Analyzing the Stock Market Price Using Machine Learning for N Number of Days.

Shree Kantesh, Prajwal Bv and Shivkumar Cm

July 20, 2022

Title

Analyzing the stock market price using machine learning for n number of days.

Keywords

LSTM,fbprophet,time series,graph analysis,stock price,prediction trends.

Abstract

This is a High-quality Indian stock exchange data from almost 20 years. It includes all giants inthe Indian stock market of both BSE and NSE. It is very challenging to forecast the stock market even if large data is available.

This data is scrapped after the large lockdown held in India due to the COVID-19 pandemic. This data is capable to cover the big recessions of the Indian market between 2008 and 2020(COVID-19).

INTRODUCTION

Data analysis is the process of gathering raw data and converting it into information that the users can use to make decisions. It entails inspecting cleansing, transforming, and modeling data to uncover valuable information, draw conclusions, and aid decision-making. In today'sbusiness world, data analysis plays an important role in making scientific decisions and assisting businesses in operating more efficiently.

Stock Market Analysis and Prediction is a project on technical analysis, visualization, and prediction using data provided by Google Finance. By looking at data from the stock market, particularly some giant technology stocks and others. Used pandas to get stock information, visualized different aspects of it, and finally looked at a few ways of analyzing the risk of a stock, based on its previous performance history. Predicted future stock prices through a Monte Carlo method.

LITERATURE SURVEY

The following papers were studied to get an overview of the techniques that were applied earlier to predict the stock market.

LSTM Fully Convolutional Networks for Time Series Classification

-Fazle Karim, Somshubra Majumdar, Houshang Darabi and Shun Chen [1]

With the proposed models, we achieve a potent improvement in the current state-of-the-art time series classification using deep neural networks. Our baseline models, with and without fine-tuning, are trainable end-to-end with nominal preprocessing and can achieve significantly improved performance.

LSTM-FCNs can augment FCN models, appreciably increasing their performance with a nominal increase in the number of parameters. An LSTM-FCN provides one with the ability to visually inspect the decision process of the LSTM RNN and provide a strong baseline on their own. Fine-tuning can be applied as a general procedure to a model to furtherelevate its performance.

Learning Long term Dependencies with Gradient Descent is difficult

-Yoshua bengio, Patrice Simard and Paolo Frasconi [10]

Recurrent networks are very powerful in their ability to represent context, often outperforming static networks. But the factor off gradient descent of an error criterion may be inadequate to train them for a task involving long-term dependencies. It has been found that the system would not be robust to input noise or would not be efficiently trainable by gradient descent when the long-term context is required. The theoretical result presented in this paper holds for any error criterion and not only from mean square error.

Improving N Calculation of the RSI Financial Indicator Using Neural Networks

-Alejandro Rodríguez-González, Fernando Guldris Iglesias, Ricardo Colomo-Palacios GinerAlor-Hernandez, Ruben Posada-Gomez [8]

There has been growing interest in Trading Decision Support Systems in recent years. Despite its volatility, it is not entirely random, instead, it is nonlinear and dynamic or highly complicated and volatile. Stock movement is affected by the mixture of two types of factors: determinant (e.g. gradual strength change between buying side and selling side) and random (e.g. emergent affairs or daily operation variations).

Stock Trend Prediction Using Simple Moving Average

Supportedby News Classification

-Stefan Lauren Dra. Harley S., M.Sc. [5]

The simple moving average is one of many time series analysis techniques. Time series analysis is a method of timely structured data processing to find statistics or important characteristics for many reasons. The simple moving average shows the stock trend by calculating the average value of stock prices for a specific duration. The prices that are used areclosing prices at the end of the day. This technique can avoid noises and therefore smooth thetrend movement.

The main objective of financial news classification is to classify and calculate each news' sentiment value. The positive news is marked by a sentiment value that is greater than 0, while negative news is marked by a less than 0 sentiment value. If there is news having 0 sentiment value, it will be omitted as its neutralism does not affect the stock trend.

