IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
IJARSCT
Impact Factor: 6.252
Volume 2, Issue 7, May 2022

Stock Prediction Webapp Using Python

Samyak Meshram , Suraj Narsale , Sahil Sangamnere , Kiran Kadam , Prof. Vrushali Paithankar
Department of Computer Engineering
Smt. Kashibai Navale College of Engineering, Pune, Maharashtra, India
Savitribai Phule Pune University, Pune, India

Abstract: The project's goal is to provide a third-party investment web application for individual investors to use to navigate the stock market. This is accomplished through the application of machine learning and artificial intelligence.Web technology that is streamlit. Several methodologies and models for stock price prediction have been developed .such as dense, feedforward neural networks, recurrent neural networks, basic linear regressions, and so on as well as linear interpolations The model architectures and hyperparameters are automatically optimised. Evolutionary algorithms are used to seek for information. For trend prediction, promising results have been discovered. The undertaking acts as a foundation for making machine learning technology more accessible to the general population in the United States. in the context of looking for investment opportunities It provides the path for new features to be added and tested. In the future, we'll be constructing AutoML models in the financial context...

Keywords: LSTM ,ATS, GRU ,ML ,SVM ,EMH ,AI, NN, ARMA, DRL, LMS ,UML, MSE ,RMSE

I. INTRODUCTION

Stock, in particular, is associated with commercialised companions and businesses that are settling in the marketization world. Stock can also be referred to as a share, which is a term that is commonly used in everyday life. It's even referred to as an investment strategy, and it's seen as a long-term investment that secures and gives ample income during retirement. Purchasing a company's stock entails owning a small portion of the corporation. People put money into it in order to earn a long-term return on their investment.

When a person buys a company stock, they are referred to as a shareholder, and they will receive a share of the profit or gain that they have invested in. The stock can be sold and bought as needed by the investor. They can share their shares with their respective or other individuals, and there are many stock brokers in the firm who are interested in doing so.

Inspired by the growing popularity of deep learning algorithms for predicting applications, these algorithms could be used to uncover hidden patterns in the trajectory of stock prices, providing additional information to retail investors when making investment decisions.

II. PROBLEM DEFINITION

Since then, stock has been an unpredictable curve in the picture. Its essence had been living and delighting for a long time. The same may be said for the company. Rather than investing and obtaining a bank loan approval, the organisation built it as a superior source of revenue generating. From a business standpoint, it is far more efficient and less stressful.

III. LITERATURE SURVEY

3.1 Stock Price Predictions

According to Y. Dai and Y. Zhang's research paper "Machine Learning in Stock Price Trend Forecasting," they employed features such as the PE ratio, PX volume, PX EBITDA, 10-day volatility, 50-day moving average, and others to estimate the next-day stock price and a long-term stock . Logistic Regression, Gaussian Discriminant Analysis, Quadratic Discriminant Analysis, and SVM are the machine learning techniques employed in the study. The accuracy ratio is calculated by dividing the number of days on which the model successfully identified the testing data by the total

IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
Volume 2, Issue 7, May 2022
IJARSCT
Impact Factor: 6.252

number of days on which the model was used. The short term model has a very poor accuracy when predicting the next day stock price; nevertheless, the Quadratic Discriminant Analysis is the best of all models, with a 58.2 percent

The larger the time frame for the long term model predicting the next n days stock values, the greater the accuracy for SVM. The SVM model's accuracy was 79.3 percent with a 44-day time span. Aside from that, it was discovered that increasing the amount of characteristics boosted accuracy.

The model's accuracy reached 79 percent when all 16 features were employed, but it plummeted to 64 percent when only 8 features were used, and 55 percent will also look into how the timeframe affects the accuracy of different models' price forecasts. Models must attain a particular level of relevance in order to be useful as a reference for users.

