An Intelligent Model for Identifying Fluctuations in the Stock Market and Predicting Investment Policies with Guaranteed Returns

Manash Sarkar, M. N. Pratima, R. Darshan, Debkanta Chakraborty, and Maroi Agrebi

1 Introduction

Stocks are critical parts of cutting-edge financial markets. The stock market is a tough forum for funding, and it calls for ample deliberation before investing money into stocks. The stock market is one of a number of important sectors that traders are dedicated and committed to. Researchers from both the economic and technical fields continue to be interested in the topic of stock market price trend indicators and predictions. Our goal is to develop a rate trend prediction system that specialises in short-term price trend prediction [1]. Modern financial systems are based primarily on fiat or decree money, which has many advantages, benefits that are due to its divisibility, sturdiness, transferability, and abundance. In addition, fiat money is not back up with a physical commodity, and governments have the power to manipulate the value of money. This may cause many problems, such as hyperinflation and income inequality [2]. Second, although people invest

---

M. Sarkar (✉)

Department of Computer Science & Engineering, Atria Institute of Technology, Bengaluru, Karnataka, India

Expert Data Scientist, Limendo, Bolzano, Italy
e-mail: manash.sarkar@atria.edu; manash.sarkar@limendo.com

M. N. Pratima · R. Darshan
Department of Computer Science & Engineering, Atria Institute of Technology, Bengaluru, Karnataka, India

D. Chakraborty
Department of Computer Science and Engineering, Dr. B. C. Roy Engineering College, Durgapur, India

M. Agrebi
LAMIH UMR CNRS 8201, Department of Computer Science, Université Polytechnique Hauts-de-France, Valenciennes, France

in stocks, they are unpredictable and unreliable. The stock rate is often affected by different factors, including investor behaviour, market sentiments, dividends, market regulations, economic developments, economic guidelines, credit situations, hobby prices, trade fee levels, and global capital flows [3]. Inventory market buying and selling are parts of an exceedingly complicated and ever-changing system where people can either accrue a fortune or lose their life savings [4]. The third problem is security: protecting the amount that people are buying and selling, current process gains, and income growth from being exposed to malicious users and fraudulent schemes. A primarily sequence-based stock price [5] is the standard nonstationary stochastic system that does not require a regular maintenance for it to stay in equilibrium. A mathematical model to describe the dynamics of this sort of procedure, after which forecasts or predictions on the price of future expenses from present values and past values, has been set up thanks to the results of numerous studies on nonlinear time collection evaluation. It also forecasts stock prices along four metrics: open fee, ultimate price, best rate, and the bottom or base amount [6]. Thus, it uses all extracted or captured guidelines at any moment for its predictions, estimating future inventory or stock market price with high accuracy.

Over the past decade and continuing today, big data technology has had a significant impact on a number of industries, and it could become ubiquitous. The financial sector, like the majority of other sectors, concentrated all of its efforts on structured information research. However, statistical records saved in various sources could be obtained with the aid of technology. Big data generate a lot of power, not only through the ability to process large volumes and types of data at fast rates but also through the capacity to charge organisations. The relationship between news articles and stock prices has already been established, but it can still be strengthened by additional data. Businesses have historically made greater investments in selection support structures. Many businesses can now obtain analytical reviews that are based on OLAP (Online Analytical Processing) structures thanks to the development of enterprise intelligence tools. However, given the frequently shifting trends in consumer behaviour and consumer markets, it may become increasingly important for businesses to move beyond OLAP machine-based analysis in their analytical reviews. In their paper, they discussed about OLAP and its significant innovations in hardware and software enable businesses to take advantage of all available data formats and support businesses in gaining business insights from those statistics. Some of the most important things to concentrate on in obtaining real-time analytics on all available factual codecs are facts.

Not just for data warehousing, data technology also includes a variety of other technologies that are used to passively track recent developments in threat analysis by using large numbers of data across many domains. This chapter aims to examine the conceptual methods that draw on literature assessments of those models and improve them. In order to address the current roadblocks to this aim, similar researchers can benefit from the information in the most recent papers. In this chapter, the authors rely on fuzzy common sense and neural networks. They also use RSI (Relative Strength Index), OS (Operating System), MACD (Moving Average Convergence/Divergence), OBV (On-Balance Volume), MSE (Mean Squared

Error), MAPE (Mean Absolute Percentage Error), and RMSE (Root Mean Squared Error) to evaluate time-series data. The chapter emphasises locating implementation challenges and how these challenges can help the trader make decisions in the stock market. The study's implication is that there is too much data to examine. Additionally, this chapter is conceptual in nature. The study's findings demonstrate that the ANFIS (Adaptive Neuro-Fuzzy Inference System) is the best model for stock market prediction, outperforming all others. The predictions and expert decisions are useful in entertainment industries [7]. The expert decision for the selection of a restaurant is also implemented in the recommendation system, where a restaurant recommendation [8] is also a useful application for expert decisions in tourism industries. Additionally, the fuzzy neural model's results show that they are more effective than other models, even though it appears that when MSE is used as a proxy for a medium, MAD (Mean absolute deviation) uses significantly less data than the other models do. As a result, it will become more adept at stock prediction. One of the methods, a reinforcement-oriented forecasting framework, which converts the solution from a typical mistake-based learning approach to an intention-directed, shape-based total learning approach, includes a trial-and-error-based total approach that was used for prediction. To make predictions on the stock market the way it ought to be, many advanced prediction techniques have been developed. At one time, some approaches were generally referred to as conventional approaches, but there were no computational techniques for hazard analysis. Pass validation, which is one of the most frequently used fact-resampling techniques to estimate actual prediction blunders and to fine-tune model parameters, is one of the techniques used to validate the model—a method for analysing the stock market, assessing how risky it is, and anticipating how it will behave so that investors can profit from their investments.

In this chapter, an attempt is made to construct a time-series prediction model to predict inventory prices. This construct is more realistic because it examines the model by using not only current algorithms but also records from the benchmark, which have made it difficult to determine whether it will outperform other algorithms. Modern-day stock values accrue from sets of statistics. The statistics gathered are modelled into diverse subcomponents or data sets, which are used to train and test the algorithm. Further, a customised intelligent version is modelled and improved to provide prediction scores for all the assessment metrics. An intelligent model is proposed to anticipate the stock market and offer an expert decision on how to invest. The proposed model additionally guarantees that an investor's amount will not be much less than their capital investment after a specific amount of time. The base amount could be fixed to the funding time. To reduce uncertainty and risk, a fuzzy logic is used, and fuzzy multiobjective optimisation is likewise used to make an optimistic decision. The proposed model predicts stock and determines the term required for a good return. A real-time stock market data set is used as a test-bed data set to validate the proposed model. Cross validation is an information resampling technique that assesses the generalisation ability of predictive models and prevents overfitting. As a result, fourfold cross validation [9] is used to gauge how well the proposed or developed model performs. Similar or comparable to earlier works,

the proposed model outperformed the models based on machine learning and deep learning.

The rest of the chapter is organised as follows: A literature survey is featured in Sect. 2. The impact of big data on the stock market is described in Sect. 3. The proposed model is described in Sect. 4. Sect. 5 shows how it is implemented. The results and discussion appear in Sect. 6. Finally, the conclusion and a brief mention of future works appear in Sect. 7.

