Stock Market Prediction Using Machine Learning

K.E.Uvagai[1], R.Logeshwari[2], G.Anuja[3], Deepa.G[4]
[1, 2, 3] Dept of Computer Science and Engineering
[4]Assistant Professor,Dept of Computer Science and Engineering
[1, 2, 3, 4] Adhi College of Engineering and Technology, Kanchipuram-631 605.

Abstract- *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 Prediction using machine learning helps you discover the future value of company stock and other financial assets traded on an exchange. The entire idea of predicting stock prices is to gain significant profits. Predicting how thestock market will perform is a hard task to do. Investment firms, hedge funds and even individuals have been using financial models to better understand market behaviour and make profitable investments and trades. A wealth of information is available in the form of historical stock prices and company performance data, suitable for machine learning algorithms to process. This project uses a long short-term memory (LSTM) and Linear regression, a particular neural network architecture and a machine learning algorithm, to predict the next-day closing price of the given stock.*

I. INTRODUCTION

GENERAL

A stock market, equity market, or share market is the aggregation of buyers and sellers of stocks (also called shares), which represent ownership claims on businesses; these may include securities listed on a public stock exchange, as well as stock that is only traded privately, such as shares of private companies which are sold to in vestors through equity crowd funding platforms. Investment is usually made with an an investment strategy in mind.

Stock exchanges may also cover other types of securities, such as fixed-interest securities (bonds) or (less frequently) derivatives, which are more likely to be traded OTC.Trade in stock markets means the transfer (in exchange for money) of a stock or security from a seller to a buyer. . Equities (stocks or shares) confer an ownership interest in a particular company. Participants in the stock market range from small individual stock investors to larger investors, who can be based anywhere in the world, and may include banks, insurance companies, pension funds and hedge funds. Their buy or sell orders may be executed on their behalf by a stock exchange trader.

Stock exchanges may also cover other types of securities, such as fixed-interest securities (bonds) or (less frequently) derivatives, which are more likely to be traded OTC.likely to be traded

Trade in stock markets means the transfer (in exchange for money) of a stock or security from a seller to a buyer. This requires these two parties to agree on a price. Equities (stocks or shares) confer an ownership interest in a particular company.

The New York Stock Exchange (NYSE) is a physical exchange, with a hybrid market for placing orders electronically from any location as well as on the trading floor. Orders executed on the trading floor enter by way of exchange members and flow down to a floor broker, who submits the order electronically to the floor trading post for the Designated market maker ("DMM") for that stock to trade the order. If a bid–ask spread exists, no trade immediately takes place – in this case, the DMM may use their own resources (money or stock) to close the difference. Once a trade has been made, the details are reported on the "tape" and sent back to the brokerage firm, which then notifies the investor who placed the order. Computers play an important role, especially for program trading

II. DEFINE NETWORK

The first step is to define your network.Neural networks are defined in Keras as a sequence of layers. The container for these layers is the Sequential class.The first step is to create an instance of the Sequential class. Then you can create your layers and add them in the order that they should be connected. The LSTM recurrent layer comprised of memory units is called LSTM(). A fully connected layer that often follows LSTM layers and is used for outputting a prediction is called Dense().

COMPILE NETWORK

Once we have defined our network, we must compile it.Compilation is an efficiency step. It transforms the simple sequence of layers that we defined into a highly efficient series of matrix transforms in a format intended to be executed on your GPU or CPU, depending on how Keras is configured.

FIT NETWORK

Once the network is compiled, it can be fit, which means adapt the weights on a training data set.Fitting the network requires the training data to be specified, both a matrix of input patterns, X, and an array of matching output patterns, y.

The network is trained using the back propagation algorithm and optimized according to the optimization algorithm and loss function specified when compiling the model.

EVALUATE NETWORK

Once the network is trained, it can be evaluated.The network can be evaluated on the training data, but this will not provide a useful indication of the performance of the network as a predictive model, as it has seen all of this data before.
We can evaluate the performance of the network on a separate dataset, unseen during testing. This will provide an estimate of the performance of the network at making predictions for unseen data in the future.

The model evaluates the loss across all of the test patterns, as well as any other metrics specified when the model was compiled, like classification accuracy. A list of evaluation metrics is returned.

