



Stock prediction and mutual fund portfolio management using curve fitting techniques

Giridhar Maji¹ · Debomita Mondal² · Nilanjan Dey³ · Narayan C. Debnath⁴ · Soumya Sen²

Received: 26 April 2020 / Accepted: 10 November 2020 / Published online: 2 January 2021
© Springer-Verlag GmbH Germany, part of Springer Nature 2021

Abstract

Investment in the share market helps generate more profit than the other financial instruments but has the threat of market risk that might lead to a high loss. This risk factor refrains many potential investors from investing in the share market directly. Instead, they invest in different mutual funds that are being managed by experienced portfolio managers. To avoid the risk factors and increase the gain, they put the accumulated capital in multiple stocks. They need to perform many calculations and predictions to overcome the uncertainties and unpredictability and need to ensure higher gains to the investors of that mutual fund. In this research work initially, a data mining based approach employs a curve fitting/regression technique to forecast the individual stock price. Based on the above analysis, we propose a framework to diversify the investment of the capital fund. This method employs buy and hold strategy using both statistical features and basic domain knowledge of the share market. The proposed framework distributes the capital first, by distributing sector-wise, and then for each sector, investing company-wise, as a diversified approach among different stocks for higher return but maintaining lower risks. Experimental results show that the proposed framework performs well and generates a good yield compared to some benchmark and ranked mutual funds in the Indian stock market.

Keywords Stock market analysis · Sector-wise stock selection · Stock price prediction · Curve fitting · Mutual fund portfolio management

1 Introduction

A company's future growth and hence return to the shareholders depends on a lot of internal and external factors. A mutual fund (MF) manager collects funds from many different clients and then invests in many different stocks, bonds, and many other money market instruments to earn a good return. The size of the MF industry in India is around INR 26.07 trillion as of June 30, 2020, and the corpus size has increased up to 4 times in the last 10 years (Association of Mutual Funds in India, 2020). It implies a whopping 400% growth in just over 10 years that indicates people's interest to invest in MFs. This sustained trend shows that general people (retail investors) are investing heavily in different MFs. One of the reasons for interest in an MF is that most investors have the appetite for higher returns but do not want to take the risk to invest in the share market directly. They feel it safer to invest in the market indirectly via some trusted MFs. Another reason for interest in MFs is the falling interest rate on the bank's fixed deposits. Again, due to the high inflation,

✉ Soumya Sen
iamsoumyasen@gmail.com

Giridhar Maji
Giridhar.Maji@gmail.com

Debomita Mondal
debomita.mondal@gmail.com

Nilanjan Dey
neelanjan.dey@gmail.com

Narayan C. Debnath
ndebnathc@gmail.com

¹ Department of Electrical Engineering, Asansol Polytechnic, Asansol 713302, India

² A.K. Choudhury School of Information Technology, University of Calcutta, Salt Lake 700106, India

³ Department of Computer Science and Engineering, JIS University, Kolkata, India

⁴ School of Computing and Information Technology, Eastern International University, Thu Dau Mot, Vietnam

the effective yield from fixed deposit becomes negligible. Hence general people look into the different investment alternatives that will give reasonable returns with minimum possible uncertainties. Generally, the investment in individual stock may earn a good profit, but it also risks an unbounded loss. Therefore, individual investors often avoid investing in the share market directly. Instead, they prefer MFs because these are managed by expert professionals (portfolio managers) and invest in multiple stocks belonging to different sectors to mitigate the risk factors.

There are many different types of *mutual funds* in the market with varied stated objectives and investment time-horizons. The fixed income MFs (aka debt funds) generally invest in secured instruments that guarantee capital protection but yield less return. Again, the equity-based MFs are the high-risk, high-return group where most of the capital goes into buying shares of companies listed in the stock exchanges. The *hybrid* and *balanced* MFs are medium risk funds where total corpus gets invested into risk-free, secured bonds, and high-risk equity stocks.

There are many other varieties of MFs such as (1) a *gold fund* that mainly invests in gold, (2) the *gilt funds* that earn from investing in fixed interest government securities, (3) the *Exchange Traded Funds* (ETF) (Agapova 2011), where a proportion of different stocks are combined and listed in the stock market for trading like a single stock, (4) *Equity-linked savings scheme* (ELSS) with a minimum lock-in period of 3 years, during which investors can not redeem. It also provides certain tax benefits.

In this research work, we consider only equity stocks though other money market instruments may very well be incorporated. Here we take different company stocks from varied business domains. This research work is generic for analysis on all the share markets and mutual funds across the world. We have chosen the Indian share market for the experiments and analysis purposes. The *SENSEX* is the indexing measure of the *Bombay Stock Exchange* (BSE) and comprises thirty well-established and well-performing companies (BSE 2019a). The *NIFTY 50* is a well-known index representing the *National Stock Exchange* (NSE) and is India's benchmark stock market index for the Indian equity market. It covers around twenty-two sectors of the Indian economy and offers investment manager exposure to the Indian market in one portfolio. Some of the companies in *NIFTY* from different sectors are ACC Ltd. (Cement), Asian Paints (Paints), Axis Bank (Private Bank), BHEL (Electric Equipment), Bharti Airtel (Telecommunication), Cipla (Pharmaceuticals), Coal India (Mining and Metals), ITC (Cigarette Hotels), Infosys (IT and Software), L&T, ONGC (Oil and Gas), SBI (public bank), TCS (IT and Software), Tata Steel (Steel), Zee entertainment (Entertainment), etc. (NSE 2019). Most of the above stocks also belong to *SENSEX*. We understand that the *SENSEX* covers almost

all the industrial sectors, and therefore tracking its performance will give the pulse of the whole stock market.

Historical stock price movement data, as well as many other time-series data related to the internal as well as external factors associated to a particular stock or a group of similar stocks (like banking stocks, Metal stocks, automobile stocks, IT stocks, etc.), can be analytically processed by different data mining tools (Liao et al. 2012; Hajizadeh et al. 2010; Zhang and Zhou 2004) to predict the future movement of the stock market. However, there is no guaranty that the result of the analysis will be accurate. As many external factors are involved, many well-performing stocks may suddenly become loss-making. One of the most convincing examples is the spreading of Coronavirus Disease (COVID-19) (Fong et al. 2020) across the world. Due to lockdown in many parts of the world, economic activities reduced drastically, resulting in a rapid fall of share prices across the globe. COVID-19 was an unplanned event and affected almost all business sectors. However, some pharmaceutical companies, gloves makers, sanitizer makers, etc. have benefited due to the sudden increase in demand for their products. It implies that it is safe to invest in a different sector to reduce the risk of investment. Even if a company from a well-performing business sector may go through some financial loss, bankruptcy, or scams that can lead to a sharp fall in the share price. It suggests investing in multiple companies belong to the same business domain. It becomes difficult for many individual investors to keep track of the several companies belonging to the different sectors and maintain a diversified portfolio. Henceforth to stay invested in the stock market with a diversified portfolio, a mutual fund is the best option for many retail investors. It gives different AMC's (Asset Management Companies) opportunities to introduce new mutual funds to the market with varying risk factors, return expectations, etc. The competition is increasing between different AMC's to attract more investors. The role of the portfolio manager is getting challenging day-by-day. In this research work, we focus on developing a framework for the portfolio manager to maintain a diversified portfolio with less risk and high return probability. The main contributions of the paper are as follows.

