2023 International Conference on Ambient Intelligence, Knowledge Informatics and Industrial Electronics (AIKIIE)

Empowering Investors: Insights from Sentiment Analysis, FFT, and Regression in Indian Stock Markets

1st Shashidhar R
Department of Electronics and Communication Engineering
JSS Science and Technology University
Mysuru, India
shashidhar.r@jssstuniv.in

2nd Aditya V
Department of Electronics and Communication Engineering
JSS Science and Technology University
Mysuru, India
vaditya06@gmail.com

3rd Srihari N
Department of Electronics and Communication Engineering
JSS Science and Technology University
Mysuru, India
sriharikashyap47@gmail.com

4th Subhash M H
Department of Electronics and Communication Engineering
JSS Science and Technology University
Mysuru, India
subhash.mh20@gmail.com

5th Krutthika Hirebasur Krishnappa
Department of Computer Science
Southern University and A&M College
Baton Rouge, Louisina, USA
krutthika.hirebas@sus.edu

***Abstract*—Sentiment Analysis has become a tool, for analyzing text data and extracting insights. In our research, we use sentiment examination techniques to calculate stock market trends in the market. We collected a dataset of news articles from two years focusing on financial news related to Indian stocks obtained through the Google News API. Using the NLTK VADER algorithm we calculate polarity scores for the news headlines to measure the sentiment associated with each article. To further improve our analysis, we introduce an approach by smearing Fast Fourier Transform and Inverse Fast Fourier Transform to analyze patterns or frequencies in the sentiment scores. This technique aims to provide features for enhancing stock market analysis. At the time we gather stock data using the Python y finance library. This includes indicators, like price changes, trading volumes, and other financial measurements. This data is essential for our regression models. We use these models to examine how the features derived from sentiment analysis and FFT components are related and how they correlate with stock price movements. By combining sentiment analysis with regression modeling the main goal of this research project is to provide insights into how news sentiment, FFT components, and stock market trends interact in the market. The results of this research endeavour could potentially equip investors with the knowledge they need to make decisions and develop strategies that contribute to participation, in the stock market.**

Keywords—Sentiment Analysis, Stock Market Prediction, Fast Fourier Transform, Regression Modeling, Indian Stock Market

I. INTRODUCTION

The Indian stock market, also known as the equity market acts as a platform where traded companies, in India raise capital by selling shares to potential investors. In India two exchanges represent this market segment; the Bombay Stock Exchange and the National Stock Exchange. These exchanges operate digitally allowing investors to trade stocks online through brokerage services. The Indian stock market plays a role in the economy as it provides companies with opportunities to raise capital and allows investors to participate in the growth of Indian businesses. This financial ecosystem is essential for facilitating access to capital and enabling investments thus contributing to progress and development.

Sentiment analysis is a branch of natural language processing that analyzes emotions expressed within written content. Concerning the stock market sentiment analysis proves valuable in evaluating prevailing sentiments found in news articles, social media posts, and other information sources that potentially impact market dynamics.

Sentiment analysis can be effectively utilized in ways, within the context of the stock market. For example, it can be used to determine whether news articles or social media posts about a stock show negative or neutral feeling. This analysis provides information for investors who want to make informed decisions about their stocks or trades. Additionally, sentiment analysis can be utilized to assess market trends. For instance, if sentiment is predominant in news articles or social media posts related to the stock market it could indicate that investors generally have a bearish outlook. Overall sentiment analysis proves beneficial for investors who want to stay informed about the stock market and make data-driven investment choices. By examining the aspects of news articles, social media posts, and other sources of information investors can gain insights, into market trends and make smarter investment decisions.

A. Overview

Google News and Yfinance are two resources, for obtaining stock-related data. Google News acts as a platform that gathers news articles from sources while Yfinance provides access to stock prices and other financial information. To gather data from Google News we utilize the Google News API to search for news articles related to stocks and refine the results to include articles published in the past two years. These collected articles can then be stored in a database or spreadsheet for analysis. Similarly for retrieving data from Yfinance we use the Yfinance API to obtain stock prices of a stock and have the option to store this price data in a database or spreadsheet. Once data is collected from both Google News and Yfinance researchers can analyze it comprehensively to identify trends in stock performance over time. Additionally, this collected data can help identify any news articles that may have influenced fluctuations, in stock prices.