III. LITERATURE REVIEW

The existing literature provides evidence that limit order book data can be used to predict short-term price movements in stock markets. This paper proposes a new neural network architecture for predicting return jump arrivals in equity markets with high-frequency limit order book data. This new architecture, based on Convolutional Long Short-Term Memory with Attention, is introduced to apply time series representation learning with memory and to focus the prediction attention on the most important features to improve performance. The data set consists of order book data on five liquid U.S. stocks. The use of the attention mechanism makes it possible to analyze the importance of the inclusion limit order book data and other input variables. By using this mechanism, we provide evidence that the use of limit order book data was found to improve the performance of the proposed model in jump prediction, either clearly or marginally, depending on the underlying stock. This suggests that path-dependence in limit order book markets is a stock specific feature. Moreover, we find that the proposed approach with an attention mechanism outperforms the multi-layer perceptron network as well as the convolutional neural network and Long Short-Term memory model.

Bayesian network is the graphical model which can represent the stochastic dependency of the random variables via the acyclic directed graph. In this study, Bayesian network is applied for the up/down analysis of the stock index. The up/down rates of the daily stock indexes in three major markets are taken as the network nodes and then, the network is determined by K2 algorithm with the K2 metric as the prediction accuracy of the network. The present algorithm is applied for predicting the up/down analysis of the daily stock indeies in 2007 and the results are compared with the traditional algorithms; Psychological line and trend estimation, which are popular algorithms which are well-known by the traders. Their accuracy comparison shows that the average correction rate of the present algorithm is almost 60%, which is almost equal or higher than them of the traditional algorithms such as the psychological line (50-59%) and the trend estimation (50-52%). Moreover, the vertical trading results reveal that the profit of the present algorithm is much greater than the others.

CONCLUSIONS

Application of Bayesian network to the up/down rate analysis of the stock index was presented in this study. The network was determined according to the K2 algorithm with K2 metric as the network score from the up/down rates of the stock indexes; FTSE100, DOW30 and Nikkei225. The network was applied for predicting the improvement in the FTSE100 in 2007. The present algorithm showed almost 60% correct answer rate, which is higher than the results by the traditional algorithms such as psychological line and the trend estimation. Although the correct answer rate of the psychological line with ó12 showed the similar accuracy, the number of investments is much smaller than that of the present algorithm. Therefore, the vertical investment revealed that total profit of the Bayesian network was much greater than the others.

IV. MODULES

- Data set collection
- Pre-Processing
- Feature Extraction
- Model training

- Testing model
- Performance Evaluation
- Prediction

DATASET COLLECTION:

Collecting data allows you to capture record of past events so that we can use data analysis to find recurring patterns.

From those patterns, you build predictive models using machine learning algorithms that look for trends and predict future changes.

Predictive models are only as good as the data from which they are built, so good data collection practices are crucial to developing high-performing models.

The data need to be error-free (garbage in, garbage out) and contain relevant information for the task at hand.

DATA CLEANING:

Data cleaning is a critically important step in any machine learning project.

In this module data cleaning is done to prepare the data for analysis by removing or modifying the data that may be incorrect, incomplete, duplicated or improperly formatted.

FEATURE EXTRACTION:

This is done to reduce the number of attributes in the dataset hence providing advantages like speeding up the training and accuracy improvements.

In machine learning, pattern recognition, and image processing, feature extraction starts from an initial set of measured data and builds derived values (features) intended to be informative and non-redundant, facilitating the subsequent learning and generalization steps, and in some cases leading to better human interpretations. Feature extraction is related to dimensionality reduction.

MODEL TRAINING:

A training model is a dataset that is used to train an ML algorithm. It consists of the sample output data and the corresponding sets of input data that have an influence on the output.

The training model is used to run the input data through the algorithm to correlate the processed output against the sample output. The result from this correlation is used to modify the model.

TESTING MODEL:

In this module we test the trained machine learning model using the test dataset

In machine learning, model testing is referred to as the process where the performance of a fully trained model is evaluated on a testing set

.

PERFORMANCE EVALUATION:

In this module, we evaluate the performance of trained machine learning model using performance evaluation criteria such as F1 score, accuracy and classification error.

In case the model performs poorly, we optimize the machine learning algorithms to improve the performance.

PREDICTION:

Prediction refers to the output of an algorithm after it has been trained on a historical dataset and applied to new data when forecasting the likelihood of a particular outcome, such as whether or not a customer will churn in 30 days.

The algorithm will generate probable values for an unknown variable for each record in the new data, allowing the model builder to identify what that value will most likely be.

