Section A-Research paper
ISSN 2063-5346

A NOVEL STUDY ON REAL-TIME PRICE PREDICTIONS STOCK FORECASTING METHOD

R. Sreedhar
Associate Professor, Department of IT
Sridevi Women's Engineering College, Telangana
rachasreedharswec@gmail.com

Beeram Indhu
BTech Student, Department of IT
Sridevi Women's Engineering College, Telangana
beeramindhu@gmail.com

Chilukuri Ananthakala
BTech Student, Department of IT
Sridevi Women's Engineering College, Telangana
chilukuriananthakala@gmail.com

Nallamothu Bhavana Chowdary
BTech Student, Department of IT
Sridevi Women's Engineering College, Telangana
bhavanachowdary39779@gmail.com

Pulipati Kavitha
BTech Student, Department of IT
Sridevi Women's Engineering College, Telangana
pulipati.kavitha@gmail.com

Rayini Ramya Sree
BTech Student, Department of IT
Sridevi Women's Engineering College, Telangana
rayini.ramya@gmail.com

ABSTRACT

Intraday trading is very common among traders because of its capacity to capitalize on price movements within a relatively short period of time. When it comes to formulating trading strategies, having access to real-time price forecasts for the subsequent few minutes can be of great use to traders. The nature of the stock market is such that it is non-stationary, complicated, noisy, chaotic, dynamic, volatile, and non-parametric. This makes it difficult to make accurate predictions in real time. Even if machine learning models are thought to be useful for stock forecasting, their hyperparameters still need to be tuned using the most recent market data in order to take into account the complexity of the market. In most cases, models are trained and tested in batches, which helps to streamline the process of error correction and accelerate the learning process. In order to ensure a high level of accuracy while making intraday stock forecasts, the models should train simultaneously and forecast for each individual instance as opposed to the entire batch. In this research, we offer a technique to estimate the stock price using the real-time stream of the live market. The strategy is based on two distinct learning approaches: incremental learning and offline–online learning. Both of these learning approaches have their advantages and disadvantages. In incremental learning, the model is retrained after each trading session to ensure that it takes into account the most recent data complexities. On the other hand, in offline-online learning, the model is retrained after each trading session to ensure that it takes into account the most recent data complexities. These techniques were utilized to analyze univariate time series, which were constructed using past stock prices, as well as multivariate time series, which took into account historical stock prices in addition to technical indications. Extensive tests were run on the eight most liquid equities that are listed on the NASDAQ stock exchange in the United States of America and the NSE stock exchange in India, respectively.

Section A-Research paper
ISSN 2063-5346

1 INTRODUCTION

The trends in stock prices are time series that are nonlinear and unstable. Investors have consistently monitored and made predictions regarding stock prices during the past 30 years [15, 25, 44] in order to increase their chances of making a profit in the stock market. In order to forecast future movements in the stock market, academics have utilized a wide variety of transaction data and have produced technical indicators [36, 48]. For the purpose of constructing time series forecasts [41, 49], statistical, economic, and other methods have been applied. For the purpose of researching variations in the stock market [16, 17], factor pricing models have been utilized. For instance, Jegadeesh and Titman [23] stated that a stock price has a tendency toward continuing in the original movement direction, and that the volume and turnover rate are both momentum elements that can be determined for the purpose of predicting a stock price's trend. Fama and French created a factor pricing model [18] to assess changes in the expected rate of return on equity in a cross section by making use of derivative indices like the total market value and the book-to-market ratio. However, as study into financial behavior has become more in-depth, researchers have discovered that people's financial actions in the market are illogical. This is because investors are affected by a variety of factors, including cultural, psychological, and other elements. The natural languages that have emerged, such as news events [2, 8, 46, 51], social media [39, 45], and stock bars [24], have evolved into the primary indicators in this area of research. In addition, researchers are consistently developing novel text embedding methods and introducing new machine learning algorithms for stock market research [12, 13]. These include convolutional neural networks (CNNs) and recurrent neural networks (RNNs).

