A Comprehensive Review of Financial Market Forecasting: From Historical Data to Sentiment-based Approaches

Komal Batool
Department of Mathematics
NED University of Engineering & Technology

Mirza Faizan Ahmed
Department of Economics & Management Sciences
NED University of Engineering & Technology

Ubaida Fatima
Department of Mathematics
NED University of Engineering & Technology

ABSTRACT

Financial market is stochastic in nature. The movement in financial market is random. One of the reasons of this fact is that the market is sensitive to multiple factors. It is autoregressive in nature, which means that it depends on its past values. Other than that, macroeconomic variables like Gross Domestic Product (GDP), interest rate, gold price or currency exchange rate also cause fluctuations in the market. Along with that the market is also sensitive to socio-political events, news, tweets and trends. The objective of this review is to understand the predictivity of financial markets based on different datasets and different training models. This paper describes a detailed review that how much features have been incorporated in order to predict the financial market and discusses the effect on predictivity of a market by changing these factors. The novelty of this paper is that it elaborates the methodologies used for the forecasting of financial market and the optimal features required for efficient prediction.

- This review paper provides a comprehensive overview of research conducted on forecasting of financial markets over past 20 years, focusing on datasets and models employed.
- The study categorizes forecasting approaches in three main methodologies: statistical modelling based forecasting, machine learning modelling based forecasting and hybrid modelling based forecasting.
- This survey aims to identify the factors that are most significant for the forecasting of financial market by categorizing the studies based on datasets: historical dataset, technical dataset and textual dataset.

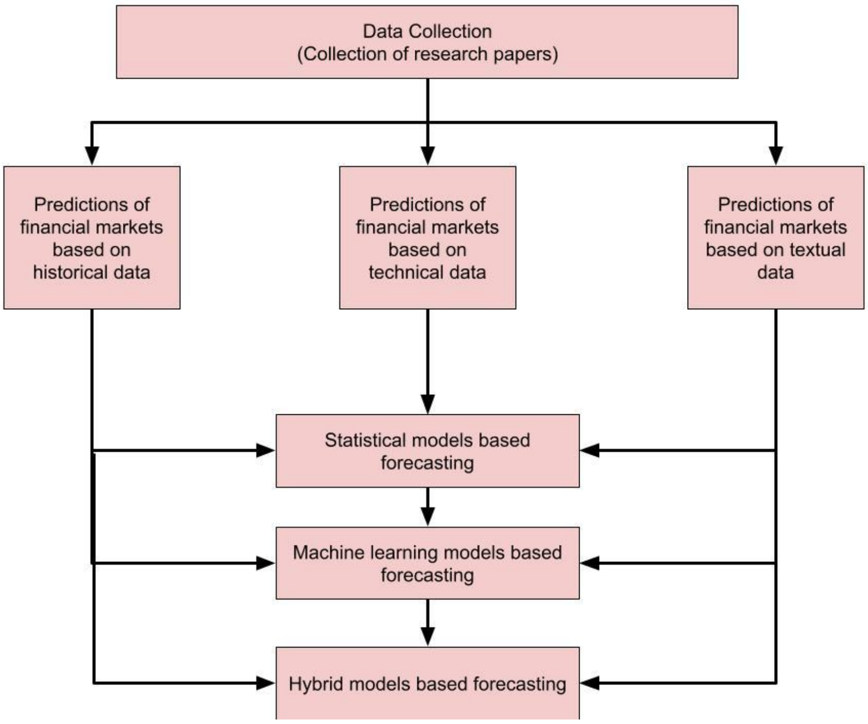
General Terms

Forecasting, Financial Markets, Mathematical Approaches

Keywords

Time series data, Stock Market, Macroeconomic variables, Machine Learning models, hybrid model, hybrid datasets, sentiment analysis.

Graphical Abstract



1. INTRODUCTION

Stock market is considered as one of the best options for the investors to invest. The first and the major step of any investment is to decide the particular investment instrument among available options. Decision making is the most crucial step as investors aim to invest in a market to have maximum return with minimum risk. As the stock market is a volatile market, it can produce major returns and major losses. Prediction of future behavior of a stock market is a strenuous task as it can be affected by several internal as well as external factors [1]. Stock market prediction is done to ease the decision making process, so that the investors can have a clear picture of future behavior of the market and can make decisions accordingly [2]. In order to predict the future behavior of the stock market, an optimal number of factors are required to be incorporated in a predictive model that affects the market directly or indirectly. From historical data, foreign exchange rate, gold prices, crude oil prices to Google trends all can make a little or obvious effect on the stock market and can be used to predict its future behavior [3] [4]. Stock market is also sensitive to financial news, political events and traders' sentiments. These factors can make a dramatic change in the price movements. This is another reason that makes the prediction of stock market a difficult task. In short, prediction of the stock market is a challenging task as it is a highly volatile market and depends on a variety of economic factors, political news and market sentiments [5].

Forecasting the stock market using econometric models based on historical data is a classical way of prediction. Time series models are widely used models in this regard. These are the linear statistical models that use past values to predict future behavior. The concept that history may repeat is the key motivation to design a model based on historical data [6].

Time series models dig out the linear relation between previous values with the current value and add randomness in it by including white noise that makes the models more realistic. Such models are called univariate models that only consider the single independent factor (historical data) to predict future results. Multivariate model on the other hand is also a time series model that considers not only historical data but also incorporate other economic factors affecting the stock market to predict the future behavior. The main assumption of both univariate and multivariate time series model is that it considers a linear relation between the independent variable and the dependent one. However, the financial time series are more often contain irregularity and nonlinearity [7].

Machine learning models can also be used to forecast the stock market. Machine learning is a branch of artificial intelligence which is based on learning and adoption of the patterns observed from the given dataset. Over the last few decades, machine learning algorithms have been widely used to predict the financial markets. Predictive models can be designed using machine learning algorithms based on both historical data and economic factors [8]. These models were found to be better predictive models as compared to that of econometric models, particularly for long-term prediction [9-11]. Secondly, unlike statistical models machine learning algorithms are capable of handling both unstructured and structured data and can generate quick conclusions [12]. Advanced machine learning models and approaches of sentiment analysis can help to get a clear picture of future movement of financial market. Thus, enable the investors to get the maximum return and minimum investment risk [13].

Deep learning models are also widely used models to understand the insights of fluctuations in financial markets. These models are quite efficient to dig out the complex relation between the financial markets with other attributes [14]. Deep learning models give improved results as compared to classical training models because of their multi-layered architecture that captures non-linear relationship among the variables [15].

Other than historical data and economic factors, the stock market is also affected by the trends, news and events [16]. As per efficient market hypothesis, financial markets are efficient in nature which means that any public or private information is translated to the market immediately [17]. This information can be collected from news, social sites, forums, tweets and trends. These factors change the sentiments of the investors and can change the decision of the investor and thus affect the stock prices [18]. Therefore, incorporating the sentiments in a predictive model is also another better approach to forecast the future behavior of the financial markets [19, 20].

The novelty of this paper is to elaborate the methodologies used to design the predictive model to forecast the future behavior of financial markets. Along with that it also discusses the factors that are helpful in the prediction of the financial market that contribute in model designing and improve the predictivity of the models. Other than historical data and economic factors, the stock market is also affected by the trends, news and events.

2. OVERVIEW

This survey is based on recent research articles published in different journals and conferences. The survey is performed to get the answers to the following research questions:

i. What are the different approaches used by the researchers to forecast the financial markets?

ii. Which attributes are used by the researchers that are found significant for the forecasting of financial markets?

iii. Does the combination of multiple factors improve the predictivity of financial market?

The key term used to perform the survey was "financial market prediction". Along with that, "time series prediction", "statistical models", "machine learning", "sentiment-based prediction", "news headlines" and "hybrid models" were used to make the searching criteria more specific.

In order to predict the financial market, we have two different approaches in research: technical analysis and fundamental analysis. In technical analysis, the financial markets are predicted using time series analysis based on historical data. Past values of the time series is the only provided data on which the predictive model is trained [21, 22]. Whereas in fundamental analysis, other factors that can affect the movement of the financial market are also incorporated while training any model. In this type of modelling the selection of features can be a challenging task to design an effective predictive model [23].

In this paper, the prediction of the financial market is categorized on the basis of the dataset used to train the model. Firstly, all those researches are filtered in which historical data is used to predict the future behavior of the financial market. After that those researches are considered in which financial markets are predicted using economic indicators. Lastly, those papers are reviewed in which textual data-based prediction of the financial market is performed that uses news, political events and investors' sentiments to forecast the future movement of the financial market.

In this survey, 145 research papers of the last twenty years are reviewed from the year 2001 to date. Figure 1 shows the count of reviewed papers with respect to their year of publication. The survey gives a brief summary of the attributes used in last 20 years by the researchers for the forecasting of financial market. It tells the significance of the features that play vital role in financial market fluctuations and thus improve the predictivity of the models. It further elaborates the techniques used for building the predictive models, the difference among them and the impact they make on the accuracy of the models.

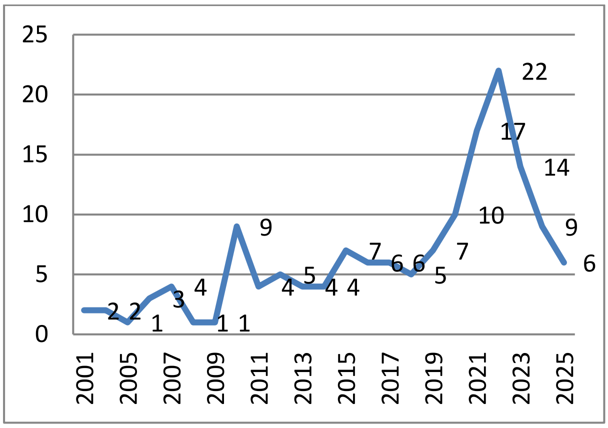


Fig 1: Year wise reviewed research papers

3. SURVEY METHODOLOGY

This section briefly elaborates multiple approaches adopted for the forecasting of financial markets. It is divided into three phases. In the first phase, recent research papers that focus on financial market forecasting are collected that are based on historical data. Significance of historical data in future prediction is reviewed along with the possible models used for the training. Three different approaches of model training are found; econometric model, machine learning model and hybrid model that can be used to design a predictive model based on historical data. 41 papers have been reviewed in this survey, which has used multiple predictive models to predict the financial market based on historical data.

