STOCK PRICE PREDICTION USING MACHINE LEARNING

Submitted by

[ ADITYA GADDHYAN ( 13000117134) ]

[ GOURAV SINGH VEESEN ( 13000117103) ]

[ ABHISHEK KUMAR GOND ( 13000117138) ]

[ ROHIT KUMAR BHAGAT ( 13000117063) ]

Under the guidance of Prof. (Mrs) Manashi De

Submitted for the partial fulfillment for the degree of Bachelor of Technology in Computer Science and Engineering



Techno Main Salt Lake

EM 4/1, Salt Lake, Sector – V, Kolkata – 700 091.

**ACKNOWLEDGEMENT**

We would like to express our sincere gratitude to Prof. (Mrs.) Manushi Das of the department of Computer Science and Engineering, whose role as project guide was invaluable for the project. We are extremely thankful for the keen interest he / she took in advising us, for the books and reference materials provided for the moral support extended to us.

Last but not the least we convey our gratitude to all the teachers for providing us the technical skill that will always remain as our asset and to all non-teaching staff for the gracious hospitality they offered us.

Place: Techno Main Salt Lake

Date: 9th December, 2020

ADITYA GADDHYAN ( 13000117134)

GOURAV SINGH VEESEN ( 13000117103)

ABHISHEK KUMAR GOND ( 13000117138)

ROHIT KUMAR BHAGAT ( 13000117063)

Department of Computer Science and Engineering

Techno Main Salt Lake

Kolkata – 700 091

West Bengal, India.

**APPROVAL**

This is to certify that the project report entitled “STOCK PRICE PREDICTION USING MACHINE LEARNING” prepared under my supervision by ADITYA GADDHYAN ( 13000117134), GOURAV SINGH VEESEN ( 13000117103), ABHISHEK KUMAR GOND ( 13000117138), ROHIT KUMAR BHAGAT ( 13000117063) *,* be accepted in partial fulfillment for the degree of Bachelor of Technology in Computer Sciences and Technology.

It is to be understood that by this approval, the undersigned does not necessarily endorse or approve any statement made, opinion expressed or conclusion drawn thereof, but approves the report only for the purpose for which it has been submitted.

|  |  |
| --- | --- |
| ………………………………………..…  Signature, Name & Designation of  Internal Guide(s) | …………………………………………  Signature, Name & Designation  of the HOD |

Table of Contents

[Introduction](#_heading=h.30j0zll) **2**

[Briefing](#_heading=h.1fob9te) 2

[Problem Domain](#_heading=h.3znysh7) 2

[Glossary](#_heading=h.2et92p0) 2

[Problem Definition](#_heading=h.tyjcwt) **3**

[Scope](#_heading=h.3dy6vkm) 3

[Exclusions](#_heading=h.1t3h5sf) 3

[Assumptions](#_heading=h.4d34og8) 4

[Related Studies](#_heading=h.2s8eyo1) **4**

[Project Planning](#_heading=h.17dp8vu) **5**

[Software Life Cycle Model](#_heading=h.3rdcrjn) 5

[Scheduling](#_heading=h.26in1rg) 6

[Cost Analysis](#_heading=h.lnxbz9) 7

[Requirement Analysis](#_heading=h.35nkun2) **7**

[Requirement Matrix](#_heading=h.1ksv4uv) 7

[Requirement Elaboration](#_heading=h.44sinio) 7

[5.2.1 ml\_01 load the dataset : load a dataset with at least 120 data entries.](#_heading=h.2jxsxqh) 7

