***BA723 Capstone Project***



***Netflix Stock Forecasting***

***By Justin Tay***

***301151554***

**Introduction**

The aim of this project is to predict and forecast Netflix’s stock price, and if we can forecast the price which model would be the most appropriate. This project will be divided into two parts in regards to our methodology:

1. What is the risk in investing in the Netflix stock?
2. If we are to forecast the Netflix stock price, which model would be the most appropriate?

The first part will consist of exploratory data analysis (EDA) regarding the Netflix stock, this will include analysing the correlation of all relevant variables from our data, seasonal decomposition, and assessing the risk and stock return of the Netflix stock.

The second part will involve generating different models and assessing which one is the most appropriate for selection by using Mean Squared Error (MSE). The models which will be generated are Moving Average (MA), ARIMA, and LSTM.

**EDA**

Time series model will be used to predict stock prices**.** The appropriate independent variables would be the Closing Prices and Date of Netflix stock. However, autocorrelation and correlation will help determine whether variables such as US interest rates and Unemployment are relevant for our modeling as well.

In regards to raw data from collecting historical stock prices, there should be no missing data. Especially with publicly available data/records such as US Fed interest rates and US unemployment rate, these data’s are readily available and verified by legitimate stakeholders.

First, we can see that the closing stock price for Netflix has been experiencing an upward trend over the years.

Chart, line chart

Description automatically generated

However despite this, we need to explore deeper into this upward trend and understand our data further. This includes understanding whether the Netflix stock price changes over time are stationary, whether the actual series is correlated with the lag series, and whether it is seasonal.

Essential to analysing stock prices as a time series is concept of correlation and autocorrelation. Correlation of variables important for our analysis in order to understand how other external variables

To measure correlation between the variables, we will be using the Pearson correlation. The Pearson correlation measures how two continuous signals co-vary over time and indicate the linear relationship as a number between -1 (negatively correlated) to 0 (not correlated) to 1 (perfectly correlated). As we can see, Netflix closing price is high correlated to several of the external variables

Table

Description automatically generated

Chart, bar chart

Description automatically generated

Autocorrelation is an important part of time series analysis. It helps us understand how each observation in a time series is related to its recent past observations. When autocorrelation is high in a time series, it becomes easy to predict their future observations. Below we have plotted the scatter plot of the stock prices with a lag of 3. We can clearly see a strong positive correlation between the two.

Chart, scatter chart

Description automatically generated

To detect whether the Netflix stock price is seasonal, we need to conduct seasonal decomposition. As we can see in the illustration below, despite there being an upward trend in Netflix prices, there is a seasonal trend to the fluctuations of the price, which aids us greatly in our forecast for Netflix stock prices. Understanding that there is a seasonal trend to the rise and fall of Netflix prices makes forecasting of Netflix stock price a lot easier.

A picture containing chart

Description automatically generated

Once we have done our initial EDA, we can begin calculating the amount of risk there is for investing in the Netflix stock. First, we must train, validate and test the data.

Graphical user interface, chart, scatter chart

Description automatically generated

Once this is done, we can calculate logarithmic return, which in turn will help us calculate the amount of risk there is in investing in Netflix stocks.

Chart, histogram

Description automatically generated

Graphical user interface, text, application, email

Description automatically generated

Graphical user interface, text

Description automatically generated

From our calculation and predictions show that it is nearly impossible that a price would drop during a day over 10% or during a year over 25%. The amount of calculated value at risk and what we got from the confidence interval shows us that Netflix is a very safe investment. Therefore, it would be recommended to invest in the Netflix stock

**Model generation and comparison**

Once that is completed, we can move on to the second part of our analysis, which is creating the models. We will be creating three separate models, namely Moving Average, ARIMA, and LSTM. The MA and ARIMA models will be univariate, as they will be using only the Netflix closing price variable. Whereas the LSTM model is multivariate, which can use multiple variables, and will be using Netflix Close, High and Low prices.

In regards to a time series model, the method used to measure the health, stability and drift of the models will be Mean Square Error (MSE). MSE measures the average prediction error made by the model when predicting the outcome for a future observation. Lower values of MSE are better. Train, test and splitting of model will be required, as well as scaling using MinMax Scaler. This is especially required for LSTM model. However, scaling might not be necessary for models of MA and ARIMA

Table

Description automatically generated

In the above illustration, we set the moving average window to 50 days.

The 50-day moving average is the leading of the three averages and is, therefore, the first line of major moving average support in an uptrend or the first line of major moving average resistance in a downtrend. Therefore for this model, we will be using a window of 50 days moving average. Indeed, there is the argument that we should also create models for various moving averages such as 10 days, 20 days and see which one has the lowest MSE and therefore select that model for our optimum MA model. However, as this is the most commonly used model for traders and analysts, we want to test this model and verify its viability. Therefore, we will continue to use the 50 day moving average.

