International Journal of Advance and Applied Research
www.ijaar.co.in
ISSN – 2347-7075 Impact Factor – 8.141
Peer Reviewed Bi-Monthly
Vol.12 No.1 September – October 2024

Applying Market News Sentiment Analysis To Stock Market Prediction

Chavan Rutuja Santosh[1] & Dr.Prasadu Peddi[2]
[1]Ph.D Research Scholar,
Department of Computer Science and Application,
Shri JJT University, Jhunjhunu, Rajasthan, India
[2]Assistant Professor and Ph.D Research Guide
Department of Computer Science and Application,
Shri JJT University, Jhunjhunu, Rajasthan, India
Corresponding Author - Chavan Rutuja Santosh
DOI - 10.5281/zenodo.15075333

Abstract:

The ability to accurately predict market movements has become increasingly important in light of the ongoing problems that have unquestionably had an impact on the financial sector. Due to the low success rate of portfolio risk management models, there is a need to investigate how risk managers can more accurately foresee the evolution of portfolios while also taking into account the systemic risks that are caused by a systemic crisis. The analysis of sentiment in natural language sentences has the potential to improve the accuracy of market predictions. This is because the sentiments of investors have an impact on the financial markets. There are a great number of investors who rely their choices on the information they obtain from newspapers or on their own instincts. The purpose of this study is to demonstrate how sentiment analysis can improve the accuracy of regression models when it comes to anticipating the opening price movements of selected equities. Our goal is to accomplish this by comparing the results and accuracy of two different market prediction scenarios that make use of regression models. One of these scenarios includes market news sentiment analysis, while the other does not include it. Through the utilization of sentiment analysis as an exogenous variable, the nonlinear autoregression model exhibits a higher degree of goodness of fit. When compared to linear models, the data suggest that polynomial autoregressions provide a better fit than linear models. It may be concluded that the utilization of sentiment scores in market modeling displays significant improvements in the efficiency of linear autoregressions.

Keywords: Sentiment Analysis; Market Prediction; Beautiful Soup; ARX; VADER

Introduction:

As a result of the inherent complexity and unpredictability of the financial business sectors, forecasting the developments of the securities exchange has always been a fundamental and challenging task in the field of financial analysis. The accuracy of these forecasts has a significant impact on the systems that are used for speculation, portfolio management, and risk management practices. Generally speaking, the estimation of the securities exchange was heavily dependent on quantitative monetary data, such as cost patterns, specialized indicators, and macroeconomic factors. On the other hand, as a result of the advent of big data and advancements in natural language processing (NLP), feeling analysis has emerged as a valuable asset that can be used in conjunction with conventional methods.

Deciphering subjective data, namely the close-to-home tone and setting of message data, is the centre of opinion analysis. This is done in order to determine how the general public feels about monetary resources. The collective brain science of market participants can be mirrored in the wealth of opinion data that can be gleaned from market gatherings, online journals, virtual entertainment, and monetary news. One of the most important contributions that financial backer sentiment makes to predictive models is the fact that it frequently influences the evolution of stock prices. A good view about a firm, for instance, might generate hopefulness, which in turn can drive stock prices higher. On the other hand, a bad emotion may motivate sell-offs, which in turn may cause costs to decrease.

Through the use of opinion analysis into financial exchange projection models, professionals are able to capture the interaction between market factors and the sentiments of financial backers. This strategy makes use of both structured mathematical data and unstructured written data, hence enhancing the predictive power of these models. This is accomplished by combining feeling ratings with autoregressive models. This study investigates the ways in which opinion research, carried out with the assistance of tools such as the Valence Mindful Word reference and Feeling Reasoner (VADER), might improve the accuracy of relapse models for securities exchange expectations. The purpose of this investigation is to demonstrate the practical value of opinion analysis in contemporary financial forecasting by comparing different models and avoiding the use of scores as exogenous factors.

When it comes to financial big data, risk directors have the ability to consider a vast array of resources, which is even more significant than the size of their portfolios. In particular, this ease can bring about a higher precision of chance forecasting, which is very useful when it comes to catching the planned gamble that is built into commercial sectors. As a result, in order to preserve the robustness of the financial business sectors and, consequently, reduce the likelihood that a fundamental gamble would materialize, monetary controllers can reap the benefits of a higher degree of precision in risk anticipation for market participants. The financial information of public organizations is routinely disseminated on a foundation that is fairly engaging, which results in an unmistakable delay. In the meantime, the rate at which monetary data will generally be reported in the news on the economy will most likely be more consistent.

