**Contents**

[1. Introduction 1](#_Toc76217036)

[2. Analysis Section 2](#_Toc76217037)

[2.1 About the data 2](#_Toc76217038)

[2.2 Data cleaning with VIS 2](#_Toc76217039)

[(1) Opening price 4](#_Toc76217040)

[(2) Closed price 4](#_Toc76217041)

[(3) High price 5](#_Toc76217042)

[(4) Low price 6](#_Toc76217043)

[(5) Volume 6](#_Toc76217044)

[2.3 Data EDA with VIS 7](#_Toc76217045)

[(1) What was the moving average of the stock MSFT? 7](#_Toc76217046)

[(2) What was the daily return of the stock on average? 7](#_Toc76217047)

[(3) What was the correlation between sales volume and stock price? 8](#_Toc76217048)

[(4) How much value do we put at risk by investing in MSFT? 10](#_Toc76217049)

[2.4 ML models used and how they work 10](#_Toc76217050)

[2.4.1 Long Short Term Memory networks 10](#_Toc76217051)

[2.4.2 Gated Recurrent Units 12](#_Toc76217052)

[3. Results 12](#_Toc76217053)

[3.1 Long Short Term Memory networks 12](#_Toc76217054)

[3.2 Gated Recurrent Units 14](#_Toc76217055)

[3.3 Model Comparison 15](#_Toc76217056)

[4. Conclusions 15](#_Toc76217057)

1. Introduction

Stock analysis is a method for investors and traders to make buying and selling decisions. By studying and evaluating past and current data, investors and traders attempt to gain an edge in the markets by making informed decisions.



There are two basic types of stock analysis: fundamental analysis and technical analysis. Fundamental analysis concentrates on data from sources, including financial records, economic reports, company assets, and market share. To conduct fundamental analysis on a public company or sector, investors and analysts typically analyze the metrics on a company’s financial statements – balance sheet, income statement, cash flow statement, and footnotes. These statements are released to the public in the form of a 10-Q or 10-K report through the database system, EDGAR, which is administered by the U.S. Securities and Exchange Commission (SEC). Also, the earnings report released by a company during its quarterly earnings press release is analyzed by investors who look to ascertain how much in revenue, expenses, and profits a company made.



Technical analysis focuses on the study of past and present price action to predict the probability of future price movements. Technical analysts analyze the financial market as a whole and are primarily concerned with price and volume, as well as the demand and supply factors that move the market. Charts are a key tool for technical analysts as they show a graphical illustration of a stock’s trend within a stated time period. For example, using a chart, a technical analyst may mark certain areas as a support or resistance level. The support levels are marked by previous lows below the current trading price, and the resistance markers are placed at previous highs above the current market price of the stock. A break below the support level would indicate a bearish trend to the stock analyst, while a break above the resistance level would take on a bullish outlook.

Technical stock analysis is effective only when supply and demand forces influence the price trend analyzed. When outside factors are involved in a price movement, analyzing stocks using technical analysis may not be successful. Examples of factors, other than supply and demand, that can affect a stock price include stock splits, mergers, dividend announcements, a class action lawsuit, death of a company’s CEO, a terrorist attack, accounting scandals, change of management, monetary policy changes, etc.

1. Analysis Section
   1. About the data

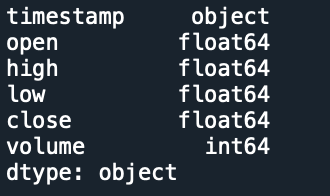
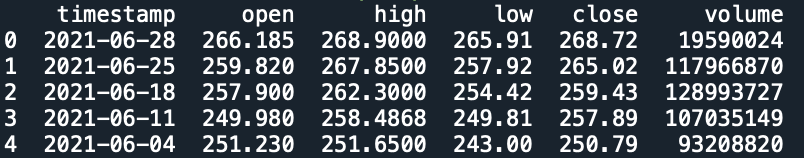
The dataset is the stock price of Microsoft company from 1999 to 2021. It is scraped from the website Alpha Vantage with its API key. The raw data contains 1130 rows and 6 columns (features). The variables include:

|  |  |
| --- | --- |
| Timestamp | The date |
| Open | The opening price of the stock |
| High | The high price of that day |
| Low | The low price of that day |
| Close | The closed price of that day |
| Volume | The amount of stocks traded during that day |

The objective is to predict future stock prices through a Long Short Term Memory (LSTM) method.

* 1. Data cleaning with VIS

First, query the data types of the columns in the dataframe using the dataframe attribute .dtypes.

﻿The .dtypes attribute indicates that the data columns in the pandas dataframe are stored as several different data types as follows:

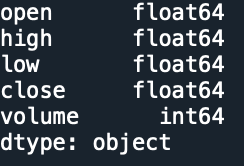
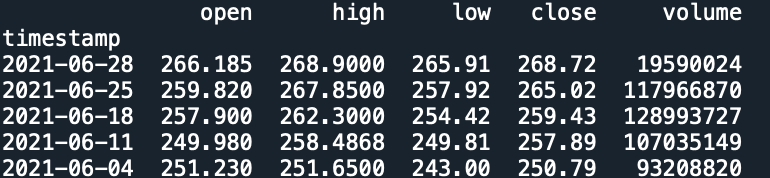
* Timestamp-object: A string of characters that are in quotes.
* Volume-int6: This is a numeric value that will never contain decimal points.
* Open/high/low/close-float64: This data type accepts data that are a wide variety of numeric formats including decimals (floating point values) and integers.

