STOCK MARKET PREDICTION USING STREAMLIT AND LSTM

Nithyasri Narasimhan
Student, Dept of CSE, SRM Institute of Science and Technology, Vadapalani, Chennai, India.

Mr.M.Rajavel,B.Tech.,M.Tech.
Assistant Professor
Dept.of Computer Science & Engg
SRMIST, Vadapalani Campus

ABSTRACT: One of the most significant activities in the financial sector is stock trading. Stock price prediction is the process of trying to forecast the price of a specific stock or share using the market's available data[1]. In this study, we created a machine learning model and LSTM to forecast the price and the training model we built operates on the 70:30 principle, where 70% of the data about a company is provided to the training section and 30% to the prediction part so the model improves over time[1]. In order to estimate the price trend of stock markets, we gathered data from Yahoo spanning 10 years and fully customized feature engineering and deep learning-based models. The technique is generally accurate at predicting stock market trends. Through careful design and evaluation of prediction term lengths, feature engineering, and data pre-processing techniques, this work enhances the field of stock analysis research in both the financial and technical spheres.

IndexTerms: Trade High, Trade Close, CNN, Machine Learning, Deep Learning, LSTM, etc.

I. INTRODUCTION

The financial market, sometimes known as the stock market, is a sophisticated, composite mechanism that enables people all over the world to buy, sell, and exchange currencies, stocks, bonds, and tax credits.[2] Through exchange or over-the-counter trading, the stock market enables investors to own shares of publicly traded companies. With this business strategy, investors can earn a living while taking little risk with their initial investments, starting new businesses, or earning large salaries.

In an effort to pinpoint the drawbacks of current strategies like ANNs (Artificial Neural Networks) and SVMs, this study analyzes the stock market forecasting methodologies now in use (Support vector machine) [4]. ANN and SVM are supervised machine learning models that can categorize new data when given new data, respectively. ANN is built on a collection of artificial neurons. While SVM is focused on data categorization, the main drawback of the model ANN was dealing with over-fitting data managing several parameters at once. This resulted in low forecasting accuracy.

The LSTM model, a subtype of RNN (Recurrent Neural Network), is used in this study because it has the advantages of learning long-term dependencies in the data and forecasting by memorizing the data[2]. The LSTM is used because it can handle huge epochs and produce output with less volatility and greater reliability. Using yahoo finance, the information was collected[1].
With data pulled from Yahoo and an application built on Streamlit, the web application shows stock visualization for a number of service and product-based businesses.

II. METHODOLOGY

We briefly discuss the RNN and LSTM design to aid understanding of the method in this study. Even while RNNs have been shown to be successful in succession prediction tasks, learning about long-term dependency can be challenging, mostly due to the vanishing gradient problem that results from the gradient propagation of the repeated network over several layers. Based on these elements, the approach fixes the visual distortion by multiplying an image of a different color by various weighting coefficients, and by embedding the watermark, various wavelet coefficients increase the watermark's resilience.

2.1 Long short-term memory network

Long short-term memory network [6] [10] (LSTM) is a type of recurrent neural network [2]. The gradient sign in a standard recurrent neural network may be increased countless times (as many times as the number of epochs) during the gradient back-spread stage thanks to the network weights associated with the connections between the neurons of the hidden layer of the recurrent neural network. This suggests that the learning cycle may be significantly impacted by the size of the loads in the change framework. The gradient signal sometimes becomes so insignificant that adaptation either becomes very moderate or stops working out and out if the loads or weights in the network are small [3] (or, more precisely, if the primary eigenvalue of the weight grid is less than 1.0). This phenomenon is known as vanishing gradients.

2.2 Working of LSTM

The special network structure known as LSTM comprises of three "gate" components [2]. The input gate, forgetting gate, and the output gate are the three gates that form the LSTM unit. When data enters the LSTM neural network through the input gate, it may get selected by rules. The forgetting gate will delete any data that does not follow the algorithm, leaving just the data that does. utilised the neural network's embedded layer of LSTM and the automated encoder (long short-term memory) [2]. To prevent gradients from exploding and vanishing, LSTM is employed in place of RNN [7]. Python is used in this project to train the model and MATLAB to minimize the input's dimensions. Data is stored and retrieved using MySQL as a dataset. The table of historical stock data includes the opening price, highest price, lowest price, closing price, transaction date, volume, and other details. This LSTM model's accuracy for this project is 89%. The historical data that were collected from the Internet and used as experimental data in this study. Three data sets were used in the studies. Finding an optimization method that consumes less resources and converges more rapidly is required. An input entryway, a neuron with a self-recurrent connection, an ignored entryway, and an output entryway are the four main parts of a memory cell [8]. With a load of 1, the self-recurrent association ensures that, barring any external impedance, a memory cell's status can remain persistent from one step to the next. The entryways have the purpose of adjusting the communication between the memory cell and its current environment. The input entrance has the ability to either allow or disallow the approaching sign from changing the memory cell's status.
III. SYSTEM ARCHITECTURE