2 Literature Survey

The objective of any investment is to earn return. Stock returns constitute one area for this chapter to look at, because many scholars have shown serious interest in this topic over the past several years. A quick evaluation of the literature assists in understanding the relevance of the content evaluation for inventory returns. Stock and other financial markets are complicated. They feature dramatic and dynamic behaviour and experience surprising booms and busts. The major challenge of this analysis is to build a device that can anticipate future fees in stock markets by taking samples of future costs. The experimental results of the study observed by Wei Yang [10] over a three-month screen showed that this version can effectively anticipate the route of the market with an average hit ratio of 87%. Similarly, on day-by-day prediction, the model was capable of predicting the open, excessive, low, and near fees of the preferred inventory every week and every month. An optimisation technique, namely principal component analysis (PCA), was implemented in a short time period for predicting stock market prices, and it includes preprocessing the stock market data set, engineering strategies, and a customised machine-learning, deep-learning machine for inventory market price trends and value predictions.

Several investigations have given rise to various selection guides that present buyers with superior predictions and analyses. Along with the wide availability of stock or price statistics, which was made possible through the Internet, the task of the investor has become more complicated, in that they now need to store, accumulate, examine, and filter out data to make accurate decisions from a diversity of information. This information consists of financial facts, real-time data, and opinions on how to effectively change the economic markets. It is far important to develop models that uncover the unique states of the market that will require making adjustments.

Quite a few interests have shifted to using various artificial intelligence techniques in the stock market, in addition to the statistical data that have been used to comprehend and anticipate fluctuations within the stock market. Technical analysis has also drawn the attention of many researchers as a means of reducing investment costs in the stock market. Therefore, knowing the market and being able to predict what will happen soon are essential skills for any investor. Ijegwa et al. [11] proposed solving this problem by working on an easy-to-use inference sign model that uses many variables to simplify the complicated marketplace. It gives reliable

hints to traders and, as a consequence, acts as a valuable selection tool. The work employs fuzzy good judgement to perform the decision-making task, based entirely on inputs from technical analysis indicators. No longer will the best techniques, methods, and algorithms be enough to best predict the stock market, but this solution comes with shortcomings, disadvantages, and risks.

A distinction between stock cost or share cost and their respective prices must be made. Price shows the buying price and the selling price, and the stock cost is an underlying price—no matter the buying and selling prices, which do not rely on a country to determine the stock price. The cognitive styles of traders and rumours frequently make it difficult to charge a fee. The role of the investor is to evaluate the investment's values, costs, and margin of safety in order to raise money, market the products, or sell them. Understanding the underlying technical method is genuinely difficult regardless of the excellence of these practices. The rate of the stock is determined by the past patterns, using a variety of techniques and tools, as suggested by Mangale et al. [12]. The approaches are typically expressed in terms of a technical method and an essential method. The essential method is used for long-term valuations. Each inventory has a price that is independent of the inventory's price; this is known as its intrinsic value. The proposed model operates through the following stages: information gathering, feature processing, fuzzy logic mapping, and stock value calculation. The mapping of the exceptional number of evaluation factors uses fuzzy logic. If-then rules are applied to the linguistic variable. The inventory price, which is used to calculate inventory worth, is affected by the fuzzy model. The dividend discount version is used to calculate the inventory fee. The purpose of the paper proposed by Alalaya et al. [13] was to take into account the projection capabilities of fuzzy logic and neural networks, as well as to combine some of the models to address various models' implementation issues in the prediction of indexes and fees for the Amman Inventory Alternate, where other researchers have demonstrated some of the differences between these measures.

Given that the stock market depends on nonstationary economic data and that the majority of the models used in this study have nonlinear structures, the Amman Stock Exchange index prices were used as a data set to examine the unique utility models. The facts are first gathered and preprocessed and then transformed from high-frequency facts to a ratio matrix. Next, the outlier set of rules identifies any anomalies in the ratio matrix, which determines the results. Although behaviour evaluation is no longer the most effective, because of its unstable nature, it is essential for all technical and social analyses. The relationship between sentiment and stock values is then examined. The developed model can afterwards be applied to forecast future stock prices. The development of this hybrid approach for forecasting inventory prices is a step toward better forecasting. In addition to quantitative data, financial news and news articles were analysed for the neural network forecasting of the proportion charge. A novel method for time-series forecasting with a simple, linear computational algorithm was proposed by Pandey et al. [14]. The suggested method uses a quantile-based fuzzy forecasting method after first predicting the trend of the future value. The suggested method is less complex than other techniques. The method's suitability for Sensex forecasting is

also investigated. We identified specific qualitative terms from our research that had favourable effects and unfavourable effects on stock price. Some of these words were frequently used in different contexts, indicating that the articles that contained them were vulnerable to price changes. After being retrieved from online databases and virtual libraries like the ACM (Association for Computing Machinery) digital library, Scopus, Kaggle, and Nifty 50, the statistical units underwent a critical analysis. Moreover, a thorough comparative analysis was completed to determine the route of structured and unstructured records for the segmentation of Nifty 50 shares. The Nifty 50 stock list was obtained from the website of the National Stock Exchange of India. The sentiment rating was derived for each tweet and each stock in the Nifty 50 by using the unstructured statistics that were downloaded from Twitter for sentiment evaluation. Twitter is a relatively new way of disseminating information, not only through short sentences but also by allowing users to highlight one specific piece of information by retweeting it.

Similar studies have characterised market performance on information: There should be no gain opportunities in stock markets, because they are defined by random walk patterns, while all new information is evaluated at any time given that prices immediately reflect that information. A crucial observation is that big data analysis should be used instead of stock market or financial fluctuations. Groups and businesses now have access to exceptional volumes and types of data thanks to cloud computing, the Internet of Things, Wi-Fi sensors, social media, and many other technologies. Innovative threat evaluation methods and applications are undergoing related developments that are tested and analysed. For an actual analysis, the data set must be preprocessed and improved. We can train the vector device on the data set and the results it produces after preprocessing the data, which was observed using the arbitrary woods technique.

To obtain precise recommendations that are based on customers' preferences, a hybrid recommendation model built on crowd search optimisation is suggested. Collaborative and content-based filtering are combined to create the hybrid recommendation described by Srakar et al. [15]. The actual time is in [16], and the data set is from the Nifty 50 stock list on Kaggle. Once investors understand the stock market and the fact that it involves a financial risk, then analysis, knowledge, advice, and prediction are all possible. Knowledge of the stock market is important because it currently makes up the majority of the stock market itself and requires a firm understanding of how stock prices are trending both now and into the future. Such knowledge could help investors to understand the upcoming inventory charge and to assess the risk. One of the resources used to train system-learning models to predict stock prices is the abundance of available stock data. Other factors that make it difficult to predict the market include a high noise-to-signal ratio and the sheer number of factors that affect inventory costs. Efforts are being made to improve and use smart devices, which include neural networks, fuzzy structures, and genetic algorithms within the discipline of financial decision-making. Thanks to the use of fuzzy rules and logic, an intelligent model has been proposed to predict stock expenses in order to guarantee returns on investment over some particular period of time, which is mathematically applied, thoroughly examined, and finally confirmed through real-time data sets.

3   Impact of Big Data in Stock Market

This section of the chapter in detail explains big data analytics, its impact on the stock market, its importance in the big data architecture, and the preparation of data sets required for further implementation, testing, and validation.

3.1   Big Data

Big data has emerged as crucial to the tech panorama. Big data, or massive information, analytics enables agencies to harness their facts and use them to discover new possibilities. Big data analytics, according to Modi et al. [17], is utilised in many fields for the precise prediction and analysis of numerous records. It facilitates the discovery of facts from huge amounts of information that provide guidance on how to use predictive analytics and personal behaviour analytics. Big data compile a collection of complex data sets for a method utilising conventional database management or storage equipment and traditional tools or conventional statistics-processing applications. The key tasks for this model include capturing, storing, analysing, and visualising those data. A massive investigation into facts is applied in certain disciplines, according to Attigeri et al. [18], for specific prediction analysis.