Plot date as strings:



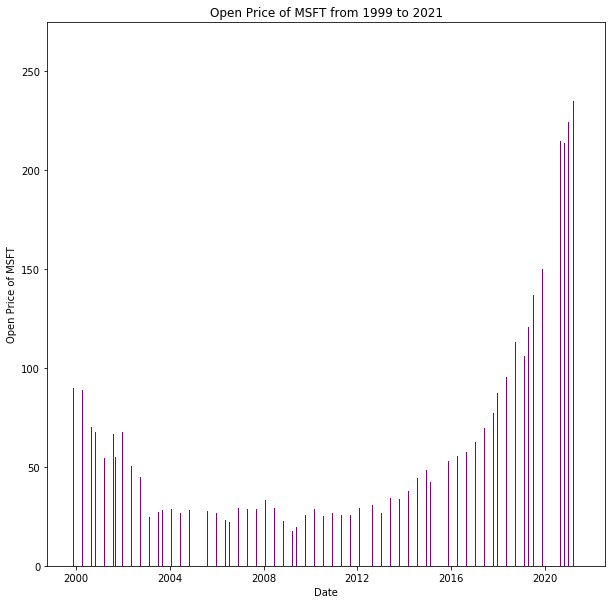
﻿Look closely at the dates on the x-axis. When plot a string field for the x-axis, Python gets stuck trying to plot the all of the date labels. Each value is read as a string, and it is difficult to try to fit all of those values on the x axis efficiently. Therefore, to avoid this problem, analysts can convert the dates from strings to a datetime object during the import of data into a pandas data frame. Once the dates are converted to a datetime object, analysts can more easily customize the dates on the plot, resulting in a more visually appealing plot.

﻿Import Date Column into Pandas Data frame As Datetime Object:

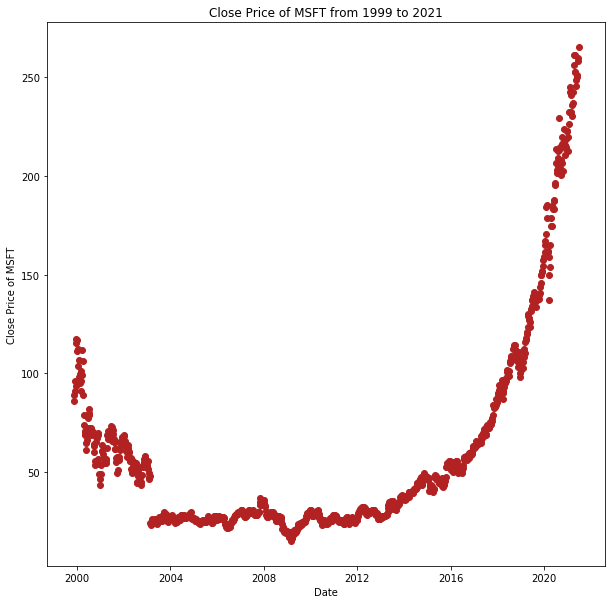
﻿To import the dates as a datetime object, analysts can use the parse\_dates parameter of the pd.read\_csv() function that allows to indicate that a particular column should be converted to a datetime object. As there is a single column that contains dates in the data, analysts can set dates as the index for the data frame with the index\_col parameter.

1. Opening price



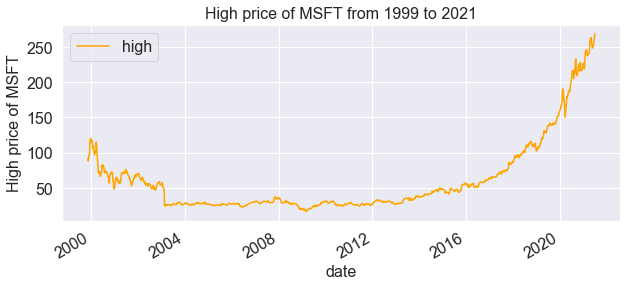
Compared with the former plot about opening price of the stock, the picture above is apparently clearer and more appealing. The opening price of MSFT decreased slowly from 2000 to 2004 and then entered a stable period. Since 2004, the price remained around 30 dollars per share. However, after 2016, the price started to increase gradually from 50 dollars to 250 dollars. On the whole, the opening price faced an upward trend.

1. Closed price



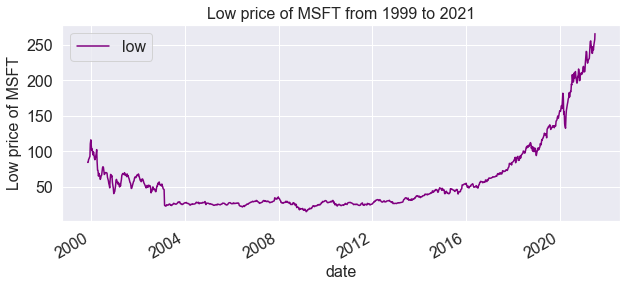
The closed price of MSFT shows a similar trend with the opening price. It decreased from 125 dollars to 25 dollars since 1999 and remained stable from 2004 to 2012. Then the price increased rapidly to over 250 dollars in 2021. In addition, the plot shows no outliers and incorrect values.

