

<Capstone Project Proposal>

Corporate Bankruptcy Prediction with Financial Statement Data

1. Problem Statement

Loan providers face significant challenges in accurately assessing the risk of corporate bankruptcy. Traditional methods, often reliant on subjective evaluations and limited data points, can lead to inaccurate assessments and substantial financial losses due to loan defaults. I plan to build a quantitative model to predict corporate bankruptcy, fulfilling the following criteria:

- Accurately predict the bankruptcy of all companies by over 90%
- Correctly identify bankrupt companies over 80% which are predicted to be bankrupt
- Predict over 75% of all bankrupt companies

2. Context

Commercial banks need to improve their ability to accurately predict corporate bankruptcy because it allows them to:

- **Reduce loan losses:** By identifying companies at high risk of bankruptcy, AB Bank can avoid lending altogether or adjust loan terms, minimizing potential defaults
- **Improve resource allocation:** Resources can be prioritized for monitoring and potentially restructuring loan agreements with companies flagged for high bankruptcy risk
- **Enhance overall risk management:** A reliable corporate bankruptcy prediction model strengthens a financial institution's risk management framework, promoting financial stability

3. Criteria for Success

The success of this project will be measured by the following criteria:

- **Model Accuracy:** Achieve at least 90% accuracy in correctly predicting overall bankruptcy.
- **Precision:** At least 80% of positive predictions (companies flagged for bankruptcy) are true positives.
- **Recall:** At least 75% of actual bankruptcies are identified by the model.

4. Scope of Solution Space

The project will focus on developing a binary classification model that predicts whether a company will go bankrupt within a specific timeframe (e.g., one year). The model will leverage a dataset of financial ratios derived from public financial statements.

5. Constraints

- **Limited Time Horizon:** The features extracted from the dataset represent financial ratios for a period immediately preceding the event (bankruptcy). This restricts the model's ability to incorporate changes in financial ratios over a more extended period, such as the previous three years. These changes could provide valuable insights into a company's financial health and bankruptcy risk.
- **Absence of Timestamps:** The absence of timestamps in the dataset prevents the model from considering the impact of business cycle fluctuations on bankruptcy events. Business cycles can significantly influence a company's financial performance and bankruptcy risk. Incorporating time-related information could enhance the model's ability to capture these dynamics.

6. Stakeholders

- **The Executives**
- **The Head of Risk Management**

7. Data Sources

- **UC Irvine Machine Learning Repository**(<https://archive.ics.uci.edu/dataset/572/taiwanese+bankruptcy+prediction>)

8. Next Steps

Building upon this proposal, the following steps will be undertaken:

- **Data Acquisition & Preprocessing:** Collect and clean financial data for the target companies
- **Exploratory Data Analysis (EDA):** Analyze the data to identify patterns and correlations between financial ratios and bankruptcy.
- **Feature Engineering:** Create new features from existing data if necessary.
- **Model Development & Training:** Train and compare multiple machine learning models for binary classification (e.g., Logistic Regression, Random Forest, Decision Trees).
- **Model Evaluation & Selection:** Evaluate model performance based on accuracy, precision, recall, and F1 score. Select the best-performing model.
- **Model Interpretation:** Analyze the model's predictions to understand the key financial factors influencing bankruptcy risk.
- **Project Documentation:** Prepare a comprehensive report summarizing the methodology, results, and recommendations.