Stock Price predictor with Recurrent neural nets

PROJECT OVERVIEW

The transition from traditional auction in 1970s to computerized transactions was fuelled with a need for an efficient access to the market. This set a stage for algorithmic and high-frequency trading. An estimated 70% of US equities in 2013 accounted for by automated trading [1]. Investment firms and hedge funds have used modelling to predict markets. They aim toward maximizing the return and spreading their risk over different financial assets. Traditional approaches can be classified as *Follow-the-Winner*, *Follow-the-Loose'*, *Pattern-Matching* and *Meta-learning*. Though these methods are efficient, their performance is dependent on the validity in different types of markets.

Quantitive strategies of hedge funds have received considerable returns over decades. Application of growing computing power and availability of big data has allowed models to identify and harvest on market inefficiencies [2]. These companies are actively looking for AI solution to outperform the benchmark indexes. In recent times a large ecosystem of start-ups has unfolded cantered around FinTech, with the aim of providing AI solutions for investment and saving.

The project aims towards utilizing Recurrent Neural Nets to predict the stock prices. The goal here is to leverage modelling abilities of neural networks for time series forecasting [3]. Time series forecasting is a difficult type of predictive modelling problem, as it has added complexity of sequence dependence among the input variables [7]. Recurrent neural networks have been quite successful in modelling time series data. LSTM is a type of RNN which can model short-term memory over a long period of time. LSTM enables us to model sequence dependence as short-term memory and incorporate it as a feature in our model.

PROBLEM STATEMENT
The goal of this project is to predict future adjusted price for a given stock across a given period. With vast amount of historical data, application of machine learning techniques

becomes more suitable here, to predict price trends for a stock being traded in a market. These models should be tested for accuracy and validated for performance. Because of the availability of wide range of historical data, a model can model can be trained over a long period of time and can make predictions with the time series.

This project aims to answer these questions

- Can a deep learning model beat a machine learning benchmark model for stock price prediction?
- Which architecture/ hyperparameters is gives an optimal result for the model?
- Can a simple deep learning model perform good enough to drive the investment strategies?

It is important to measure the quality of a model by quantifying its performance. Here we are going to use *Mean Squared Error* (MSE) and *Root Mean Squared Error* (RMSE). It is the difference between the price predicted by a model and actual price, a stock was trading on a day. Mathematical formula of RMSE is as depicted in the Fig 1. Using RMSE gives us consistency while comparing with other studies. Also, we will use graphical visualizations of different models to find degree of fit, ability to incorporate volatility and lag across time and other intrinsic micro trends from dataset.

$$RMSE = \sqrt{\frac{\sum_{i=1}^{N}(Predicted_i - Actual_i)^2}{N}}$$

Fig 1. Root Mean Square Error (RMSE)

ANALYSIS

*DATA EXPLORATION & EXPLORATORY VISUALIZATION*

In this project we are going to analyse stocks of Apple (AAPL), Google (GOOG), Amazon (AMZN) and SPY stock indices. The historical data for these entities was fetched from Yahoo finance application and is available in repository inside data folder. In this report we are going to discuss Apple stocks, to make the report brief and interesting. Although, a separate notebook for each entity having detailed analysis can be found inside Project folder. All the images of the report can be found in Project/images folder.

Here, we are using pandas to load a csv file for a given stock (in our case 'data/AAPL.csv'). Date column is parsed while loading and set as an index. As we are going to use 'Rolling mean and Adjusted closing price as a feature for our model, all the columns except this were

dropped. For rolling mean, window size of 10 is selected. It is evident from the Fig 2 that resolution of window size 10 is better than the other window size.