1. High price



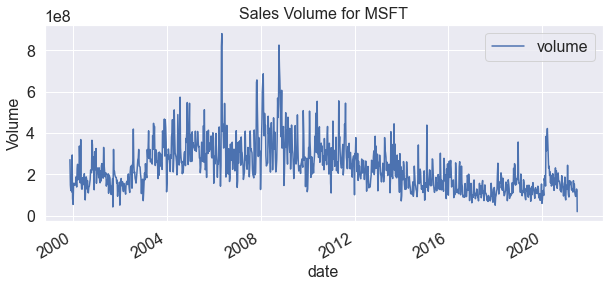
From 1999 to 2004, the high price of MSFT decreased from 125 dollars to 25 dollars. Since 2004, the range of high price was between 25 and 50 dollars. However, after 2016, the stock price increased significantly and reached its highest point at around 260 dollars. Moreover, the plot shows no outliers and incorrect values.

1. Low price



Like what shows in high price plot, the low price of MSFT also decreased gradually from 1999 to 2004 and increased dramatically from 2016 to 2021. The lowest price in 1999 was around 90 dollars and the lowest price in 2021 is around 260 dollars. There is no outliers and incorrect values in the plot.

1. Volume

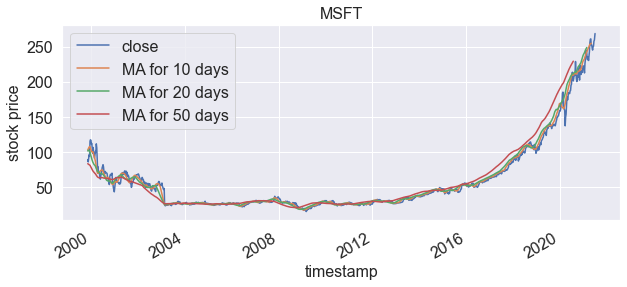


The overall trend of sales volume is downward from 1999 to 2021. It reached its peak in 2006, but dropped dramatically in the later period. Though the volume fluctuated between 0.2 trillion and 0.8 trillion from 2006 to 2016, it remained around 0.1 trillion since 2016.

According to the plots of each variable, no outliers and incorrect values show. There is no missing data neither. Therefore, the dataset is clean now.

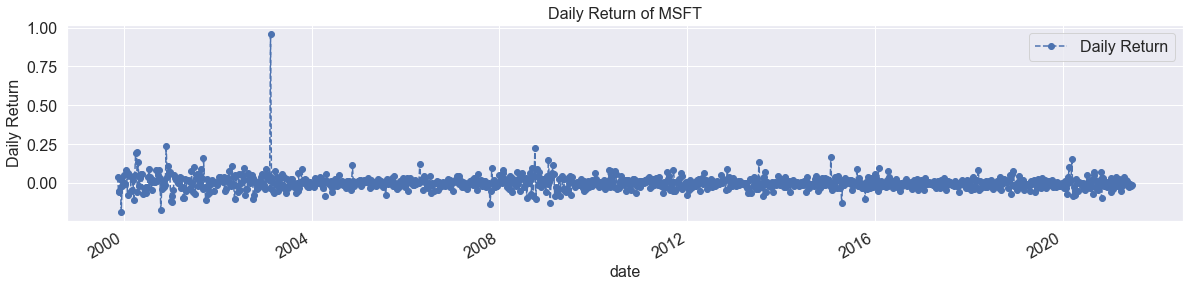
* 1. Data EDA with VIS

1. What was the moving average of the stock MSFT?



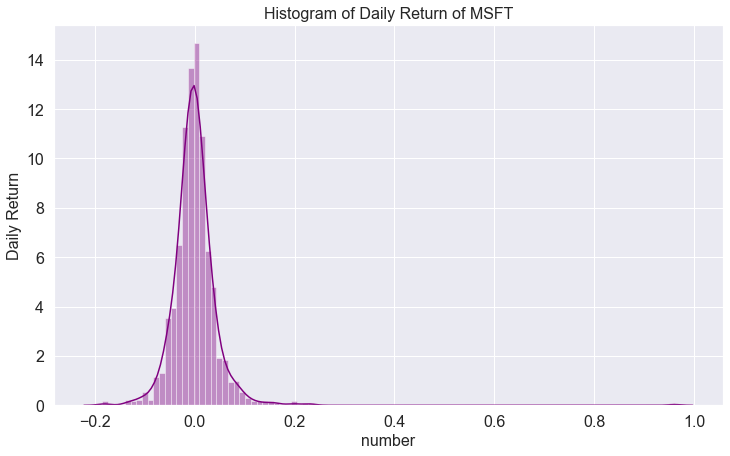
The moving average for 10 days is relatively smoother than the moving average for 20 or 50 days. It decreased from 80 dollars to 25 dollars from 1999 to 2004 and remained stable from 2004 to 2016. Then it constantly increased to 230 dollars in 2021. The moving average for 50 days fluctuated more than that for 20 days.