2023 International Conference on Ambient Intelligence, Knowledge Informatics and Industrial Electronics (AIKIIE) | 979-8-3503-1646-9/23/$31.00 ©2023 IEEE | DOI: 10.1109/AIKIIE60097.2023.10390502

II. Literature Survey

Eric utilized sentiment analysis methods, in his study. He gathered information from Twitter during the week of the 2020 US presidential election and also throughout the COVID-19 pandemic period. To transform and represent this data as signals Eric employed FIR (Finite Impulse Response) filtering techniques. To detect changes, in signals over time, he relied on changepoint detection methods. Additionally, Eric performed data interpolation. Examined recurring patterns using Fast Fourier Transform and Fast Fourier Transform -based filtering techniques [1].

The research paper authored by Jing and colleagues introduces a method that combines knowledge methods, with investor sentiment investigation to progress the accurateness of predicting stock prices. Their model utilizes deep learning algorithms. Sentiment analysis to capture and analyze the sentiments expressed by investors through sources like social media and news articles. The study's experimental results demonstrate the effectiveness of this approach in forecasting stock prices providing insights into the application of deep learning and sentiment analysis in financial forecasting. Overall, this research contributes, to advancing stock market prediction methods by considering the impact of investor sentiment. [2]

Xingchen focused on using Natural Language Processing techniques to analyze the sentiment articulated in financial news reports about 87 companies that were extensively covered by Reuters over seven years. The study yielded results. Firstly, it was discovered that in industry sectors when there is a positive or negative sentiment towards one specific company in the media.

Furthermore, the research established a moderate connection between heightened media sentiment and abnormal market returns and increased market volatility. [3]

De Oliveira et al. Recently published a paper that outlines a methodology, for enhancing investment strategies in the stock market. They propose combining sentiment analysis with network techniques to analyze investor sentiment from sources and make predictions about stock market trends using deep learning algorithms. The study highlights how this approach can significantly improve investment decision-making by leveraging sentiment-driven predictions. Overall their research contributes to the field of investment strategies by offering a methodology that integrates sentiment analysis and deep learning resulting in more accurate and well-informed portfolio generation, within the Brazilian stock market. [4]

The study conducted by Pooja and her colleagues focuses on using sentiment analysis, on media to improve stock market forecasting using learning techniques. This research explores how sentiment analysis methods can be combined with the deep learning algorithms to enhance the accuracy of market predictions. The findings highlight the significance of integrating social media sentiment analysis with learning to improve the effectiveness of stock market prediction and decision-making processes. [5]

The research conducted by Rubi and colleagues explores the use of sentiment analysis techniques to predict stock prices. They examine how sentiment analysis can be applied to analyze data from sources, like news articles and social media posts aiming to extract insights related to sentiments that can be used for forecasting stock prices. This investigation highlights the potential of sentiment analysis as a tool for predicting stock market trends. It enhances our understanding of how sentiments impact stock prices providing insights for developing accurate prediction models. Overall this research emphasizes the role of sentimental analysis in advancing methodologies for predicting stock prices, that can have implications, for making informed investment decisions. [6]

The study conducted by László and colleagues explores the field of predicting changes, in stock values by analyzing the emotions expressed in stock news headlines. This research investigates how sentiment analysis techniques can assess the tone conveyed in news headlines about stocks and its subsequent impact on stock value fluctuations. The outcomes, from this learning, significantly contribute to our understanding of how sentiment influences stock markets offering insights that can guide the development of models utilizing sentiment analysis of news headlines[7,8].

The research paper written by Ali and his colleagues focuses on sentiment analysis techniques used in analyzing data, from social media platforms to predict stock price movements[9,10]. The study thoroughly examines how sentiment analysis can extract information from social media discussions about stocks and evaluate its effectiveness in forecasting stock price trends. This research highlights the importance of analyzing sentiments expressed on media platforms for predicting stock market trends. It provides insights into the relationship between social media sentiment and subsequent changes in stock prices[11,12]. The findings of this study contribute significantly to improving models by incorporating sentiment analysis based on data, from media. This enhancement helps investors make decisions by providing better predictions of stock price movements[13].

