

Data Mining Techniques for Analysing Prediction of Time Series Data in Stock Trading by Using Big Data Analytics

Peta Mahesh¹, Gade Shivaram Reddy²

^{1,2}. IT Department, Guru Nanak Institutions Technical Campus, Hyderabad, India

Article Info

Volume 9, Issue 6

Page Number : 521-530

Publication Issue

November-December-2022

Article History

Accepted : 01 Dec 2022

Published : 15 Dec 2022

ABSTRACT

Big data analytics can be used in many domains for accurate prediction and analysis of the large amount of data. They facilitate the discovery of significant information from large data, which is hidden otherwise. Time collection information is not unusual place in information units has emerge as one of the focuses of present day research. The prediction of time collection may be found out via the mining of time collection information, in order that we are able to reap the improvement procedure and regularity of social financial phenomena reflected via way of means of time collection, and extrapolate to are expecting its improvement trend. More and extra interest has been paid to time collection prediction withinside the era of huge information. It is the fundamental software of time collection prediction to as it should be are expecting the trend. In this paper, we introduce diverse time collection autoregressive version, transferring average version. As the time collection prediction in preferred scenarios, the ARIMA is implemented to the hazard prediction of the National SME Stock Trading in aggregate with specific scenarios. The case research display that the effects of our evaluation are essentially regular with the real situation, which has significantly helped the prediction of financial risks. In this paper, we describe an approach for analysis of the stock market to understand its volatile nature and predict its behavior to make profits by investing in it. We first provide literature survey of past works on this domain. Then we provide a methodology of our approach which contains data collection and machine learning algorithms. In this paper, we describe an approach for analysis of the stock market to understand its volatile nature and predict its behavior to make profits by investing in it. We first provide literature survey of past works on this domain. Then we provide a methodology of our approach which contains data collection and machine learning algorithms.

Keywords: Data mining, ARIMA, AM, big data, time series, forecast, stock

I. INTRODUCTION

Stock marketplace is producing substantial quantity of precious buying and selling records. This paper offers a evaluate on numerous records mining strategies implemented to inventory buying and selling records to research and expect inventory trend, define of proposed version for inventory prediction machine primarily based totally on technical, essential and outside environmental factors. This paper awareness on information inventory marketplace associated statistics for traders and examine National Stock Exchange records to expect inventory destiny movements. Review on Technical indicators, fee primarily based totally and quantity primarily based totally, that are broadly used to research inventory records. Stock analyst, traders and inventory agents are attempting to research inventory buying and selling records to expect destiny motion of stocks. Data mining is used to extract significant formerly unknown guidelines and hidden styles in records.

An economic change has vital functionalities: First is to inspire the method for the corporations through techniques for which they could change. The second is to kind out and cope with the setting, wherein change can appropriately happen. Contributing and benefitting from the marketplace has by no means been fundamental, and this is due to clean vulnerability and tremendously unpredictable nature of the marketplace i.e. shares/values can probable enhance and fall in price quickly. Instability is a authentic share of the scattering of profits for a given protection or marketplace file. Usually, the better the unpredictability, the extra risky the protection.

Recorded instability likewise 'acknowledged unpredictability' is the instability of true prices of fundamental stocks. They have tested to be maximum trying out but gratifying and superb for venture. There are distinct examinations at the behavior of the marketplace. Specifically, subordinates, for example,

fates and options have taken a variety of considerations, recently. Anticipating those subordinates isn't always sizeable for the hazard the board purposes. Other than that unique expectation of the marketplace's route can help economic professionals with gaining massive blessings with a restricted amount of capital. Financial change forecast may be visible as a hard time-dependent expectation [10].

There are several variables which are persuasive at the economic markets, along with political occasions, cataclysmic events, economic conditions, etc. In spite of the multifaceted nature of the tendencies in marketplace prices, marketplace behavior isn't always completely arbitrary. Rather, it is represented through a totally nonlinear dynamical framework.