1. What was the daily return of the stock on average?



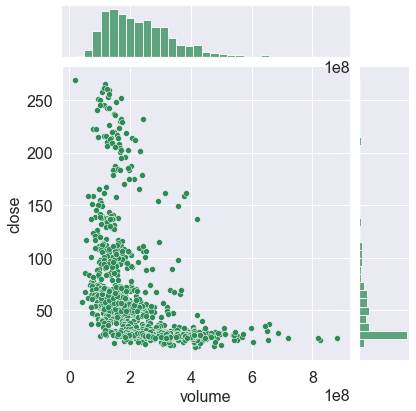
To analyze the risk of the stock, the analysts can take a closer look at the daily changes of the stock, and not just its absolute value. The plot shows that the daily return keeps stable from 1999 to 2021. However, it is extremely high in 2004

Then, take an overall look at the average daily return using a histogram.



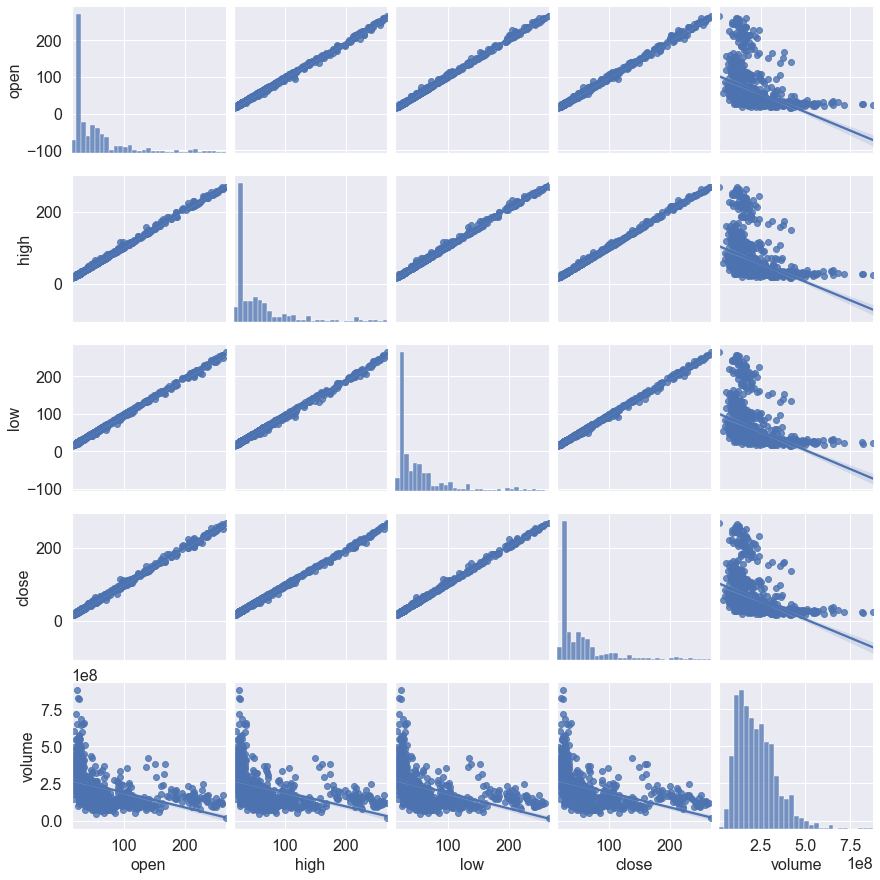
The histogram of daily return shows that the daily return of MSFT within 22 years is mainly around 14 dollars. The overall trend looks like normally distributed.

1. What was the correlation between sales volume and stock price?

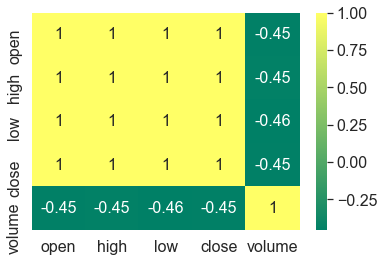


The above plot displays the relationship between the volume and closed price. It shows that when the closed price of MSFT is below 50 dollars, the volume is likely to be over than 0.4 trillion. When the price is above 150 dollars, the volume will be lower than 0.4 trillion. Therefore, the stock price and volume are negatively correlated with each other.

Seaborn and pandas make it very easy to repeat this comparison analysis for every possible combination of price of stocks and volume. Analysts can use sns.pairplot() to automatically create this plot



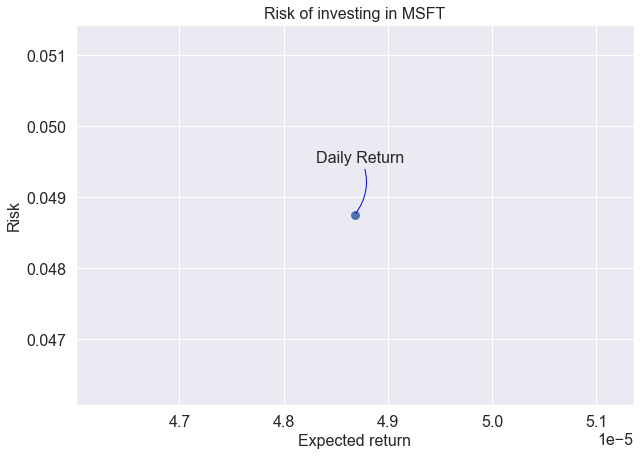
The plot provides a quick glance to show all the relationships on volume between all kinds of stock price. The opening price, low price, high price, and closed price shares a linear relationship between each other, which means that they are strongly positively correlated. The volume has negative relationships with these prices, which reflects that when the stock price is low, people are more likely to buy the stocks.