Bollen and Mao conducted a research study that focused on the impact of Twitter, on stock market sentiment. Their research explores how Twitter's mood can potentially predict stock market behaviour. They carefully analyzed the sentiments expressed in tweets over a while and used this data to make predictions about changes in the stock market. The argument of their paper emphasizes that Twitter provides insights into sentiment and can influence how the stock market behaves. The researchers used a machine learning algorithm to evaluate the sentiment of tweets related to the stock market and other financial topics. With this data, they created a "mood index" that could anticipate shifts in the stock market. To validate their model's effectiveness Bollen and Mao examined data from the period between October 2008 and December 2008 which was marked by economic turmoil. Impressively their model demonstrated precision in predicting changes in the stock market. In summary, this paper suggests that Twitter is a tool for forecasting fluctuations, in the stock market. However, it also acknowledges limitations that need to be taken into consideration. The authors point out that Twitter sentiment can be affected by factors such, as events and media attention which makes it difficult to pinpoint the exact influence of mood, on the stock market[14].

The research, which was published in the journal Applied Computer Systems investigates how sentiment analysis methods can be used to forecast changes, in stock market prices in Ghana. This study explores how analyzing the sentiment expressed in news articles can help understand fluctuations in stock prices within the market. By examining

the emotions conveyed in these articles the research demonstrates that sentiment analysis can be a tool for predicting stock market trends and gaining insights into the association among sentiment and stock price dynamics in Ghana. The findings of this study provide evidence of the practicality of using sentiment analysis to forecast stock market movements offering information about sentiments and stock prices, in the Ghanas market environment [15].

III. Methodology

In this section authors explain the clear picture of the anticipated procedure, in this view figure 1 displayed the block diagram of the anticipated method.

