

# Predicting Probability of Default for First-Hand Car Loan Applications

Students: İsmail Berat Düzenli, Mehmet Emin Bayrak, Fatih Arda Zengin, Buse Keleş

Faculty Member : Öznur Taştan

Company Advisor : Burak Yüksel

## Executive Summary

- **Objective:** Develop an explainable and accurate Probability of Default (PD) model to predict credit risk for first-hand car loan applications.
- **Importance:** Helps financial firms maximize profits by lending to creditworthy applicants while avoiding likely defaulters.
- **Challenges:** Imbalanced dataset, limited domain knowledge in finance, balancing accuracy and explainability.

## Problem Statement

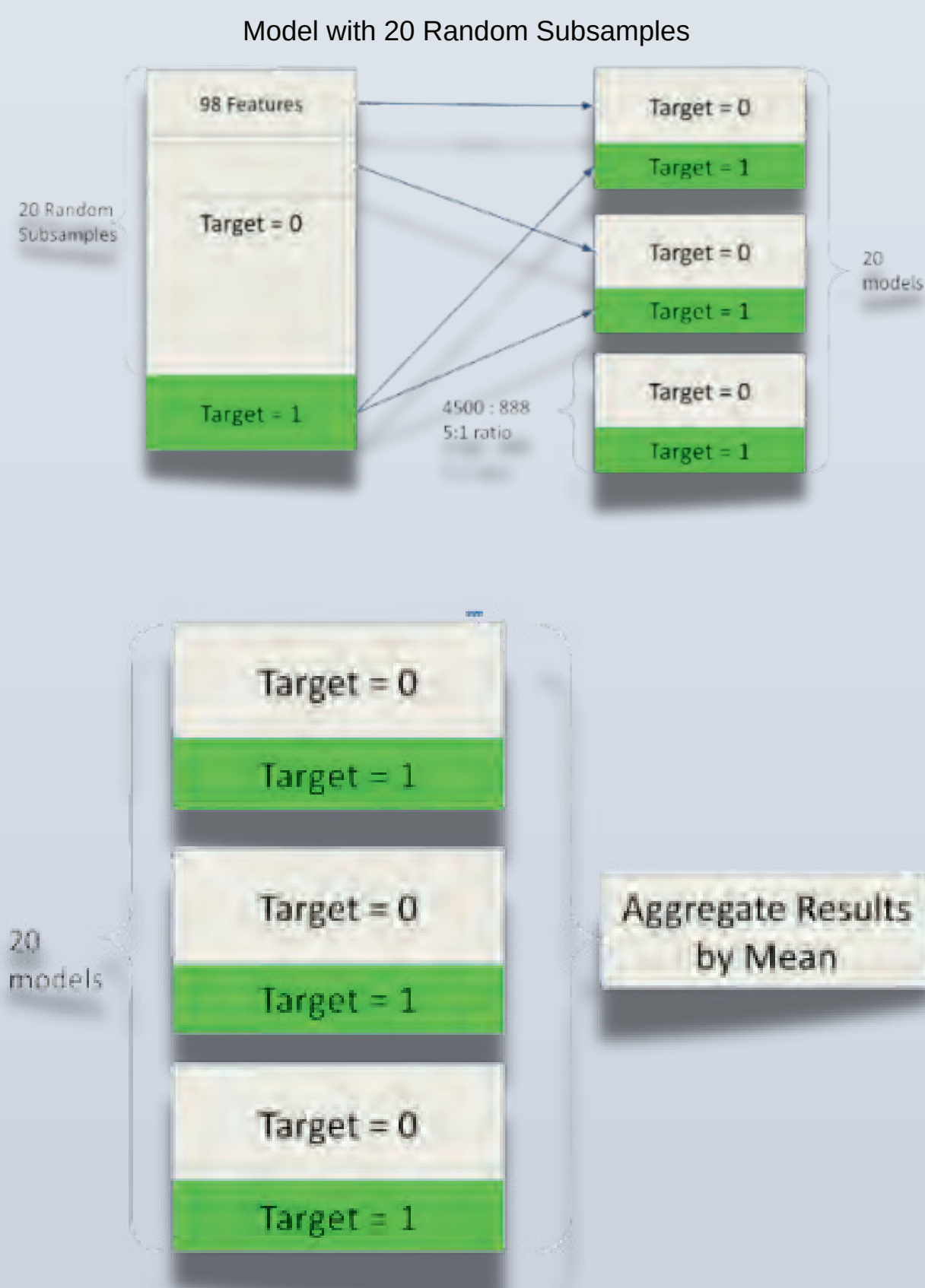
- **Dataset:** 96,443 instances, 1,096 targets, 304 features.
- **Focus:** Identifying influential features, creating an explainable model, and enhancing model accuracy.