Due to the growing popularity of high-frequency trading, the procedures that are used for portfolio sizing and stock selection are highly stringent. In point of fact, the information that investors can view on a continuous basis is associated with stock trading data. This includes information such as the opening, most elevated, least, and closing prices of stock, in addition to many specialized indicators, etc. Furthermore, as a result of the growing availability of electronic data, investors are now able to make use of their decision-making feelings in light of text data that can be obtained within financial news on the internet. This information can then be incorporated into the evaluation of the value of the company's security.

Old-fashioned and contemporary methodologies can be divided into two categories, depending on the data that was proved to be able to predict resource costs. This is due to the fact that resource costs incorporate a number of different characteristics into their value. Consequently, the primary methodology might incorporate such data as stock data boundaries and "monetary record and profit and misfortune explanation boundaries" [1].

In the meantime, the two are put together in [2] by the analysis of the organization, the investigation of the industry, the macroeconomic indicators, the political conditions, the topographical and meteorological conditions, and the political conditions. During this interim period, the term "specialized examination" refers to the research of several specialized indicators, such as costs [3, feeling, crude data, volume, cycle, unpredictability, stream of assets [2], or other specialized indicators [1].

Despite the fact that a significant portion of the specialized and essential data are presented in an organized manner in the conventional methodologies, the cutting-edge approaches may be able to perform on unstructured data sources. These sources are primarily obtained through electronic financial news, virtual entertainment, websites, online gatherings, and other electronic communications [4]. Because there are more websites and online clients, it is typically difficult to locate and organize the data that is relevant to the situation. The process of "scratching" a website in order to extract data from it is referred to as "web scratching." Although it is theoretically possible to scratch other data sources, such as record papers, it is actually possible to do so. On the other hand, the vast majority of scratching jobs are often carried out on website pages.

As is the case with problems pertaining to conduct finances, costs are merely a perceived worth [5]. There is a rational readiness to investigate the impact that the perspectives of society have on the costs of resources. This technique is known as evaluation mining, and it involves identifying feelings (whether positive or negative) through the use of words.

The concept of financial backer sensation is currently being utilized by a number of academics in order to calculate the growth of stock value and portfolio advancement [6-8]. In their article, [9] compiled analyses that were pertinent to the exploitation of web-based information for the purpose of estimating advances in the securities exchange. According to their specific findings, it appears that the field of organization informing may have some prospective applications in the field of financial estimation.

In addition, a number of academics have started combining data from other sources in order to evaluate the profitability of the endeavor. There have been a number of stock-cost forecasting studies directed, which have utilized a variety of data sources. These tests include those directed by [10-13], amongst others. Additionally, to the best of our knowledge, a few academics have suggested the utilization of big data in order to investigate stock selection and portfolio development; nevertheless, the viability of this concept has not been proved (i.e., [7]). As a result, we would like to demonstrate how data science methodologies could be employed to identify stocks that are suitable for investment in a protections market that has a great deal of resources.

The new piece of writing suggests a number of different ways for conducting securities exchange expectations, which are useful for both essential and specialist evaluation. There are a variety of artificial intelligence calculations that include arrangement methods such as support vector machines, k-closest neighbors, strategic relapses, innocent Bayes, choice tree grouping, and arbitrary timberland characterization; relapse procedures such as polynomial relapse, basic straight relapse, choice tree relapse, irregular woods relapse, and backing vector relapse; fluffy rationale calculations; profound brain organizations; hereditary calculations; and artificial brain networks [14-18]. It was discovered in [15] that there are two examples of acknowledgment processes, specifically layout coordinating and perceptually significant focuses (PIPs).

The AI market-expectation process was proposed to include additional stages, which can be summarized into three stages, as illustrated in Figure 1.

1. Dataset preparation: reports, blogs, online news, social media, or market data [3,4];
2. Data preprocessing [3]:
   - Dimensionality reduction,
   - feature representation, and
   - feature selection (with regard to natural language processing);
3. Evaluation and modeling.

