# Stock Prediction

*Research Paper*

Pothireddy Venumadhavi
21011101086
Artificial Intelligence and Data Science


Shiv Nadar University, Chennai
20 January 2023

## Elaborated the stocks

Modeling and Forecasting of the financial market have been an attractive topic to scholars and researchers from various academic fields. The financial market is an abstract concept where financial commodities such as stocks, bonds, and precious metals transactions happen between buyers and sellers. In the present scenario of the financial market world, especially in the stock market, forecasting the trend or the price of stocks using machine learning techniques and artificial neural networks are the most attractive issue to be investigated. As Giles explained, financial forecasting is an instance of signal processing problem which is difficult because of high noise, small sample size, non-stationary, and non-linearity. The noisy characteristics mean the incomplete information gap between past stock trading price and volume with a future price. The stock market is sensitive with the political and macroeconomic environment. However, these two kinds of information are too complex and unstable to gather. The above information that cannot be included in features are considered as noise. The sample size of financial data is determined by real-world transaction records. On one hand, a larger sample size refers a longer period of transaction records; on the other hand, large sample size increases the uncertainty of financial environment during the 2 sample period. In this project, we use stock data instead of daily data in order to reduce the probability of uncertain noise, and relatively increase the sample size within a certain period of time. By non-stationarity, one means that the distribution of stock data is various during time changing. Non-linearity implies that feature correlation of different individual stocks is various. Efficient Market Hypothesis was developed by Burton G. Malkiel in 1991. In Burton's hypothesis, he indicates that predicting or forecasting the financial market is unrealistic, because price changes in the real world are unpredictable. All the changes in prices of the financial market are based on immediate economic events or news. Investors are profit-oriented, their buying or selling decisions are made according to most recent events regardless past analysis or plans.

The argument about this Efficient Market Hypothesis has never been ended. So far, there is no strong proof that can verify if the efficient market hypothesis is proper or not. However, as Yaser claims, financial markets are predictable to a certain extent. The past experience of many price changes over a certain period of time in the financial market and the undiscounted serial correlations among vital economic events affecting the future financial market are two main pieces of evidence opposing the Efficient Market Hypothesis. In recent years, machine learning methods have been extensively researched for their potentials in forecasting and prediction of the financial market. Multi-layer feed forward neural networks, SVM, reinforcement learning, relevance vector machines, and recurrent neural networks are the hottest topics of many approaches in financial market prediction field. Among all the machine learning methods, neural networks are well studied and have been successfully used for forecasting and modeling financial market. "Unlike traditional machine learning models, the network learns from the examples by constructing an input-output mapping for the problem at hand. Such an approach brings to mind the study of nonparametric statistical inference; the term "nonparametric" is used here to signify the fact that no prior assumptions are made on a statistical model for the input data", according to Simon. As Francis E.H. Tay and Lijuan Cao explained in their studies, Neural networks are more noise tolerant and more flexible compared with traditional statistical models. By noise tolerance, one means neural networks have the ability to be trained by incomplete and overlapped data. Flexibility refers to that neural networks have the capability to learn dynamic systems through a retraining process using new data patterns.

## User technologies

Python was the language of choice for this project. This was an easy decision for the multiple reasons.

1. Python as a language has an enormous community behind it. Any problems that might be encountered can be easily solved with a trip to Stack Overflow. Python is among

the most popular languages on the site which makes it very likely there will be a direct answer to any query.

2. Python has an abundance of powerful tools ready for scientific computing. Packages such as Numpy, Pandas, and SciPy are freely available and well documented. Packages such as these can dramatically reduce, and simplify the code needed to write a given program. This makes iteration quick.

3. Python as a language is forgiving and allows for programs that look like pseudo code. This is useful when pseudocode given in academic papers needs to be implemented and tested. Using Python, this step is usually reasonably trivial.

However, Python is not without its flaws. The language is dynamically typed and packages are notorious for Duck Typing. This can be frustrating when a package method returns something that, for example, looks like an array rather than being an actual array. Coupled with the fact that standard Python documentation does not explicitly state the return type of a method, this can lead to a lot of trials and error testing that would not otherwise happen in a strongly typed language.

- Long short-term memory is a recurrent neural network introduced by Sepp Hochreite and Jurgen Schmidhuber in 1997. LSTM is designed to forecast, predict and classify time series data even long time lags between vital events happened before

- LSTMs have been applied to solve several of problems; among those, handwriting Recognition and speech recognition made LSTM famous.

- . LSTM has copious advantages compared with traditional back-propagation neural networks and normal recurrent neural networks. The constant error back propagation inside memory blocks enables in LSTM ability to overcome long time lags in case of problems similar to those discussed above; LSTM can handle noise, distributed representations, and continuous values; LSTM requires no need for parameter fine-tuning, it works well over a broad range of parameters such as learning rate, input gate bias, and output gate bias.

The objective of our project can be generalized into two main parts. We examine the feasibility of LSTM in stock market forecasting by testing the model with various configurations.