- At first, we do a business level analysis on the individual stock to perform a data mining-based approach by applying curve fittings (regression) to find the accuracy of the prediction curve.
- In the next level, based on these predictions, a framework is derived for the portfolio managers to invest in diversified stocks to increase the return on investment.

The rest of the paper is organized as follows. Related works are discussed in Sect. 2, followed by the proposed framework and methodology employed for mutual fund portfolio

management in Sect. 4. Experimental results with BSE stock market data are presented in Sect. 5. Finally, the study is concluded in Sect. 6 with the future direction of work.

2 Related work

A large number of studies have been conducted on stock exchange index movement forecast and individual stock price prediction using data mining and statistical analysis (Chikhi and Diebolt 2010). Stock market movement and the impact of different intrinsic and external factors such as gold price, crude oil price, political stability, dollar exchange rate, bank interest rate, etc. on the stock market are discussed in Mondal et al. (2018). A review of the stock market is presented in Rusu and Rusu (2003), Atsalakis and Valavanis (2009) and Preethi and Santhi (2012). Authors have identified three different categories of analysis: fundamental analysis, technical analysis, and efficient market theory.

1. The fundamental analysis performed on the basis of macroeconomic factors such as prevailing interest rate, inflation data, industry/sector a company belongs to, profit/loss of the company, the amount of dividend paid, brand value, crude oil price (Zhu et al. 2014), gold price (Weng et al. 2020), foreign capital inflow etc. It employs statistical models and uses financial algorithms to the financial statements and other periodical market information to estimate the share price of the company.
2. The underlying assumption of the technical analysis is that stock price variation itself incorporates the effects of all economic, political (Narang 2015), financial, and external factors. Technical analysis studies the short-term changes in stock price with the expectation that historical stock prices behavior will continue in the near future. Accurate and clean historical data such as days closing price, the dividend paid, bonus share issued, split ratio, buyback of shares, etc., are abundant from stock exchanges. Hence, it becomes promising to employ data mining techniques to predict the future trends of stock prices for investors on these datasets.
3. Important properties of the efficient market theory say (1) market prices are known to all stakeholders as soon as they change, (2) all related market information is freely available to all participants, (3) share prices reflect their fundamental values, and (4) stock prices follow a random walk.

Many neuro-fuzzy (Asadi 2019; Chandar 2019) models were proposed and different techniques like *Artificial Neural Network* with backpropagation (Mingyue et al. 2016; Singh and Mohanty 2018), *Recurrent Neural Network* (RNN) (Zhao et al. 2020; Ghosh et al. 2019), evolutionary techniques like

genetic algorithms (Weng et al. 2020; Ding et al. 2020), the *firefly algorithm* combined with the *support vector regression* (Zhang et al. 2019), the *linear regression* (Altay and Satman 2005; Ji et al. 2020) etc. were applied to predict stock prices. Zhang and Zhou (2004) reviewed data mining in the context of financial application from both technical and application perspectives. Then they compared different data mining techniques to solve financial problems and discuss important data mining issues involved in specific financial applications. Most of the above models other than regression are complex in design, use many parameters which are difficult to select, model settings such as the number of hidden layers, initial values of variables, use of optimizers, etc. They require a very long training phase, and the problem of over-fitting or under-fitting exists too.

Chandar (2019) utilized *discrete wavelet transform* (DWT) to decompose time series data and combined it with an adaptive fuzzy inference system to forecast the closing price of a stock. Data visualization (Moere and Offenhuber 2009) is important to represent the movement of the stocks and to understand the performance using the visual tools. This aspect enables the ambient computing (Curran 2011) by integrating cognitive processing with the visualization for better human interaction and understanding of the system.

Gottschlich and Hinz (2014) proposed a *decision support system* design that enables investors to include the crowd's recommendations in their investment decisions and use it to manage a portfolio. Here a serious concern is the presence of noise in the data. Common investors (here the crowd) have much less in-depth knowledge regarding the intricacies of the stock market and stock price movement based on sporadic rumors or speculation does not yield stable growth or returns. Hence it may provide higher gains for a very short period but may not be a beneficial tool for a stable mutual fund portfolio management.

Qian and Rasheed (2007) employed the *Hurst exponent* to select a time window when prediction probability would be higher and then used different machine learning techniques to be trained with data from that time period. They used different heuristics to generate the parameters for classification and tested on the *Dow Jones Industrial Average* index to achieve an average accuracy of 65%. In the context of the present study, it would not be wise to adapt as an MF portfolio that runs throughout the year, and it can not be open to investment for a selected time window. Also, the complexity of the above method will be rigorous when used for a bucket of different stocks.

Park et al. (2019) used stock indices movement data of eight countries and used them to predict S&P index movement. They have considered a total of 10 years day's closing index values and calculated their movement as *upward* and *downward* to use different binary classifiers such as *support*

vector machine (SVM), *k*-nearest neighbor (KNN), probabilistic neural network (PNN) etc. Ananthi and Vijayakumar (2020) have recently applied KNN based regression model on different market indicator like *MACD*, *RSI*, *Bollinger bands*, and *candle stick* patterns to generate buy/sell signals. The main limitation here is the forecasting horizon as the above work can only predict for a few days in advance, but we observe that regression based tools yield comparable prediction without large computational burden.

Maji et al. (2017) considered lagged correlation among the upward and downward movement of a sectoral index with some day-lag and applied association rule mining to predict future movement with a forecasting horizon the same as the day-lag. Authors have used *Association Rule Mining* (ARM) to find the correlation between different stock prices and generate association rules between different stock prices. Baralis et al. (2017) employed frequent itemset mining to supporting *buy-and-hold* investors in technical analyses by automatically identifying promising sets of high-yield yet diversified stocks to buy. Both of the above works forecast for a very limited time window, a few days in the future, and most useful for short term traders.

A stock market portfolio recommender system has been proposed in (Paranjape-Voditel and Deshpande 2013). Sankar et al. (2015) proposed a trust-based stock recommender system utilizing social network analysis techniques. They have utilized the stock holding and buying/selling information from other trusted mutual funds to recommend stocks. This approach is entirely dependent on the collective wisdom of other fund managers. Some authors propose recommender systems based on users' contextual short term information and sentiment analysis (Richthammer and Pernul 2018; Gottschlich and Hinz 2014). All of the stock recommender systems are again useful to individual regular traders for short term gains.

Majority of the research works discussed above focused on the share market directly by forecasting the future price or movement direction. In almost all the cases (except Maji et al. 2017; Paranjape-Voditel and Deshpande 2013), many financial indicators (short-term, long-term, macro-economic, and micro-economic parameters) have been used to predict the future price.

Our framework is based on the assumption that historical stock prices incorporate the impact of all such parameters. Hence, we simplify our framework by only considering historical stock prices as a proxy to all such financial parameters. A point to note is that investors who invest in the share market look into the closing price of different periods to check the growth or decay for that period. The closing price of the stock is the standard representation of the stocks for historical analysis.