Analysts could also do a correlation plot, to get actual numerical values for the correlation between the stock price and the volume. Same as the result in the previous pair plot, the correlation plot above shows the strongest correlation among all kinds of stock price and demonstrates the negative relationship between the volume and them.

1. How much value do we put at risk by investing in MSFT?

There are many ways analysts can quantify risk, one of the most basic ways with the information on daily percentage returns is by comparing the expected return with the standard deviation of the daily returns.

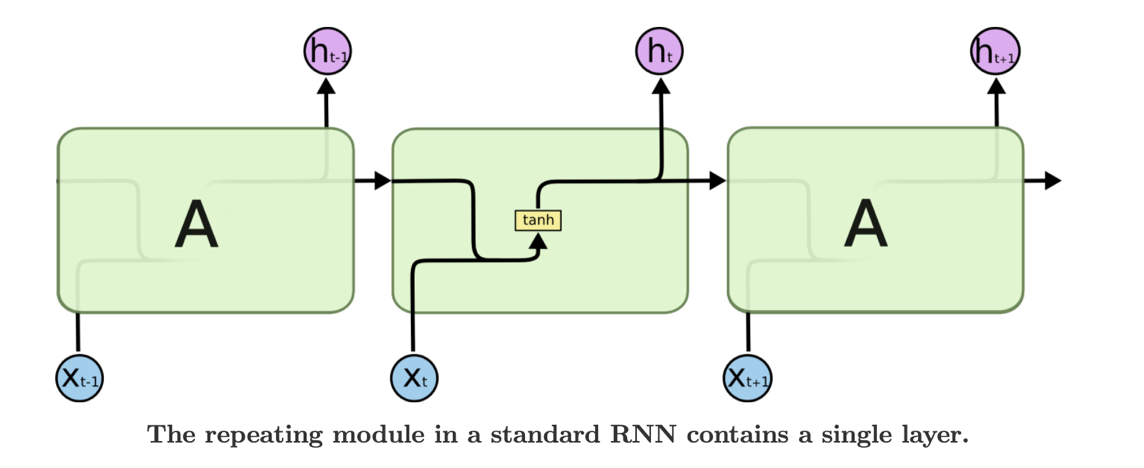


The risk has a positive relationship with the expected return. When the expected return increases, the risk will also increase. When the expected return becomes close to 0.49 million, the risk will increase at a faster speed. Besides, the expected return of MSFT is between 0.48 million and 0.49 million. The risk is around 0.049.

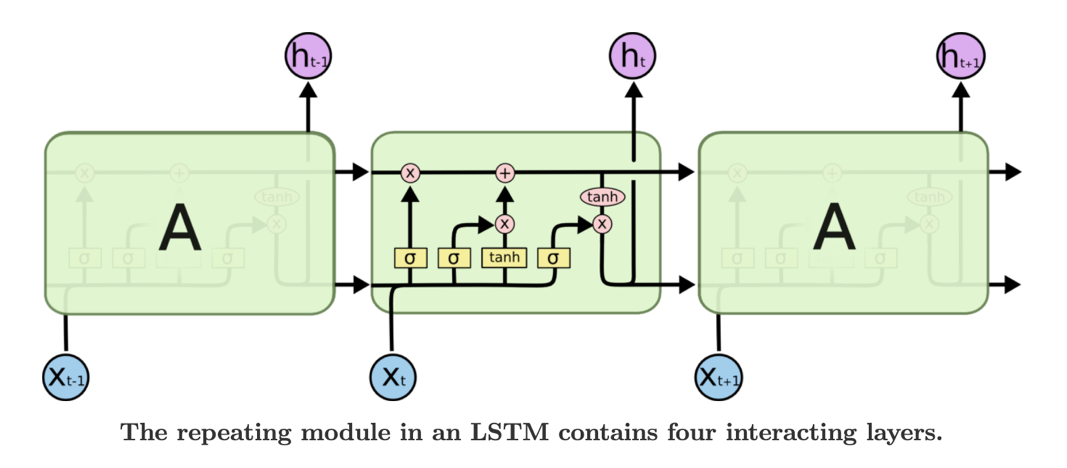
* 1. ML models used and how they work
     1. Long Short Term Memory networks

Long Short Term Memory networks – usually just called “LSTMs” – are a special kind of RNN, capable of learning long-term dependencies. They are explicitly designed to avoid the long-term dependency problem. Remembering information for long periods of time is practically their default behavior, not something they struggle to learn.

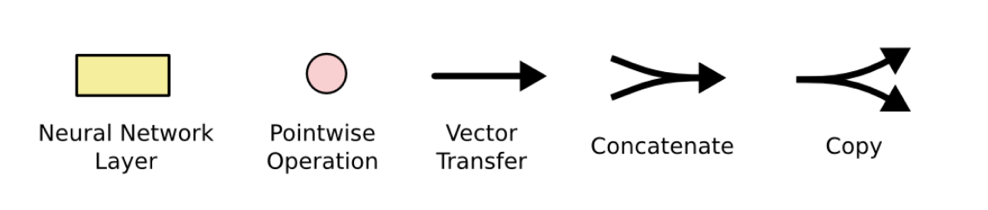
All recurrent neural networks have the form of a chain of repeating modules of neural network. In standard RNNs, this repeating module will have a very simple structure, such as a single tanh layer.



