COMS4995 – Applied Machine Learning: Project Proposal

Title: Predicting Fortune 500 Company Financial Health Trajectory

Group Members: Yu-Heng Chi, Param Sejpal, Jessica Villanueva, Yihan Yang, Zhiyi Zhang

Objective: Develop a model that predicts company performance based on historical financial data using metrics related to cash flow, debt ratios, net income, revenue, and stock price trends.

Abstract: This model will predict the financial health trajectories of the largest corporations, where financial health is indicated by financial performance metrics and events like bankruptcy or restructuring. We can predict if a company will go bankrupt or enter a restructuring event using classification, and these market events are based on financial metrics that can be predicted using regression.

Approach: We will compile a dataset based on financial statements from a Kaggle dataset and market data from Yahoo! Finance.

1. Data from this dataset of financial statements of major companies from 2009-2023 will be used to analyze Net Income, Gross Profit, EBITDA, and other common metrics from company balance sheets and 10-K annual reports.
2. We will look the following KPIs from Yahoo! Finance: Stock price declining over 20% YTD, revenues declining over 5% year over year (fiscal-to-fiscal), Negative Net Income, Negative Cash Flow From Operations, Debt/EBITDA above 5x, EBITDA/Interest Expense below 1.1x, debt maturing in 2 years, and Free Cash Flow.

Proposed ML techniques: For classifying sudden market events, will implement **logistic regression** and **gradient boosting**. For predicting financial metric values themselves, we will apply **linear regression** and time series methods. We will perform the following phases of the supervised learning framework.

- We will create the dataset with the above metrics and preprocess the data, which will include handling missing data, encoding, and scaling.
- We will (1) split the data, (2) determine hyperparameters, (3) train the models, (4) evaluate performance, and (5) deploy the model.
- The implementation of certain methods will depend on the nature of the data; for instance, we will likely implement stratified splitting.
- We will predict financial health (classification) and metrics (regression) using the models above.

Conclusion: Machine learning techniques like logistic regression, gradient boosting, and linear regression will be used to determine how each feature impacts company performance. Observations on company financial health trajectories can be made based on the results obtained from the above models.