Predicting and Analyzing price of Stock Market using LSTM in RNN

line 1:1st Sivaganesh M
line 2: *Computer Science and Engineering*
line 3: *Paavai Enginnering College*
line 4: Namakkal,india
line 5: ganashmuruganpec@paavai.edu.in

line 1: 3rd Karthika M
line 2: Computer Science and Engineering
line 3: Paavai Enginnering College
line 4: Namakkal,india
line 5:karthika180402@gmail.com

line 1: 2nd Haripriya R
line 2: Computer Science and Engineering
line 3: Paavai Enginnering College
line 4: Namakkal,india
line 5: haripriya130602@gmail.com

line 1: 4th Kiruthika R A
line 2: : Computer Science and Engineering
line 3: Paavai Enginnering College
line 4: Namakkal,india
line 5: kiruthikakirthirakirk@gmail.com

***Abstract*—Predicting stock market prices is a complex task that traditionally involves extensive human-computer interaction.due to the correlated nature of stock prices,conventional batch processing methods cannat be utilized efficiently for stock market analysis.it is near impossible to predict stock prices,owing to the volatility of factor that play a major role in the movement of prices.However,it is possible to make an educated estimate of prices.the inputis given in the form of company.**

I. INTRODUCTION

The main objective of the project is to see in which precision a Machine learning algorithm can predict the stock market values. The existing system has low accuracy where overfitting occurs due to improper use of the input parameters. The proposed system is better at predicting as it gives accurate prediction and forecasting output, which is not the case with traditional models. The project deals with the stock market, which is considered to be very dynamic and complex in nature Predicting stock prices is an uncertain task that is modeled using machine learning to predict the return on stocks.Long Short-term Memory is a kind of Recurrent Convolutional Network that is extremely powerful in sequence prediction problems because they're able to store past information .This is important in our case because the previous price of a stock is crucial in predicting its future price. An accurate prediction of future prices may lead to a higher yield of profit for investors through stock investments. As per the predictions, investors will be able to pick stocks that may give a higher return.automatic compliance to electronic requirements that facilitate the concurrent or later production of electronic products, and (3) conformity of style throughout a conference proceedings. Margins, column widths, line spacing, and type styles are built-in; examples of the type styles are provided throughout this document and are identified in italic type, within parentheses, following the example. Some components, such as multi-leveled equations, graphics, and tables are not prescribed, although the various table text styles are provided. The formatter will need to create these components, incorporating the applicable criteria that follow.

1.1Technologies used

Project Domain - Machine Learning

Machine learning is an application of artificial intelligence (AI) that provides systems the ability to automatically learn and improve from experience without being explicitly programmed.Machine learning techniques and algorithms focus on the development of computer systems that can easily access data and make it readily available for using it for improvement

A SOME MACHINE LEARNING METHODS

Machine learning algorithms have been categorized into two main subcategories- supervised and unsupervised.

Supervised algorithms require supervision by someone who has machine learning skills to provide input and receive the desired output. In addition to that, he/she will be involved in furnishing feedback about the accuracy. Once the model training is complete, the algorithm will be applied to new data.

Unsupervised algorithms do not require any training with the data. However, they do make use of an iterative approach. This approach is called Deep Learning.

Unsupervised learning algorithms are also referred to as neural networks. These networks are used wherever the complexity is more and can perform more tasks than supervised learning systems. These neural networks progress by combing through Dataset for training and automatically identify correlations between variables present in the dataset. Once the model is trained, the algorithm can then use its associations to test the data.