Because of advancements in information technology, graphical indicators have emerged as significant quantitative evaluation indicators of the stock market. These graphical indicators, which include time-sharing diagrams, average lines, and K-lines, are used to analyze stock price trends in order to obtain more intuitive stock price change trends. Traders are able to recognize existing trends as well as future trends with the assistance of moving averages, which were first proposed by Joseph E. Granville [20]. Moving averages can help traders spot excessive delays in trends that are poised to reverse. K-charts are a visual representation that conveys an intuitive understanding of the movements of stock values through the use of patterns, colors, forms, and other components. Therefore, K-line charts [6] have been the approaches that have received the greatest amount of attention as an important tool for assisting investors in their decision-making. K-charts have been used in a significant number of research projects by a significant number of researchers. These researchers have mostly utilized the time series similarities across K-lines [28] and found patterns [50] in order to forecast the movements of stock values. Indicators that are represented by graphs are garnering an increasing amount of attention as an essential component. Research is also centered on determining how graphical indicators can be integrated with more traditional data-based indicators.

In the recent years, a significant amount of attention has been paid by researchers to the processing, integration, and analysis of disparate types of data in relation to the stock market. For instance, Tab et al. [47] recommended utilizing a tensor to substitute a series vector for the purpose of realizing the integration of data information to model market information. They also employed an event-driven mechanism for the purpose of balancing the heterogeneity of various forms of data used for data prediction. The appropriate features were extracted from preprocessed multi-source data by Chai [7], who then utilized an extended hidden Markov model to model the quantified features. This was done in order to capture the underlying time dependency in the data, which was then applied to the task of financial time series prediction. There have been many studies on the fusion and processing of heterogeneous data in the stock market [26, 31, 52, 53], and mainstream methodologies have been applied in those studies; however, there are still some limitations.

2 LITERATURE REVIEW

A prediction of the future movement of the value of a stock on a financial exchange is what is meant by the term "stock market forecast." Investors have a tremendous deal of potential for profit from the accurate forecasting of share prices, and being able to properly predict the movement of share prices within a short period of time can result in substantial earnings. There have been a number of different approaches suggested for forecasting the market and offering direction for decision-making. Instead of being a stochastic process, the movement of stock prices can be interpreted as a discrete time series that is founded on clearly defined numbers that are gathered at predetermined intervals of time. It is necessary for the time-series data to be steady in order to construct the forecasting model. The differencing method can be used to get stationary data from a non-stationary time series. This can be accomplished by applying the method. On the other hand, the differencing technique will exclude any information in the time series that relates to an existing trend. In this particular field, a variety of approaches, such as statistical methods and models based on machine learning, may be utilized. In most cases, statistical models begin with the premise that there is a linear correlation structure present among the values of the time series. According to Alves et al. (2018), the nature of the time series of the stock market is non-linear, volatile, chaotic, and very noisy. Traditional statistical methods include the autoregressive method (AR), the moving average model (MA), the combination of AR and MA, sometimes known as the autoregressive moving average model (ARMA), and the autoregressive integrated moving average (ARIMA). The statistical properties of the ARIMA model, in addition to the well-known Box-Jenkins model-building process, are largely responsible for the model's widespread use. However, ARIMA models are unable to capture nonlinear patterns, and it is not always possible to replicate complicated real-life problems using linear models (Zhang 2003). The researchers came up with the Grader causality test, which extends the investigation into a multivariate time-series analysis rather than a univariate one. A multivariate time-series forecasting model was developed by using the vector autoregressive moving average (VARMA). This model can represent Vector Moving Average (VMA) and Vector Autoregressive (VAR) models in a flexible manner (Liu et al. 2021). Pellegrini et al. (2011) apply the ARIMA-GARCH model to the process of forecasting a financial series. This model takes into account conditional variances and is based on the generalized autoregressive conditional heteroscedastic (GARCH) model. Given that the ARIMA-GARCH models never converge to homoscedastic intervals, the prediction intervals produced by these models may not be sufficient.