Fig 2. Comparison of different window size for Rolling mean
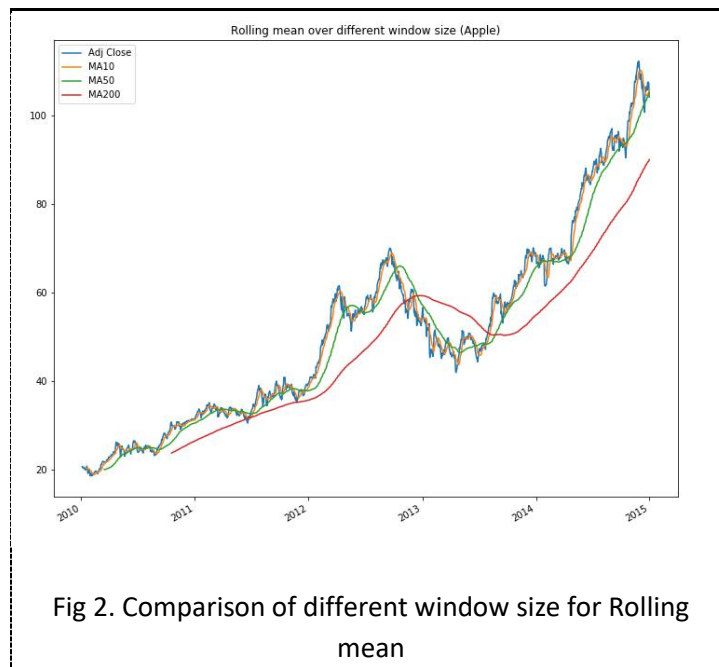
We are using past five-year data to create and validate models. Data ranges from Jan 2010 to Jan 2015. Fig. 3 depicts Adjusted closing price for stocks in these date range. For Adjusted closing price it looks like google has always been much more valuable as a company. But to really understand this we must look at total market cap of the company not just the stock prices. This information though absent, can be derived from total value traded for a given period. In Fig 3, we can also see volume trade plots for different stock, using both values, market cap of a company can be approximated.

Fig 3. Adjusted closing price for Apple, Google, Amazon and SPY(left)
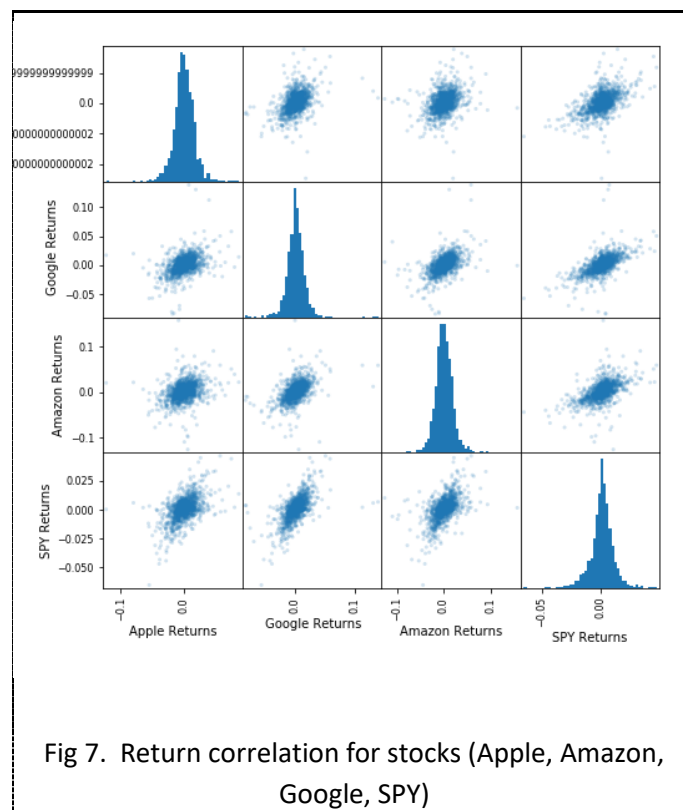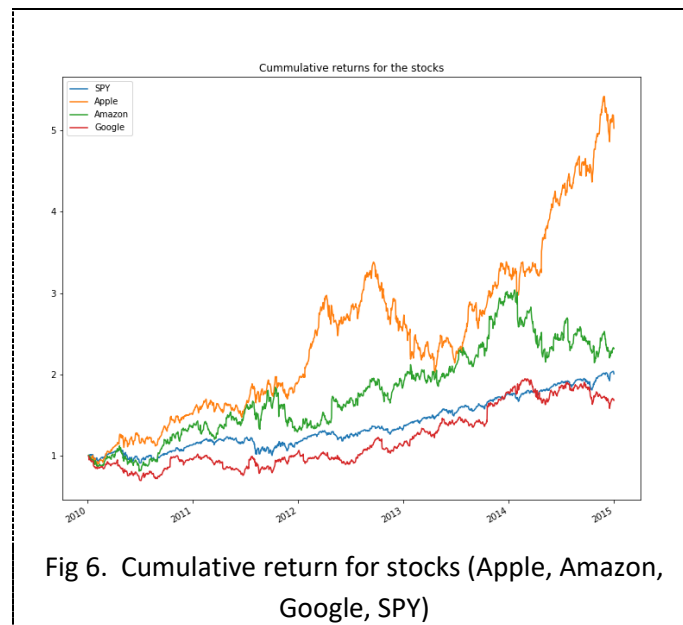Volume of the stocks traded (right).
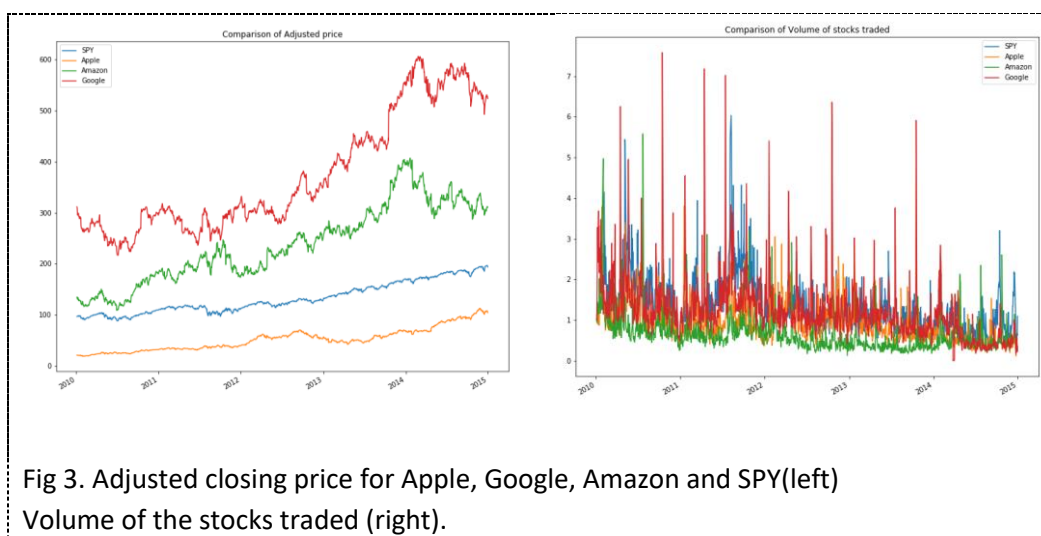
These entities share a common domain of software industry. We can explore this correlation. Correlated entities react in a similar fashion to external events to the industry they belong. For example, local government policies, effect of season and climates etc. From Fig4, we can observe that these is a certain degree of correlation between these stocks.

