Case Study Analysis of Stock Market Using Algorithms

Prince Nathan S[1], Onkar Saudagar[2], Rutika Shinde[3]

[1]PG Student, Department of Data Science and Analytics,
National Institute of Electronics and Information Technology, Chennai, India

[2]Student, Department of Information Technology,
Pune Institute of Computer Technology, Pune, India

[3]Student, Department of Information Technology,
Sinhgad Institute of Technology and Science, Pune, India

---***--

ABSTRACT:
Data Science consists of data cleaning, arranging it in the proper manner, and then analyzing it to get the desired output. It is an interdisciplinary field of study that uses data for various research and reporting purposes to derive insights and meaning out of that data. In today's economy, it is very important to predict the stock prices and their ups and downs. As we are seeing a lot of people are getting closure towards the stock or the bitcoin-like crypto's.

So there are techniques so that one can predict the appropriate result. There are many algorithms that are available to trace the movement of the market. Big data, Data mining, Data visualization all helps to analyze the data and predict the outcome.

Keywords: Data Mining, Algorithms, NN, ML, Time Series, ARIMA

I. INTRODUCTION:
Data Science is a technology that uses algorithms, processes to get knowledge from raw data. It is made up of machine learning, data mining, neural network.

It consists of math, statistics, analysis, and relatable information Stock market is an aggregation of investors. It claims ownership of stocks.

The Stock Market Analysis using data science is the main motive behind this seminar. Here we will get information about Neural Network, Time Series Data, Data mining,

Outliers, Various Algorithms, Pseudocodes used will help to understand the model.

II. LITERATURE SURVEY:
Similar work has been done in the field of the Stock Market in using Data Science.

A different research paper uses different methods as discussed below:-

STOCK MARKET TREND PREDICTION USING NEURAL NETWORKS
The paper will go into detail about data prediction methods and approaches used throughout the world to learn the variety of data in the world. Here knowledge is gained from sample datasets. Neural network models the behavior of known systems without being given any rules or models.

STOCK MARKET PREDICTION WITH HISTORICAL TIME SERIES DATA
Time series is an observation of data at equal intervals. It includes data extraction and preprocessing, Prediction with LSTM(long short term memory) on time series data. The characteristics of time series are stationarity, trend, and seasonality.

STOCK MARKET PRICE FORECASTING BY NEWS MINING
In mining, proper feature selection and extraction of data is very important. Here word segmentation is used and also many metrics for feature selection.

UNSUPERVISED OUTLIER DETECTION IN TIME SERIES DATA
Outliers are such observations in a data set that seems inconsistent with the remaining of the data set. They deviate too much from actual values in the data set,

behave as outsiders in the group. Identification of outliers may give a lot of useful knowledge.

III. KEY TOPICS

STOCK MARKET

A stock market is a gathering of buyers and sellers of stocks (also known as shares), which reflect ownership claims on firms. Stocks can be exchanged publicly or privately, such as shares of private companies offered to investors through equity crowd funding platforms. Stockbrokers and computerized trading platforms are the most common ways to invest in the stock market

In stock markets, trade refers to the transfer of a stock or security from a seller to a buyer in exchange for money. This necessitates a price agreed between the two parties. Equities (stocks or shares) represent a share of a company's ownership.

Small individual stock investors to huge institutional investors, which can be headquartered anywhere in the world and include banks, insurance firms, pension funds, and hedge funds, all participate in the stock market. A stock exchange trader may execute their buy or sell orders on their behalf.

The goal of a stock exchange is to provide a marketplace for the exchange of securities between buyers and sellers. The exchanges facilitate price discovery by providing real-time trading information on the listed assets.

Along with debt markets, which are often more onerous but do not trade publicly, the stock market is one of the most important avenues for firms to raise money. Businesses can become publicly listed and raise more financial capital for expansion by selling shares of ownership in a public market. The liquidity that an exchange provides to investors allows security holders to sell securities swiftly and easily. Investing in stocks has this advantage over other less liquid investments such as real estate and other immovable assets.

For example, rising stock prices are linked to increasing corporate investment and vice versa. As a result, central banks prefer to keep a watch on the stock market's regulation and behavior, as well as the smooth operation of financial system ac-activities in general. Central banks exist to ensure financial stability.