Predicting the destiny prices is finished depending on the specialized investigation, which concentrates the marketplace's interest utilizing beyond prices and the alternative marketplace records. The marketplace exam is in logical inconsistency with the Efficient Market Hypothesis (EMH). EMH turned into created in 1970 through economic professional Eugene Fama whose principle stated that it's far not possible for an investor to outperform the marketplace as all of the to be had statistics is already there withinside the inventory prices. If the EMH turned into true, it'd be not possible to use device studying techniques for marketplace prediction. Nevertheless, there are numerous a success technical analyses in the economic international and the wide variety of research acting in instructional literature which are the usage of device studying strategies for marketplace prediction [9].

Time series data mining comes from the need of people to visualize data models according to their abilities. People rely on complex methods to perform these tasks. In fact, we can ignore small fluctuations to get the conceptual model and distinguish different

time models based on the similarity between models. The main time series related tasks include content-based querying, anomaly checking, pattern recognition, prediction, clustering, classification and segmentation. A large number of decision-making problems cannot be separated from prediction in various research fields of the natural sciences and social sciences, forecast is the basis of the decision-making [1]. Therefore, we mainly explored the time series data analysis and prediction. Time series prediction methods are divided into traditional time series prediction methods and machine learning methods. The traditional time series forecasting method refers to predicting the trend development of future time series only based on the trend development of historical time series.

At present, there are many methods for analyzing and predicting factors related to the relationship between supply and demand in the financial market, but the effect of this method is not obvious. We conducted a time series analysis of the financial and economic fields and used the ARIMA model to predict the risks of the National SME Stock Trading (New Third Board). The final results are basically consistent with the actual results, and good prediction results have been obtained.

II. RELATED WORK

In one of the approaches, the primary fetches the statistics after which preprocess it, after that it's far converted from excessive-frequency statistics to a ratio matrix after which the outlier set of rules reveals the anomalies in it. Then the predictions are made primarily based totally on the role of the anomalies and the end result is evaluated. The opinions on actual alternate statistics display that this method is extra powerful in predicting than the alternative conventional statistics mining algorithms. [1] To make a earnings through making an investment withinside the inventory marketplace there may be a want for

extensive making plans because of diverse uncertainties and its unstable nature. One manner to research developments of the inventory marketplace and to make a profitable funding is to remember the historic volatility of shares. One of the proposed approaches is to accumulate historic statistics of eight years (2009-2016) from the NSE internet site and additionally gift statistics of shares and indexes.

The ultimate charge of the inventory is taken into consideration on this method. After amassing statistics, it's far organized and processed the usage of Apache Hive. After that general deviation turned into calculated for the quarter, 4 – 12 months, 8 to 12 months length and turned into as compared with Nifty 50 index and the inventory with more general deviation turned into picked. Based in this evaluation for quarterly and 12 months smart returns on shares, regular shares and promising organizations may be found out which turned into established the usage of contemporary statistics on shares. [3] Another manner to assist buyers is to decide while the low and excessive expenses of shares arise with a view to parent out while to purchase and promote shares. For this one of the proposed answers is to apply a feed-ahead neural network.

ReLU turned into used as an activation characteristic and an adaptive second estimation optimizer turned into used. The statistics turned into amassed for nine years, turned into pre-processed and divided into education and check dataset. Then it turned into fed into diverse fashions like neural networks, ARIMA version, Support vector machines, Multi-layer perceptron version, etc. after that outcomes of diverse techniques were as compared with real statistics through calculating the suggest error, root-suggest-rectangular error, etc. Is turned into observed out that the feedforward neural networks gave the first-rate accuracy for opening the charge of a inventory. [4] Big Data research is applied basically in different

divisions for genuine prediction and exam of the large statistics sets.

