Applied Neural Networks: Stock Price Prediction
Executive Summary

1. **Overview**
   We develop neural networks that can accurately forecast stock prices based on various features and compare this model with other predicting models for predicting stock prices using historical data.

2. **Data**
   We obtained the S&P 500 stock data from https://www.kaggle.com/datasets/camnugent/sandp500. The data includes 5 years of stock data across 500 companies from 2013-02-08 to 2018-02-07. The data is clean without any missing or undefined entries. Among the 500 companies, we chose Advance Auto Parts (AAP) to apply to our models. The data has a trend but no seasonality.

3. **Implementation**
   Each of the models we used forecast a company's closing stock price one day into the future using all the previous stock closing prices from that 5 years time. All of our methods involve a sliding window approach where a fixed number of days prior to the day we're trying to predict is used.

   Programming language used: Python
   Built-in libraries used: sklearn, pandas, numpy and matplotlib

   Prediction Models Implemented:

- Naive Prediction: Given a day for which we wish to predict a stock's closing price, we predict it to be the closing price from the day before. Here, the window size is 1.

- Sliding-window Linear Regression model: For any day we wish to predict, we take W days prior and do a linear regression to make our prediction. A window size of 5 was used.

- LSTM Model: For any day we wish to predict, we consider frames of data from the W days preceding that day and then train a neural network on that window of data to make our prediction. A window size of 7 was used.

4. **Model Performance**

   We evaluated our models using four-fold time series cross validation on the first 80% of data. Thus, we partitioned this 80% of data into four "folds" and used the first fold strictly for training. For folds two,three, and four, we took the mean squared error (MSE) between the actual data and the predicted data on these folds. We then took the

average of the three MSE calculations as the final evaluation metric. The model with the lowest time series cross-validation (TSCV) error was then used to predict the test data.

LSTM Neural Network TSCV Error: 11.02$
Sliding Windows Linear Regression TSCV Error: 10.58$
Naive Prediction TSCV Error: 6.86$

Naive Prediction Test Error: 7.54$

5. **Future Work :**

Currently, our predictions rely solely on stock data from a single company (AAP). Moving forward, we can incorporate data from multiple companies to develop a more comprehensive and accurate model.

Most importantly, our methods do not take into account other significant factors such as political events and economic conditions that can affect stock price movements. They also do not take any probabilistic models of the stock market into account. They are general methods which can apply to any time series data (e.g weather). These can be incorporated into future prediction models.

Furthermore, we can test and refine our models using stocks from markets that are economically diverse and significantly different from the US market. This will help ensure our models are robust and applicable in a variety of global contexts.