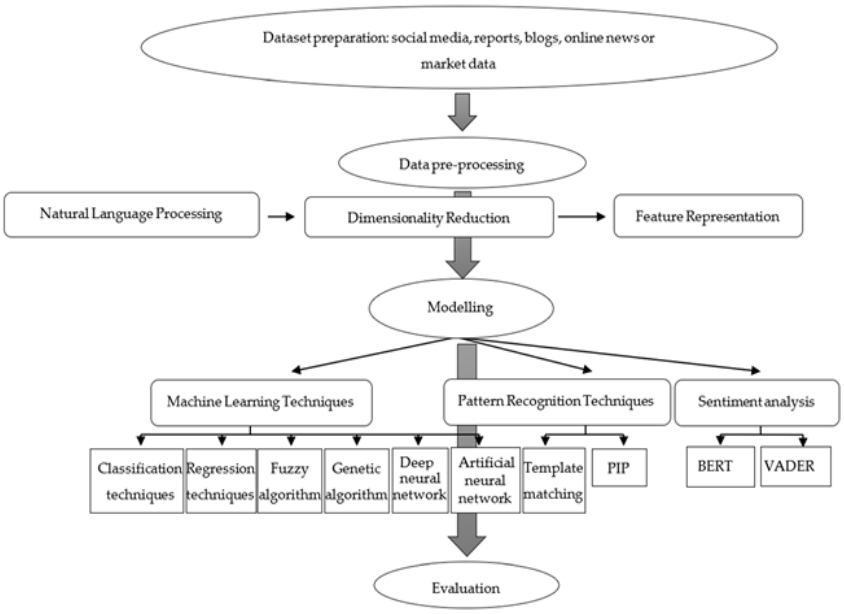

Figure 1. Applying machine learning to the process of market prediction.

The following techniques were delegated for the purpose of determining highlights in [3]: sack of words; n-grams as continual successions of words; hereditary calculations; and province streamlining. A few examples of highlight portrayal approaches that could be utilized in the meanwhile include data gain (IG), chi-square measures (CHI), archive recurrence (DF), precision adjusted (ACC2), term recurrence opposite record recurrence (TF-IDF), paired/Boolean (0/1) or opinion esteem.

Opinion inquiry is one of the regular language handling (NLP) tasks [19], and it refers to the process of analyzing human feelings that are conveyed through messages. In light of the fact that there are many different degrees of granularity, such as reports, sentences, or viewpoints, the primary objectives of opinion analysis are to obtain classes according to extremity (positive/negative/impartial articulation), point grouping (determining whether an articulation is subjective or objective), and incongruity location (determining whether an expression is unexpected). As a result of the growth of informal communities and their implementation in a variety of businesses, such as consumer goods and medical services, statistical analysis of financial feelings has a significant deal of potential applications.

Opinion investigation may for the most part incorporate some element portrayal procedures alongside AI methods (e.g., the pack of-words model in addition to help vector mama chines [20]), however a few models have been broadly utilized, specifically Pre-training of Profound Bidirec-tional Transformers (BERT) and the

Valence Mindful Word reference for Feeling Thinking (VADER) [21-23].

Additionally, the exploration configuration is presented in Figure 2, while the next section includes the strategic technique that was utilized to consolidate the feeling examination in market expectations. We utilized an autoregression using an exogenous element model (ARX), in addition to quadratic and cubic relapses, in order to arrive at our motivation. At the beginning of the third section, the results are presented, beginning with charts that cover the opinion score for each and every stock that was dissected. Segment 3.1 also includes the introduction of a scatter plot, which is used to construct the connection between the opinion score and the stock starting cost. On the other hand, the significance of coordinating the sensation score into these relapses is highlighted in Segment 3.2, which is also where the relapse models are presented and discussed. The discussion of the results and a correlation with other studies that are similar is offered in the fourth segment, while the continuation of the discussion focuses on the primary conclusions.

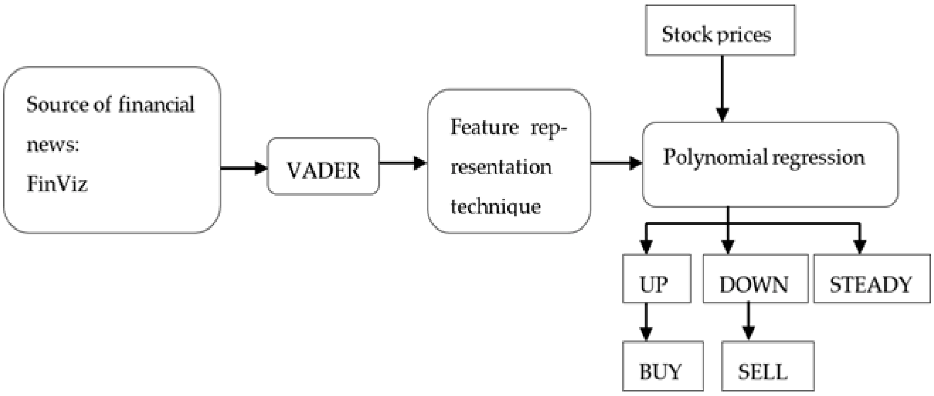

Figure 2. Research design.