3.1.1   Big Data Architecture

Big data architectures have been developed to carry out the gathering, processing, and analysis of statistics that are too large or complex for conventional database systems. Depending on the skills of the clients and their equipment, there are different thresholds at which businesses step into the world of facts. It may mean hundreds of gigabytes for some people, while it may mean many terabytes for others. The means of handling large data sets improves along with the tools for doing so. As opposed to focusing on just the size of the facts, though they tend to be large, this era increasingly focuses on the value that can be extracted from data sets through advanced analytics (Fig. 1).

3.2   Structure of Big Data

According to Rajakumar et al. [19], both structured data and unstructured data have played significant roles in the decision-making processes for daily buying and selling activities. Gathering information reports is a preferred starting point for the decision-making process of a buying-and-selling activity, which leads to the creation
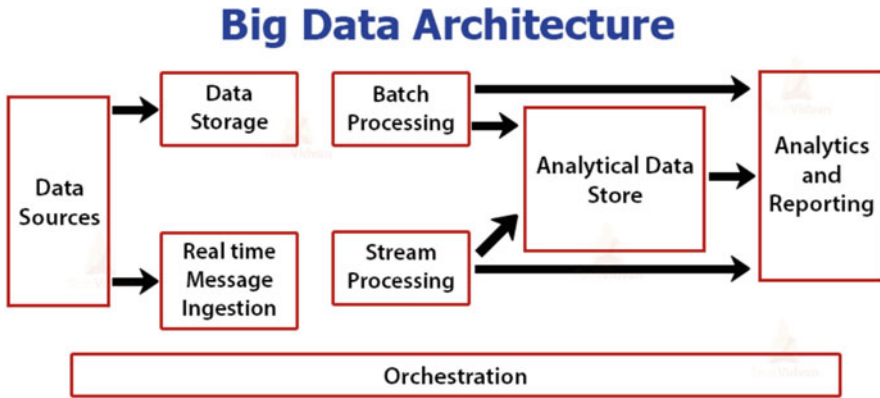
Big Data Architecture

Fig. 1 Big data architecture

of the buy, keep, and sell funding method decision. A specific array of resources, including the agency database, publications, the press, radio, television, and the Internet, are used to collect information bulletins and business moves. Information from businesses include their planning, financial statistics (EPS, PE), income potentials and expectations, shares, returns, financial statements, and economic feedback. Marketers or other individuals spread exciting and informative messages. Every message is thoroughly investigated, taking into consideration variables such as economic expertise, experiences, backgrounds, and cognitive biases. On the basis of this analysis, a decision that takes into account risk tendencies and trading objectives is made. Finally, the effects or outcomes are assessed, and a recommendation to buy, hold, or sell is made. Rouf et al. [20] have used machine learning in inventory price prediction to find patterns in data. Typically, stock markets generate a sizeable amount of structured and unstructured heterogeneous statistics. It is now possible to quickly examine more-complex heterogeneous records and produce more-accurate results by using machine-learning algorithms.

3.2.1 Structured Data

Mathematics follows a prescribed format and structure, and for that reason, it is simple to use to examine and analyse the statistics. It collectively conforms to or is represented in a tabular format, with relationships in between special rows (features) and columns (records). Structured query language (SQL) databases are commonplace examples of structured records that rely on how facts may be saved, processed, and accessed.

Nifty 50 is a collection or a basket of the 50 most-energetic stocks on the national inventory alternate of India, which acts as a benchmark for the general movement of the inventory market, according to Sujata Suvarnapathaki [21]. The Nifty 50 is in the form of dependent records, and traders are regularly interested in knowing the

standard behaviour of stocks. Having a prior idea of the worth of an inventory or the movement of an inventory would benefit buyers. A stock market analysis allows buyers to make informed decisions on whether to invest, not invest, or maintain a current investment. For this purpose, traders can also rely on historical data on the essential analysis parameters and the technical analysis parameters influencing the behaviour of stocks.

3.2.2 Unstructured Data

Unstructured data are statistics that come in unknown shapes, cannot be stored by using conventional methods, and cannot be analysed until they have been converted into established formats. Multimedia content material like audio files, motion pictures, videos, and pictures are examples of unstructured data. Recently, unstructured data have been growing faster than other sorts of big data. With the advent of social media, the democratisation of value, and more-efficient information systems, data has never been so easy to gather and so accessible. Each day, around 500 million tweets are publicly sent around the world, and more than one billion customers are on Facebook.

One of the most promising prospects for finance is thought to be big data, specifically for managing risk. The financial industry developed dashboards to manipulate risks by using a strong framework of social media scrutiny. The definition of *big data* is a combination of structured data and unstructured data (such as inventory costs, heartbeats, and images) produced in real time, according to Sanger et al. [22, 23]. These authors brought attention to the fact that qualitative records can be duplicated or can be reflected in statistics that have not been properly taken into account when calculating inventory fees. In my experience, tweets are similar to or fit the definition of massively produced unstructured data, and social media has become the subject of many studies in the social sciences and economics.

3.2.3 Semistructured Data

Semistructured data are a form of big data that do not conform to the formal shape needed by data models. However, these data come with a few types of organisational tags or distinctive markers that assist in cutting up or splitting semantic factors. JSON documents or XML documents store these data. This class of data includes facts that are significantly easier to examine than those in unstructured records. A range of tools can analyse and process XML documents or JSON documents, decreasing the complexity of the analysing system. Semistructured statistics can be converted by using the packages of big data from a given economic quarter, which include social media analyses, net analytics, risk management, fraud detection, and security intelligence.

According to Bach et al. [24], text mining or textual content analytics is one method for obtaining information from a vast array of big data. The goal of text

content mining, which is also known as text data mining or text analytics, is to examine textual data, such as emails, reviews, texts, webpages, reviews, and legitimate documents in order to extract data, transform it into records, and make it useful for various types of decision-making. In its final form, text mining may be used for evaluation, visualisation (through maps, charts, and thought maps), integrating information from databases or warehouses, machine learning, and many other purposes. Text mining includes linguistic, statistical, and system-learning techniques.

3.3   Big Data in the Stock Market

At present, the market actions and behaviours are converting faster than ever, making the market even more unpredictable. Businesses must verify the alternative enterprise techniques and enforce them with top-of-the-line enterprise solutions, according to Lima et al. [25, 26]. In the era of top-of-the-line technological innovation, the various types of records can be improved by using the latest information technology, and data statistics is one of the most precious inputs for automated systems, according to Hasan et al. [27, 28]. Financial markets and technological evolution have affected every human interest over the past few years. Big data generation has become a fundamental part of the financial services enterprise and could continue to pressure future innovation, according to Razin E [29]. Analysing big data can enhance predictive modelling to better estimate the costs and outcomes of investments. Gaining access to big data and advanced algorithmic expertise improves the precision of predictions and the ability to mitigate the inherent risks of buying and selling efficiently. Data may be reviewed and applications advanced to regularly update information in order to make accurate predictions. Financial businesses use large numbers of data to reduce operational risk, combat fraud, significantly reduce information asymmetry issues, and meet regulatory and compliance requirements.

Contributing to and reaping the benefits from the market have not been easy, because of the market's obvious vulnerability and unpredictable nature. Stocks can quickly rise and fall in value, according to Jaweed et al. [30]. Stability is only a percentage of the earnings that are dispersed for a particular safety or market report. Usually, the more unpredictable something is, the less safe it is.

