Proposal for DATA318 Final Project

Project Name: Influence of Tweets on Stock Prices

Team members: Jake Peters & Uyanga (Soka) Naranbaatar

**What is our target variable?**

The target variable is Tesla stock and DOGE coin stock. The predictor variables we are going to use are the sentiments of ELON MUSK’s tweet,

**Why is it interesting?**

Both of us are interested in investing in stocks, and we hypothesize that there might be correlations between the price change of the stocks and Elon Musk’s opinion about entities that he has tweeted about. We also think that his tweets will impact the stock price within a matter of hours from his tweet, which is why we can use the same day-metrics to join our predictors and target variables together.

**What is your plan for analysis?**

The main idea is to predict whether Tesla Inc (TSLA) and cryptocurrency Dogecoin (DOGE) stock prices go up or down on a given day based on Elon Musk’s tweets involving those companies and the sentiment analysis of the emotions of the words in his tweets, such as excitement, anger, trust, etc. There will be 5 or more different sentiments that we can use as qualitative variables to predict the resulting impact on the stock prices, such as anger, sadness, excitement, and other variables.

Alternatively, we could also use functions from R packages that translate these sentiments to qualitative predictors, which we would then use to create a linear model to predict the Stock Prices.

The plan is to use the classification and regression models. We are using Support Vector Machines, Logistic Regression, and Naive Bayes models to predict whether the stock stock price goes up or down based on the different sentiments of the tweets. Linear Regression will be used to predict the amount of price change.

**How will they be evaluated?**

The data will be split into a training set and a testing set, in addition to the cross validation for the training set. This data will then be plugged into our various models. We will then compare the results from different models and see which ones most accurately predict the price of the stocks on a given day.