

## A project by Team Pandas

Trading has been one of the earliest activities of mankind and continues to evolve. The first modern stock trade began when a company first publicly traded in order to raise capital. It has become a regular practice in how the most modern company set up their business. Stock price indicates its current value to buyers and sellers. In most cases, when there is more buying, it indicates that the value is high. a high opening price might indicate that the value is low or that sometimes the value is high. For this project, we investigated three companies and predicts the accuracy of a stock price a day ahead and the company that is undervalued

### Aims and Objective

The aim of the project is to perform an EDA(Exploratory Data Analysis) on the stock prices of three companies (Amazon, Apple, and Microsoft) and build a model that will predict the next day's stock price. The model should be deployed to make live predictions and can be achieved by training a machine learning model with the past closing price.

### Flow process of this project



- Data Source
- Exploratory data analysis (EDA)
- Feature engineering
- Model building
- Model Deployment

- **Data source**

This dataset is obtained from Kaggle [New York Stock Exchange | Kaggle](#). It focuses on the historical stock prices of tech companies.

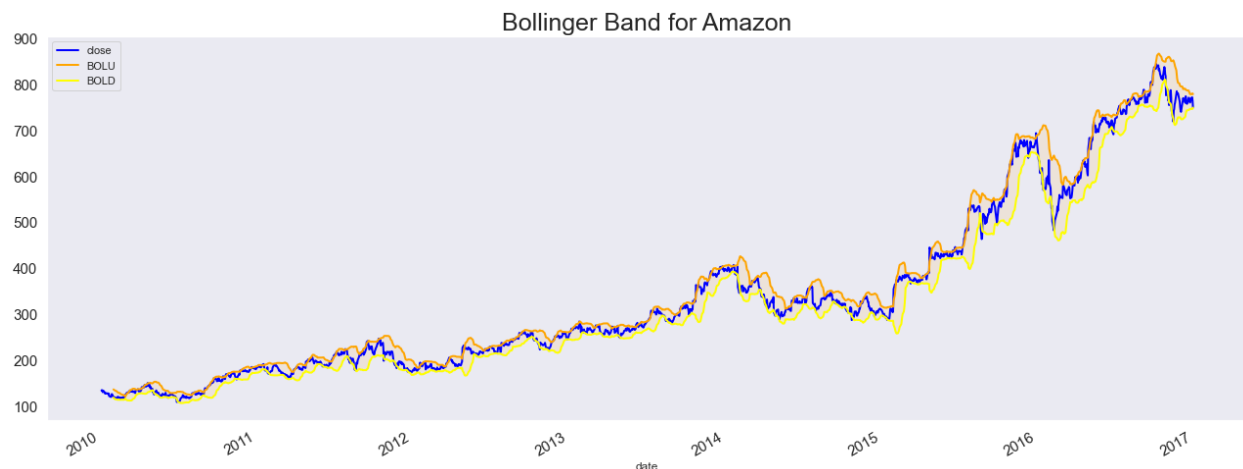
This dataset consists of 851264 rows and 7 columns, it contains the opening and closing of stock prices, the low and high price of stock price, and its volume.

- **Exploratory data analysis**

This is a crucial step, it examines how the value of different variables are distributed, discovers the patterns, and checks assumptions with graphical representations.

The Bollinger Band can be used to determine the volatility of an asset and identify the trend using overbought and oversold strategies. The tightness of the Bollinger band shows that the asset is less volatile because the tightness of Bollinger is inversely proportional to Volatility.