Research Methodology:

For the purpose of carrying out our inquiry, we made use of the FinViz platform in order to acquire financial news on a variety of dynamic stocks. In addition to that, we made use of Python, which was the second most popular programming language in the year 2020, following C [24]. The most significant update to this programming language, which was first introduced in 1991, took place in 2008 with the release of Python 3.0. This language has undergone subsequent revisions over the course of its history. Python provides a comprehensive standard library that includes a variety of useful tools, such as Lovely Soup, which is a package that may be utilized for scraping and parsing site data [25].

Following the identification of the data source, we executed a Python script that makes use of Delightful Soup to scrape article titles from FinViz. FinViz is a platform that allows users to investigate the securities market through the use of an internet browser [26]. Subsequently, we utilized VADER to execute the opinion investigation, and we utilized Pandas (the Python data examination package) to break down and return the ensuing feeling investigation scores for the titles of the monetary articles.

Wonderful Soup is the most well-known package for scraping and processing site data of all the available options. The creators of the library claim that it is capable of deciphering any information. In order to accomplish this, BeautifulSoup employs easy algorithms and Pythonic terms in order to construct a parse tree that is both secure and easily accessible. One of the benefits of utilizing BeautifulSoup is that it produces an interpretation of the data that has been parsed to UTF-8, which is a design that is widely used on the web [27]. For the purpose of data collection, we utilized a web scrubber that was constructed with the help of various tools that were accessible in the BeautifulSoup library.

The application of a model is essential in order to steer the research of feelings. "Valence Mindful Word reference for Opinion Thinking" (also known as "VADER") is a straightforward rule-based approach that is used for the assessment of general feelings. In addition to being sensitive to extremes as well as the intensity of feelings, this model is also capable of being applied to text data that has not been tagged. The National Language Toolkit (NLTK) bundle is what people remember most about VADER. This bundle covers a stage for the construction of Python applications that enables working with human language data [28]. The advantages of traditional emotion dictionaries, such as LIWC (semantic request and word count), are enhanced by VADER, which shares and improves upon these advantages. When compared to LIWC, VADER is distinguished by its ability to summarise a greater number of domains and its greater receptivity to emotional expressions in the context of web-based entertainment environments. Hutto and Gilbert were given the opportunity to plan and observe a number of lexical features that are particularly sensitive to feeling under conditions similar to those of a microblog. The VADER, in addition to eleven other highly regarded opinion investigation devices, continued to proceed [22]. As a result, VADER can be applied for the purpose of conducting a sentiment analysis of titles pertaining to financial news that is disseminated in the online environment and shared through virtual entertainment. Nonetheless, it is crucial to make reference to that there were different dates in which the picked stocks were not covered by the significant money news distributors that FinViz collects its data from; consequently, the feeling score was 0.

In the demonstrating stage, two direct autoregressions were carried out once the opinion score was collected. One of the autoregressions did not include an exogenous component, while the other included one (the exogenous variable being the feeling score). The results of Condition (1) and Condition (2), respectively, revealed the following:

$$y = \beta1 + \beta2yt-1 + \beta3yt-2 + \beta4yt-3 + \beta5yt-4 + \varepsilon \qquad (1)$$

$$y = \beta1 + \beta2yt-1 + \beta3yt-2 + \beta4yt-3 + \beta5yt-4 + \beta6x + \beta7xt-1 + \beta8xt-2 + \varepsilon \qquad (2)$$

where β1 represents the catch, β2 represents the slant, yt−1/2/3/4 represents the stock-opening-cost change factors in the earlier days, x and xt−1 represent the feeling score exogenous components in the present and earlier days, and ε represents the blunder term. y represents the change in the stock cost at time t. In the mean time straight, quadratic and cubic relapses were utilized in order to analyze the relationship between the sensation score and the change in the stock opening price. Condition (3), Condition (4), and Condition (5) each introduce the connection for the direct, quadratic, and cubic autoregressions. In addition, Condition (5) introduces the link for each independent variable:

$$y = \beta1 + \beta2x + \varepsilon \qquad (3)$$

$$y = \beta1 + \beta2x + \beta3x2 + \varepsilon \qquad (4)$$

$$y = \beta1 + \beta2x + \beta3x2 + \beta4x3 + \varepsilon \qquad (5)$$

The variables y, β1, β2, β3, and β4 represent the opaque borders, x represents the opinion score variable, and ε represents the blunder term. The stock opening cost change is being represented by y. Is the change in stock cost at time t, denoted by Y.