We observe that financial time series regression tools are the most widely used and simple to compute. Many recent

studies utilized regression techniques in stock price forecasting. To keep the computations simple and easy, without compromising on forecasting accuracy, we have employed regression tools in the initial prediction phase of our proposed framework of MF portfolio management.

3 Problem statement and contribution

Formation of a mutual fund portfolio to avoid the risk of investment on individual/few stocks by investing in multiple stocks belonging to different business sectors. It is difficult to manage and analyze many stocks at a time. Henceforth a framework is required to work with multiple stocks from various business sectors to increase the profit over time but reducing the risk factors.

Contributions

- Analysis is performed only on the closing stock price but it includes the effect of various parameters such as dividend, bonus, split, etc. This simplifies the overhead of analysing with multiple parameters and avoids the use of complex tools.
- Diversified portfolio management by analyzing multiple stocks (company-wise) from different business sectors (sector-wise).
- Analysis and comparison of the different business sectors performance wise.
- A systematic portfolio management framework to increase the return over a longer period but minimizing the risks.
- Flexible to support the analysis on individual stock, group of stocks, business domains and that for any time period.

4 Proposed framework to forecast and allocate fund across sectors

This section derives a framework for portfolio managers to invest for many different time horizons in stocks to maximize investors' gain. Here, the analysis is done on multiple sectors to ensure diversified portfolio management to reduce investors' risk.

As already discussed that there are various perspectives of investors while investing their funds to gain more profit. Share markets are highly sensitive, and there are numerous factors directly or indirectly associated with it that control the market sentiment (Gottschlich and Hinz 2014). A group of investors foresees an appreciation in stock valuation so, they buy stocks at the current market price and expect to sell them in future and book profit. Again, at the same time, another group of investors assumes that stock prices will fall

in the future so, they sell at the current market price, and in the future, they may buy again at dips. In reality, there exists simultaneously many sellers and buyers in the market due to the different perception of future stock valuation. In the future, any one group of investors will benefit, and others may have to incur a loss. Many different statistical techniques and time-series analyses have been used with mixed success. We aim to use well-tested regression techniques as a starting tool to screen individual stocks along with domain knowledge in bucketing company stocks into industrial sectors. Therefore, the objective is to propose an analytical approach to predict the share price of different companies and invest total capital across various diversified sectors to earn better profit percentages and mitigate the overall risk. Moreover, we aim to maximize the profit over time as the investors prefer to get a higher return for long term investments.

4.1 Analysis of different statistical method and finding a suitable one

Statistics being the body of methods meant for the study of numerical data, the first step in any statistical inquiry must be a collection of relevant numerical data. Once data are collected, a knowledge discovery may analyze the behavior of the data. However, as the different data are different, a suitable statistical method is to be identified for data processing belongs to a particular application.

In this present work, one of the primary challenges is finding a preferable method that would help us forecast/predict the companies' share value. However, in reality, it is impossible to predict the share value of the companies accurately, as several issues control the movement of the share market. Many parameters may not be directly related to the up/down of the share market, but indirectly related to the share market by affecting parameters responsible for changing the share value directly. Even some new cases/issues (e.g., COVID-19, terrorist attacks, etc.) may evolve in the future that is unknown at the time of analysis. The proposed prediction model will be based on historical prices of shares as it reflects the effects of all events and parameters that determine the stock prices.

There are several statistical methods such as *standard deviation*, *linear regression*, non-linear regression, *correlation*, *time series analysis*, etc. that are useful for prediction. Standard deviation is the simplest one. The result produced by the standard deviation is used for basic calculation. It gives a high error rate for complex data. Therefore, for more satisfactory precision, it is not a suitable method.

In correlation analysis, the concern is the mutual relationship between the two variables. It uses a measure of the interdependence of the two variables, known as the correlation coefficient.

Because of the close connection between the correlation coefficient and linear regression, it is clear that the former can serve as a satisfactory measure of the strength of the relationship between the two variables only when that relationship is of the linear type. Hence a low value of the correlation coefficient does not rule out the possibility that the variables are related in some other manner. However, as the problem domain is the capital market, it might be possible that the correlation coefficient is small in the measure. Therefore, while dealing with real financial time series data, it is not possible to guarantee that the correlation coefficient remains greater than a particular value all the time. For this limitation, correlation is not considered in the present work as a tool of prediction.

In the share market, the stock prices of companies change over time, and it does not maintain any specific pattern. Different companies follow different patterns of change. Not only this, but a particular company may also show varying patterns over time. Therefore, it is necessary to understand the pattern of change in the share price. Once the values of the shares plotted against time, different types of curves get generated. In mathematical analysis, several forms of curves are identified over time. Among these numerous forms, we intend to look for the one that best fits the given pattern.

There are several forms of curves used in mathematical analysis. Out of those commonly used are listed below in Table 1. Now the best-fitted form of curve is identified by comparing the actual values with the values generated by solving the regression equation associated with each type of curve. In this problem domain, the independent variable is the time (x) and the dependent variable is the closing stock price (y).

This comparison helps to calculate the error value for each curve defined above. Now the curve for which the minimum error value is obtained is identified as the best-fitted curve to represent the share movement of that particular company.

Regression analysis is very much useful, as it helps us in the prediction, and forming the curves which help to compute the error values. In regression analysis, one of the two variables (say x) is the *independent* variable and the other (y) as the *dependent* variable, and the objective is to be investigate the dependence of the variable y on the independent variable x . The major problem in the case of linear regression is to express the relationship between y and x through a mathematical function such as a linear equation and, then only it is possible to use the resulting equation to predict y in terms of x .

Non-linear regression is another variation of regression techniques. But in the case of non-linear regression, there is no hard and first rule to establish the relationship between two variables in a linear pattern like linear regression. The

Table 1 Different well-known curves with their general form (equation)

Sl. no	Curve model	General form (equation)
1	Straight line/Poly (1)	$y = p_1 \times x + p_2$
2	Parabolic curve/Poly (2)	$y = p_1 \times x^2 + p_2 \times x + p_3$
3	Cubic curve/Poly (3)	$y = p_1 \times x^3 + p_2 \times x^2 + p_3 \times x + p_4$
4	Poly(n)	$y = p_1 \times x^n + p_2 \times x^{n-1} + \dots + p_n$
5	Exponential curve (1)	$y = a \times e^{b \times x}$
6	Exponential curve (2)	$y = a \times e^{b \times x} + c \times e^{d \times x}$
7	Geometric curve (1)	$y = a \times x^b$
8	Geometric curve (2)	$y = a \times x^b + c$
9	Fourier curve (1)	$y = a_0 + a_1 \times \cos(w \times x) + a_2 \times \sin(w \times x)$
10	Fourier curve (2)	$y = a_0 + a_1 \times \cos(x \times w) + b_1 \times \sin(x \times w) + a_2 \times \cos(2 \times x \times w) + b_2 \times \sin(2 \times x \times w)$
11	Logarithmic curve	$\log(y) = f(x)$; where $f(x)$ is linear in x