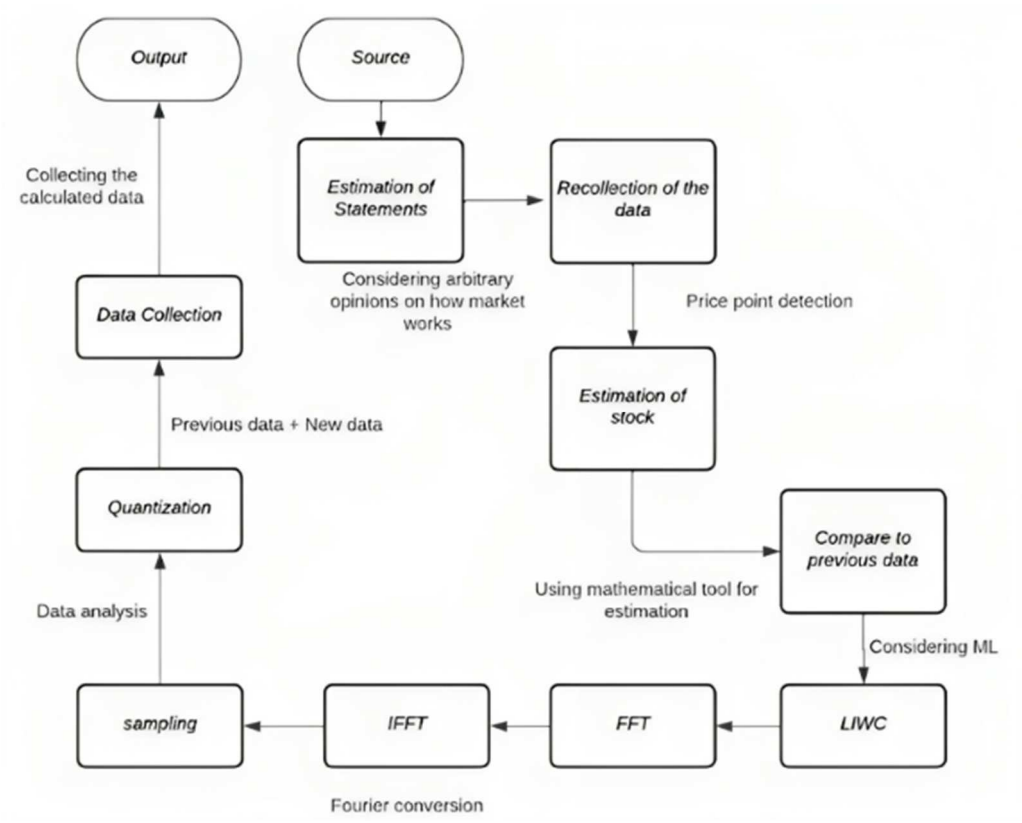

Fig. 1. Block Diagram

A. Source

This represents the source of data for stock market analysis. It could be financial databases, APIs, news sources, or any other relevant data sources.

B. Data Collection

This step involves gathering the necessary data from the chosen sources. It could include financial statements, historical stock prices, news articles, and other relevant information.

C. Estimation of Statements

In this step, various financial reports like income reports, balance slips, and cash run accounts are estimated based on the collected data. This provides a comprehensive view of the company's financial performance.

D. Recollection of Data

This step involves reviewing and validating the collected data to ensure accuracy and reliability. It may involve data cleaning, removing outliers, or filling in missing values.

E. Estimation of Stock

Using the collected and validated data, the estimation of stock includes analysing several issues such as economic performance, industry movements, and market conditions to estimate the value or potential performance of a specific stock.

F. Compare Previous Data

This step compares the current data with historical data to identify patterns, trends, and changes over time. It helps in understanding the stock's performance and making comparisons to make informed decisions.

G. VADER

It is a tool designed to analyze the emotions expressed in written content. It is widely used in Natural Language Processing. Text Analytics to determine whether a text segment conveys negative or neutral sentiment. VADER provides an automated way to assess the sentiment polarity of a given text.

H. Fast Fourier Transform

It is a mathematical algorithm used to transform time-domain data (such as stock price series) into its frequency-domain representation. It helps identify dominant frequencies or periodic components in the data.

I. Inverse Fast Fourier Transform

It is the inverse operation of the *Fast Fourier Transform*. It transforms the frequency-domain representation back to the

time-domain representation, allowing the reconstructed data to be analyzed.

J. Sampling

Sampling refers to selecting a subset of the data for analysis. It can help reduce computational complexity and focus on specific time intervals or data points of interest.

K. Quantization

Quantization involves converting continuous data into discrete values. It can be used to simplify data representation or reduce the number of levels or categories in the analysis.

L. Output

The climax of this analysis presents outcomes, projections or suggestions that have been derived from the actual examination. It's important to note that the schematic image provided is merely a depiction of the technique, the specific methods and tools used may differ depending on the needs of the analysis and personal preferences.

M. Datasets sources

- News data is taken from Google news API which has a collection of most of the newspaper publishers.
- The historical stock market data is taken from the yfinance library in Python.

IV. Procedure/Algorithm

A. Linear Regression

Statistical modelling, through regression, allows us to establish a connection, between a variable and one or more independent variables. We can express the variable as a combination of the independent variables accounting for some additional error terms as it is assumed that these variables have a linear relationship. The basic form of a linear regression model with one autonomous adjustable is given by the equation:

$$y = \beta_0 + \beta_1 x \tag{1}$$

where:

- y is a dependent variable,

- x is an independent variable,

- β_0 is a y-intercept or the constant term,

- β_1 is slope or coefficient of independent variable x,

$$\beta_1 = \frac{\Sigma((x_i - \bar{x})(y_i - \bar{y}))}{\Sigma((x_i - \bar{x})^2)} \tag{2}$$

$$\beta_0 = \bar{y} - \beta_1 \bar{x} \tag{3}$$

where:

- Σ represents the sum of,

- xi and yi are individual values of independent and dependent variables respectively.

- $\bar{x}$ and $\bar{y}$ are means of independent and dependent variables respectively.

