**FE-582: Foundations of Financial Data Science**

**Group 9:** Tashveen Kaur, Abhinav Ganguly, Clinton Nwokike

# Project Overview:

**Project Description** March 21, 2023

The objective of the project is to analyze financial market data by applying data collection, data preparation, feature extraction, data cleaning, analytical processing and algorithms. Data clustering, data classification, outlier analysis, and data mining techniques may be implemented. The data can be obtained from the financial markets (stocks, financial statements, etc.) or from text data sources (twitter, news, etc.). You may implement new methods and argue the advantage of them over traditional methods.

# Problem Statement:

Can we predict bankruptcy using Machine Learning?

# Background:

One of the primary objectives of credit risk assessment is predicting business insolvency. Particularly since the financial crisis of 2007–2008, it has elevated in importance for most financial institutions, professionals, and scholars.

Bankruptcy or corporate failure can hurt both the individual company and the global economy. Business practitioners, investors, governments, and academic academics have long sought techniques to identify the risk of business failure in order to reduce the economic losses associated with bankruptcies.

To summarize, forecasting insolvency is an important task for many financial institutions. The purpose is to forecast the likelihood of a corporation going bankrupt. Effective prediction models are required by financial institutions in order to make suitable lending decisions.

Several models for predicting bankruptcy were able to be developed thanks to recent developments in machine learning (ML).

A recent glance at these bankruptcy activities is the bankruptcy of Silicon Valley Bank recently and Signature Banks. These recent bank failures led us to solve the issue that was in the mind of various investors about the prediction of these banks and companies and to create a model that could accurately predict the failure.

# Our Approach:

Let's try to predict whether a given company will go **bankrupt** or not. We will use a variety of classification models to make our predictions. The data set we have chosen uses various ratios like (ROA, Gross Margin, Operating Profit, etc.)and other quantitative metrics like ( Net Income or Equity to Liability ) that are used as variables. We will utilize different models to test our approach.

# Data:

The data was collected from the Taiwan Economic Journal from 1999 to 2009. Bankrupt companies were identified based on the business regulations of the Taiwan Stock Exchange.

Source - Deron Liang and Chih-Fong Tsai, deronliang '@' gmail.com; cftsai '@' mgt.ncu.edu.tw, National Central University, Taiwan  
The data was obtained from UCI Machine Learning Repository: <https://archive.ics.uci.edu/ml/datasets/Taiwanese+Bankruptcy+Prediction>

**Data Volume**:

* Time Period: 1999–2009

**2. Data Variety**:

* Across different industries (electronic manufacturing, retail, shipping, tourism...)
* Each industry has sufficient amount of companies in the similar size in order to do the comparison
* 95 features (X1-X95, business regulations of Taiwan Stock Exchange)
* 1 label (bankrupt or not)

**Description of Variables**

Updated column names and description to make the data easier to understand (Y = Output feature, X = Input features)