They allow the revelation of vital statistics from large informational indexes. In this paper, a method on Cloudera-Hadoop primarily based totally statistics pipeline is proposed to carry out investigations for any scale and type of statistics, wherein US shares are tested to foresee each day will increase depending on non-stop statistics from Yahoo Finance. The Apache Hadoop Big-Data Framework is used to cope with extensive informational collections through disseminated stockpiling and making ready, shares from the US monetary alternate are picked and their regular gain statistics is remoted into making ready and check series to expect the shares with excessive every day pickups utilizing Machine Learning module of Spark. [5] Another proposed manner turned into to research monetary information and social media statistics to construct a prediction version that makes use of large statistics processing techniques, gadget gaining knowledge of and social media analytics for predicting inventory marketplace developments.

It shows that sentiment evaluation helps diverse evaluation techniques. Using social media contents with numeric statistics facilitates the excellent of enter and offers higher predictions [6] One of the proposed approaches for brief time period prediction is to apply a technique primarily based totally on hierarchical clustering, stepwise regression and ANN version for figuring out comparable historic patterns for shares and to are expecting every day inventory charge through optimal big variables the usage of characteristic selection. additionally, the processing is carried out the usage of a large statistics framework primarily based totally on R and Hadoop. and the accuracy is decided the usage of RMSE values of inventory items.[8]

Time series data is encountered in every aspect of the scientific eld [7]. A time series is a series of observations taken in chronological order. For

examples, a time series can be constituted by the closing price of a stock A on each trading day from June 1, 2015 to June 1, 2016; a time series can be constituted by the daily maximum temperature in a certain place; The station's environmental detection data records consist of a time series and so on. With the rapid development of big data, more and more time series data are stored in computers, so that we have a huge amount of time series data. Faced with these time series data, people want to reveal the information existing in these series data sets through effective methods or techniques. Today, the study of time series data has been rapidly developed and has become an important research direction in data mining. We can discover the inherent rules of things change and provide a reference for relevant people through the study of time series data.

III. TIME SERIES DATA ANALYSIS AND FORECAST

In this paper, a three-step analysis method for time series data analysis is proposed, Firstly, the data is pre-processed, which includes stationary processing of time series that are in an unstable state. Secondly, the pre-processed data is tested for stationarity. Finally, the prediction model is used to predict the probability distribution in the same time period in the future.

DATA PROCESSING

A time series can be considered stable when it has no systematic changes in the mean (no trend), no systematic changes in the variance and periodic changes strictly eliminated. The time series can be further subdivided into strict stationary and weak stationary.

For all time t , any positive integer k and any k positive integers t_1, t_2, \dots, t_k , the joint distribution of $r_{t_1+t_2}, \dots, r_{t_k}$ is the same as the joint distribution of $r_{t_1+Ct}, r_{t_2+Ct}, \dots, r_{t_k+Ct}$, we call the time series $f_{rt} g$

to be strictly stable and the joint distribution of $r_{t1} + r_{t2} + \dots + r_{tk}$ remains unchanged under the translational transformation of time. The above time series are strong stationary time series, but the time series we use are generally weak stationary sequence.

A weakly stationary sequence r_t must satisfy the following two conditions: $E r_t / D (\mu \text{ is constant})$. Variance $Cov(r_t) + r_{t-1} / D l, l$ only depends on l (l is any integer). For weakly stationary time series, the mean and the covariance of r_t and r_{t-1} do not change with time. We usually call a stationary sequence is weakly stationary in financial data. Differential operation is usually used to achieve the stable condition when the time series is not stable. The difference (forward here) is to find the difference between the value r_t of the time series r_t at time t and the value r_{t-1} at time $t-1$. Let us consider it as dt , it is a first-order difference. If the same operation is performed on the new sequence fdt , it is a second-order difference. Generally, non-stationary time series can be processed through d -time difference to be as stationary or as approximate as stationary time series.