As noted, the 50-day moving average is widely used because it works well. The more accurate a moving average is as a trend indicator, the more useful it is for traders and analysts. The ideal moving average shows a level that price will not likely violate on a mere temporary retracement, thus possibly giving a false market reversal signal. It can also be used to place a trailing stop on an existing market position.

Additionally, it is helpful if the moving average is a level that price will approach on retracements and can, therefore, be used to make additional market entries. Through trial and error using various moving averages, the 50-day moving average has served these purposes well. Below is an illustration of the moving average for Netflix’s stock for 50-day moving average.

Chart, line chart, histogram

Description automatically generated

**ARIMA**

ARIMA is an acronym that stands for AutoRegressive Integrated Moving Average.

This type of model is becoming a popular tool for data scientists to employ for forecasting future demand, such as sales forecasts, manufacturing plans or stock prices. In forecasting stock prices, for example, the model reflects the differences between the values in a series rather than measuring the actual values.

Before we can begin ARIMA modelling, we must conduct the Augmented Dickey-Fuller Test. The ADF test is a common statistical test used to test whether a given Time series is stationary or not. We can achieve this by defining the null and alternate hypothesis.

Null Hypothesis: Time Series is stationary. It gives a time-dependent trend.

Alternate Hypothesis: Time Series is non-stationary. In another term, the series doesn’t depend on time.

ADF or t Statistic < critical values: Accept the null hypothesis. Time series is stationary.

ADF or t Statistic > critical values: Failed to reject the null hypothesis. The time series is non-stationary

Graphical user interface, text, application

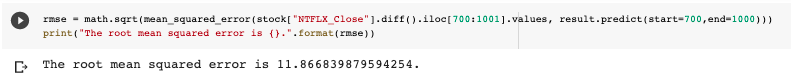
Description automatically generated

After conducting ADF test, we can see that the p-value is 0.938, which is more than 0.05. Therefore, failure to reject the null hypothesis means that the time series is non-stationary.

Table

Description automatically generated

Afterwards, we can calculate the MSE for ARIMA model and compile the model to compare price prediction against actual price.



Chart, line chart

Description automatically generated

**LSTM**

Long short-term memory (LSTM) is an artificial recurrent neural network (RNN) architecture used in the field of deep learning. In regards to LSTM, we are using Multi-variate model as this can allow us to also take into account other closely correlated variables such as Netflix High and Low prices. These are also the variables we will use (which we could not use previously)

Chart

Description automatically generated with medium confidence

Table

Description automatically generated

Once we have successfully compiled the model, we can illustrate our forecast against the actual price and calculate the MSE

Chart

Description automatically generated

From this, we can see how the LSTM has the lowest MSE compared to MA and ARIMA. Therefore we can conclude that there is a low risk to investing in the Netflix stock and that the LSTM model would be the best model to use out of the three we’ve created and measured against.

**Resources used:**

For BA706 project:

Kaggle. (2018, February 23). *Telco Customer Churn*. https://www.kaggle.com/blastchar/telco-customer-churn

SAS. (2021). *SAS Enterprise Miner*. SAS Support. <https://support.sas.com/en/software/enterprise-miner-support.html>

For BA723 Capstone Project

Netflix Historical Data taken from: <https://finance.yahoo.com/quote/NFLX/history?p=NFLX>   
NasDaq Historical Data taken from: <https://finance.yahoo.com/quote/%5EIXIC?p=%5EIXIC>

US Federal Interest Rate taken from: <https://tradingeconomics.com/united-states/interest-rate>

US Unemployment Rate taken from: <https://tradingeconomics.com/united-states/unemployment-rate>

Kiat, K. C. (2019, July 3). *Financial analytics - exploratory data analysis of Stock Data*. Medium. Retrieved December 13, 2021, from https://towardsdatascience.com/financial-analytics-exploratory-data-analysis-of-stock-data-d98cbadf98b9

Mitchell, C. (2021, December 7). *How to use a moving average to buy stocks*. Investopedia. Retrieved December 13, 2021, from https://www.investopedia.com/articles/active-trading/052014/how-use-moving-average-buy-stocks.asp

Phi, M. (2020, June 28). *Illustrated guide to LSTM's and GRU's: A step by step explanation*. Medium. Retrieved December 13, 2021, from https://towardsdatascience.com/illustrated-guide-to-lstms-and-gru-s-a-step-by-step-explanation-44e9eb85bf21

Prabhakaran, S. (2021, November 22). *Arima model - complete guide to time series forecasting in python: ML+*. Machine Learning Plus. Retrieved December 13, 2021, from https://www.machinelearningplus.com/time-series/arima-model-time-series-forecasting-python/