DATA SCIENCE

Data science is an interdisciplinary field that employs scientific methods, pro- cedures, algorithms, and systems to extract knowledge and insights from noisy, structured, and unstructured data, as well as to apply that knowledge and actionable insights across a wide range of application areas. Data mining, machine learning, and big data are all connected to data science.

Data science is a "concept that unifies statistics, data analysis, informatics, and related approaches" in order to use data to understand and analyze actual occurrences." Within the framework of mathematics, statistics, computer science, information science, and domain knowledge, it employs techniques and theories from a variety of domains. Data science, on the other hand, is distinct from computer science and information science. Jim Gray, the recipient of the Turing

The award, envisioned data science as the "fourth paradigm" of science (empirical, the- theoretical, computational, and now data-driven), claiming that everything about research is changing" due to the impact of information technology and the data flood.

MARKET PREDICTION

The act of attempting to anticipate the future value of business stock or other financial instruments traded on an exchange is known as a stock market

prediction. A successful forecast of a stock's future price could result in a large profit. Stock prices, according to the efficient-market theory, represent all currently accessible information, and any price fluctuations that are not based on newly revealed information are thus fundamentally unpredictable. Others disagree, and individuals who hold this position claim to have a variety of ways and techniques that allow them to obtain future pricing information.

Stock market forecasting has moved into the technology arena since the invention of the digital computer. Artificial neural networks (ANNs) and Genetic Algorithms are the most often used techniques (GA).ANNs can be conceived of as approximators for mathematical functions. The feed-forward network, which uses the backward propagation of errors algorithm to update the network weights,is the most frequent type of ANN used for stock market prediction. The time requirement is another type of ANN that is better for stock prediction.

TIME SERIES

A time series is a set of data points that are indexed (or listed or graphed) in time order in mathematics. A time series is a collection of images taken at evenly spaced intervals. As a result, it's a series of discrete-time data.

Methods for studying time series data in order to extract useful statistics and other aspects of the data are referred to as time series analysis. The employment of a model to predict future values based on previously observed values is known as time series forecasting. This form of analysis isn't commonly referred to as "time series analysis," which refers to relationships between points in time within a single series. Interrupted time series analysis is used to find changes in a time series' evolution from before to after some action that may alter the underlying variable.

ARIMA

An autoregressive integrated moving average (ARIMA) model is a generalization of an autoregressive moving average (ARMA) model in statistics and econometrics, particularly in time series analysis. These models are used to fit time series data in order to better understand or anticipate future points in the series (forecasting). When data shows signs of non-stationarity in the sense of mean (but not variance/auto covariance), an initial differencing step (equivalent to the "integrated" element of the model) can be applied one or more times to eliminate the non-stationarity of the mean function, ARIMA models are used. We are motivated to make stationary a non-stationary time series, e.g., by using differencing, before we can use the ARMA model, because it is theoretically sufficient to describe a regular wide-sense stationary time series, according to World's decomposition theorem.

The AR in ARIMA refers to the fact that the evolving variable of interest is regressed on its own lagged values. The MA section of the equation denotes that the regression error is a linear combination of error terms whose values happened simultaneously and at different times in the past. The letter I (for "integrated") denotes that the data values have been replaced with the difference between their current values and their former values (and this differencing process may have been performed more than once). Each of these features is designed to help the model fit the data as closely as feasible.
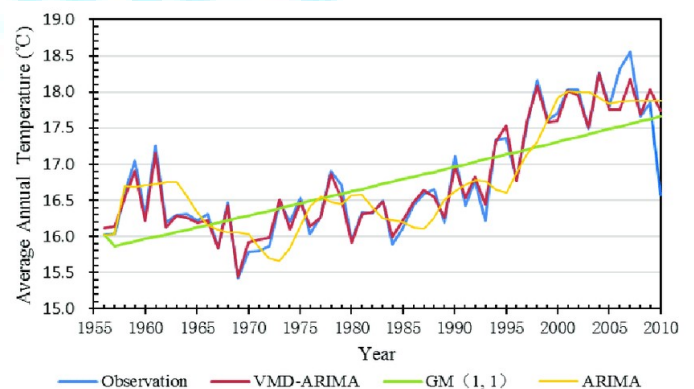
IV. APPLICATIONS

USE OF ALGORITHM

Data science makes considerable use of algorithms. An algorithm is a set of rules that must be followed in order to complete a task. You've probably heard about how algorithms are used to purchase and sell stocks. Algorithmic trading is when software creates

rules for things like when to purchase and sell stocks.