The traditional time-series forecasting methods are able to capture linear correlations and produce satisfactory results when applied to a small dataset. However, the performance of these algorithms is not particularly good when used to time-series that are both huge and complex, such as time-series pertaining to the stock market (Liu et al. 2021). As a direct

Section A-Research paper
ISSN 2063-5346

consequence of this, researchers in this field have begun to place an increased emphasis on machine learning and deep learning methodologies. In their research on forecasting stock values, Javed Awan et al. (2021) made use of machine learning algorithms and sentiment analysis. According to the findings, the linear regression, the extended linear regression, and the random forest produce more accurate findings than the decision tree does. Cao and Tay's study from 2001, Kim's study from 2003, and Maguluri and Ragupathy's study from 2020 are three examples of research that employed linear and non-linear support vector machines (SVMs) to anticipate financial time series. On the other hand, these models suffer from the issue of overfitting, and the algorithms are not very good at making predictions using vast datasets. According to Behera and colleagues' (2020) research, support vector regression has a higher level of accuracy when compared to other models. An end-to-end framework was developed by Tuarob et al. (2021), and it consists of three different sub-models: Davis-C, Davis-A, and Davis-V. Davis-C is responsible for the collecting of data linked to stocks in real-time; Davis-A is responsible for analysis; and Davis-V is responsible for visualization. Their methodology indicates that a mix of machine learning algorithms performs significantly better than a single machine learning algorithm when compared to other solo machine learning algorithms. Vijh et al. (2020) constructed two models: the first one predicts the price trends for the next day based on historical data, while the second one predicts the price trends for the following month based on historical data. Both of these models use past data. In order to forecast the trend based on volume volatility, sentiment, and continuous up/down movement, they utilized Logistic Regression, Support Vector Machines, and Boosted Decision Trees.

For anticipating movements in the stock market over the past few years, the use of deep learning algorithms has grown increasingly prominent. According to Kumar et al. 2021, these methods are able to extract meaningful characteristics from complicated and inconsistent data without relying on the expertise of humans and can even identify underlying nonlinearities. Deep learning has been employed by a number of industry professionals to improve stock forecasts and generate money for shareholders. Deep learning techniques such as artificial neural networks (ANN), convolutional neural networks (CNN), long-short-term memory (LSTM), hybrid algorithms, and others contribute to superior results in financial time-series forecasting compared to statistical and machine learning techniques. Vijh et al. (2020) investigated the ANN and Random Forest on multivariate time-series on five stocks to forecast the next day's closing price utilizing features such as the previous day's open price, closing price, Moving Average, Highs, and Lows. This was done in order to predict the price at which the market will close the next day. A stock forecasting approach that is a mix of CNN and LSTM was proposed by Lu et al. (2020).

3. METHODOLOGY

In this work, time-series forecasting models are investigated for their potential to accurately predict stock prices by making use of high-frequency data collected at 15-minute intervals. The methodology that has been proposed makes use of two distinct learning strategies, namely learning in increments and learning both offline and online. Time series that are either univariate or multivariate can be analyzed using these methods. The stock prices were used to construct the univariate time series, while exponential moving averages (EMAs) and volume-weighted average prices (VWAPs) were added to the stock prices to create the multivariate time series. The univariate time series was then used to create the multivariate time series. The

ISSN 2063-5346

incremental model is continually updated whenever it obtains fresh instances of the stock price from the live feed of the stock market. This ensures that the model accurately reflects current market conditions. On the other hand, the Offline–Online learning model requires retraining following each trading session it participates in. The model needs to be retrained so that it can adjust to the current market patterns, as well as the volatility and seasonality of the market. A graphical illustration of the process is displayed in Figure 2.