Yahoo Finance is used as the input source for the data [3]. The information is made up of stock information for a number of product- and service-based businesses, which displays the market closing rates for each of their individual stocks. For preprocessing, the data is divided into two unique sections: training data and testing data. The machine learning model's input data will be the bifurcated data[3]. The data is utilised to build an LSTM model, which will be used to forecast stock value. The model's correctness is then verified by a number of rounds to make sure the strategy used resolves the issues that were in the prior models. Before the needed accuracy is attained in reliable numbers, the procedure proceeds through multiple iterations. Once the model has received enough training, streamlit is used to build a web application that gives the user the chance to use the model and forecast the data of their invested stocks[3]. The user can choose between companies that offer products and services using the radio buttons provided by the web app. Additionally, a forecast for the next 30 days at most can be obtained. A slider is offered for this purpose, allowing the prediction to be changed from one day to a whole month. With the help of this specific feature, the investor would be able to decide whether to keep the stock in the market, purchase additional shares of related companies, or sell it in order to cut losses. Utilizing Google Colab and Streamlit, the Stock Market Prediction Using LSTM (Web Application) is effectively implemented[3].
IV. MODULE IDENTIFICATION

4.1 Training Data

This prototype was made to function in a methodical frame style. In the conclusion of every trading day, a new neural network was created, which meant that a new set of weights was determined using a new batch of training and corroboration data. The data from trade that occurred before the current day is used for training, and the performance of the model is confirmed using data from the previous week. The most recent model will be used to make all forecasts the following day. The preferred supervised learning method was chosen since it works with time series: the LSTM neural network (Long Short-Term Memory), the current neural network capable of categorising input data taking into consideration the previous instances. The model's LSTM input layer, which will receive input from both technical indicators and price data, and feed an output layer using sigmoid activation, was built using Google's TensorFlow software. Ten years of data, from 2011 to 2021, are used to train the LSTM, based upon the suggested method will forecast the first month of the following year based on associations found or recorded by the LSTM of Google stock price. The packages required for data processing are numpy, matplotlib, pandas, keras, and tensor flow. In fact, the approach is dealing with a regressor because it is attempting to forecast a continuous variable outcome.

4.2 Data Pre-processing and Implementation

Prior to employing the LSTM to calculate the stock price, we must preprocess this data. Apply the fit transform function to the values in our data to transform them. The data is scaled using the min-max scaler so that all of the price values are on the same scale. Then, we divide the data into two groups, using 70% for training and 30% for testing, and we assign each group to a different variable. The sequence for training and testing is created using a function. The LSTM model is created in the following phase. The Sequential model from Keras is imported along with the necessary libraries for this study. In our model, we employ drop out for regularization between the two LSTM layers. With a 10% dropout, fifty units are assigned in the LSTM parameter. The loss function used by Adam Optimizer to optimize the issue is mean squared error [7]. Our LSTM network uses mean absolute error as the metric because it is related to time-series data.
V. RESULTS AND DISCUSSIONS

The length of the data and the number of epochs both significantly influenced the testing outcomes after our training data generated a range of outcomes [4]. For instance, the outcomes would be as follows if we adjusted the dataset for the model. Since it was previously said that the data is divided into training data and data used for prediction, it consists of the close market price and the test data. The LSTM model is trained using test data collected over a period of 100 days, ensuring that the model's predictions continue to be accurate. A results study shows that as the dataset size increases, both models perform more accurately[3]. With additional data and by using gradient descent, the model can produce more patterns and its layer weights can be adjusted more precisely [6] [9]. The model's accuracy in forecasting the close market rate of any given stock is demonstrated by the fact that there is a very small difference between the projected data and the actual data[3]. Since the data is so big, it is normalised in order to speed up training while keeping the model's accuracy. The green and orange lines indicate the outcomes of the data visualisation, and the blue lines highlight the test data.
VI. CONCLUSION

Emerging Machine Learning techniques and its dominant algorithms have penetrated the latest developments in market research and stock market forecasting. We therefore estimate any designated firm’s stock closing price in this paper using a web-based application that we constructed using the LSTM algorithm [1]. There are drawbacks to using neural networks that are made up of LSTM (Long-Short Term Memory Models), such as the fact that RNN is still highly challenging to train even after adding LSTM (long/short memory). The primary cause of the adoption of time embedding transformer architecture to predict Stocks was the model's inability to choose which data to preserve or discard when the input stream grew larger [3].

VII. REFERENCES

[2] STOCK PRICE PREDICTION submitted by Somaraju Dinesh, Adduri Maruthi Siva Rama Raju, Sasumana Rahul, Oruganti Naga Sandeep