Big data are fundamentally changing both how investors make investment decisions and how the world's stock markets operate. When given records, computers are now able to make precise predictions and decisions that are similar to those of humans. This allows them to execute trades at high speeds and with high frequency. A technical analysis and a fundamental analysis of a prediction are available options. According to Sean Mallon [31], a technical evaluation is carried out by applying device learning to historical data on inventory costs, and a fundamental evaluation is carried out by conducting a sentiment analysis on social media data.

More so than ever before, social media data have significant effects and can help forecast stock market trends. The strategy involves gathering data from social media and extracting the expressed sentiments. The relationship between sentiment and stock value inventory is then examined.

3.4 Nature of Dynamic Data in the Stock Market

After being recorded, dynamic data continuously change or vary in order to preserve their integrity. The stock market is a true, dynamic, extremely complex system that is constantly changing. The stock price is volatile and dynamic in nature. The pipeline, the zoom info [32], and dynamic data enhance business-wide communication and alignment. Access to the same customer data is available throughout the entire organisation. Consistency in sales, marketing, and branding is easier to achieve if a data set is dynamic.

The quantitative extraction method for the financial market system indicates that the dynamic characteristics, the dynamic process [33], and the mechanism of the financial market system's dynamic characteristics are part of an evolution of monitoring abnormal behaviour in the financial market's bubble, crisis, and collapse process but are not enough. But before investing, certain factors that influence stock prices must be considered, such as market trends, price actions, etc., according to John et al. [34]. Data scientists are exploring how to predict market trends. Stock price behaviour needs to be analysed as a sequence of discrete time data. The linear model is typically used to empirically test the effectiveness of the stock market, but the price time series is chaotic.

4 Proposed Model (Fig. 2)

4.1 Objectives

- To implement an expert system to predict the uncertainty of market risk.
- To propose a model that guarantees that a base amount will remain unchanged while marketing in the long run, thus increasing the efficiency of trade.
- To protect against fraud by ensuring that users abide by the norms and regulations set by governments for safe trading.

4.2 Mathematical Implementation

The mathematical implementation is carried out by using mathematics, operations, and logic. At the beginning, a statistical analysis is performed to analyse the data.
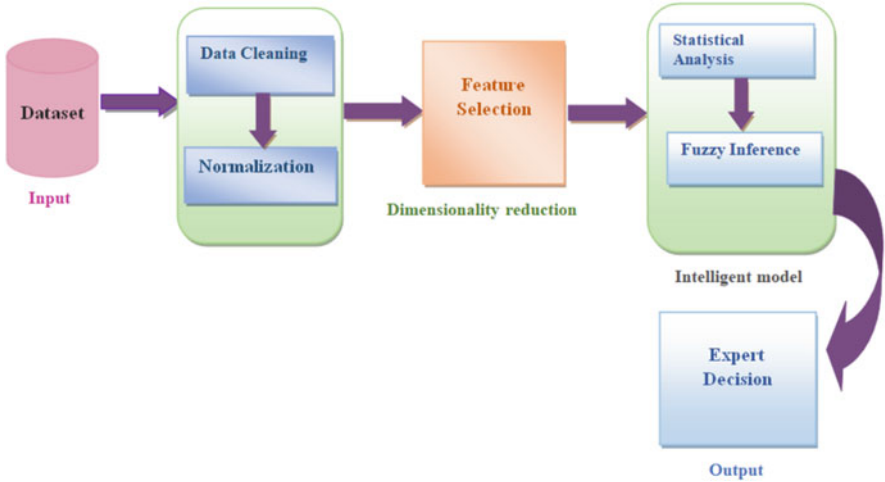
Fig. 2 Proposed Model

After analysing the data, a fuzzy inference is used to abate the uncertainty of market fluctuations.

4.2.1 Statistical Analysis

In this section, a sequential statistical analysis is performed. In this study, five companies are considered for validation. Through estimation, the expected outcome can be found in an interval that is predicted to include the unknown features of the companies' shares. The unknown features have some specific probabilities.

Let us assume that $C_1, C_2, C_3, \ldots, C_n$ constitute a random sample of companies from a set of companies. The unknown feature of a company is θ.

Two special limits, which serve as the lower and upper limits to the value of confidence, are denoted as L_1 and L_2. The confidence limit defined as

$$P\left(L_1 \leq \theta \leq L_2\right) = C \tag{1}$$

where C is the coefficient of the confidence interval.

The confidence limit for unknown features μ is considered as 95% for this research. Therefore, the normal distribution can be determined by using Eq. (2):

$$Z = \frac{\bar{x} - \mu}{\sigma / \sqrt{n}} \tag{2}$$

where $\bar{x}$ is the sample and μ and σ^2 / n are the mean and the variance, respectively.

In a bell curve, the area under the standard normal curve occupies 95%, and the critical region occupies 5%:

$P\left[-1.96 \leq \frac{\bar{x}-\mu}{\sigma/\sqrt{n}} \leq 96\right] = 0.95$. Here, 95% of the cases belong with the inequalities: $\left[-1.96 \leq \frac{\bar{x}-\mu}{\sigma/\sqrt{n}} \leq 96\right]$.

$\bar{x}-1.96\frac{\sigma}{\sqrt{n}} \leq \mu \leq \bar{x}+1.96\frac{\sigma}{\sqrt{n}}$. Thus, the interval $(\bar{x}-1.96\frac{\sigma}{\sqrt{n}}, \quad \bar{x}+1.96\frac{\sigma}{\sqrt{n}})$ is the 95% confidence interval for μ. The limit of μ is $\bar{x} \pm 1.96\frac{\sigma}{\sqrt{n}}$. This limit is modelled after hypothesis testing. During the hypothesis testing, H_0 and H_1 are the null hypothesis and an alternative hypothesis, respectively.

$$
\begin{aligned}
H_0 &: \{\theta \leq 0.95\} \\
H_1 &: \{\theta > 0.95\}
\end{aligned}
\tag{3}
$$

4.2.2 Fuzzy Inferences

Fuzzy logic is implemented in this research to control the uncertainty in the market value of the share units. Because the share price may fluctuate within a fraction of a second, it is very difficult to use for market prediction.

Assume that S is a set of blue-chip stocks. A new company C may or may not belong in S. The objective of this research is to determine whether the stocks of the new company C are blue-chip stocks.

The membership grade of C is $\mu_{\underset{\sim}{S}}(C)$ when C is mapped in S:

$$
\mu_{\underset{\sim}{S}} : (C) \rightarrow \{0, 1\}
\tag{4}
$$

where $c \in C$.

If the c does not belong in S, then it belongs in a complementary set of S.

$$
S \cap S' = \varphi
\tag{5}
$$

Therefore, it is determined that S and S' are disjoined sets. Considered two data sets, A and B, are in a universe of two discourses, U_1 and U_2, respectively. The multiple-input-single-output (MISO) rule base is applied with n rules.

$$
\text{If } A_{i1}, \text{ and } A_{i2}, \text{ and} \ldots. \text{and } A_{in}, \text{ then } B_i
\tag{6}
$$

where n is the number of input variables. The jth rules can be calculated by using Eq. (7):

$$
I_j = A_i \rightarrow B_i A_i = A_{i1} \cap A_{i2} \cap \ldots.. A_{in} = \cap_{j=1}^{n} A_{ij}
\tag{7}
$$

The summation of the implementation I_{Total} of n can be determined by using Eq. (8):

$$I_{\text{Total}} = R_1 \cup R_2. \cdots \cup R_N = \cup_{i=1}^{N} R_i = \cup_{i=1}^{N} A_i \rightarrow B_i \tag{8}$$

where R denotes the fuzzy relation of $U_1 \times U_2$.

