International Review of Economics and Finance 109 (2026) 105423

Contents lists available at ScienceDirect

International Review of Economics and Finance

journal homepage: www.elsevier.com/locate/iref

Sentiment-based stock price prediction in developing countries: Evidence from Iran

Eshagh Jahangiri *, Marco Corazza

Department of Economics, Ca Foscari University of Venice, Venice, Veneto, Italy

ARTICLE INFO

JEL classification:
G11
C45
G17
C53

Keywords:
Stock price movement prediction
Developing country
Tehran stock exchange
Sentiment analysis
ParsBERT
Support vector machine

ABSTRACT

This study examines the predictive ability of public (investor) sentiment from web financial news and social media within the Tehran Stock Exchange (TSE) between November 1, 2020, and July 31, 2022. The study examines the influence of Persian-language sentiment data on movements in stock prices across various time horizons by employing a fine-tuned ParsBERT large language model architecture and a Support Vector Machine to predict the stock price movement over 1-day, 7-day, 30-day, and 90-day intervals. Findings show that the forecasting performance of public sentiment from web financial news exceeds that of social media in short-term predictions; however, social media sentiment provides more reliable and accurate projections for medium- to long-term intervals. Furthermore, combining sentiment data and technical indicators significantly enhances forecast accuracy. These findings highlight the unique role of sentiment analysis in understanding investor behavior in politically influenced and developing economies.

1. Introduction

The importance of the stock market is not hidden from anyone, and its effect on the economic growth of each country is well documented (Mamun et al., 2018; Rioja & Valev, 2014). Consequently, interest in predicting the price of stocks has increased among researchers and traders (Jishag et al., 2020; Lawal et al., 2020). The prediction of stock market trends seeks to forecast future stock prices, allowing investors to come to knowledgeable investment choices. The movement of the stock market depends on various factors, including economic conditions, political situations, corporate performance, news, and the behavior of traders (Hong, 2020). Thus, because of these influencing factors, the nature of the stock market is highly stochastic and volatile, and the prediction of movements becomes a challenging task. However, due to its impact on optimizing return on investment (ROI), it is a crucial responsibility for both traders and researchers.

There is extensive literature on stock price prediction methods due to their importance and profitability. Previously, the focus of the literature was more on the advanced and complex models, but as stated by Nardo et al. (2016), recent studies have shifted their attention from model sophistication to diverse sources of information, mainly textual data such as social media posts, web financial news, and blogs.

There are two main approaches for stock price prediction that are widely used among professionals and traders: technical analysis and fundamental analysis (Nti et al., 2020b). The former uses historical prices and volumes of one stock to predict its future price trend (Agarwal et al., 2019; Vijh et al., 2020) and the latter focuses on an analysis of economic aspects and information on industries and companies for the prediction of stock prices (Boachie et al., 2016; Ghaznavi et al., 2016). Based on the literature, most of the

* Correspondence to: Cannaregio 283, 30121, Venice, Italy.
E-mail address: eshagh.jahangiri@unive.it (E. Jahangiri).

https://doi.org/10.1016/j.iref.2026.105423
Received 15 April 2025; Received in revised form 31 January 2026; Accepted 20 May 2026
Available online 25 May 2026

studies (about 66%) used a technical analysis approach for stock price prediction followed by fundamental analysis (about 23%), and the rest on both of them (Nti et al., 2020b).

Web financial news and social media messages, such as ones on the platform X (formerly known as Twitter) seem to be effectively used for stock price trend forecasts, which can also affect stocks' returns (Mehta et al., 2021; Mu et al., 2023; Ren et al., 2020). Pagolu et al. (2016) claim that a company's share worth is influenced not only by past stock price but also by contemporary events, news, and forthcoming developments. Behavioral finance states that principal signals for successful market forecasting can be observed from news and views, making use of networks like social media websites to improve predictive accuracy (Li et al., 2017).

Researchers and traders cite specific instances illustrating the effect of web financial news and social media messages on stock price fluctuations, including the tweet from an account on April 23, 2013, stating, "Breaking: Two Explosions in the White House and Barack Obama is injured", resulting in a decline in the Dow Jones Industrial Average (DJIA) index. In the past decade, tweets of Elon Musk, CEO of Tesla car company, showed their ability to move stock price.[1] For example, his tweet on May 1, 2020, "Tesla stock price is too high imo"[2] has resulted in the 10% decrease of Tesla's stock price on that day.

Investor sentiment indicates market participants' perspectives on expected cash flows relative to an objective benchmark, namely the true intrinsic value of the asset being traded (Metawa et al., 2019; Zhang, 2008). Liu et al. (2023) analyze the synergistic connection between public sentiment and stock prices. They exhibited a substantial positive correlation, wherein an increase in stock price corresponds with a rise in investor sentiment and a decrease in stock price correlates with a decline in investor sentiment. They have also discovered that public sentiment on social media is proactive, which theoretically supports the use of investor sentiment for stock market prediction. Hence, public sentiment is a crucial factor of stock price fluctuations (Khan et al., 2022).

Bustos and Pomares-Quimbaya (2020) shows an increase in using social media and news to extract investor sentiment for stock price prediction. There are different tools for deriving public (investor) sentiment from news or social media messages that people can use for stock price prediction. Natural Language Processing (NLP) is crucial in analyzing written content. NLP is a field dedicated to enabling communication between humans and computers. NLP enables computers to comprehend, decode, and produce natural language in a manner analogous to that of humans (Hirschberg & Manning, 2015). Sentiment analysis is a crucial component of NLP, which leverages computational methods to identify and classify the sentiment of textual data, including web financial news or social media messages (Picasso et al., 2019).

The literature on stock price prediction using textual data illustrates conflicting results. Some studies, such as Baker and Wurgler (2006) support the predictive capabilities of these data, while some earlier studies argue that social data lacks the ability to predict stock price movement (Antweiler & Frank, 2004). Therefore, using textual data (social media, web financial news, etc.) in stock price prediction remains a topic of ongoing debate for researchers. Consequently, as expressed by Ashtiani and Raahemi (2023), the effort of stock market forecasting has not yet fully leveraged news and social media, and it needs more attention and studies to provide more results regarding their impact.

The paper by Nti et al. (2020b) shows that the majority of studies in stock market fluctuations forecast employing sentiment analysis focused on developed countries, and there is little research in this field in developing nations. Furthermore, the extent of stock market inefficiency and irrationality of investor behavior varies among countries, necessitating further research on forecasting stock markets through sentiment modeling in developing nations by scholars and investors (Agarwal et al., 2019).

While there are some studies about forecasting stock prices by sentiment analysis on the stock markets of developing countries, Iran is economically and politically distinct, necessitating further examination. The United States and European Union have imposed significant sanctions on various sectors of Iran, such as international trade, banking, shipping industry, and Energy/petroleum industry, resulting in several countries ceasing trade in goods, including oil and other commodities, from Iran.[3] The Iranian banking system is presently not fully integrated with the global financial network, and international transactions are hindered as most foreign assets have been inaccessible due to sanctions. The Iranian economy is marked by substantial government intervention, close to a state-controlled system. The most recent Index of Economic Freedom positioned Iran at 169 out of 176 countries in the ranking.[4]

A limited amount of research exists about stock price movement prediction focused on the Tehran Stock Exchange (TSE), Iran, utilizing sentiment analysis (Ashtiani & Raahemi, 2023). The Tehran Stock Exchange (TSE) is notably different from other exchanges in developing nations due to its significant integration with the government and state-owned enterprises. The Iranian government has progressively depended on the TSE to generate revenue and mitigate budget deficits resulting from sanctions. This establishes an unusual environment in which the state influences changes in market behaviors, resulting in hazards such as market manipulation and artificial valuation fluctuations. A notable distinguishing characteristic is the large number of retail investors, rendering the TSE more vulnerable to volatility in comparison to other exchanges. The distinctive combination of political influence, state intervention, and increased individual engagement indicates that the TSE's performance is deeply linked to wider economic and political dynamics in Iran, making it a vital case for comprehending the impact of public sentiment in stock market fluctuations in politically restricted markets.[5]

The "Price Limit Domain" rule is a significant distinction between the TSE and other stock markets. It says that stock prices cannot move more than 2% from the previous day's closing price either way. Consequently, the maximum daily return can fluctuate by ±2%, remaining constant and independent of the trading day's duration. The NYSE has a comparable regulation known as the Limit

[1] https://www.cnbc.com/2021/01/29/elon-musks-tweets-are-moving-markets.html.
[2] https://twitter.com/elonmusk/status/1256239815256797184
[3] https://www.worldbank.org/en/country/iran/overview
[4] https://www.heritage.org/index/pages/all-country-scores
[5] https://www.mei.edu/publications/iranian-governments-risky-stock-market-bet

Up/Limit Down (LULD) Plan[6]; however, its range is not as constrained as that of the TSE and is contingent upon the time of day, indicating a degree of flexibility. Furthermore, under the LULD scheme, the price limit is determined based on the average price from the preceding five-minute interval, but in the TSE, it is assessed in relation to the preceding day's closing price.

While the TSE is subject to unique constraints such as strict price limits and heavy government intervention, these factors create a unique environment for testing sentiment analysis. In a market where traditional fundamental analysis is often obscured by recurrent political shocks and sanctions, investor sentiment becomes a primary driver of market behavior, making it a critical case study for high-intervention emerging markets (Moradi et al., 2021; Tohidi, 2022).

This study, based on the above discussions, assesses the effect of investor sentiment through two main sources, social media messages and financial news datasets, on stock price fluctuations in Tehran Stock Exchange (TSE),[7] Iran. This research empirically investigates whether information from social media and web-based financial news enhances the prediction of stock price changes, examines the added value of investor sentiment data in improving stock prediction models that rely solely on historical prices, and evaluates which source of investor sentiment, social media or financial news, offers greater predictive power.

Although this research is not the first study to consider the investor sentiment for predicting stock market fluctuations at TSE, it possesses three characteristics that make it novel: (i) a comparison among the predictive ability of investor sentiment extracted from social media and financial news is performed, which up to the author's knowledge, marking the first such analysis at the TSE; (ii) while most studies just analyzed the usefulness of investor sentiment in stock price prediction, this paper assesses whether the information extracted from social media and financial news provides additional value beyond those that are already available from the historical prices; (iii) this work is, up to the author's knowledge, the first research utilizing a Large Language Model (LLM) to analyze and extract information from social media messages and web financial news to forecast changes in stock prices on the TSE.

The remainder of the paper is organized as follows: Section 2 presents some theoretical concepts and reviews some related studies to this research. Section 3 outlines the methodology and data used in this research. Section 4 presents the results, and the paper concludes in Section 5 with some suggestions for further work.

2. Background and related literature

These days, news websites, social media networks, forums, and blogs are great information sources for various purposes in various industries. There is also great interest from academia, with over 2300 publications using this information in the last 15 years. Therefore, many applications have emerged to use the public opinion extracted from these sources with the help of sentiment analysis (Rodríguez-Ibánez et al., 2023).

This section provides the theoretical framework for this study and a review of analogous studies employing sentiment analysis to predict stock price fluctuations.