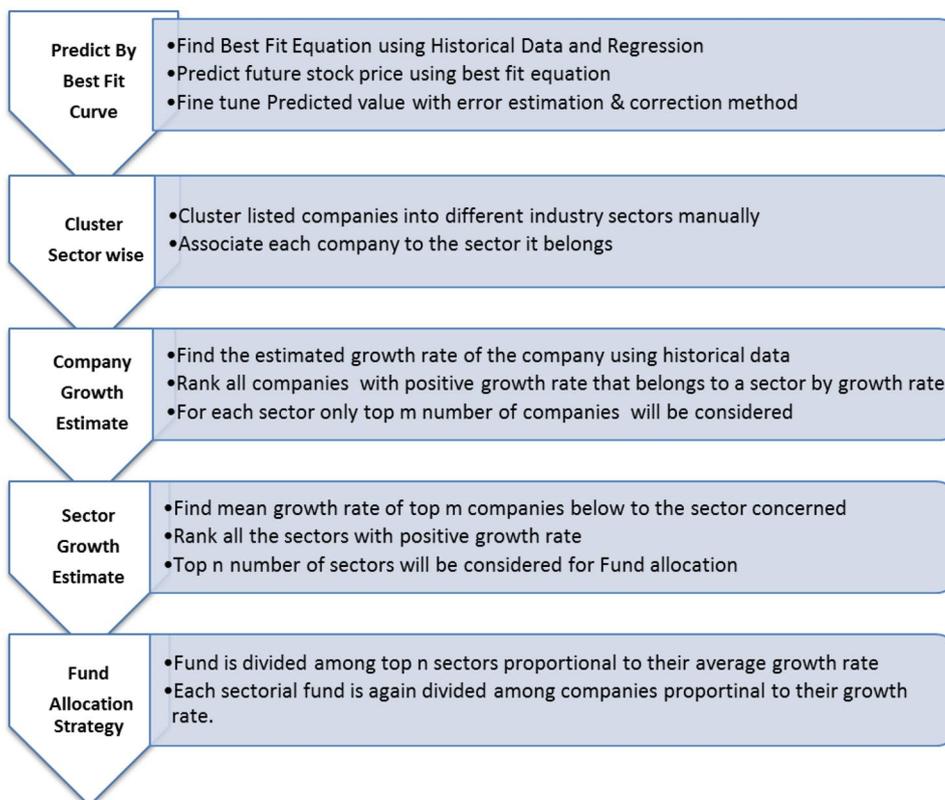
relationship between the two variables in non-linear regression is non-linear.

4.2 Methodology

One of the main objectives is to predict the share price of an individual company. Actual share prices plotted in a 1-month interval up to the current year. The present-day stock price of a company is estimated using regression on the historical price data (say last P monthly closing price). The

regression method is applied to solve the different types of curve-fitting models and compare those values with the current one (actual market price). Next, the difference between the estimated price and the actual price is computed. The percentage of difference gives the error rate for each type of curve. The minimum squared error percentage value would be chosen and, therefore, the corresponding curve would be identified as the best-fitted curve, and the corresponding share value is considered for further processing. The best fit curve then predicts the stock price for the next time period

Fig. 1 Overview of different phases of the proposed investment strategy



for the company. The overall flow of the methodology is depicted in Fig. 1.

In order to get better prediction value refinement is done on the predicted value generated so far (using the best-fit regression line) by removing some error to some extent, by implementing an error estimation technique discussed below.

In the case of the share market, the impact of the share price of recent years is more influential than the previous years. However, a long-time analysis is also required to understand the trend of the particular stock. With these issues as critical consideration, It uses data for a longer time with a higher impact on recent prices in the proposed formulation. It puts more weight in recent years and gradually decrease the weight for previous years. Then after calculating an error measure, It adjusts on the error value with the predicted value for better estimation.

It gives a set of predicted values of different companies, and that is the input data set for further activities. Therefore, it is needed to allocate the fund into the market such that the net return is comparatively higher. Investing total funds into a single company would not always maximize the returns as a company showing a higher growth-rate might not continue the same in the future. The same is also true for the companies having a low-profit percentage at present, might show better return in the future due to different factors such as the launching of new products, new investment from the investors, acquisition, etc. Therefore, the capital should be invested across many companies (diversified investment) that belong to diverse industry sectors to maintain better returns while reducing the risk. The proposed strategy is to diversify into many sectors for better portfolio management.

After predicting the share values, the companies are clustered sector-wise for diversified fund allocation. These different sectors with a different rate of growth are identified. Therefore, we propose a mathematical approach to allocate funds sector-wise.

Before the allocation of funds, it needs to calculate the growth rate of all individual companies belongs to a sector. The calculation is on the growth rate between the two same time period for all the companies and to give more focus on the recent growth rate as compared to the older period. Some weighting factors are set for the previous periods that would get multiplied with their corresponding growth rate where their sum up gives the overall growth of that company. Similarly, we need to calculate the growth rate of all the companies within a sector. Then, the mean of the growth rate of all the companies except those with negative growth would reflect the overall growth rate of a particular industry sector. Likewise, the net growth rate of all the sectors is calculated. The philosophy here is to allocate a bigger chunk of the fund to the high growth sectors and less to the

moderately growing sectors. The same logic is applied while selecting the candidate companies within an industry sector.

4.3 Algorithm for diversified fund allocation across Companies as well as sector-wise to maximize the net return

In the proposed methodology, the prediction of the current share value is done based on the data from the previous p months. The *month-wise weight* (Y_i) is used for calculating the impact of growth/fall of the price several times in the proposed algorithm for consecutive years to compute the error values and in case of the computing growth rate of an individual company in subsequent years. It is calculated for the i th month using Eq. (1):

$$Y_i = \frac{2 * (p - i + 1)}{p * (p + 1)}. \quad (1)$$

Different industry sectors are identified by consulting financial news sources and NSE and BSE web portals. Companies that belong to these sectors are also available from the same sources. Historical stock prices of those listed companies are also available from NSE and BSE portal (BSE 2019b).

STEP 1: [PREDICTION OF STOCK PRICE OF SELECTED COMPANIES]

1. A: [IDENTIFYING BEST FITTED CURVE AND DATA SET GENERATION]

- 1.A. i: Collect the historical stock prices of a company for the last p months. This constitutes the initial dataset for the model.
- 1.A. ii: Solve different curve-fitting models by a regression method to predict the stock price for different time periods between p and 0 months (present). Therefore, generating some predicted share values of that company across the different time period.
- 1.A. iii: Calculate the percentage deviation of the forecasted values from the actual values.
- 1.A. iv: Choose the curve as best fitted for which the rate of deviation (*Root Mean Square Error (RMSE)*) is lowest and R-squared value is maximum then put the predicted values of the corresponding curve in our data set.
- 1.A. v: Predict the share value (for historical data for which actual prices are also known) after a certain period (when we would need to withdraw our fund may be after 3, 6, 12 months in future) by solving the best-fitted curve using regression. Insert the predicted share value into the dataset.

- 1.A. vi: Repeat the steps from 1.A.i to 1.A.v for every company to generate our data set.

1. B. [FINE-TUNING OF PREDICTED STOCK PRICE BY ERROR ESTIMATION]

- 1.B. i: Calculate the deviation of predicted share value from the actual value at each considered time period
- 1.B. ii: Suppose the deviations at P different time periods are, \dots, d_p . Where d_k is the deviation of share value corresponding to its actual share value k period (months) earlier.
- 1.B. iii: Calculate the *Company Net Error (CNE)* by the following formula:

$$CNE_j = Y_1d_1 + Y_2d_2 + \dots + Y_kd_k + \dots + Y_p d_p,$$

where CNE_j is the *Company Net Error* of the j th company (where $j = 1 \dots m$). Y values for each time period are calculated following Eq. (1).