5 Implementation

Implementation science is the study of techniques to improve the uptake, implementation, and translation of a study's findings into ordinary practices. The implementation of research seeks to recognise and work within the actual geographical areas of world situations, instead of looking at and manipulating conditions or removing effects.

5.1 Data Preparation

Stock market–related records are diverse, so first, similar works from a survey of economic studies are compared against stock price trends to conduct a statistical analysis in order to determine which guidelines to follow. After accumulating the records, the statistical structure of the data set is described, by following Shah et al. [35]. At this stage, the data collected from the general public and civic records are organised. A description of the data set includes the data shape, records, and tables in every class of statistics. Second, we compare similar works from the survey of economic studies to determine which instructions to follow for the collection of statistics. The data set is a group of inventory-request statistics from certain agencies (Akhtar et al. [36]). After gathering the information, we organised the records in the data set. Contributing to and benefiting from the market have by no means been essential or fundamental, because of the unpredictable nature of the market: Stocks, shares, and values are likely to quickly rise or fall in price. Recorded instability is likewise regarded as unpredictability from the instability in the actual charges for fundamental shares. As an alternative, it is represented by using a totally nonlinear dynamical framework.

In this chapter, the Nifty 50 data of columns and rows, which are data sets, are used to validate the proposed model. The information series approach is one of many critical approaches. It is carried out by filtering the statistics and preprocessing data; it is supported by training the model; and it requires checking the set of rules by using extraordinary sets of statistics. The data sets have been taken from Kaggle as real-time data sets: fifteen features from the Nifty 50 stock market records (2000–2021) have been reduced to six variables: date, low fee, excessive fee, commencing fee, final price, and extent. The data sets have been normalised,

and all the capabilities are represented in their diverse units. There is no statistical inconsistency but rather data integrity.

5.2 Data Cleaning and Data Preprocessing

We nowadays check whether statistics have a null cost or a price and whether unknown types or styles of records have been wiped clean, crammed, or stuffed into statistical formulations (Karim et al. [37]). For example, if it changed into a null cost from a corresponding characteristic, then it is checked for a variety of values, such as discrete and classifier costs. If it is a class price, then calculate the median cost; otherwise, calculate the mean cost, and place it in the null location.

5.2.1 Data Normalisation

Normalisation changes the values of various numeric columns within the data set to one unit of measure, which further enhances the overall performance of our model (Roshan Adusumilli [38]). Normalisation shows how to effectively organise structured statistics within a database. It also includes tables presenting, depicting, and establishing relationships between them and defining guidelines for relationships. Inconsistency, uncertainty, and redundancy can be assessed by tests based entirely on these rules, thus adding flexibility to the database.

5.3 Fuzzy Inference

Fuzzy rules are used to ensure that certain steps are taken in the event of ambiguity or uncertainty (Abbasi et al. [39]). It can rework many concepts, variables, and ambiguous and obscure systems into mathematical models and pave the way for argumentation, inference, manipulation, and decision-making in the event of uncertainty. A dynamic gadget that uses fuzzy sets, fuzzy logics, and/or analogous mathematical frameworks is referred to as a fuzzy machine. Some approaches can be used to construct a fuzzy inference device, wherein one or more inputs or antecedents can be used to generate one or more outputs or consequents. Technical indicators are used to reveal fee movements (adjustments in rates); they form the input parameters in the fuzzy machine. Inventory statistics are taken from a special inventory exchange and used for the work, as can be seen in Table 1. The past values of open, high, low, and ultimate charges and the volumes of a particular stock are recorded for a series of days and stored in a database to train the machine.

Table 1 Actual values for various features

Date	Opening rate	Low rate	High rate	Closing rate	Volume
(a) Actual values for Axis Bank					
02-04-2018	506.75	492.2	508.5	499.05	7,120,498
03-04-2018	494	493.7	506	501.55	6,053,789
04-04-2018	504.5	490.25	504.5	492.25	6,047,813
05-04-2018	498.4	495.2	504.65	503.1	6,682,587
06-04-2018	502.95	495.2	502.95	500.2	3,812,133
Standard deviation		2.1149468	2.08044		
(a) displays the actual values for the features of Axis Bank from 02-04-2018 to 06-04-2018, for which the standard deviation of the low rate and that of the high rate have been calculated					
(b) Actual values for Axis Bank					
09-07-2018	515.1	515.1	529.3	525.25	6,295,104
10-07-2018	527.95	522	534.9	532.55	5,281,760
11-07-2018	532.4	530.8	542.35	536.55	6,432,494
12-07-2018	539.7	535.3	545.7	538.35	4,128,221
13-07-2018	539.7	519.45	541	523	4,169,212
Standard deviation		8.3132	6.52687		
(b) displays the actual values for the features of Axis Bank from 09-07-2018 to 13-07-2018, for which the standard deviation of the low rate and that of the high rate have been calculated					
(c) Actual value for Axis Bank					
03-12-2018	632	621.6	633	624.8	5,914,250
04-12-2018	624.95	618.6	627.55	623.25	8,382,005
05-12-2018	620.9	613.4	623.95	614.95	9,728,083
06-12-2018	610.25	598.25	613.85	600.05	7,519,849
07-12-2018	602.4	592.35	609.5	603.95	12,262,327
Standard deviation		12.8738	9.71279		
(c) displays the actual values for the features of Axis Bank from 03-12-2018 to 07-12-2018, for which the standard deviation of the low rate and that of the high rate have been calculated					

In this chapter, the above data set tables are taken from Axis Bank and are related to the stock market features that are used for building a model. They are divided into three sets of random samples taken over a period of eight months, and afterwards, the respective standard derivations of all fifteen random samples are calculated

5.3.1 Axis Bank

The Table 2 depicts the mean of the standard deviation calculated in Table 1 at three intervals of random samples taken over a period of eight months in 2018. The mean of the standard deviation of the low and high rates has been calculated to be 4.61682, which is used to establish a model.

Table 2 Mean and standard deviation for three intervals (Axis Bank)

Intervals	Low rate	High rate
1st interval	2.1149468	2.08044
2nd interval	8.3132	6.52687
3rd interval	12.8738	9.71279
Standard deviation	5.40016	3.83348
Mean		4.61682

Table 3 Mean and standard deviation for three intervals (Tata Steel)

Intervals	Low	High
1st interval	10.9415	12.6731
2nd interval	6.81887	9.12113
3rd interval	7.12811	5.16207
Standard deviation	2.29614	3.72795
Mean		3.012045

Table 4 Mean and standard deviation for three intervals (Titan)

Intervals	Low	High
1st interval	25.8135	17.1588
2nd interval	13.7267	8.46692
3rd interval	15.8859	17.133
Standard deviation	6.44066	5.01083
Mean		5.725745

5.3.2 Tata Steel

Similarly, the standard deviations of the low and high rates have been calculated for Tata Steel's stock market features at three intervals, which appear in Table 3, depicting the mean of the standard deviation calculated at three intervals of random samples taken over a period of seven months in 2018. The mean of the standard deviation of the low and high rates has been calculated to be 6.61882 and 3.012045, which is used to establish a model.

5.3.3 Titan

Similarly, the standard deviations of the low and high rates have been calculated for Titan's stock market features at three intervals, which appear in Table 4, depicting the mean of the standard deviation calculated at three intervals of random samples taken over a period of twelve months in 2018–2019. The mean of the standard deviation of the low and high rates has been calculated to be 5.725745, which is used to establish a model.