Fig 4. Correlation matrix for stocks (Apple, Amazon, Google, SPY)
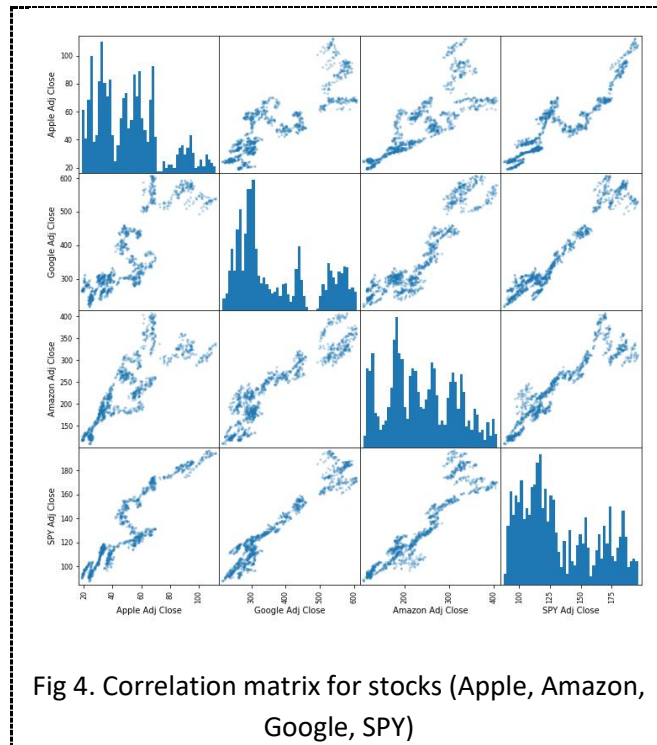
Daily percentage return is very helpful for predicting daily volatility of a given stock. It basically indicates, if we bought a stock today and sell it tomorrow, how much gain or loss it will make us. In Fig5, we can observe daily percentage returns of our stocks. We can observer all the stocks which we are working has similar degree of volatility.