In the second phase research papers with different training attributes are highlighted. External factors that make an impact on the financial markets are studied. The aim of this study is to understand the importance of economic factors or external factors in the movement of financial markets and their contribution in designing predictive models for the forecasting of financial markets. 26 papers have been reviewed in this survey that uses different models to predict financial market based on macroeconomic variables.

In the third phase, price movement in the financial market is observed based on trends, news headlines, political and socio-economic events. 43 research papers have been reviewed that focus on the textual data-based prediction of the financial market. The objective of this review is to identify that how these factors affect the financial market of any particular economy and whether incorporating these factors in designing a predictive model is beneficial or not. Figure 2 represents the count of research papers reviewed for the survey in which multiple features have been used for financial market prediction.

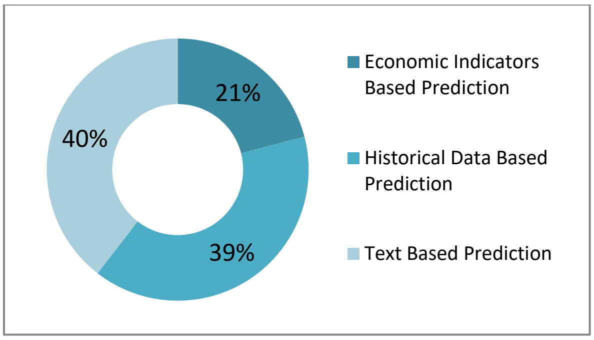


Fig 2: Feature wise reviewed research papers

The main reason behind dividing this systematic literature review into three phases is to observe the factors affecting the financial market separately and to determine how strongly they are correlated to the financial market. The second objective is to identify the recent approaches picked by the researchers to predict the financial market and which attributes are preferred to be included in the predictive model. Figure 3 represents the research papers studied during a survey that shows the data selection priorities of researchers with respect to the time. Figure 4 represents the complete breakage of the reviewed research papers using different approaches of model designing based on different data types.

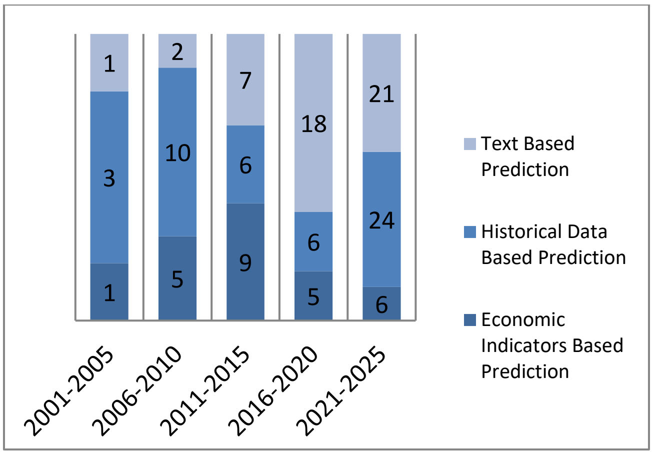


Fig 3: data and year breakage

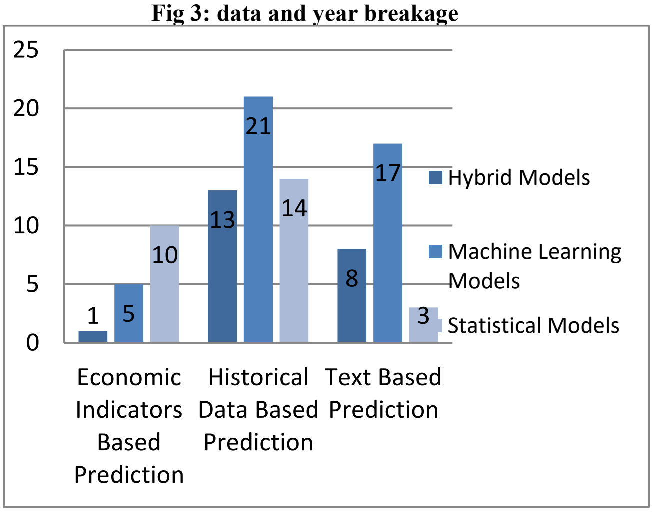


Fig 4: Count of reviewed research papers using different data types

3.1 Financial Market Prediction Based on Historical Data

Prediction of a stock market based on historical data is a classical method of the prediction of a market. Historical data is used to design a predictive model to forecast the future prices of the stocks with the belief that the market will behave in the same manner in future as it moved in the past.

Time series modelling is one of the approaches to forecast the future movements of a financial market based on historical data. Time series data or sequential data is numeric data collected at a specific time point in a given time interval. Both econometrics models and machine learning models can be used for time series analysis. The objective of time series prediction techniques is to predict the future behavior of the series based on past values [24]. In a stock market, the quotes given are Open, Close, High, Low and Volume collected at evenly spaced time points. These quotations can be used to design the predictive model such as a time series of closing price can be used as a predictor variable to forecast the future closing price of the stock [25]. Multiple techniques can be used for time series prediction including econometric models and machine learning models. Both approaches have been widely used to forecast the financial market.

3.1.1 Historical Data Based Statistical Models

Statistical models like the Autoregressive (AR) model, moving average (MA) model, Autoregressive moving average (ARMA) model and Autoregressive integrated moving average (ARIMA) model are the widely used models for the forecasting of time series. For financial time series, these models are used to predict the future expected return [26, 27]. While dealing with the time series data, two statistical measures are required to be predicted; the mean and variance, considered as the expected return and the risk associated with that return [28, 29]. Return and risk are equally significant for the prediction as a trader or an investor is always keen to know the future return along with the risk associated with it [29].

It is one of the properties of volatility that there is a co-movement among different stock markets. It is evident that the stocks with similar volatility have a tendency of greater co-movement than the stocks with different volatility [30]. Autoregressive Conditional Heteroscedasticity (ARCH) model is one of the time series models that is used to forecast the volatility or risk associated in a financial market. Generalized ARCH (GARCH) is a modified form of ARCH model [31]. Prediction of risk is equally significant as that of prediction of return in the financial market as the traders who invest in the market aim to get the maximum return with the minimum uncertainty [24]. [32] designed a hybrid ARMA-GARCH model on daily returns from NASDAQ stock exchange from 2000 to 2016 and designed a predictive model. It was observed that the designed hybrid model can forecast the daily return of the NASDAQ stock exchange with the error level of 1%.

[33] used ARIMA model to predict the time series of Italian banks' deposits. It was found that ARIMA model is good to predict the volume of deposits in the bank along with the changes and the time trends. It was also observed that this model is useful to deal with the seasonality in time series data. [34] studied the relationship between china stock market and global financial market statistically and depicted a unidirectional relationship.

A predictive model was designed to forecast the volatility of Nifty Realty Index and found that GARCH(1,1) model is the best fitted model to understand the past volatility calculated using squared residuals and is best to predict future volatility [35].Twelve years historical data was used by [36] to forecast the volatility of DJIA index using log-ARCH model. [37] used ARIMA model to predict the stock price of Apple stock based on historical data. It was found that the designed predictive model is good for short-term prediction and can be used for decision making.

From the literature review, it is concluded that the statistical/ econometric models are widely used models to forecast the future price of financial instruments based on current and historical prices. The predictions of these models are quite helpful for the investors to make investment decisions.

3.1.2 Historical Data Based Machine Learning Models

Statistical models like ARMA, ARCH and GARCH assumed the linear relation between the past and the future values. These models also consider the time series as stationary [38]. On the contrary, machine learning models are effective for non-stationary time series and can dig out the non-linear relationship between past and future values [8, 39]. These models are therefore widely used to predict the financial markets based on historical data to extract the complex non-linear relationship [40]. Several machine learning algorithms can be used in this regard like artificial neural network (ANN), support vector machine (SVM), random forest and so on. In research, these models have been widely used to forecast financial markets. Financial markets, particularly equity markets are non-linear in nature. Therefore, machine learning models are more effective for future predictions. [41] studied the significance of artificial intelligence in the field of finance by considering US financial markets. [42] compared multiple machine learning models with econometric models to predict the price movement of crypto currency. A comparison analysis was done among K-Nearest Neighbor (KNN), Auto ARIMA and Facebook's Prophet (Fbprophet) for crypto currency price prediction. [43] studied the role of deep learning models in predicting financial markets and concluded that the deep learning models are better predictive model as compared to the traditional models as they are more capable to explain the non-linear relation in a data. [44] also studied the role of deep learning in the field of finance and demonstrated the efficiency of deep learning models for the forecasting of stock market. Reinforcement learning is another technique that has been widely used in the domain of finance. [45, 46] have used deep reinforcement learning for financial market analysis

- **Artificial Neural Network**

Artificial neural network is one of the machine learning algorithms which is a human brain inspired algorithm consisting of artificial neurons similar to neurons of the human brain. It consists of three layers; input, hidden and output [47]. This model is capable to learn the patterns from non-linear data and therefore, performs better in the prediction of non-linear time series. [48] performed a recent survey in the domain of machine learning algorithms to predict the stock market and concluded that artificial intelligent models are the dominant machine learning models to predict the financial time series. [47] designed a predictive model using ANN for the Indian stock market and found ANN as an efficient model to predict Bombay stock exchange. A random walk model was compared with the probabilistic neural network model (PNN) by [49] to predict returns from the Taiwan stock index and found that PNN model performed better as compared to the other model. A feed-forward neural network and radial basis neural network with back propagation was designed to predict National Stock Exchange (NSE) indices and found that the feed-forward NN outperformed back propagation NN with the accuracy of almost 100% compared to the other model with 80% accuracy for trend prediction. On the contrary, radial basis NN showed better performance for stock price prediction than a feed-forward neural network [50]. [51] designed a predictive model based on a back propagation neural network to forecast the stock exchange of Thailand index. Mean Absolute Percentage Error (MAPE) was used to calculate the error level of the model and found the least error of 2% to predict the stock index. Using the past three values nine days ahead Athens stock exchange indicator was predicted by [52]. The back propagation neural network was designed to predict the time series that predict the future values with the root mean square error (RMSE) of 0.0024. A group of researchers forecasted the volatility of the stock market of China Pakistan Economic Corridor (CPEC) linked countries namely KSE-100 (Pakistan), SSE-100 (China), KASE (Kazakhstan), TADAWUL (Kingdom of Saudi Arabia), KLSE (Malaysia), MOEX (Russia), CAC40 (France), BIST (Turkey) and FTSE (United Kingdom). A machine learning model Non-linear AutoRegressive Neural Network (NAR) was used along with the traditional GARCH family models based on historical data from 2014 to 2021. It was found that the NAR model outperforms the traditional GARCH models [53].