V. EXISTING SYSTEM

The research work done by V Kranthi Sai Reddy Student, ECM, Sreenidhi Institute of Science and Technology, Hyderabad, India. In the finance world stock trading is one of the most important activities.

This paper explains the prediction of a stock using Machine Learning. The technical and fundamental or the time series analysis is used by the most of the stockbrokers while making the stock predictions.

In this paper we propose a Machine Learning (ML) approach that will be trained from the available stocks data and gain intelligence and then uses the acquired knowledge for an accurate prediction.

In this context this study uses a machine learning technique called Support Vector Machine (SVM) to predict stock prices for the large and small capitalizations and in the three different markets, employing prices with both daily and up-to-the-minute frequencies.

DISADVANTAGES

- Low accuracy
- Based on SVM – machine learning
- It is difficult to select a "good" kernel function.
- Large data sets require a long training time.

PROPOSED SYSTEM

Stock market prediction is basically defined as trying to determine the stock value and offer a robust idea for the people to know and predict the market and the stock prices.

In this project, we used Long short-term memory network (LSTM) and linear regression to predict the future stock price.

Long short-term memory network (LSTM) is a particular form of recurrent neural network (RNN).

ADVANTAGES

- High accuracy.
- Ability to learn context- specific temporal dependence.
- Provide a reliable stock price forecasting with the highest prediction accuracy.
- Stock markets help companies to raise capital.

VI. FEASIBILITY STUDY

The feasibility of the project is analyzed in this phase and business proposal is put forth with a very general plan for the project and some cost estimates. During system analysis the feasibility study of the proposed system is to be carried out. This is to ensure that the proposed system is not a burden to the company. For feasibility analysis, some understanding of the major requirements for the system is essential.

ECONOMICAL FEASIBILITY

This study is carried out to check the economic impact that the system will have on the organization. The amount of fund that the company can pour into the research and development of the system is limited. The expenditures must be justified. Thus the developed system as well within the budget and this was achieved because most of the technologies used are freely available. Only the customized products had to be purchased.

TECHNICAL FEASIBILITY

This study is carried out to check the technical feasibility, that is, the technical requirements of the system. Any system developed must not have a high demand on the available technical resources.This will lead to high demands on the available technical resources. This will lead to high demands being placed on the client. The developed system must have a modest requirement, as only minimal or null changes are required for implementing this system.

SOCIAL FEASIBILITY

The aspect of study is to check the level of acceptance of the system by the user. This includes the process of training the user to use the system efficiently. It help to many business organization to predict the correct accuracy.

SYSTEM SPECIFICATIONS

HARDWARE REQUIREMENTS

System	Pentium i3 Processor
Moniter	15'' LED
Hard disk	500 GB
Input Devices	Keyboard,Mouse
RAM	2 GB

SOFTWARE SPECIFICATION

SI NO	NAME OF THE COMPONENT	NAME OF THE REQUIREMENT
1	Operating system	Windows 10
2	Server side script	PYTHON 3.6
3	IDE	SPYDER
4	Libraries used	keres
5	Framework	Flask

VII. SOFTWARE DESCRIPTION

Python is a very popular general-purpose interpreted, interactive, object-oriented, and high-level programming language. Python is dynamically-typed and garbage-collected programming language.It was created by Guido van Rossum during 1985- 1990. Like Perl, Python source code is also available under the GNU General Public License **Python** is a MUST for students and working professionals to become a great Software Engineer specially when they are working in Web Development Domain. I will list down some of the key advantages of learning

Python:

Python is Interpreted − Python is processed at runtime by the interpreter. You do not need to compile your program before executing it. This is similar to PERL and PHP.

Python is Interactive − You can actually sit at a Python prompt and interact with the interpreter directly to write your programs.

Python is a Beginner's Language − Python is a great language for the beginner-level programmers and supports the development of a wide range of applications from simple text processing to WWW browsers to games.

CHARACTERITISTICS OF PYTHON

- It supports functional and structured programming methods as well as OOP.
- It can be used as a scripting language or can be compiled to byte-code for building large applications.
- It provides very high-level dynamic data types and supports dynamic type checking.
- It supports automatic garbage collection.
- It can be easily integrated with C, C++, COM, ActiveX, CORBA, and Java.

APPLICATIONS OF PYTHON

Easy-to-learn − Python has few keywords, simple structure, and a clearly defined syntax. This allows the student to pick up the language quickly.

Easy-to-read − Python code is more clearly defined and visible to the eyes.

Easy-to-maintain − Python's source code is fairly easy-to-maintain.