TIME SERIES PREDICTION MODEL

The correlation coefficient is actually the angle between the two vectors in the vector space and the covariance is the expected value (or mean) of the product of their deviations from their individual expected values. The correlation coefficient is equal to 1 or $r-1$ when the two vectors are parallel (In particular, 1 means the same direction, $r-1$ means the reverse). If the two vectors are perpendicular and the cosine of the included angle is equal to 0, it means that the two vectors are uncorrelated. The smaller the angle between the two vectors, the closer the absolute value of the correlation coefficient is to 1, and the higher the correlation between the two vectors. The linear correlation between the two vectors is

measured by correlation coefficient. In the stable time series.

first g , the linear correlation between r_t and its past value r_{t-1} is measured by autocorrelation coefficient. The correlation coefficient between r_t and r_{t-i} is called the autocorrelation coefficient of spacing l of r_t , which is usually recorded as p_l specific:

$$\rho_1 = \frac{Cov(r_t, r_{t-1})}{\sqrt{Var(r_t) Var(r_{t-1})}} = \frac{Cov(r_t, r_{t-1})}{Var(r_t)}$$

The above formula uses the property of weak stationary: $Var r_t / D Var r_{t-1}$. For r_t samples of stationary time series, then the autocorrelation coefficient of the samples with an interval of 1 is estimated as:

$$\hat{\rho}_1 = \frac{\sum_{t=l+1}^T (r_t - \bar{r})(r_{t-1} - \bar{r})}{\sum_{t=1}^T (r_t - \bar{r})^2}$$

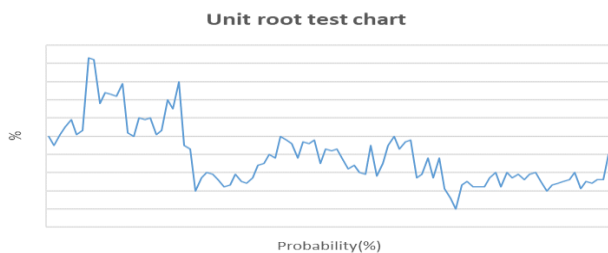
A series of autocorrelation sequences $P_1, P_2, P_3, \dots, P_n$ is called the sample autocorrelation function of r_t . We consider that the time series is completely uncorrelated when all the values in the autocorrelation function are 0. Therefore, we often need to check whether multiple autocorrelation coefficients are 0.

IV. ARIMA PREDICTION MODEL

So far, we have focused on stationary sequences. We can consider using the ARIMA model if the sequence is nonstationary. The ARIMA can be used for statistics and artificial intelligence [12]. ARIMA has only one more letter "I" than ARMA, which means that it has one more level of connotation than ARMA. A non-stationary sequence can be transformed into a stationary time series after d times of difference. For the specific value of d , we first perform a stationary test on the sequence after the first difference. Then we will continue to make the difference if it is still

non-stationary until the test is stationary after d times. Finally, the specific value of d is calculated.

ADF is a common unit root test method [13]. Its original hypothesis is that the sequence has a unit root, and the sequence is non-stationary. It is necessary to be significant at a given confidence level and reject the original hypothesis for a stable time series data. According to Table 1 and Figure 2 above, we assume the original hypothesis that the sequence has a unit root. The original hypothesis cannot be rejected because we can see that the value of p-value is 0.1704489, which is much larger than the significant level. Therefore, the daily index series of the Shanghai Stock Index is non-stationary.



Because it can be seen that the p-value is very close to 0 and the original hypothesis is rejected. The value of d for the original sequence can be 1 because the sequence is stable after one difference. An ARMA model can be built from the differential sequence after the value of d is determined. At present, ARIMA has been widely used in various elds [14]. Next, we will use the ARIMA model to analyze example in the financial eld.

	value
Test Statistic Value	-2.30472
p-value	0.170449
Lags Used	1
Number of Observations Used	379
Critical Value(1%)	-3.44772
Critical Value(5%)	-2.8692
Critical Value(10%)	-2.57085

Table 1. Unit Root Inspection Table.

DF test results	Thresholds of different significance levels		
	1%	5%	10%
-2.456a	-3.4457	-2.8682	-2.5703
-18.961b			

Table 2. Unit Root Test Results.

probability interval (%)	Probability distribution of daily increase					
	<-5	<-3	<-1	>1	>3	>5
value (%)	2.14	6.9	23.33	23.81	10.24	4.76

Table 3. Daily Increase Probability Distribution Table.

