STOCK MARKET PREDICTION AND ANALYSIS USING MACHINE LEARNING ALGORITHMS

Vasuki Rohilla[1], Snehal Gore[2], Akshara Garad[3], Sumedh Dhengre[4]

[4]Professor, Computer Dept. AISSMS College Of Engineering, Pune, Maharashtra, India
[1,2,3] AISSMS College Of Engineering, Pune, Maharashtra. India

---***---

Abstract - *The goal of Stock Market Prediction is to forecast the future worth of a company's financial stocks. The nature of the stock market movement has always been ambiguous for investors because of various influential factors. This research aims to use machine learning and deep learning algorithms to reduce the risk of trend prediction considerably.*

Machine learning is a recent trend in stock market prediction technologies that provide projections based on the values of current stock market indices by training on their prior values. To develop accurate predictions, machine learning employs a range of models. The research focuses on stock value prediction using Linear regression, LSTM-based machine learning, and other ML models*. There are several elements to examine, including open, close, low, high, and volume. The evaluation results will show that one of the models will outperform other prediction models for continuous data by a significant margin.*

Key Words: LSTM, Linear Regression, Stock Market Indices, Recurrent Neural Network, Stacked LSTM.

1. INTRODUCTION

Financial markets are extremely volatile, and they create enormous volumes of data on a regular basis. It is the most widely traded financial instrument, and its value fluctuates rapidly. Stock prices are forecasted to determine the worth of a company's stock or other financial instruments traded on stock exchanges in the future. The stock market allows investors to purchase shares of publicly traded corporations through exchange or over-the-counter trading. This market has provided investors with the opportunity to make money and live a prosperous life by investing small quantities of money at low risk compared to the risk of starting a new business or the necessity for a high-paying job. Many factors influence stock markets, resulting in market uncertainty and excessive volatility. Although humans can take orders and transmit them to the market, automated trading systems (ATS) run by computer programs can submit orders faster and more accurately than humans. However, implementing risk strategies and safety measures based on human judgments is essential to evaluate and control the performance of ATSs. When developing an ATS, many factors are taken into accounts, such as the trading strategy to be used, complex mathematical functions that reflect the state of a specific stock, machine learning algorithms that allow for the prediction of future stock value, and specific news about the stock being studied. However, numerous factors influence the stock market, including political events, economic conditions, and traders' expectations.

Researchers from a range of sectors, including computer science and business, are studying stock market projections. Researchers have experimented with a number of tactics and algorithms, as well as a mix of indications, to anticipate the market. The attribute that defines a prediction model is determined by factors that influence market performance.

Time-series prediction is a commonly utilized technique in many real-world applications, including weather forecasting and financial market forecasting. It predicts the result in the following time unit using continuous data over a period of time. In practice, many time series prediction algorithms have proven to be effective. Recurrent Neural Networks (RNN) and their special types, Long-short Term Memory (LSTM) and Gated Recurrent Units, are now the most often used algorithms (GRU).

This paper proposes to use LSTM, Linear Regression, and other ML models as ML tools for predicting the stock market prices for the next 30 days. Many people try to predict stock values, but it's a difficult task. Although perfect accuracy is unlikely, even simple linear models such as Linear Regression can be surprisingly close. Time series are used to represent stock trading data, and LSTM has the ability to learn extended observation sequences. This paper will also help us learn which machine learning algorithm will give the most accurate prediction. Long Short Term Memory (LSTM) networks are a type of recurrent neural network that can solve linear problems. A deep learning technique is LSTM. To learn very lengthy sequences, long-term memory (LSTM) units are required. The gated recurrent system in this form is more general. Because LSTMs address the evanescent gradient issue, they are more benign than other deep learning algorithms like RNN or classical feed-forward.

2. LITERATURE SURVEY

A blindfolded monkey throwing darts at a newspaper stock listing should do as well as any investment professional, according to Princeton University economist Burton Malkiel, who argues in his 1973 book, If the market is genuinely efficient, and a share price reflects all aspects as soon as they are made public, a blindfolded monkey tossing darts at a newspaper stock listing should do as well as any investing specialist, according to "A Random Walk Down Wall Street." However, let us not conclude that this is simply a stochastic or random process with few prospects for machine learning. Let's see if you can at least model the data so that the predictions you make correlate with the actual behavior of the data. In other words, you don't need the exact stock values of the future, but the stock price movements (that is if it is going to rise or fall in the near future).

