**UNIVERSITY OF WESTMINSTER**

**Msc Big Data Technologies**

Module Code: 7BUIS008W.2

Module Title: Data Mining and Machine Learning

Module Leader: Dr. Panagiotis Chountas

Module Tutor: Dr. Natalia Yerashenia

Assignment: Coursework 2

Student ID: W1848650

Student Name: Mohammadamin Shams

Table of Contents

[1 Abbreviations 2](#_Toc125310532)

[2 Predict Price of Netflix Stock with LSTM Neural Network 2](#_Toc125310533)

[2.1 Data Analysis 2](#_Toc125310534)

[2.1.1 Understanding The Data and Pre-processing 3](#_Toc125310535)

[2.2 Data Visualization 3](#_Toc125310536)

[2.3 Long Short-Term Memory Model 4](#_Toc125310537)

[3 References 8](#_Toc125310538)

# Abbreviations

|  |  |
| --- | --- |
| LSTM | Long Short-Term Memory |

# Predict Price of Netflix Stock with LSTM Neural Network

In this part we want to talk about the task B of the coursework 2. We have a data set that contains Netflix stocks price from 2002 until 17 of April 2022. Our goal is to use this data and split it into two groups of testing and training and then train our model to predict the stocks of Netflix. This approach is not that easy because we only going to use close value of the market and we want to predict the same feature and we need to change the format of data. After we changed the format of data, we only need to feed the model with training and evaluate our model based on the test set. In the next parts we are going to explore the data first and then start reshaping and creating our LSTM neural network.

## Data Analysis

To create LSTM Neural network first we need to check our data and start doing preliminary stages for a better data analysis. First part is importing all the necessary libraries and classes and import the Netflix data set.

Text

Description automatically generated

Figure 2. Importing the dataset and necessary libraries

### Understanding The Data and Pre-processing

After importing our dataset, we need to do some preliminary analysis to understand the dataset, understand data type of each column and see if there is null values or not. In data set there is no null values and we also can get a good report for the dataset by using profilereport function from pandasprofiling library. In the figure below we can see all the preliminary stages that has been done to understand the data.

Table

Description automatically generatedGraphical user interface, application

Description automatically generated

Figure 2. Preliminary stages to understand the data

Our Netflix data has 5009 rows and 7 columns. These 7 columns are related to the stock price of Netflix for a specific date. Opening value and closing values are obvious the values for the time that markets open for trading and closes. High and low is the highest and lowest price of a specific date. The adjusted closing price amends a stock's closing price to reflect that stock's value after accounting for any corporate actions. Volume is the total shares of the Netflix traded that day.

## Data Visualization

Now we need to visualize the stock for Netflix and only using close market value of each day for the last year. To do this, we only need to choose date column and close value from our initial dataset for visualization. After that we need to change date column from object to date with the help of pandas library and to\_datetime function and set it as the data set index and then plot our dataset only for the last year with matplotlib library and plot function. Below we can see the code and the last year stock price for Netflix close value of their share.

Graphical user interface, text, application

Description automatically generatedGraphical user interface, chart

Description automatically generated

Figure 2. Netflix share price in a plot

In the last year we can see that Netflix stock first went up to the all-time high at 700 dollars but just after two months after that their stock has tumbled around 40 percent and dropped to 400 dollars. Based on CNN business news, Netflix after surging to the top of the streaming mountain is struggling to climb higher. Netflix is raising their prices and one of the reasons of this sudden drop especially is raising prices in north America. High prices can be problematic for Netflix because their rivals are pushing hard to get more subscribers by creating new content and offering low prices and Netflix may even struggle more and their stock price may even drop more. One of the Netflix plans for preventing drop in stock prices is getting more subscribers by trying to expand their company in the Asian countries like India and China, we can see this by new contents from Asian countries and well-developed contents for their test.

## Long Short-Term Memory Model

Time series prediction problems are more difficult than regression predict models because time series adds the complexity of a sequence dependencies to our model. Long Short-Term Memory neural networks are well-crafted to predict targets based on a time-series. One of the problems for neural networks was the duration between important events in a time series. They couldn’t detect them and because these events were based on the duration, the problem of gradient vanishing would show up and LSTMs were defined to issue this problem.