2.1. Theoretical foundation

Natural Language Processing (NLP) has recently gained popularity because of its wide range of applications in various fields (Khurana et al., 2023). Using NLP techniques, unstructured financial news or social media data can be transformed into a structured form, which can then be used for extracting the investor sentiment regarding a specific stock or the overall market. There are various applications of NLP such as chatbots and virtual assistants (e.g. ChatGPT) which significantly help humans in many ways like finding the answers to their questions about various things and writing code for various programming languages and many more.

Sentiment analysis mainly uses two approaches (Drus & Khalid, 2019). Lexicon-based sentiment analysis is one of them, and it is a pre-developed dictionary of tagged words. The corpus of the dictionary includes words that are tagged as positive, negative, or neutral. This method breaks down the sentence into individual words and then searches for these words in the lexicon (dictionary) to find their sentiments (Taboada et al., 2011). Aggregating each word's sentiment infers the sentiment of the sentence. If the aggregate of positive sentiment scores dominates, the sentence is classified as positive, and negative otherwise. If none outweighs the other, it may be classified as neutral. Despite its simplicity, the lexicon-based approach has some limitations, such as its inability to account for word meanings based on their usage (Yadav & Vishwakarma, 2020). For example, the word "goat" is animal; however, recently, in the context of football, it is considered as "Greatest Of All Time", which means the greatest football player of all time and is a positive sentiment. However, the lexicon-based method cannot identify the context and classify the sentiment of words based on that.

The other approach for sentiment analysis is Machine Learning (ML) algorithms, which determine the sentiment using statistical models (Singh & Jaiswal, 2023). Machine learning techniques are widely used for sentiment analysis to overcome some disadvantages of lexicon-based methods. ML algorithms contain supervised learning algorithms, such as Naive Bayes or Support Vector Machines, and deep learning approaches, namely Recurrent Neural Networks or Transformers. These methods are able to understand the context-specific meanings of words which was the limitation of lexicon-based models. ML methods can classify the sentiments of large volumes of textual data, which is helpful for applications in finance that a high volume of financial news is produced daily.

[6] https://www.luldplan.com

[7] https://tse.ir

Large Language Models (LLMs) are a type of Artificial Intelligence (AI) that uses a machine learning (deep learning) approach to understand and produce natural language like humans (Min et al., 2023). LLMs are trained on massive amounts of textual data to understand context, translate languages, summarize information, and do tasks like sentiment analysis. LLMs have profoundly transformed the study environment of sentiment analysis by establishing a novel paradigm that leverages deep learning and extensive pre-training on varied textual data. Numerous studies have underscored the significance of LLMs in sentiment analysis, accentuating their efficacy in capturing contextual nuances, comprehending sarcasm, and managing complex linguistic patterns (Chen et al., 2020; Zhang et al., 2018b).

Some well-known LLMs include Open AI's GPT (Brown et al., 2020), Google's BERT (Devlin et al., 2018), and Meta's Llama (Touvron et al., 2023). Open AI's GPT chatbot and virtual assistant, known as ChatGPT, was released in November 2022 and quickly gained popularity.

Emotions have an effect on investors' trading decisions and can change them (Zaleskiewicz & Traczyk, 2020). Thus, it is vital to study the sentiment of web financial news and social media messages that affect investors' perceptions about stock exchange markets. Thinking of utilizing the investor sentiment for financial predictions has been started by the introduction of Keynes Beauty Contest Analogy in the early twentieth century. This analogy was proposed by Maynard Keynes in 1936, it was a contest to select the six most beautiful faces from a hundred pictures. The best strategy is to choose what others might think of the most beautiful faces and choose them. This thought applies also in investing that investors and traders buy stocks, which they think maybe other investors might buy those stocks. This is the effect of investor (public) sentiment on financial markets (Barber & Odean, 2008). By using NLP approaches (sentiment analysis), useful information can be extracted from texts and classified based on their sentiments, positive, negative, or neutral.

It is worth noting that one of the innovative aspects of this research is the use of a LLM for sentiment analysis. To the best of the author's knowledge, this is the first study employing a LLM for sentiment analysis of financial textual data in the Persian language.

2.2. Related literature

The three primary textual datasets for extracting investor sentiment using sentiment analysis for stock price predictions are web financial news, social media posts, and web search queries (Fataliyev et al., 2021). This study concentrates on the first two sources, examining related research that has utilized them. This section includes four subsections for the examination of the relevant literature. The first subsection will analyze studies that have utilized social media for stock market prediction, followed by a subsection that reviews papers employing web financial news for the same purpose. The remaining subsections discuss studies undertaken in developing countries and those focused on the Tehran Stock Exchange (TSE).

2.3. Social media

Bollen et al. (2011) examined the forecasting power of social media, tweets, for the Dow Jones Industrial Average (DJIA) and it showed that its effect is statistically significant and the sentiment of tweets can predict the movement of DJIA with around 86.7% accuracy. Another paper, Sun et al. (2016), utilized *StockTwits*[8] platform to extract the investors' sentiment towards a market for stock price movement prediction. Their proposed model predicted the stock price movement with 51.7% accuracy on average.

Coyne et al. (2017) used posts published by potential investors and traders on the *StockTwits* platform to forecast stock price changes. They provided three models for the stock price prediction using various NLP techniques. Their top-performing approach reached an accuracy of 65% in forecasting stock price changes. The model for predicting stock price movements was introduced by Hasan and Fong (2018), tested on the Saudi Stock Market utilizing a lexicon-based technique for sentiment analysis, achieving a minimum recall of 69% and a maximum of 96%.

Zhou et al. (2018) demonstrated the feasibility of predicting the Chinese stock market through Granger causality tests and correlation analysis. They utilized the Chinese tweets on the website *Weibo*, similar to the X platform, and assessed the effect of various online emotions, such as joy, sadness, and others, on the Chinese stock market. Sagala et al. (2020) demonstrated the significance of sentiment assessment in forecasting stock price fluctuations. They demonstrated the highest accuracy by integrating elements of historical data and social media sentiment while employing the Support Vector Machine algorithm.

2.4. Web financial news

Financial news from well-known news websites is another important source that many researchers and investors have used for stock price movement prediction. Various research has proven the effectiveness of news in predicting stock price trends, such as Shah et al. (2018) and Zhang et al. (2018a). Another research on stock market movement forecast employing news is Feuerriegel and Gordon (2018) in which they showed the long-term effect of news on stock price trends. They illustrated that in the long-term, the error of prediction by using financial news is decreased.

Mate et al. (2019) utilized New York Times news articles' headlines to extract public sentiment to forecast the stock price movement. They have used lexicon-based sentiment analysis to get the sentiment of news articles. The paper Jishag et al. (2020) examined the effect of financial news reported on various online sources, such as *nasdaq.com* and *Yahoo Finance* and provided

[8] www.stocktwits.com

the model which could predict the stock price movement with around 89% accuracy. They used the lexicon-based approach for extracting the sentiment of the news.

The study by Li (2020) examined the impact of news on predicting stock price trends. It used financial news from Reuters and the *Reddit* Forum to test how well financial news can represent changes in stock prices. They used the Long Short-Term Memory (LSTM) model, which is a common machine learning algorithm used in financial predictions to figure out how stock prices would move in the future. Eck et al. (2021) used financial news provided by the German financial news provider named *OnVista* to predict stock price movement utilizing various machine learning methods, and they achieved the highest accuracy, 70%, with the Support Vector Machine (SVM) algorithm. The stock price has been predicted using various models, including Random Forest (RF), Logistic Regression (LR), and Naive Bayes (NB).

One recent paper in this field is Kabbani and Usta (2022), which used the big data of financial news from 27 major news websites consisting 2.7 million articles from 2016 to 2020. They utilized lexicon-based sentiment analysis to derive the emotion of the news. They predicted stock price movements using three different machine learning methods: Logistic Regression, Random Forest, and Gradient Boosting Machine. The Random Forest model achieved the highest accuracy of 63%.

2.5. Developing countries

Gunduz and Cataltepe (2015) predicted the changes in the Borsa Index of Istanbul utilizing Turkish news. They used the Naive Bayes classifier to predict the movement and employed the Multi-Layer Perceptron (MLP) to categorize the news mood. For training a MLP classifier, they manually labeled several news articles to serve as training set.

Gálvez and Gravano (2017) proposed the Random Forest model to forecast the stock price changes in the Buenos Aires Stock Exchange using traders' comments posted on online message boards. They achieved the best performance when they combined technical indicators with message board data (62.5% accuracy).

Nti et al. (2020a) assessed the performance of predicting stock price movement using various textual sources, such as Twitter and financial news websites, in the Ghana Stock Exchange. They used lexicon-based approach for sentiment analysis and found that social media data (Twitter) is more accurate than financial news websites for stock price trend prediction. The sentiment of these sources has been found to have higher accuracy in forecasting stock price movements for longer horizons.

The study by Carosia et al. (2020) examined the impact of tweet sentiments influencing fluctuations in the Brazilian stock market. It established the correlation between investor sentiment and market dynamics. They have demonstrated that expanding the window size enhances the accuracy of forecasting stock market movements.

2.6. Tehran stock exchange (TSE)

Derakhshan and Beigy (2019) employed social media comments to predict stock price fluctuations in the Tehran Stock Exchange (TSE). The method discussed is Latent Dirichlet Allocation with Part-of-Speech tagging (LDA-POS), which classifies user comments according to parts of speech, including nouns, adjectives, verbs, and prepositions, thereby extracting sentiment information relevant to topics. The research employed this methodology on a Persian dataset obtained from the *Sahamyab* social network, resulting in a prediction accuracy of 55.33% for prominent shares in TSE. The LDA-POS algorithm surpassed traditional price-only and LDA-based models, exhibiting a 2.33% enhancement above baseline methods, so illustrating the efficacy of sentiment-based features for stock market prediction within the Iranian context.

Tajmazinani et al. (2022) analyzes the effect of integrating technical analysis with Persian news mood on forecasting Iranian stock prices. This paper presents the HESNEGAR lexicon for sentiment analysis, filling in the gaps of Persian-specific tools (Asgarian et al., 2018). The writers collected and analyzed news stories about stocks in Iran using the HESNEGAR lexicon to find out how people felt about different topics. Convolutional Neural Networks (CNNs) were employed to evaluate a price-only, news-sentiment, and hybrid model that integrates both. They demonstrated that the combined method outperformed the other models, which demonstrates the value of integrating technical indicators with public opinion in stock trading and prediction.

Ebrahimian et al. (2023) focuses on predicting stock prices in the Tehran Stock Exchange (TSE) with a hybrid methodology that integrates sentiment analysis of social media with technical analysis. This study employs the sentiment analysis method, which involves acquiring and classifying user remarks from the Persian social media platform, *Sahamyab*. The sentiment classification models were trained and evaluated by the researchers whom manually classified a sample of 2344 tweets into positive and negative classifications using machine learning algorithms such as Naive Bayes, Support Vector Machines (SVM), and Decision Trees. It checks models using two different sets of criteria: Technical Analysis (TA) alone, and Technical Analysis and Sentiment Analysis (TASA) together. Both of these use the Support Vector Regression (SVR) method. The results show that sentiment study, even though it gives us information about market trends, does not significantly improve the accuracy of stock price predictions when used with technical indicators.

