## **Task Report Cos30018 Option B**

B.1: Set Up

Name: Le Bao Nguyen

Student Id: 104169837

- I. Setting up the environment:
  - We started downloading the code v0.1 and P1 from the canvas:

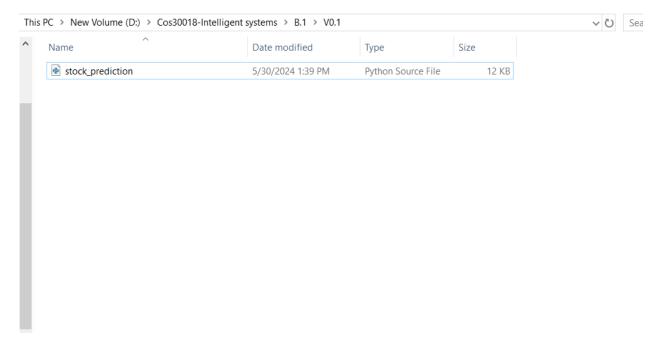


Figure 1: Downloading the v0.1 code from canvas into the computer.

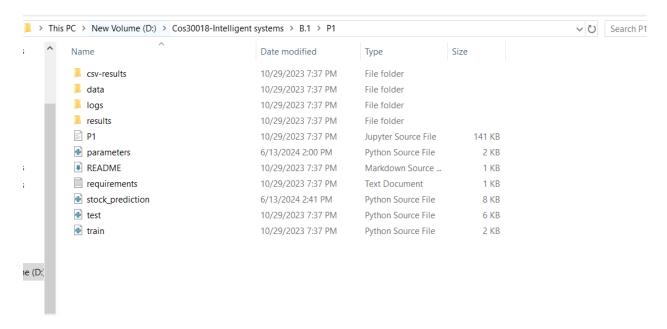


Figure 2: Downloading the P1 code from canvas into the computer.

We upload both v0.1 and P.1 folder into the google drive to link to gg drive (we can still run the code even if it's not link to the google drive or not):

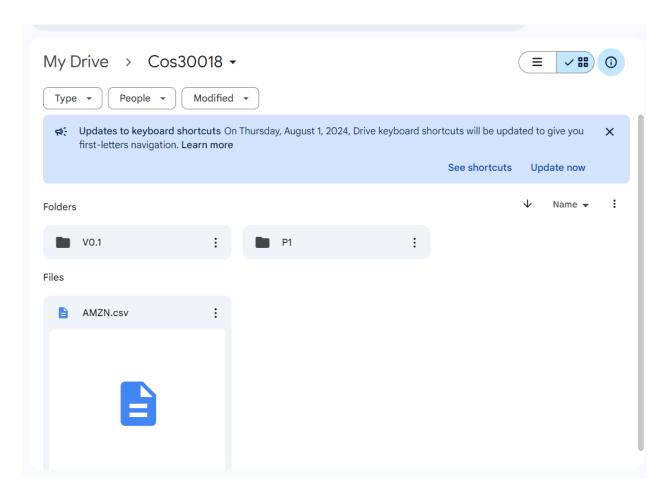


Figure 3: Uploading the v0.1 and P1 code into folder in google drive.

- After that, we open the gg collab and start to link to our gg drive that contain folders and direct to the folder locations:



Figure 4: Linking the v0.1 and P1 code from google drive to google collab.

- We import the libraries or extensions that we need to run the code and use "!pip install ...."install the libraries if they are still not available in the gg collab:

Figure 5: Importing and installing libraries.

- After completing import the resources, we start to implement code into the gg collab to run it but we implement separately to test out if the code run or not:

```
[2] DATA_SOURCE = "yahoo"

COMPANY = "TSLA"

# start = '2012-01-01', end='2017-01-01'

TRAIN_START = '2015-01-01'

TRAIN_END = '2020-01-01'

data = yf.download(COMPANY, start=TRAIN_START, end=TRAIN_END, progress=False)

[3] PRICE_VALUE = "close"

scaler = MinMaxScaler(feature_range=(0, 1))

scaled_data = scaler.fit_transform(data[PRICE_VALUE].values.reshape(-1, 1))
```

Figure 6: Implementing the code separately to check if it works or not (1).