1.2Existing System

Traditional approaches to stock market analysis and stock price prediction include fundamental analysis, which looks at a stock's past performance and the general credibility of the company itself, and statistical analysis, which is solely concerned with number crunching and identifying patterns in stock price variation. The latter is commonly achieved with the help of Genetic Algorithms (GA) or Artificial Neural Networks (ANN's), but these fail to capture the correlation between stock prices in the form of long-term temporal dependencies. Another major issue with using simple ANNs for stock prediction is the phenomenon of exploding / vanishing gradient, where the weights of a large network either become too large or too small (respectively), drastically slowing

their convergence to the optimal value. This is typically caused by two factors:

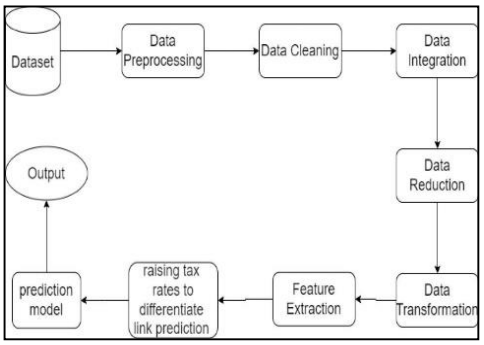
weights are initialized randomly, and the weights closer to the end of the network also tend to change a lot more than those at the beginning. An alternative approach to stock market analysis is to reduce the dimensionality of the input data and apply feature selection algorithms to shortlist a core set of features (such as GDP, oil price, inflation rate, etc.) that have the greatest impact on stock prices or currency exchange rates across markets. However, this method does not consider long-term trading strategies as it fails to take the entire history of trends into account; furthermore, there is no provision for outlier detection.

1.3 Proposed System

We propose to use LSTM (Long Short Term Memory) and Random Forest algorithms to provide efficient stock price prediction.. LSTMs and Random Forest are capable of learning very long-term dependencies and they work tremendously well on a large variety of problems. We use the dropout technique in our model to avoid overfitting. In this project, we perform both numerical analysis and textual analysis on the stocks and news dataset to try to predict the future price of the stock. The numerical analysis will be performed by treating the stock trend as a time series and we try to forecast future prices by observing the prices over the last x number of days.

In the textual analysis we perform sentiment analysis of the news articles and learn the influences of news on stock prices. Finally, predictions from these two models will be used as input to a merged model to output final predictions. The changes in the stock market is not always be in a regular pattern or not always follow the continuous cycle. Based on the companies and sectors, the existence of the trends and the period of their existence will differ.

The analysis of this type of cycles and trends can offer a more profit to the investors. In future work, we add more stock market data and compare more models to improve the accuracy of predicted stock prices. In the future, for better accuracy models can be trained with more varied and detailed data. Also, other algorithms along with proposed can be used to create a new hybrid model.



1.3.1 PROBLEM STATEMENT

Due to the advancement of technologies, there are many prediction systems that are implemented in the field of stock market, but Units every system lacks efficiency and accuracy.

2.LITERATURE REVIEW

- Xiongwen Pang, Yanqiang Zhou, Pan Wang, Weiwei Lin, "An innovative neural network approach for stock market prediction", 2020, The Journal of Supercomputing volume 76, pages2098–2118
- Adil Moghar, "Stock Market Prediction Using LSTM Recurrent Neural Network," 2020, Procedia Computer Science, Volume 170, pg no. 1168-1173
- Samuel Olusegun Ojo; Juliana Adeola Adisa; Pius Adewale Owolawi; Chunling T. Du; 2020, "Stock Market Behaviour Prediction using Long Short-Term Memory Network and Gated Recurrent Unit International Conference on Computational Science and Computational Intelligence (CSCI).
- Mehrnaz Faraz; Hamid Khaloozadeh; Milad Abbasi, 2020, "Stock Market Prediction-by-Prediction Based on Autoencoder Long Short- Term Memory Networks", 28th Iranian Conference on Electrical Engineering (ICEE)

3.SYSTEM DESIGN

This diagram is nothing but a simple description of all the entities that have been incorporated into the system. The diagram represents the relations between each of them and involves a sequence of decision-making processes and steps. You can simply call it a visual or the whole process and its implementation. All functional correspondences are explained in this diagram.

Figure 3.1 System Architecture

3.1.1.PROCESS MODEL

3.1.2.DFD

A. Data Flow Digram Level 0

- This is basically a contextual diagram,also referred to as a cotext diagram.it only represents the top level or the 0 Level in the Whole process.
- It gives an abstraction kind of view and shows the whole system as a single process and its relationship to externalities.