Research on TSE indicates that there is still an unanswered question regarding how well and when to use social media and financial news to predict stock prices. Additionally, most studies have primarily focused on lexicon-based or machine learning approaches to sentiment analysis. This research seeks to enhance stock price forecasting by examining the impact of social media messages and web financial news through advanced sentiment analysis, investigating whether incorporating investor sentiment data improves prediction models based solely on historical prices, and evaluating the forecasting effectiveness of these two primary sentiment sources. Additionally, it introduces a fine-tuned large language model (LLM) for the sentiment analysis of Persian financial textual data, marking a novel contribution in this domain.

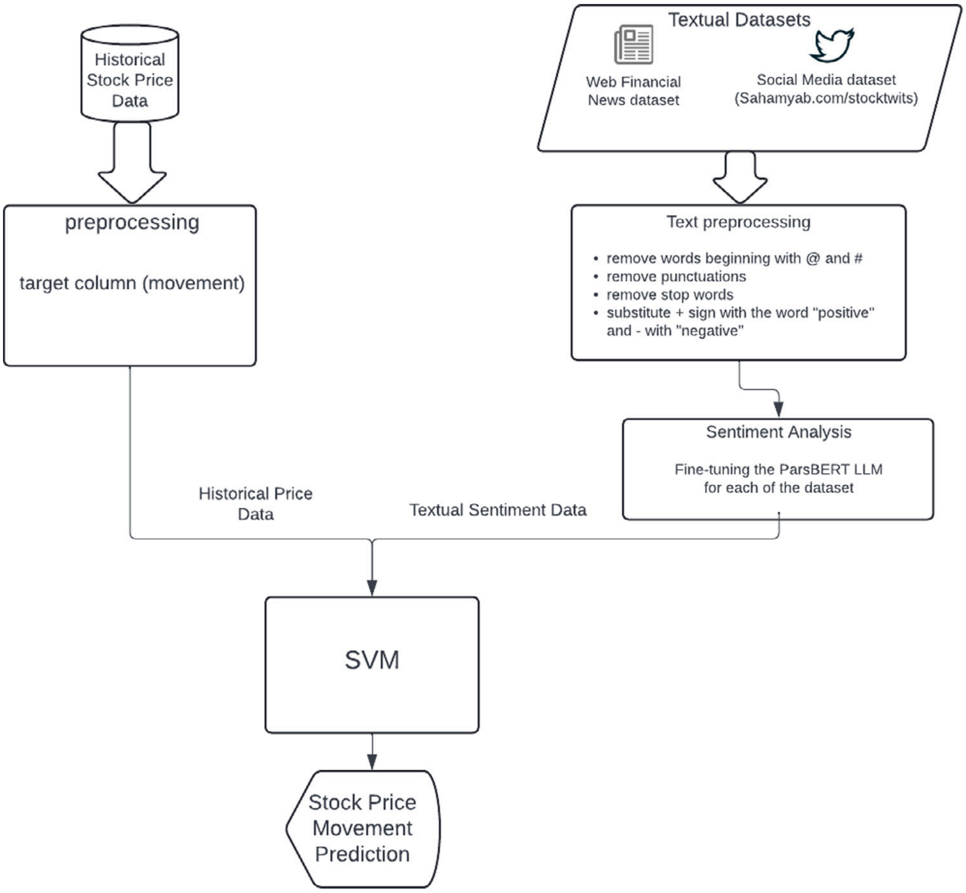

Fig. 1. Architectural framework of the stock price movement prediction system. The workflow illustrates the transition from raw data collection (historical prices, web financial news, and social media) through text preprocessing and sentiment extraction using fine-tuned ParsBERT models, predicting in binary trend classification using a Support Vector Machine (SVM) algorithm.

3. Methodology

As discussed in the previous section, the majority of prior research examined the effect of investor mood on the stock market of developed countries (Nti et al., 2020b). However, sentiment analysis models, especially for one country's social media posts, might not be applicable to other countries due to differences in languages, expression, and sentiment signals (Dang et al., 2020). As a result, it might seem hard to understand or judge people's feelings from a global viewpoint because numerous factors significantly influence feelings. These mostly include differences in national beliefs and behaviors, laws and rules that are different in each country, and the level of development of the stock market, among other things. Consequently, sentiment analysis might appear more straightforward with countries or regions with a common culture or familiarity than on a worldwide scale.

This research evaluates the ability of investor sentiment to predict stock price movements in the Tehran Stock Exchange (TSE). In contrast to earlier studies on TSE, this research employs a cutting-edge sentiment analysis method for getting investor sentiment from news and social media and evaluates their predictive efficacy in forecasting stock price movements.

This section presents the methodological framework of the study. It begins with a description of the dataset, including data collection, preprocessing, and feature extraction. An explanation of the sentiment analysis approach follows and provides details of the fine-tuning process of a language model for sentiment classification. The predictive modeling approach is then presented, together with the evaluation criteria applied to evaluate its performance.

Fig. 1 presents the framework of this research. The framework consists of four main steps: data collection, data preprocessing, sentiment analysis, and machine learning models. These steps are explained in detail in the following subsections.

3.1. Dataset description

The data used in this study are composed of three datasets, two textual datasets (D_{social}, D_{news}) and one historical price dataset (D_{price}). This research concentrates on the five of the largest firms, determined by their market capitalization, and most actively

Table 1
Selected securities from the Tehran Stock Exchange (TSE) in this study. The sample includes five major firms (KHODRO, FOOLAD, FARS, SHASTA, FAMELI) representing the automotive, steel, petrochemical, and investment sectors, alongside the TSE Total Index (TEDPIX).

Tickers	Company name
KHODRO	Iran Khodro Automobile
FOOLAD	Mobarakeh Steel Company
FARS	Persian Gulf Petrochemical Industries
SHASTA	The Social Security Investment Company
FAMELI	National Iranian Copper Industries Co.

traded (determined by their trading volume) on the TSE. The primary rationale for selecting these stocks is that the TSE is not a very active market, and using social media and financial news for predicting how stock prices will move requires stocks with high trading liquidity and significant media coverage on news websites and social media platforms. Consequently, five stocks that ensure substantial transactions and sufficient financial textual data were identified to extract investor sentiment (Gálvez & Gravano, 2017). Table 1 presents the names and tickers of selected stock companies. The TSE total index is included in the analysis to evaluate how investor sentiment influences the entire market.

These five companies were chosen because they represent industries most impacted by Iran's particular economic circumstances. They were also chosen based on their market capitalization and news coverage. Companies like KHODRO and FOOLAD, for example, are highly sensitive to international sanctions and government pricing policies. This sensitivity guarantees sentiment-driven data in the financial news and social media conversation around them.

The period selected for assessing the influence of investor sentiment on stock price prediction in this research is from November 1, 2020, to July 31, 2022. This period was chosen for two reasons. First, although historical prices, social media platforms for the stock market, and financial news websites have been accessible since around 2010, Internet access was limited to personal computers, preventing investors from expressing their opinions at any time using their mobile devices. The reliability of a sentiment analysis model is contingent upon the quality of the textual data.

Second, as mentioned in the introduction 1, the Iranian government wields considerable power over the Tehran Stock Exchange (TSE), holding 35% of its companies directly and another 40% through pension funds and investment corporations. Consequently, the revenue derived from the stocks of these enterprises constitutes a significant source of income for the government. Following the imposition of U.S. sanctions and its withdrawal from the 2015 Iran nuclear agreement, the Iranian government encountered a substantial budget deficit, exemplified by a 38% shortfall from March 2018 to March 2019 (Moradi et al., 2021). Consequently, the Iranian government opted to address the deficit by divesting its stocks in the Tehran Stock Exchange.

Since 2017, with substantial government backing and television advertising, the number of active TSE traders surged by 400% in 2019 compared to 2018 (reached 10 million market participants). Most newcomers to TSE were unfamiliar with the functioning of stock markets and fundamental facts regarding them. Furthermore, during these years, the TSE total index experienced an extraordinary gain; for instance, between March 2018 and March 2019, the TSE total index surged by 187%. Consequently, the substantial governmental assistance and the involvement of traders lacking fundamental knowledge of the stock market resulted in abnormal market conduct from 2017 to 2020. The operation ended in one of the most significant collapses in the history of the TSE in May 2020. Accordingly, the period following this anomalous market behavior has been included in the analysis.

The process of data gathering and preprocessing for each of the dataset in the period of analysis is discussed in the following subsections.

3.1.1. Web financial news

For the financial news dataset (D_{news}), data were scraped and collected from four leading news websites in Iran, *Boursepress*, *Bourse24*, *Eghtesad News*, and *Boursenews*, spanning the period from November 1, 2020, to July 31, 2022. Three of the identified websites for scraping financial news are exclusively focused on TSE related news, whereas the *Eghtesad News* website covers a wider array of financial and economic news. However, only its TSE news part has been scraped. Following the web scraping process, only three dataset columns have been chosen for analysis: "Date", "Content", and "Tags". The "Content" column does not include the title (headline) of the news and contains the main body of news stories. To filter news related to the stocks examined in this research, only news that contains the stocks' ticker in either "tags" or "content" columns were selected. Table 2 presents statistics pertaining to this dataset for each ticker.

3.1.2. Social media dataset

The social media dataset (D_{social}) contains user comments from the website , the Iranian counterpart of . This website serves as the primary social media platform for TSE in Iran. Users express their views on stocks and the market on this website, employing the # symbol followed by the stock symbols (e.g. #KHODRO in Persian) in their messages to clearly indicate the stocks under discussion. Users can categorize their messages on this platform with "optimistic" or "pessimistic" labels. The optimistic label indicates the user's expectation of an increase in the stock's price, while the pessimistic label signifies an opposite expectation. The original dataset contains several columns; however, only three are used in this research: "Date", "Content", and "Label". The D_{social} encompasses the same timeframe as the D_{news} to provide a comparative analysis of their performance throughout the same interval. The comments regarding the five stocks are selected by utilizing the # sign together with their respective tickers. Table 3 presents statistics pertaining to the social media dataset for each ticker.

Table 2
Descriptive statistics for the web financial news dataset (Nov 2020–July 2022). This table summarizes the daily distribution (Minimum, Median, Mean, and Maximum) of news articles per transaction date and the total volume of articles processed for each ticker to generate sentiment features.

Tickers	Company name	The number of news				Total news
		Min	Median	Mean	Max	
KHODRO	Iran Khodro Automobile	1	7	7	23	2711
FOOLAD	Mobarakeh Steel Company	0	8	8	21	3260
FARS	Persian Gulf Petrochemical Industries	0	3	4	12	1471
SHASTA	The Social Security Investment Company	0	1	2	51	638
FAMELI	National Iranian Copper Industries Co.	0	1	1	5	334
TEDPIX	The TSE Total Index	22	86	97	303	40,163

Table 3
Descriptive statistics for the social media (Sahamyab) dataset (Nov 2020–July 2022). This table presents the volume of user comments per trading day and the total count of messages used to calculate the daily average sentiment for individual stocks and the broader market index.

