

STOCK MARKET PREDICTION USING MACHINE LEARNING

M.Sivaganesh¹, R.Shankar², M.Tony David Raj³, S.Udaya kiran⁴

¹Assistant Professor, ²BE-IVth Year, ³BE-IVth Year, ⁴BE-IVth Year

Department of Computer Science and Engineering, Paavai Engineering College, Pachal, Namakkal.

Abstract — The stock market price prediction and classification are Stock market price prediction is a common problem in finance that involves using historical data to forecast future prices of stocks or other financial instruments. The goal of stock market price prediction is to identify profitable trading opportunities and make informed investment decisions, to resolve these issues different machine learning algorithm is implemented for predict the model accuracy of stock market price level. The problem statement for stock market price prediction involves identifying the factors that influence stock prices and developing a model to predict future prices based on historical data. Some of the factors that may influence stock prices include economic indicators, company financials, news and events, and market sentiment. LSTM (Long Short-Term Memory) is a type of recurrent neural network that is well-suited for time series prediction tasks, including stock market price prediction. Here's a step-by-step guide on how to build an LSTM model for stock market price prediction LSTM models can provide useful insights and predictions, they are not foolproof, and the stock market can be unpredictable and subject to sudden changes. Therefore, it is important to use a combination of different models and techniques, continually update the models with new data, and apply rigorous testing and validation methods to ensure the accuracy and reliability of the predictions.

1. INTRODUCTION

A stock market is a public market where you can buy and sell shares for publicly listed companies. The stocks, also known as equities, represent ownership in the company. The stock exchange is the mediator that allows the buying and selling of shares. Stock Price Analysis using machine learning helps you to discover the real time value of company stock and other financial assets traded on an exchange. The development in stock market prediction has gained high significance among expert analysts and investors. Analyzing stock market movements and price actions are extremely difficult due to the noisy environment in the market. The complication of stock prices changes many factors that include announcements of quarterly earnings and market news. The stock market indices are calculated based on their market capitalization. Accurate forecasting of the stock market is therefore a very difficult task by changing the market world. The researchers and market analysts have been keen on developing and testing of stock market behavior.

2 RELATED WORKS

2.1 STOCK MARKET FORECASTING

In the stock market, the investor shows interest in profit by investing some money in the stock market. The stock market has shown investor interest due to advanced applications where 12 prediction may lead to prosperous market forecasting. Predicting movements of the stock market precisely depends on advance information. The tools which are used for stock market forecasting can track and control the market which can be used to make the right decisions. The stock market needs to handle several pieces of information on industrial stocks which cover the entire financial market. These are adjusted according to the business status investors who consider sales and purchase. Several factors affecting the market position are the future estimation income, a news release on earnings and changes in management, etc. Therefore, accurate prediction of the stock market helps investors in making better decisions. Through ML techniques the investor can earn more money with high risk. The process of the stock market.

2.2 MACHINE LEARNING

Predictive analytics tools are powered by several different models and algorithms that can be applied to wide range of use cases. Determining what predictive modeling techniques are best for your company is key to getting the most out of a predictive analytics solution and leveraging data to make insightful decisions in the statistical context, Machine Learning is defined as an application of artificial intelligence where available information is used through algorithms to process or assist the processing of statistical data. While Machine Learning involves concepts of automation, it requires human guidance. Machine Learning involves a high level of generalization in order to get a system that performs well on yet unseen data Year 13 instances. Machine learning is a relatively new discipline within Computer Science that provides a collection of data analysis techniques. Some of these techniques are based on well-established statistical methods (e.g. - logistic regression and principal component analysis) while many others are not. Most statistical techniques follow the paradigm of determining a particular probabilistic model that best describes observed data among a class of related models. Similarly, most machine learning techniques are designed to find models that best fit data (i.e. they solve certain optimization problems), except that these machine learning models are no longer restricted to probabilistic ones. Therefore, an advantage of machine learning techniques over statistical ones is that the latter require underlying probabilistic models while the former do not. Even though some machine learning techniques use probabilistic models, the classical statistical techniques are most often too stringent for the oncoming Big Data era, because data sources are increasingly complex and multi-faceted. Prescribing probabilistic models relating variables from disparate data sources that are plausible and amenable to statistical analysis might be extremely difficult if not impossible. Machine learning might be able to provide a broader class of more flexible alternative analysis methods better suited to modern sources of data. It is imperative for statistical agencies to explore the possible use of machine learning techniques to determine whether their future needs might be better met with such techniques than with traditional ones.