VISUALIZING AND UNDERSTANDING RECURRENT NETWORKS

-Andrej Karpathy, Justin Johnson, Li Fei-Fei [4]

Character-level language models have been used as an interpretable test bed for analyzing thepredictions, representations training dynamics, and error types present in Recurrent Neural Networks. In particular, the qualitative visualization experiments, cell activation statistics, and comparisons to finite-horizon n-gram models demonstrate that these networks learn powerfully, and often interpretable long-range interactions on real-world data.

that further architectural innovation may be needed to address the remaining errors

Motivation

As we are in the 21st century relying on human efforts to indicate which stock skyrocket or which stock become the prey for bulls because money causes emotion and in emotion humans tend to take a wrong decision therefore by making use of machine learning i.e ability of a machine to learn from data and machine learning to make the task easier for prediction as it’s booming in today's world it has become trend cuz data analysts and data scientist needs to process the sequence of raw data to give meaningful information or prediction.

The machines had made the human life easier by replacing the humans in so many fields and increasing the efficiency of work and also lessen the cost of production, the purpose of using machine learning in stock price prediction is the ability of the machine to fetch the real-time data and trained algorithm which give accurate information required for the investor with 0 emotions.

Financial markets provide a unique platform for trading and investing, where trades can be executed from any device that can connect to the Internet. With the advent of stock markets, people have the opportunity to have multiple avenues to make their investments grow. Not only that, but it also gave rise to different types of funds like mutual funds, hedge funds, and index funds for people and institutions to invest money according to their risk appetite.

Problem Domain

The stock market appears in the news every day. You hear about it every time it reaches a new high or a new low. The rate of investment and business opportunities in the Stock market can increase if an efficient algorithm could be devised to predict the short-term price of an individual stock.

Previous methods of stock predictions involve the use of Artificial Neural Networks and Convolution Neural Networks which have an error loss of an average of 20%.

In this paper, we will see if there is a possibility of devising a model using a Recurrent Neural Network which will predict stock price with a less percentage of error. And if the answer turns out to be YES, we will also see how reliable and efficient will this model be.

Problem Formulation

The most interesting task is to predict the market. So many methods are used for completingthis task. Methods vary from very informal ways to many formal ways a lot. This tech. are categorized as:

- Prediction Methods
- Traditional Time Series
- Technical Analysis Methods

- Machine Learning Methods
- Fundamental Analysis Methods
- Deep Learning

The criteria for this category are the kind of tool and the kind of data that these methods areconsuming to predict the market. What is mutual to the technique is that they are predicting and hence helping the market's future behavior

Solution Methodologies or Problem Solving

CHOOSING THE DATASET

For this project, we chose Google stocks. Google stock is a large index traded on the New York stock exchange. All companies in the index are large publicly traded companies, leaders in each of their sectors. The index covers a diverse set of sectors featuring companies such as Microsoft, Visa, Boeing, and Walt Disney. It is important to use a predefined set of companies rather than a custom-selected set so that we do leave ourselves open to methodology errors or accusations of fishing expeditions. If we had selected a customset of companies, it could be argued that the set was tailored specifically to improve our results. Since the project aims to create a model of stock markets in general. Google was chosen because it is well known. The components provided a good balance between available data and computational feasibility.

GATHERING THE DATASETS

A primary dataset will be used throughout the project. The dataset will contain the daily percentage change in stock price. Luckily, daily stock price data is easy to come by. Googleand Yahoo both operate websites that offer a facility to download CSV files containing a full 14 daily price history. These are useful for looking at individual companies but cumbersome when accessing large amounts of data across many stocks. For this reason, Quandl was used to gather the data instead of using Google and Yahoo directly. Quandl is a free-to-use website that hosts and maintains vast amounts of numerical datasets with a focusspecifically on economic datasets, including stock market data which is backed by Google and Yahoo. Quandl also provides a small python library that is useful for accessing the database programmatically. The library provides a simple method for calculating the daily percentage change daily in prices.