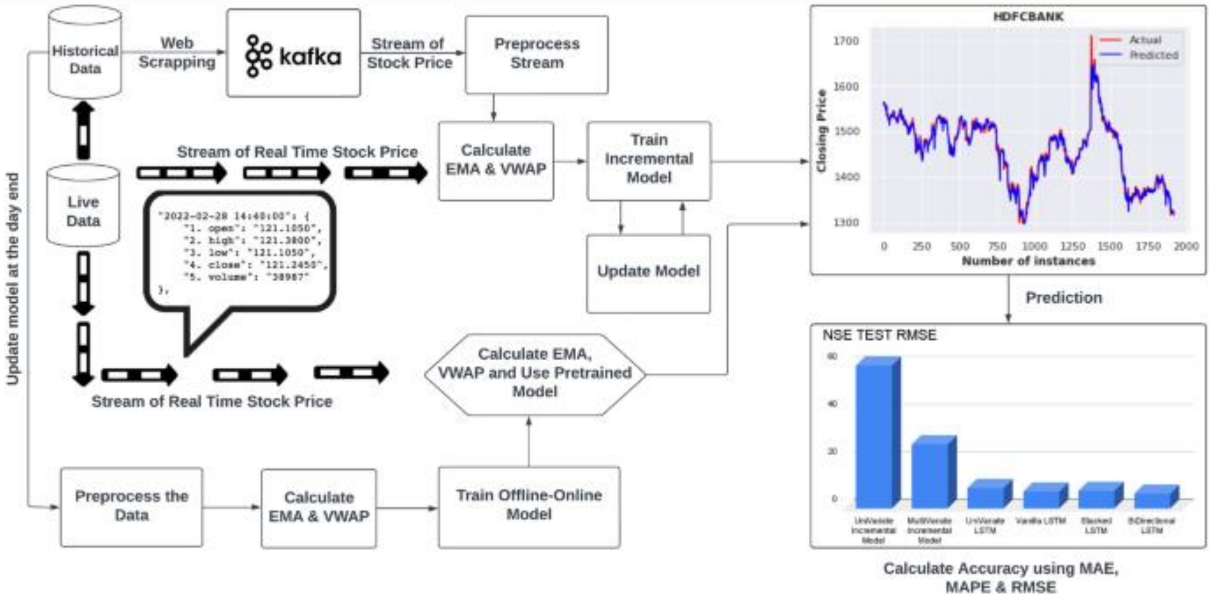

Fig 1: Proposed model

Technical indicators

Traders apply technical charts, which involve the analysis of price actions as well as technical indicators, in order to improve their accuracy in stock price prediction. Intraday traders employ a variety of technical indicators, such as MACD and RSI, to help them decide whether or not it is the right time to buy or sell a certain stock. However, EMA(d) and VWAP are both viable indicators to use when attempting to catch current trends. The exponential moving average (EMA) is calculated by taking the average price of a stock over the most recent d data points and weighting it exponentially. This gives more weight to the most recent data points' prices. Because EMAs are so focused on more recent price activity, they have a greater propensity to react more swiftly to changes in price.

Construction of graph data

The primary focus of this piece of writing is research on the process of constructing subgraph data. The researchers choose trading data, news about the stock market, and graphical indicators to use in the construction of subgraph data, and the combined results of these three forms of subgraph data are used to generate graph data. Five different trading days were chosen to serve as nodes in the construction of this subgraph, as can be seen in the trade indicator subgraph presented in Fig. 2(a). Six different indices are included as properties of the node: the starting index, the highest index, the lowest index, the closing index, the trading volume, and the trading value. Edges are placed between the nodes of neighboring trading days in accordance with the continuity of the trading days that are close

ISSN 2063-5346

together. A random integer in the range [0,1] [27] is used as the initial weight of the edge. The node feature matrix for the transaction data subgraph is a 5*6 matrix (5 trading day nodes, where each node has 6 features [1]), and the edge weight matrix for the transaction data subgraph is a 5*2 matrix (the number of edges is established between trading days c25 52). The opening index, maximum index, minimum index, closing index, trading volume, and trading quantity for each of the six indicators make up the node's six characteristics, which are referred to as the six node characteristics. The stock market news of each trading day is taken as an indicator subgraph, each news item is taken as a node of the corresponding subgraph, and the related news text word vector is taken as the node feature in Fig. 2(b), which depicts the stock market news subgraph. The similarity of the news texts is used to determine the weight of the edges that are put between the news texts in order to construct the link that exists between the news texts.