There have been numerous articles that have investigated this topic and have been quite helpful in the development of my paper. However, I have noticed the following comments on the articles:

(1) The leakage of test data into training inputs or the normalisation of mixed test/train data

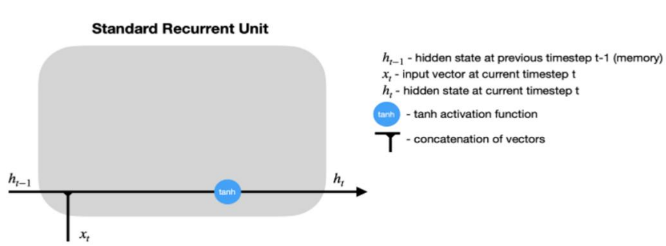
This occurs when a model relies on data that isn't currently available, but produces better-than-expected results because to the future data it already has.

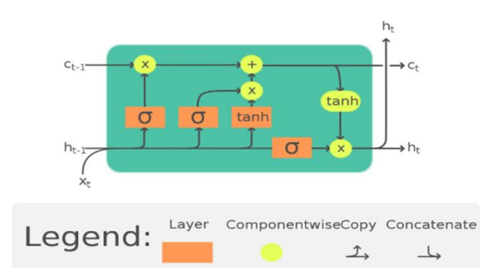
(2) Difficulty estimating the price for the next day

Many articles claim to have discovered a means to generate money using machine/deep learning, however the difficulty is that they are having trouble generating a generalizable profit, such as the next day's anticipated pricing.

III. LONG SHORT-TERM MEMORY (LSTM)

Hochreiter and Schmidhuber initially proposed long short-term memory [8] in 1997 to overcome the aforementioned issues. Lst memory is a neural network architecture that addresses the challenge of learning to remember information over a time interval by incorporating memory cells and gate units. Memory cells, each of which has a cell state that stores previously encountered information, are used in a typical formulation. The output is determined by a combination of the cell state (which is a representation of the prior information) and the cell state is updated every time an input is sent into the memory cell. The updated cell state and the new input can be used to compute the new output when another input is passed into the memory cell.





Long-Term Memory Short-Term Memory The Recurrent Neural Network algorithm is part of the deep learning algorithm family. Because of the feedback links in its architecture, it is a recurrent network. Its capacity to process the complete sequence of data gives it an advantage over standard neural networks. input gate, output gate, and forget gate are all part of its architecture.

IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
Volume 2, Issue 7, May 2022
Impact Factor: 6.252

V. SOFTWARE REQUIREMENTS SPECIFICATION

PROJECT SCOPE

The flexibility and top marks that one can present are referred to as features. Considering finance and understanding about it provided insight into the fiscal and stock markets. So the idea's selling point was handling and automating the resource, which other agents are profiting from. There are numerous models that flow in the market that are attempting to create a resource and provide predictability to the majority of it accurate, but everything is not the same, and the results are not perfect. The efficiency fluctuates with the stock market and its forecasting.

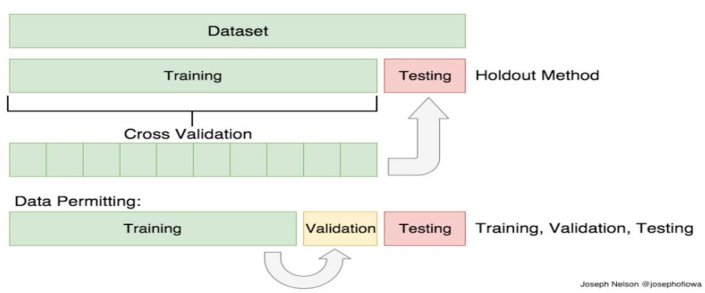
USER CLASSES AND CHARACTERISTICS

Normal users and advanced users are divided into two categories. Users who want to see more information about a model's historical (test set) performance as well as more information about the architecture and inputs can enable advanced user mode in the settings page and access those data in each individual stock

ASSUMPTIONS AND DEPENDENCIES

a DATASET

This is the first and most important module to complete before beginning the project. The dataset is a collection of data that has been joined together to indicate data variations over time for further estimation, as well as the source of the resources and their outcome for subsequent review. It generates the result optimization and provides a reasonable time frame for customising and obtaining the derivation flow.



