

Stock Price Prediction and Analysis

Anudeep Verma^{#1}

Faisal Aziz^{*2}

Priyanshu Jain^{#3}

Venkatakrisnan VK^{#4}

IIT Bombay

Center of Data Science and Machine Intelligence

DS 203

I. INTRODUCTION

This project aims to predict the stock prices given the data of the closing price values of the previous days. We are using various models like -

- Auto-regression model
- LSTM (Long Short-Term Memory)
- N-beats
- Linear Regression

II. DATA DESCRIPTION

The Dataset includes Apple Stock Prices from 2009-01-01 to today (2021-25-11)

There are 6 attributes:

1. **Open** : Datatype - float64 and have zero null values
2. **High** : Datatype - float64 and have zero null values
3. **Low** : Datatype - float64 and have zero null values
4. **Close** : Datatype - float64 and have zero null values
5. **Adjusted Close** : Datatype - float64 and have zero null values
6. **Volume** : Datatype is I64 and have zero null values

Each attribute has 3248 entries. Our **aim** is to predict the Close attribute from the data given.

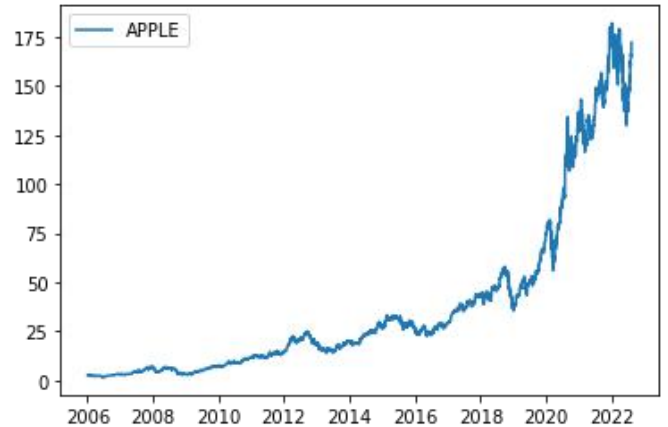


Fig. 1 The Dataset

III. EXPLORATORY DATA ANALYSIS

An easy way to comply with the conference paper formatting requirements is to use this document as a template and simply type your text into it.

A. Simple Moving Averages

Moving Average (MA-x) is basically the movement of the mean closing value of the previous 'x' days.

We have used MA-7 and MA-21 for this project. Also, we have used this MA-21 for the calculation of the upper and lower Bohlinger Bands.

B. Exponential Moving Averages

It is a type of moving average (MA) that places greater weight and significance on the most recent data points. The exponential moving average (EMA) is also referred to as the exponentially weighted moving average

We have used EMA-12 and EMA-26 for this project.

C. Bohlinger Bands

These are used to analyse where the stock prices have peaked or troughed, such that its probability of the stock price falling down or going up respectively is high enough. Hence from a trader's perspective –

- Sell Point: If price crosses upper band
- Buy Point: If price crosses lower band

This helps us visualize the probability of the price movement based on the current location of the price inside the band.

It is calculated :

- ✧ Upper band= $(MA-21) + 2*(SD-20)$
- ✧ Lower band= $(MA-21) - 2*(SD-20)$



Fig. 2 MA, Exponential MA and Bollinger Bands for APPLE stock

D. Fourier Transform

We also applied fourier transform over closing price of stock, taking increasing number of components of the series.

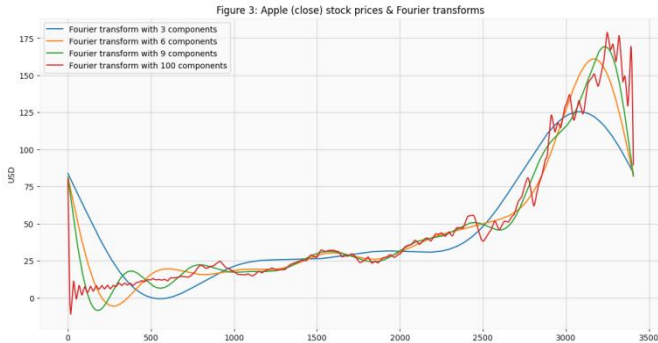


Fig. 3 Apple (close) stock price & Fourier Transform

E. Candle Stick Graph

Finally, visualized the data into a more familiar candlestick pattern.

Overall the graph looks as follows:



Fig. 4 Candle Stick Graph

IV. MODELS USED

We implemented models as stated above as follows:

➤ Linear Regression

We iterate through 4 to 11 features to find the optimum number of features to take for test dataset prediction and we got the optimal number as 7 to be used for predicting closing price of Apple stock. They were : 12ema, MACD, 20sd, upper_band, lower_band, ema and momentum.