Aistis Raudys proposed a negative weight moving average-based optimal stock price smoothing weighting technique. The moving average, on the other hand, has a lag. With the widespread usage of deep learning technology in recent years, many domestic and international researchers have begun to employ deep learning to perform stock prediction research.

Volodymyr Turchenko and colleagues proposed using the stock price of Fiat as the research object and an MLP neural network to predict stock price in the short term. The amount of parameters required to use the MLP neural network is, however, excessively enormous, and scalability is limited. As a result, Avraam Tsantekidis and others offered a CNN-based stock price forecast. CNN recognizes neurons' local connections and weight sharing while retaining significant properties and reducing a vast number of unnecessary ones. It outperforms the MLP model in terms of learning results. CNN, on the other hand, has some limits. CNN focuses on spatial mapping and offers some advantages when it comes to image processing. It isn't really appropriate for learning time series. The LSTM model is a variant of the RNN model that has three control units: the forgetting gate, the input gate, and the output gate.

The control unit in the model will make judgments on the information as it enters, leaving the conforming information and deleting the non-conforming information. LSTM can tackle the problem of lengthy sequence dependence in neural networks using this principle.

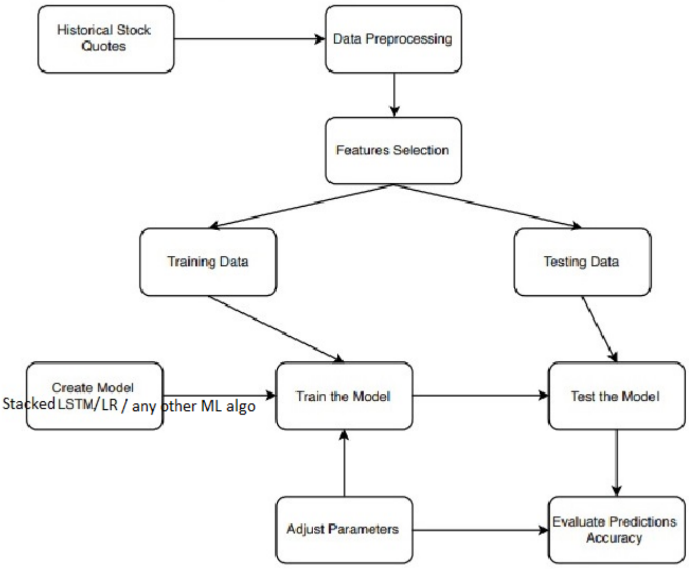
It is also mentioned that stock value prediction by methods for Multi-Source multiple instances learning unequivocally foreseeing the protections trade is a difficult task, but the web has turned out to be a useful tool in making this task less difficult, due to the related course of action of the data, it is certainly not difficult to evacuate certain inclinations right now, it is less difficult to establish associations between different variables and, for the time being, it is less difficult to establish associations between different variables and, for the time being. The use of some different options from specific legitimate data and the use of different strategies, such as the use of a feeling analyzer, to suggest a remarkable relationship between the emotions of individuals and how they are influenced by the enthusiasm for express stocks, is how budgetary trade information can be adequately predicted. One of the more notable aspects of the wanted approach was extracting major events from the news to analyze how they affected stock prices. Trade prediction protection is also mentioned: employing historical data analysis. The stock or offer expense can be predicted using historical data and examples, but counts are required to predict expenses. The traditional frameworks are only concerned with the type of element chosen for forecasting. The latter is commonly accomplished using Genetic Algorithms (GA) or Artificial Neural Networks (ANN), but they fail to build a relationship between their stock costs as long-distance transient dependencies.

RautSushrut et al. proposed that supervised movement be determined using financial index data. Portfolio modeling is a computational analytical approach used in the financial industry. A discussion of statistical AI technique has been addressed; the use of SVM methodology has been demonstrated in the study, as well as the application of tactical methodologies to anticipate stock values.