An algorithm may be programmed to buy a stock if it lowers 8% in value during the time of the day, or to sell it if it loses 10% of its value since it was initially purchased. Algorithms are created to work without the need for human interaction. Bots are a term that has been used to describe them.

TRAINING AND TESTING OF DATASET

Training is the process of using data to teach a machine how to respond in machine learning and data science. We can develop a model for learning. This machine learning methodology allows a computer to make accurate predictions based on the data it has gathered in the past. If you want to teach a machine to forecast the future of stock prices, you'll need a model of the previous year's stock prices to use as a starting point.

The data from January to October would be the training set. Then we'll do some testing in November and December. By analyzing how the stocks performed from January to October, our machine should have learned. We'll now ask it to forecast what should have occurred in November and December of that year. The machine's forecasts will be compared against real-world pricing. We're trying to reduce the amount of variation between what the model predicts and the actual data as we tweak our training model.

USE OF MODELS

A time series is a collection of data, in this case, the value of a stock, that has been indexed across time. This time span could be broken into hourly, daily, monthly, or even minute increments. To collect price data, a time series model is developed using machine learning and deep learning models. The data must be examined before being fitted to the model. This is what allows for the prediction of future stock prices over a specified period.

When it comes to the stock market or stocks in general, a machine learning model can be given financial data such as the P/E ratio, total debt, volume, and other factors to assess whether is stock is a good investment. A model can evaluate if it is

time to sell, hold, or purchase a stock based on the financials we provide. A model can be able to forecast something so complicated that it misses the link between the feature and the target variable. Over fitting is the term for this. Under fitting occurs when a model's fit to the data is insufficient, resulting in overly simplistic predictions.

Over fitting is a problem when a model has trouble identifying stock market trends and so is unable to adjust to future events. When a model predicts the basic average price based on the stock's full history, it is said to be under fitting.

V. CONCLUSION

Data science is the application of a variety of tools, algorithms, and machine learning techniques to uncover hidden patterns in raw data. We can learn a lot about a company's financial health by using data science and forecasting methods. Forecasting is crucial because future event prediction is a critical input into many sorts of planning and decision-making processes, including finance, industrial process control, and risk management.

Time series analysis is useful in financial forecasting and analysis, and it may be used in any field. Time series analysis is used in finance to forecast financial variables such as stock prices, assets, and commodities. Stocks are the most volatile investment, with a high risk of loss but a large return if properly managed in a portfolio. Capital markets are exchanges for buying and selling equity and debt securities. They also include operations such as stock offering and trading, as well as the issue of public company shares.

REFERENCES

1. Jung-Hua Wang and Jia-Yann Leu, "Stock Market Trend Prediction Using ARIMA-based Neural Networks", Proceedings of International Conference on Neural Networks (ICNN'96).

2. M. Kesavan, J. Karthiraman, Rajadurai T. Ebenezer, S. Adhithyan, "Stock Market Prediction with Historical Time Series Data and Sentimental Analysis of Social Media Data", 2020 4th International Conference on Intelligent

Computing and Control Systems (ICICCS).

3. Xiangyu Tang, Chunyu Yang, Jie Zhou, "Stock Price Forecasting by Combining News Mining and Time Series Analysis", IEEE/WIC/ACM International Joint Conference on Web Intelligence and Intelligent Agent Technology.

4. R. M. Kapila Tharanga Rathnayaka, D.M.K.N Seneviratna, Wei Jianguo, Hasitha Indika Arumawadu, "A hybrid statistical approach for stock market forecasting based on Artificial Neural Network and ARIMA time series models", International Conference on Behavioral, Economic and Socio-cultural Computing (BESC).