Figure 7: Implementing the code separately to check if it works or not (2).

```
model.add(Dropout(0.2))

# The Dropout layer randomly sets input units to 0 with a frequency of
# rate (= 0.2 above) at each step during training time, which helps
# prevent overfitting (one of the major problems of ML).

model.add(LSTM(units=50, return_sequences=True))
# More on stacked LSTM:
# https://machinelearningmastery.com/stacked-long-short-term-memory-networks/

model.add(Dropout(0.2))
model.add(Dropout(0.2))
model.add(Dropout(0.2))

model.add(Dropout(0.2))

# Prediction of the next closing value of the stock price

# We compile the model by specify the parameters for the model
# See lecture Week 6 (COS30018)
model.compile(optimizer='adam', loss='mean_squared_error')
```

Figure 8: Implementing the code separately to check if it works or not (3).

Figure 9: Implementing the code separately to check if it works or not (4).

- II. Testing the code v0.1 and P1:
  - We run the v0.1 code and train through 25 epochs before creating the charts to show the accurate and predicted data of TSLA stock. After creating the chart, the model predicts the stock price for the next day after the test period.
  - The Sequential API from Keras is used to build the model. To avoid overfitting, it is composed of three LSTM layers, each of which is followed by a Dropout layer.

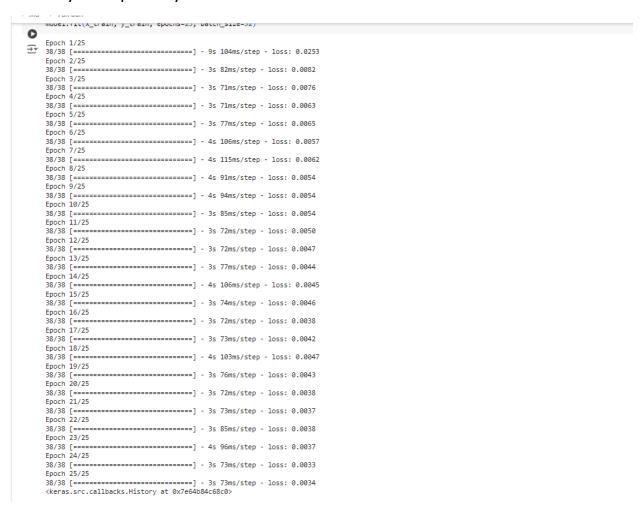


Figure 10: Training the model through 25 epochs before printing the chart.

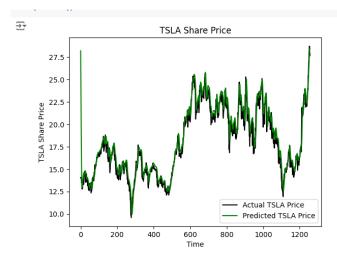


Figure 11: The prediction chart after training.

шпе

Figure 12: Printing the prediction for next day.

- When we run the P1 code, it has to go through 500 epochs to train which take a long time to do it so we decide to reduce it to 15 epochs by editing the epochs in the parameter.py.

```
| IESI_SIZE = 0.2
| [2] # features to use
       FEATURE_COLUMNS = ["adjclose", "volume", "open", "high", "low"]
        # date now
       date_now = time.strftime("%Y-%m-%d")
        ### model parameters
        N_LAYERS = 2
        # LSTM cell
        CELL = LSTM
        # 256 LSTM neurons
       UNITS = 256
        # 40% dropout
       DROPOUT = 0.4
        # whether to use bidirectional RNNs
        BIDIRECTIONAL = False
        ### training parameters
        # mean absolute error loss
        # LOSS = "mae"
        # huber loss
        LOSS = "huber loss"
        OPTIMIZER = "adam"
        BATCH_SIZE = 64
        EPOCHS = 500
        # Amazon stock market
        ticker = "AMZN"
        ticker_data_filename = os.path.join("data", f"{ticker}_{date_now}.csv")
        # model name to save, making it as unique as possible based on parameters
        model_name = f"{date_now}_{ticker}-{shuffle_str}-{scale_str}-{split_by_date_str}-\
        {LOSS}-{OPTIMIZER}-{CELL.__name__}-seq-{N_STEPS}-step-{LOOKUP_STEP}-layers-{N_LAYERS}-units-{UNITS}"
        if BIDIRECTIONAL:
           model_name += "-b"
```

Figure 13: We need to replace the epochs from 500 to 15 for faster train.