2.3 WEB SCRAPING

Web scraping is an automatic method to obtain large amounts of data from websites. Most of this data is unstructured data in an HTML format which is then converted into structured data in a spreadsheet or a database so that it can be used in various applications. There are many different ways to perform web scraping to obtain data from websites. These include using online services, particular API's or even creating your code for web scraping from scratch. Many large websites, like Google, Twitter, Facebook, Stack Overflow, etc. have API's that allow you to access their data in a structured format. This is the best option, but there are other sites that don't allow users to access large amounts of data in a structured form or they are simply not that technologically advanced. In that situation, it's best to use Web Scraping to scrape the website for data. So, when a web scraper needs to scrape a site, first the URLs are provided. Then it loads all the HTML code for those sites and a more advanced scraper might even extract all the CSS and JavaScript elements as well. Then the scraper obtains the required data from this HTML code and outputs this data in the format specified by the user. Mostly, this is in the form of an Excel spreadsheet or a CSV file, but the data can also be saved in other formats, such as a JSON file.

3. EXISTING SYSTEM

Time series forecasting consists of a research area designed to solve various problems, mainly in the financial area. Support Vector Machine-59% Linear Regression-61.2% Naive Bayes-35%. We implemented a Random Forest approach to predict stock market prices. Random Forests are very effectively implemented in forecasting stock prices, returns, and stock modeling. We outline the design of the Random Forest with its salient features and customizable parameters. We focus on a certain group of parameters with a relatively significant impact on the share price of a company. With the help of sentiment analysis, we found the polarity score of the new article and that helped in forecasting accurate result. Although share market can never be predicted with hundred per-cent accuracy due to its vague domain, this paper aims at proving the efficiency of Random Forest for forecasting the stock prices. We considered historical data about the stock prices of a publicly listed company to implement machine learning algorithms in predicting the future stock price of a company, starting with simple algorithms like averaging and linear regression.

