

Thesis Topic Registration Form

Student's Data:

Student's Name: Kafi MD Abdullah Hel

Student's Neptun code: N06WMD

Educational Information:

Training programme: Computer Science BSc

I have an internal supervisor

Internal Supervisor's Name: *Md. Easin Arafat*

Supervisor's Home Institution: *Department of Data Science and Engineering*

Address of Supervisor's Home Institution: *1117, Budapest, Pázmány Péter sétány 1/C.*

Supervisor's Position and Degree: *PhD Candidate*

Thesis Title: Interactive Machine Learning Model Explanation Web Tool for Credit Risk Prediction

Topic of the Thesis:

(Upon consulting with your supervisor, give a 150-300-word-long synopsis of your planned thesis.)

Problem to Be Solved

Financial institutions struggle to accurately predict whether loan applicants will repay or default. Traditional credit scoring methods rely on rigid rules that miss intricate patterns in applicant data, while modern machine learning models though more accurate function as "black boxes" with opaque decision-making processes. This project addresses the need for a credit risk prediction system combining high accuracy with transparent, understandable explanations.

Motivation

Explainable AI in lending serves multiple critical purposes. Regulatory bodies increasingly demand that financial institutions justify automated decisions, particularly those affecting individuals' access to credit. Transparency builds trust with applicants who deserve to understand approval or rejection reasons. Interpretable models help institutions identify potential biases and ensure fair treatment across demographic groups. This project demonstrates practical skills in data science and web development while addressing real-world ethical and compliance challenges.

Where It Is Applied

The platform benefits commercial banks evaluating personal loans, microfinance organizations serving underbanked populations, and digital lending platforms processing high-volume applications. Beyond commercial use, it serves educational purposes helping students and researchers understand how machine learning models weigh factors when assessing creditworthiness.

How It Will Be Implemented

Using publicly available credit datasets, multiple classification algorithms (Logistic Regression, Random Forest, XGBoost) will be trained and evaluated. The best-performing model will be integrated into a Flask web application where users input applicant information and receive instant predictions. SHAP or LIME frameworks will generate visual explanations showing how individual features influence decisions, compared against traditional scoring benchmarks.

Expected Outcomes

The project delivers a functional web tool providing real-time credit risk predictions with clear visual explanations. Users will understand which factors drive decisions, while performance metrics demonstrate improved accuracy over conventional approaches. Success metrics include prediction accuracy, explanation clarity, and enhanced user trust in automated lending decisions.

Budapest, 2025. 10. 05.