Tickers	Company name	The number of messages				Total messages
		Min	Median	Mean	Max	
KHODRO	Iran Khodro Automobile	0	72	85	497	33,214
FOOLAD	Mobarakeh Steel Company	0	19	26	211	10,233
FARS	Persian Gulf Petrochemical Industries	0	3	5	58	1757
SHASTA	The Social Security Investment Company	0	48	71	748	24,634
FAMELI	Persian Gulf Petrochemical Industries	0	18	22	227	8676
TEDPIX	The TSE Total Index	0	120	127	513	52,672

Table 4
Sample of raw historical price data for ticker KHODRO. Values are in Iranian Rial (IRR). Columns include daily Open, High, Low, Close, Adjusted Close, and Volume.

Date	Open	High	Low	Close	Adj Close	Volume
2022–07–25	1883	1912	1883	1883	1885	275289570
2022–07–26	1900	1920	1870	1889	1889	176883538
2022–07–27	1900	1920	1880	1901	1903	81490702
2022–07–30	1918	1922	1811	1823	1843	154575743
2022–07–31	1823	1839	1751	1773	1770	240516933

3.1.3. Historical price dataset

The *finpy-tse* Python module is employed to import the historical stock price data on TSE, (D_{price}), for five stocks and the TSE total index. The D_{price} is utilized to predict the stock price behavior, whether it will increase or decrease, over the next n days, based on factors such as the sentiment of financial news on the web or social media. Table 4 illustrates a sample of D_{price} dataset for the ticker KHODRO imported using *finpy-tse* Python module.

Three widely used technical indicators identified in the literature, Moving Average (MA), Relative Strength Index (RSI), and Moving average convergence/divergence (MACD), are incorporated into the stock price movement prediction model to evaluate their efficacy in comparison to investor sentiment and to see whether their accuracy may be improved by incorporating investor sentiment. The technical indicators are calculated as follows:

14-day Moving Average (MA) is calculated as the mean of the closing prices during the preceding 14 days:

$$MA_{14} = \frac{1}{14}\sum_{i=0}^{13} P_{t-i}$$

where P_t is the closing price at time t.

14-day Relative Strength Index (RSI) is calculated as follows:

1. Calculate the average gains and losses over the past 14 days:

$$\text{Average Gain}_{14} = \frac{1}{14}\sum_{i=0}^{13} G_{t-i}$$

$$\text{Average Loss}_{14} = \frac{1}{14}\sum_{i=0}^{13} L_{t-i}$$

where G_{t-i} and L_{t-i} are the gains and losses on day $t-i$, respectively.

2. Compute the Relative Strength (RS):

$$RS_{14} = \frac{\text{Average Gain}_{14}}{\text{Average Loss}_{14}}$$

3. Calculate the RSI:

$$RSI_{14} = 100 - \frac{100}{1 + RS_{14}}$$

Moving Average Convergence Divergence (MACD) is computed utilizing two exponential moving averages (EMAs):

1. Calculate the 12-day EMA and 26-day EMA:

$$EMA_{12} = P_t \cdot \frac{2}{13} + EMA_{12(t-1)} \cdot \left(1 - \frac{2}{13}\right)$$

$$EMA_{26} = P_t \cdot \frac{2}{27} + EMA_{26(t-1)} \cdot \left(1 - \frac{2}{27}\right)$$

2. Compute the MACD line:

$$MACD = EMA_{12} - EMA_{26}$$

Consistent with numerous studies in the literature (Batra & Daudpota, 2018; Nti et al., 2020a), stock returns are employed to determine whether stock prices will increase or decrease over the subsequent n days. For instance, if $n = 7$ and the specified date is t, the stock return is computed for the date t and $t + 7$. In this research, a negative stock return is denoted as 0; a positive stock return shows an increase in stock price, expressed as 1. The stock return for stock S at time t for n days in the future is calculated as follows:

$$\text{Stock Return}(S, t) = \left(\frac{\text{Closing Price}_{t+n}}{\text{Closing Price}_t}\right) - 1$$

The model forecasts how stock prices will move across 1-day (Daily), 7-day (Weekly), 30-day (Monthly), and 90-day (Quarterly) periods to evaluate the impact of public (investor) sentiment over multiple timeframes.

3.2. Sentiment analysis

Numerous techniques in NLP are utilized for sentiment analysis, including lexicon-based models (e.g. TextBlob (Nemes & Kiss, 2021)), as well as machine learning and deep learning methodologies, such as pre-trained large language models (LLMs). LLMs are often more accurate because of their training on millions of texts and pages (Howard & Ruder, 2018; Kotelnikova et al., 2021). This research includes the fine-tuning of a large language model (LLM) to determine the emotions of web financial news and social media in the Persian language.

Fine-tuning LLMs involves altering a pre-trained model, such as GPT or BERT, to enhance its performance on a particular final job by doing further training on a smaller, task-specific dataset. LLMs are initially trained on an extensive, varied corpus to acquire a comprehensive understanding of language, incorporating grammar, context, and semantics. During fine-tuning, the model sees data that more accurately reflects the intended application, such as sentiment analysis, legal document comprehension, or domain-specific question-and-answer tasks, enabling it to acquire nuances and specialized knowledge relevant to that function. Fine-tuning modifies the model's weights slightly, enhancing its performance for the specific application while preserving the general information acquired during the first training. This method is efficient as it utilizes the considerable linguistic expertise of the foundational LLM, enhancing its capability to meet particular requirements while minimizing the necessity for substantial data and computational resources (Minaee et al., 2024).

A ParsBERT LLM, a monolingual language model based on Google's BERT architecture, is fine-tuned on web financial news and social media messages about stocks to extract the sentiment of the Persian textual data about the stock market. ParsBERT LLM, as introduced by Farahani et al. (2021), is pre-trained on an extensive Persian corpus comprising over 3.9 million documents, 73 million sentences, and 1.3 billion words. In this research, separate fine-tuning of the ParsBERT is performed on the D_{social} and D_{news} datasets.

The following two subsections provide a detailed description of the fine-tuning process of the ParsBERT LLM for each textual dataset: web financial news and social media.

3.2.1. Fine-tuning ParsBERT for web financial news dataset

To prepare the dataset required for fine-tuning the ParsBERT LLM for web financial news, 6757 articles were randomly picked from the scraped financial news and subsequently annotated by two Iranian master's students specializing in finance. The annotation process involved manually labeling each news article based on its sentiment, positive, negative, or neutral, towards the stock market. The students were provided with clear annotation guidelines to ensure consistency and minimize subjective biases. The annotated (labeled) dataset was imbalanced; therefore, a balanced subset of appropriate size was extracted. Table 5 presents the comparison of original imbalanced and balanced datasets.

The subsequent preparation processes have been executed on the balanced dataset to prepare it for the fine-tuning process:

Table 5
Sentiment label distribution for the ParsBERT training set. This table details the counts of optimistic, pessimistic, and neutral labels in the news dataset. The "Balanced" column represents the subset used to fine-tune the LLM to ensure impartial classification across all sentiment categories.

Label	Dataset	
	Unbalanced dataset	Balanced dataset
Optimistic	3655	849
Pessimistic	849	849
Neutral	2253	849
Total	**6757**	**2457**

Table 6
Performance metrics and hyperparameters for the fine-tuned ParsBERT sentiment model for Web Financial News.

Dataset	Num of epochs	Learning rate	Batch size	Precision	Recall	Accuracy	F1 score
Financial News	5	2e−5	16	0.673	0.676	0.68	0.67

1. Removal of URLs and HTML tags;
2. Removal of special characters;
3. Removal of punctuations;
4. Removal of Persian stop-words;
5. Removal of English characters (if any);
6. Lemmatizing and tokenization of words;

ParsBERT LLM was fine-tuned for sentiment analysis of web financial news using the labeled dataset. The fine-tuning process has been conducted with several hyperparameters, using Hugging Face's transformers library and implemented in PyTorch. The dataset was split into 80%, 10%, and 10% for training, validation, and testing. The training was performed in a GPU-accelerated environment. The AdamW optimizer with a weight decay of 0.01 was used to prevent overfitting, and cross-entropy loss was employed as the objective function.

Multiple hyperparameter configurations were evaluated during fine-tuning to identify the best configuration. The number of epochs ranged from 3 to 5, with shorter training cycles (3 epochs) resulting in underfitting. Likewise, several learning rates were examined, such as 2e-3 and 2e-1. The optimal performance was achieved with a learning rate of 2e-5 since higher values resulted in considerable volatility in loss values. The batch size was varied at 8, 16, and 32. A batch size of 16 was ultimately selected as the optimal convergence between stability and processing efficiency.

A linear learning rate scheduler with warm-up steps was applied during fine-tuning to improve convergence. Moreover, gradient clipping was employed to prevent exploding gradients which ensures solid updates to model parameters. The model was trained for five epochs, with early stopping implemented based on validation loss and F1-score to mitigate overfitting. Table 6 displays the best result obtained.

The fine-tuned ParsBERT model for web financial news is employed to determine the sentiment of unlabeled news articles.

3.2.2. Fine-tuning ParsBERT for social media

For D_{social}, some of the messages have already been labeled by users as "optimistic" and "pessimistic". The label *optimistic* indicates that users anticipate an increase in stock prices, while *pessimistic* means the contrary. The labeled messages are used to fine-tune the ParsBERT LLM for the classification of unlabeled messages as either optimistic or pessimistic. The labeled dataset for fine-tuning is composed of 9318 bullish and 9318 bearish messages. The subsequent preprocessing steps have been undertaken to prepare the messages:

1. Remove URLs and HTML tags;
2. Words beginning with @ and # have been removed;
3. Remove punctuations;
4. a single or sequence of the sign '+' and the sign '-' have been substituted with the word "positive" and "negative" respectively;
5. Persian stop-words have been eliminated;
6. Lemmatizing and tokenization of words;

The preprocessing steps have been done using *Hazm* library in Python which is a version of NLTK (Natural Language Toolkit) for the Persian (Farsi) language.

ParsBERT LLM was fine-tuned for the sentiment analysis using preprocessed social media dataset (). A series of experiments with various hyperparameter configurations were conducted to identify the optimal settings for training stability and performance. The number of epochs varied between 3 and 5, and learning rates were tested, including 2e-5, 2e-3, and 1e-5. The batch size was

Table 7
Training hyperparameters and performance metrics of the fine-tuned ParsBERT model for sentiment classification on the social media dataset.

Dataset	Num of epochs	Learning rate	Batch size	Precision	Recall	Accuracy	F1 score
Social Media	3	1e−5	16	0.78	0.775	0.78	0.776

experimented with at 8, 16, and 32. The final model that trained with 3 epochs, a learning rate of 1e-5, and a batch size of 16 demonstrated the best performance. Table 7 shows the best result achieved.

The unlabeled messages inside the designated timeframe for the research are labeled with the fine-tuned ParsBERT LLM model.