. Quandl also provides a small python library that is useful for accessing the database programmatically. The library provides a simple method for calculating the daily percentagechange daily in prices.

Technical Analysis Methods

Technical analysis is used to attempt to forecast the price movement of virtually any tradable instrument that is generally subject to forces of supply and demand, including stocks, bonds, futures, and currency pairs. Technical analysis can be viewed as simply the study of supply and demand forces as reflected in the market price movements of a security. It is most commonly applied to price changes, but some analysts may additionally track numbers other than just prices, such as trading volume or open interest figures.

Over the years, numerous technical indicators have been developed by analysts in an attempt to accurately forecast future price movements. Some indicators are focused primarily on identifying the current market trend, including support and resistance areas, while others are focused on determining the strength of a trend and the likelihood of its continuation.

Commonly used technical indicators include trendlines, moving averages, and momentum indicators such as the moving average convergence divergence (MACD) indicator.

Data Model

Data collection is the process of gathering and measuring information from countless different sources. To use the data, we collect to develop practical machine learning solutions.

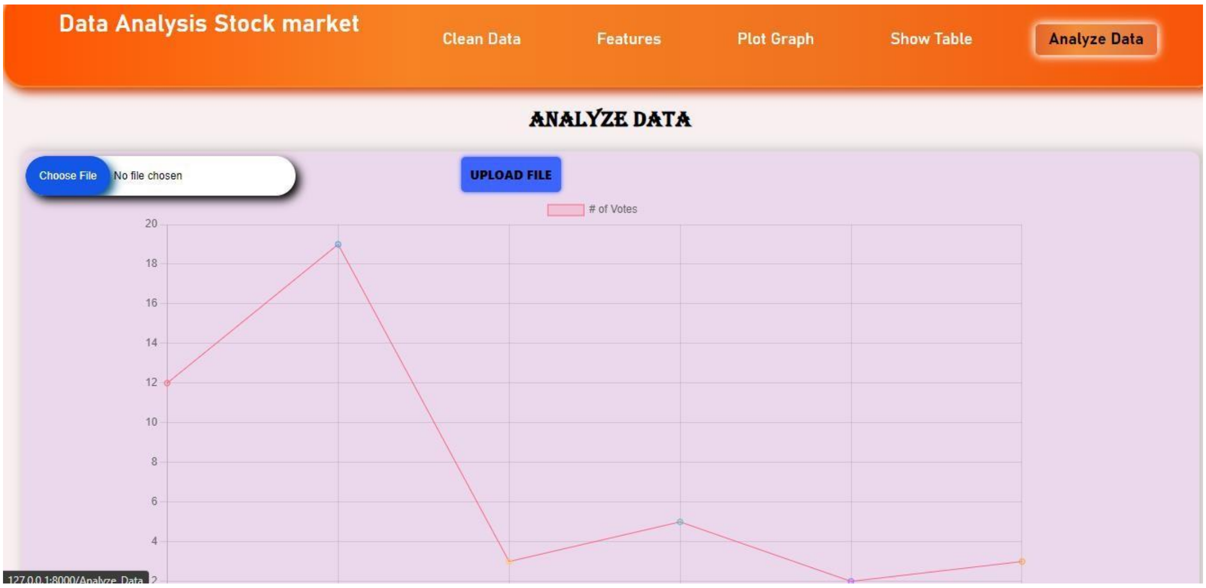
Collecting data allows you to capture a record of past events so that we can use data analysisto find recurring patterns. From those patterns, you build predictive models using machine learning algorithms that look for trends and predict future changes.