- 1.B. iv: Add the CNE_j to the corresponding predicted share value of the company to generate *FPV (Final Predicted Value)* and insert *FPV* into our data set.
- 1.B. v: Repeat step 1.B.i to 1.B.iv until *FPV (Final Predicted Value)* of all the companies get calculated.

STEP 2: [BUSINESS DOMAIN IDENTIFICATION]

Group all listed companies into different industry sectors or domains based on the business areas of the company with the help of domain experts.

STEP 3: [SELECTION OF SECTORS AND COMPANIES BELONG TO THAT SECTOR FOR INVESTMENT]

3. A: [COMPANY-WISE WITHIN A SECTOR]

- 3.A. i: Pick up a company from a particular sector.
- 3.A. ii: Find the percentages of the growth rate of the company for different time periods with respect to the month immediately earlier. Calculate this into present day from the initial time period (say p months ago). Suppose the growth rate between i th previous month and $(i - 1)$ th previous month is Gr_i , where $i = 1$ to P , considering current month as 0th month. Therefore, Gr_i is the growth rate of $(i - 1)$ th time period w.r.t its immediate earlier month i.e. i th month. To maximize the impact of current growth over the growth of older year, we would develop a mathematical formula stated below.

Suppose the growth rates of a company are Gr_1, Gr_2, \dots, Gr_p respectively from present to P years earlier.

- 3.A. iii: Calculate the *Company Net Growth Rate (CNGR)* by the following formula:

$$CNGR_j = Y_1Gr_1 + Y_2Gr_2 + \dots + Y_iGr_i + \dots + Y_pGr_p$$

Where $CNGR_j$ is the *Company Net Growth Rate* of the j^{th} company (where $j=1$ to m). Y_i is calculated following Eq. (1).

- 3.A. iv: Repeat step 3.A.i to 3.A.iii until *Company Net Growth Rate (CNGR)* of all the companies of that particular sector gets calculated.
- 3.A. v: Consider only the companies having positive growth rate to invest our fund and discard all the companies having a negative growth rate for that time period.

3. B: [SECTORWISE]

Calculate the net growth rate of a particular sector by finding the mean value of the growth rate of all the companies belong to that sector.

3. C: LOOP

Repeat step 3.A and 3.B for each sector.

STEP 4: [ALLOCATION OF FUND]

In the case of fund allocation, the motive is to allocate more funds in such sectors and companies having better growth rate over the sectors and the companies having a lower rate of growth to enlarge overall profit. Say overall fund is F .

4. A. [SECTORWISE]

- 4.A. i. Find out the Sector Multiplying Factor (*SMF*) by the following formula:

$$SMF = \frac{100}{\sum_{i=1}^n G_i}$$

Where G_i is the growth rate of sectors S_i , n is the number of sectors selected for investment.

- 4.A. ii: Determine the sector-wise fund to be invested by the mathematical formula given below:

$$SA_i = G_i * SMF$$

Where SA_i denotes sector-wise percentage allocation. Thus sector-wise allocation is given by

$$SFA_i = F * SA_i$$

- 4.A. iii: Repeat step 4.A.i to step 4.A.ii for all the selected sector.
- 4. B. [COMPANYWISE] Repeat for each sector S_i where $i = 1$ to n Let each sector S_i consists of m number of companies C_1 to C_m with growth percentages of G_1 to G_m respectively.
 - 4.B. i: Find out the Company Multiplying Factor (CMF) by the following formula:

$$CMF = \frac{100}{\sum_{j=1}^m G_j}$$

Where G_j is the growth rate of a sector containing m number of companies C_j (where $j = 1 \dots m$) respectively.
 - 4.B. ii: Determine the company wise fund to be invested by the mathematical formula given below: $CA_k = g_k * CMF$ for sector C_i . (where $k = 1 \dots m$) Where CA_k denotes company wise allocation percentage wise Thus company wise allocation is given by $SCA_k = SFA_i * CA_k$
 - 4.B. iii: End of Repeat Step 4B

5 Experimental results

In this experiment, 137 months of (Aug 2004–Dec 2015) closing stock prices of 20 different companies from BSE along with their dividend paid, bonus share issued, share split, etc. (These parameters are used for the mutual fund profit calculation not for share price forecasting) are used as the training set. Data from January 2016 to Dec 2016 has been used for validation, evaluation, and comparison of the proposed portfolio with other CRISIL (CRISIL 2020; Annapoorna and Gupta 2013) ranked mutual funds and benchmark indexes. For the experiments, we first perform regression on different company’s historical closing prices and select the best fitting curve for further processing. As an example, we have shown different curves fitted with the stock prices of TCS in Fig. 2. The trend line gives the curve equation and corresponding coefficient values. We choose the curve with minimum RMS error for each of the fitted curves. Table 2 shows the sample calculations of the best curve fitting of TCS with different parameter values like coefficients of fitted curve, RMSE, and R-squared value. We compute the same with the other 19 companies and the best fit curve for each of these companies along with the different parameters are shown in Table 3. Next, we allocate the fund into multiple industrial sectors based on the cumulative performance of the stocks from each such sectors as presented in Table 4. We have considered a total capital of Rs.

Fig. 2 Different curves fitting with monthly closing stock prices of TCS

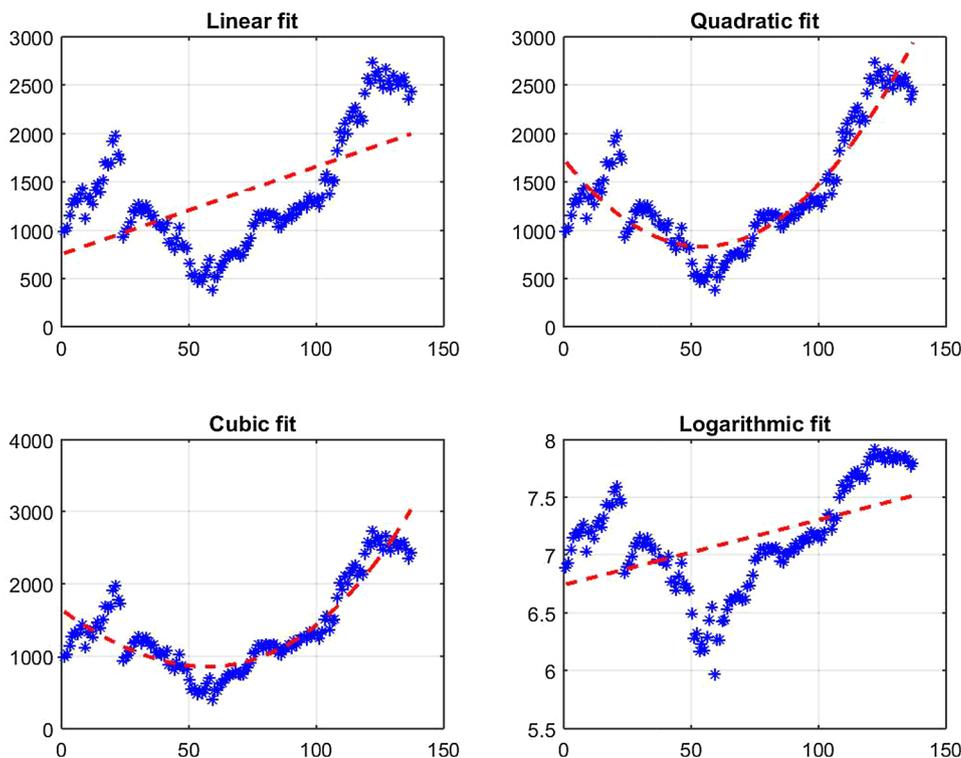


Table 2 Error percentage with a different form of the curve-fitting model for TCS (sector = IT)