Linear regression can be extended to multiple independent variables by adding additional terms with their corresponding coefficients. The general equation for a multiple linear regression is:

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_p x_p + \varepsilon \tag{4}$$

B. Lasso Regression

Lasso regression, a form of regression also known as L1 regularization enhances the least squares method by incorporating a penalty component. This approach promotes sparsity, in the coefficients allowing certain ones to become zero. Doing it improves the accuracy of the model and facilitates variable selection. This attribute makes Lasso regression beneficial, for feature selection and reducing model complexity. In Lasso regression, the objective function is modified by adding a regularization term that penalizes the magnitude of the coefficients. The modified objective function to be minimized is:

$$\frac{1}{2n}\Sigma\left(y_i - \beta_0 - \sum\beta_j x_{ij}\right)^2 + \alpha\sum|\beta_j| \tag{5}$$

where:

- y_i is a observed value of dependent variable for ith data point,

- β_0 is a y-intercept or a constant term,

- β_j represents coefficient for jth independent variable,

- x_{ij} is a value of jth independent variable for ith data point,

- α is a regularization parameter that controls the strength of the penalty term,

- Σ represents the sum of all data points.

C. Ridge Regression

Ridge regression, also known as L2 regularization is a technique used in regression to improve accuracy. It adds a penalty term to the least squares approach. Similar, to Lasso regression Ridge regression can assist with feature selection. Unlike Lasso regression Ridge regression doesn't require sparsity. Only reduces the coefficients, towards zero without setting them to zero. By introducing a regularization component that penalizes the magnitude of the coefficients Ridge regression adjusts the function.

The modified objective function to be minimized is:

$$\frac{1}{2n}\Sigma\left(y_i - \beta_0 - \sum\beta_j x_{ij}\right)^2 + \alpha\sum\beta_j^2 \tag{6}$$

where the variables have the same meaning as in the ordinary least squares method and Lasso regression.

D. Support Vector Machine

The fundamental principle of SVMs is to maximize the generalizability of the model by identifying an ideal hyperplane with the highest margin of separation between data points of various classes. In terms of mathematics, SVM seeks to solve the following problem:

$$\text{minimize:} \frac{1}{2}\|w\|^2 + C\Sigma\varepsilon_i \tag{7}$$

$$\text{subjectto: } y_i = (w \cdot x_i + b) \geq 1 - \varepsilon_i \,\forall\, I \; and \; \varepsilon_i \geq 0 \,\forall\, i \tag{8}$$

where:

- w is a weight vector orthogonal to hyperplane...

- b is a bias term.

- C is a regularization parameter that controls trade-off between maximizing margin and allowing misclassifications,

- (x_i, y_i) represents the training examples, with x_i being the feature vector and y_i being the class label (+1 or -1),

- ξ_i are slack variables that allow for misclassifications or points that fall within the margin or on the wrong side of the hyperplane.

E. Performance Evaluation Criteria

These metrics assess the performance of your linear regression model. The high R-squared values and the relatively low values of MAE, MSE, and RMSE indicate that your model performs well and has a good fit to the data.

V. Result And Discussion

We have taken HDFC Bank as a sample case to show the result Figure 2 displays the sentiment analysis without FFT and Figure 3 displays the sentiment analysis with FFT.