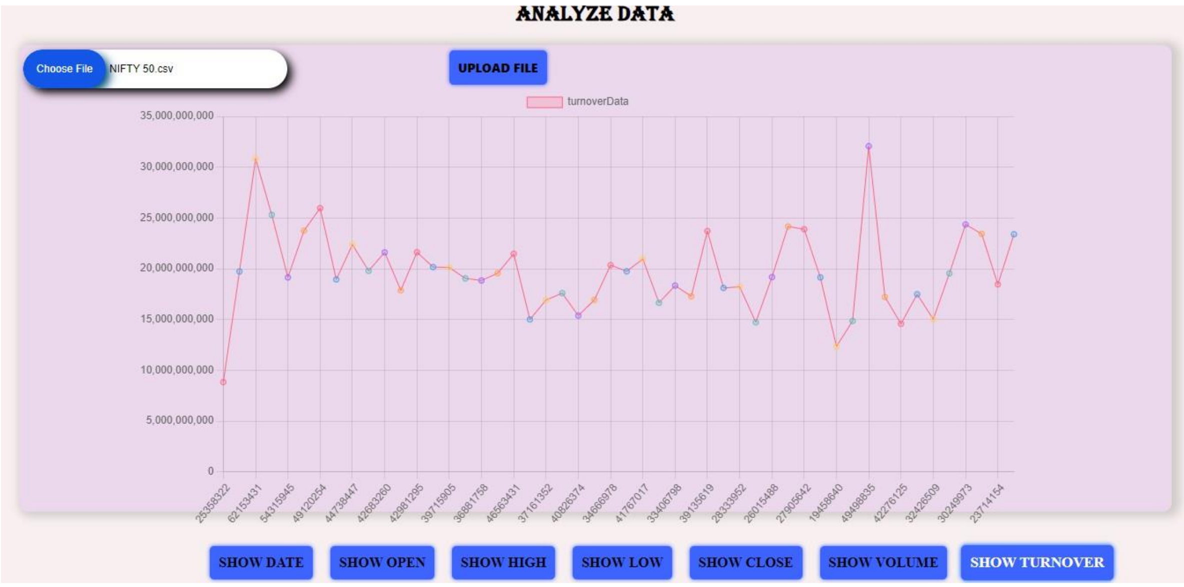
The Indian Stock Analysis official website is the principal basis of data for this project. The data was web scrapped from the website and kept in the appropriate format using a python library called beautiful soup. The dataset has the columns

	Date	Open	High	Low	Close	Volume	Turnover	P/E	P/B	Div Yield
0	2000-01-03	1482.15	1592.90	1482.15	1592.20	25358322	8.841500e+09	25.91	4.63	0.95
1	2000-01-04	1594.40	1641.95	1594.40	1638.70	38787872	1.973690e+10	26.67	4.76	0.92
2	2000-01-05	1634.55	1635.50	1555.05	1595.80	62153431	3.084790e+10	25.97	4.64	0.95
3	2000-01-06	1595.80	1639.00	1595.80	1617.60	51272875	2.531180e+10	26.32	4.70	0.94
4	2000-01-07	1616.60	1628.25	1597.20	1613.30	54315945	1.914630e+10	26.25	4.69	0.94
5	2000-01-10	1615.65	1662.10	1614.95	1632.95	45013949	2.375350e+10	26.57	4.74	0.93
6	2000-01-11	1633.25	1639.90	1548.25	1572.50	49120254	2.596950e+10	25.59	4.57	0.96
7	2000-01-12	1572.30	1631.55	1571.70	1624.80	38364961	1.895000e+10	26.44	4.72	0.93
8	2000-01-13	1627.85	1671.15	1613.65	1621.40	44738447	2.237610e+10	26.38	4.71	0.93
9	2000-01-14	1622.15	1627.40	1591.40	1622.75	43292009	1.979980e+10	26.41	4.71	0.93

regarding the company name,number of stocks sold on particular day,opening and closing price,upper circuit and lowercircuit.

Comparison of Results:





Dropout is a valuable feature to assist in improving this, and optimizing the selection of dropouts still couldn't guarantee good validation results. Despite the metrics of sensitivity, specificity, and precision indicating good performance The

LSTM outperformed the RNN marginally, but not significantly, However, the LSTM takes considerably longer to train, and performance benefits gained from the parallelization of machine learning algorithms on a GPU are evident with a 70.7% performance improvement for training the LSTM model, the ability to predict using streaming data should improve the model.