### **For Amazon**

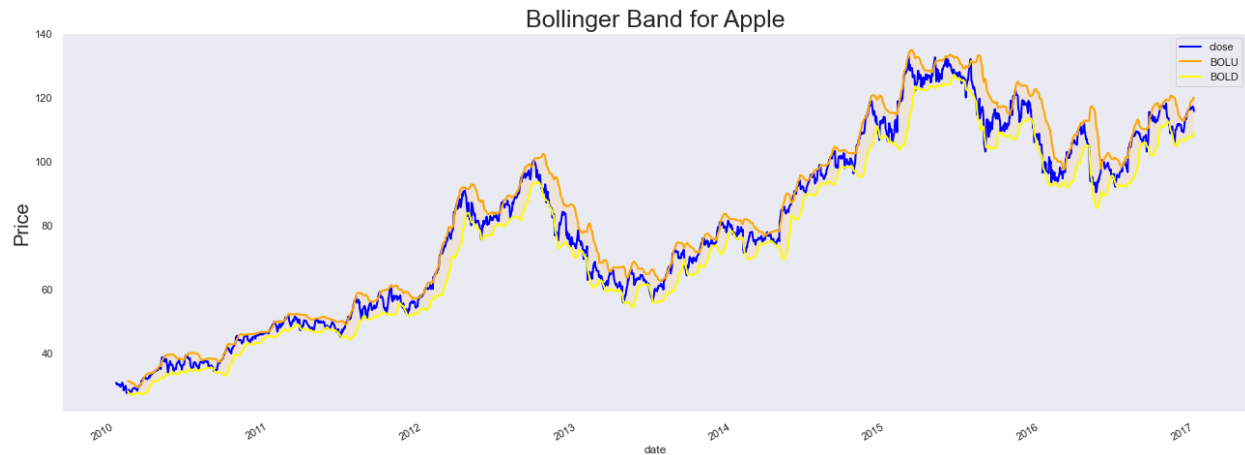


Amazon has a very tight Bollinger which shows that the market was less volatile. The Chart above shows the Bollinger band of amazon, it has a better visual so as to be able to visualize the point of bounce and pullback. The price of assets breaking below the lower band indicates oversold which was responsible for the bounce. On the other hand, a point around may show a pullback when the price crosses the upper band.

### **For Apple,**

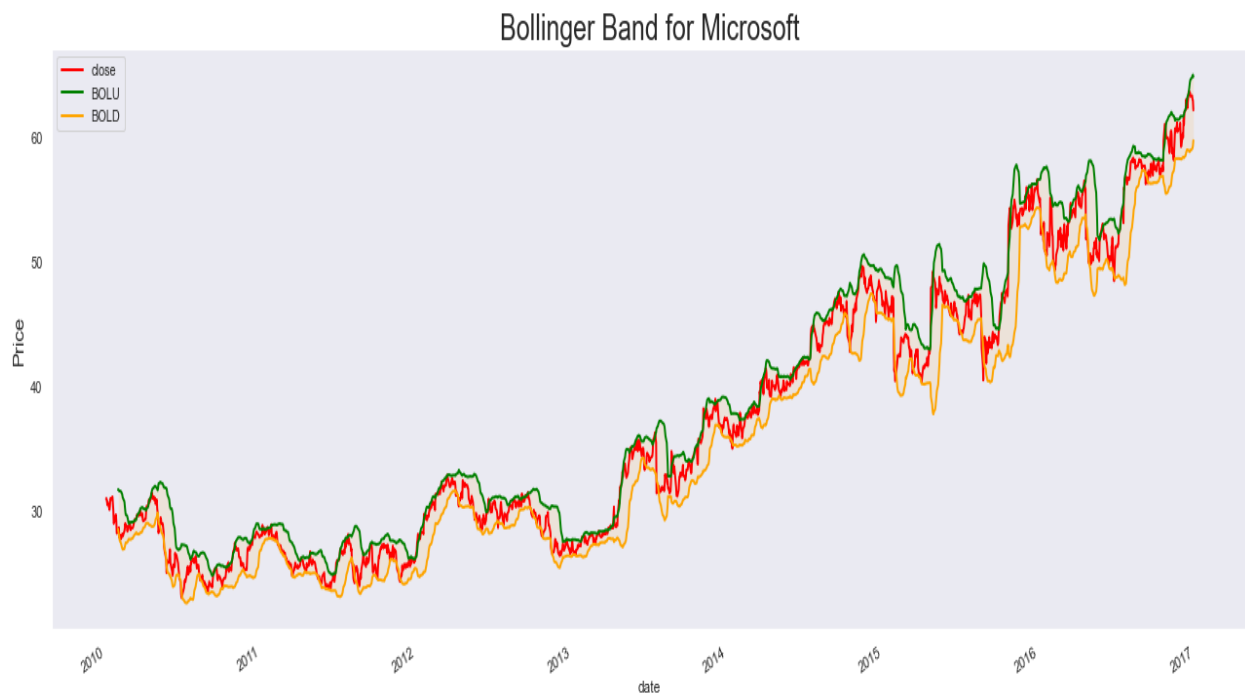
The Bollinger Band is not too tight and this indicates high volatility. The Chart below shows the Bollinger band of Apple, it has a better visual so as to be able to visualize the point of bounce and pullback. The price of assets breaking below the lower band indicates oversold which was