The unit root evaluation of the upward push primarily based totally at the inventory fee inside a sure time frame can decide the steadiness of the upward push collection. The possibility distribution of various rises and falls withinside the equal length withinside the destiny may be inferred primarily based totally at the ancient distribution of the upward push while the series is stable, in order that the fascinated events put together plans for excessive conditions that significantly have an effect on the extent of net really well worth of funds. In latest years, the OTC New Third Board has evolved hastily in China. We can find that the New Third Board marketplace has traits after cautious observation. Firstly, the general marketplace fee volatility is significantly better than that of the Shanghai and Shenzhen markets.

Secondly, the volatility distribution is significantly rightward. The fluctuation chance of person shares is frequently launched quickly and violently due to the fact there may be no restrict of the day by day restrict system. In the following, we cognizance at the sensible problems withinside the Chinese NEEQ inventory marketplace. The time collection evaluation approach changed into used to estimate the distribution possibility of destiny upward push and fall primarily based totally at the differential time collection of day by day upward push and fall of inventory prices.

V. ANALYTICAL METHOD

We used the three-step evaluation approach proposed above for evaluation: In the first degree, a non-desk

bound series is transformed into a desk bound time collection with the aid of using differential processing. In the second one degree, we use the ADF unit root check to check whether or not the time collection is stable. In the 0.33 degree, the probability distribution of various rises and falls within the equal term primarily based totally at the ancient distribution of rises and put together is inferred for excessive conditions that could significantly have an effect on the extent of NAV.

STEP 1: statistics preparation (statistics preprocessing). The time collection is defined as a chain of quantitative observations at consecutive times. In the evaluation of financial time collection, the fee time collection itself is commonly volatile, now no longer completely random distribution, and has apparent autocorrelation. At the equal time, the regulation of fee distribution can also additionally also alternate suddenly because of a whole lot of factors, in order that the regulation set up within the beyond degree might not nonetheless maintain within the destiny. Therefore, it's far commonly invalid to investigate the fee time collection immediately in an try to find the regulation or regression formula. We pre-procedure the time collection earlier than making use of the ARMA version if the series is non-desk bound. Generally, the approach for managing volatile time collection is to make first order distinction of the time collection.

A non-stationary sequence can be transformed into a stationary sequence after d times of difference. The specific value of d depends on the structure of the stationarity test after the time series difference. we will continue to make the difference if it is still non-stationary until the test is stationary after d times. The relative ratio of the stock prices (the relative increase) is more concerned about the absolute value of stock price changes, so that the ratio method is generally used in the analysis of financial product price time series [16]. At the same time, the stock price

difference will continue to increase or decrease accordingly after the stock price continues to rise or fall. Therefore, it is proposed to use the natural logarithm of the ratio of adjacent variables in the time series of stock prices to perform first-order difference processing.

A distinguished benefit of this approach is that the first-order distinction collection yet received from this approach is about same to the inventory rate increase, that could be without delay used for the opportunity prediction of the destiny inventory rate distribution. In this paper, we use the ratio approach to cope with time collection.

STEP 2: Stationarity check. We practice the unit root take a look at to the logarithmic upward push collection. Our intention is to research the stationarity of the residuals to decide if the ARMA model is a right version for them. The authentic speculation of the unit root take a look at is to check whether or not the collection is stationary. Then, negating the authentic speculation manner that the collection (or the differential collection on this example) is stationary. Specially, we use ADF (Augmented Dicky-Puller) to test whether or not the time collection is stable. STEP 3: Prediction of the opportunity distribution of risk. We can put together for excessive conditions that critically affect asset benefit degrees with the opportunity distribution of rate the following day. The simple concept of the ARMA version is to combine the AR and MA models in order that the wide variety of parameters used is saved small.