Figure 14: The training duration when we set to 500 epochs.

- After we set the epochs to 15 and train the model, we print the chart and results by running the test.py.

```
4/15
            | Figoria | 1 | 
            ch 12: val_loss did not improve fr
             Epoch :
85/85
            -----] - ETA: 0s - loss: 6.0415e-04 - mean_absolute_error: 0.0177
■
```

Figure 15: The training duration when we set to 15 epochs.

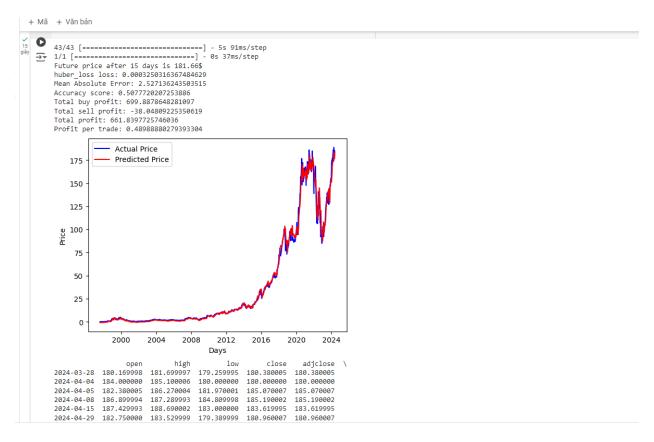


Figure 16: The prediction chart when after train model set to 15 epochs.

- III. The understanding of the code base v0.1:
- 1. Import libraries.

```
import numpy as np
import matplotlib.pyplot as plt
import pandas as pd
import pandas_datareader as web
import datetime as dt
import tensorflow as tf
import yfinance as yf

from sklearn.preprocessing import MinMaxScaler
from tensorflow.keras.models import Sequential
from tensorflow.keras.layers import Dense, Dropout, LSTM, InputLayer
```

Figure 17: Importing the libaries.

- Before we start running the code, we need to implement the libraries. These are the libraries required for date-time modification (datetime), machine learning (tensorflow, keras), data visualization (matplotlib), and data management (numpy, pandas, pandas\_datareader). The purpose of yfinance is to get historical stock data.

2. Set data source and company.

```
DATA_SOURCE = "yahoo"
COMPANY = "TSLA"
```

Figure 18: Setting the data resources from yahoo and get the TESLA data.

- To retrieve the stock data for Tesla, "DATA\_SOURCE" is set to "yahoo" (using Yahoo Finance), and "COMPANY" is set to "TSLA" (Tesla).
- 3. Define training period and fetch data.

```
TRAIN_START = '2015-01-01'
TRAIN_END = '2020-01-01'

data = yf.download(COMPANY, start=TRAIN_START, end=TRAIN_END, progress=False)
```

Figure 19: Setting the time to train and fetch the TESLA data.

- The data source, the ticker symbol of the business whose stock values are being examined, and the date ranges for the training and testing datasets are all defined in this section using constants. The stock data will be retrieved from Yahoo Finance, as indicated by the setting of "DATA\_SOURCE" to "yahoo".
- 4. Prepare the data.

```
[ ] PRICE_VALUE = "Close"

scaler = MinMaxScaler(feature_range=(0, 1))

scaled_data = scaler.fit_transform(data[PRICE_VALUE].values.reshape(-1, 1))
```

Figure 20: Scaling the data.

- For predicted purposes, the stock's closing price is chosen. Using "MinMaxScaler", the data is scaled to a range of 0 to 1, which enhances the performance of the model.

## 5. Prepare training data.

```
# Number of days to look back to base the prediction
PREDICTION_DAYS = 60 # Original

# To store the training data
x_train = []
y_train = []

scaled_data = scaled_data[:,0] # Turn the 2D array back to a 1D array
# Prepare the data
for x in range(PREDICTION_DAYS, len(scaled_data)):
    x_train.append(scaled_data[x-PREDICTION_DAYS:x])
    y_train.append(scaled_data[x])

x_train, y_train = np.array(x_train), np.array(y_train)
x_train = np.reshape(x_train, (x_train.shape[0], x_train.shape[1], 1))
```

Figure 21: Setting the training data part.