responsible for the bounce. On the other hand, a point around June shows a pullback when the price crosses the upper band.



## For Microsoft,

Microsoft has a narrow Bollinger which shows the volatility of the asset. The Chart below shows the Bollinger band of Microsoft, it has a better visual so as to be able to visualize the point of bounce and pullback. The price of assets breaking above the Upper band indicates a point of overbought which was technically responsible for the pullback afterward. On the other hand, a point around July shows a bounce when the price crosses the lower band.



- **Feature Engineering**

In this stage, we filled out a closed column and reshaped the data in three dimensions. After which, the data were scaled using the MinMax Scaler.

This method was chosen to help deal with outliers and also to optimize machine learning performance.

- **Model building and deployment**

Time series is a special type of data set in which one or more variables are measured over time. In time series, however, observations are measured over time. Each data point in the data set corresponds to a point in time. This means that there is a relation between different data points of the dataset. This has important implications for the types of machine learning algorithms that can apply to the time series dataset. LSTM MODEL was developed and deployed for live predictions.

### ***LSTM (Long Short-Term Memory)***

LSTMs are Recurrent Neural Networks. Neural Networks are very complex machine learning models that pass input data through a network. Each node in the network learns a very simple operation. The neural network consists of many such nodes. The fact that the model can use a large number of simple nodes makes the overall prediction very complex. Neural Networks can therefore fit very complex and nonlinear data sets.

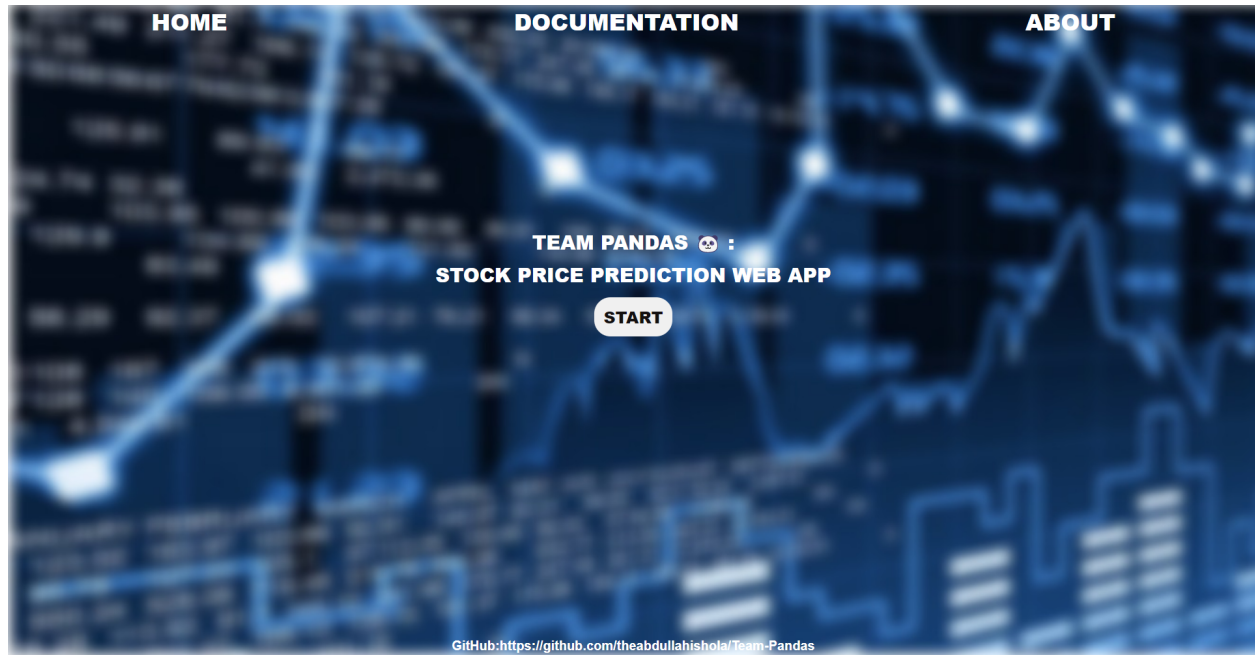
RNNs are a special type of Neural network, in which the network can learn from sequence data. This can be useful for multiple use cases, including understanding time series (which are clearly sequences of values over time), but also text (sentences are sequences of words).