First step to create our model is fixing the random number seed to ensure that our results are reproducible. Second part is changing the data type of the values and put them as floats because LSTM works better with them. Next step is normalising the data, because our close value is not scaled and LSTM is sensitive to the scale of the data, specially when sigmoid function is used. So we need to scale our data to range between 0 and 1 which can be done with the help of MinMaxscaler from sklearn.preproccesing library. In the figure below these steps are shown. we only used the data for last 6 year because when we used the whole time the outcome was not good but choosing the last 6 year created a better prediction. we guess one of the reasons maybe the overfitting and another one maybe the actual short- and long-term feature to the LSTM which will consider prices in the beginning of Netflix creation and that is going to be problematic because now the price is really changed from back then.

Graphical user interface, text, application, email

Description automatically generated

Figure 2. Data preparation for LSTM

Second part is forming our test and train set. Our data is in timely manner, and we cannot split it with tes\_train\_split function, this time we need to create test size and train size which will give us the number of rows which until that part we have 70% of the data for train set and the starting row for another 30% of the data for test set is test size. Figure below shows how train and test set is created.

Graphical user interface, application, Word

Description automatically generated

Figure 2. Train and test split

We can see that our train set is 1057 observations and test is 454 observations.

Now we need to create our input and output for LSTM model. The way we want to predict is only use one day as input and then we want to predict the following day as an output, so our time step is one. we can use other time steps but for the sake of argument we only used one. Then we need to create a function that gets the number of time steps with data set and form the two input and output, the function is created in a way that by changing the look back get another X and y, for example we could change the look back and use 7 days as input to predict the eighth day. The code is shown in the fig below.

Graphical user interface, text

Description automatically generated

Figure 2. Forming input and output for LSTM

Last stage before creating the model is reshaping input because LSTM get the inputs in a specific form. This is the form accepted by LSTM:

[samples, time steps, features]

To do this we use reshape function from NumPy library and put desired shape we want.

Graphical user interface, text, application

Description automatically generated

Figure 2. Reshape data for LSTM usage

After the data is ready, we can build our model and predict the test data and evaluate our predictions. In the figures below the code and results provided.

Graphical user interface, text, application

Description automatically generated

Figure 2. Build the LSTM model

In figure above we created our LSTM model. Our LSTM consists of one visible layer that gets one input, one hidden layer with 5 LSTM neurons and one output layer which condense everything and creates one output. Our model does this 100 times(epochs=iteration) for our data set to get abetter result with a batch size of 1. Verbose is only used to see which epoch is happening and how much each epoch had changed the data. we can put it 0 to not see the progress or put it 2 to see what happens in each epoch and see the progress too.

Table

Description automatically generated

Figure 2. Model summary

Now the model is created we can fit the train and test and see the accuracy of our model. Before we evaluate the model, we need to inverse scaling and then evaluate the model by measuring the root mean squared of the error between prediction and actual values. As we can see in the next fig, our model is well trained and in the training data has the average root mean squared of 7.28 and square root of that is 2.68. Which indicates our model is doing great. For testing data, we have 23.36 for average root mean squared error and the root of it is 4.83 and shows our model predict test data extremely good.

Graphical user interface, text, application

Description automatically generated

Figure 2. Model predictions and evaluation

In the last part we can put all the things together and visualize them in one plot. To do that again first we should reshape the test and training sets and also inverse scale the actual values of the dataset and put all together in one plot. The code for that is in the fig below.

Graphical user interface, text, application, email

Description automatically generated

Figure 2. Code for plotting the output of model and real value

The blue colour is the actual price and the orange is train set prediction by model and the green is the test set prediction by LSTM model. We can see that our model created a good result in a way that train predictions and actual price look really the same.

Chart, histogram

Description automatically generated

Figure 2. Prediction of LSTM and real values for the last 6 years

# References

Brownlee, J. (2018). How to Develop LSTM Models for Time Series Forecasting. [online] Machine Learning Mastery. Available at: <https://machinelearningmastery.com/how-to-develop-lstm-models-for-time-series-forecasting/>

keras.io. (n.d.). The Sequential model. [online] Available at: <https://keras.io/guides/sequential_model/>.

finance.yahoo.com. (n.d.). Netflix, Inc. (NFLX) Stock Historical Prices & Data - Yahoo Finance. [online] Available at: <https://finance.yahoo.com/quote/NFLX/history?p=NFLX>.

‌