LSTMs also have this chain like structure, but the repeating module has a different structure. Instead of having a single neural network layer, there are four, interacting in a very special way.



In the diagram below, each line carries an entire vector, from the output of one node to the inputs of others. The pink circles represent pointwise operations, like vector addition, while the yellow boxes are learned neural network layers. Lines merging denote concatenation, while a line forking denote its content being copied and the copies going to different locations.



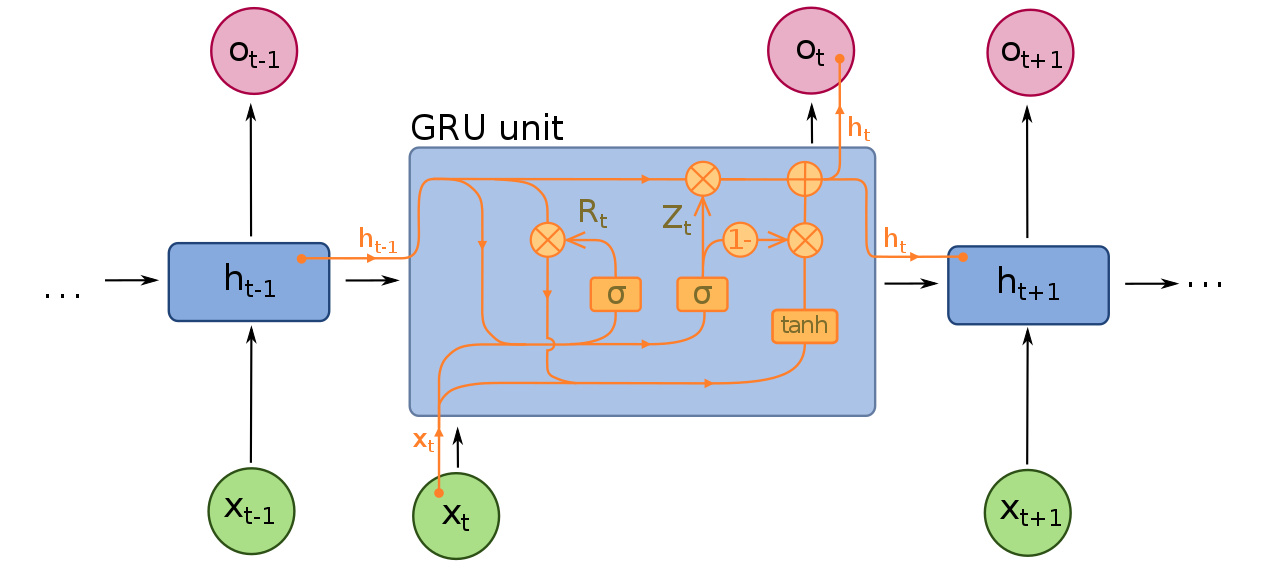
The key to LSTMs is the cell state, the horizontal line running through the top of the diagram. The cell state is kind of like a conveyor belt. It runs straight down the entire chain, with only some minor linear interactions. It’s very easy for information to just flow along it unchanged. What’s more, the LSTM does have the ability to remove or add information to the cell state, carefully regulated by structures called gates. Gates are a way to optionally let information through. They are composed out of a sigmoid neural net layer and a pointwise multiplication operation.

2.4.2 Gated Recurrent Units

In simple words, the GRU unit does not have to use a memory unit to control the flow of information like the LSTM unit. It can directly make use of all hidden states without any control. GRUs have fewer parameters and thus may train a bit faster or need less data to generalize. But, with large data, the LSTMs with higher expressiveness may lead to better results.

Also, GRUs address the vanishing gradient problem (values used to update network weights) from which vanilla recurrent neural networks suffer. If the grading shrinks over time as it back propagates, it may become too small to affect learning, thus making the neural net untrainable. If layer in a neural net can’t learn, RNN’s can essentially “forget” longer sequences.

GRUs solve this problem through the use of two gates, the update gate and reset gate. These gates decide what information is allowed through to the output and can be trained to retain information from farther back. This allows it to pass relevant information down a chain of events to make better predictions.