Curve model	Equation	Coefficients (range)	R-squared	Adjusted R-squared	RMS error
Linear	$f(x) = p_1 * x + p_2$	p1 = 9.066 (6.916, 11.22) p2 = 753.5 (582.6, 924.5)	0.34	0.3351	503.2
Quadratic	$f(x) = p_1 * x^2 + p_2 * x + p_3$	p1 = 0.3072 (0.276, 0.3383) p2 = - 33.32 (- 37.76, - 28.89) p3 = 1736 (1603, 1868)	0.83	0.8257	257.7
Cubic	$f(x) = p_1 * x^3 + p_2 * x^2 + p_3 * x + p_4$	p1 = 0.0007091 (- 0.0001824, 0.001601) p2 = 0.1604 (- 0.02677, 0.3475) p3 = - 25.19 (- 36.32, - 14.05) p4 = 1640 (1462, 1818)	0.83	0.8276	256.3
Exp order 1	$f(x) = a * e^{b*x}$	a = 685.5 (578.2, 792.7) b = 0.009011 (0.007458, 0.01056)	0.46	0.4559	455.2
Exp order 2	$f(x) = a * e^{b*x} + c * e^{d*x}$	a = 1526 (1347, 1705) b = - 0.02264 (- 0.03131, - 0.01396) c = 129.5 (54.18, 204.8) d = 0.02288 (0.01847, 0.02728)	0.81	0.806	271.8
Power order 1	$f(x) = a * x^b$	a = 410.7 (214.6, 606.7) b = 0.296 (0.186, 0.406)	0.16	0.1484	569.5
Power order 2	$f(x) = a * x^b + c$	a = 1.371e-07 (- 6.151e-07, 8.893e-07) b = 4.758 (3.63, 5.885) c = 1023 (941.2, 1105)	0.72	0.7136	330.3

Table 3 CNE, FPV and CNGR (%) of different companies from selected sectors

Sl #	Sector	Company name	Best fit curve	CNE	FPV	CNGR (%)
1	IT	TCS	Poly (2)	-0.66814	2987.489	1.03496
2	IT	INFOSYS	Fourier (2)	-86.9938	1206.09	1.129897
3	IT	Wipro	Poly (2)	-0.01535	609.6597	0.529457
4	IT	Mind Tree	Poly (2)	-0.08897	1685.753	1.166362
5	BANKING	SBI	Poly (2)	0.29379	2382.164	2.604578
6	BANKING	ICICI	Poly (2)	-0.00057	1519.11	2.438129
7	BANKING	PNB	Fourier (1)	-3.75158	592.5136	1.997507
8	BANKING	HDFC	Poly (2)	0.144772	5438.641	2.702206
9	METAL	SAIL	Fourier (1)	-0.97805	82.02259	2.210792
10	METAL	Tata Steel	Poly (3)	0.021196	329.3907	1.947975
11	METAL	Hindalco	Fourier (1)	78.25338	1192.171	1.113968
12	METAL	Vedanta	Poly (3)	-0.19517	1357.714	3.187939
12	TELECOM	IDEA	Fourier (1)	0.052979	152.4592	1.545887
14	TELECOM	AIRTEL	Exp (2)	-0.54232	652.5065	2.526714
15	TELECOM	RCOM	Poly (2)	-0.01265	76.3301	0.347679
16	TELECOM	TATA Comm	Fourier (1)	-0.00725	482.0565	1.969281
17	PETRO	RIL	Poly (3)	0.25506	2058.618	2.165733
18	PETRO	HPCL	Power (2)	8.237011	813.3963	0.864871
19	PETRO	IOCL	Fourier (1)	-0.50527	399.1524	0.811133
20	PETRO	Petronet LNG	Poly (2)	-0.00323	205.7566	2.686628

1000,000 for investment by the portfolio manager to evaluate the return on investment at different time periods. Finally, Table 5 shows the fund allocated to each company from each sector depending on their ranking based on expected earning.

The main motive is to allocate more funds to the companies with a higher growth rate than those with a comparatively lower growth rate within a particular sector to earn a better profit. Therefore, the percentage of total funds allocated, along with the amount of funds allocated to an

Table 4 Sector-wise fund allocation

Sl #	Sector name	Sector growth rate (G_i)	% of Fund allocated to a sector $(SA_i = G_i \times SMF)$	Amount of fund allocated to a sector (Approx) $SFA_i = F \times SA_i$
1	IT	0.965169	11.0994435	110,000
2	BANKING	2.435605	28.0094575	280,000
3	METAL	2.115169	24.32443775	243,000
4	TELECOM	1.59739	18.36998788	183,000
5	PETRO	1.632091	18.76904938	187,000

individual company within a particular sector (say, IT) shown in Table 4. For example, the fund allocated to the IT sector is Rs. 1,10,000/-. Therefore, this amount of funds further allocated across various companies within the IT sector shown in the first four rows of Table 5.

Let us assume that funds allocated to different companies as per the framework, and stocks brought in December 2015, and the funds remain invested for the next 30 months. For simplicity, let us also assume that no more fund inflow or redemption took place in-between. Next, we calculated the market value of the stocks bought at the end of every successive quarter. The dividend amounts, given by the companies from time to time, also added to arrive at the final value of the invested capital. Next, the absolute percentage return on investment is deduced for all subsequent quarters. Detailed calculation sheet considering the actual

stock prices from December 2015 to June 2018 placed in the Supplementary Material. We compared the gains with some well-established funds' performance during the same period. HSBC Large Cap Equity Fund—Regular Plan (G) is ranked first by CRISIL (2020) for the period, and also, Axis Focused 25 Fund—Direct Plan (G) placed at rank-1 in diversified equity MFs in the last 3 years. Table 6 presents the comparative results between our proposed methodology and the top-performing MFs. Figure 3 represents the comparative performance graphically. We observe from the bar chart that proposed MF generates quite a good income in terms of capital appreciation as well as return on investment. We have followed a *buy-and-hold strategy* during the evaluation of the investment framework. It means that during the whole period of 30 months, no stocks have been sold out partially on peaks, and no new stocks are brought on