- ***Long Short-Term Memory***

Long-Short Term Memory (LSTM) is another artificial neural network designed particularly for sequential data. This model is capable of dealing with the long sequential data as compared to other neural network models [54]. [55] used daily stock indices of S&P 500 from 1992 to 2005 and designed the LSTM model along with other neural network models. It was found that LSTM outperformed random forest, deep neural network (DNN) and logistic regression.

- ***Support Vector Machine***

Support vector machine is a machine learning algorithm that can be used for both regression and classification. This model can be used for time series prediction along with other machine learning applications [56]. [57] has proved that this model is a powerful predictive tool to forecast the stock market. [58] compared GARCH (1,1), neural network and support vector machine to predict the volatility of six different indices and found that the GARCH (1,1) and support vector machine are better predictive models than back propagation neural network. In accordance with the above literature review, it is found that both econometric/statistical models and machine learning models have been widely used to forecast the future movement of the financial market. It is found that both approaches are equally significant for forecasting. One approach may outperform the other depending on the dataset or the measuring tool of accuracy. As [59] compared Econometric models like ARIMA, SARIMA and Self Exciting Threshold AutoRegressive (SETAR) model with machine learning model NAR and found that NAR outperforms other models based on RMSE and Mean Absolute Error (MAE), but if the comparison is based on Mean Absolute Percentage Error (MAPE) SARIMA is found to be the best model.

3.1.3 Historical Data Based Hybrid Models

Hybrid model is the one in which two or more models are ensembled to design a single predictive model with the aim to design a better predictive model that adds up the performance of all the models used in it. These models combine the accuracy of different separate models and can contribute to achieve higher accuracy and precision [60]. [61] in a literature review of multiple research papers concluded that the hybrid models are capable of obtaining more accurate and promising results in forecasting of financial time series as compared to individual predictive models. Econometric/ Statistical models are good in capturing the linear relationship in a time series. If these models are combined with non-linear machine learning algorithms like neural networks, then the combined hybrid model can be proved to be a better predictive model that has the capability of analysing both linear and non-linear data patterns in a financial series [62]. [63] highlighted the significance of hybrid models for the forecasting of financial market. [64] forecasted the time series of Canadian lynx data and the IBM stock price indices using the hybrid model of ARIMA and ANN and concluded that the hybrid model performs more effectively than the ordinary ARIMA and ANN model. It was concluded that the designing of a hybrid model is an effective way to improve the accuracy achieved by either of the models separately. [65] used ARIMA and GARCH models to design a hybrid predictive model to predict the NSE index. The hybrid ARIMA-GARCH model was compared with ordinary ARIMA and GARCH and found that the ARIMA-GARCH hybrid model reduces the error and improves the performance over other models. Hybrid DSS based Hyper-Kernel SVM (HKSVM) was designed by [66] using historical financial data to forecast future movements. [67] designed a hybrid model using recursive empirical model decomposition (REMD) and long short-term memory (LSTM) for time series forecasting and concluded that the designed hybrid model is more effective than the individual REMD and LSTM models. It was found that the hybrid model has increased the accuracy by 20% compared with the LSTM model. [13] designed hybrid model of Neural network AutoRegressive (NNAR) model and GARCH using three different ways to predict the volatility of KSE-100 index and concluded that the linear combination of GARCH and NNAR is the best hybrid model to predict the volatility of KSE-100 index for both short and long term prediction. In 2022 a hybrid model by using the combined approach of Long Short-Term Memory, Auto encoder, and Deep Neural Networks (LSTM-AE-DNNs) was designed by [68] to predict Dow Jones daily stock index and concluded that the designed hybrid model is more efficient for the prediction then the individual models. A hybrid model of Multivariate Artificial Neural Network (MANN) and Dynamic Conditional Correlation GARCH (DCC-GARCH) was developed to forecast the volatility of five stock markets S&P 500, FTSE-100, KSE-100, KLSE and BSESN. It was concluded that the hybrid MANN-DCC-GARCH model is a better predictive model as compared to individual MANN and DCC-GARCH [69]. [70] used combined CNN with LSTM model to design a hybrid predictive model for the forecasting of stock price using historical data of OHLCV (Open High Low Close Volume). [71] used Generative Adversarial Network (GAN) with CNN to analyse inter stock correlations. 10 years historical data of multiple stock markets was considered in this study.

Figure 5 represents the count of year wise research papers reviewed that use historical dataset for the forecasting of financial market. Table 1 represents the complete summary of datasets used in the reviewed researches used for the financial market prediction. Table 2 represents the modelling approach used for the forecasting of financial markets based on past data.

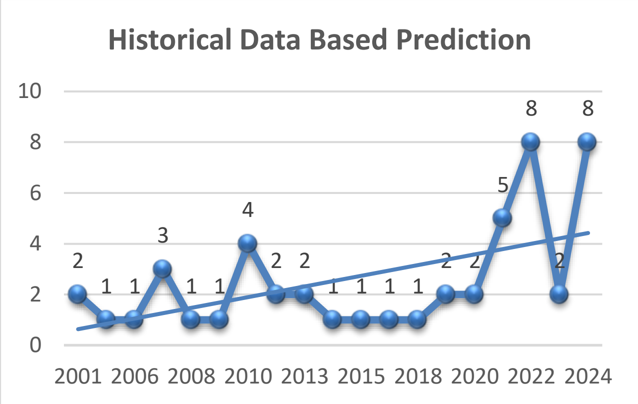


Fig 5: Year wise papers reviewed for historical data-based forecasting

Table 1. Summary table of historical data-based market prediction approach

Ref	Year	Data Employed
[49]	2003	Taiwan stock index.
[47]	2006	Indian stock market.
[52]	2007	Athens stock exchange indicator
[50]	2008	National Stock Exchange (NSE)
[33]	2010	Italian banks' deposits
[51]	2010	stock exchange of Thailand
[64]	2013	Canadian lynx data and the IBM stock price indices
[65]	2015	NSE index.
[55]	2018	Daily stock indices of S&P 500 from 1992 to 2005.
[32]	2022	Daily returns of NASDAQ stock exchange from 2000 to 2016.
[35]	2022	Nifty Realty Index
[37]	2022	Stock price of Apple.
[53]	2022	Daily volatility of KSE-100 (Pakistan), SSE-100 (China), KASE (Kazakhstan), TADAWUL (Kingdom of Saudi Arabia), KLSE (Malaysia), MOEX (Russia), CAC40 (France), BIST (Turkey) and FTSE (UK)
[13]	2022	Daily closing price of KSE-100 index.
[68]	2022	Dow Jones daily stock index
[69]	2022	Stock price of S&P 500, FTSE-100, KSE-100, KLSE and BSESN.

Table 2. Summary table of historical data-based market prediction approach

Ref	Year	Model	Model Type
[49]	2003	Random walk model and probabilistic neural network model (PNN)	Machine Learning Model
[47]	2006	ANN	Machine Learning Model
[50]	2008	A feed-forward neural network and radial basis neural network with back propagation	Machine Learning Model
[58]	2009	GARCH (1,1), neural network and SVM	Statistical and Machine Learning Model
[33]	2010	ARIMA	Statistical Model
[51]	2010	Back propagation neural network	Machine Learning Model
[64]	2013	ARIMA-ANN	Hybrid Model
[65]	2015	ARIMA-GARCH	Hybrid Model
[55]	2018	LSTM, deep neural network and logistic regression	Machine Learning Model
[59]	2021	ARIMA, SETAR and NAR	Statistical and Machine Learning Model
[32]	2022	ARIMA-GARCH	Statistical Model
[35]	2022	GARCH(1,1)	Statistical Model
[37]	2022	ARIMA	Statistical Model
[53]	2022	Non-linear AutoRegressive Neural Network	Machine Learning Model
[67]	2022	REMD and LSTM	Hybrid Model
[13]	2022	NNAR-GARCH	Hybrid Model
[68]	2022	LSTM, Auto encoder, and Deep Neural Networks (LSTM-AE-DNNs)	Hybrid Model
[69]	2022	Multivariate Artificial Neural Network (MANN) and Dynamic Conditional Correlation GARCH (DCC-GARCH).	Hybrid Model
[66]	2023	Hybrid DSS based Hyper-Kernel SVM (HKSVM)	Hybrid Model

3.2 Financial Market Prediction Based on Macroeconomic Variables

Financial market is sensitive to several economic factors like exchange rate, tax, crude oil prices, interest rate, US dollar rate and so on [72, 73]. These features can also be used along with historical data to forecast financial markets. [74] examined the causal relations between the stock price and macroeconomic variables. Industrial output, inflation and exchange rate were taken as basic macroeconomic factors to investigate the long and short-term relation with the stock market index. It was observed that in the short run there is unidirectional and bidirectional relation among the variables whereas in long run there is a cointegration relationship between stock price macroeconomic variables. A group of researchers in 2017 investigated the forecast ability of stock returns of S&P 500, NASDAQ and DJIA with respect to the Treasury bill interest rate. It was concluded that the forecasting is favorable for short-term prediction of five days and long-term prediction of 30 days [75]. [76] claimed that the foreign investor's decisions can be affected by local investments and therefore the macroeconomic environment should be strictly monitored. It was also claimed that changes in the macroeconomic variables had a significant effect on the stock market performance. The relation between the oil prices, exchange rates and prices of emerging stocks was studied and a strong correlation was found among them. It was found that the increase in the price of emerging stock causes the increase in oil prices. It was also observed that the positive shocks to oil prices decrease the stock prices of emerging markets and the US dollar exchange rate in the short run [77]. [78] studied the relation between the oil price and the gold price and concluded that the fluctuations in oil price influence gold and other metals in the market. In 2010 six macroeconomic indicators including gold price, US Dollar exchange rate, interest rate on T-bills, interest rate on deposits and closing price of 4 indices were used to predict the return on stock price index of the Istanbul Stock Exchange (ISE). It was concluded that using macroeconomic factors, the monthly price index can be successfully predicted with the accuracy of 98% [79]. [80] studied the relationship between the prices of precious metals and the exchange rate and found a significant long run relationship between them. [81] showed that the macroeconomic factors are the leading predictor to forecast the price of thirteen US sector indices.

