**Stock Price Prediction**

6341 Applied Machine Learning

Professor: Dr. Zixuan Meng

Group 10

Group Members: Wei-Chien Cheng(wxc200000), Chien-Chun Liao(cxl210005), Jane(Jie) Song(jxs210024), Long Wan(lxw190011), Qiyang Liu(qxl210002)

We will analyze the top 20 companies in the technology industry and build our model to train a combination of machine learning (ML) models and predict their stock price. More specifically, we collect firm-level, industry-level, macroeconomic variables, daily stock transaction data, and news information as inputs or predictors for 20 years.

Our dataset is a merged dataset consisting of news data from Associated Press, Yahoo News, technical feature data from Yahoo Finance. Firm-level data from Capital IQ and Yahoo Finance, including sales, net-income, gross-income, tax rate, R&D, D&A, EBITDA, P/E, EPS, Momentum(Past 12-month return), Gross Profits-to-Assets, Asset Growth, Investment-to-Assets, Net Operating Assets, etc. Industry-level will include Industry Earnings, Industry Size, etc. Macro Variables will consist of IP, CPI, NASDAQ, SP500, and DOW Jones. And daily stock transaction variables, which include Open, Close, High, Low, Adj Close, Volume. Our data will contain approximately 50,000 observations and 30 attributes.

We will utilize simple linear regression, polynomial linear regression, KNN, Support Vector Regression (SVR), LASSO regression, and random forest. Further, we will perform hyperparameter tuning and cross-validation to choose top models, ensembling those models to a final model using meta-learning or soft voting.