**CSIT 456: Advanced Techniques in Data Science**

**Project Report**

**Project Title:** Predicting Daily Closing Price Variation for Apple Stock

**Team Members:** Aaron Hayward, Jhony Ortiz, and Sai Moturi

**Description**

**Basic Information**

The motivation behind this project was to analyze the trend in stock prices for Apple Inc. from 100 days prior to May 5th, 2021; this was done to predict if the closing prices were higher or lower than the opening prices for this time frame. These predictions assisted in better understanding the overall health of the Apple stock during the pre-mentioned time period.

**Project Objectives**

The primary questions we are attempting to answer for this project include the following. How many days experienced a net decrease from the opening to closing stock prices? Conversely, how many days experienced a net increase from the opening to closing stock price? Finally, what was the net movement in stock after totaling the daily stock price changes, from the beginning to end, of the 100 day period.

**Description of the Data Set**

The data set we are utilizing originated from the historical data of Yahoo finance, specifically on it’s Apple Stock page. It contains the stock history for Apple over the last 20 years. For the purposes of this project, we analyzed only the data from December 10th of 2020 to May 5th of 2021. The time span covers 100 days in total.

Link: [Apple Inc. (AAPL) Stock Historical Prices & Data - Yahoo Finance](https://finance.yahoo.com/quote/AAPL/history/)

**Data Analysis Using Machine Learning**

**Data Preparation**

In order for us to prepare the data set, it was necessary to clean it first. Instead of using all the data on the csv file, we used only the 100 days time frame mentioned prior from when we first found the data. We added a column called price\_variation to determine how much the price changed positively or negatively and that was done by subtracting the open price for each day from the close price for each day. We then created different features to be used for the data models.

**Models Selection**

We used Pandas, Numpy, Matplotlib, and Seaborn within this dataset. The Decision Tree and K Nearest Neighbors were the machine learning models used to predict the outcomes of the current dataset. The data science technique we utilized for analysis was Data Classification.

**Models Implementation**

To be able to train the models, we had to split the data to a target and a feature dataset and then split

the data to 65% training and 35% testing. For decision tree, we created and trained the model, before

tuning it we got an accuracy of 1.0. To tune the hyperparameters, we used GridSearchCV to do so

and the accuracy was the same 1.0. For KNN, we imported KneighborsClassifier to get the test

score and it came out to 0.97. To tune it, we import StratifiedKFold and cross\_val\_score and the

best n\_neighbors came out to 5. Once testing the score again we got a score of 0.97 again.

**Evaluation**

The overall accuracy came out to be much higher than we initially expected it to be. The decision

tree model happened to give a better accuracy than KNN. More specifically, the Decision Tree scored

an accuracy of 100% while the KNN scored an accuracy of 97.1%. Post hypertuning there was no

increase for either, with each accuracy staying the same. The Decision Tree utilized Gridsearch

whilst the KNN utilized the cross validation.

**Conclusion**

The results of this project have shown that from the dataset analyzed, 57 days in total experienced a net decrease from opening to closing stock prices. Conversely, 43 days experienced a net increase from the opening to closing stock value. In total, after the studied 100 day period there was a net negative decrease from opening to closing prices by approximately $13.97.

**Important Notice**

The results and accuracy of this project contain different numbers than those originally contained in

the Google Slides presentation. This is because after we presented, we made an edit to our target

data set code that vastly increased the accuracy of the entire project. The new code states that:

df['Close\_Change']= np.where(df['Close'] > df['Open'], 1, 0)

# Shift to the next day and compare it. If greater, set value to one, else 0

**Works Cited**

Cheng, A. (2020, February 11). *Machine Learning: Step-By-Step*. Medium. <https://towardsdatascience.com/machine-learning-step-by-step-6fbde95c455a>.

Yahoo! (2021, May 17). *Apple Inc. (AAPL) Stock Historical Prices & Data*. Yahoo! Finance. <https://finance.yahoo.com/quote/AAPL/history/>.

YouTube. (2020). *Stock Prediction Using Python & Machine Learning*. *YouTube*. <https://www.youtube.com/watch?v=tQx-XGPecBA>.