b DATA ABSTRACTION

Abstraction is the process of selecting the best resource for categorising the above dataset and learning the most from it. The data abstraction is an important aspect of the flow. All of the data is comprised of a large number of chunks that, when processed, can limit the yield result as well as the computational mean. As a result, the data yield has to be derivative given the resources available.

c TRAINING DATASET

The machine has to be trained, which is where the training data comes in handy. Thousands of machine learning algorithms exist. The easiest way to put machine learning into practise is to provide the outcome and content in order to extract what is required within the time frame

d TEST DATASET

These are the data sets that produce the output once the data has been learned from. This is the output result of the test generation. Each phase of testing generates results. The testing phase is also known as the development phase. Now a new set of datasets is passed that is designed to be similar to the training dataset, and the efficiency of the dataset is determined.

DOI: 10.48175/IJARSCT-4359

IJARSCT
ISSN (Online) 2581-9429
IJARSCT
Impact Factor: 6.252
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
Volume 2, Issue 7, May 2022

VI. FUNCTIONAL REQUIREMENTS

SYSTEM FEATURE

Functional requirements are concerned with the software's functionality from an engineering standpoint. It improves and describes the component flow as well as the structural flow of the same.

The functional statement is concerned with categorising raw datasets and learning from the same dataset. the datasets are grouped into clusters, and their impairment is assessed for efficiency. The data is cleansed when the dataset is cleaned, and the machine learns and identifies the pattern specified for it. It then goes through numerous iterations to provide output.

EXTERNAL INTERFACE REQUIREMENTS

SOFTWARE INTERFACES

The entire training takes place in a Jupyter notebook environment. In order to train models and run the evolution process, Google Colaboratory [21], which provides an easy-to-use Jupyter notebook environment and a free GPU service, is employed in addition to each team member's personal PC. Streamlit is a Python framework that makes it simple to construct and share custom web apps for machine learning and data research.

NONFUNCTIONAL REQUIREMENTS

PERFORMANCE REQUIREMENTS

It's a feature of stock prediction software that describes how responsive it is to different user interactions.

SAFETY REQUIREMENTS

It specifies the software's user interface in terms of how easy it is to understand the user interface of stock prediction software for any type of stock trader and other stock market stakeholders.

SECURITY REQUIREMENTS

It defines CORS security, which allows resources on a web page to be requested from a domain other than the one that served the first resource.

SOFTWARE QUALITY ATTRIBUTES

Maintaining the highest possible accuracy in closing stock prices in the quickest time possible using available data

SYSTEM REQUIREMENTS

SOFTWARE REQUIREMENTS

- Operating system : Windows & Linux
- IDE : Jupyter Notebook
- Data Set : .csv file
- Visualization : mat plot lib, pandas.
- Server : Web Server with HTTP .

HARDWARE REQUIREMENTS

- Processor : Intel i5 or above
- RAM : Minimum 225MB or more.
- Hard Disk : Minimum 2 GB of space
- Input Device : Keyboard
- Output Device : Screens of Monitor or a Laptop

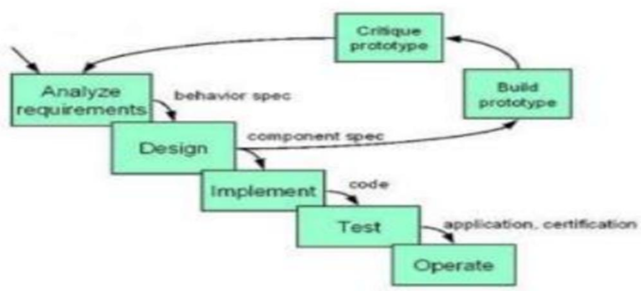
IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
IJARSCT
Impact Factor: 6.252
Volume 2, Issue 7, May 2022

ANALYSIS MODELS: SDLC MODEL TO BE APPLIED

Developing a system involves not only code writing but also many other activities such as requirement gathering, document writing, testing and others. System development is not an easy task as all the activities have to be done in a well- planned manner. Hence, the development methodologies are guidelines or blueprint to develop the system effectively and efficiently

