MJRET

OPEN ACCESS

Multidisciplinary Journal of Research in Engineering and Technology
www.mjret.in

eISSN:2348-6953

STOCK MARKET PREDICTION ANALYSIS USING MACHINE LEARNING ALGORITHMS

Mr. Paresh Kharche, Mr. Nilesh Sanap, Mr. Ajay Vetti, Mr. Omkar Kumbhar, Prof Rathod V.U.

Computer
JSPM's Imperial College of Engineering and Research, Wagholi.
Wagholi Pune, India

Abstract: *Stock Market data has been used by many people to extract the profit from the market. Due to rise in technology and artificial intelligence, it has become easy to do analysis and predict the future prices based on the past performance of the given stock. Market prediction is in the forefront of development of deep learning algorithms. With the advancement in neural networks, there are different ways to predict the stock price. Artificial intelligence techniques are being used in conjunction with data mining to solve a real-world problem. Deep learning algorithm has increased the profits of countless people in the past few years. We are going to use deep learning model to predict the price of the stock. We propose a system that uses stacked LSTM model to predict the future prices. Our main aim is to see how this simple implementation will bring acceptable results.*

Keywords*: Time Series Analysis, Deep learning, recurrent neural network, Convolutional neural network, Long Short Term Memory, Stocks forecasting*

1. INTRODUCTION

The stock exchange is thought for its volatility, randomness, and unpredictability. It is a chaotic place with an improbably large endlessly dynamic stream of information that makes predicting and functioning on those predictions to form a profit terribly arduous. it's truly one in all the most difficult tasks in times series prediction. This capstone's main goal is to check and apply deep learning techniques to the stock market so as to predict stock behaviour and therefore act on those predictions to avoid investment risk and generate profit. The goal is to be achieved by Deep learning so as to require advantage of pre- built neural networks models. Predictions square measure then tested against actual historical stock value

knowledge. In order to try to therefore, several tools are going to be accustomed accurately reach the objectives of this project. As for the model and languages to be used, when intensive research the programming language to be used for implementation is going to be Python, this can be due to its flexibility and therefore the availableness of pre-built models and opensource notably helpful libraries which will facilitate us with our goal and perhaps even enhance results. In addition, this paper can be a straightforward example of the foremost fitting model (the one that yields the most effective results) within the case of your time series prediction that is definitely the LSTM model that stands for Long Short-Term Memory. Compared to a standard deep neural network, its effectiveness is because of the addition of a vital part in statistic predictions, the memory part. Moreover, this report will have a lot of advanced examples of LSTM models, concerning the targeted stocks.

2. LITERATURE SURVEY

[1] Mustain Billah, Sajjad Waheed, Abu Hanifa, they examined that Predicting closing stock worth accurately may be a difficult task. Computational power-assisted systems are established to be a useful tool for stock prediction like Artificial Neural Network (ANN), adaptational Neuro Fuzzy abstract thought System (ANFIS) etc. Latest analysis works prove that adaptational Neuro Fuzzy abstract thought System shows higher results than Neural Network for stock prediction. During this paper, Associate in Nursing improved Levenberg Marquardt (LM) coaching algorithmic rule of artificial neural networks has been planned. Improved Levenberg Marquardt algorithmic rule of neural networks will predict the doable day end closing stock worth with less memory and time required, provided previous historical securities market knowledge like gap worth, highest worth, lowest worth, total share listed. Moreover, Associate in Nursing improved lumen algorithmic rules will predict dayend stock worth fifty-three less error than ANFIS and ancient lumen algorithms. It conjointly needs half-hour less time, fifty-four less memory than ancient lumen and forty seventh less time, fifty-nine less memory than ANFIS.

[2] Abidatul Izzah, Yuita Arum Sari, Ratna Widyastuti, Toga Aldila Cinderatama, in this work, Stock Prediction is developed in each of 2 studies, economics, and data processing. Stock predictions got special attention because of its importance for making a more practical and economical designing. During this study, Improved Multiple regression (IMLR) was designed into a mobile application primarily based humanoid platform for stock worth prediction. IMLR may be a hybrid Multiple regression with Moving Average technique. The app was in-built many steps, that area unit demand analysis, system style, implementation, and testing. Knowledge was collected from the finance.yahoo.com page with class "Jakarta Composite Index (JKSE)"that was mechanically taken by victimization Yahoo Finance API. during this

app, users not solely may see daily stock history however conjointly stock worth predictions in real time

3. BACKGROUND

1. Machine Learning

Machine learning is a growing technology which enables computers to learn automatically from past data. Machine learning uses various algorithms for building mathematical models and making predictions using historical data or information. Currently, it is being used for recognition, speech-recognition, email filtering, Facebook-tagging, recommender system, and many more. I. Supervised Learning: Supervised learning is the type of machine learning in which machines are trained using well "labelled" training data, and on basis of that data, machines predict the output. The labelled data means some input data is already tagged with the correct output.