Y - Bankrupt?: Class label  
X1 - ROA(C) before interest and depreciation before interest: Return On Total Assets(C)  
X2 - ROA(A) before interest and % after tax: Return On Total Assets(A)  
X3 - ROA(B) before interest and depreciation after tax: Return On Total Assets(B)  
X4 - Operating Gross Margin: Gross Profit/Net Sales  
X5 - Realized Sales Gross Margin: Realized Gross Profit/Net Sales  
X6 - Operating Profit Rate: Operating Income/Net Sales  
X7 - Pre-tax net Interest Rate: Pre-Tax Income/Net Sales  
X8 - After-tax net Interest Rate: Net Income/Net Sales  
X9 - Non-industry income and expenditure/revenue: Net Non-operating Income Ratio  
X10 - Continuous interest rate (after tax): Net Income-Exclude Disposal Gain or Loss/Net Sales  
X11 - Operating Expense Rate: Operating Expenses/Net Sales  
X12 - Research and development expense rate: (Research and Development Expenses)/Net Sales  
X13 - Cash flow rate: Cash Flow from Operating/Current Liabilities  
X14 - Interest-bearing debt interest rate: Interest-bearing Debt/Equity  
X15 - Tax rate (A): Effective Tax Rate  
X16 - Net Value Per Share (B): Book Value Per Share(B)  
X17 - Net Value Per Share (A): Book Value Per Share(A)  
X18 - Net Value Per Share (C): Book Value Per Share(C)  
X19 - Persistent EPS in the Last Four Seasons: EPS-Net Income  
X20 - Cash Flow Per Share  
X21 - Revenue Per Share (Yuan ¥): Sales Per Share  
X22 - Operating Profit Per Share (Yuan ¥): Operating Income Per Share  
X23 - Per Share Net profit before tax (Yuan ¥): Pretax Income Per Share  
X24 - Realized Sales Gross Profit Growth Rate  
X25 - Operating Profit Growth Rate: Operating Income Growth  
X26 - After-tax Net Profit Growth Rate: Net Income Growth  
X27 - Regular Net Profit Growth Rate: Continuing Operating Income after Tax Growth  
X28 - Continuous Net Profit Growth Rate: Net Income-Excluding Disposal Gain or Loss Growth  
X29 - Total Asset Growth Rate: Total Asset Growth  
X30 - Net Value Growth Rate: Total Equity Growth  
X31 - Total Asset Return Growth Rate Ratio: Return on Total Asset Growth  
X32 - Cash Reinvestment %: Cash Reinvestment Ratio  
X33 - Current Ratio  
X34 - Quick Ratio: Acid Test  
X35 - Interest Expense Ratio: Interest Expenses/Total Revenue  
X36 - Total debt/Total net worth: Total Liability/Equity Ratio  
X37 - Debt ratio %: Liability/Total Assets  
X38 - Net worth/Assets: Equity/Total Assets  
X39 - Long-term fund suitability ratio (A): (Long-term Liability+Equity)/Fixed Assets  
X40 - Borrowing dependency: Cost of Interest-bearing Debt  
X41 - Contingent liabilities/Net worth: Contingent Liability/Equity  
X42 - Operating profit/Paid-in capital: Operating Income/Capital  
X43 - Net profit before tax/Paid-in capital: Pretax Income/Capital  
X44 - Inventory and accounts receivable/Net value: (Inventory+Accounts Receivables)/Equity  
X45 - Total Asset Turnover  
X46 - Accounts Receivable Turnover  
X47 - Average Collection Days: Days Receivable Outstanding  
X48 - Inventory Turnover Rate (times)  
X49 - Fixed Assets Turnover Frequency  
X50 - Net Worth Turnover Rate (times): Equity Turnover  
X51 - Revenue per person: Sales Per Employee  
X52 - Operating profit per person: Operation Income Per Employee  
X53 - Allocation rate per person: Fixed Assets Per Employee  
X54 - Working Capital to Total Assets  
X55 - Quick Assets/Total Assets  
X56 - Current Assets/Total Assets  
X57 - Cash/Total Assets  
X58 - Quick Assets/Current Liability  
X59 - Cash/Current Liability  
X60 - Current Liability to Assets  
X61 - Operating Funds to Liability  
X62 - Inventory/Working Capital  
X63 - Inventory/Current Liability  
X64 - Current Liabilities/Liability  
X65 - Working Capital/Equity  
X66 - Current Liabilities/Equity  
X67 - Long-term Liability to Current Assets  
X68 - Retained Earnings to Total Assets  
X69 - Total income/Total expense  
X70 - Total expense/Assets  
X71 - Current Asset Turnover Rate: Current Assets to Sales  
X72 - Quick Asset Turnover Rate: Quick Assets to Sales  
X73 - Working capitcal Turnover Rate: Working Capital to Sales  
X74 - Cash Turnover Rate: Cash to Sales  
X75 - Cash Flow to Sales  
X76 - Fixed Assets to Assets  
X77 - Current Liability to Liability  
X78 - Current Liability to Equity  
X79 - Equity to Long-term Liability  
X80 - Cash Flow to Total Assets  
X81 - Cash Flow to Liability  
X82 - CFO to Assets  
X83 - Cash Flow to Equity  
X84 - Current Liability to Current Assets  
X85 - Liability-Assets Flag: 1 if Total Liability exceeds Total Assets, 0 otherwise  
X86 - Net Income to Total Assets  
X87 - Total assets to GNP price  
X88 - No-credit Interval  
X89 - Gross Profit to Sales  
X90 - Net Income to Stockholder's Equity  
X91 - Liability to Equity  
X92 - Degree of Financial Leverage (DFL)  
X93 - Interest Coverage Ratio (Interest expense to EBIT)  
X94 - Net Income Flag: 1 if Net Income is Negative for the last two years, 0 otherwise  
X95 - Equity to Liability

# Procedure and Methods:

* Load the data.
* EDA
* Data Visualization
* Building models:

Logistic Regression

Naive Bayes

KNN

Decision Tree

Random Forest

SVM

XG Boost

CatBoost

ANN

* Result
* Conclusion

**EDA (Exploratory Data Analysis)**

A picture containing text, receipt

Description automatically generatedA picture containing graphical user interface

Description automatically generatedA picture containing diagram

Description automatically generatedTable

Description automatically generatedA close-up of a document

Description automatically generated with medium confidence

### Relevant Papers

Liang, D., Lu, C.-C., Tsai, C.-F., and Shih, G.-A. (2016) Financial Ratios and Corporate Governance Indicators in Bankruptcy Prediction: A Comprehensive Study. European Journal of Operational Research, vol. 252, no. 2, pp. 561-572.  
<https://www.sciencedirect.com/science/article/pii/S0377221716000412>