Project Proposal

Credit Eligibility Prediction Using Machine Learning

Maria Chmite   Hind Bertit   Rayane Mestari

Supervised by: Dr. Yousra Chtouki

Fall 2025
School of Science & Engineering
Al Akhawayn University

1   Problem Definition and Motivation

Credit approval in retail banking is traditionally based on manual rules and human judgment, which can lead to slow processing times, inconsistent decisions, and potential bias. As the volume of loan applications increases, banks require automated, scalable, and data-driven decision-making tools.

The goal of this project is to leverage machine learning techniques to build an automated system capable of predicting whether a client is eligible for a consumer credit product. Such a system can improve decision consistency, reduce operational costs, and support fairer credit decisions.

2   Related Work Overview

Previous studies in credit risk modeling show that ensemble-based machine learning methods, particularly Random Forest and Gradient Boosting models (XGBoost), consistently outperform traditional statistical approaches such as logistic regression. Recent literature also highlights the lack of complete end-to-end implementations that include structured methodologies (e.g., CRISP-DM) and real-world deployment considerations.

Based on this review, this project adopts XGBoost as the primary model and emphasizes a full CRISP-DM workflow, including deployment and evaluation using industry-standard metrics.

3   Project Objectives

The main objectives of this project are:

- To build a machine learning model that predicts credit eligibility using client demographic, geographic, and product-ownership data;

- To compare multiple machine learning models and justify model selection based on empirical results;

- To deploy the final model in an interactive web-based application accessible to non-technical users.

4   Machine Learning Task

This project focuses on a supervised learning task:

- **Type:** Binary classification;

- **Target variable:** Credit eligibility (approved / not approved).

5   Hypotheses

The project tests the following hypotheses:

- **H1:** XGBoost will outperform Random Forest and Logistic Regression in terms of ROC-AUC and F1-score.

- **H2:** Applying feature engineering and class imbalance handling will significantly improve recall for eligible clients.

6   Dataset Overview

The project uses proprietary retail banking data containing anonymized client information, including demographic attributes, product usage, and geographic features. The dataset exhibits severe class imbalance, requiring dedicated preprocessing and resampling techniques.

7   Proposed Methodology

The project follows the CRISP-DM methodology:

- Business understanding;

- Data understanding and preparation;

- Modeling and evaluation;

- Deployment of the final solution.

8   Expected Outcomes

Expected outcomes include:

- A validated machine learning model for credit eligibility prediction;

- A deployed interactive dashboard for real-time inference;

- A documented evaluation of model performance and limitations.