From the research survey it is found that macroeconomic factors are the key indicators that can be used to predict the future behavior of the financial market. These factors are significantly correlated to the prices of financial instruments like equity, stock and currencies and incorporating these factors in predictive models can enhance and improve the predictivity of the model.

In order to predict the future movement of the financial market based on economic factors statistical models, machine learning models and hybrid models can be used.

3.2.1 Economic Factors Based Statistical Models

Several statistical models can be used to extract the relation between the financial market and other economic factors to design a better predictive model. Multivariate time series is one of the approaches that can be used to forecast the series of financial markets based on other financial time series. [82] used a nonlinear multivariate MS-ARMA-GARCH approach to predict the volatility of the Tehran Stock Exchange based on macroeconomic factors: inflation rate, money growth rate and exchange rate. [12] predicted crude oil price for five years ahead based on three economic factors: CPI, GDP and gold price using the econometric model Auto Regressive Integrated Moving Average with Exogenous Input (ARIMAX) along with other models. It was concluded that ARIMAX is the best predictive model with the least RMSE of 8.71. Econometric/ statistical model ARMA and E-GARCH was used by [44] to measure the impact of NIFTY, energy index and stock price index on the price of crude oil. In 2016 the volatility of Chinese stock market was predicted based on US economic factors using regressive models. It was found that the Chinese stock market is significantly correlated to the US economic variables [83].[76] investigated the reaction between macroeconomic variables and NSE. Three macroeconomic variables; Lending Interest rate, Inflation Rate and 91 day T-bill were used to predict the NSE indices using a regression model. [84] used the econometric model ARMA-EGARCH and found a significant effect of bit coin, gold price and exchange rate on crude oil prices. [85] forecasted the volatility of crude oil price using VAR (Vector AutoRegressive) model based on six econometric factors. VAR (Vector AutoRegressive) model was also used to investigate the interaction of Crude Oil Price, Consumer Price Level and Exchange Rate in Nigeria [86]. [87] used Exponential GARCH (EGARCH) to investigate the factors affecting the price of gold in the United States with the data from 2003 to 2016. It was found that there is a strong, linear and negative relation among gold return, Dollar return, oil price and silver price. [88] used econometric model GARCH and Structured VAR (SVAR) models to study the effect of financial, political and economic risk on the risk and return of the stock market. Macroeconomic factors including foreign exchange rate of hard currencies, interest rate and inflation rate were used along with the weighted average monthly indices of the companies listed on the Nairobi Securities Exchange in Kenya from January 2008 to December 2012. A multivariate regression analysis was performed to investigate the relation between the selected macroeconomic variables and the stock indices [89].

3.2.2 Economic Factors Based Machine Learning Models

Along with econometric models, machine learning models have been widely used to forecast price movement in financial markets based on economic factors. These models have the capability to extract the nonlinear relation between the macroeconomic factors and the prices of financial instruments. [90] studied Artificial Neural Network (ANN) model to review its application in the prediction of the financial market based on economic variables. In 2005 a predictive model using ANN was developed based on 31 financial and economic factors to forecast stock market returns [91]. [92] examined the effectiveness of incorporating external factors in designing a predictive model to forecast the stock market. ANN is used to design a model based on external indicators, such as commodity prices and currency exchange rates, in predicting movements in the Dow Jones Industrial Average (DJIA) index. [93] used 10 technical indicators to predict the future movement of Istanbul Stock Exchange. ANN and SVM were used for the prediction, and it was concluded that the average performance of ANN model is better than that of SVM. [94] used artificial neural network to examine the effectiveness of combined approach of technical analysis, fundamental analysis and of time series to predict price behavior.

3.2.3 Economic Factors Based Hybrid Models

Based on economic factors, several econometric series models have been used to forecast the financial market. Along with that multiple machine learning algorithms have also been used for the prediction of the financial market. A combined approach, in which two or more time series models or machine learning models ensembles for better prediction can be used for future forecasting of markets. [95] used the Wolf's sunspot data, the Canadian lynx data, and the British pound/US dollar exchange rate data from 1700 to 1987 to design a hybrid model by combining a statistical model ARIMA and a machine learning model Neural Network to take advantage of the unique strength of ANN and ARIMA in linear and non-linear modelling. It was concluded that the designed hybrid model is a better predictive model which is more effective and can improve forecasting accuracy.

Figure 6 represents the year wise number of research papers reviewed for this survey in which macroeconomic variables have been used for the financial market prediction. Table 3 shows a compete summary of the datasets used for macroeconomic based forecasting in the reviewed research papers. Table 4 represents the summary of model types used for financial market forecasting based on economic indicators.

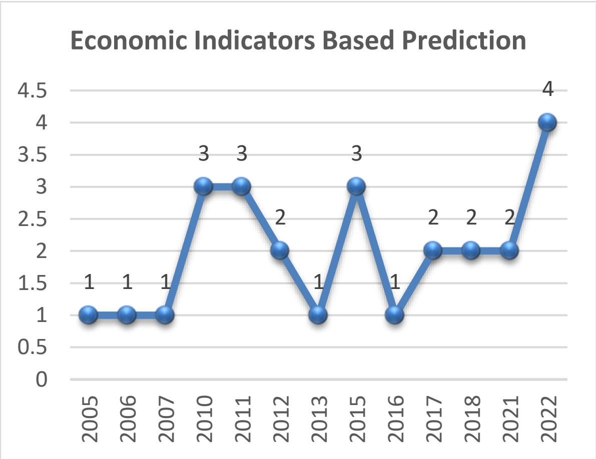


Fig 6: Year wise papers reviewed for macroeconomic variables-based forecasting

Table 2: Summary table of macroeconomic variables-based market prediction approach

Ref	Year	Data
[95]	2003	Wolf's sunspot data, the Canadian lynx data, and the British pound/US dollar exchange rate.
[92]	2005	Dow Jones Industrial Average (DJIA) index.
[93]	2010	Istanbul Stock Exchange along with 10 technical indicators.
[76]	2012	NSE indices, lending interest rate, inflation rate and 91 day T-bill
[86]	2015	Crude Oil Price, Consumer Price Level and Exchange Rate
[89]	2015	Foreign exchange rate of hard currencies, interest rate and inflation rate and monthly indices of the companies listed on the Nairobi Securities Exchange in Kenya.
[87]	2017	Gold return, Dollar return, oil price and silver price.
[84]	2021	Bit coin price, gold price and exchange rate on crude oil prices
[12]	2022	crude oil price, CPI, GDP and gold price
[44]	2022	NIFTY, energy index, stock price and price of crude oil.

Table 3: Summary table of models used for macroeconomic variables-based prediction

Ref	Year	Model	Model Type
[95]	2003	ANN-ARIMA	Hybrid Model
[91]	2005	ANN	Machine Learning Model
[92]	2005	ANN	Machine Learning Model
[93]	2010	ANN and SVM	Machine Learning Model
[94]	2013	ANN	Machine Learning Model
[86]	2015	VAR	Statistical Model
[84]	2021	ARMA-EGARCH	Statistical Model
[12]	2022	ARIMAX	Statistical Model
[44]	2022	ARMA and E-GARCH	Statistical Model
[85]	2022	VAR	Statistical Model
[88]	2022	GARCH and Structured VAR	Statistical Model

3.3 Text based Prediction of Financial Market

It is believed that the trends, financial news and political events make a significant impact on the sentiments of the traders and thus these factors affect the financial markets. Incorporating these factors while training a model can enhance the predictivity of the model. Google trends, tweets, local and international news and events are considered as the major attributes that have a direct impact on financial markets and can make a drastic change in the market. Sentiment based prediction of the stock market can be done using multiple forms of data like news data, tweets, trends and so on. It is found that the stock market is highly correlated with company's sentiments and tweets [96]. News and events have a significant impact on sales and prices of products and incorporating these factors in a training model can enhance the predictive of the models [97]. The significance of news for future prediction was studied by [98] and concluded that news data is one of the most significant attributes for future forecasting. [99] claimed that sentiment mining analysis can result in more predictable models for future forecasting. [100] discussed the impact of emotions on financial markets and concluded that the stock price movement is affected by the volume of tweets related to that stock or company. The importance of online emotions in predicting the stock market in China was investigated in 2016. Over 10 million stock-relevant tweets were analyzed, and it was concluded that various online emotions like joy, fear, disgust and sadness can contribute to predict five attributes of the stock market in China [101]. [102] performed a comprehensive review to understand how major stock market indices react to emotions extracted from financial news. [103] quantified the information from financial news using text mining analysis and concluded that the accuracy of predictive model can be enhanced by incorporating news data in the model. [104] claimed that the combination of all datasets including historical data, macro-economic variables and text based data, instead of using selected features, yields better predictive results.

Based on news and events financial markets can be forecasted using multiple approaches like using statistical models, machine learning models, deep learning models and hybrid models. These models have been frequently used for financial market prediction using news and events.

3.3.1 Textual Data Based Statistical Models

As far as statistical or econometric models are concerned, time series models are the best used models for this purpose. Time series model has been widely used econometric model for future forecasting whether it is financial data, weather data, sales or purchases, political data or as well as medical data. Sentiment analysis has also been done using classical time series models. [105] used the ARMA model for public opinion evaluation. ARMA model was used for multidimensional sentiment analysis on large scale textual data related to COVID-19. An asymmetric GARCH model was designed to predict the volatility of the Indian stock exchange based on positive and negative sentiments [106]. In 2015 it was investigated if stock market is sensitive to oil news, if yes then to what extent crude oil prices are driven by news. Regression models like Instrumental Variable Regression and Vector Autoregressive (VAR) Model was used to study the relation of news with WTI crude oil prices and a visible effect of news sentiments was noticed on oil market [107].