A broad standard library − Python's bulk of the library is very portable and cross-platform compatible on UNIX, Windows, and Macintosh.

PYTHON –FUNCTIONS

A function is a block of organized, reusable code that is used to perform a single, related action. Functions provide better modularity for your application and a high degree of code reusing.As you already know, Python gives you many built-in functions like print(), etc. but you can also create your own functions. These functions are called user-defined.

MACHINE LEARNING WITH PYTHON

Machine Learning (ML) is basically that field of computer science with the help of which computer systems can provide sense to data in much the same way as human beings do. In simple words, ML is a type of artificial intelligence that extract patterns out of raw data by using an algorithm or method. The key focus of ML is to allow computer systems to learn from experience without being explicitly programmed or human intervention.

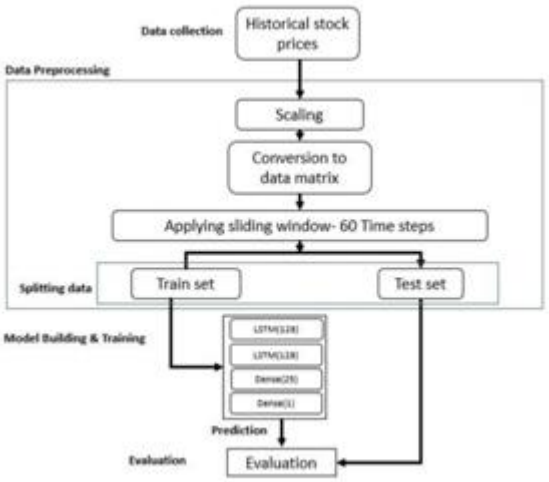
NEED FOR MACHINE LEARNING

Human beings, at this moment, are the most intelligent and advanced species on earth because they can think, evaluate and solve complex problems. On the other side, AI is still in its initial stage and haven't surpassed human intelligence in many aspects. Then the question is that what is the need to make machine learn? The most suitable reason for doing this is, "to make decisions, based on data, with efficiency and scale".

Lately, organizations are investing heavily in newer technologies like Artificial Intelligence, Machine Learning and Deep Learning to get the key information from data to perform several real-world tasks and solve problems. We can call it data-driven decisions taken by machines, particularly to automate the process. These data-driven decisions can be used, instead of using programming logic, in the problems that cannot be programmed inherently. The fact is that we can't do without human intelligence, but other aspect is that we all need to solve real-world problems with efficiency at a huge scale. That is why the need for machine learning arises.

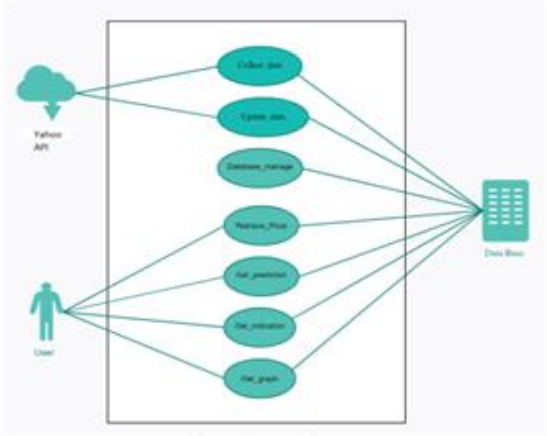
VIII. SYSTEM DESIGN

SYSTEM ARCHITECTURE

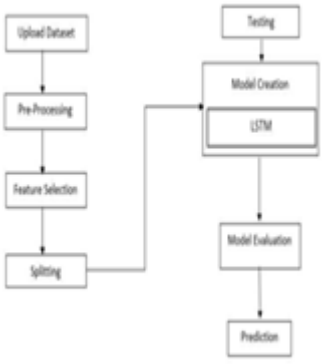


USE CASE DIAGRAM

A use case is a methodology used in system analysis to identify, clarify and organize system requirements. The use case is made up of a set of possible sequences of interactions between systems and users in a particular environment and related to a particular goal. A use case document can help the development team identify and understand where errors may occur during a transaction so they can resolve them.