3.3. Prediction model

The Support Vector Machine (SVM) algorithm is chosen for stock price trend forecast because of its frequent application across numerous studies (Derakhshan & Beigy, 2019; Pimprikar et al., 2017). Based on Nti et al. (2020b), SVM is among the top three most used algorithms for stock price movement forecast. SVM is particularly well-suited for financial classification problems because of its ability to handle high-dimensional data, mitigate overfitting through margin optimization, and perform well with relatively small datasets, which is often the case in financial studies.

SVM is a supervised machine learning algorithm that seeks to find an optimal hyperplane that best separates different classes in a given feature space. Stock price movement forecasting is framed as a binary classification problem, aiming at predicting whether the stock price will increase or decrease.

Different feature configurations, including technical indicators and investor sentiment derived from social media and financial news websites, are employed to forecast movements at daily, weekly, monthly, and quarterly intervals. The following subsections explain the methodology of extracting investor sentiment from social media and web financial news.

3.3.1. Social media features

A fine-tuned ParsBERT LLM is employed to label social media messages as optimistic or pessimistic, putting them in the "Social Media Sentiment" column of D_{social}. Messages which are labeled as optimistic are marked with the label 1, whereas those considered pessimistic are assigned the label 0. For each date t and the stocks, the average sentiment of all messages is computed to reflect the investors' sentiment for the stocks on that day, derived from social media messages, and is represented by the column "Daily Social Media Sentiment". The trading decisions made by investors on day t may be influenced by the events or happenings from previous days; thus, the three lags of "Daily Social Media Sentiment" are incorporated as features in the stock price movement prediction model, represented as "Lagged Social Media Sentiment" for days $t-1$, $t-2$, and $t-3$. Consequently, three specified features will be employed to forecast stock price fluctuations utilizing public sentiment derived by social media communications.

3.3.2. Web financial news features

After labeling the web financial news as optimistic, pessimistic or neutral using fine-tuned ParsBERT LLM as our "News Sentiment" column of D_{news}, the corresponding values are transformed to −1, 0, and 1 for pessimistic, neutral, and optimistic, respectively. To calculate the investors' sentiment from web financial news concerning stocks on each date, the mean sentiment of all news for each date is computed and represented by the column "Daily News Sentiment". Similar to social media features, the three lags of "Daily News Sentiment" are added to the stock price movement prediction model using investor sentiment derived from web financial news dataset. The three lagged sentiment indicators, denoted as "Lagged News Sentiment" for days $t-1$, $t-2$, and $t-3$.

3.4. Evaluation metrics

The forecasting of stock price fluctuations comprises a matter of binary classification, as it encompasses only two categories: upward or downward. The primary evaluation criteria for these problems are accuracy, recall, precision, and F1-score (Balasubramanian et al., 2024). According to the literature, these four evaluation metrics are among the most frequently utilized in the domain of stock price forecast (Ashtiani & Raahemi, 2023).

In classification problems, model performance is assessed using the following fundamental metrics:

- True Positive (TP): The model correctly predicts an upward movement when the stock price actually increases.
- True Negative (TN): The model correctly predicts a downward movement when the stock price actually decreases.
- False Positive (FP): The model incorrectly predicts an upward movement when the stock price actually decreases.
- False Negative (FN): The model incorrectly predicts a downward movement when the stock price actually increases.

The evaluation metrics are computed as follows:

Accuracy

$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN}$$

Table 8
Stock price movement prediction results using only Web Financial News sentiment. The SVM model classifies price movements (1 for rise, 0 for fall) across four horizons: Daily, Weekly, Monthly, and Quarterly. Metrics are derived from five-fold cross-validation.

Tickers	Timeframe	Precision	Recall	Accuracy	F1-score
Khodro	Daily	0.4778	0.4585	0.4585	0.4497
	Weekly	0.5223	0.4893	0.4893	0.4899
	Monthly	0.5608	0.5079	0.5079	0.4943
	Quarterly	0.5987	0.5207	0.5207	0.5265
Foolad	Daily	0.4437	0.4481	0.4481	0.4405
	Weekly	0.4906	0.4911	0.4911	0.4869
	Monthly	0.5456	0.5392	0.5392	0.5257
	Quarterly	0.5690	0.5544	0.5544	0.5496
Fars	Daily	0.5461	0.5438	0.5438	0.5280
	Weekly	0.5260	0.5358	0.5358	0.5247
	Monthly	0.4599	0.4566	0.4566	0.4513
	Quarterly	0.5198	0.4908	0.4908	0.4776
Shasta	Daily	0.5401	0.5441	0.5441	0.5347
	Weekly	0.5304	0.5118	0.5118	0.5053
	Monthly	0.4908	0.4298	0.4298	0.4422
	Quarterly	0.7066	0.5318	0.5318	0.5647
Fameli	Daily	0.5103	0.5077	0.5077	0.4921
	Weekly	0.5849	0.5772	0.5772	0.5496
	Monthly	0.4910	0.4975	0.4975	0.4767
	Quarterly	0.5229	0.5177	0.5177	0.4955
TEDPIX	Daily	0.4704	0.4805	0.4805	0.4646
	Weekly	0.5750	0.5707	0.5707	0.5532
	Monthly	0.4879	0.5024	0.5024	0.4851
	Quarterly	0.5358	0.5390	0.5390	0.5337

Recall

$$\text{Recall} = \frac{TP}{TP + FN}$$

Precision

$$\text{Precision} = \frac{TP}{TP + FP}$$

F1-score

$$\text{F1-Score} = \frac{2 \cdot \text{Precision} \cdot \text{Recall}}{\text{Precision} + \text{Recall}}$$

4. Results

This section provides an overview of the paper's findings. The SVM prediction model was employed using cross-validation with five splits on the specified five stocks and the TSE total index, incorporating three sets of features as previously described in the previous section. The importance and usefulness of using the cross-validation technique for machine learning models is supported in the literature (Ogundunmade et al., 2022).

4.1. Results of using public sentiment only

This section demonstrates the predictive efficacy of the SVM model utilizing only investor sentiment obtained through web news and social media datasets across various time horizons. This section evaluates the predictive ability of the two primary channels of investor sentiment.

Table 9
Stock price movement prediction results using only Social Media sentiment. The SVM model classifies price movements (1 for rise, 0 for fall) across four horizons: Daily, Weekly, Monthly, and Quarterly. Metrics are derived from five-fold cross-validation.

Tickers	Timeframe	Precision	Recall	Accuracy	F1-score
Khodro	Daily	0.4359	0.4220	0.4220	0.4175
	Weekly	0.4897	0.4871	0.4871	0.4793
	Monthly	0.4652	0.4587	0.4587	0.4527
	Quarterly	0.6840	0.6613	0.6613	0.6682
Foolad	Daily	0.4928	0.4962	0.4962	0.4857
	Weekly	0.5627	0.5570	0.5570	0.5470
	Monthly	0.5602	0.5595	0.5595	0.5570
	Quarterly	0.4769	0.4532	0.4532	0.4544
Fars	Daily	0.4639	0.4670	0.4670	0.4599
	Weekly	0.4284	0.4405	0.4405	0.4267
	Monthly	0.5237	0.5122	0.5122	0.4948
	Quarterly	0.5897	0.5492	0.5492	0.5537
Shasta	Daily	0.4543	0.4622	0.4622	0.4453
	Weekly	0.4896	0.4971	0.4971	0.4851
	Monthly	0.6602	0.6465	0.6465	0.6426
	Quarterly	0.6191	0.5064	0.5064	0.5416
Fameli	Daily	0.5106	0.5055	0.5055	0.4887
	Weekly	0.5957	0.5954	0.5954	0.5756
	Monthly	0.4669	0.4615	0.4615	0.4623
	Quarterly	0.4609	0.4615	0.4615	0.4523
TEDPIX	Daily	0.4121	0.4244	0.4244	0.4020
	Weekly	0.5665	0.5659	0.5659	0.5630
	Monthly	0.5684	0.5659	0.5659	0.5567
	Quarterly	0.4473	0.4488	0.4488	0.4409

4.1.1. *Results of using web financial news features*

First, web financial news features are employed to extract public sentiments for incorporation into the stock price movement prediction model. The model's predictive performance on the web financial news dataset is illustrated in Table 8.

Prediction accuracy tended to improve as the time horizon extended, as shown in Table 8. This implies that using investor sentiment extracted from web financial news may be more effective in capturing long-term trends. Khodro's accuracy increases from 45.85% in the 1-day period to 52.07% in the 90-day period, which is consistent with the trend observed in other securities, including Foolad and Shasta.

The accuracy statistic for the TSE Index exhibits a significant upward trend with extended timeframes which indicates an enhanced predictive performance over a longer duration. The accuracy begins at 48.05% for the 1-day interval, suggesting that the sentiment data is slightly superior to random guessing in forecasting short-term price fluctuations. Nonetheless, it markedly increases to 57.07% over the 7-day interval, indicating that investor sentiment becomes more pronounced in reflecting weekly patterns.

4.1.2. *Results of using social media features*

Table 9 summarizes the prediction performance of using social media features to derive investor sentiments.

The accuracy metric in Table 9 indicated that social media public sentiment predicts market behavior well for medium-term periods (7-day to 30-day). Investor sentiment expressed on social media may have a delayed but significant impact on stock prices, as illustrated by the fact that Fars and Shasta stocks exhibit their highest accuracy within the 30-day period. However, certain inconsistencies, such as the significant decrease in Foolad's accuracy during the 90-day period, suggest that social media sentiment may not be able to maintain its predictive power over extended periods.

The accuracy of the TSE Index is initially low at 42.44% for the 1-day period, but it substantially improves to 56.59% in the 7-day and 30-day periods, a significant improvement. Nevertheless, the 90-day period accuracy experiences a significant decline to 44.88%, in contrast to the financial news results. This suggests that social media sentiment is effective for short- to medium-term predictions (up to 30 days), but its utility decreases considerably over extended periods, as observed in Table 9. This trend can be related to the characteristics of social media sentiment, which primarily reflects immediate market reaction to news, speculation, and investor emotions. Over longer durations, fundamental financial elements, such as earnings reports, macroeconomic data, and institutional trading strategies, predominate, thereby diminishing the impact of sentiment-driven fluctuations.

The accuracy of utilizing investor sentiment to predict stock price movements varies depending on the source used to derive the investor sentiment, web financial news or social media, and the timeframe. Public sentiment showed that it can be beneficial, in most cases, as models can achieve accuracy exceeding 50%, which suggests its predictive capabilities. The Efficient Market Hypothesis (EMH), particularly the weak form, posits that prediction models based on public information can achieve a maximum of approximately 50% accuracy, as stock prices are anticipated to follow a standard random walk. Financial news-based public

Table 10
Aggregate prediction performance averaged across all sample securities. These summaries provide the mean Precision, Recall, Accuracy, and F1-Score for the entire study, highlighting the predictive value of different data sources across short- and long-term horizons.