3.3.2 Textual Data Based Machine Learning Models

Machine learning models are the most frequently used models for financial market prediction. These models have the capability to train on multiple types of datasets like textual data, numeric data, images, audio and video. This is one of the reasons why such models are used for sentiment based predictions [108]. [109] studied the effectiveness of machine learning models for stock market prediction based on historical data and public tweets. [110] analyzed the relationship between social media messages and stock price and found a strong correlation between number of daily messages and volume of trade and a negative association between trade volume and financial indicators. [111] used multiple machine learning algorithms like Support Vector Machine (SVM), Multi-Layer Perceptron (MLP) and Radial Basis Function (RBF) to predict KSE-100 index based on external factors including Oil rates, Interest rate, Gold & Silver rates and Foreign Exchange (FEX) rate and sentiment features from NEWS and social media feed. It was found that MLP outperforms other machine learning models. In 2016 Support Vector Machine (SVM) model was used for financial market prediction and it was concluded that the forecasting models can be more efficient and effective if social media mining is combined along with other information [112]. [3] attempted to predict the Pakistan Stock exchange using Google trends along with macroeconomic factors by quantifying the semantics of the international market. Multiclass neural network and multiclass decision tree were used to design the predictive models. It was found that the multiclass decision tree outperforms multiclass neural network with an accuracy of 94%. Ten influential companies from NASDAQ were forecasted based on sentiment analysis using deep learning [113]. [114] forecasted the Hong Kong exchange market based on technical indicators and news sentiments using machine learning algorithms. It was found that LSTM outperforms Multiple Kernel Learning (MKL) and SVM in forecasting the financial market. A group of researchers in 2015 used macroeconomic indicators and domestic trends as public mood indicators to forecast the daily volatility of S&P 500. Predictive model was designed using LSTM for the prediction [115]. Using the same model, LSTM KSE-100 index was predicted based on multiple types of investor sentiments, using categorized financial news and historical data and categorized financial to understand the impact of financial news on stock market [116]. [117] proposed a web-based sentiment analysis approach. This technique was used to extract information from web text to forecast the oil prices based on compound, neutral, positive and negative sentiments. LSTM model was again used to predict the closing price of stock based on sentiment analysis [118]. [119] predicted the stock market based on multiple types of investor sentiments. Different machine learning models like LSTM, SVM and CNN were used to predict the financial market based on text mining [120]. [121] used random forest classifier to predict the direction of stock market index based on sentiment scores which were classified using Bidirectional Encoder Representations (BERT). Similarly, Recurrent Neural Network (RNN) model was used with character language for inter and intraday stock price prediction of Korean stock market [122]. [123] applied BERT to analyze the sentiments of Chinese stock review and achieved higher prediction accuracy. BERT model was compared by [124] with other machine learning models like LSTM, logistic regression, SentWordNet and found that BERT outperforms all other models in sentiment-based forecasting. [125] used deep learning models for opinion based prediction of stock market.

3.3.3 Textual Data Based Hybrid Models

Hybrid models have been widely used for the prediction of financial markets based on textual data. Semantic based, sentiment based and event extraction based techniques are used in this regard. A hybrid SVM-LSTM-GRU was designed by [126] to forecast the stock market based on news sentiments. [105] designed a hybrid model of Latent Dirichlet Allocation (LDA) and ARMA to predict the public opinion sentiments. It was found that the designed hybrid model LDA-ARMA has the average error of less than 5.64% and is effective to apply on large scale public sentiment evolution. ARIMA-GARCH was combined with LSTM for the prediction in which historical data was treated by ARIMA-GARCH whereas news sentiments and event embeddings were taken as the input of the LSTM model to predict the price of crude oil. It was found that the combined model is an effective predictive model to forecast crude oil price [127]. [128] designed a hybrid model of CNN and LSTM to forecast the bit coin price based on market sentiments. LSTM was also combined with Relational Graph Convolutional Network (RGCN) to understand the interconnectivity among the news using the news headlines to find out the relation of news headlines with stock price movement [129]. [113] ensembled multimodal AdaBoost and LSTM to predict the bit coin price based on trading data, block chain information and media sentiment. A hybrid system was proposed by [130] that apply text mining on social media data mining on the past stock dataset to enhance the prediction performance. In 2023, an emotion enhanced convolutional neural network (ECNN), the denoising autoencoder (DAE) models, and long short-term memory model (LSTM) were combined by [131] to design a hybrid model for the forecasting of risk and return of a stock market. [132] used hybrid deep learning models along with meta-learning model based on sentiments dataset and other macroeconomic features for multiple financial markets including cryptocurrency and equity markets.

Figure 7 represents the year wise papers reviewed for this survey that are using text-based data for the forecasting of financial market. Table 5 summarizes the dataset used in the studied research papers for text-based prediction. Table 6 shows a complete summary of different approaches to design a model for financial market prediction based on textual dataset.

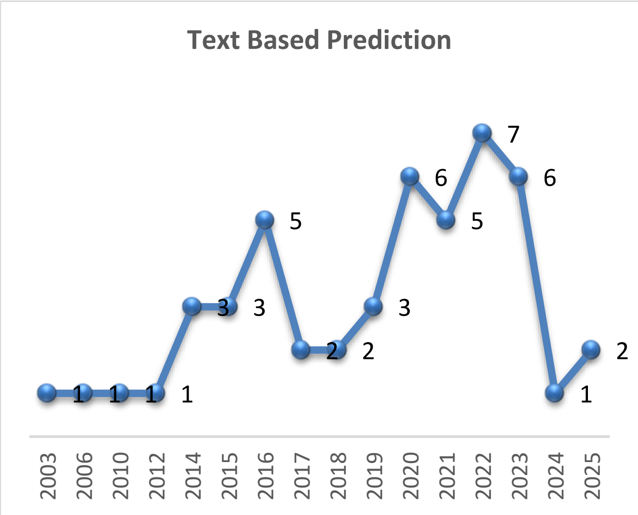


Fig 7: Year wise papers reviewed for text-based forecasting

Table 4: Summary table of text-based market prediction approach

Ref	Year	Data
[107]	2015	WTI crude oil prices
[115]	2015	Macroeconomic indicators, domestic trends and daily volatility of S&P 500.
[111]	2016	KSE-100 index, Oil rates, Interest rate, Gold & Silver rates and Foreign Exchange (FEX) rate and sentiment features from NEWS and social media feed.
[3]	2017	Pakistan Stock exchange, Google trends and macroeconomic factors
[122]	2017	Korean stock market.
[106]	2020	Indian stock exchange.
[114]	2020	Hong Kong exchange market, technical indicators and news sentiments
[105]	2021	Large scale textual data related to COVID-19.
[123]	2021	Chinese stock review
[116]	2023	KSE-100 and categorized financial news

Table 5: Summary table of models used for text-based prediction

Ref	Year	Model	Model Type
[105]	2021	ARMA	Statistical Model
[107]	2015	VAR	Statistical Model
[115]	2015	LSTM	Machine Learning Model
[111]	2016	SVM, MLP and RBF	Machine Learning Model
[99]	2016	SVM	Machine Learning Model
[3]	2017	Multiclass neural network and multiclass decision tree	Machine Learning Model
[122]	2017	RNN	Machine Learning Model
[106]	2020	Asymmetric GARCH	Statistical Model
[114]	2020	LSTM, MKL and SVM	Machine Learning Model
[118]	2020	LSTM	Machine Learning Model
[129]	2021	RGCN and LSTM	Hybrid Model
[133]	2021	BERT	Machine Learning Model
[124]	2021	BERT, LSTM, logistic regression and SentWordNet	Machine Learning Model
[105]	2021	LDA and ARMA	Hybrid Model
[127]	2022	LSTM, SVM and CNN.	Machine Learning Model
[120]	2022	Random forest classifier and BERT	Machine Learning Model
[121]	2022	SVM-LSTM-GRU	Hybrid Model
[126]	2022	CNN and LSTM	Hybrid Model

[128]	2022	Multimodal AdaBoost and LSTM	Hybrid Model
[113]	2023	LSTM	Machine Learning Model

3.4 Multiple NLP Approaches for textual Data

Text based forecasting of financial market can be classified into three different categories of Natural Language Processing (NLP): semantic based, sentiment based and event extraction based [134].

3.4.1 Semantic Based Forecasting of Stock Market

Semantic based analysis is one of the approaches in NLP to collect features from textual data for further modelling. Financial market forecasting using semantic based approach can be used to identify the interconnectivity impact between the textual data and financial market fluctuations. This approach has a limitation of processing of long-text data [134]. Multiple NLP techniques can be used for the extraction of semantics from textual data. Bag of Words (BOW) has been used by [135] and [136] that break up the text into list of words to quantify the words. [137] used Term frequency-inverse document frequency (TF-IDF) that calculates the terms present in a textual data and identify the importance of each term in a document. Pakistan Stock Exchange was predicted using multiclass decision tree based on Google trends and macroeconomic factors in which semantics of international market was quantified [3].

3.4.2 Sentiment Based Forecasting of Stock Market

Sentiment analysis is an NLP technique that is used to dig out the emotions, opinions or attitude of people towards a certain topic or event [138]. It gives result in a form of polarities; positive, negative, neutral or sarcastic [139]. Traders' sentiments play a vital role in the fluctuation of prices in stock market and can result in heavy loss and profits [140]. Sentiment based learning can be classified into three approaches; firstly, the lexicon-based approach that uses sentiment dictionaries to identify the polarities of each word. Secondly, the machine learning approach that trains the sentiment classifiers using different techniques like n-gram, BOW or TF-IDF and lastly, deep learning approaches that use neural networks to automatically extract the feature representation for sentiment measure [141]. LSTM model was used by [118] for the prediction of financial market based on sentiment analysis. [121, 133] analyse the different financial markets based on sentiment analysis using BERT model.

3.4.3 Event Extraction Based Forecasting of Stock Market

Event extraction is a technique that focuses to dig out the essential event information from textual dataset and represent the extracted data in a structured form. In other words, event extraction is used to convert the unstructured natural languages into structured one [142].

FTSE 50 stock index was predicted by [143] based on events from Reuters News articles extracted by ViewerPro system that filters irrelevant news and identifies events through pattern matching. Monitoring the factors that cause the occurrence of extreme events can lead to make effective entry and exit strategy for investors and can therefore play a significant role in decision making [144]. [145] designed an event based hybrid model for the prediction of two stock markets S&P 500 index and NIFTY with the prediction accuracy of 80% to 86%.The relationship between events and financial indicators for the prediction of S&P 500 was studied by [146] using dilated causal convolution networks with attention (Att-DCNN). Table 7 summarizes the multiple NLP approaches used for forecasting of financial market.