Data set introduction

In order to carry out the tests, this study makes use of the China Stock Market & Accounting Research (CSMAR) database. The stock market chart data for the Shanghai Composite Index, China Securities Index (CSI) 300, and Shenzhen Composite Index are constructed with the help of this database's data table, which spans the time period from 2013-01-11 to 2019-11-25 and contains the information. Dgl is utilized in the process of constructing graph data and carrying out the convolution operation that corresponds to it. In order to establish the connections between different news texts, the news index subgraph treats each individual piece of news as a node in the graph. The word vector of the news text serves as the associated node feature, and the degree of similarity between each piece of news serves as the edge weight. Each day that a market is open constitutes a subgraph. In order to convert the news texts into 200-dimensional word vectors for use as node features, the genism library is utilized. The weights of the edges in the news subgraph are determined by a similarity model based on the similarities that exist between the various news texts.

CONCLUSION AND FUTURE WORK

This study investigates incremental and offline-online learning strategies for the purpose of forecasting stock prices on the NASDAQ and NSE. The models that were used for this study were trained on the most recent stock data while the stock's time-series was continuously updated from the live market feed. This allowed the models to fine-tune their hyperparameters based on the changes that occurred in the stock's time-series during the trading sessions. The results of this study were published in the journal Applied Financial Economics. We chose the exponential moving average (EMA) and the volume weighted average price (VWAP), in addition to the stock price, as characteristics to take into consideration when developing an efficient multivariate time-series dataset. This decision was reached after conducting an in-depth investigation of numerous technical indicators that contribute to improved price prediction. In addition, the system was analyzed using RMSE, MAE, and MAPE to determine how well it performed when applied to the top 8 companies listed on the NSE and NASDAQ, respectively. The effectiveness of the EMA and VWAP in forecasting stock prices is demonstrated by the fact that all of the models improved their predictions when applied to multivariate time series.

REFERENCES

1. Arasu A, Widom J. Resource sharing in continuous sliding-window aggregates[EB/OL]. [2019-10-02]. https://www.microsoft.com/en-us/research/wp-content/uploads/2016/02/sharing.pdf

Section A-Research paper
ISSN 2063-5346

2. Atkins A, Niranjan M, Gerding E (2018) Financial news predicts stock market volatility better than close price[J]. J Financ Data Sci 4(2):120–137

3. Belov G, Scheithauer G (2006) A branch-and-cut-and-price algorithm for one-dimensional stock cutting and two-dimensional two-stage cutting[J]. Eur J Oper Res 171(1):85–106

4. Box GEP, Jenkins GM, Reinsel GC et al (2015) Time series analysis: forecasting and control[M]. Wiley, Hoboken

5. Bruna J, Zaremba W, Szlam A et al (2013) Spectral networks and locally connected networks on graphs[J]. arXiv preprint arXiv:1312.6203 6. Bulkowski TN (2012) Encyclopedia of Canlestick charts[M]. Wiley, Hoboken 7. Chai L, Xu H, Luo Z et al (2020) A multi-source heterogeneous data analytic method for future price fluctuation prediction[J]. Neurocomputing 418:11–20 8. Chan WS (2003) Stock price reaction to news and no-news: drift and reversal after headlines[J]. J Financial Econ 70(2):223–260 9. Chen Y, Hao Y (2017) A feature weighted support vector machine and k-nearest neighbor algorithm for stock market indices prediction [J]. Expert Syst Appl 80:340–355 10. De Gooijer JG, Hyndman RJ (2006) 25 years of time series forecasting[J]. Int J Forecast 22(3):443–473 11. Defferrard M, Bresson X, Vandergheynst P (2016) Convolutional neural networks on graphs with fast localized spectral filtering[C]. In: Proceedings of Advances in Neural Information Processing Systems, 3844–3852