

Reflexive Report: Credit Default Prediction Using Machine Learning

Daniz Hajizada

April 2025

1 Introduction

For my final year project, I proposed and developed a machine learning pipeline to predict default using real-world banking data from 30,000 Taiwanese credit card users. I worked independently, without a preset schedule or fixed expectations, meaning I had to design everything from scratch — from technical implementation and report structuring to managing my own time and workflow.

Initially, I had only a basic understanding of the banking environment and the nature of implementation of machine learning in such a field. To strengthen my knowledge, I had to dive into the industry, study financial history, and how risk is managed. For example, my initial simple thoughts were that banks could be bankrupt if not properly giving out loans, but through further research, I realised bankruptcy processes in banks are far more complex and structured. Some of the reading I did for technical material for my literature review highlighted the importance of a Z-score, which in simpler terms is an algorithm which can predict bankruptcy. This all meant that I was in the right direction. In short, this process of gaining knowledge and putting it into a good use involved a lot of independent reading, technical experimentations, and supervisor guidance.

2 Key Experiences and Feelings

At the start, I felt overwhelmed by the complexity of banking systems and uncertain about whether I could build something useful without a detailed specification and contribute in some sort of way to the overall research field. There were moments of frustration, especially when I couldn't figure out how to create meaningful features from repayment data or when a model, I expected to perform well, overfitted badly at the end. However, there were also moments of excitement — when I got to see the feature importance chart because I was genuinely interested and curious about which features could actually impact the decision making and the thought process of a machine while making such decisions. Working with my supervisor helped relieve stress. His metaphors — like

”having a recipe but not the right ingredients” — made me rethink a couple of designs and put me in the right direction. His comments were extremely helpful and helped me improve a lot. Sometimes I could not see the full picture and I did not understand what I needed to do in order to achieve completeness while writing up my report. He provided me with a set of guiding questions, which helped me determine what should be included, omitted, or elaborated upon in the text to ensure the reader understands my point.

One of the key experiences, as mentioned earlier, was working with real-time banking data, as it was interesting to see which features banks usually have of users. One critical decision during the project came when choosing to include sequential delay scores and binned categories as part of my feature engineering strategy. This choice was based on both intuition and ongoing analysis, but it wasn’t straightforward. At several points, I had to exclude features when they caused multicollinearity or weakened model performance.

3 Analysis

If I had the opportunity to start over again, I would draft a rough initial roadmap to guide my stages better and reduce early uncertainty. I would also invest more time into exploring relevant literature earlier to ensure that feature creation was more informed and avoid building redundant variables. These two were a major problem back then, which delayed my overall project progression.

4 Personal Evaluation and Future Value

Overall, I believe I performed moderately well, particularly considering the self-driven and open-ended nature of the project. If I were to redo it, I would approach some stages with greater strategy and technical preparation, but I am proud of the final outcome.

My report walks the reader through the entire modelling process - from data overview and simple preprocessing stages to advanced stacked model predictions - and can serve not only as a reminder of my work but as a useful resource for others doing similar projects.

Importantly, this project confirmed my interest in the application of machine learning within finance and proved to me that I can independently manage my time and complete a complex research problem given the specific deadline. The experience has also given me a portfolio-worthy project that I can confidently reference when applying for positions in data science.