The most popular RNN design, according to M. Roondiwala et al, is the Long Short Term Memory. LSTM introduces a memory cell, a processing device that replaces traditional artificial neurons in the secret network layer. Networks may efficiently link memory and distant input in time using these memory cells, making them suited for dynamically capturing data structure across time with a high predictive limit. As mentioned in the article, stock predictions can also be made on NIFTY50 shares. Data collection is one of the most critical processes, followed by model training, and testing the strategy with various data sets is required.

3. PROPOSED SYSTEM

In brief, we shall first obtain historical data from the market. The data must then be extracted for data analysis, divided into testing and training data, and the algorithm must be trained to forecast the price. Finally, the data must be visualized.

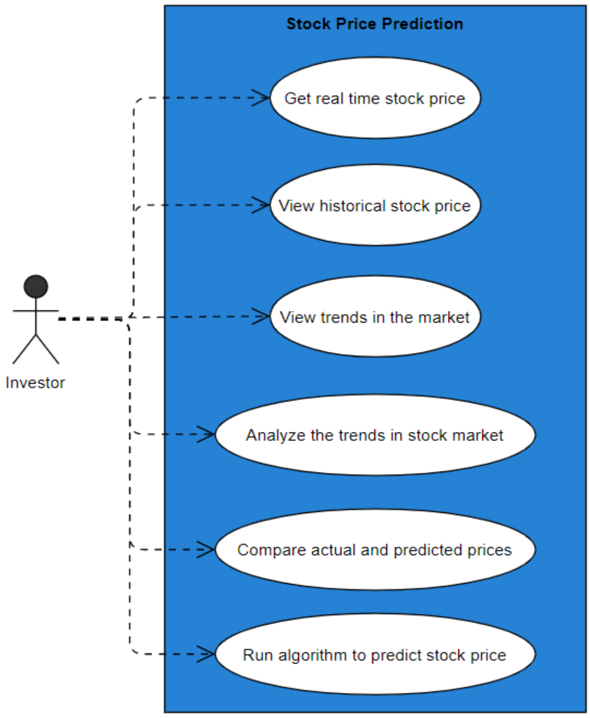


ARCHITECTURE DIAGRAM

First, we will be gathering information from the market. Tiingo API, Yahoo, and Google Finance are all good places to look for stock market data. These websites provide APIs through which stock datasets from multiple firms can be downloaded by simply inputting parameters.

Then data must then be preprocessed. The practice of preparing raw data for use in a machine learning model is known as data preprocessing. It's the first and most important stage in building a machine learning model. When working on a machine learning project, we don't always have access to clean and prepared data. Because it necessitates translating raw data into a fundamental configuration, this phase represents a substantial leap in information mining. The information collected from the source will be contradictory, fragmentary, and contain errors. The information will be purified during the preprocessing step, and then highlights scaling will be required to limit the factors. When the values of the features are closer together in machine learning algorithms, the algorithm has a better probability of being trained correctly and faster, however when the data points or feature values are far apart, it will take longer to grasp the data and the accuracy will be lower.

Such, if the data in any circumstance comprises data points that are far apart, scaling is a strategy for bringing them closer together, or, to put it another way, scaling is used to generalize data points so that the space between them is reduced. As we also know that LSTM and Linear Regression are sensitive to the scale of the data hence we will be using a min-max scaler for scaling the data.


Stock Price Prediction
Get real time stock price
View historical stock price
View trends in the market
Analyze the trends in stock market
Compare actual and predicted prices
Run algorithm to predict stock price
Investor

USE CASE DIAGRAM

The next step that we will be doing is splitting the data into testing and training datasets, creating the model as in the case of Stacked LSTM and we will be training the model according to it. The observations in the training set provide the learning experience for the algorithm. The test set is a collection of data used to assess the model's performance using a performance metric. It's critical that no observations from the training set make it into the test set. If the test set contains examples from the training set, it will be impossible to identify whether the algorithm has learned to generalize from it or has simply memorized it. A software that generalizes successfully will be able to complete tasks with fresh data effectively. A software that memorizes the training data by learning an overly complex model, on the other hand, could be able to reliably predict the values of the response variable for the training set, but not for fresh examples. Memorizing the training set is called over-fitting. A software that memorizes its observations may not perform effectively because it may memorize noise or coincidental relations and structures. Many machine learning algorithms struggle with balancing memorization and generalization, or over-fitting and under-fitting.