A. Sentiment Analysis with filtering

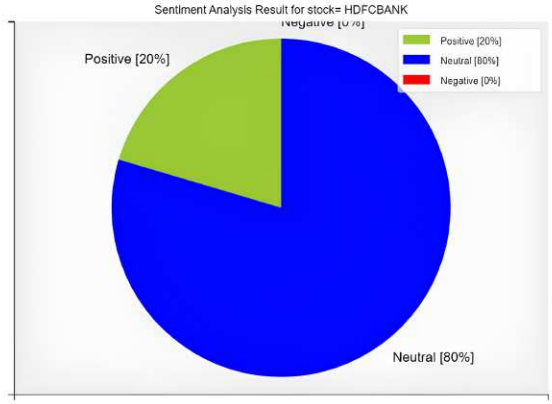

Fig. 2. Sentiment Analysis without FFT

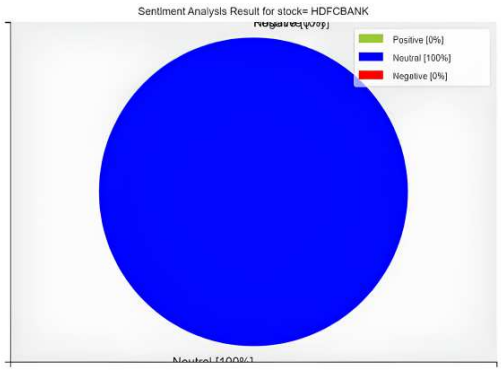

Fig. 3. Sentiment analysis with FFT

The reason for the difference in results after filtering is as follows

When working with datasets it is essential to have an understanding of the Fast Fourier Transform and its counterpart the Inverse Fast Fourier Transform. These processes have an impact on the resulting data values so comprehending their properties is crucial. The FFT is a procedure used to convert time domain signals, such as text data into their frequency domain equivalents. This transformation divides the signal into sinusoidal components with frequencies. By analyzing the amplitudes and phases of these components in the domain representation we gain insights. On the other hand, the IFFT operates by converting the frequency domain representation back, into its original time domain form. It reconstructs the signal using information obtained from the FFT frequency domain components. When applying FFT to text data it essentially converts sequences of characters into numbers that represent amplitudes and phases of frequency components. This transformation can unveil patterns or structures in data that are associated with frequencies. However, it's important to note that text data doesn't naturally exhibit periodicity or oscillatory behaviour required for frequency analysis.

Consequently, the way text data is represented in the domain might not offer easily understandable information. The transformed data can appear as a combination of values, with sizes and phases. When you use the inverse fast Fourier transform (IFFT) on the frequency domain representation obtained from the fast Fourier transform (FFT) you aim to reconstruct the text data. However, since the frequency domain representation may include values with varying sizes and phases there may be differences between the reconstructed data and the original text data. This reconstruction process involves some level of approximation. This could lead to disparities compared to the input.

It's important to note that in theory FFT and IFFT operations are generally considered lossless. However, due to limitations in finite precision computational capabilities, there can be errors or rounding issues that contribute to discrepancies between the original and reconstructed data. In summary when applying FFT and IFFT to text data resulting values can deviate from the input due to these operation's nature and inherent characteristics of information. It's crucial to evaluate whether frequency analysis is appropriate, for your use case and interpret outcomes accordingly. Table I shows comparison result of different machine learning algorithms and Table II shows the actual price and predicted price of the proposed model.

TABLE I. Comparison Of Different Ml Algorithms

Metrics	Support Vector Regression	Ridge Regression	Lasso Regression	Linear Regression
Mean Absolute Error (MAE)	21.14	7.77	11.48	7.83
Mean Squared Error (MSE)	602.66	91.21	221.88	95.62
Root Mean Squared Error (RMSE)	24.54	9.55	14.89	9.77

TABLE II. Actual Price vs. Predicted Price

Company	Date	Predicted Price (Rs)	Actual Price (Rs)

Adani Enterprises	10 /7/23	2402	2410
	11 /7/23	2419	2422
	12 /7/23	2411	2387
	13 /7/23	2386	2362
	14 /7/23	2371	2376
Infosys	10 /7/23	1330	1329
	11 /7/23	1340	1348
	12 /7/23	1336	1333
	13 /7/23	1347	1365
	14 /7/23	1395	1425
HDFC Bank	10 /7/23	1661	1656
	11 /7/23	1655	1648
	12 /7/23	1637	1632
	13 /7/23	1645	1641
	14 /7/23	1644	1635
TCS	10 /7/23	3289	3271
	11 /7/23	3279	3272
	12 /7/23	3250	3259
	13 /7/23	3331	3340
	14 /7/23	3449	3514