Source	Timeframe	Precision	Recall	Accuracy	F1-score
Social Media	Daily	0.4551	0.4550	0.4550	0.4481
	Weekly	0.5116	0.5115	0.5115	0.5071
	Monthly	0.5441	0.5427	0.5427	0.5372
	Quarterly	0.5784	0.5762	0.5762	0.5734
	Average	0.5223	0.5213	0.5213	0.5164
Financial News	Daily	0.4906	0.4906	0.4906	0.4802
	Weekly	0.5316	0.5316	0.5316	0.5225
	Monthly	0.5455	0.5435	0.5435	0.5313
	Quarterly	0.5924	0.5908	0.5908	0.5873
	Average	**0.5400**	**0.5391**	**0.5391**	**0.5303**

Table 11
Performance of the prediction model (SVM) using Technical Indicators (TI) for various stocks and timeframes. Features include 14-day Moving Average (MA), Relative Strength Index (RSI), and MACD.

Tickers	Timeframe	Precision	Recall	Accuracy	F1-score
Khodro	Daily	0.3790	0.4564	0.4564	0.3935
	Weekly	0.4665	0.5051	0.5051	0.4616
	Monthly	0.7685	0.6769	0.6769	0.6387
	Quarterly	0.8530	0.7923	0.7923	0.7943
Foolad	Daily	0.5829	0.5363	0.5363	0.4910
	Weekly	0.4657	0.5164	0.5164	0.4789
	Monthly	0.7519	0.7280	0.7280	0.7231
	Quarterly	0.8082	0.7734	0.7734	0.7726
Fars	Daily	0.4011	0.4416	0.4416	0.3819
	Weekly	0.5953	0.5070	0.5070	0.4569
	Monthly	0.5804	0.5245	0.5245	0.4486
	Quarterly	0.5565	0.5026	0.5026	0.4735
Shasta	Daily	0.3168	0.4281	0.4281	0.3455
	Weekly	0.4319	0.4596	0.4596	0.3764
	Monthly	0.3166	0.4069	0.4069	0.3448
	Quarterly	0.9279	0.9050	0.9050	0.9084
Fameli	Daily	0.4567	0.5126	0.5126	0.4500
	Weekly	0.4788	0.4871	0.4871	0.4500
	Monthly	0.5840	0.5858	0.5858	0.5496
	Quarterly	0.5147	0.5099	0.5099	0.4854
TEDPIX	Daily	0.5485	0.4977	0.4977	0.4535
	Weekly	0.6895	0.6082	0.6082	0.5518
	Monthly	0.7335	0.6153	0.6153	0.5560
	Quarterly	0.8591	0.7910	0.7910	0.7641

sentiment is more effective at capturing broader market trends over extended timeframes, such as 90 days, which is why it tends to perform better. Conversely, social media sentiment is particularly effective in short- to medium-term periods, such as the 7-day and 30-day intervals, as it accurately represents the sentiment of short-term investors and instantaneous public reactions. Consequently, news is more appropriate for making long-term predictions, whereas social media is more effective at capturing short-term movements.

To provide a clearer comparison, Table 10 reports the overall performance of the stock price movement prediction model using social media- and web financial news-based investor sentiment by averaging across all stocks and the total index results. On average, investor sentiment extracted from web financial news performs better than that of social media.

4.2. Results of using public sentiment with technical indicators

This section aims to evaluate whether the incorporation of investor sentiment into a model that solely utilizes technical indicators enhances its performance.

Table 11 presents the predictive performance of the model using only technical indicators (TIs). The table indicates that the model utilizing TIs outperforms the model based on investor sentiment.

Table 12 illustrates the results of the model using both technical indicators and investor sentiment as features for forecasting stock price fluctuations.

The findings indicate that integrating Technical Indicators (TI) with investor sentiment metrics (such as web financial news or social media) typically improves stock prediction accuracy. In the case of Khodro stock, the incorporation of news enhances 1-day

Table 12
Prediction results for hybrid models combining Technical Indicators (TI) with sentiment data. This table compares how the addition of financial news or social media sentiment features affects the classification accuracy and F1-score for individual stocks and the TEDPIX index.

Tickers	Feature	Timeframe	Precision	Recall	Accuracy	F1-score
Khodro	TI + News	Daily	0.4054	0.4374	0.4374	0.3953
		Weekly	0.5033	0.5207	0.5207	0.4869
		Monthly	0.7609	0.7107	0.7107	0.6984
		Quarterly	0.8386	0.7861	0.7861	0.7893
	TI + Social Media	Daily	0.3543	0.4012	0.4012	0.3579
		Weekly	0.4323	0.4610	0.4610	0.4080
		Monthly	0.6968	0.6168	0.6168	0.5910
		Quarterly	0.8638	0.8146	0.8146	0.8171
Foolad	TI + News	Daily	0.5272	0.5291	0.5291	0.5210
		Weekly	0.5299	0.5266	0.5266	0.5166
		Monthly	0.7424	0.7114	0.7114	0.7014
		Quarterly	0.7967	0.7620	0.7620	0.7523
	TI + Social Media	Daily	0.5314	0.5266	0.5266	0.5160
		Weekly	0.4905	0.5367	0.5367	0.4983
		Monthly	0.7409	0.7215	0.7215	0.7160
		Quarterly	0.8209	0.7899	0.7899	0.7877
Fars	TI + News	Daily	0.4836	0.4826	0.4826	0.4757
		Weekly	0.5192	0.4510	0.4510	0.4096
		Monthly	0.4938	0.5248	0.5248	0.4518
		Quarterly	0.4724	0.4815	0.4815	0.4682
	TI + Social Media	Daily	0.4333	0.4379	0.4379	0.4321
		Weekly	0.4497	0.4406	0.4406	0.4190
		Monthly	0.5738	0.5674	0.5674	0.5135
		Quarterly	0.5452	0.5272	0.5272	0.5193
Shasta	TI + News	Daily	0.4740	0.4590	0.4590	0.4064
		Weekly	0.5537	0.5146	0.5146	0.4718
		Monthly	0.5417	0.4352	0.4352	0.3862
		Quarterly	0.7681	0.6982	0.6982	0.7179
	TI + Social Media	Daily	0.3335	0.4266	0.4266	0.3599
		Weekly	0.5984	0.5320	0.5320	0.4750
		Monthly	0.4666	0.4676	0.4676	0.4115
		Quarterly	0.7562	0.6457	0.6457	0.6762
Fameli	TI + News	Daily	0.5159	0.5126	0.5126	0.4840
		Weekly	0.4934	0.4817	0.4817	0.4743
		Monthly	0.5284	0.5199	0.5199	0.4994
		Quarterly	0.5222	0.5074	0.5074	0.4778
	TI + Social Media	Daily	0.5459	0.5207	0.5207	0.4879
		Weekly	0.5006	0.4921	0.4921	0.4833
		Monthly	0.5412	0.5201	0.5201	0.4870
		Quarterly	0.5133	0.4950	0.4950	0.4870
TEDPIX	TI + News	Daily	0.5402	0.5268	0.5268	0.5056
		Weekly	0.6766	0.6122	0.6122	0.5651
		Monthly	0.7235	0.6268	0.6268	0.5796
		Quarterly	0.8325	0.7610	0.7610	0.7302
	TI + Social Media	Daily	0.5585	0.5537	0.5537	0.5474
		Weekly	0.6513	0.6244	0.6244	0.5998
		Monthly	0.5938	0.6000	0.6000	0.5449
		Quarterly	0.8572	0.7927	0.7927	0.7672

accuracy from 45.64% (utilizing only TI) to 52.07%. However, optimal results are attained with both TI and tweets, reaching 86.18% over a 90-day span. A comparable pattern is noted for Foolad, wherein TI combined with tweets yields the best accuracy of 78.99% over 90 days, in contrast to 77.34% when employing solely TI. Fars and Shasta, however, exhibit less continuous advancements. For example, Fars's accuracy reaches a maximum of 54.92% with only TI during the 90-day period, with minimal enhancement with news (52.77%) and tweets (51.93%). Shasta demonstrates the best accuracy combining TI and social media at 76.81% over 90 days, suggesting that although social media investor sentiment enhances performance, its effect fluctuates based on the stock's sensitivity to sentiment effects.

The incorporation of sentiment elements markedly improves the performance of the TSE Index. The 1-day accuracy with simply TI is 49.77%, which improves to 61.22% when incorporating news, and further rises to 75.37% with the addition of social media

Table 13
Aggregate prediction performance averaged across all sample securities. These summaries provide the mean Precision, Recall, Accuracy, and F1-Score for the entire study, highlighting the predictive value of different data sources across short- and long-term horizons.

Source	Timeframe	Precision	Recall	Accuracy	F1-score
Technical Indicator (TI)	Daily	0.4353	0.4655	0.4622	0.4487
	Weekly	0.5145	0.5366	0.5335	0.5186
	Monthly	0.6642	0.6532	0.6524	0.6478
	Quarterly	0.7545	0.7711	0.7695	0.7639
	Average	0.5921	0.6066	0.6044	0.5948
TI + Social Media	Daily	0.4376	0.4671	0.4650	0.4508
	Weekly	0.5265	0.5446	0.5432	0.5319
	Monthly	0.7123	0.7028	0.7021	0.6962
	Quarterly	0.8095	0.8231	0.8205	0.8157
	Average	0.6215	0.6344	0.6327	0.6237
TI + Financial News	Daily	0.4402	0.4692	0.4673	0.4529
	Weekly	0.5184	0.5397	0.5385	0.5247
	Monthly	0.7011	0.6915	0.6910	0.6853
	Quarterly	0.7984	0.8103	0.8076	0.8031
	Average	0.6145	0.6277	0.6261	0.6165

public sentiment features. This trend persists in extended timeframes, with optimal outcomes observed in the 90-day interval, where accuracy increases from 79.10% (TI alone) to 83.25% (TI and financial news), ultimately achieving 85.72% when TI is combined with social media. This indicates that both financial news and social enhance the model's predictive capacity for the TSE Index; however, social media have a more significant influence, implying that real-time social media sentiment reflects relevant details more closely associated with the TSE index.

Overall, adding investor sentiment into technical indicators enhanced prediction accuracy for most stocks and the TSE Index, with social media public sentiment often showing a more significant influence than financial news articles. This pattern demonstrates that although technical indicators proficiently reflect historical price movements, public sentiment, particularly from social media, enhances them by providing real-time information. Consequently, in a developing nation, investor sentiment (derived from social media or web financial news) provides novel information that is not already reflected in historical price data.

To better illustrate the impact of adding investor sentiment to the model which relies only on TIs as features, Table 13 presents the overall performance. The results indicate that, on average, both social media and financial news improves the performance on average.

4.3. Discussion

The comparison of stocks' results using social media public sentiment (Table 9) against financial news public sentiment (Table 8) revealed varied outcomes contingent upon the specific stock and time horizon. For Khodro, the use of social media public sentiment yielded superior long-term accuracy (66.13% for 90 days) relative to financial news (52.07% for 90 days), suggesting that social media was more adept at identifying prolonged trends. Likewise, for Foolad, social media features achieved greater accuracy (55.95% for 30-day) than financial news (53.92%). For Shasta, the performance of using financial news sentiment surpassed that of social media throughout most timeframes, especially in the 90-day period, where accuracy is 53.18% for news compared to 50.64% for social media. This indicated that for certain stocks, financial news may be more important in reflecting long-term investor sentiment and price movements.