[Design](#_heading=h.z337ya) 8

[Technical Environment](#_heading=h.3j2qqm3) 9

[Hierarchy of Modules](#_heading=h.1y810tw) 9

[Detailed Design](#_heading=h.4i7ojhp) 9

[Test Plan](#_heading=h.1ci93xb) **10**

[Conclusion](#_heading=h.3whwml4) **10**

[Project Benefits](#_heading=h.2bn6wsx) 10

[Future Scope for improvements](#_heading=h.qsh70q) 10

[References / Bibliography](#_heading=h.3as4poj) **11**

[APPENDIX A – Published Paper (optional)](#_heading=h.1pxezwc) **12**

[APPENDIX A – Prototypes (optional)](#_heading=h.49x2ik5) **12**

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| IntroductionBriefing The stock market is an essential component of the nation’s economy, where most of the capital is exchanged around the world. Therefore, the stock market’s performance has a significant influence on the national economy. It plays a crucial role in attracting and directing the distributed liquidity and savings into optimal paths. In this way, the scarce financial resources could be adequately allocated to the most profitable activities and projects.  Accurate prediction of stock market returns is a very challenging task due to the volatile and non-linear nature of the financial stock markets. The core idea behind the project is to create a robust and effective stock price prediction system using different tools and parameters. The output will contain the prediction using LSTM and alongside also show the forecast using regression and moving average. Problem Domain The stock price prediction clubbed with some other algorithms is widely used for algorithm based trading where in the field of FO and stocks trading, the computer automatically buys and sells stock based on real time prediction and calculation. The efficiency of these applications directly depends on the efficiency of the forecast. So this forecast system can be further used for algorithm based trading.  So we will be using deep learning techniques to forecast the price of NIFTY stocks on a real time basis which will be in the form of a web app.  Also, the small traders or individuals whose primary job is not trading, can use the forecast and likewise place their bids using this application. Glossary  |  |  | | --- | --- | | **Terms** | **Definition** | | LSTM | 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 | | Moving average | Moving Average is a calculation used to analyze data points by creating a series of averagesof different subsets of the full data set. | | Regression | Regression is a statistical method used in finance, investing, and other disciplines that attempts to determine the strength and character of the relationship between one dependent variable and a series of other variables. | | Web application | A web application (or web app) is application software that runs on a web server,Web applications are accessed by the user through a web browser with an active internet connection. | | Neural Networks | Neural networks are widely parallel interconnected networks composed of simple adaptive  units whose organization can simulate the interaction of biological nervous systems with real-world  objects | | NIFTY | Nifty is an abbreviation of National Stock Exchange Fifty, it is the broad index of National Stock Exchange (NSE) | | Deep learning | Deep learning is an AI function that mimics the workings of the human brain in processing data for use in detecting objects, recognizing speech, translating languages, and making decisions. |   *Table 1.3.1: Glossary* Problem DefinitionScope We aim to build a dynamic and robust system with good efficiency for forecasting the prices of stock based on LSTM. Finally the web app will be efficient enough to forecast the NIFTY 50 stocks on a real time basis. Exclusions With little tweaks, it can also be used to predict the prices of FO shares and commodities but here we will build it in a way that it can be used with stock. (FO needs other parameters like expiry, strike price, premium etc for forcast) Assumptions More or less, since we are aiming to predict the prices and the data needed for training is 60-90 days, so we assume that the company is at least 60-90 days old. Related Studies Multiple methods are used in the finance sector from olden days like moving average and regression line. But with Google's use of LSTM, the forecasts have become more accurate and continuous research in algorithms to achieve a higher accuracy is being made. The use of simpler algorithms like moving average and regression is being replaced with LSTM and auto ARIMA. There are some more algorithms like k nearest neighbours.  Using a Keras Long Short-Term Memory (LSTM) Model to Predict Stock Prices. LSTMs are very 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.  We will use MinMaxScaler from Sci-kit Learn to normalise the data. The data is not normalized and the range for each column varies, especially Volume. Normalizing data helps the algorithm in converging i.e. to find local/ global minimum efficiently.  As is known, both macroeconomic factors and financial series inherent changes can influence the stock index price. Xiong et al. applied Long Short Term Memory neural network to model SP500 index volatility with Google domestic trends as the indicators of the macroeconomic factors.  There are three types of gates in the LSTM cell to protect and control cell state. Every gate has an expression of σ ( W i X + b i ) . The range of sigmoid layer output is (0,1), which indicates how much of each component in C t − 1 should be passed. If the output is 0, it means that no pass is allowed while the output of 1 representing all pass.    Fig 3.1 Long Short Term Memory (LSTM) model structure. C t is the cell state and h t is the hidden state. σ represents the sigmoid activation function and tanh represents the tanh activation function. ⊗ means the element-wise product and ⊕ means concatenation operation.   * **Forget Gate(f):** It determines to what extent to forget the previous data. * **Input Gate(i):** It determines the extent of information to be written onto the Internal Cell State. * **Output Gate(o):** It determines what output(next Hidden State) to generate from the current Internal Cell State.   The basic work-flow of a Long Short Term Memory Network is similar to the work-flow of a Recurrent Neural Network with only difference being that the Internal Cell State is also passed forward along with the Hidden State. Project PlanningSoftware Life Cycle Model The Waterfall Model was the first Process Model to be introduced. It is also referred to as a **linear-sequential life cycle model**. It is very simple to understand and use. In a waterfall model, each phase must be completed before the next phase can begin and there is no overlapping in the phases.    