Final performance on test data was as follows:

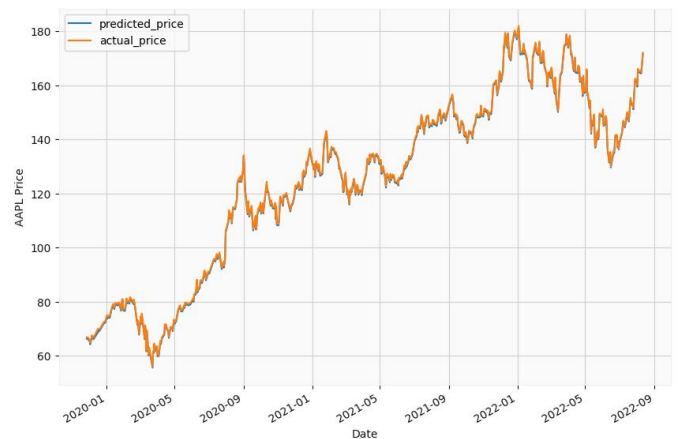


Fig. 5 Linear regression on test data

➤ LSTM

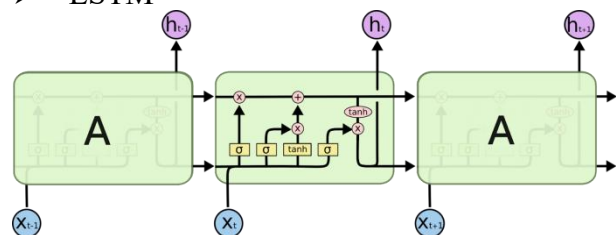


Fig. 6 LSTM architecture

Here, we used two LSTM layers having 500 neurons each and two dense layers having 16 and 1 neuron each for making prediction.

Further, we trained the model using different batch sizes- 64, 32 & 8 and found that the best performance comes for 64 batch size training.

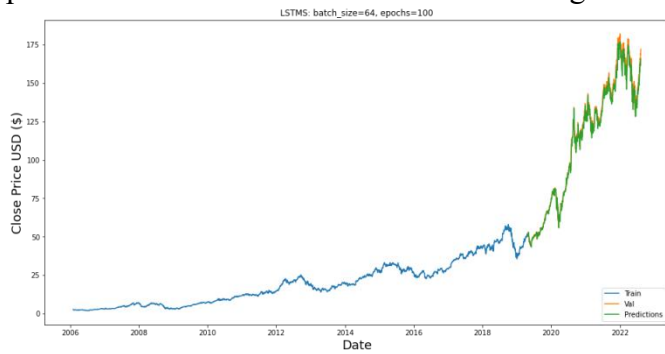


Fig. 7 LSTM on test data

➤ NBeats

It was a paper implementation of: “*N-BEATS: Neural basis expansion analysis for interpretable time series forecasting* published in *ICLR 2020*”

It had following advantages:

- ✧ Interpretable
- ✧ Applicable without modification
- ✧ Fast to train

N-beats uses entire window of past values as input. It has feed-forward network along with residual stacking mechanism, which allows it to stack many blocks without the risk of gradient vanishing. It also has the advantage of boosting.

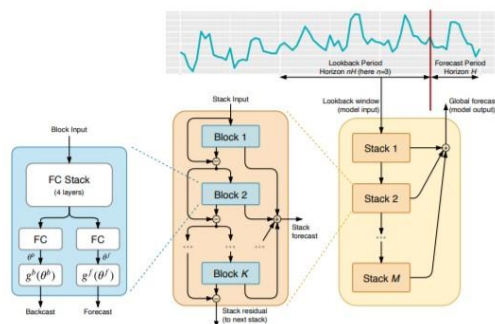


Fig. 8 NBeats architecture

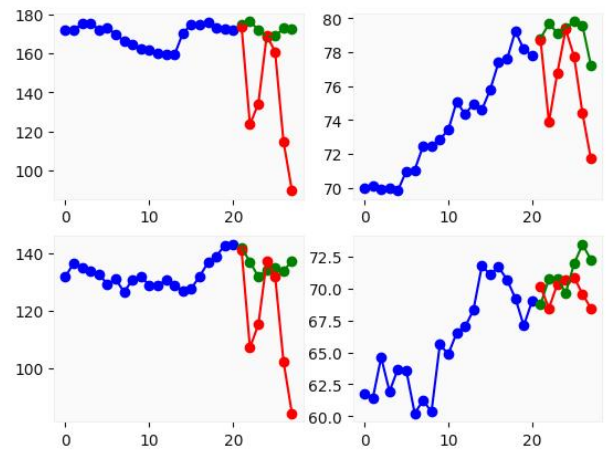


Fig. 9 NBeats on train and test data

➤ Autoregressive

It predicts future prices based on past performance. We forecast the interested variable using linear combination of past values of the variable. The term autoregression indicates that it is a regression of the variable against itself.

AR(1) is one in which the current value is based on the immediately preceding value, while an AR(2) process is one in which the current value is based on the previous two values.

$$y_t = c + \phi_1 y_{t-1} + \phi_2 y_{t-2} + \dots + \phi_p y_{t-p} + \epsilon_t,$$

Fig. 10 AR(P)

We took the lag of 6 while training the model. Meaning it accounts values upto 6 periods back.