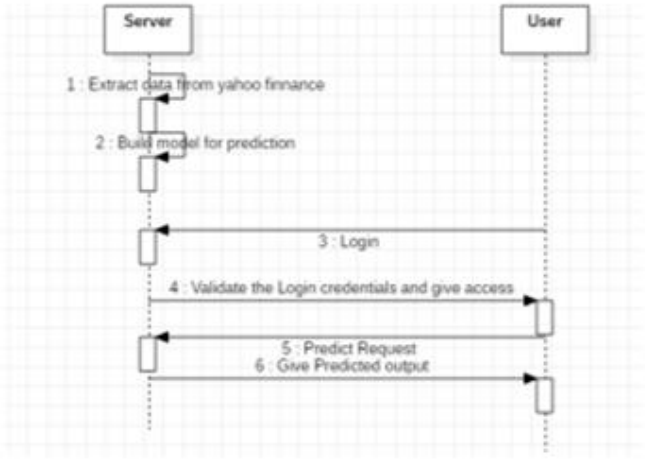


BLOCK DIAGRAM



SEQUENCE DIAGRAM

A sequence diagram simply depicts interaction between objects in a sequential order i.e. the order in which these interactions take place. We can also use the terms event diagrams or event scenarios to refer to a sequence diagram. Sequence diagrams describe how and in what order the objects in a system function. These diagrams are widely used by businessmen and software developers to document and understand requirements for new and existing systems.



REFERENCES

[1] Y. Mäkinen, J. Kanniainen, M. Gabbouj, and A. Iosifidis, ‘‘Forecasting jump arrivals in stock prices: New attention-based network architecture using limit order book data,’’ Quant. Finance, vol. 19, no. 12, pp. 2033–2050, Dec. 2019.

[2] Y. Zuo and E. Kita, ‘‘Stock price forecast using Bayesian network,’’ Expert Syst. Appl., vol. 39, no. 8, pp. 6729–6737, Jun. 2012, doi: 10.1016/j.eswa.2011.12.035.

[3] A. Kazem, E. Sharifi, F. K. Hussain, M. Saberi, and O. K. Hussain, ‘‘Support vector regression with chaos-based firefly algorithm for stock market price forecasting,’’ Appl. Soft Comput., vol. 13, no. 2, pp. 947–958, Feb. 2013, doi: 10.1016/j.asoc.2012.09.024.

[4] G. B Fang, ‘The features of volatility clustering parametric and nonparametric analysis about China’s stock marcket,’’ Technol. Econ., no. 10, 2007.

[5] T.-L. Chen and F.-Y. Chen, ‘‘Examining stock index return with pattern recognition model based on cumulative probability-based granulating method by expert knowledge,’’ Granular Comput., vol. 4, no. 4, pp. 671–685, Oct. 2019.

[6] Singh, S., Madan, T. K., Kumar, J., & Singh, A. K. (2019, July). Stock market forecasting using machine learning: Today and tomorrow. In 2019 2nd International Conference on Intelligent Computing, Instrumentation

and Control Technologies (ICICICT) (Vol. 1, pp. 738-745). IEEE.

[7] Yu, Pengfei, and Xuesong Yan. "Stock price prediction based on deep neural networks." Neural Computing and Applications 32.6 (2020): 1609-1628.

[8] Nayak, Aparna, MM Manohara Pai, and Radhika M. Pai. "Prediction models for Indian stock market." Procedia Computer Science 89 (2016): 441-449.

[9] Ghosh, Achyut, Soumik Bose, Giridhar Maji, Narayan Debnath, and Soumya Sen. "Stock price prediction using LSTM on Indian share market." In Proceedings of 32nd international conference on, vol. 63, pp. 101-110. 2019.

[10] Wei, Dou. "Prediction of stock price based on LSTM neural network." 2019 International Conference on Artificial Intelligence and Advanced Manufacturing (AIAM). IEEE, 2019.

[11] Sadia, K. H., Sharma, A., Paul, A., Padhi, S., & Sanyal, S. (2019). Stock market prediction using machine-learning algorithms. Int. J. Eng. Adv. Technol, 8(4), 25-31.

[12] Usmani, Mehak, Syed Hasan Adil, Kamran Raza, and Syed Saad Azhar Ali. "Stock market prediction using machine learning techniques." In 2016 3rd international conference on computer and information sciences (ICCOINS), pp. 322-327. IEEE, 2016.

[13] Cont., R. Empirical properties of asset returns: stylized facts and statistical issues. Quant. Finance 2001, 1, 223–236. [14]Cheng, J.; Huang, K.; Zheng, Z. towards Better Forecasting by Fusing Near and Distant Future Visions. In Proceedings of the Thirty-Fourth AAAI Conference on Artificial Intelligence (AAAI), New York, NY, USA, 7–12 February 2020