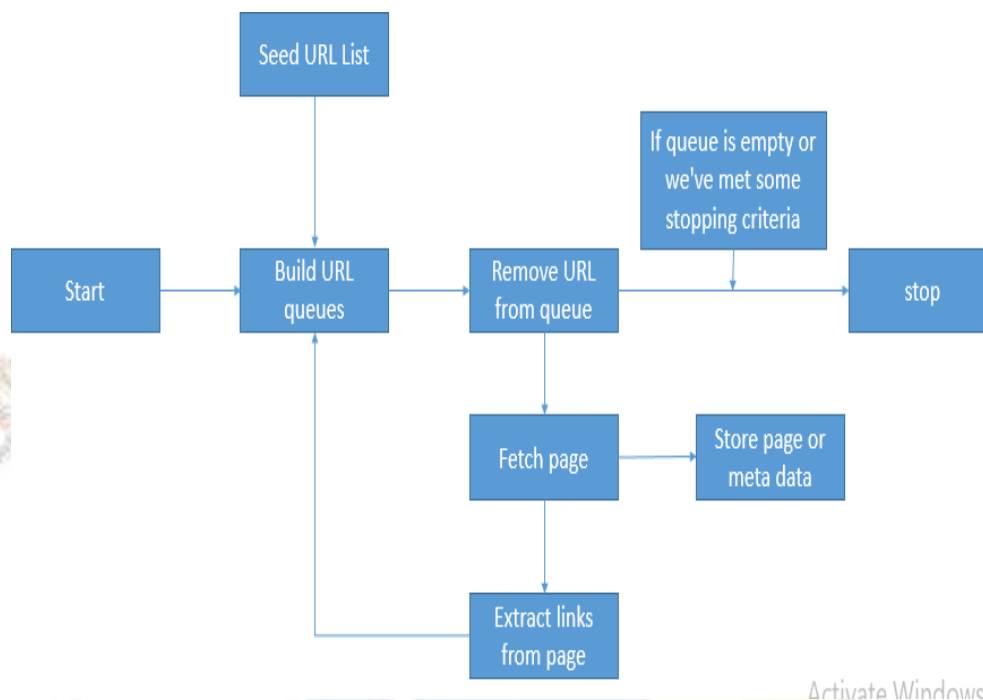
4. PROPOSED SYSTEM

Web scraping is the process of gathering information from the Internet. Even copying and pasting the lyrics of your favorite song is a form of web scraping. However, the words "web scraping" usually refer to a process that involves automation. Some websites don't like it when automatic scrapers gather their data, while others don't mind. The selected studies use the hybrid approach for stock market forecasting. Selected study S3 proposed a hybrid method that is a Web scraping compound with a rough approach and S8 proposed a

hybrid method that combines Web scraping with GA to improve GA performance in stock market forecasts. Another study S13 combined the statistical technique discrete wavelet transform with ML 25 algorithm web scraping for stock market prediction.

5. SYSTEM ARCHITECTURE

A system architecture is the conceptual model that defines the structure, behavior, and more views of a system. An architecture description is a formal description and representation of a system, organized in a way that supports reasoning about the structures and behaviors of the system.



5. Algorithm

LONG SHORT-TERM MEMORY

LSTM stands for Long Short-Term Memory, which is a type of recurrent neural network (RNN) architecture that is used for processing sequential data, such as time-series data. LSTM networks were introduced to address the vanishing gradient problem, which is a common issue in traditional RNN where the gradients become too small to update the weights in the network during back propagation. LSTM networks overcome this problem by using a special gating mechanism that allows them to selectively remember and forget information over time. The key components of an LSTM network are the input gate, forget gate, output gate, and cell state. The input gate controls which information is stored in the cell state, the forget gate controls which information is discarded from the cell state, and the output gate controls which information is outputted from the cell state. LSTM networks have been used in a variety of applications, such as speech recognition, image captioning, and stock price prediction. They are particularly useful for tasks that involve long-term dependencies, where information from the past can affect the prediction at a later time. LSTM networks are commonly used in stock market analysis and prediction due to their ability to model time series data with long-term dependencies. The architecture of an LSTM network is particularly well-suited to capturing the complex relationships between stock prices over time, which can be influenced by a variety of factors such as news events, economic indicators, and market trends. To use LSTM for stock market analysis, the network is typically trained on historical price and volume data, as well as other relevant variables such as technical indicators and economic data. The goal of the training is to learn a model that can accurately predict future stock prices based on the historical data. Once the model is trained, it can be used to make predictions on new data, such as daily or weekly stock prices. The predicted prices can then be compared to the actual prices to evaluate the accuracy of the model. This can be useful for identifying potential trading opportunities or for developing a long-term investment strategy. That stock market analysis is a complex and unpredictable field, and there is no guarantee that the LSTM model will always make accurate predictions. Additionally, the accuracy of the model can be affected by factors such as changes in market conditions, unexpected news events, or other unforeseen factors. Therefore, it's important to use caution and to always consider multiple sources of information when making investment decisions

5.1 TESTINGThe purpose of testing is to discover errors. Testing is the process of trying to discover every conceivable fault or weakness in a work product. It provides a way to check the functionality of components, sub-assemblies, assemblies and/or a finished product it is the process of exercising software with the intent of ensuring that the Software system meets its requirements and user expectations and does not fail in an unacceptable manner. There are various types of test. Each test type addresses a specific testing requirement.

5.1.1 Unit Testing

Unit testing involves the design of test cases that validate that the internal program logic is functioning properly, and that program inputs produce valid outputs. All decision branches and internal code flow should be validated. It is the testing of individual software units of the application .it is done after the completion of an individual unit before integration. This is a structural testing, that relies on knowledge of its construction and is invasive. Unit tests perform basic tests at component level and test a specific business process, application, and/or system configuration. Unit tests ensure that each unique path of a business process performs accurately to the documented specifications and contains clearly defined inputs and expected results.