Table 6: Summary Table of forecasting of financial markets using different NLP approaches

Ref	Year	Model	NLP Approach
[135]	2006	BOW	Semantic Based
[143]	2014	ViewerPro	Event Based
[136]	2015	BOW	Semantic Based
[137]	2016	TF-IDF	Semantic Based
[3]	2017	Multi class decision tree	Semantic Based
[122]	2017	BERT	Sentiment Based
[118]	2020	LSTM	Sentiment Based
[123]	2021	BERT	Sentiment Based
[121]	2022	BERT	Sentiment Based
[146]	2020	Att-DCNN	Event Based

4. CONCLUSION

This survey is performed in order to have a clear picture of the methodologies; researchers have been adopting in order to predict financial markets. This survey includes the extensive study of forecasting of financial market based on multiple features and using different approaches. It has been found that financial market is sensitive to historical values, macroeconomic variables, news and events. Therefore, all these features contribute in order to have affective prediction of financial markets.

The extensive study of the review paper of almost two decades showed that the prediction of financial market based on historical and macroeconomic features is a classical approach but text-based forecasting of financial market was not very common technique before last few years. Financial market is sensitive to tweets, trends, news headlines and related discussions in other social media platform. Therefore, natural language-based prediction of financial market improves the predictivity of the market and incorporating the latest NLP techniques in this domain will help the researchers in getting efficiency in their results.

It concludes that as the financial market is sensitive to multiple factors therefore all these factors contribute in the efficient forecasting of financial markets. Incorporating different types of datasets, fusion of datasets and hybrid construction increase the predictivity of the future patterns of the market. Along with that advanced deep learning models along with classical statistical and machine learning models also improve the forecastability.

5. LIMITATIONS

It is a bigger picture of methodologies adopted for the forecasting of financial markets but not all approaches have been covered.

6. DECLARATION OF INTEREST

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

7. FUNDING DECLARATION

No Funding.

8. ACKNOWLEDGMENTS

Our thanks to our prestigious university, NED University of Engineering & technology for providing us research resources and platform.

9. REFERENCES

[1] Ankiewicz, M., *Effectiveness of Investing in Socially Responsible Companies during the Covid-19 Pandemic.* Procedia Comput Sci, 2021. **192**: p. 4732-4740.

[2] Huy, D.T.N. and N.T. Hang, *Factors that affect Stock Price and Beta CAPM of Vietnam Banks and Enhancing Management Information System–Case of Asia Commercial Bank.* Revista Geintec-Gestao Inovacao E Tecnologias, 2021. **11**(2): p. 302-308.

[3] Ahmed, F., et al., *Financial market prediction using Google Trends.* International Journal of Advanced Computer Science and Applications, 2017. **8**(7): p. 388-391.

[4] Dassanayake, W., *Critical comparison of statistical and deep learning models applied to the New Zealand Stock Market Index.* 2022.

[5] Ghadekar, P., et al., *Multi-day Window for Stock Movement Prediction and Financial News Classification for Predicting Market Sentiments.* International Journal of Next-Generation Computing, 2021. **12**(5).

[6] Murphy, J.J., *Technical analysis of the financial markets: A comprehensive guide to trading methods and applications*. 1999: Penguin.

[7] Rady, E., H. Fawzy, and A.M.A. Fattah, *Time series forecasting using tree based methods.* J. Stat. Appl. Probab, 2021. **10**(1): p. 229-244.

[8] Schnaubelt, M., *A comparison of machine learning model validation schemes for non-stationary time series data*. 2019, FAU Discussion Papers in Economics.

[9] Srivastava, A.K., et al., *Design of Machine-Learning Classifier for Stock Market Prediction.* SN Computer Science, 2022. **3**(1): p. 88.

[10] Ahmed, N.K., et al., *An empirical comparison of machine learning models for time series forecasting.* Econometric reviews, 2010. **29**(5-6): p. 594-621.

[11] Jin, X., L. Wei, and Q. Zhang, *The Stock Price Prediction Based on Time Series Model, Multifactorial Regression, Machine Learnings.* BCP Business & Management, 2022.

[12] Bhagat, V., M. Sharma, and A. Saxena. *Modelling the nexus of macro-economic variables with WTI Crude Oil Price: A Machine Learning Approach*. in *2022 IEEE Region 10 Symposium (TENSYMP)*. 2022. IEEE.

[13] Batool, K., M.F. Ahmed, and M.A. Ismail, *A hybrid model of machine learning model and econometrics' model to predict volatility of KSE-100 Index.* Reviews of Management Sciences, 2022. **4**(1): p. 225-239.

[14] Asl, M.M. and M. Kolahkaj, *Predicting Stock Prices in the Iranian Stock Market Using Convolutional Neural Network Optimization.* 2023.

[15] Hung, B.T., P. Chakrabarti, and P. Chatterjee, *Stock Prediction Using Multi Deep Learning Algorithms*, in *Computational Intelligence for Modern Business Systems: Emerging Applications and Strategies*. 2023, Springer. p. 97-113.

[16] Tajmazinani, M., et al., *Modeling stock price movements prediction based on news sentiment analysis and deep learning.* Annals of Financial Economics, 2022. **17**(01): p. 2250003.

[17] Malkiel, B.G., *Efficient market hypothesis*, in *Finance*. 1989, Springer. p. 127-134.

[18] Batool, K., U. Fatima, and M.F. Ahmed, *Trend Prediction of DJIA index based on News Extraction from Yahoo Finance.* International Journal of Computer Applications, 2025. **975**: p. 8887.

[19] Islam, M.Z., M.M.H. Chowdhury, and M.M. Sarker, *The impact of big data analytics on stock price prediction in the Bangladesh stock market: a machine learning approach.* International Journal of Science and Business, 2023. **28**(1): p. 219-228.

[20] Li, Q., et al., *A multimodal event-driven LSTM model for stock prediction using online news.* IEEE Transactions on Knowledge and Data Engineering, 2020. **33**(10): p. 3323-3337.

[21] Nam, K. and N. Seong, *Financial news-based stock movement prediction using causality analysis of influence in the Korean stock market.* Decision Support Systems, 2019. **117**: p. 100-112.

[22] Soni, S., *Applications of ANNs in stock market prediction: a survey.* International Journal of Computer Science & Engineering Technology, 2011. **2**(3): p. 71-83.

[23] Vui, C.S., et al. *A review of stock market prediction with Artificial neural network (ANN)*. in *2013 IEEE international conference on control system, computing and engineering*. 2013. IEEE.

[24] Chandra, A. and M. Thenmozhi, *On asymmetric relationship of India volatility index (India VIX) with stock market return and risk management.* Decision, 2015. **42**: p. 33-55.

[25] Rustam, Z. and P. Kintandani, *Application of support vector regression in indonesian stock price prediction with feature selection using particle swarm optimisation.* Modelling and Simulation in Engineering, 2019. **2019**(1): p. 8962717.

[26] Anaghi, M.F. and Y. Norouzi. *A model for stock price forecasting based on ARMA systems*. in *2012 2nd International Conference on Advances in Computational*

Tools for Engineering Applications (ACTEA). 2012. IEEE.

[27] Singh, S., K.S. Parmar, and J. Kumar, *Soft computing model coupled with statistical models to estimate future of stock market.* Neural Computing and Applications, 2021. **33**(13): p. 7629-7647.

[28] Fryzlewicz, P., *Lecture notes: Financial time series, arch and garch models.* University of Bristol, 2007.

[29] Tran, T., et al. *Neu-stock: stock market prediction based on financial news*. in *Proceedings of the 2nd International Conference on Human-centered Artificial Intelligence (Computing4Human 2021). CEUR Workshop Proceedings*. 2021.

[30] López-García, M.N., et al., *Volatility Co-movement in stock markets.* Mathematics, 2021. **9**(6): p. 598.

[31] Engle, R., *GARCH 101: The use of ARCH/GARCH models in applied econometrics.* Journal of economic perspectives, 2001. **15**(4): p. 157-168.

[32] Arashi, M. and M.M. Rounaghi, *Analysis of market efficiency and fractal feature of NASDAQ stock exchange: Time series modeling and forecasting of stock index using ARMA-GARCH model.* Future Business Journal, 2022. **8**(1): p. 14.

[33] Piscopo, G., *Italian deposits time series forecasting via functional data analysis.* Banks & bank systems, 2010(5, Iss. 1): p. 12-19.

[34] Najaf, K. and A. Chin, *The impact of the China Stock market on global financial markets during COVID-19.* International Journal of Public Sector Performance Management, 2024. **13**(1): p. 100-114.

[35] Jain, D. and S.K. Mittal, *Modeling Stock Market Return Volatility-Garch Evidence from Nifty Realty Index.* 2022.

[36] Mattera, R. and P. Otto, *Network log-ARCH models for forecasting stock market volatility.* International Journal of Forecasting, 2024. **40**(4): p. 1539-1555.

[37] Padma, A.P. and A.K. Mishra, *Forecasting on Stock Market Time Series Data Using Data Mining Techniques.* Dogo Rangsang Research Journal, 2022. **9**(1): p. 351-358.

[38] Cao, L. and F.E. Tay, *Financial forecasting using support vector machines.* Neural Computing & Applications, 2001. **10**: p. 184-192.

[39] Tealab, A., H. Hefny, and A. Badr, *Forecasting of nonlinear time series using ANN.* Future Computing and Informatics Journal, 2017. **2**(1): p. 39-47.

[40] Álvarez-Díaz, M., *Is it possible to accurately forecast the evolution of Brent crude oil prices? An answer based on parametric and nonparametric forecasting methods.* Empirical Economics, 2020. **59**(3): p. 1285-1305.

[41] Odonkor, B., et al., *Integrating artificial intelligence in accounting: A quantitative economic perspective for the future of US financial markets.* Finance & Accounting Research Journal, 2024. **6**(1): p. 56-78.