## Data Preprocessing

- Target variable creation.
- Data split: 80% training, 20% testing.
- Date conversion, categorical encoding.
- Addressing abnormalities, introducing new features.
- Log transformation of skewed features.

## Model Development

- **Algorithm:** LightGBM
- **Approach:**
  - Undersampled dataset into 20 subsamples with a 5:1 ratio.
  - Trained 20 LightGBM models.
  - Cross-validation for parameter tuning.



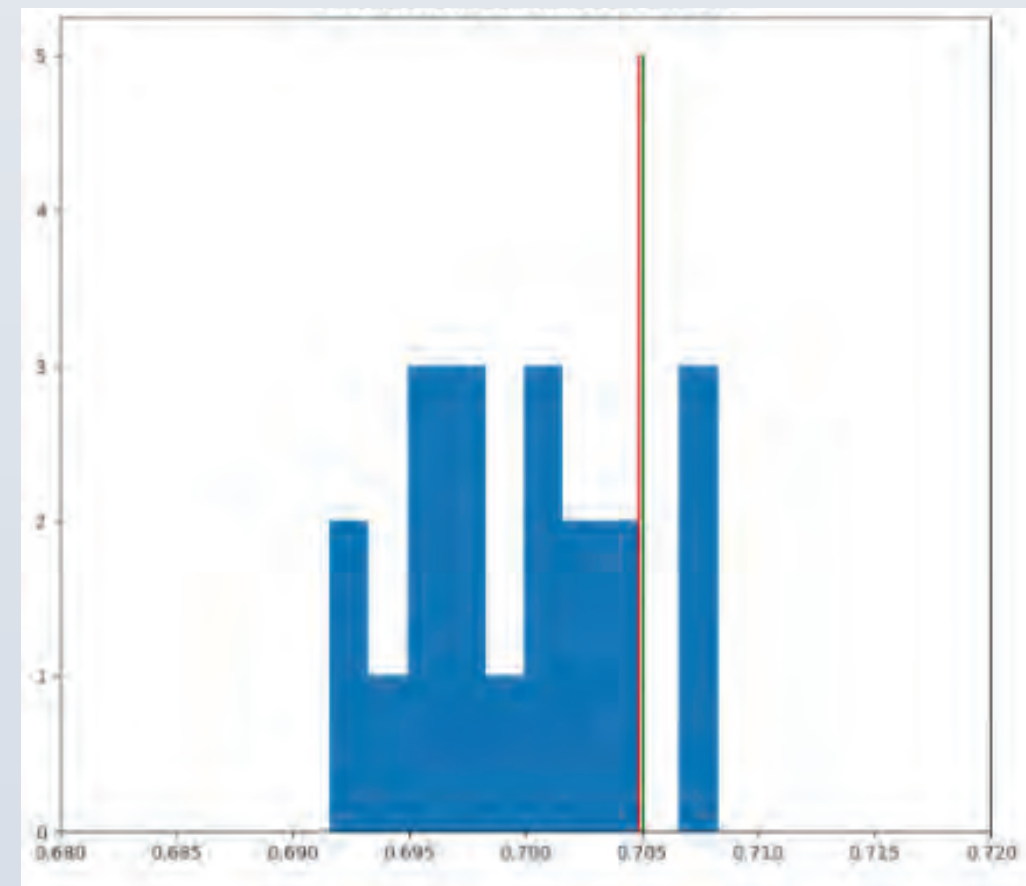
## Explainability Enhancement

- **Techniques:**
  - SHAP (feature importance, local/global interpretability).
  - LIME (local explanations).
  - DiCE (counterfactual generation).
- **Approach:** Generating counterfactual instances to explain model decisions.

## Results

- **Performance:**
  - Gini coefficient: 0.7
  - F1 score: 0.125 (low, highlighting class imbalance challenges).
  - Precision-Recall Curve: Trade-offs between precision and recall.
  - ROC Curve: High AUC indicating strong diagnostic capability.
- **Discussion:** Identified challenges and proposed improvements (e.g., alternative ensemble techniques, hyperparameter tuning, additional data sources).

Gini Score On Submodels on Test Set



## Impact

- **Innovation:** Combines high accuracy with explainability, addressing "black-box" issue in ML models.
- **Advantages for Koç Finans:**
  - Improved credit risk assessment accuracy.
  - Reduced need for manual assessment.
  - Potential compliance with future regulations.
  - Competitive market advantage.

## Conclusion and Future Work

- **Achievements:** Successful predictive performance, implementation of counterfactual analysis.
- **Future Steps:** Address limitations, refine feature selection, improve precision and recall, enhance model explainability.