Figure 3.2.1 A. Data Flow Diagram Level 0

B.Data Flow Diagram Level 1

Level DFDs represent the complete system as a single process. it notates every process and sub-process that comes together in a sequence to form the complete system. This along with and 2-level data flow diagrams comprise the fundamental system model.

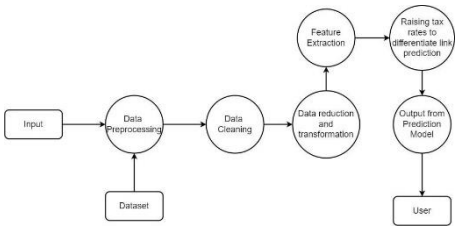


Figure 3.2.1 B.Data Flow Digaram Level 1

4.MODULES

4.1.Data Preprocessing Module

In Machine learning, Data pre-processing is an most important step and to determine some predictions , this step can't be excluded .Data preprocessing comprises of 4 steps . The first step is the data cleaning in which the duplicate, incorrectly formatted , corrupted data will be fixed or removed. The second step is the data integration is which it combines multiple sources data into single view. The third step is the data reduction step in which the data are encoded ,scaled sorted if needed .The final step is the data transformation is which the data is transformed into a required format

4.2. Feature Extraction Module

A few obstacles can be involved in approaching network data sets using machine learning models, including undefined euclidian distance, extracting relevant features, unequal data classes due to the size of the actual networks, or embedding graphs in low vector space while maintaining the structure to read networks.We offer a variety of research techniques to remove features in-network data sets based on metrics such as the most powerful methods, possible methods, and learning graph representation. Some of our contributions include raising tax rates to differentiate link prediction methods and continue to present important network data collections to learn the link prediction problem.

4.3 Prediction Module

The final module is the prediction model.This module predicts the output using the given dataset and produces the results with this the customer can be benefited most.so they know how the stocks rate going to be in future they can invest in the area where they are benefited.Results of testing were remarkable marks in terms of usability and learning impact of the system.

5. ALGORITHMS

5.1.LSTM

Long Short-Term Memory (LSTM) networks are a type of recurrent neural network capable of learning order dependence in sequence prediction problems. This is a behavior required in complex problem domains like machine translation, speech recognition, and more. LSTMs are a complex area of deeplearning.

5.2. RANDOM FOREST

Random forest is a Supervised Machine Learning Algorithm that is used widely in Classification and Regression problems. It builds decision trees on different samples and takes their majority vote for classification and average in case of regression.

6. SYSTEM REQUIREMENTS

6.1. Hardware Requirements

Microsoft Server enabled computers, preferably

workstations

Higher RAM, of about 4GB or above

Processor of frequency 1.5GHz or above

6.2. Software Requirements

Python 3.6 and higher
Anaconda software

7.TESTING METHODOLOGIES

There are many different types of testing methods or techniques used as part of the software testing methodology. Some of the important testing methodologies are:

7.1.System Testing

Testing is performed to identify errors. It is used for quality assurance. Testing is an integral part of the entire development and maintenance process. The goal of the testing during phase is to verify that the specification has been accurately and completely incorporated into the design, as well as to ensure the correctness of the design itself. For example the design must not have any logic faults in the design is detected before coding commences, otherwise the cost of fixing the faults will be considerably higher as reflected. Detection of design FAULTS can be achieved by means of inspection as well as walkthrough.