[42] Parikh, H., N. Panchal, and A. Sharma. *Cryptocurrency Price Prediction Using Machine Learning*. in *Proceedings of the 6th International Conference on Advance Computing and Intelligent Engineering: ICACIE 2021*. 2022. Springer.

[43] Gao, P., R. Zhang, and X. Yang, *The application of stock index price prediction with neural network.* Mathematical and Computational Applications, 2020. **25**(3): p. 53.

[44] Ghosh, A. and S. Banerjee, *Exploring the relevance of crude oil prices and installed generation capacity in prognosticating the NIFTY energy index.* Millennial Asia, 2023. **14**(4): p. 560-581.

[45] Zou, J., et al., *A novel deep reinforcement learning based automated stock trading system using cascaded lstm networks.* Expert Systems with Applications, 2024. **242**: p. 122801.

[46] Qiu, Y., R. Liu, and R.S. Lee, *The design and implementation of a deep reinforcement learning and quantum finance theory-inspired portfolio investment management system.* Expert Systems with Applications, 2024. **238**: p. 122243.

[47] Dase, R. and D. Pawar, *Application of Artificial Neural Network for stock market predictions: A review of literature.* 2010.

[48] Krollner, B., B. Vanstone, and G. Finnie. *Financial time series forecasting with machine learning techniques: A survey*. in *European Symposium on Artificial Neural Networks: Computational Intelligence and Machine Learning*. 2010.

[49] Chen, A.-S., M.T. Leung, and H. Daouk, *Application of neural networks to an emerging financial market: forecasting and trading the Taiwan Stock Index.* Computers & Operations Research, 2003. **30**(6): p. 901-923.

[50] Charkha, P.R. *Stock price prediction and trend prediction using neural networks*. in *2008 first international conference on emerging trends in engineering and technology*. 2008. IEEE.

[51] Sutheebanjard, P. and W. Premchaiswadi. *Stock exchange of Thailand index prediction using back propagation neural networks*. in *2010 Second International Conference on Computer and Network Technology*. 2010. IEEE.

[52] Hanias, M., P. Curtis, and J. Thalassinos, *Prediction with neural networks: the Athens stock exchange price indicator.* European Journal of Economics, Finance and Administrative Sciences, 2007. **9**: p. 21-27.

[53] Fraz, T.R., S. Fatima, and M. Uddin, *Comparing the forecast performance of nonlinear models and machine learning process. An empirical evaluation of GARCH family and NAR models in the light of CPEC.*

[54] Moghar, A. and M. Hamiche, *Stock market prediction using LSTM recurrent neural network.* Procedia computer science, 2020. **170**: p. 1168-1173.

[55] Fischer, T. and C. Krauss, *Deep learning with long short-term memory networks for financial market predictions.* European journal of operational research, 2018. **270**(2): p. 654-669.

[56] Ukil, A., *Intelligent systems and signal processing in power engineering*. 2007: Springer Science & Business Media.

[57] Hu, Z., J. Zhu, and K. Tse. *Stocks market prediction using support vector machine*. in *2013 6th international*

conference on information management, innovation management and industrial engineering. 2013. IEEE.

[58] Hossain, A., et al. *Comparison of GARCH, neural network and support vector machine in financial time series prediction*. in *Pattern Recognition and Machine Intelligence: Third International Conference, PReMI 2009 New Delhi, India, December 16-20, 2009 Proceedings 3*. 2009. Springer.

[59] Davidescu, A.A., S.-A. Apostu, and A. Paul, *Comparative analysis of different univariate forecasting methods in modelling and predicting the romanian unemployment rate for the period 2021–2022.* Entropy, 2021. **23**(3): p. 325.

[60] Hajirahimi, Z. and M. Khashei, *Hybridization of hybrid structures for time series forecasting: A review.* Artificial Intelligence Review, 2023. **56**(2): p. 1201-1261.

[61] Hajirahimi, Z. and M. Khashei, *Hybrid structures in time series modeling and forecasting: A review.* Engineering Applications of Artificial Intelligence, 2019. **86**: p. 83-106.

[62] Wang, J.-J., et al., *Stock index forecasting based on a hybrid model.* Omega, 2012. **40**(6): p. 758-766.

[63] Ge, Q., *Enhancing stock market Forecasting: A hybrid model for accurate prediction of S&P 500 and CSI 300 future prices.* Expert Systems with Applications, 2025. **260**: p. 125380.

[64] Wang, L., et al., *An ARIMA-ANN hybrid model for time series forecasting.* Systems Research and Behavioral Science, 2013. **30**(3): p. 244-259.

[65] Babu, C.N. and B.E. Reddy. *Selected Indian stock predictions using a hybrid ARIMA-GARCH model*. in *2014 International conference on advances in electronics computers and communications*. 2014. IEEE.

[66] Alsalamah, M., *HKSVM-DSS: novel machine learning-based approach for decision support system in stock market.* Inf. Sci. Lett, 2023. **12**(5): p. 2041-2053.

[67] Yang, Y., C. Fan, and H. Xiong, *A novel general-purpose hybrid model for time series forecasting.* Applied Intelligence, 2022. **52**(2): p. 2212-2223.

[68] Masalegou, S.M.B., et al., *A Stock Market Prediction Model Based on Deep Learning Networks.* Journal of System Management (JSM), 2022. **8**(4): p. 1-17.

[69] Fatima, S. and M. Uddin, *On the forecasting of multivariate financial time series using hybridization of DCC-GARCH model and multivariate ANNs.* Neural Computing and Applications, 2022. **34**(24): p. 21911-21925.

[70] Zhao, Q., Y. Hao, and X. Li, *Stock price prediction based on hybrid CNN-LSTM model.* 2024.

[71] Ma, D., et al., *VGC-GAN: A multi-graph convolution adversarial network for stock price prediction.* Expert Systems with Applications, 2024. **236**: p. 121204.

[72] Haleh, H., B.A. Moghaddam, and S. Ebrahimijam, *A new approach to forecasting stock price with EKF data fusion.* International Journal of Trade, Economics and Finance, 2011. **2**(2): p. 109.

[73] Nassirtoussi, A.K., T.Y. Wah, and D.N.C. Ling, *A novel FOREX prediction methodology based on fundamental data.* African Journal of Business Management, 2011. **5**(20): p. 8322.

[74] Karacaer, S. and A. Kapusuzoglu, *Investigating causal relations among stock market and macroeconomic variables: Evidence from Turkey.* Journal of Economic & Management Perspectives, 2010. **4**(3): p. 501.

[75] Joseph, A., M. Larrain, and C. Turner, *Daily stock returns characteristics and forecastability.* Procedia computer science, 2017. **114**: p. 481-490.

[76] Oriwo, E.A., *The relationship between macro economic variables and stock market performance in Kenya*. 2012.

[77] Basher, S.A., A.A. Haug, and P. Sadorsky, *Oil prices, exchange rates and emerging stock markets.* Energy economics, 2012. **34**(1): p. 227-240.

[78] Zhang, C. and X. Tu, *The effect of global oil price shocks on China's metal markets.* Energy Policy, 2016. **90**: p. 131-139.

[79] Boyacioglu, M.A. and D. Avci, *An adaptive network-based fuzzy inference system (ANFIS) for the prediction of stock market return: the case of the Istanbul stock exchange.* Expert Systems with Applications, 2010. **37**(12): p. 7908-7912.

[80] Yaqoob, T. and J. Iqbal, *Are precious metals hedge against financial and economic variables?: evidence from cointegration tests.* The Journal of Asian Finance, Economics and Business, 2021. **8**(1): p. 81-91.

[81] Weng, B., et al., *Macroeconomic indicators alone can predict the monthly closing price of major US indices: Insights from artificial intelligence, time-series analysis and hybrid models.* Applied Soft Computing, 2018. **71**: p. 685-697.

[82] Heidari, H., A. Refah-Kahriz, and N. Hashemi Berenjabadi, *Dynamic Relationship between Macroeconomic Variables and Stock Return Volatility in Tehran Stock Exchange: Multivariate MS ARMA GARCH Approach.* Quarterly Journal of Applied Theories of Economics, 2018. **5**(2): p. 223-250.

[83] Chen, J., et al., *Chinese stock market volatility and the role of US economic variables.* Pacific-Basin Finance Journal, 2016. **39**: p. 70-83.

[84] OZTURK, M.B.E. and S.C. Cavdar, *The contagion of COVID-19 pandemic on the volatilities of international crude oil prices, gold, exchange rates and Bitcoin.* The Journal of Asian Finance, Economics and Business (JAFEB), 2021. **8**(3): p. 171-179.

[85] Dąbrowski, M.A., et al., *The role of economic development for the effect of oil market shocks on oil-exporting countries. Evidence from the interacted panel VAR model.* Energy Economics, 2022. **110**: p. 106017.

[86] Obioma, B. and C. Eke, *An empirical analysis of crude oil price, consumer price level and exchange rate interaction in Nigeria: A vector autoregressive (VAR) approach.* American Journal of Economics, 2015. **5**(3): p. 385-393.

[87] Erdoğdu, A., *The most significant factors influencing the price of gold: An empirical analysis of the US market.* Economics, 2017. **5**(5): p. 399-406.

[88] Keshavarz, H. and M. Rezaei, *The Effect of Economic, Financial and Political Risk on the Risk and Return of*

Tehran Stock Exchange. Monetary & Financial Economics, 2022. **28**(22): p. 127-152.

[89] Kitati, E., E. Zablon, and H. Maithya, *Effect of macro-economic variables on stock market prices for the companies quoted on the nairobi securities exchange in Kenya.* International Journal of Sciences: Basic and Applied Research, 2015. **21**(2): p. 235-263.

[90] Li, Y. and W. Ma. *Applications of Artificial Neural Networks in Financial Economics: A Survey*. in *2010 International Symposium on Computational Intelligence and Design*. 2010.

[91] Enke, D. and S. Thawornwong, *The use of data mining and neural networks for forecasting stock market returns.* Expert Systems with Applications, 2005. **29**(4): p. 927-940.

[92] O'Connor, N. and M.G. Madden. *A Neural Network Approach to Predicting Stock Exchange Movements using External Factors*. in *Applications and Innovations in Intelligent Systems XIII*. 2006. London: Springer London.

[93] Kara, Y., M. Acar Boyacioglu, and Ö.K. Baykan, *Predicting direction of stock price index movement using artificial neural networks and support vector machines: The sample of the Istanbul Stock Exchange.* Expert Systems with Applications, 2011. **38**(5): p. 5311-5319.