VI. Conclusion

In times there has been a surge, in research focused on understanding how we can predict stock market returns. Researchers in this field have been using computational and soft computing techniques to analyze sentiments and determine if they can assist in forecasting stock market movements. Previous studies have already made progress in improving the accuracy of predictions. According to literature incorporating online information sources along with public sentiment could further enhance the precision of market forecasts. This study explores how the attitude of the public towards news both positive and negative extracted from the internet can be utilized to predict fluctuations in stock prices. Our primary methodology for this investigation is regression. The results strongly support the assertion that Ridge Regression outperforms its counterparts, exhibiting a superior fit for the dataset at hand (as shown in Table I). To conduct our experiments, we collected stock data from January 2022 to March 2023 from four companies listed on the stock exchange. The results confirm that public sentiments can indeed be used to predict stock market behaviour. Notably by employing signal processing techniques to reduce noise in the news data, we observe an improvement in the accuracy of our model when predicting stock prices for a day (as shown in Table II). This emphasizes the potential for advancements in stock prediction models, through incorporating data sources related to the stock market. Results of study are really important for policymakers because they help us understand how monetary policy affects the markets. It also reminds institutions like the Reserve Bank of India to be aware of the effects that their statements can have on financial market uncertainty. Lastly, the research suggests that combining both technical approaches to predicting markets, in developing economies is a promising area worth exploring in the future.

References

[1] Mayor, E., Bietti, L.M. & Canales-Rodríguez, E.J. Text as signal. A tutorial with case studies focusing on social media (Twitter). Behav Res (2022). https://doi.org/10.3758/s13428-022-01917-1

[2] Jing, Nan, Zhao Wu, and Hefei Wang. "A hybrid model integrating deep learning with investor sentiment analysis for stock price prediction." Expert Systems with Applications 178 (2021): 115019.

[3] Wan, X., Yang, J., Marinov, S. et al. Sentiment correlation in financial news networks and associated market movements. Sci Rep 11, 3062 (2021). https://doi.org/10.1038/s41598-021-82338-6

[4] de Oliveira Carosia, Arthur Emanuel, Guilherme Palermo Coelho, and Ana Estela Antunes da Silva. "Investment strategies applied to the Brazilian stock market: a methodology based on sentiment analysis with deep learning." Expert Systems with Applications 184 (2021): 115470.

[5] Mehta, Pooja, Sharnil Pandya, and Ketan Kotecha. "Harvesting social media sentiment analysis to enhance stock market prediction using deep learning." PeerJ Computer Science 7 (2021): e476.

[6] Gupta, Rubi, and Min Chen. "Sentiment analysis for stock price prediction." 2020 IEEE conference on multimedia information processing and retrieval (MIPR). IEEE, 2020.

[7] Nemes, László, and Attila Kiss. "Prediction of stock values changes using sentiment analysis of stock news headlines." Journal of Information and Telecommunication 5.3 (2021): 375-394.

[8] Derakhshan, Ali, and Hamid Beigy. "Sentiment analysis on stock social media for stock price movement prediction." Engineering Applications of Artificial Intelligence 85 (2019): 569-578.

[9] J. Bollen and H. Mao. Twitter mood as a stock market predictor. IEEE https://1drv.ms/w/s!AjFtyI9W-WqzgrkKHpLF1QcgScFzogComputer, 44(10):91–94.

[10] Nti, Isaac Kofi, Adekoya, Adebayo Felix and Weyori, Benjamin Asubam. "Predicting Stock Market Price Movement Using Sentiment Analysis: Evidence From Ghana" Applied Computer Systems, vol.25, no.1, 3920, pp.33-42.

[11] Štrimaitis, Rokas, Pavel Stefanovič, Simona Ramanauskaitė, and Asta Slotkienė. "Company recommendation model: Empowering the accounting system and publicly available data." (2022).

[12] Kushwaha, Bijay Prasad, Atul Shiva, and Vikas Tyagi. "How Investors' Financial Well-being Influences Enterprises and Individual's Psychological Fitness? Moderating Role of Experience under Uncertainty." Sustainability 15, no. 2 (2023): 1699.

[13] Kumari, Sharda, Bibhas Chandra, and Jamini Kanta Pattanayak. "Personality traits and motivation of individual investors towards herding behaviour in Indian stock market." Kybernetes 49, no. 2 (2019): 384-405.

[14] Pandey, Dharen Kumar, and Rahul Kumar. "Lockdown, unlock, stock returns, and firm-specific characteristics: the Indian tourism sector during the Covid-19 outbreak." Current Issues in Tourism 25, no. 7 (2022): 1026-1032.

[15] Ashtiani, Matin N., and Bijan Raahmei. "News-based intelligent prediction of financial markets using text mining and machine learning: A systematic literature review." Expert Systems with Applications (2023): 119509.