The next step for LSTM is to create a stacked LSTM model and train it. An LSTM model with numerous LSTM layers is known as a stacked LSTM architecture. An LSTM layer above sends a sequence of values to the LSTM layer below, rather than a single value. One output per input time step, as opposed to one output time step for all input time steps. A 3D input is required for each LSTM memory cell. Each memory cell in an LSTM outputs a single value for the whole sequence as a 2D array when it processes one input sequence of time steps. Hence we have to change the dimensions from 2D to 3D.

The next step will be to predict the test data and plot the output. But before that, we need to scale up our data to get accurate results and transform it back to its original form. Now we will plot the graph and check its performance matrices. Every machine learning pipeline incorporates performance metrics. They inform you how far you've come and give you a score. So, we calculate the RMSE. The square root of the average of the squared difference between the target value and the value predicted by the regression model is the Root Mean Squared Error(RMSE).

The next step is to predict the future data for 30 days and plot the output and calculate the performance metrics. We will then compare the loss and accuracy for both Linear Regression, Stacked LSTM, and other ML algorithms used and display the result for which is the most accurate.

4. CONCLUSIONS

In this paper, a survey on various methods to predict stock prices has been done. A comparison study of stock price time series prediction based on the Stacked LSTM, Linear Regression, and other ML models is proposed. We will compare the accuracy of the results of Linear Regression and Stacked LSTM and other ML models and find out which should be preferred and why. We have also stated the step-by-step process for prediction for the next 30 days.

REFERENCES

[1] Yulian Wen, Peiguang Lin and Xiushan Nie at 2020 IOP Conf. Ser.: Mater. Sci. Eng "Research of Stock Price Prediction Based on PCA-LSTM Model"

[2] I. Parmar et al., "Stock Market Prediction Using Machine Learning," 2018 First International Conference on Secure Cyber Computing and Communication (ICSCCC), 2018

[3] "Stock Price Prediction Using Long Short Term Memory" by Raghav Nandakumar1, Uttamraj K R, Vishal R, Y V Lokeswari at IRJET, Volume: 05 Issue: 03, Mar-2018

[4] Pascanu, Razvan, Tomas Mikolov, and Yoshua Bengio. "On the difficulty of training recurrent neural networks." International Conference on Machine Learning. 2013.

[5] Loke. K.S. "Impact Of Financial Ratios And Technical Analysis On Stock Price Prediction Using Random Forests", IEEE, 2017.

[6] Stock Price Prediction Using LSTM by Pramod BS, Mallikarjuna Shastry P. M. at research gate

[7] M. Nabipour, P. Nayyeri, H. Jabani, S. S. and A. Mosavi, "Predicting Stock Market Trends Using Machine Learning and Deep Learning Algorithms Via Continuous and Binary Data; a Comparative Analysis," in *IEEE Access*, vol. 8

[8] X. Yuan, J. Yuan, T. Jiang, and Q. U. Ain, "Integrated Long-Term Stock Selection Models Based on Feature Selection and Machine Learning Algorithms for China Stock Market," in *IEEE Access*, vol. 8,2022

[9] "Optimal negative weight moving average for stock price series smoothing", Aistis Raudys, CIFEr, London, UK: IEEE, 2014: 239-246.

[10] X. Shao, D. Ma, Y. Liu, Q. Yin. "Short-term forecast of the stock price of multi-branch LSTM based on K-means", 2017 4th International Conference on Systems and Informatics (ICSAI).

[11] Loke. K.S. "Impact Of Financial Ratios And Technical Analysis On Stock Price Prediction Using Random Forests", IEEE, 2017.

[12] S. Selvin, R. Vinaya Kumar, 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.

[13] Eapen J, Bein D, Verma "A. Novel deep learning model with CNN and bi-directional LSTM for improved stock market index prediction." 2019 IEEE

[14] "Stock Price Correlation Coefficient Prediction with ARIMA-LSTM Hybrid Model" Hyeong Kyu Choi, B.A Student Dept. of Business Administration, Korea University

BIOGRAPHIES

Vasuki Rohilla
Computer Engineering Final Year Student
AISSMS COLLEGE OF ENGINEERING

Snehal Gore
Computer Engineering Final Year Student
AISSMS COLLEGE OF ENGINEERING

Akshara Garad
Computer Engineering Final Year Student
AISSMS COLLEGE OF ENGINEERING