COMS4995 – Applied Machine Learning: Project Proposal

Title: Predicting Fortune1000 Company Financial Health Trajectory

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

Objective: Predict Fortune1000 financial health trajectories based on key financial metrics from 2024. Implement machine learning techniques for making predictions and observing company performance.

Abstract: This model will predict the financial health trajectories of Fortune 1000 companies, where financial health is indicated by market capitalization ("Market Cap") value. We can predict Market Cap value using regression techniques, and we will additionally investigate company health using classification based on EBITDA margins.

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

1. Data from this Kaggle dataset of 2024 financial statements of Fortune 1000 Companies 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 use KPIs from Yahoo! Finance on EBITDA. This metric will be used for regression.

Proposed ML techniques: For predicting financial metric values themselves, we will apply linear regression and a neural network. However, we will also use other models besides these baseline ones to investigate our results (gradient boosting regression and random forest regression). 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. Additionally, we may later implement SMOTE, as we will see during classification that the majority class of Fortune1000 companies is "healthy."

Conclusion: In the real world, a company is classified based on sudden market events, like bankruptcy or restructuring. Observations on company financial health trajectories can be made based on the results obtained from the above models, from which we will predict financial metrics (regression) and categorize companies by financial health (classification). Specifically, regression will be used to predict Market Cap for 2024 and classification will be used to categorize companies by financial performance.