Table 5 Threshold values of three companies' data sets

Company	Low rate	High rate
Axis Bank	5.40016	3.83348
Tata Steel	2.29614	3.012045
Titan	6.44066	5.725745
Mean	4.71232	4.190423

5.3.4 Threshold Value

The threshold value calculations for the dataset features of low rate and high rate for the AXIS BANK, TATASTEEL, and TITAN firms are shown in Table 5. The results are 4.71232 for the mean low rate across the three companies and 4.190423 for the high rate, which are then utilised to create fuzzy rules.

5.3.5 Parameters

Standard deviation: Σ
Lesser than: $<$
Greater than: $>$
Greater than or equal to: $\geq$
Lesser than or equal to: $\leq$
Equal to: $=$

5.3.6 Fuzzy Rules

Fuzzy rules are developed on the basis of the threshold values in Table 5.

Rule 1: *IF* s.d Σ of low rate is <4.71232 *THEN* returned amount will be $<$ assured amount (loss).

Rule 2: *IF* s.d Σ of low rate is $=$ to 4.71232 *THEN* returned amount will be $=$ assured amount.

Rule 3: *IF* s.d Σ of low rate is >4.71232 and <4.190423 *THEN* returned amount will be $\geq$ assured return.

Rule 4: *IF* s.d Σ of high rate is >4.190423 *THEN* the return amount will be surely be $>$ assured return (profit).

6 Results and Discussion

In Fig. 3, the graphs depict the opening rate, low rate, high rate, and closing rate against the volume for Axis Bank, taken from fifteen random samples at three intervals, which are further used for computation.

In Fig. 4, the graphs depict the opening rate, low rate, high rate, and closing rate against the volume for Tata Steel, taken from fifteen random samples at three intervals, which are further used for computation.

Fig. 3 Actual values for Axis Bank over a period of seven months. (**a**) Actual value for Axis Bank (2-04-2018 to 06-04-2018). (**b**) Actual value for Axis Bank (09-07-2018 to 13-07-2018). (**c**) Actual value for Axis Bank (03-12-2018 to 07-12-2018)
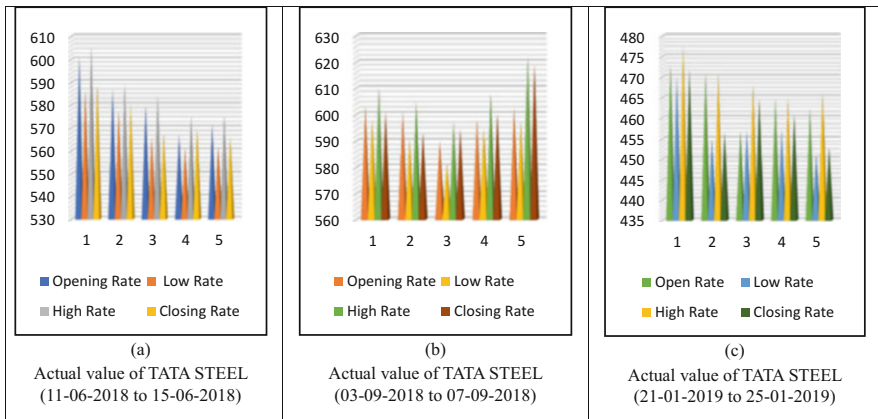
Fig. 4 Actual values of Tata Steel over a period of seven months. (**a**) Actual value for Tata Steel (11-06-2018 to 15-06-2018). (**b**) Actual value for Tata Steel (03-09-2018 to 07-09-2018). (**c**) Actual value for Tata Steel (21-01-2019 to 25-01-2019)
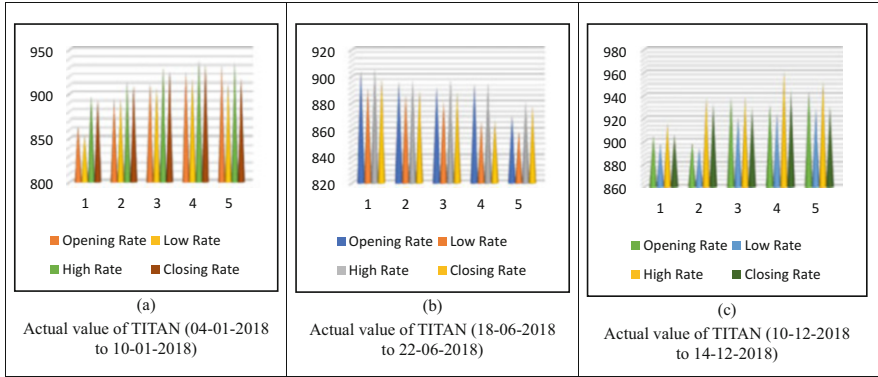
Fig. 5 Actual values for Titan over a period of twelve months. (**a**) Actual value for Titan (04-01-2018 to 10-01-2018). (**b**) Actual value for Titan (18-06-2018 to 22-06-2018). (**c**) Actual value for Titan (10-12-2018 to 14-12-2018)
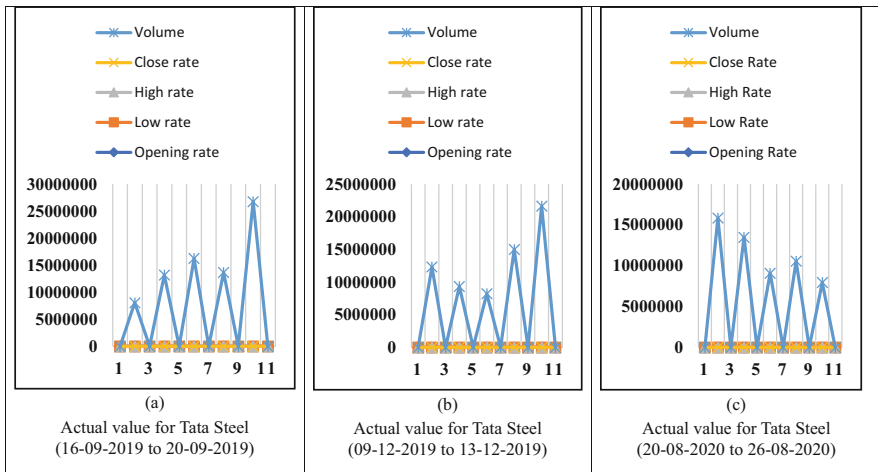

Fig. 6 Actual values the Tata Steel over a period of eleven months. (**a**) Actual value for Tata Steel (16-09-2019 to 20-09-2019). (**b**) Actual value for Tata Steel (09-12-2019 to 13-12-2019). (**c**) Actual value for Tata Steel (20-08-2020 to 26-08-2020)

In Fig. 5, the graphs depict the opening rate, low rate, high rate, and closing rate against the volume for Titan, taken from fifteen random samples at three intervals, which are further used for computation.

In Fig. 6, the graphs depict the opening rate, low rate, high rate, and closing rate against the volume for Tata Steel, taken from fifteen random samples at three intervals, which are further used for computation.