The aim of a supervised learning algorithm is to find a mapping function to map the input variable(x) with the output variable(y).

TYPES OF SUPERVISED LEARNING

1. Random Forest:

Random Forest is a popular type Supervised Machine Learning algorithm. It can be used as both Classification and Regression in ML. It is solely based on ensemble learning, which is a process of combining the multiple models to solve difficult problems.

Random Forest is a classifier that contains a number of decision trees on various subsets of the given dataset and takes the average to improve the predictive accuracy of that dataset.

Ex. Banking, Medicine, Marketing etc.

2. Naïve Bayes:

Naïve Bayes is also a supervised Machine Learning algorithm, which is based on bayes theorem and is used for solving classification problems.

It is mostly used in text classification which includes a high-dimensional training dataset. Naive Bayes classifier is one of the simple, most effective and probabilistic classification algorithms which predicts in the basis of probability of an object.

Ex: Spam Filtration, Sentiment analysis etc.

3. Decision Tree:

Decision Tree is a supervised learning technique that can be used for both classification and Regression problems, but mostly it is preferred for solving Classification problems. It is a tree-structured classifier, where internal nodes represent the features of a dataset, branches represent the decision rules and each leaf node represents the outcome.

It is a graphical representation for getting all the possible solutions to a problem/decision based on given conditions.

In a Decision tree, there are two nodes, which are the Decision Node and Leaf Node. Decision nodes are used to make any decision and have multiple branches, whereas Leaf nodes are the output of those decisions and do not contain any further branches.

4. Logistic Regression:

Logistic regression is one of the most popular Machine Learning algorithms, which comes under the Supervised Learning technique. It is used for predicting the categorical dependent variable using a given set of independent variables.

Logistic regression predicts the output of a categorical dependent variable. Therefore, the outcome must be a categorical or discrete value. It can be either Yes or No, 0 or 1, true or False, etc. but instead of giving the exact value as 0 and 1, it gives the probabilistic values which lie between 0 and 1.

Logistic Regression can be used to classify the observations using various types of data and can easily determine the most effective variables used for the classification.

5. Evaluation metrics:

Evaluation metrics explain the performance of a model. An important aspect of evaluation metrics is their capability to discriminate among model results.

4. PROPOSED SYSTEM

In the Major project Machine Learning and deep learning algorithm, libraries to perform prediction of the particular given stock. we have used two main libraries i.e. Keras and Tensor flow. The methodology of the major project is divided into five steps. They are as follows:

1. System Architecture

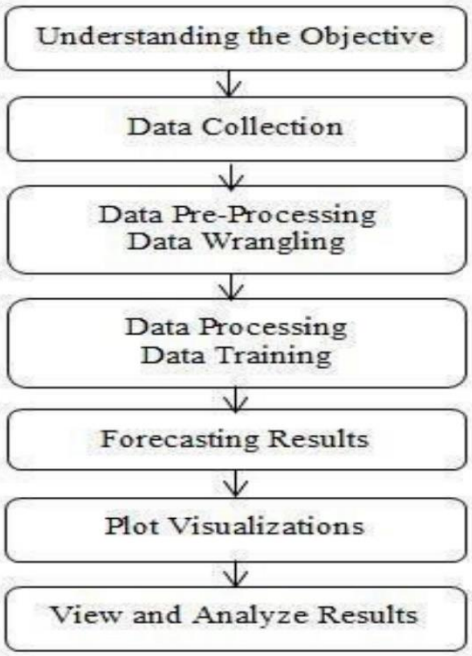

Fig.1: Architecture of Stock Market Prediction

The first step will be collecting the stock data e.g. AAPL i.e. Apple stocks. So we will be using many different types of library in our methodology such as pandas-data reader which is a remote data access to extract data from various internet sources into pandas data frame and for the reading the data, is a financial data platform that makes high quality financial tools available to all and it has a REST and Real Time Data API with this library it helps us to access the data, NumPy is used for working with arrays and it also has functions for working in domain of linear algebra, Fourier transform, and matrices, Matplotlib which is used to plot on map of stock prices, sklearn is the most useful and robust library for deep learning. For this program, LSTM are sensitive to the scale of the data so we apply Min Max scalar to convert the data to 0 and 1. The second step will be pre-processing the data so will train the data and then test it. By using the LSTM method, we will be splitting the dataset into train and test split. The data will be split into time steps so the test part we will take x_train and y_train and so the train part we will take y_train and y_test. After doing this, we have to reshape our x_train into 3 dimensions which is required for LSTM.

The third step will be to create a stacked LSTM model. A Stacked LSTM is an architecture that can be defined as multiple LSTM layers. LSTM gives sequential output rather than single value output. The fourth step is to predict the test data and plot the output. so first we will check the performance metrics and then transform back to original form. Then we will calculate RMSE (Root mean square error) performance metrics and test the data RMSE.

After doing the test data we will plot the output in that we have to shift the train & test predictions for plotting. Test data of predicted output shows that it is basically a green colour line output. Blue colour line output shows it is a complete dataset. Training dataset is how the prediction is gone shows that it is basically orange colour output. We have also used scaler inverse transform; it will transform the data such that its distribution will have a mean value 0 and standard deviation of 1.

Fig.2: Predicted Output for Stock Market Prediction

The fifth step is to predict the future 30 days and again plot the output. So basically, we will take 100 days data before to predict 101th data i.e. next to 30 days. Test data will be taken from 1 to 101 (i.e. 100 indexes) and predicting the data will be taken from 101 to 131 (i.e. 30 days). After doing that we will be getting a new 30 days.

