Stock Market Prediction using ML & ANN on Big Data Models

1st Apurv Harkhani, 2nd Rithik Nair, 3rd Sushmit Palan, 4th Rohit Rane, 5th Prof. Priyanka Deshmane

1,2,3,4 Student, Computer Engineering, Alamuri Ratnamala Institute of Engineering & Technology, Maharashtra, India

5Professor, Computer Engineering, Alamuri Ratnamala Institute of Engineering & Technology, Maharashtra, India

Abstract - Stock market forecasting is one of the most difficult tasks in computing. False rumors, stock calculations, bad behavior, business people's opinions, business rumors, etc. There are many such predictions. We explore statistics as game changers on the field. According to the theory of market value, the consequences of the nature of this set are already marked in the stock price, when all the information about the business and the market economy is instantly available to all participants/employees. Therefore, it is said that the historical price can predict its future movement only under the influence of all market conditions. Therefore, we use machine learning (ML) techniques on historical price data to determine the future, taking past stock prices as the final assessment of all relevant issues. Machine learning techniques can find patterns and insights we haven't seen before that can be used to make predictions. We propose to analyze and predict the future growth of the company using the LSTM (Long Short Term Memory) model.
Keywords: Stock market prediction, Machine Learning, LSTM, Recurrent neural network

I. INTRODUCTION

Stock forecasting can be very profitable and a source of motivation for most researchers in the field. Most researchers use research papers or analytics to forecast the market. While analytics focuses on analyzing price trends to predict future prices, fundamental analysis relies on analyzing false data such as financial news and earnings data. Today, more and more valuable business information is being disseminated over the Internet. This paints a picture of the importance of word mining techniques in extracting valuable information to analyze business behavior. Although many articles review forecasting methods as analytical techniques, few focus on the use of paper mining. Unlike other recent reviews that focus on discussing different social media marketing strategies, this study aims to compare different learning patterns (ML) of historical data to find out which method is better for predicting which type of data. type and quantity. This study also reviews the detailed literature and future research for each review, highlighting the latest findings and future directions. Most researchers in the field rely on market forecasts, which are very useful. Researchers often run clinical trials or critical diagnoses. Fundamental analysis relies on the evaluation of intangible information such as financial news and earnings data, while analytical analysis focuses on analyzing the future value of cost-based estimates. Important business information is increasingly available online. This paints a picture of the need for data mining techniques to extract valuable information for studying business applications. While a few studies focus on the use of mining tools, many articles analyze predictive strategies based on analytical techniques.

II. LITERATURE SURVEY

Existing research in stock price prediction and time series prediction can be roughly divided into two groups: three groups based on operating strategy and the use of variables used to model the problem. Models using bivariate or multivariate regression on cross-sectional data are the first of most study categories [5]. Because of their simplicity and the error in the linearity assumptions they make, these models often do not perform very well. The second category predicts stock prices using time series elements and other economic methods such as vector autoregressive (VAR), autoregressive distribution lag (ARDL), autoregressive integrated moving average (ARIMA), Granger causality tests, and quantile Regression [1-3]. Proposals using machine learning, deep learning and natural language processing are in the third category of work [4-6]. Aside from the process of working using learned skills as the subject matter, one of the disadvantages of current stock forecasting theories is their inability to make accurate forecasts. Very good and fast pattern change movement. We aim to solve the problems in this study by taking advantage of models based on machine learning and deep learning to create a highly reliable, reliable and accurate framework for stock index forecasting. In particular, we propose a deep learning model based on the LSTM network and evaluate its performance in predicting future stock prices.

III. PROPOSED WORK

Stock market forecasting is a big topic and there are many areas in which we can work, but the purpose of each model is to check whether the model fits the data and the results and correct predictions. All models have some similar effects in that they need the names of listed companies to predict, hold, and sell the three conditions of the market, and provide each company's business data for comparison with the stock code stored in the machine. Complete data management until final preparation of data for other machine learning classifiers that will predict tokens and assign products. To show the cost of the overhead model, the current production of a particular firm will be shown and with the image, the model will be different. Now that you have the data, it's time to build a time series forecasting model that we will use the Bootstrap and Django framework. Build a neural network (RNN) and add multiple LSTM cells to the RNN so the model can learn from the data over time. Money does not match, and sometimes valuable information comes directly. Traditional time series methods, such as ARIMA and GARCH models, only work when the series is stationary, otherwise assumptions need to be made before the returns of the series (or some other changes). However, when this mode is used during operation, serious problems arise as security cannot be guaranteed as new information is added. This solution uses an unsecured network. Additionally, neural

networks can discover relationships between data and use them to predict (or classify) new data. One problem with LSTMs is that models are created with a parameter. Since the LSTM cannot be used to evaluate hypothesis testing.