5.1.2 Functional Testing

Functional tests provide systematic demonstrations that functions tested are available as specified by the business and technical requirements, system documentation, and user manuals. Functional testing is centered on the following items:

1. **Valid Input** : Identified classes of valid input must be accepted
2. **Invalid Input** : identified classes of invalid input must be rejected.
3. **Functions** : identified functions must be exercised.
4. **Output** : identified classes of application outputs must be exercised.
5. **Systems/Procedures** : interfacing systems or procedures must be invoked.

Organization and preparation of functional tests is focused on requirements, key functions, or special test cases. In addition, systematic coverage pertaining to identify Business process flows; data fields, predefined processes, and successive processes must be considered for testing. Before functional testing is complete, additional tests are identified and the effective value of current tests is determined.

5.1.3 Integration Testing

Integration tests are designed to test integrated software components to determine if they actually run as one program. Testing is event driven and is more concerned with the basic outcome of screens or fields. Integration tests demonstrate that although the components were individually satisfaction, as shown by successfully unit testing, the combination of components is correct and consistent. Integration testing is specifically aimed at exposing the problems that arise from the combination of components.

5.1.4 White box testing:

White Box Testing is a testing in which in which the software tester has knowledge of the inner workings, structure and language of the software, or at least its purpose. It is used to test areas that cannot be reached from a blackbox level.

5.1.5 Black Box Testing

Black Box Testing is testing the software without any knowledge of the inner workings, structure or language of the module being tested. Black box tests, as most other kinds of tests, must be written from a definitive source document, such as specification or requirements document, such as specification or requirements document. It is a testing in which the software under test is treated, as a black box you cannot “see” into it. The test provides inputs and responds to outputs without considering how the software works.

6. CONCLUSION

The work demonstrated the potential use of machine learning in analyzing the stock market based on the company name, previous price and current prices. The developed webpage is user friendly and the accuracy of predictions are above 90 percent. The process was adopted for all the area to improve and authenticate the price of the products which are useful for the country’s GDP growth. The various strategies used in stock market divisions divided by mathematical strategies and ML strategies. The purpose behind this survey

is to classifying the current techniques related to adapted methodologies, used various datasets, performance matrices, and applying techniques, most dominant journals using 30 investigative articles. The techniques used in the stock market prediction are categorized in different ML algorithms. For improving the prediction accuracy, some of the selected studies use the hybrid approaches in the stock market. Web scraping techniques are widely used approach for achieving the success ful stock market prediction. These techniques can design for monitoring and surveillance of the whole stock market. The big challenge that stock market prediction face is that most current techniques cannot be identified with the aid of historical data on stocks.

8. FUTURE WORK

In the future, we will strive to improve the system for making a reliable stock market system that is more reliable and accurate.

9. REFERENCES

- [1] Sahaj Singh Maini, Govinda.K, " Stock Market Prediction using Data Mining Techniques," IEEE International Conference on Intelligent Sustainable Systems, 2017.
- [2] Si, J., Mukherjee, A., Liu, B., Li, Q., Li, H., & Deng, X, Stock Market Prediction On The Basis Of Sentiments Of Twitter Feeds , 2013.
- [3] Mittal, A., & Goel , A., Sentiment Analysis On Twitter Feeds to Discover The Interrelationship Among “Public Sentiment “And The “Market Sentiment, 2012
- [4] Gidofalvi, G., & Elkan, C, Predict Short-term Movement Of Stock Price Using Financial News Articles, 2003.
- [5] Olivier C., Blaise Pascal University: “Neural network modeling for stock movement prediction, state of art”. 2007
- [6] Leng, X. and Miller, H.-G. : “Input dimension reduction for load forecasting based on support vector machines”, IEEE International Conference on Electric Utility Deregulation, Restructuring and Power Technologies (DRPT2004), 2004.
- [7] Nayak, S.C. : “Index prediction with neuro-genetic hybrid network: A comparative analysis of performance”, IEEE International Conference on Computing, Communication, and Applications (ICCCA), pp. 1-6, 2012.
- [8] Fazel Z., Esmail H., Turksen B.: “A hybrid fuzzy intelligent agent-based system for stock price prediction”, International Journal of Intelligent Systems, 2012.
- [9] Honghai Y., Haifei L.: “Improved Stock Market Prediction by Combining Support Vector Machine and Empirical Mode Decomposition”, Fifth International Symposium on Computational Intelligence and Design (ISCID), 2012