Following the execution of the straight, quadratic, and cubic relapses, we proceeded to execute a nonlinear autoregression with an exogenous component (NARX), as should be evident in Condition (6):

$$y = \beta_1 y_{t-1} + \beta_2 y_{t-2} + \beta_3 y_{t-3} + \beta_4 y_{t-4} + \beta_5 x + \beta_6 x^2 + \varepsilon \quad (6)$$

According to the equation, y represents the original cost change in time t, x represents the opinion score, and yt-1/2/3/4 represents the initial cost changes in time t−1, t−2, t−3, and t−4 respectively.

In addition, we utilized the data that we had collected in order to carry out direct autoregressions with and without an exogenous component. In this process, the sensation score was accumulated with the market capitalization weight. The results indicate that the models with equal weight would be recommended to determinant coefficients for the training data; hence, we made sure to keep an equivalent load for each of the relapses that were remembered for the current investigation.

The period of time that was investigated included the period of time from August to September 2022, which is a total of 37 days. In order to compile all of the data, similar weight was utilized.

As a result of their prevalence, as well as the fact that everybody who has any interest in the securities market and puts resources into this file is presented to them, a significant majority of the organizations that took part in the study were selected from the S&P 500 indices. Due to the fact that Solidarity Advances will be conducting its first initial public offering (first sale of stock) on September 17, 2020 [29], this particular organization was selected. In addition, we included SONY, a company that is not based in the United States, in order to observe the differences in terms of opinion scores between firms that are located within the United States and those that are located outside of the United States. There was a separation between The New York Stock Exchange and the market data that was provided by the businesses.

Analysis:

Sentiment Analysis:

Following the process of scratching the message from FinViz for each selected organization and analyzing the feeling of the articles from August and September 2022, we came to the conclusion that a typical opinion score in the insight was around 0.06, and a normal unpredictability score was approximately 0.18. Figure 3 presents the feeling scores for each stock, and from this figure, we can see the angles that accompany each of the scores.

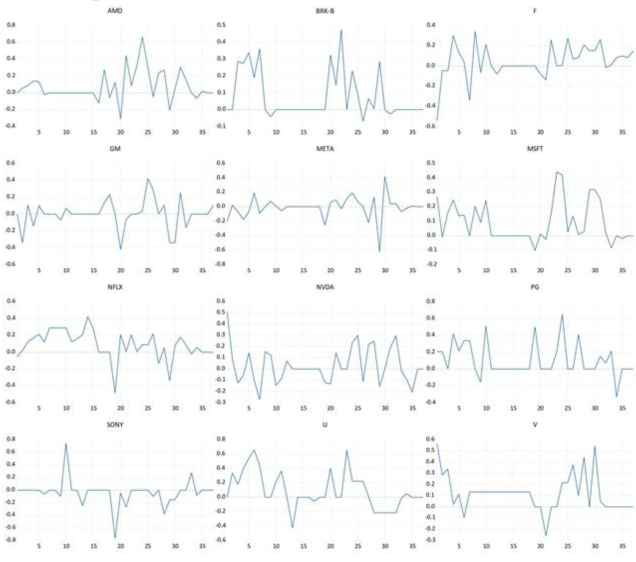

Figure 3. Sentiment analysis charts

The average sensation score for AMD (High level Miniature Gadgets) stock was a positive one, with a value of 0.08, which was higher than the complete normal value, and an instability score of 0.18.

On the whole, the feeling score in the news that was associated with BRK-B (Berkshire Hathaway Inc. Class B) stock was 0.11, which indicates that there was a good feeling in the news that was associated with this stock. With a score of 0.202, which was considered to be a generally high one, the unpredictability score was higher than the normal instability level of 0.18. In addition, the normal opinion score for the stock of F (Portage Engine Organization) was 0.04, which indicates that the news associated with this stock is generally favorable. Additionally, the stock had a below optimal unpredictability of 0.17. Regarding Microsoft Corporation (MSFT), a favorable sentiment was discovered to be associated with stock, with an average opinion score of 0.09 and the least amount of instability being 0.14 within the time period that was investigated. For the stocks of GM (General Engines) and META (Meta, Inc., formerly Facebook, Inc.), the typical feeling scores were -0.001 and -0.001, indicating pessimistic opinions in the news associated with these stocks. Additionally, both of these equities had an unpredictability of 0.17 and 0.16. In the mean time, the normal opinion score for NFLX (Netflix) stock was 0.09, indicating a positive emotion in the news connected with this stock. Additionally, the unpredictability score was 0.14 during the examined period, which was lower than the normal instability score of 0.18. In the same vein as Microsoft, the news associated with PG (Procter and Bet) received a score of 0.11 for positive opinion and a score of 0.20 for unpredictability, both of which were higher than what was anticipated.

