**CS-483**

**Group Project 1**

**Group: JET = NaN**

Members: Julian Keller

Tianyi Wang

Eric Gioe

The historical timeframe chosen by the team was determined to be one year of data (253 points) for the training set, and one month (24 points) for the prediction dataset. The reason behind the single year for the training dataset was to eliminate the seasonality of the stock market. Every year there is a similar repeatable pattern from November to April, with the conclusion of the fourth quarter of the fiscal year and the beginning of the new fiscal year. It was our goal to only look at the latest year of data, so that we eliminate this phenomenon. One month of data was chosen for the prediction dataset in order to provide a sufficient amount of data for the model to perform well. It is difficult for regression models to perform complex data prediction without enough data.

It was decided that each of the group members was to complete a different model to predict stock data. The goal of this was to try three different methods to determine the most effective in predicting the stocks. The three models that were chosen were Long Short-Term Memory (LSTM), K-Nearest Neighbors (KNN) with Linear Regression, and Autoregressive Integrated Moving Average (ARIMA). The methods of predicting the data were compared using an error measurement for the same feature of the same stock that was chosen. The method with the least error was chosen for analyzing all of the data. The Long Short-Term Memory was the model that had the least amount of error in the data prediction

Julian – Long Short-Term Memory

Tianyi – K-Nearest Neighbors with Linear Regression

Eric – Autoregressive Integrated Moving Average