- The model forecasts the price for the following day using data spanning 60 days. Two arrays are created from the data: "x\_train" (input) and "y\_train" (output). As required by the LSTM model, the "x\_train" array is reshaped to be three-dimensional.

## 6. Build the model.

```
model = Sequential() # Basic neural network
model.add(LSTM(units=50, return_sequences=True, input_shape=(x_train.shape[1], 1)))

[ ] model.add(Dropout(0.2))
model.add(LSTM(units=50, return_sequences=True))
model.add(Dropout(0.2))
model.add(LSTM(units=50))
model.add(Dropout(0.2))

model.add(Dropout(0.2))

model.add(Dense(units=1))
model.compile(optimizer='adam', loss='mean_squared_error')

[ ] model.fit(x_train, y_train, epochs=25, batch_size=32)
```

Figure 22: Building the model after training.

The Sequential API from Keras is used to build the model. To avoid overfitting, it is composed of three LSTM layers, each of which is followed by a Dropout layer. The anticipated stock price is produced by the last Dense layer. The Mean Squared Error loss function and the Adam optimizer

are used to construct the model. After that, it is trained for 25 epochs with a batch size of 32 using the training data.

7. Load and prepare test data.

```
# Load the test data
TEST_START = '2020-01-02'
TEST_END = '2022-12-31'

test_data = yf.download(COMPANY, start=TRAIN_START, end=TRAIN_END, progress=False)

test_data = test_data[1:]

actual_prices = test_data[PRICE_VALUE].values

total_dataset = pd.concat((data[PRICE_VALUE], test_data[PRICE_VALUE]), axis=0)

model_inputs = total_dataset[len(total_dataset) - len(test_data) - PREDICTION_DAYS:].values
model_inputs = model_inputs.reshape(-1, 1)
model_inputs = scaler.transform(model_inputs)
```

Figure 23: Loading the test data.

- Similar to how the training data is retrieved and prepared, as well is the test data. For the duration of the test, "actual\_prices" contains the actual stock prices. The scaled values needed for prediction are stored in the "model\_inputs" array.
- 8. Prepare test inputs and predict.

```
x_test = []
for x in range(PREDICTION_DAYS, len(model_inputs)):
    x_test.append(model_inputs[x - PREDICTION_DAYS:x, 0])

x_test = np.array(x_test)
    x_test = np.reshape(x_test, (x_test.shape[0], x_test.shape[1], 1))
# TO DO: Explain the above 5 lines

predicted_prices = model.predict(x_test)
    predicted_prices = scaler.inverse_transform(predicted_prices)
```

Figure 24: Preparing the test input and begin to predict.

 In order to create input sequences for the LSTM model to make predictions, the test data is prepared in the same way as the training data. These sequences are initially stored in an empty list called "x\_test". For the test data, the stock prices are predicted by the trained LSTM model. Predicted prices are generated in a series based on the test sequences that have been developed. The "MinMaxScaler"'s inverse transform function is then used to scale these forecasted prices back to their initial range.

Visualize the results.

```
[ ] plt.plot(actual_prices, color="black", label=f"Actual {COMPANY} Price")
    plt.plot(predicted_prices, color="green", label=f"Predicted {COMPANY} Price")
    plt.title(f"{COMPANY} Share Price")
    plt.xlabel("Time")
    plt.ylabel(f"{COMPANY} Share Price")
    plt.legend()
    plt.show()
```

Figure 25: Printing the result and prediction chart.

- The script uses matplotlib.pyplot to plot the actual and predicted stock prices in order to evaluate the effectiveness of the model. Plotting the expected costs is done in green, and plotting the actual prices is done in black. The x- and y-axes of the plot are labeled "TSLA Share Price," in accordance with its title.
- 10. Predict the next day.

ппе

Figure 26: Printing the prediction for the next day.

- Finally, following the test period, the model predicts the stock price for the following day. This illustrates how the model may forecast future events using the most recent data.
- IV. The Git hub link:

https://github.com/lebaonguyen123456/Cos30018---Option-B-Stock-prediction