Table 5 Allocation percentage of fund company-wise within all sectors

Sl #	Sector	Sector fund	Companies within sector	Company growth rate	CMF	% of Sector fund allocated to the company $CA_k = g_k * CMF$ (Approx.)	Amount of fund allocated to company $SCA_k = SFA_i * CA_k$ (Approx.)
1	IT	1100000	TCS	1.034960	25.9	26.81	29,486
2			INFOSYS	1.129897		29.26	32,190
3			WIPRO	0.529457		13.71	15,084
4			Mind Tree	1.166362		30.21	33,229
5	Banking	280000	SBI	2.604578	10.3	26.73	74,856
6			ICICI	2.438129		25.03	70,072
7			PNB	1.997507		20.50	57,408
8			HDFC	2.702206		27.74	77,662
9	Metal	243000	SAIL	2.210792	11.8	37.68	91,561
10			TATA Steel	1.947975		23.02	55,948
11			Hindalco	1.113968		13.17	31,994
12			Vedanta	3.187939		37.68	91,561
13	Telecom	183000	IDEA	1.545887	15.7	24.20	44,274
14			AIRTEL	2.526714		39.54	72,366
15			RCOM	0.347679		5.44	9,957
16			TATA Comm	1.969281		30.82	56,401
17	Petro	187000	RIL	2.165733	15.3	33.17	62,035
18			HPCL	0.864871		13.25	24,773
19			IOCL	0.811133		12.42	23,234
20			Petronet LNG	2.686628		41.15	76,956

ABSOLUTE % RETURN

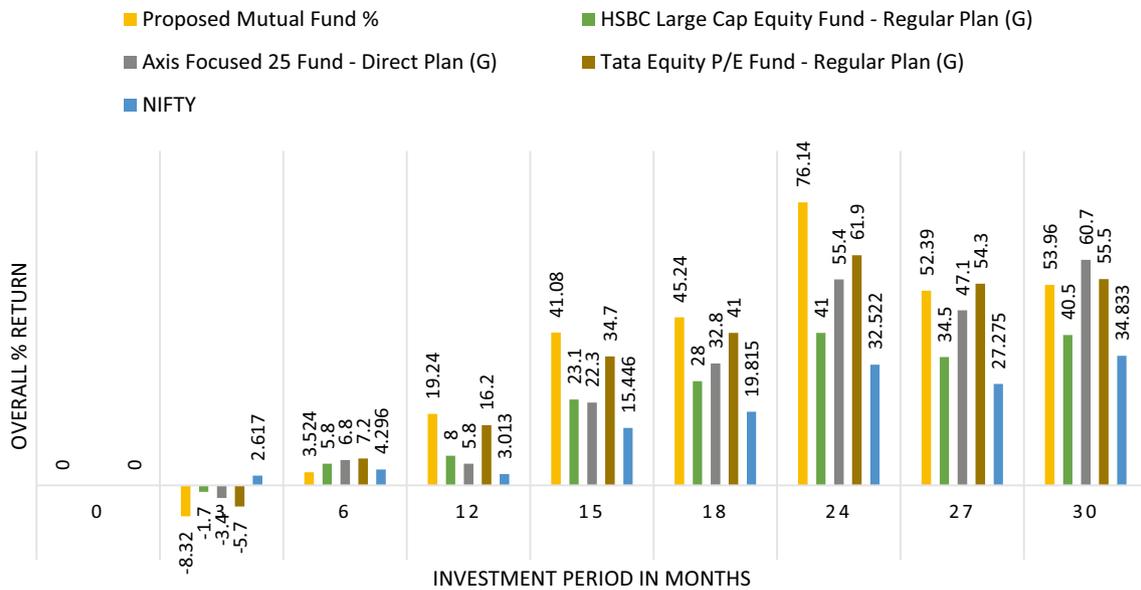


Fig. 3 Absolute overall gain/return (%) on investment over the period of 30 months for proposed mutual fund and three other well-established funds. (HSBC large cap equity fund regular plan (G) is

ranked-1 MF by (CRISIL 2020) and Axis focused 25 fund direct plan is ranked-1 in diversified equity MFs in last 3 years)

Table 6 Absolute percentage return of the proposed mutual fund in comparison to some well-ranked mutual funds along with NIFTY benchmark index for a period of 30 months

Month-year	Total amount	Duration of investment (months)	% Return in proposed MF (%)	% Return in HSBC large cap equity fund—regular plan (G)	% Return in axis focused 25 fund—direct plan (G)	% Return in tata equity P/E fund—regular plan (G)	NIFTY points	% Return in Nifty index
12-2015	1,000,000	0	0	–	–	–	7946	0
03-2016	918,759	3	– 8.32	– 1.7	– 3.4	– 5.7	7738	2.6
06-2016	1,035,236	6	3.524	5.8	6.8	7.2	8288	4.3
12-2016	1,165,869	12	19.24	8	5.8	16.2	8186	3.0
03-2017	1,293,750	15	41.08	23.1	22.3	34.7	9174	15.5
06-2017	1,292,870	18	45.24	28	32.8	41	9521	19.8
12-2017	1,479,314	24	76.14	41	55.4	61.9	10,531	32.5
03-2018	1,245,890	27	52.39	34.5	47.1	54.3	10,114	27.3
06-2018	1,221,441	30	53.96	40.5	60.7	55.5	10,714	34.8

dips. The above ensures that our framework does not depend much on the market timing of the fund managers. Our return could have been more if we would have re-invested the dividend amounts in the same stock, but in this study, we have kept the dividend amount aside and never invested for simplicity. If we carefully observe the results in Fig. 3, we see that during the first 3 months almost all the MFs including the proposed scheme have negative returns and after 6 months all go to green and increase subsequently. The period is from January 2016 to June 2016 and certain global events

caused those bearish patterns such as economic slowdown of China (Hu et al. 2018), the Brexit issue (Mora-Cantalops et al. 2019) and volatile crude oil price.

6 Conclusion

This paper proposes a novel methodology to build up a mutual fund portfolio to increase the profit over time and avoid risks. The people with the goal of the investment

generally look for higher returns over time. The proposed methodology helps build a robust portfolio over time by analyzing the historical data based on regression analysis. Experimental results show that it performs as good as India's top-performing mutual funds or even beat them over a long period. The majority of the research works in this domain are focused on predicting the individual stock price. Here we extend and reframe the problem by identifying the different business sectors and then compute the price of every stock selected for these sectors and finally group these stocks in such a cumulative manner so that the risk of investment is reduced but the percentage of return increases. The proposed method can be used by portfolio managers to benchmark mutual funds' performances for a considerably long period. Though the experiment performed on the Indian share market data, it is also applicable to different share markets across the globe. This framework has the advantage of being less dependent on market-timing of buy/sell by expert fund managers. We have excluded the options of buying on dips and selling on ups to enhance profit margin. Instead, we have focused on making it robust against market volatility by using a *buy-and-hold* strategy. Dividends received are also not re-invested in the framework for simplicity. We have shown that it is possible to gain a decent return with much less market timing dependence and complex financial indicator analysis. The *high net worth individuals* (HNI) could also leverage the proposed framework for their portfolio management. In our future study, strategies other than *buy-and-hold* could be explored by adding complexity to the framework. Every quarter buy/sell strategy could be incorporated based on the forecasted change in the upcoming months. The dividend amount could also be used to hedge the market risk by investing in gold/government securities/bonds, etc. We can study the curve's changing pattern for varying periods for a specific domain/company for better prediction.