5. RESULTS/ SCREENSHOTS

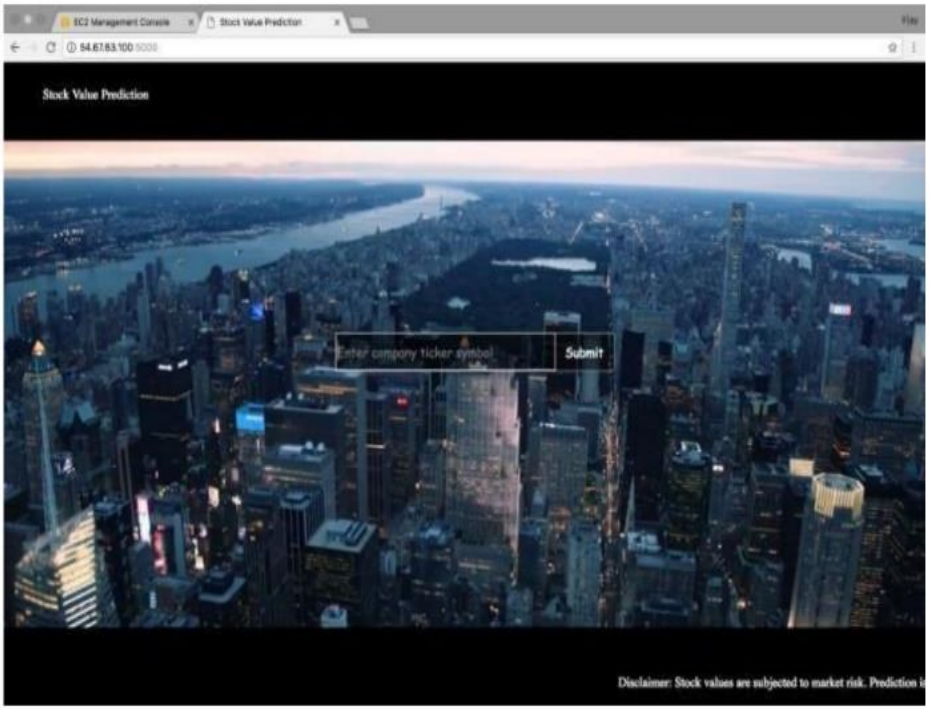

Fig.3: Home Page

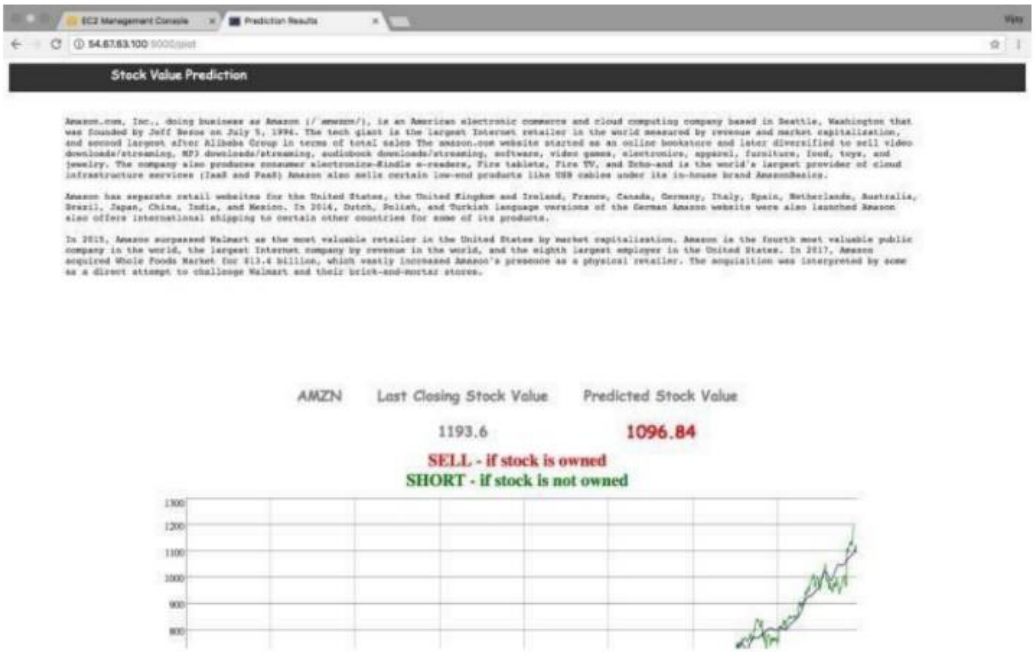

Fig.4: The Prediction with Company's Wiki Entry Gist

Fig.5: The Predicted Stock Price for tomorrow. The figure also shows the actual stock price against the predictions

Fig.6: Snippets of the top 10 news about the company

Fig.7: Backend Processing of Data in Prophet

6. CONCLUSION

The deep learning models by itself shows more accurate by using them in deep learning algorithm. We would recommend choosing Kaggle as the source rather than a forum since Kaggle is more factual and has a higher degree of accuracy of the datasets. Our future work would include researching more on the topic of LSTM model for forecasting as a source in conjunction with stock market data and highly specialized prediction algorithms the analysts use. Possibly try and implement the same using python at a small scale. to be thinking critically. This way we can help people make more informed decisions and they will not be fooled into thinking what others want to manipulate them into believing.

REFERENCES

[1] *Mustain Billah, Sajjad Waheed, Abu Hanifa, 2016. "Stock Market Prediction Using an Improved Training Algorithm of Neural Network", IEEE.*

[2] *Abidatul Izzah, Yuita Arum Sari, Ratna Widyastuti, Toga Aldila Cinderatama, 2016. "Mobile App for Stock Prediction Using Improved Multiple Linear Regression", IEEE.*

[3] *R. Yamini Nivetha, Dr. C. Dhaya, 2017. "Developing a Prediction Model for Stock Analysis", IEEE.*

[4] *N. Srinivasan, C.L akshmi, 2017. "Stock Prediction and Analysis Using Intermittent Training Data With ANN", ICIIECS.*

[5] *K. Hiba Sadia, Aditya Sharma, Adarsh Paul, SarmisthaPadhi, Saurav Sanyal, 2019. "Stock Market Prediction Using Machine Learning Algorithms", IJEAT.*

[6] *Aditya Menon, Shivali Singh, Hardik Parekh, 2019. "A Review of Stock Market Prediction Using Neural Networks", IEEE.*

[7] *Tejas Mankar, Tushar Hotchandani, Manish Madhwani, Akshay Chidrawar, Lifna C.S, 2019. "Stock Market Prediction based on Social Sentiments using Machine Learning", IEEE.*