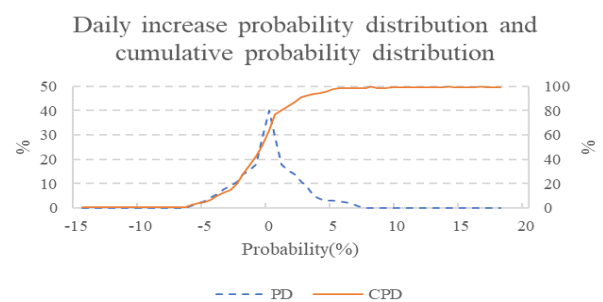


Figure 2. Daily increase probability distribution and cumulative probability distribution.

It may be visible from desk 2 that the life of unit root withinside the ultimate fee time collection can't be denied even at the confidence stage of 10%, in order that the ultimate fee time collection is largely non-stationary collection may be rejected to have a unit root while the confidence stage is significantly better than 1%, that's that the time collection of logarithmic booms and reduce is largely stable. Therefore, we are able to use the time collection evaluation to predict the destiny possibility of distribution primarily based totally at the inventory's beyond advantage information. The following is a forecast of the distribution of the advantage withinside the subsequent 1 switch day primarily based totally on historic advantage information.

As may be visible from the day by day possibility distribution of ultimate expenses has a feature of peaks and lengthy tails. When carried out to intraday TCO buying and selling, we should be aware of placing affordable prevent loss expenses to avoid big losses resulting from small possibility events. In addition to the day-by-day boom distribution probability, we can't immediately use the cumulative distribution of the smaller length boom stage possibility while we want to attain the boom distribution possibility of different time periods. For example, it isn't feasible to attain a weekly boom horizontal distribution possibility or a month-to-month boom horizontal distribution possibility from the day-by-day boom horizontal distribution possibility via way of means of a super imposed manner. The accurate processing technique is to immediately find the logarithmic first-order distinction of the weekly closing fee series to attain the time collection of the logarithmic upward push of the weekly ultimate fee and manner the series. For example, facts at the weekly logarithmic upward push time collection of Jindalai are proven. They may be

acquired different distribution chances via way of means of interpolation withinside the distribution possibility desk or via way of means of taking intersection factors at the possibility distribution curve.

Assuming that the possibility of prevalence is much less than 5% as the buying and selling condition, via way of means of making horizontal strains p D 5% and pD95% at the ordinate to intersect the distribution curve at factors (Representing the possibility distribution factors of immoderate decline and immoderate boom, respectively), taking factors at the abscissa can get an immoderate drop the corresponding abscissa is 4.5%, and the corresponding abscissa is 4.35%. In different words, there's a 84% possibility that the inventory's person shares will upward push or fall between 4.5% and 4.35%. Therefore, It may be sold after which offered intraday TCO buying and selling while the inventory falls via way of means of greater than 4.5% at the identical day; It may be offered after which sold intraday TCO buying and selling while a inventory rises greater than 4.35% at the identical day.

With the non-stop development on time collection facts mining technology, its utility has been prolonged to financial evaluation and it could nicely expect the dangers withinside the financial eld withinside the destiny. We analyzed the time collection facts and its numerous prediction models, then carried out the ARIMA prediction version to the evaluation of financial. The paper proposed a time collection evaluation technique to expect the destiny upward push thru the historic upward push and fall possibility distribution curve. The possibility of falling distribution makes an awesome prediction for severe conditions which can significantly have an effect on the extent of internet belongings in actual lifestyles in order that we will infer what's going to take place withinside the destiny thru time collection facts sequences. As the studies continues, greater

technology can be considered to increase the scope of prediction and enhance the accuracy. We count on the time-collection predictions could have greater programs in financial. By accurate price predictions, we will enhance reaction measures to dis-cowl viable emergencies; We can enhance the prediction through including the spatial size below the aggregate of time and space. For example, the higher aid utilization for customers and taxis may be generated through predicting the number of taxi rides in a sure location of Didi taxi. This also can correct financial predictions, such as, it could assist managers Reasonably specify method through predicting the amount of cash offered and sold, etc. The facts mining of time collection facts whose steering and assist [25] to real manufacturing and lifestyles turns into increasingly important.