[94] de Oliveira, F.A., C.N. Nobre, and L.E. Zárate, *Applying Artificial Neural Networks to prediction of stock price and improvement of the directional prediction index – Case study of PETR4, Petrobras, Brazil.* Expert Systems with Applications, 2013. **40**(18): p. 7596-7606.

[95] Zhang, G.P., *Time series forecasting using a hybrid ARIMA and neural network model.* Neurocomputing, 2003. **50**: p. 159-175.

[96] Pagolu, V.S., et al. *Sentiment analysis of Twitter data for predicting stock market movements*. in *2016 international conference on signal processing, communication, power and embedded system (SCOPES)*. 2016. IEEE.

[97] Tseng, K.-K., et al., *Price prediction of e-commerce products through Internet sentiment analysis.* Electronic Commerce Research, 2018. **18**(1): p. 65-88.

[98] Asur, S. and B.A. Huberman. *Predicting the Future with Social Media*. in *2010 IEEE/WIC/ACM International Conference on Web Intelligence and Intelligent Agent Technology*. 2010.

[99] Bing, L., K.C.C. Chan, and C. Ou. *Public Sentiment Analysis in Twitter Data for Prediction of a Company's Stock Price Movements*. in *2014 IEEE 11th International Conference on e-Business Engineering*. 2014.

[100] Peterson, R.L., *Trading on sentiment: The power of minds over markets*. 2016: John Wiley & Sons.

[101] Zhou, Z., J. Zhao, and K. Xu. *Can Online Emotions Predict the Stock Market in China?* in *Web Information Systems Engineering – WISE 2016*. 2016. Cham: Springer International Publishing.

[102] Saurabh Kamal, S.S., *A Comprehensive Review on Summarizing Financial News Using Deep Learning.* arXiv preprint, 2021.

[103] Li, X., et al., *Enhancing quantitative intra-day stock return prediction by integrating both market news and stock prices information.* Neurocomputing, 2014. **142**: p. 228-238.

[104] Batool, K. and U. Fatima, *Multi-Modal Data Driven Algorithm for Efficient Trade Market Prediction.* 2025.

[105] Zhuang, M., et al., *Analysis of public opinion evolution of COVID-19 based on LDA-ARMA hybrid model.* Complex & Intelligent Systems, 2021. **7**(6): p. 3165-3178.

[106] Paramanik, R.N. and V. Singhal, *Sentiment Analysis of Indian Stock Market Volatility.* Procedia Computer Science, 2020. **176**: p. 330-338.

[107] Feuerriegel, S., S.F. Heitzmann, and D. Neumann. *Do Investors Read Too Much into News? How News Sentiment Causes Price Formation*. in *2015 48th Hawaii International Conference on System Sciences*. 2015.

[108] Batool, K. and U. Fatima. *Analysis of High Frequency Markets: Global News Impact and Efficient Market Hypothesis Insights*. in *2025 International Conference on Artificial Intelligence, Computer, Data Sciences and Applications (ACDSA)*. 2025. IEEE.

[109] Sohrab Mokhtari, K.K.Y., Jin Liu, *Effectiveness of Artificial Intelligence in Stock Market Prediction based on Machine Learning.* arXiv preprint arXiv:2107.01031 2021.

[110] Garcia-Lopez, F.J., I. Batyrshin, and A. Gelbukh, *Analysis of relationships between tweets and stock market trends.* Journal of Intelligent & Fuzzy Systems, 2018. **34**(5): p. 3337-3347.

[111] Usmani, M., et al. *Stock market prediction using machine learning techniques*. in *2016 3rd International Conference on Computer and Information Sciences (ICCOINS)*. 2016.

[112] Wang, Y. and Y. Wang. *Using social media mining technology to assist in price prediction of stock market*. in *2016 IEEE International Conference on Big Data Analysis (ICBDA)*. 2016.

[113] Bouktif, S., A. Fiaz, and M. Awad, *Augmented Textual Features-Based Stock Market Prediction.* IEEE Access, 2020. **8**: p. 40269-40282.

[114] Li, X., P. Wu, and W. Wang, *Incorporating stock prices and news sentiments for stock market prediction: A case of Hong Kong.* Information Processing & Management, 2020. **57**(5): p. 102212.

[115] Ruoxuan Xiong, E.P.N., Yuan Shen, *Deep Learning Stock Volatility with Google Domestic Trends.* arXiv preprint arXiv:1512.04916 2015. **3**.

[116] Shazia Usmani , J.A.S., *LSTM based stock prediction using weighted and categorized financial news.* PloS one, 2023. **18**(3).

[117] Zhao, L.-T., et al., *Forecasting oil price using web-based sentiment analysis.* Energies, 2019. **12**(22): p. 4291.

[118] Jin, Z., Y. Yang, and Y. Liu, *Stock closing price prediction based on sentiment analysis and LSTM.* Neural Computing and Applications, 2020. **32**(13): p. 9713-9729.

[119] Deng, S., et al., *Dynamic forecasting of the Shanghai Stock Exchange index movement using multiple types of investor sentiment.* Applied Soft Computing, 2022. **125**: p. 109132.

[120] Lin, W.-C., C.-F. Tsai, and H. Chen, *Factors affecting text mining based stock prediction: Text feature representations, machine learning models, and news platforms.* Applied Soft Computing, 2022. **130**: p. 109673.

[121] Fazlija, B. and P. Harder *Using Financial News Sentiment for Stock Price Direction Prediction*. Mathematics, 2022. **10**, DOI: 10.3390/math10132156.

[122] dos Santos Pinheiro, L., Mark Dras, *Stock market prediction with deep learning: A character-based neural language model for event-based trading.* Proceedings of the Australasian Language Technology Association Workshop, 2017.

[123] Mingzheng, L., et al., *Sentiment analysis of Chinese stock reviews based on BERT model.* Applied Intelligence, 2021. **51**(7): p. 5016-5024.

[124] Alaparthi, S. and M. Mishra, *BERT: A sentiment analysis odyssey.* Journal of Marketing Analytics, 2021. **9**(2): p. 118-126.

[125] Albahli, S. and T. Nazir, *Opinion mining for stock trend prediction using deep learning.* Multimedia Tools and Applications, 2025. **84**(19): p. 21249-21272.

[126] Kumar, R., et al. *Emotion Analysis of News and Social Media Text for Stock Price Prediction using SVM-LSTM-GRU Composite Model*. in *2022 International Conference on Computational Intelligence and Sustainable Engineering Solutions (CISES)*. 2022.

[127] Liu, J. and X. Huang, *Forecasting Crude Oil Price Using Event Extraction.* IEEE Access, 2021. **9**: p. 149067-149076.

[128] Fakharchian, S., *Designing a forecasting assistant of the Bitcoin price based on deep learning using market sentiment analysis and multiple feature extraction.* Soft Computing, 2023. **27**(24): p. 18803-18827.

[129] Li, W., et al. *Modeling the stock relation with graph network for overnight stock movement prediction*. in *Proceedings of the twenty-ninth international conference on international joint conferences on artificial intelligence*. 2021.

[130] Sharaf, M., et al., *An efficient hybrid stock trend prediction system during COVID-19 pandemic based on stacked-LSTM and news sentiment analysis.* Multimedia tools and applications, 2023. **82**(16): p. 23945-23977.

[131] Zhao, Y. and G. Yang, *Deep learning-based integrated framework for stock price movement prediction.* Applied Soft Computing, 2023. **133**: p. 109921.

[132] Batool, K., M.M. Baig, and U. Fatima, *Accuracy and Efficiency in Financial Markets Forecasting Using Meta-Learning Under Resource Constraints.* Machine Learning with Applications, 2025: p. 100681.

[133] Li, M., et al., *Sentiment analysis of Chinese stock reviews based on BERT model.* Applied Intelligence, 2021. **51**(7): p. 5016-5024.

[134] Cheng, W.K., et al., *A review of sentiment, semantic and event-extraction-based approaches in stock forecasting.* Mathematics, 2022. **10**(14): p. 2437.

[135] Mittermayer, M.-A. and G.F. Knolmayer. *Newscats: A news categorization and trading system*. in *Sixth international conference on data mining (ICDM'06)*. 2006. Ieee.

[136] Luss, R. and A. d'Aspremont, *Predicting abnormal returns from news using text classification.* Quantitative Finance, 2015. **15**(6): p. 999-1012.

[137] Dadgar, S.M.H., M.S. Araghi, and M.M. Farahani. *A novel text mining approach based on TF-IDF and Support Vector Machine for news classification*. in *2016 IEEE International Conference on Engineering and Technology (ICETECH)*. 2016. IEEE.

[138] Chen, X., et al., *Sentiment analysis for stock market research: A bibliometric study.* Natural Language Processing Journal, 2025. **10**: p. 100125.

[139] Liu, B., *Sentiment analysis: Mining opinions, sentiments, and emotions*. 2022: Nota.

[140] Ahmed, M., *Stock Investors Sentiments and the Predictive Power of CBOE Volatility Index on Standard & Poor's 500 Returns.* 2023.

[141] Yadav, A. and D.K. Vishwakarma, *Sentiment analysis using deep learning architectures: a review.* Artificial Intelligence Review, 2020. **53**(6): p. 4335-4385.

[142] Xiang, W. and B. Wang, *A survey of event extraction from text.* IEEE Access, 2019. **7**: p. 173111-173137.

[143] Nuij, W., et al., *An automated framework for incorporating news into stock trading strategies.* IEEE transactions on knowledge and data engineering, 2013. **26**(4): p. 823-835.

[144] Rai, A., et al., *Detection and forecasting of extreme events in stock price triggered by fundamental, technical, and external factors.* Chaos, Solitons & Fractals, 2023. **173**: p. 113716.

[145] Bhanja, S. and A. Das, *A Black Swan event-based hybrid model for Indian stock markets' trends prediction.* Innovations in Systems and Software Engineering, 2024. **20**(2): p. 121-135.

[146] Daiya, D., M.-S. Wu, and C. Lin. *Stock movements prediction that integrates heterogeneous data sources using dilated causal convolution networks with attention*. in *ICASSP 2020-2020 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. 2020. IEEE.