LSTMs are a specific type of RNN. They have proven useful for time series forecasting on multiple occasions. They require some data and are more complicated to learn than supervised models.

LSTM is the preferred model. Other models applicable include ARIMA, Prophet, DeepAR.

The model was deployed on a dedicated website for users to draw insights before making decisions as to the three top tech companies

Link: <https://theabdullahishola.github.io/HDS/>



## Results



date	Close	Prediction
2014-11-26	333.570007	324.091461
2014-11-28	338.640015	325.856598
2014-12-01	326.000000	327.670654
2014-12-02	326.309998	328.113281
2014-12-03	316.500000	327.845886
2014-12-04	316.929993	326.285309
2014-12-05	312.630005	324.252960
2014-12-08	306.640015	321.754517

Using AMAZON as an example, the difference between the Predicted and actual values is not large which reflects the accuracy of the model. The model has an RMSE of 29.78.

## Conclusion and recommendation

Insights about the stock market are needed to grow a financially stable business, hence the need for analysis and use of machine learning algorithms to help with business decisions through forecasting.

After an in-depth Analysis of the stocks, A deep learning algorithm, LSTM, was used to make predictions to help forecast the prices of stocks. However, deep learning models work better with larger datasets, hence for better analysis, a larger dataset will be preferred.

## REFERENCES

1. Joos Korstanje | **How to select a Model For Time Series** | Updated July 22nd, 2022  
<https://neptune.ai/blog/select-model-for-time-series-prediction-task>
2. Wikipedia| **Long short-term memory (LSTM)**| Updated 16 August 2022  
[https://en.wikipedia.org/wiki/Long\\_short-term\\_memory](https://en.wikipedia.org/wiki/Long_short-term_memory)

## **TEAM MEMBERS**

Abdulhfiz Abdus-salaam  
Abdullah Ishola Hammed  
Adesokan Sulaimon Adewale  
Agbata Dikson Ukolojo  
Ashinze Emmanuel Chidi  
Bukola Zainab Agoro  
Chibuike Lawrence Orji-oko  
Esther Chizitere Amadi  
Favour Emmanuel  
Favour Uwaishe  
Flora Okezue  
Joseph Komolafe  
Modupeoluwa Adedeji  
Muhammed Balogun  
Oghenekaro Oghoroko  
Okwudiri Godswill Nwagu  
Oluwasayo Akinkunmi  
Poucheu Mbonda Florian-wilfred  
Quadiri Omotesho  
Ramsey Ithacius Njema  
Subomi Oyewole  
Tunde Egunjobi  
Victoria Ihenetu