The TSE Index demonstrated higher performance when utilizing financial news public sentiment compared to social media across most timeframes. This suggests that the TSE Index was mainly influenced by conventional financial news sources rather than social media sentiment, especially regarding short-term and long-term forecasts. Consequently, financial news data served as a more reliable and effective sentiment indicator for predicting the total market index than social media.

Overall, financial news sentiment outperformed social media sentiment for numerous stocks, particularly over shorter timeframes such as 1-day and 7-day intervals, and in forecasting the TSE Index. For certain specific stocks, such as Khodro and Foolad, social media sentiment was more effective over longer periods, revealing extra insights that may not be as apparent in conventional news sources. This indicated that the ability to utilize social media or news sentiment is dependent upon the particular stock and the forecasting timeframe.

These differences in performance are likely rooted in the distinct mechanisms of information diffusion within the Iranian market. The short-term predictive power of financial news can be attributed to its role as an authoritative signal for state-led policy shifts. In a state-controlled economy where the government manages a significant portion of market capitalization, official news outlets act as the primary conduits for structured reports on sanctions and regulation. Professional investors react immediately to these official declarations, which lead to rapid market adjustments within a 1-day to 7-day window. This is especially apparent in the TSE Index (TEDPIX), which is more closely aligned with official narratives and economic disclosures than with informal social discourse.

Conversely, the dominance of social media over longer horizons is fundamentally linked to the institutional "Price Limit Domain" rule, which restricts daily fluctuations to $\pm 2\%$. This constraint explicitly delays the process of price discovery; since the market cannot instantly reflect the full magnitude of a sentiment shock in a single day, public discourse on platforms like Sahamyab creates a persistent narrative momentum that takes weeks to be fully incorporated into the stock price. Moreover, the significant growth of retail investors, which rose by 400% in 2019 and frequently possess an inadequate basic understanding of financial markets, introduces a psychological dimension to market activity that technical indicators alone cannot capture. These market participants frequently utilize specialized social platforms like Sahamyab to inform their trading decisions, making social data a potent leading indicator for individual stock volatility. The model produces more reliable forecasts across all time periods by combining sentiment with technical indicators to account for both mathematical historical trends and the behavioral irrationality of a retail-heavy base.

Based on the results, web financial news sources serve as a marginally superior predictor, providing greater accuracy across a wider array of conditions and for the index.

5. Conclusion and further work

Recent research and publications regarding stock price prediction models have been on the rise (Bustos & Pomares-Quimbaya, 2020). Diverse methodologies and features have been employed for stock market forecasting. In response to the rise in Internet usage, particularly on social media platforms, academics and traders attempt to improve the performance of stock market prediction models by integrating online information as features (Agarwal et al., 2019).

This study assessed the effect of investor sentiment, sourced from web financial news and social media messages, on stock price movement forecasting in the Tehran Stock Exchange (TSE) in Iran. This research utilizes two fine-tuned Persian Large Language Models (LLMs) based on the ParsBERT architecture to comprehensively analyze Persian-language financial texts for stock market forecasting. To the best of the author's knowledge, this is the first research to propose a fine-tuned LLM for analyzing financial texts in Persian (Farsi) language. The findings of this research addresses three important research questions.

The findings indicate that investor sentiment, especially from the two mentioned sources, can substantially improve prediction accuracy, particularly when integrated with technical indicators. This suggests that textual data possesses significant usefulness in explaining the psychological and behavioral aspects of trading decisions that are insufficiently represented by historical price data alone. The analysis of two sentiment sources revealed slight differences in their prediction power based on the timeframe range. In short-term forecasts, web financial news surpassed social media in anticipating stock price patterns, probably due to the structured format of news items and their focus on market-relevant information. Conversely, for medium- to long-term forecasts, social media sentiment showed enhanced consistency and precision. This may be attributed to social media's capacity to represent more types of genuine investor opinions and the expansive market narratives that develop over time, making it more adept at expressing long-term market sentiment.

Iran's unique political and economic environment, marked by substantial government intervention and a series of sanctions, profoundly influences investor behavior and the dynamics of the TSE. The data demonstrate that the TSE is highly responsive to fluctuations in public sentiment, as evidenced by both web financial news and social media platforms mirroring similar sensitivity. The findings suggest that sentiment analysis may offer enhanced forecasting capabilities in economies significantly influenced by external political and economic issues compared to more developed markets. This emphasizes the significance of context-specific models in employing sentiment analysis for financial forecasts in developing countries.

This research validated that investor sentiment can offer further insights beyond those present in historical price data via technical indicators. This indicates that sentiment and technical indicators contain complementary facets of stock price fluctuations. The research also emphasized that not all stocks react similarly to identical sentiment sources. For certain stocks, social media sentiment produced a more significant impact, whilst for others, web financial news proved to be more significant. This variability highlights the necessity of customizing sentiment-based stock price prediction models for particular securities and indicates that the investor base, stock liquidity, and market capitalization may affect the extent to which various sentiment sources influence price movements. This information is essential for investors formulating stock-specific trading techniques based on sentiment data.

This research demonstrates that sentiment analysis is a crucial instrument for comprehending market dynamics in developing countries, especially those characterized by minimal transparency and significant retail investor participation. Although the Iranian context is distinct due to its political and economic isolation, the success of the ParsBERT-SVM framework suggests that sentiment data captures a critical 'psychological dimension' of the market. This dimension is particularly crucial in emerging economies, because historical pricing data frequently does not capture the full effects of political shocks or artificial restraints, such as the TSE's daily price limits. By capturing these constrained market dynamics, our study offers a practical framework for applying sentiment analysis in other volatile, high-intervention economies.

5.1. Further work

There are some limitations in this research that needs further work. Future studies should investigate the integration of a wider range of sentiment sources, including investor forums, such as count and comments on social media messages and comments on news articles, to get a more holistic understanding of investor sentiment. Furthermore, incorporating geopolitical events, macroeconomic indicators, and policy changes into the research will improve the comprehension of their impact on sentiment and price movements in stocks. Furthermore, performing cross-market analyses across developed and developing countries would clarify the different roles of sentiment in different economic environments.

Table 14
Definition and sources of variables used in the empirical analysis. This table provides a reference for all features used in the SVM prediction models, including their origin and calculation basis.

Variable	Description	Data source
$S\_sentiment$	Daily average sentiment score derived from social media posts.	Sahamyab.com
$S\_sentiment_{t-n}$	Lags (1, 2, and 3 days) of the social media sentiment score.	Calculated
$N\_sentiment$	Daily average sentiment score derived from web financial news.	News scrapers
$N\_sentiment_{t-n}$	Lags (1, 2, and 3 days) of the web financial news sentiment score.	Calculated
MA_{14}	14-day Moving Average of the stock's adjusted closing prices.	Calculated
RSI_{14}	14-day Relative Strength Index used to measure price momentum.	Calculated
$MACD$	Moving Average Convergence Divergence for trend-following.	Calculated
$Target$	Binary price movement (1 for an increase, 0 for a decrease).	Calculated
$Close$	Daily adjusted closing price of the security in Iranian Rial (IRR).	Finpy-tse module

Funding

No funding was received for conducting this study.

Ethical approval

Not applicable.

Declaration of competing 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.

Appendix A

See Table 14.

Data availability

Data will be made available on request.

References

Agarwal, S., Kumar, S., & Goel, U. (2019). Stock market response to information diffusion through internet sources: A literature review. *International Journal of Information Management*, *45*, 118–131.

Antweiler, W., & Frank, M. Z. (2004). Is all that talk just noise? the information content of internet stock message boards. *The Journal of Finance*, *59*(3), 1259–1294.

Asgarian, E., Kahani, M., & Sharifi, S. (2018). Hesnegar: Persian sentiment wordnet. *Signal and Data Processing*, *15*(1), 71–86.

Ashtiani, M. N., & Raahemi, B. (2023). News-based intelligent prediction of financial markets using text mining and machine learning: A systematic literature review. *Expert Systems with Applications*, *217*, Article 119509.

Baker, M., & Wurgler, J. (2006). Investor sentiment and the cross-section of stock returns. *The Journal of Finance*, *61*(4), 1645–1680.

Balasubramanian, P., Chinthan, P., Badarudeen, S., & Sriraman, H. (2024). A systematic literature survey on recent trends in stock market prediction. *PeerJ Computer Science*, *10*, e1700.

Barber, B. M., & Odean, T. (2008). All that glitters: The effect of attention and news on the buying behavior of individual and institutional investors. *The Review of Financial Studies*, *21*(2), 785–818.

Batra, R., & Daudpota, S. M. (2018). Integrating stocktwits with sentiment analysis for better prediction of stock price movement. In *2018 international conference on computing, mathematics and engineering technologies (iCoMET)* (pp. 1–5). IEEE.

Boachie, M. K., Mensah, I., Frimpong, A., & Ruzima, M. (2016). Interest rate, liquidity and stock market performance in ghana. *International Journal of Accounting and Economics Studies*, *4*(1), 46–51.

Bollen, J., Mao, H., & Zeng, X. (2011). Twitter mood predicts the stock market. *Journal of Computational Science*, *2*(1), 1–8.

Brown, T., Mann, B., Ryder, N., Subbiah, M., Kaplan, J. D., Dhariwal, P., Neelakantan, A., Shyam, P., Sastry, G., Askell, A., et al. (2020). Language models are few-shot learners. *Advances in Neural Information Processing Systems*, *33*, 1877–1901.

Bustos, O., & Pomares-Quimbaya, A. (2020). Stock market movement forecast: A systematic review. *Expert Systems with Applications*, *156*, Article 113464.

Carosia, A. E. d. O., Coelho, G. P., & Silva, A. E. A. d (2020). Analyzing the brazilian financial market through portuguese sentiment analysis in social media. *Applied Artificial Intelligence*, *34*(1), 1–19.

Chen, L.-C., Lee, C.-M., & Chen, M.-Y. (2020). Exploration of social media for sentiment analysis using deep learning. *Soft Computing*, *24*(11), 8187–8197.

Coyne, S., Madiraju, P., & Coelho, J. (2017). Forecasting stock prices using social media analysis. In *2017 IEEE 15th intl conf on dependable, autonomic and secure computing, 15th intl conf on pervasive intelligence and computing, 3rd intl conf on big data intelligence and computing and cyber science and technology congress (DASC/piCom/dataCom/cyberSciTech)* (pp. 1031–1038). IEEE.

Dang, N. C., Moreno-García, M. N., & De la Prieta, F. (2020). Sentiment analysis based on deep learning: A comparative study. *Electronics*, *9*(3), 483.

Derakhshan, A., & Beigy, H. (2019). Sentiment analysis on stock social media for stock price movement prediction. *Engineering Applications of Artificial Intelligence*, *85*, 569–578.

Devlin, J., Chang, M.-W., Lee, K., & Toutanova, K. (2018). Bert: Pre-training of deep bidirectional transformers for language understanding. arXiv preprint arXiv:1810.04805.

Drus, Z., & Khalid, H. (2019). Sentiment analysis in social media and its application: Systematic literature review. *Procedia Computer Science*, *161*, 707–714.

Ebrahimian, K., Abbasi, E., Alam Tabriz, A., & Mohammadzadeh, A. (2023). Stock price forecasting with support vector regression based on social network sentiment analysis and technicl analysis. *International Journal of Finance & Managerial Accounting*, *8*(29), 53–64.