Testing is one of the important steps in the software development phase. Testing checks for the errors, as a whole of the project testing involves the following test cases:

Static analysis is used to investigate the structural properties of the Source code.
Dynamic testing is used to investigate the behavior of the source code by executing the program on the test data

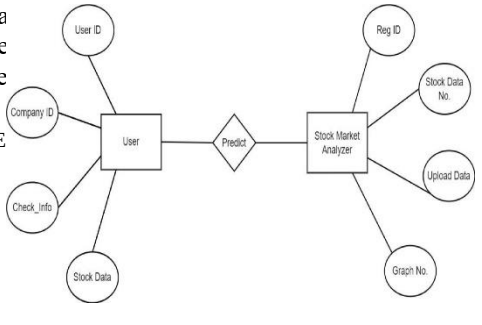
7.2.UNIT TESTING

Unit testing is conducted to verify the functional performance of each modular component of the software. Unit testing focuses on the smallest unit of the software design (i.e.), the module. The white-box testing techniques were heavily employed for unit testing.

7.3.FUNCTIONAL TESTING

Functional test cases involved exercising the code with nominal input values for which the expected results are known, as well as boundary values and special values, such as logically related inputs, files of identical elements, and empty files.
Three types of tests in Functional test:
Performa
Stress Te
Structure

7.4. INTE



Integration testing is a systematic technique for construction the program structure while at the same time conducting tests to uncover errors associated with interfacing. i.e., integration testing is the complete testing of the set of modules which makes up the product. The objective is to take untested modules and build a program structure tester should identify critical modules. Critical modules should be tested as early as possible. One approach is to wait until all the units have passed testing, and then combine them and then tested. This approach is evolved from unstructured testing of small programs. Another strategy is to construct the product in increments of tested units. A small set of modules are integrated together and tested, to which another module is added and tested in combination. And so on. The advantages of this approach are that, interface dispenses can be easily found and corrected.The major error that was faced during the project is linking error. When all the modules are combined the link is not set properly with all support files.Then we checked out for interconnection andthe links. Errors are localized to the new module and its intercommunications. The product development can be staged, and modules integrated in as they complete unit testing. Testing is completed when the last module is integrated and tested.

8.Conclusion

Thus, for the better accuracy result of stock market prediction , data preprocessing has to be done properly and in data preprocessing process data cleaning and data visualization are the most important things.. Most importantly, the above assessment not only helped us to predict the outcome but also gave us valuable insights into the type of data, which could be used in the future to train our differentiators in the best possible way

REFERENCES

[1]Pratheeth S, Vishnu Prasad R, 2021, Stock Price Prediction using Machine Learning and Deep Learning, 2021 IEEE Mysore Sub SectionInternational Conference (MysuruCon).

[2]Xiongwen Pang, Yanqiang Zhou, Pan Wang, Weiwei Lin, "An innovative neural network approach for stock market prediction", 2020, The Journal of Supercomputing volume 76, pages 2098–2118.

[3]Adil Moghar, "Stock Market Prediction Using LSTM Recurrent Neural Network," 2020, Procedia Computer Science, Volume 170, pg no. 1168-1173.

[4]Ishita Parmar, Navanshu Agarwal, Sheirsh Saxena, 2018, "Stock Market Prediction Using Machine Learning," 2018 First International Conference on Secure Cyber Computing and Communication (ICSCCC).

[5]Samuel Olusegun Ojo; Juliana Adeola Adisa; Pius Adewale Owolawi; Chunling T. Du; 2020, "Stock Market Behaviour Prediction using Long Short-Term Memory Network and Gated Recurrent Unit International Conference on Computational Science and Computational Intelligence (CSCI).

[6]Mehrnaz Faraz; Hamid Khaloozadeh; Milad Abbasi, 2020, "Stock Market Prediction-by-Prediction Based on Autoencoder Long Short- Term Memory Networks", 28th Iranian Conference on Electrical Engineering (ICEE).

[7]Agustinus Bimo Gumelar; Haryati Setyorini; Derry Pramono Adi; Sengguruh Nilowardon, 2020, "Boosting the Accuracy of Stock Market Prediction using XGBoost and Long Short-Term Memory", International Seminar on Application for Technology of Information and Communication (iSemantic).

[8]R. Shanmuga Priya & C. Sruthi, 2021, "Stock Price Prediction Based on Deep Learning Using Long Short-Term Memory", Part of the Lecture Notes in Electrical Engineering book series (LNEE,volume 792), Springer.