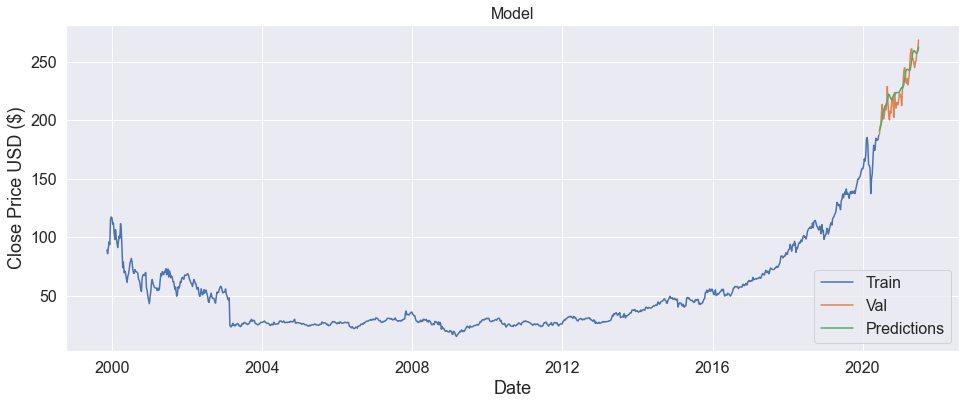


1. Results
   1. Long Short Term Memory networks



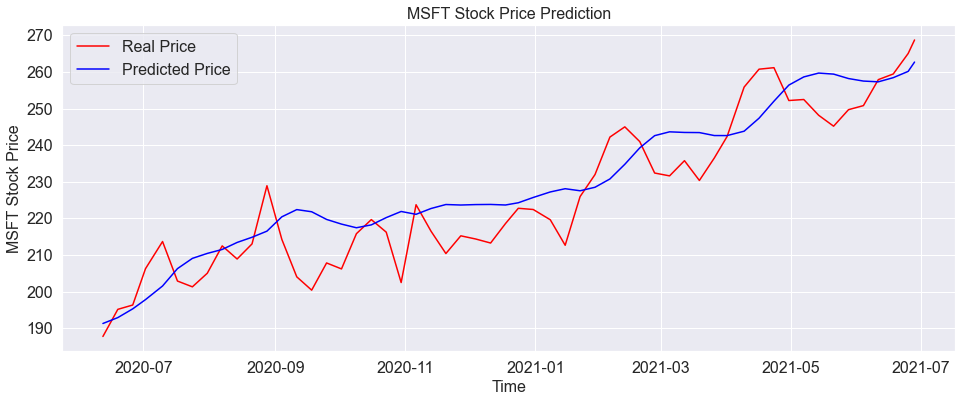
First, check the history of closed stock price of MSFT:

* From 2000 to 20004, the stock price fluctuated to decline from 125 dollars to 25 dollars.
* From 2004 to 2016, it kept stable around 20-50 dollars.
* After 2016, the stock price of MSFT increased dramatically to over 250 dollars.



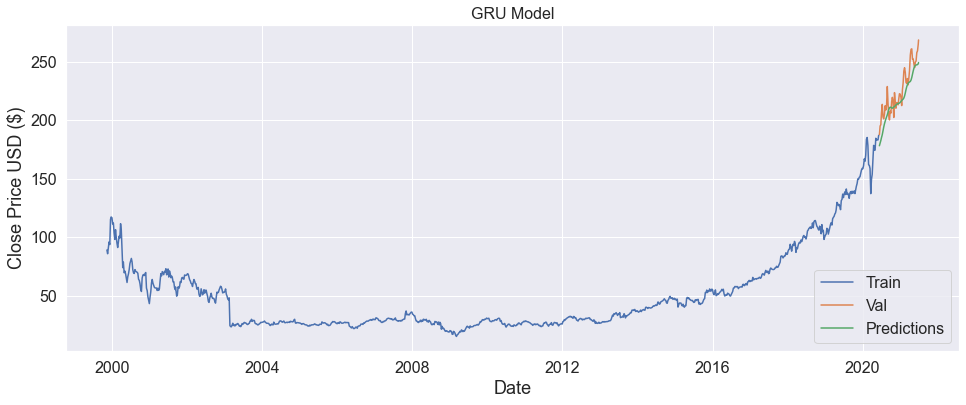
The line plot shows the predictions of MSFT stock price in the future. The green line reflects that the price will still increase to about 260 dollars with the current growth rate.

As for the model evaluation, the real price and predicted price of MSFT will be plotted and compared as below.



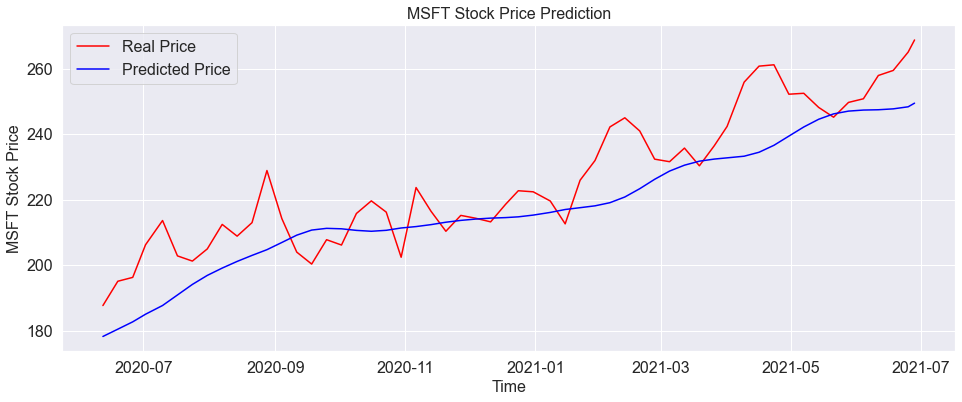
The comparison shows that the line of predicted price roughly accords with the line of real price. Both of then face an upward trend. However, the blue line looks smoother than the red line, which means that the model did not perform well for the short period and failed to depict the fluctuations between years. The root mean squared error of the LSTM model is 9.0798.

* 1. Gated Recurrent Units



The GRU model shows a similar result with the LSTM model. It also predicts that the stock price of MSFT will increase to about 260 dollars.