PROTOTYPE MODEL

The prototype model requires the involvement of users or clients in the early phases, such as the requirement gathering phase and design phase, as this can help to develop the best design before implementation. Based on Chandra (2015), the prototype model is suitable for projects with unclear requirements as clients can add the requirements during the design phase. it is best to be used in the projects that have less understanding of specifications so that the development team can alter the requirements from time to time



SYSTEM IMPLEMENTATION PLAN

This project may be run on standard computer hardware. We used Intel I5 processor with 8 GB RAM and a 2 GB Nvidia Graphic Processor. It also has two cores that run at 1.7 GHz and 2.1 GHz, respectively. The first phase, which takes 10-15 minutes, is followed by the testing phase, which takes only a few seconds to make predictions and calculate accuracy. Data pre-processing, model training, and prediction are all done with Python 3.5 in Google Colab.
Windows 8 and higher, Linux-based OS, or MAC OS

Planning Phase

This is the phase that lists out all the activities and tasks that will be performed in this project and set the time frames to complete it. The planning phase is to ensure that the project can be conducted and completed smoothly. In this project, a project plan which includes Work Breakdown Structure and Gantt Chart will be created in this planning phase. Work Breakdown Structure lists down all the sub-tasks of this project while Gantt Chart set the time frames to complete each task in the Work Breakdown Structure

Analysis phase

The analysis phase is to study, analyse and gather the requirements of the project. The requirements related to the algorithm of predicting future stock prices, are gathered through researching papers and also some advice from the supervisor. The requirements of the time series algorithms and evaluation methods of the algorithms will be analysed.

Design phase

This phase aims to design the system in detail based on the requirements analysed in the previous phase. In this phase, the workflow of the tasks and a brief picture of the system's interface will be developed

IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
Impact Factor: 6.252
Volume 2, Issue 7, May 2022

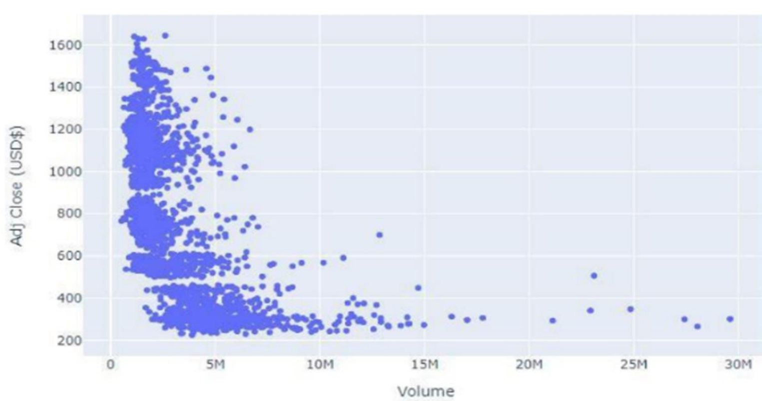
Implementation phase

After designing what the system should look like, it is followed by an implementation phase to implement the prototype by referring to the design. A prototype will be evaluated in performance and user experience. Then the prototype will go through the analysis phase, design phase and implementation phase again to refine the prototype based on the evaluation.

Deployment phase

When the prototype is being evaluated to a stage satisfaction level is high, then it will undergo deployment phase. The prototype is implemented into a system and testing of the system will be conducted. Lastly, the system will be deployed Why do we place such a high value on adjusted closeness? This price includes dividends, stock splits, rights offers, and any other events that occur after markets shut. As a result, this is the feature that we would try to forecast because it is more dependable than the other measures that are subject to non-market variables.

We can detect a general upward trend in the adjusted close price in the graph. We looked at the relationship between the various features to identify what features might be contributing to this. We notice that adjusted close has a negative connection with volume. As a result, it appears that volume will become an important component in our models.