fig 4.1.1 The following illustration is a representation of the different phases of the Waterfall Model. Scheduling  Cost Analysis Since the project is not complete, it is not possible to provide the cost analysis. Requirement AnalysisRequirement Matrix    |  |  |  |  |  |  |  |  | | --- | --- | --- | --- | --- | --- | --- | --- | | **Rqmt ID** | ***Requirement Item*** | ***Requirement Status*** | ***Design Module*** | ***Design Reference (section# under project Report)*** | **Test Case**  **Number** | ***Technical Platform of Implementation*** | ***Prototype prepared ?*** | | ml\_01 | load the dataset | In-progress | MLC | 5.5.1 | T\_ML\_01 | Python | No | | ml\_02 | build the machine learning code | In-progress | MLC | 5.5.1 | T\_ML\_02 | Python | No | | ml\_03 | check for the efficiency | In-progress | MLC | 5.5.1 | T\_ML\_03 | Python |  | | WA\_01 | build web app | In-progress | WAC | 5.5.2 |  | Flask and JS | No | | WA\_02 | make the UI friendly | In-progress |  |  |  |  |  | | WA\_03 | check the exceptions (like missing input from user) | In-progress | WAC | 5.5.2 | T\_WA\_01 | Flask and JS | No | | WA\_04 | integrate | In-progress |  |  |  |  |  | | WA\_05 | the web app should show the stock price | In-progress | DPC | 5.5.2 | T\_WA\_02 | Flask and JS |  | | dp\_01 | deploy on cloud | In-progress | DPC | 5.5.3 |  |  |  | | dp\_02 | Automatic pulling of real time dataset | In-progress | DPC | 5.5.3 | T\_WA\_03 |  |  | | dp\_03 | acceptance testing | In-progress | DPC | 5.5.3 | T\_WA\_04 |  |  | |  |  |  |  |  |  |  |  |  Requirement Elaboration 5.2.1 ml\_01 load the dataset : load a dataset with at least 120 data entries. 5.2.2 ml\_02 build the machine learning code : build a fully functional ML code.  5.2.3 ml\_03 check for the efficiency : test on the dataset and check for the efficiency.  5.2.4 WA\_01 build web app : build a robust web app.  5.2.5 WA\_02 make the UI friendly: make the UI user friendly so that the user does not find it cumbersome to use.  5.2.5 WA\_03 check the exceptions (like missing input from the user): implement proper exception and error handling.  5.2.6 WA\_04 integrate : integrate the webapp and the ml code.  5.2.7 WA\_05 the web app should show the stock price : check that the output of the web app is in correct format and matches with the one in unit testing.  5.2.8 dp\_01 deploy on cloud : deploy on GCP along with proper script to load some dataset.  5.2.9 dp\_02 Automatic pulling of real time dataset : the data of symbols that we offer to forecast should be pulled everyday after the market closes in order to carry out real time forecasting.  5.2.10 rq\_11 acceptance testing: final testing of everything on different platform. Design   6.3.1 The Workflow of the project Technical Environment For machine learning, we will use python3 along with its rich libraries like numpy, pandas, sklearn and keras. For the web interface, the frontend will be on Vanilla JS alone and the backend on flask. For the cloud, we will use GCP and for the CI/CD, we will here use jenkins. The VCS used will be git and the code will be hosted on github. For the database, we will use MongoDB.  **Resource requirements**  **Server side:**   * MongoDB atlas cluster of atlest 50GB * GCP VM or compute engine instance with 10GB of space and Ubuntu LTS OS [Alternatively AWS EC2 instance can be used for the same]   **User side**  Any mobile or computer device to access the web page. Hierarchy of Modules   6.2.1 hierarchy of modules Detailed Design The design is divided into three parts: the machine learning code, the web app and the deployment. the following are the sub sections:  Load the sample dataset of some company with at least 120 days of data where we will use 60-90 days of data or training and the rest for testing purposes. Also after data cleaning and pre processing, we will build LSTM. To build our model we are going to use the LSTM RNN, our model uses 80% of data for training and the other 20% of data for testing. For training we use mean squared error to optimize our model [2]. We will try training and testing on different epochs and will go with the one with the most feasible result. For the development purpose we will be using the data of a handful companies and in the production, we will have three months data starting from the current date of nifty stocks.  We will build the web app and integrate it. We will first check the efficiency of our web app to handle and point out errors like missing input or improper input. After that we will move on to test the code. If it gives the same result as it used to in unit testing, we will carry out some final rounds of tests.  Here we will deploy our code on GCP and test it using the web browser, using its public IP. We will carry out platform testing on different browsers. Also, if possible, we will try to set up a CI/CD pipeline for automatic deployment and integration where any changes in remote repository will trigger a change in production automatically. Test Plan The exact test cases are yet to be decided but the test will go as follows: ConclusionProject Benefits The project can be used in predicting the prices of stock and can also be used in algorithm based trading. Future Scope for improvements We can use different important parameters like implied and historic volatility, take the result of Auto ARIA and k nearest neighbour and moving average in consideration to improve the efficiency. Also, a more robust algorithm based on deep learning can be used to take different market sentiments into consideration and predict unusual movements.  References / Bibliography  1. Moghar, A., & Hamiche, M. (2020). *Stock Market Prediction Using LSTM Recurrent Neural Network. Procedia Computer Science, 170, 1168–1173.* doi:10.1016/j.procs.2020.03.049 [<https://doi.org/10.1016/j.procs.2020.03.049> ] 2. Nelson, D. M. Q., Pereira, A. C. M., & de Oliveira, R. A. (2017). *Stock market’s price movement prediction with LSTM neural networks. 2017 International Joint Conference on Neural Networks (IJCNN).* doi:10.1109/ijcnn.2017.7966019 [<https://doi.org/10.1109/AIAM48774.2019.00113>] 3. Predicting stock prices with LSTM [<https://towardsdatascience.com/predicting-stock-price-with-lstm-13af86a74944>] 4. The Application of Stock Index Price Prediction with Neural Network *Mathematical and Computational Applications* (ISSN 2297-8747; ISSN 1300-686X for printed edition) is an international [peer-reviewed](https://www.mdpi.com/editorial_process) open access journal on the applications of the mathematical and/or computational techniques published quarterly online by MDPI from Volume 21 Issue 1 (2016). [<https://www.mdpi.com/journal/mca>]  APPENDIX A – Published Paper (optional) None APPENDIX A – Prototypes (optional) None |
|  |