VI. CONCLUSION

The time collection evaluation proposes a technique of estimating the distribution possibility of destiny inventory rate fluctuations primarily based totally on historic rate information, thereby we will avoid the hazard of massive losses resulting from intraday trading. The technique of without delay judging the destiny rate fashion primarily based totally at the historical rate frequently lacks sufficient reliability due to the fact rate collection are frequently unstable. We first carry out a restored difference processing at the inventory rate collection, that's equal to converting the studies item from the inventory rate itself to the alternate cost of the inventory rate. Then test the stationarity and decide the stationarity of the boom collection through unit root evaluation of the logarithmic boom of the inventory rate inside a sure length of time. Finally, we follow ARMA (3,3) to the solid differential series to attain the historic distribution rule. We use this rule to attain the possibility distribution of various rises and falls withinside the equal length of time withinside the destiny, in order that we will put together for severe

conditions which could significantly have an effect on the extent of internet asset cost.

VII. REFERENCES

- [1]. L. Zhao and L. Wang, "Price Trend Prediction of Stock Market Using Outlier Data Mining Algorithm," in 2015 IEEE Fifth International Conference on Big Data and Cloud Computing, Dalian, China, 2015, pp. 93–98.
- [2]. M.D. Jaweed and J. Jebathangam, "Analysis of stock market by using Big Data Processing Environment" in International Journal of Pure and Applied Mathematics, Volume 119.
- [3]. Book "Data Mining Techniques", by Michael J.A. Berry and Gordon S. Linoff, 2nd edition.
- [4]. Han, J., Kamber, M., Jian P. (2011). "Data Mining Concepts and Techniques". San Francisco, CA: Morgan Kaufmann Publishers.
- [5]. Sang C. Suh, Ed. Practical Applications of Data Mining.
- [6]. Pring, M.J. (1991). Technical analysis explained. New York: McGraw-Hill.
- [7]. Hajizadeh E., Ardakani H., and Shahrabi J., "Application of data mining techniques in stock markets: A survey", Journal of Economics and International Finance Vol. 2(7), pp. 109-118, July 2010.
- [8]. D. Zhang, "High-speed train control system big data analysis based on fuzzy RDF model and uncertain reasoning," Int. J. Comput., Commun. Control, vol. 12, no. 4, p. 577, Jun. 2017.
- [9]. W. Xu, L. Liu, Q. Zhang, and P. Liu, "Location decision-making of equipment manufacturing enterprise under dual-channel purchase and sale mode," Complexity, vol. 2018, Dec. 2018, Art. no. 3797131.
- [10]. J. Y. Chen, "Thrown under the bus and outrunning it! The logic of DiDi and taxi drivers' labour and activism in the on-demand economy," New Media Soc., vol. 20, no. 8, pp. 26912711, Aug. 2018.

- [11]. L. Zhang, J. Lu, J. Zhou, J. Zhu, Y. Li, and Q.Wan, ``Complexities' day-to-day dynamic evolution analysis and prediction for a Didi taxi trip network based on complex network theory," Mod. Phys. Lett. B, vol. 32, no. 9, Mar. 2018, Art. no. 1850062.
- [12]. Y. Lu and X. Xiong, ``Topic analysis of microblog about `didi taxi' based on K-means algorithm," Amer. J. Inf. Sci. Technol., vol. 3, no. 3, p. 72, 2019.
- [13]. J.-C. Kim and K. Chung, ``Mining based time-series sleeping pattern analysis for life big-data," Wireless Pers. Commun., vol. 105, no. 2, pp. 475489, Mar. 2019.

Cite this article as :

Peta Mahesh, Gade Shivaram Reddy, "Data Mining Techniques for Analysing Prediction of Time Series Data in Stock Trading by Using Big Data Analytics", International Journal of Scientific Research in Science and Technology (IJSRST), Online ISSN : 2395-602X, Print ISSN : 2395-6011, Volume 9 Issue 6, pp. 521-530, November-December 2022. Available at doi : <https://doi.org/10.32628/IJSRST2215112>
Journal URL : <https://ijsrst.com/IJSRST2215112>