Volume on Balance

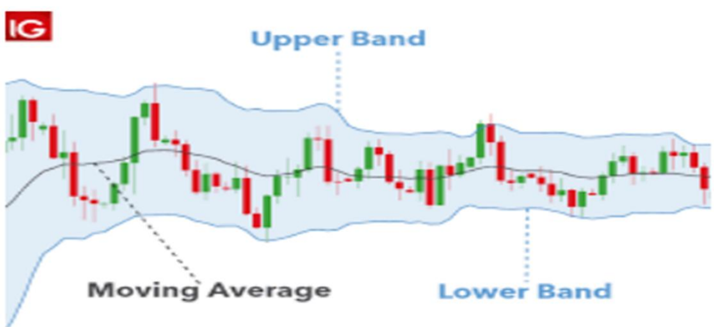
This metric is a momentum technical indicator that uses changes in volume as a proxy for stock price changes. When the current day's closing price is greater than yesterday's, we add volume, and when the current day's price is lower, we

IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
Impact Factor: 6.252
Volume 2, Issue 7, May 2022

deduct it. As demand grows, a positive volume pressure will result in higher prices, whereas a negative volume pressure will finally result in lower prices.

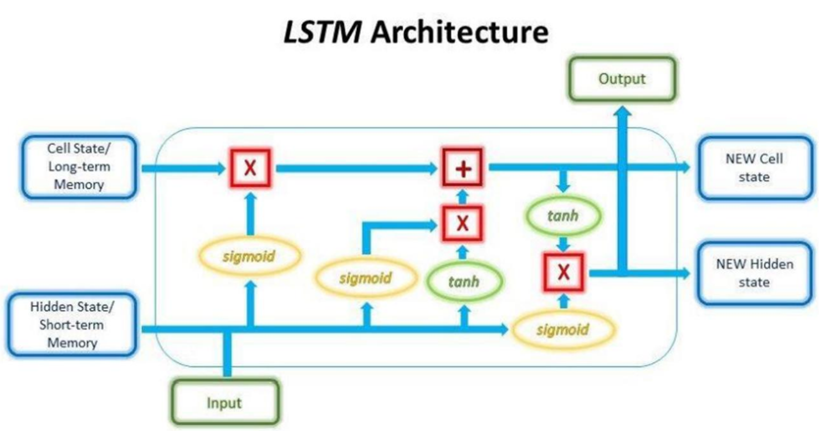
2) Moving average with an exponential component (EMA)
The exponential moving average (EMA) is a moving average that gives recent values more weight than those from a longer time period. This reduces the amount of lag and enables for a quicker reaction to price changes.

Bollinger bands
A Bollinger band is a pair of lines that are 2 deviations apart from the simple moving average of a time series. Because the majority of price action occurs within this band, it limits the potential of the next day's price. The tighter the bands, the more volatile the price movement, and hence the greater the chance that a current trend is ending or even reversing. It's time to arrange our data as input for our models after we've added the features!



VII. SYSTEM DESIGN

SYSTEM ARCHITECTURE
The system's design is based on a client-server model, with the server and client being loosely linked.



OTHER SPECIFICATIONS

ADVANTAGES

While 1-day stock price predictions are very close to actual stock prices, 10-day stock price predictions diverge significantly from actual stock prices. This demonstrates that machine learning techniques fail to accurately anticipate stock prices for retail investors.

Nonetheless, some of the models have been demonstrated to beat forecasts based on random walks, as discussed in 6.2, and so may still be useful as a reference for more experienced investors who may compare the results to their own analysis findings to spot relevant trends.

APPLICATIONS

Finally, a web application for stock price prediction was created in order to provide consumers with a better understanding of the stock market trend. To discover the best prediction models for the online application, LSTM algorithms were tested. LSTM is the most accurate of the five time series algorithms, and it was chosen as the prediction model for the stock price prediction mobile app.

VIII. CONCLUSION

On the Yahoo financial dataset, two approaches were used in this paper: LSTM and Regression. Both strategies have shown an improvement in prediction accuracy, resulting in good outcomes. The use of recently introduced machine learning algorithms to stock prediction has generated promising results, indicating that they can be used in profitable exchange schemes. It has led to the conclusion that utilising machine learning techniques, it is possible to predict the stock market with greater accuracy and efficiency.