**Instructions (remove the following after your report is completed):**

1. Already included ToC in this template. After completing the document, ensure to right-click on ToC 🡪 Update field 🡪 Update entire table for automatically updating the ToC.
2. **Citations** should be used for all referred texts using appropriate numbers within the square bracket for all mapped references under Section 8: References. You should check any standard journal paper for typical use of citations.
3. Depending on the type of your project, sections can be altered to this generic template.
4. Except under TOC, **Font** Style=”Times New Roman”, Font Size=”12” and Alignment=”Justified” should be uniformly used for the project documentation. Needless to say **spellchecker** should be used.
5. Team should perform reasonable numbers of **proof reading** for avoiding unintentional errors and factual discrepancies before appearing in project viva.
6. **1** spiral bound copy of the project report will have for **submission** at the time of the examination.

Figure 1.1: **Sample figure**

1. For all **figures**, captions should be bold with centrally aligned and should be positioned below the figures, e.g.

Using MS-Word features, insert figures and tables after they are cited in the text so that they can automatically come after inserting / updating TOC.

Use a text box to insert a graphic (which is ideally a 300 dpi TIFF or EPS file, with all fonts embedded) because, in an MS Word document, this method is somewhat more stable than directly inserting a picture.

To have non-visible rules on your frame, use the MSWord “Format” pull-down menu, select Text Box > Colors and Lines to choose No Fill and No Line.

1. For all **tables**, captions should be bold with centrally aligned and should be positioned above the tables, e.g.

*Table 1: Sample Table*

|  |  |  |  |
| --- | --- | --- | --- |
| **Table Head** | **Table Column Head** | | |
| ***Table column subhead*** | ***Subhead*** | ***Subhead*** |
| copy | More table copya |  |  |

a. Sample of a Table footnote. (Table footnote)

1. If you have already prepared a **prototype**, indicate the same under Requirement Matrix and Detailed design. Please also specify prototype details under Appendix showing codes, screens, test data, sample output and detailed steps of compilation, execution and setups (if any).
2. If you have published related paper(s) in a standard **journal** / presented in a recognized **conference**, please ensure to refer the same under Section 8: References as well as including communication on your paper(s) acceptance / publishing note under the Appendix section. You should also show appropriate documentation at the time of project viva.