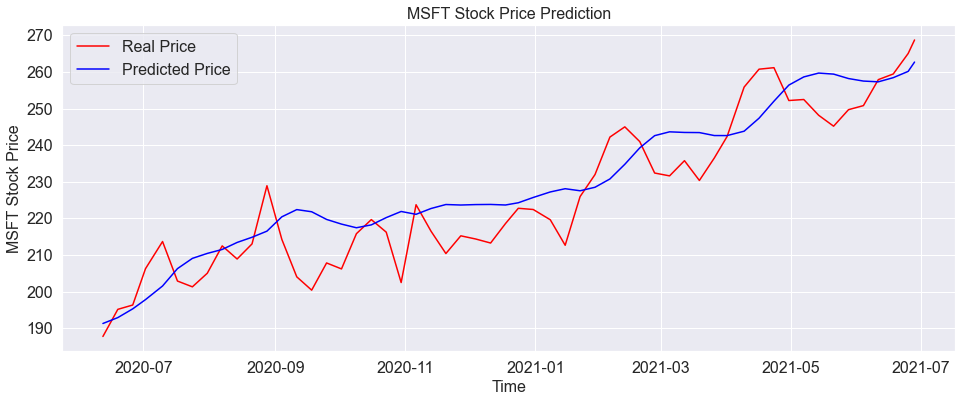
As for the model evaluation, the real price and predicted price of MSFT will be plotted and compared as below.



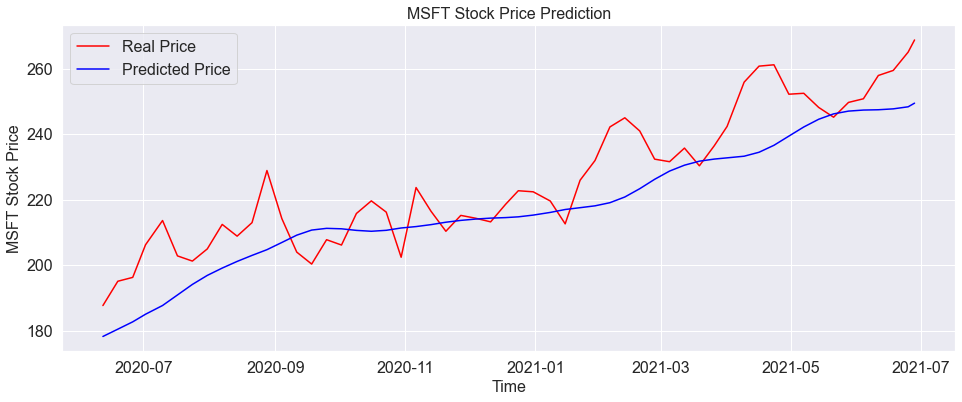
The comparison shows that the line of predicted price roughly accords with the line of real price. Both of then face an upward trend. However, the blue line seems lower than the red line, which means that the model tends to underestimate the stock price. What’s more, the blue line looks smoother than the red line, which reflects that the model did not perform well for the short period and failed to depict the fluctuations between years. The root mean squared error of the GRU model is 12.2692.

* 1. Model Comparison

**LSTM model:**



**GRU model:**



Comparing the performance of two models, LSTM seems to perform better than GRU. In the first picture, the blue line is also closer to the red line than the blue line to the red line in the second plot. However, both of them failed to depict the fluctuations of stock price from 2020 to 2021.

Root mean squared error:

|  |  |  |
| --- | --- | --- |
|  | LSTM | GRU |
| Root mean squared error | 9.0798 | 12.2692 |

The root means squared error table also shows that LSTM outperforms GRU in this case as the RMSE of LSTM is lower than that of GRU. The LSTM mode is more accurate when using datasets with longer sequences and displays much greater volatility throughout its gradient descent compared to the GRU model.

For the advantages of the GRU model, it is able to train 3.84% faster than the LSTM model. For future work, different kernel and recurrent initializers could be explored for each cell type. To sum up, LSTM is good at longer sequences while GRU has used fewer tensor operations and takes less time to train. The results of the two, however, are almost the same.

1. Conclusions

The stock price of MSFT decreased from 2000 to 2004 and then entered a stable period. Since 2004, the price remained around 30 dollars. After 2016, the price started to increase dramatically from 50 dollars to 250 dollars. On the whole, the price faced an upward trend.

In recent news, Microsoft introduced its Windows 11 personal computer operating system. Microsoft stock rose 0.5%. Windows 11 features a refreshed design with a new user interface and Start menu. It also provides PC performance improvements and integrates the Teams videoconferencing app. Besides, Microsoft announced a deal to buy Nuance Communications (NUAN) for $19.7 billion on April 12. The acquisition of Nuance will give Microsoft more heft in the health-care sector. Microsoft stock rose a fraction on the news. And, on March 31, Microsoft received a contract to supply over 120000 Microsoft HoloLens augmented-reality headsets to the U.S. Army. The contract is part of the Integrated Visual Augmentation System (IVAS) program. The deal could be worth up to $21.88 billion over 10 years. MSFT stock climbed 1.7% on the report. On Feb. 24, Microsoft announced three new industry-specific cloud offerings. They included versions of Microsoft Cloud for financial services, manufacturing, and nonprofits. It also previewed a version for retail. Plus, it provided the first update to Microsoft Cloud for Healthcare. Microsoft stock rose 0.6% on the news.



Combing the news and model analysis, the stock price of Microsoft is predicted to go on increasing to about 260 dollars.