Fig 5. Daily percentage return for stocks (Apple, Amazon, Google, SPY)
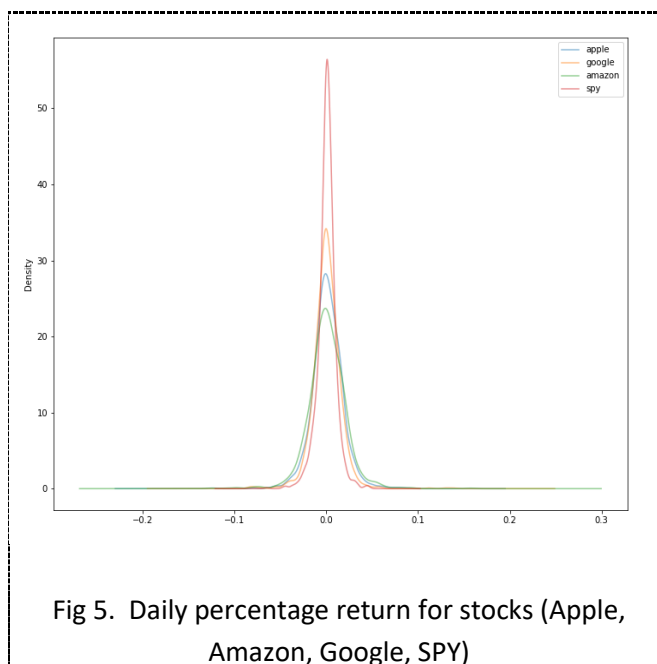
While daily returns are useful, it doesn't give the investors an immediate insight to the gains they have made till date, especially in cases when stocks are very volatile. Cumulative returns are computed relative to day investment was made. If cumulative returns are above 1 we are making profit and vice versa. Let's explore cumulative returns for our stocks. From Fig6, it can be observed that Apple gave a good cumulative return compared to other stocks. In Fig7, we can observer the correlation between returns of different stocks.

Fig 6. Cumulative return for stocks (Apple, Amazon, Google, SPY)

Fig 7. Return correlation for stocks (Apple, Amazon, Google, SPY)

ALGORITHMS AND TECHNIQUE

Lorem ipsum dolor sit amet, consectetuer adipiscing elit, sed diam nonummy nibh euismod tincidunt ut laoreet dolore magna aliquam erat volutpat. Ut wisi enim ad minim veniam, quis nostrud exerci tation ullamcorper suscipit lobortis nisl ut aliquip ex ea commodo consequat.

Nam liber tempor cum soluta nobis eleifend option congue nihil imperdiet doming id quod mazim placerat facer possim assum. Typi non habent claritatem insitam; est usus legentis in iis qui facit eorum claritatem. Investigationes demonstraverunt lectores legere me lius quod ii legunt saepius. Duis autem vel

BENCHMARK

Lorem ipsum dolor sit amet, consectetuer adipiscing elit, sed diam nonummy nibh euismod tincidunt ut laoreet dolore magna aliquam erat volutpat. Ut wisi enim ad minim veniam, quis nostrud exerci tation ullamcorper suscipit lobortis nisl ut aliquip ex ea commodo consequat.

METHODOLOGY

DATA PREPROCESSING

Lorem ipsum dolor sit amet, consectetuer adipiscing elit, sed diam nonummy nibh euismod tincidunt ut laoreet dolore magna aliquam erat volutpat. Duis autem

IMPLEMENTATION

Lorem ipsum dolor sit amet, consectetuer adipiscing elit, sed diam nonummy nibh euismod tincidunt ut laoreet dolore magna aliquam erat volutpat. Duis autem vel eum iriure dolor in hendrerit in vulputate velit esse molestie consequat, vel

REFINEMENT

Lorem ipsum dolor sit amet, consectetuer adipiscing elit, sed diam nonummy nibh euismod tincidunt ut laoreet dolore magna aliquam erat volutpat. Ut wisi enim ad minim veniam, quis nostrud exerci tation ullamcorper suscipit lobortis nisl ut aliquip ex ea commodo consequat.

RESULTS

MODEL EVALUATION AND VALIDATION

Lorem ipsum dolor sit amet, consectetuer adipiscing elit, sed diam nonummy nibh euismod tincidunt ut laoreet dolore magna aliquam erat volutpat. Duis autem

JUSTIFICATION

Lorem ipsum dolor sit amet, consectetuer adipiscing elit, sed diam nonummy nibh euismod tincidunt ut laoreet dolore magna aliquam erat volutpat. Duis autem vel eum iriure dolor in hendrerit in vulputate velit esse molestie consequat, vel

CONCLUSION

FREE FORM VISUALIZATION

Lorem ipsum dolor sit amet, consectetuer adipiscing elit, sed diam nonummy nibh euismod tincidunt ut laoreet dolore magna aliquam erat volutpat. Duis autem

IMPROVEMENTS

Lorem ipsum dolor sit amet, consectetuer adipiscing elit, sed diam nonummy nibh euismod tincidunt ut laoreet dolore magna aliquam erat volutpat. Duis autem vel eum iriure dolor in hendrerit in vulputate velit esse molestie consequat, vel