In Fig. 7, the graphs depict the opening rate, low rate, high rate, and closing rate against the volume for Titan, taken from fifteen random samples at three intervals, which are further used for computation.
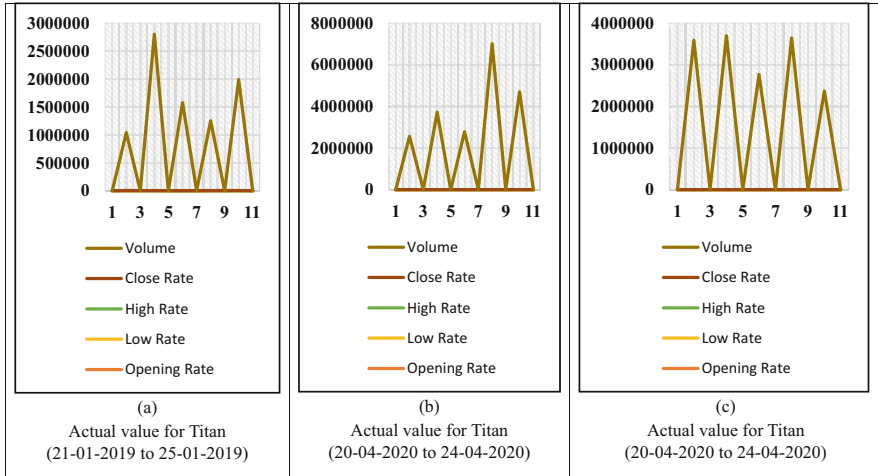
Fig. 7 Actual values for Titan over a period of sixteen months. (**a**) Actual value for Titan (21-01-2019 to 25-01-2019). (**b**) Actual value for Titan (20–04-2020to 24-04-2020). (**c**) Actual value for Titan (20-04-2020 to 24-04-2020)
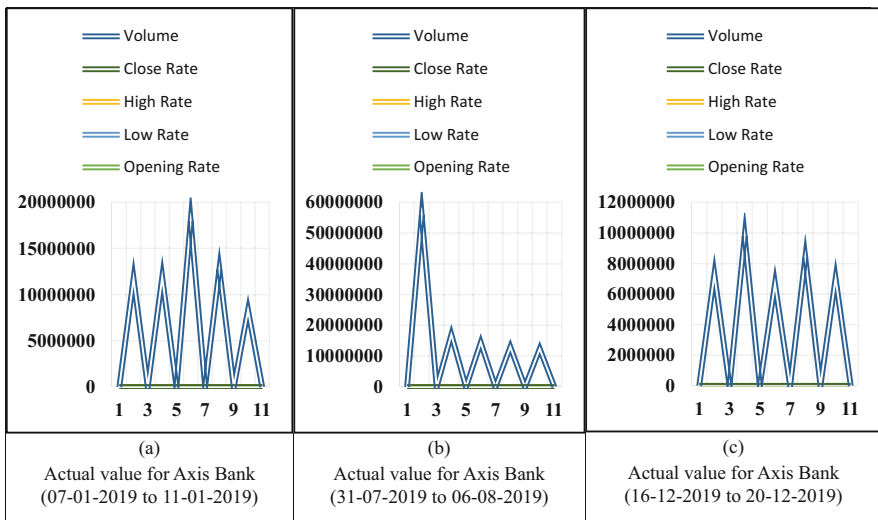

Fig. 8 Actual values for Axis Bank over a period of twelve months. (**a**) Actual value for Axis Bank (07-01-2019 to 11-01-2019). (**b**) Actual value for Axis Bank (31-07-2019 to 06-08-2019). (**c**) Actual value for Axis Bank (16-12-2019 to 20-12-2019)

In Fig. 8, the graphs depict the opening rate, low rate, high rate, and closing rate against the volume for Axis Bank, taken from fifteen random samples at three intervals, which are further used for computation.

Table 6 Values of accuracy, recall, and precision for various training and testing strategies

Training and testing strategy	Strategy 1 (50%–50%)	Strategy 2 (60%–40%)	Strategy 3 (70%–30%)	Strategy 4 (80%–20%)	Average
Accuracy	92.78	96.88	96.46	95.55	95.42
Recall	91.22	94.06	94.54	94.82	93.66
Precision	95.76	96.24	96.04	97.26	96.325
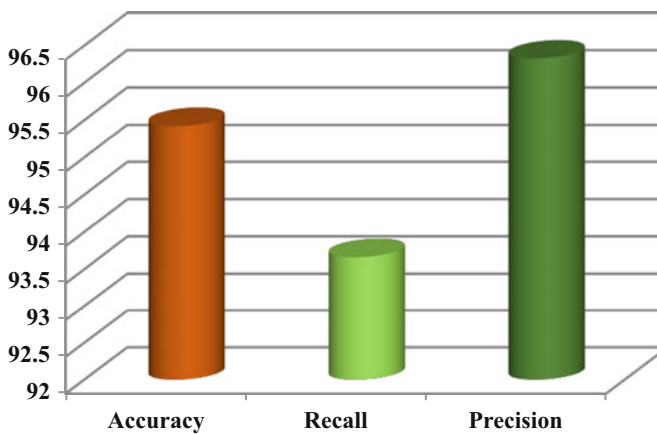
Fig. 9 Visualisation of the performance metrics

6.1 Performance Analysis

In this section of the chapter, an analysis of performance is carried out on real-time data sets. Several parameters of the confusion matrix [40] are used to evaluate the performance of the proposed model. Using the provided data set, a fourfold cross validation [41] is carried out. The data set is not balanced, so error correction is required to improve the model's accuracy. The proposed model is used as a monitoring system in the healthcare sector, and recall and precision are key performance indicators. Table 6 features the performance metrics for four strategies, the ratios from the training and testing data, and the accuracy, recall, and precision values for various training and testing strategies.

The average values of the performance metrics are visualised in Fig. 9.

7 Conclusion

One of the most significant stages in the financial market is online share trading, where people can invest their money as capital on long- or short-term bases.

The purpose of the research in this chapter is to discover an expert solution that yields a guaranteed return on investment. Numerous blue-chip businesses are in the stock market. Blue-chip companies typically trade on a significant stock exchange, such as the New York Stock Exchange, though this is not strictly required. They are frequently listed in market indexes. The stocks are very liquid because both individual and institutional investors frequently trade them on the market. Because there will always be a buyer on the other side of the transaction, investors who suddenly need money can sell their stock with confidence. Thanks to this research, an intelligent framework for choosing investment firms that offers guaranteed returns can be suggested. To create a pilot surveyor to examine the data set, a statistical data analysis is carried out. To reduce the uncertainty in market fluctuations, fuzzy logic is used. The accuracy, recall, and precision values for the model's performance are 95.42, 93.66, and 96.325, respectively. In the future, this model might incorporate blockchain technology to increase the security of online trading.

References

1. Shen, J. & Omair Shafiq, M. (2020). Short-term stock market price trend prediction using a comprehensive deep learning system. Springer Open, *Journal of Big Data, 7*, Open access, Article number: 66.
2. Tanwar, S., Patel, N. P., Patel, S. N., Patel, J. R., Sharma, G., & Davidson, I. E. (2020). Deep learning-based cryptocurrency price prediction scheme with inter-dependent relations. *IEEE Access, 9*, 1345–1356.
3. Liu, H., Qi, L., & Sun, M. (2022, June). Short-term stock price prediction based on CAE-LSTM method, ResearchGate. Hindawi access, *Wireless Communications and Mobile Computing, 2022*(S1), 1–7. https://doi.org/10.1155/2022/4809632
4. Banerjee, S., Neha Dabeeru, R., & Lavanya. (2020). Stock market prediction. *International Journal of Innovative Technology and Exploring Engineering (IJITEE), 9*(9), 506–509. ISSN: 2278–3075.
5. Li, H., Dagli, C.H., & Enke, D. (2007, May). Short-term stock market timing prediction under reinforcement learning schemes. In *IEEE Xplore Conference on Approximate Dynamic Programming and Reinforcement Learning. ADPRL 2007. IEEE International Symposium*. pp. 233–237. https://doi.org/10.1109/ADPRL.2007.368193
6. Rao, P. S., Srinivas, K., & Krishna Mohan, A. (2020, May). A survey on stock market prediction using machine learning techniques. https://doi.org/10.1007/978-981-15-1420-3_101
7. Sarkar, M., Roy, A., Badr, Y., Gaur, B., & Gupta, S. (2021). An intelligent music recommendation framework for multimedia big data: A journey of entertainment industry. *Studies of Big Data, Springer Nature Singapore, 2*, 39–67.
8. Roy, A., Banerjee, S., Sarkar, M., Darwish, A., Elhosen, M., & Hassanieen, A. E. (2018). Exploring New Vista of intelligent collaborative filtering: A restaurant recommendation paradigm. *Journal of Computational Science, Elsevier, 27*, 168–182.
9. Berra, D. (2018, January). Cross-validation, ResearchGate. Reference module in life sciences, pp. 1–3. https://doi.org/10.1016/B978-0-12-809633-8.20349-X
10. Yang, W. (2007). Stock price prediction based on fuzzy logic. In *2007 International Conference on Machine Learning and Cybernetics*, pp. 1309–1314. https://doi.org/10.1109/ICMLC.2007.4370347

11. Vo, M. T., Vo, A. H., Nguyen, T., Sharma, R., & Le, T. (2021). Dealing with the class imbalance problem in the detection of fake job descriptions. *Computers, Materials & Continua, 68*(1), 521–535.