Eck, M., Germani, J., Sharma, N., Seitz, J., & Ramdasi, P. P. (2021). Prediction of stock market performance based on financial news articles and their classification. In *Data management, analytics and innovation: proceedings of ICDMAI 2020, volume 2* (pp. 35–44). Springer.

Farahani, M., Gharachorloo, M., Farahani, M., & Manthouri, M. (2021). Parsbert: Transformer-based model for persian language understanding. *Neural Processing Letters*, *53*, 3831–3847.

Fataliyev, K., Chivukula, A., Prasad, M., & Liu, W. (2021). Stock market analysis with text data: A review. arXiv preprint arXiv:2106.12985.

Feuerriegel, S., & Gordon, J. (2018). Long-term stock index forecasting based on text mining of regulatory disclosures. *Decision Support Systems*, *112*, 88–97.

Gálvez, R. H., & Gravano, A. (2017). Assessing the usefulness of online message board mining in automatic stock prediction systems. *Journal of Computational Science*, *19*, 43–56.

Ghaznavi, A., Aliyari, M., & Mohammadi, M. (2016). Predicting stock price changes of tehran artmis company using radial basis function neural networks. *Int Res J Appl Basic Sci*, *10*(8), 972–978.

Gunduz, H., & Cataltepe, Z. (2015). Borsa istanbul (bist) daily prediction using financial news and balanced feature selection. *Expert Systems with Applications*, *42*(22), 9001–9011.

Hasan, A. A., & Fong, A. C. (2018). Sentiment analysis based fuzzy decision platform for the saudi stock market. In *2018 IEEE international conference on electro/information technology* (pp. 0023–0029). IEEE.

Hirschberg, J., & Manning, C. D. (2015). Advances in natural language processing. *Science*, *349*(6245), 261–266.

Hong, S. (2020). A study on stock price prediction system based on text mining method using lstm and stock market news. *Journal of Digital Convergence*, *18*(7).

Howard, J., & Ruder, S. (2018). Universal language model fine-tuning for text classification. arXiv preprint arXiv:1801.06146.

Jishag, A., Athira, A., Shailaja, M., & Thara, S. (2020). Predicting the stock market behavior using historic data analysis and news sentiment analysis in r. In *First international conference on sustainable technologies for computational intelligence: proceedings of ICTSCI 2019* (pp. 717–728). Springer.

Kabbani, T., & Usta, F. E. (2022). Predicting the stock trend using news sentiment analysis and technical indicators in spark. arXiv preprint arXiv:2201.12283.

Khan, W., Ghazanfar, M. A., Azam, M. A., Karami, A., Alyoubi, K. H., & Alfakeeh, A. S. (2022). Stock market prediction using machine learning classifiers and social media, news. *Journal of Ambient Intelligence and Humanized Computing*, 1–24.

Khurana, D., Koli, A., Khatter, K., & Singh, S. (2023). Natural language processing: State of the art, current trends and challenges. *Multimedia Tools and Applications*, *82*(3), 3713–3744.

Kotelnikova, A., Paschenko, D., Bochenina, K., & Kotelnikov, E. (2021). Lexicon-based methods vs. bert for text sentiment analysis. In *International conference on analysis of images, social networks and texts* (pp. 71–83). Springer.

Lawal, Z. K., Yassin, H., & Zakari, R. Y. (2020). Stock market prediction using supervised machine learning techniques: An overview. In *2020 IEEE Asia-Pacific conference on computer science and data engineering* (pp. 1–6). IEEE.

Li, H. (2020). Related research on news sentiment tendency and stock price fluctuation. In *E3S web of conferences*: *vol. 214*, (p. 03001). EDP Sciences..

Li, B., Chan, K. C., Ou, C., & Ruifeng, S. (2017). Discovering public sentiment in social media for predicting stock movement of publicly listed companies. *Information Systems*, *69*, 81–92.

Liu, Q., Lee, W.-S., Huang, M., & Wu, Q. (2023). Synergy between stock prices and investor sentiment in social media. *Borsa Istanbul Review*, *23*(1), 76–92.

Mamun, A., Ali, M. H., Hoque, N., Mowla, M. M., & Basher, S. (2018). The causality between stock market development and economic growth: Econometric evidence from bangladesh. *International Journal of Economics and Finance*, *10*(5), 212.

Mate, G., Siddhant, A., Rutuja, K., & Maitreyi, M. (2019). Stock prediction through news sentiment analysis. *Journal of Architecture & Technology*, *11*(8), 36–40.

Mehta, P., Pandya, S., & Kotecha, K. (2021). Harvesting social media sentiment analysis to enhance stock market prediction using deep learning. *PeerJ Computer Science*, *7*, e476.

Metawa, N., Hassan, M. K., Metawa, S., & Safa, M. F. (2019). Impact of behavioral factors on investors' financial decisions: case of the egyptian stock market. *International Journal of Islamic and Middle Eastern Finance and Management*, *12*(1), 30–55.

Min, B., Ross, H., Sulem, E., Veyseh, A. P. B., Nguyen, T. H., Sainz, O., Agirre, E., Heintz, I., & Roth, D. (2023). Recent advances in natural language processing via large pre-trained language models: A survey. *ACM Computing Surveys*, *56*(2), 1–40.

Minaee, S., Mikolov, T., Nikzad, N., Chenaghlu, M., Socher, R., Amatriain, X., & Gao, J. (2024). Large language models: A survey. arXiv preprint arXiv:2402.06196.

Moradi, M., Appolloni, A., Zimon, G., Tarighi, H., & Kamali, M. (2021). Macroeconomic factors and stock price crash risk: do managers withhold bad news in the crisis-ridden iran market?. *Sustainability*, *13*(7), 3688.

Mu, G., Gao, N., Wang, Y., & Dai, L. (2023). A stock price prediction model based on investor sentiment and optimized deep learning. *IEEE Access*.

Nardo, M., Petracco-Giudici, M., & Naltsidis, M. (2016). Walking down wall street with a tablet: A survey of stock market predictions using the web. *Journal of Economic Surveys*, *30*(2), 356–369.

Nemes, L., & Kiss, A. (2021). Prediction of stock values changes using sentiment analysis of stock news headlines. *Journal of Information and Telecommunication*, *5*(3), 375–394.

Nti, I. K., Adekoya, A. F., & Weyori, B. A. (2020a). Predicting stock market price movement using sentiment analysis: Evidence from ghana. *Applied Computer Systems*, *25*(1), 33–42.

Nti, I. K., Adekoya, A. F., & Weyori, B. A. (2020b). A systematic review of fundamental and technical analysis of stock market predictions. *Artificial Intelligence Review*, *53*(4), 3007–3057.

Ogundunmade, T. P., Adepoju, A. A., & Allam, A. (2022). Stock price forecasting: Machine learning models with k-fold and repeated cross validation approaches. *Mod Econ Manag*, *1*.

Pagolu, V. S., Reddy, K. N., Panda, G., & Majhi, B. (2016). Sentiment analysis of twitter data for predicting stock market movements. In *2016 international conference on signal processing, communication, power and embedded system* (pp. 1345–1350). IEEE.

Picasso, A., Merello, S., Ma, Y., Oneto, L., & Cambria, E. (2019). Technical analysis and sentiment embeddings for market trend prediction. *Expert Systems with Applications*, *135*, 60–70.

Pimprikar, R., Ramachadran, S., & Senthilkumar, K. (2017). Use of machine learning algorithms and twitter sentiment analysis for stock market prediction. *International Journal of Pure and Applied Mathematics*, *115*(6), 521–526.

Ren, Y., Liao, F., & Gong, Y. (2020). Impact of news on the trend of stock price change: An analysis based on the deep bidirectiona lstm model. *Procedia Computer Science*, *174*, 128–140.

Rioja, F., & Valev, N. (2014). Stock markets, banks and the sources of economic growth in low and high income countries. *Journal of Economics and Finance*, *38*, 302–320.

Rodríguez-Ibánez, M., Casánez-Ventura, A., Castejón-Mateos, F., & Cuenca-Jiménez, P.-M. (2023). A review on sentiment analysis from social media platforms. *Expert Systems with Applications*, *223*, Article 119862.

Sagala, T. W., Saputri, M. S., Mahendra, R., & Budi, I. (2020). Stock price movement prediction using technical analysis and sentiment analysis. In *Proceedings of the 2020 2nd Asia Pacific information technology conference* (pp. 123–127).

Shah, D., Isah, H., & Zulkernine, F. (2018). Predicting the effects of news sentiments on the stock market. In *2018 IEEE international conference on big data (big data)* (pp. 4705–4708). IEEE.

Singh, N., & Jaiswal, U. C. (2023). Sentiment analysis using machine learning: A comparative study. *ADCAIJ: Advances in Distributed Computing and Artificial Intelligence Journal*, *12*, e26785.

Sun, A., Lachanski, M., & Fabozzi, F. J. (2016). Trade the tweet: Social media text mining and sparse matrix factorization for stock market prediction. *International Review of Financial Analysis*, *48*, 272–281.

Taboada, M., Brooke, J., Tofiloski, M., Voll, K., & Stede, M. (2011). Lexicon-based methods for sentiment analysis. *Computational Linguistics*, *37*(2), 267–307.

Tajmazinani, M., Hassani, H., Raei, R., & Rouhani, S. (2022). Modeling stock price movements prediction based on news sentiment analysis and deep learning. *Annals of Financial Economics*, *17*(01), Article 2250003.

Tohidi, M. (2022). Market sentiment and stock market volatility: Evidence from tehran stock exchange. *Interdisciplinary Journal of Management Studies*, *15*(4), 701–720.

Touvron, H., Lavril, T., Izacard, G., Martinet, X., Lachaux, M.-A., Lacroix, T., Rozière, B., Goyal, N., Hambro, E., Azhar, F., et al. (2023). Llama: open and efficient foundation language models. arXiv preprint arXiv:2302.13971.

Vijh, M., Chandola, D., Tikkiwal, V. A., & Kumar, A. (2020). Stock closing price prediction using machine learning techniques. *Procedia Computer Science*, *167*, 599–606.

Yadav, A., & Vishwakarma, D. K. (2020). Sentiment analysis using deep learning architectures: a review. *Artificial Intelligence Review*, *53*(6), 4335–4385.

Zaleskiewicz, T., & Traczyk, J. (2020). Emotions and financial decision making. *Psychological Perspectives on Financial Decision Making*, 107–133.

Zhang, C. (2008). *Defining, modeling, and measuring investor sentiment*.

Zhang, J., Cui, S., Xu, Y., Li, Q., & Li, T. (2018a). A novel data-driven stock price trend prediction system. *Expert Systems with Applications*, *97*, 60–69.

Zhang, L., Wang, S., & Liu, B. (2018b). *Deep Learning for Sentiment Analysis: A Survey. Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*, *8*(4), Article e1253.

Zhou, Z., Xu, K., & Zhao, J. (2018). *21*, *Tales of emotion and stock in china: volatility, causality and prediction* (pp. 1093–1116). World Wide Web.