Working of LSTM:
The LSTM algorithm uses three "gate" models and has a unique network topology. An LSTM cell has three gates: an input gate, a memory gate, and an output gate. The data will be selected by the code when it enters the LSTM network. Information that does not fit the algorithm will be deleted from the memory gate, only the information that fits the algorithm will be deleted. Historical data includes metrics such as open price, highest price, lowest price, closing price, transaction date and trading volume.
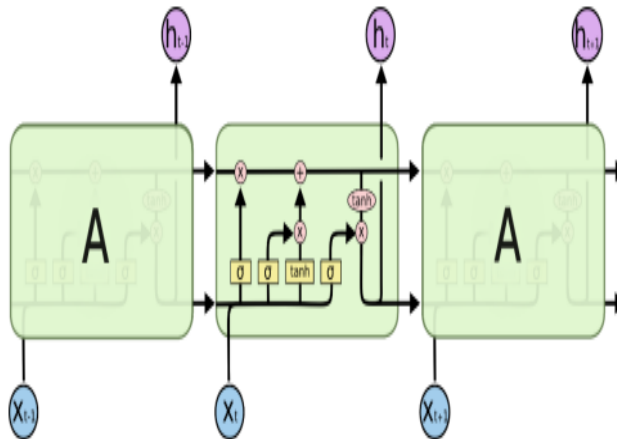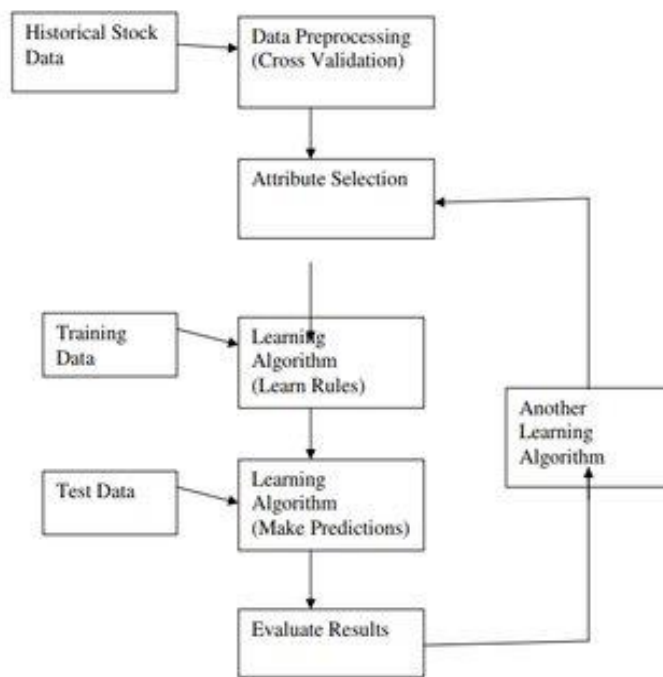

Fig1: LSTM Architecture

Fig 2: Prediction Using Historical Data

IV. METHODOLOGY

Forecasting the stock market can seem like a daunting task because there are so many unknowns. Although there are many models in machine learning, this article focuses on the most important models and uses them to make predictions.

A) Long Short Term Memory (LSTM)
An (advanced version) LSTM of a recurrent neural network (RNN) stores information from previous states. These differ from RNNs in that they take into account long-term dependencies, whereas RNNs try to make connections between new and current ideas.

B) Data Analysis for Machine Learning

Pandas: Pandas is an open source library designed primarily for speed and social reasons or for document annotations. It provides a set of data and methods for working with time series and numerical data. The NumPy library is the core of this library. Pandas is fast and provides users with great performance and productivity.

NumPy: NumPy, short for Numerical Number, is a Python library for calculating and manipulating the properties of multiple variables and linear arrays. It is a Python extension module mainly developed in C language. There are many ways to do numerical calculations. Many powerful data types are available in NumPy, including many arrays and matrices. The best calculations for arrays and matrices are done using this data structure.