The stock market prediction system can be enhanced in the future by using a much larger dataset than the one now in use. This would aid in improving the precision of our prediction models.

ACKNOWLEDGMENT

It brings us great pleasure to deliver the "Stock Prediction Webapp using Python" B.E project report.

I did want to take this opportunity to express my gratitude to Prof. Vrushali Paithankar, my internal guide, for providing me with all of the assistance and advice I required. I owe them a debt of gratitude for their unwavering support.Their wise advice was quite beneficial. R.H.Borhade Head of Computer Engineering Department, Smt. Kashibai Navale College of Engineering, is also to be thanked for his invaluable assistance and ideas.

Finally, we would like to express our gratitude to Dr.A.V.Deshpande for providing numerous resources for Our Project, including a laboratory with all necessary software platforms and a continuous Internet connection.

REFERENCES

[1] F. a. o. Eugene, "Efficient capital markets: a review of theory and empirical work," Journal of finance, vol. 25, no. 2, pp. 383-417, 1970.

[2] Z. A. Farhath, B. Arputhamary and L. Arockiam, "A Survey on ARIMA Forecasting Using Time Series Model," Int. J. Comput. Sci. Mobile Comput, vol. 5, pp. 104-109, 2016.

[3] S. Wichaidit and S. Kittitornkun, "Predicting SET50 stock prices using CARIMA (cross correlation ARIMA)," in 2015 International Computer Science and Engineering Conference (ICSEC), IEEE, 2015, pp. 1-4.

[4] D. Mondal, G. Maji, T. Goto, N. C. Debnath and S. Sen, "A Data Warehouse Based Modelling Technique for Stock Market Analysis," International Journal of Engineering & Technology, vol. 3, no. 13, pp. 165-170, 2018.

[5] G. Maji, S. Sen and A. Sarkar, "Share Market Sectoral Indices Movement Forecast with Lagged Correlation and Association Rule Mining," in International Conference on Computer Information Systems and Industrial Management, Bialystok, Poland, Sprigner, 2017, pp. 327-340

[6] M. Roondiwala, H. Patel and S. Varma, "Predicting stock prices using LSTM," International Journal of Science and Research (IJSR), vol. 6, no. 4, pp. 1754-1756, 2017.
[7] T. Kim and H. Y. Kim, "Forecasting stock prices with a feature fusion LSTM-CNN model using different representations of the same data," PloS one, vol. 14, no. 2, p. e0212320, April 2019.
[8] S. Selvin, R. Vinayakumar, E. A. Gopalkrishnan, V. K. Menon and K. P. Soman, "Stock price prediction using LSTM, RNN and CNN-sliding window model," in International Conference on Advances in Computing, Communications and Informatics, 2017.
[9] S. Hochreiter, "Untersuchungen zu dynamischen neuronalen Netzen," Diploma, Technische Universität München, vol. 91, no. 1, 1991. SKNCOE, Department of Computer Engineering 2021-2022 49
[10] Y. Bengio, P. Simard, P. Frasconi and others, "Learning long-term dependencies with gradient descent is difficult," IEEE transactions on neural networks, vol. 5, no. 2, pp. 157-166, 1994.
[11] S. Hochreiter and J. Schmidhuber, "LSTM can solve hard long time lag problems," in Advances in neural information processing systems, NIPS, 1997, pp. 473--479.
[12] S. Hochreiter, "The vanishing gradient problem during learning recurrent neural nets and problem solutions," International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems, vol. 6, no. 2, pp. 107-116, 1998.
[13] J. Schmidhuber, D. Wierstra, M. Gagliolo and F. Gomez, "Training recurrent networks by evolino," Neural computation, vol. 19, no. 3, pp. 757-779, 2007.
[14] L. Pasa and A. Sperduti, "Pre-training of recurrent neural networks via linear autoencoders," in Advances in Neural Information Processing Systems, NIPS, 2014, pp. 3572-3580.
[15] J. Chen and N. S. Chaudhari, "Segmented-memory recurrent neural networks," IEEE transactions on neural networks, vol. 20, no. 8, pp. 1267-1280, 2009.