References

- Agapova A (2011) Conventional mutual index funds versus exchange-traded funds. *J Financ Mark* 14(2):323–343
- Altay E, Satman MH (2005) Stock market forecasting: artificial neural network and linear regression comparison in an emerging market. *J Financ Manag Anal* 18(2):18
- Ananthi M, Vijayakumar K (2020) Stock market analysis using candlestick regression and market trend prediction (CKRM). *J Ambient Intell Humaniz Comput*. <https://doi.org/10.1007/s12652-020-01892-5>
- Annapoorna M, Gupta PK (2013) A comparative analysis of returns of mutual fund schemes ranked 1 by CRISIL. *Tactful Manag Res J* 2(1):1–6
- Asadi S (2019) Evolutionary fuzzification of ripper for regression: case study of stock prediction. *Neurocomputing* 331:121–137
- Association of Mutual Funds in India (2020) Industry trends June 2020. <https://www.amfiindia.com/Themes/Theme1/downloads/home/Industry-Trends.pdf>. Accessed 19 July 2020
- Atsalakis GS, Valavanis KP (2009) Surveying stock market forecasting techniques-Part II: soft computing methods. *Expert Syst Appl* 36(3):5932–5941
- Baralis E, Cagliero L, Garza P (2017) Planning stock portfolios by means of weighted frequent itemsets. *Expert Syst Appl* 86:1–17
- BSE (2019a) Bombay stock exchange, India. <http://www.bseindia.com>. Accessed 30 Nov 2019
- BSE (2019b) BSE India stock monthly closing price dataset. <http://www.bseindia.com/markets/equity/EQReports/StockPrChHistori.aspx?scripcode=512289&flag=sp&Submit=G>. Accessed 30 Sept 2019
- Chandar SK (2019) Fusion model of wavelet transform and adaptive neuro fuzzy inference system for stock market prediction. *J Ambient Intell Humaniz Comput* 1–9
- Chikhi M, Diebolt C (2010) Nonparametric analysis of financial time series by the kernel methodology. *Qual Quantity* 44(5):865–880
- CRISIL (2020) Mutual fund ranking. <https://www.crisil.com/en/home/what-we-do/financial-products/mf-ranking.html>. Accessed 19 July 2020
- Curran K (2011) Ubiquitous developments in ambient computing and intelligence: human-centered applications. *Artif Intell* 2021:229
- Ding S, Cui T, Xiong X et al (2020) Forecasting stock market return with nonlinearity: a genetic programming approach. *J Ambient Intell Human Comput* 11:4927–4939. <https://doi.org/10.1007/s12652-020-01762-0>
- Fong SJ, Dey N, Chaki J (2020) Artificial intelligence for coronavirus outbreak. Springer, New York. <https://doi.org/10.1007/978-981-15-5936-5>
- Ghosh A, Bose S, Maji G, Debnath N, Sen S (2019) Stock price prediction using LSTM on Indian share market. In: Proceedings of 32nd international conference on, vol 63, pp 101–110
- Gottschlich J, Hinz O (2014) A decision support system for stock investment recommendations using collective wisdom. *Decis Support Syst* 59:52–62
- Hajizadeh E, Ardakani HD, Shahrabi J (2010) Application of data mining techniques in stock markets: a survey. *J Econ Int Financ* 2(7):109–118
- Hu S, Liu M, Fong S, Song W, Dey N, Wong R (2018) Forecasting China future MNP by deep learning in behavior engineering and applications. Springer, New York, pp 169–210
- Ji M, Zhang K, Wu Q (2020) Introducing a hybrid model SAE-BP for regression analysis of soil temperature with hyperspectral data. *Int J Ambient Comput Intell (IJACI)* 11(3):66–79
- Liao SH, Chu PH, Hsiao PY (2012) Data mining techniques and applications—a decade review from 2000 to 2011. *Expert Syst Appl* 39(12):11303–11311
- Maji G, Sen S, Sarkar A (2017) Share market sectoral indices movement forecast with lagged correlation and association rule mining. In: IFIP International conference on computer information systems and industrial management, Springer, pp 327–340
- Mingyue Q, Cheng L, Yu S (2016) Application of the artificial neural network in predicting the direction of stock market index. In: 2016 10th International conference on complex, intelligent, and software intensive systems (CISIS), IEEE, pp 219–223
- Moere AV, Offenhuber D (2009) Beyond ambient display: a contextual taxonomy of alternative information display. *Int J Ambient Comput Intell (IJACI)* 1(2):39–46
- Mondal D, Maji G, Goto T, Debnath NC, Sen S (2018) A data warehouse based modelling technique for stock market analysis. *Int J Eng Technol* 7(313):165–170
- Mora-Cantalalops M, Sánchez-Alonso S, Visvizi A (2019) The influence of external political events on social networks: the case of

- the Brexit twitter network. *J Ambient Intell Human Comput*. <https://doi.org/10.1007/s12652-019-01273-7>
- Narang V (2015) Impact of politics on sensex. *Int J Econ Bus Rev* 3(1):165–168
- NSE (2019) National stock exchange, India. <http://www.nseindia.com>. Accessed 30 Nov 2019
- Paranjape-Voditel P, Deshpande U (2013) A stock market portfolio recommender system based on association rule mining. *Appl Soft Comput* 13(2):1055–1063
- Park M, Lee ML, Lee J et al (2019) Predicting stock market indices using classification tools. *Asian Econ Financ Rev* 9(2):243–256
- Preethi G, Santhi B (2012) Stock market forecasting techniques: a survey. *J Theor Appl Inf Technol* 46(1):24
- Qian B, Rasheed K (2007) Stock market prediction with multiple classifiers. *Appl Intell* 26(1):25–33
- Richthammer C, Pernul G (2020) Situation awareness for recommender systems. *Electron Commerce Res* 20:783–806. <https://doi.org/10.1007/s10660-018-9321-z>
- Rusu V, Rusu C (2003) Forecasting methods and stock market analysis. *Creative Math* 12:103–110
- Sankar CP, Vidyaraj R, Kumar KS (2015) Trust based stock recommendation system—a social network analysis approach. *Proced Comput Sci* 46:299–305
- Singh N, Mohanty S (2018) Short term price forecasting using adaptive generalized neuron model. *Int J Ambient Comput Intell (IJACI)* 9(3):44–56
- Weng F, Chen Y, Wang Z, Hou M, Luo J, Tian Z (2020) Gold price forecasting research based on an improved online extreme learning machine algorithm. *J Ambient Intell Human Comput* 1–11
- Zhang D, Zhou L (2004) Discovering golden nuggets: data mining in financial application. *IEEE Trans Syst Man Cybern Part C (Appl Rev)* 34(4):513–522
- Zhang J, Teng YF, Chen W (2019) Support vector regression with modified firefly algorithm for stock price forecasting. *Appl Intell* 49(5):1658–1674
- Zhao J, Zeng D, Liang S, Kang H, Liu Q (2020) Prediction model for stock price trend based on recurrent neural network. *J Ambient Intell Human Comput*. <https://doi.org/10.1007/s12652-020-02057-0>
- Zhu HM, Li R, Li S (2014) Modelling dynamic dependence between crude oil prices and Asia-Pacific stock market returns. *Int Rev Econ Financ* 29:208–223

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.