12. Sachan, S., Sharma, R., & Sehgal, A. (2021). Energy efficient scheme for better connectivity in sustainable mobile wireless sensor networks. *Sustainable Computing: Informatics and Systems, 30*, 100504.

13. Ghanem, S., Kanungo, P., Panda, G., et al. (2021). Lane detection under artificial colored light in tunnels and on highways: An IoT-based framework for smart city infrastructure. *Complex & Intelligent Systems*. https://doi.org/10.1007/s40747-021-00381-2

14. Sachan, S., Sharma, R., & Sehgal, A. (2021). SINR based energy optimization schemes for 5G vehicular sensor networks. *Wireless Personal Communications*. https://doi.org/10.1007/s11277-021-08561-6

15. Priyadarshini, I., Mohanty, P., Kumar, R., et al. (2021). A study on the sentiments and psychology of twitter users during COVID-19 lockdown period. *Multimedia Tools and Applications*. https://doi.org/10.1007/s11042-021-11004-w

16. Azad, C., Bhushan, B., Sharma, R., et al. (2021). Prediction model using SMOTE, genetic algorithm and decision tree (PMSGD) for classification of diabetes mellitus. *Multimedia Systems*. https://doi.org/10.1007/s00530-021-00817-2

17. Priyadarshini, I., Kumar, R., Tuan, L. M., et al. (2021). A new enhanced cyber security framework for medical cyber physical systems. *SICS Software-Intensive Cyber-Physical Systems*. https://doi.org/10.1007/s00450-021-00427-3

18. Priyadarshini, I., Kumar, R., Sharma, R., Singh, P. K., & Satapathy, S. C. (2021). identifying cyber insecurities in trustworthy space and energy sector for smart grids. *Computers & Electrical Engineering, 93*, 107204.

19. Singh, R., Sharma, R., Akram, S. V., Gehlot, A., Buddhi, D., Malik, P. K., & Arya, R. (2021). Highway 4.0: Digitalization of highways for vulnerable road safety development with intelligent IoT sensors and machine learning. *Safety Science, 143*, 105407, ISSN 0925-7535.

20. Sahu, L., Sharma, R., Sahu, I., Das, M., Sahu, B., & Kumar, R. (2021). Efficient detection of Parkinson's disease using deep learning techniques over medical data. *Expert Systems*, e12787. https://doi.org/10.1111/exsy.12787

21. Suvarnapathaki, S. (2022). Using unstructured data with structured data for segmentation of nifty 50 stocks. *JETIR, 9*(6).

22. Sanger, W., & Warin, T. (2016). High frequency and unstructured data in finance: An exploratory study of Twitter. *JGRCS 2016, 7*(4).

23. Tetlock, P. C., Saar-Tsechansky, M., & Macskassy, S. (2008). More than words: Quantifying language to measure firms' fundamentals. *The Journal of Finance, 63*(3), 1437–1467.

24. Peji'c Bach, M., Krsti, Ž., Seljan, S., & Turulja, L. (2019). Text mining for big data analysis in financial sector - A literature review. *Sustainability, 11*, 1277. https://doi.org/10.3390/su11051277

25. Lima, L., Portela, F., Santos, M. F., Abelha, A., & Machado, J. (2015). *Big data for stock market by means of mining techniques*. Springer Science and Business Media. https://doi.org/10.1007/978-3-319-16486-1_67

26. Santhosh Baboo, L., & Renjith Kumar, P. (2013). Next generation data warehouse design with big data for big analytics and better insights. *Global Journal of Computer Science and Technology, 13*(7).

27. Morshadul Hasan, M., Popp, J., & Olah, J. (2020). Current landscape and influence of big data on finance. *Journal of Big Data, 7*, Article No: 21.

28. Choi, T.-M., & Lambert, J. H. (2017, August 11). *Advances in risk analysis with big data.* https://doi.org/10.1111/risa.12859

29. Razin, E. (2015, December 3). Big buzz about big data: 5 ways big data is changing finance. *Forbes*.

30. Jaweed, M. D., & Jebathangam, J. (2018). Analysis of stock market by using Big Data Processing Environment. *International Journal of Pure and Applied Mathematics, 119*(10), 81–86.

31. Mallon, S. *Big data analytics has potential to massively disrupt the stock market*. https://www.smartdatacollective.com/big-data-analytics-has-potential-to-massively-disrupt-stock-market/
32. The Pipeline, The ZoomInfo. https://pipeline.zoominfo.com/marketing/dynamic-data
33. Yang, P., & Hou, X. (2022). Research on dynamic characteristics of stock market based on big data analysis. *ResearchGate, Hindawi, Discrete Dynamics in Nature and Society, 2022*, Article ID 8758976, 1–8. https://doi.org/10.1155/2022/8758976
34. John, J., & Joseph, B. (2022). Stock price prediction using LSTM with dynamic data sets. *Proceedings of the National Conference on Emerging Computer Applications (NCECA), 4*(1), 625–628. https://doi.org/10.5281/zenodo.6938228
35. Shah, A., Patel, P., & Vora, D. (2020). Dynamic approach to stock trades using ML techniques. *International Research Journal of Engineering and Technology (IRJET), 07*(12), 608–610, e-ISSN: 2395-0056.
36. Akhtar, M. M., Zamani, A. S., Khan, S., AliShatat, A. S., Dilshan, S., & Samdan, F. (2022). Stock market prediction based on statistical data using machine learning algorithms. *Journal of King Saud University – Science, Science Direct, 34*(4), 101940.
37. Karim, R., Alam, M. K., & Hossain, M. R. (2021, August). Stock market analysis using linear regression and decision tree regression. In *IEEE Conference: 2021 1st International Conference on Emerging Smart Technologies and Applications (eSmarTA)*. https://doi.org/10.1109/eSmarTA52612.2021.9515762
38. Adusumilli, R. (2019). Machine learning to predict stock prices, Published in towards Data Science, Dec 26.
39. Abbasi, E., & Abouec, A. (2008). Stock price forecast by using neuro-fuzzy inference system. *World Academy of Science, Engineering and Technology, 46*, 320–323.
40. Visa, S., Ramsay, B., Ralescuand, A., & van der Knaap, E. (2011). Confusion matrix-based feature selection. In *Proceedings of the 22nd Midwest Artificial Intelligence and Cognitive Science Conference*, Cincinnati, pp. 16–17.
41. Jung, Y., & Hu, J. (2015). A k-fold averaging cross-validation procedure. *Journal of Nonparametric Statistics, 27*(2), 1–13.