Justification of the Results

- There are some null values in the dataset in the columns such as winner, city, venue, etc.
- the presence of these null values, the classification cannot be done accurately.
- To replace the null values in different columns with dummy values.
- This step is the main part where we can eliminate some columns of the dataset that are not useful for the estimation of the match-winning team.
- This is estimated using feature importance.
- The considered attributes have the following feature importance.

The data which has been collected is used for visualizing fora better understanding of the information.

70.7% performance improvement for training the LSTM model, the ability to predict using streaming data should improve the model.

Conclusion

Financial markets provide a unique platform for trading and investing, where trades can be executed from any device that can connect to the Internet. With the advent of stock markets, people have the opportunity to have multiple avenues to make their investments grow. Not only that, but it also gave rise to different types of funds like mutual funds, hedge funds, andindex funds for people and institutions to invest money according to their risk appetite.

Governments of most countries invest a part of their healthcare, employment, or retirement funds into stock markets to achieve better returns for everyone. Online trading services have already revolutionized the way people buy and sell stocks. The financial markets have evolved rapidly into a strong and interconnected global marketplace. These advancements bring forth new opportunities and the data science techniques offer many advantages, but they also pose a whole set of new challenges. In this paper, we propose a taxonomy of computational approaches to stock market analysis and prediction, present a detailed literature study of the state-of-the-art algorithms and methods that are commonly applied to stock market prediction, and discuss some of the continuing challenges in this area that require more attention and provide opportunities for future development and research. Unlike traditional systems, the stock market today is built using a combination of different technologies, such as machine learning, expert systems, and big data which communicate with one another to facilitate more informed decisions. At the same time, global user connectivity on the internet has rendered the stock market susceptible to customer sentiments, less stable due to developing news, and prone to malicious attacks. This is where further research can play an important role in paving the way for how stock markets will be analyzed and made more robust in the future. A promising research direction is to explore various algorithms to evaluate whether they are powerful enough to predict for the longer term because markets act like weighing machines in the long run having less noise and more predictability. Hybrid approaches that combine statistical and machine learning techniques will probably prove to be more useful for stock prediction.

References:

- Abu-Mostafa, Yaser S., and Amir F. Atiya. 1996. Introduction to financial forecasting. Applied Intelligence 6: 205–13.
- Arévalo, Rubén, Jorge García, Francisco Guijarro, and Alfred Peris. 2017. A dynamic trading rule based on filtered flag pattern recognition for stock market price forecasting. Expert Systems with Applications 81: 177–92.
- Babu, M. Suresh, N. Geethanjali, and B. Satyanarayana. 2012. Clustering Approach to Stock Market Prediction. International Journal of Advanced Networking and Applications 3: 1281
- Ballings, Michel, Dirk Van den Poel, Nathalie Hespeels, and Ruben Gryp. 2015.

Evaluating multiple classifiers for stock price direction prediction. Expert Systems with Applications 42: 7046–56. [CrossRef]

- Bao, Wei, Jun Yue, and Yulei Rao. 2017. A deep learning framework for financial time series using stacked autoencoders and long-short term memory. PLoS ONE 12: e0180944. [CrossRef] [PubMed]
- Cervelló-Royo, Roberto, Francisco Guijarro, and Karolina Michniuk. 2015. Stock market trading rule based on pattern recognition and technical analysis: Forecasting the DJIA index with intraday data. Expert Systems with Applications 42: 5963–75. [CrossRef]
- Creighton, Jonathan, and Farhana H. Zulkernine. 2017. Towards Building a Hybrid Model for Predicting Stock Indexes. Paper presented at the 2017 IEEE International Conference on Big Data (Big Data), Boston, MA, USA, December 11–14.
- Fama, Eugene F. 1970. Efficient Capital Markets: A Review of Theory and Empirical Work. The Journal of Finance 25: 383–417. [CrossRef]