bank Credit Approval Prediction

2. Group Members:

Oğuzhan ALPERTÜRK – e231575@metu.edu.tr

Arda ÇAVUŞOĞLU – e244824@metu.edu.tr

Yağızcan PANÇAK – e269776@metu.edu.tr

3. Problem Description: The goal of this project is to predict whether a bank customer's credit application will be approved based on their demographic and financial attributes. Credit approval is an important decision-making problem in the financial sector, where transparency and interpretability are essential. Many real-world systems require not only high predictive accuracy but also understandable decision criteria to ensure fairness and trust. We aim to analyze the trade-off between predictive performance and interpretability across classical machine learning models.

4. Dataset: We will use the UCI Credit Approval Dataset, a publicly available dataset containing customer attributes such as age, income, employment status, and past credit history. However, for the sake of privacy, these variables have been switched with random symbols. The dataset includes both categorical and numerical features and a binary target label indicating approval or rejection.

5. ML Methods: We plan to apply and compare the following classical machine learning models:

- Logistic Regression: a linear and interpretable baseline.
- Decision Tree: rule-based model offering transparent decision paths.
- Random Forest: ensemble approach with higher performance and feature importance analysis.

The project will focus on both predictive accuracy and model interpretability, using visualization of feature importances and decision rules.

6. Evaluation: We will evaluate models using Accuracy, ROC-AUC, and interpretability-based metrics. Interpretability will be examined through:

- Coefficient analysis (for Logistic Regression)
- Decision rules and tree visualization (for Decision Tree)
- Feature importance (for Random Forest)

All models will be evaluated using 5-fold cross-validation to ensure generalization.

7. References:

1. Quinlan, J. R. "Induction of Decision Trees." Machine Learning, 1986.
2. Breiman, L. "Random Forests." Machine Learning, 2001.
3. UCI Machine Learning Repository: Credit Approval Dataset.
4. Ribeiro, M. T. et al. "Why Should I Trust You?" KDD, 2016.

8. Work Plan:

- 1-2 Week: Literature review, dataset preprocessing.
- 1-2 Week: Implement and train baseline models.
- 1-2 Week: Evaluate performance and interpretability.
- 1 Week: Final analysis and report preparation.