Sckit Learn: Sckit-learn is mostly developed in Python and heavily utilises NumPy for high-performance array and linear algebra operations. To enhance performance, several fundamental algorithms are also written in Cython. Support vector machines, logistic regression, and linear support vector machines are implemented using wrappers built in Cython for LIBSVM and LIBLINEAR, respectively. It might not be viable to expand these functions with Python in such circumstances.

Bootstrap: An HTML, CSS, and JS package called Bootstrap makes it easy to build educational websites instead of complex online apps. The main reason it's included in a web project is to use Bootstrap's color, size, font and layout options for the project. Therefore, the important point to be considered is whether the manufacturers have a responsibility such as selection. When Bootstrap is integrated into a project, all HTML components have a simple manifest template. Therefore, text, messages and paper products are regularly displayed in all online browsers. Developers can use the CSS classes provided in Bootstrap to further customize the appearance of their content.Wormhole attack in wireless sensor network can disturb the routing process and ultimately degrade network performance. In this paper, we have presented existing wormhole attack types and their detection mechanism. Wormhole detection in a dynamic WSN setting is an open research area. A good research direction for wormhole detection is integration of trust based systems and time or distance bounding wormhole detection techniques.

Django: Unlike building complex online apps, creating educational websites is easier using an HTML, CSS, and JS package called Bootstrap. Using Bootstrap's color, size, and font and layout options for the project is a huge benefit of incorporating it into a web project. So the most important thing is whether the developers are responsible for supporting this possibility. When Bootstrap is used in a project, all HTML elements contain style declarations. As a result, text, words and content are displayed in all web browsers. Developers can take advantage of the CSS classes provided by Bootstrap to enhance the originality of their content.

V. CONCLUSION

In general, the stock forecasting task is very difficult and cannot be performed with high accuracy. But machine learning techniques can provide good market predictions that investors can benefit from. Also, there are many ways to find the most accurate stock price prediction model. The number of neural network nodes can be increased and we can get better results using neural networks. By analyzing several methods, we found that LSTM (Long Short Term Memory) provides the most accurate information about stock prices in real time. It is concluded that the stock market can be more predictable and efficient by using machine learning techniques. In the future, trade forecasting can be further improved using larger dataset than currently used. This will help improve the precision of our indicated models. Machine learning models can be checked to verify the accuracy they produce

REFERENCES

[1]Adebiyi, A., Adewumi, O., and Ayo, C. K. Ayo: Stock Price Prediction Using the ARIMA Model. In: Proc. of the Int. Conf. on Computer Modelling and Simulation, Cambridge, UK, pp. 105 - 111 (2014)

[2]Mondal, P., Shit, L., and Goswami, S.: Study of Effectiveness of Time Series Modeling (ARMA) in Forecasting Stock Prices. Int. Journal of Computer Science, Engineering and Applications, 4, 13 - 29 (2014)

[3]Jarrett, J. E. and Kyper, E.: ARIMA Modelling with Intervention to Forecast and Analyze Chinese Stock Prices. Int. Journal of Engineering Business Management, 3(3), 53 – 58 (2011)

[4]Mostafa, M.: Forecasting Stock Exchange Movements Using Neural Networks: Empirical Evidence from Kuwait. Expert Systems with Applications, 37, 6302 - 6309 (2010)

[5]Wu, Q., Chen, Y., and Liu, Z.: Ensemble Model of Intelligent Paradigms for Stock Market Forecasting. In: Proc. of the IEEE 1st Int. Workshop on Knowledge Discovery and Data Mining, Washington DC, USA, pp. 205 – 208 (2008)

[6]Siddiqui, T. A. and Abdullah, Y.: Developing a Nonlinear Model to Predict Stock Prices in India: An Artificial Neural Networks Approach. IUP Journal of Applied Finance, 21(3), 36– 39 (2015)

[7]Anurag Sinha Department of computer science, Student, Amity University Jharkhand Ranchi, Jharkhand (India), 834001 - Stock Market Prediction Using Machine Learning

[8]V Kranthi Sai Reddy Student, ECM, Sreenidhi Institute of Science and Technology, Hyderabad, India - Stock Market Prediction Using Machine Learning.

[9]Dharmaraja Selvamuthu, Vineet Kumar and Abhishek Mishra Department of Mathematics, Indian Institute of Technology Delhi, Hauz Khas, New Delhi 110016, India - Indian stock market prediction using artificial neural networks on tick data.

[10]Huicheng Liu Department of Electrical and Computer Engineering Queen's University, Canada - Leveraging Financial News for Stock Trend Prediction with Attention-Based Recurrent Neural Network.