The mean feeling score for NVDA (Nvidia) was 0.03, which indicates that there is a positive attitude regarding the news that is associated with this stock. Additionally, the stock had an unpredictability score of 0.16, which was rather inexpensive. The news that was related to V (Visa) stocks had the highest increased positive opinion score, which was 0.12, and the unpredictability score was 0.17. During the mean time period, the normal opinion score for U (Solidarity Innovations) was a positive one with a value of 0.09, and it had the highest level of unpredictability, which was 0.25. The news that was associated with Sony received the lowest opinion score, which was a value of -0.04, while its instability stayed better than expected, with a value of 0.20 according to the evaluation.

In terms of the unpredictability of opening costs, Figure 4 illustrates the period instability for each stock that was analyzed individually.

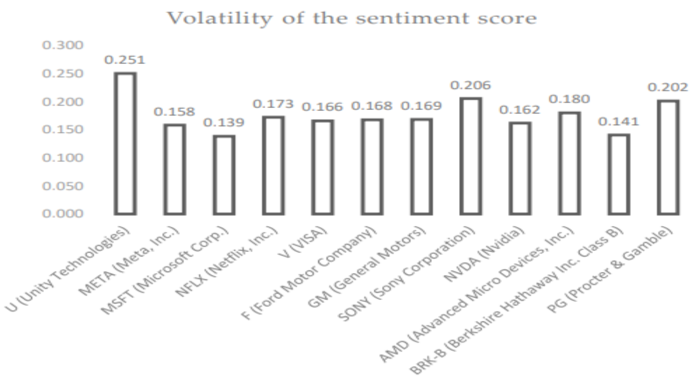

Figure 4. During the time under consideration, the opening price of the stock was subject to fluctuation.

In spite of the fact that the U (Solidarity) stock had the highest quality deviation (instability) of 0.25, the MSFT (Microsoft Corp.) stock had the least amount of unpredictability, which was 0.14. Temporarily, the initial price of AMD stock increased on 19.08.2022 from that of 18.08.2022. This was due to the fact that it introduced the most notable feeling score for Cutting edge Miniature Gadgets (AMD) on 19.08.2022, which was 0.66, and the most elevated normal opinion score, which was 0.12, during the period that was mentioned earlier. This may suggest that the impression of the news may have played a role in bringing about this transformation.

Discussion:

The findings of this research shed light on the significant role that opinion analysis plays in improving projections for the securities exchange and working to improve the accuracy of relapse models. The significance of combining subjective and quantitative data for more reliable deciding is highlighted in the study through the incorporation of opinion ratings obtained from financial news with autoregressive models. According to the findings of the sentiment analysis, the selected stocks exhibited a variety of opinion scores and volatilities, which highlighted the fluctuating market characteristics that they possessed. It is remarkable that stocks such as Visa and Procter & Gamble, which had the highest positive emotion scores, substantial areas of strength for reflected certainty, and moderate volatilities, were among the most successful. On the other hand, Solidarity Innovations demonstrated the most significant unpredictability, regardless of whether or not they received a positive sentiment score. This demonstrates that opinion alone is not sufficient to fully explain cost developments, and that other market factors include significant influence.

It was established that the usage of nonlinear relapse models, such as quadratic and cubic autoregressions, was superior to the utilization of straight models in terms of capturing the intricate connections that exist between opinion scores and stock prices. As evidenced by the higher R-squared values, the nonlinear autoregressive model with an exogenous factor (NARX) demonstrated the advantages of incorporating emotional analysis into predictive statistical modeling. Based on these observations, it is clear that the financial business sectors are inherently nonlinear, and it is imperative that high-level exhibiting tactics be used in order to achieve precise expectations. In addition, stock-specific pieces of information shed light on the complex influence that opinion has on the performance of the market. For example, Microsoft had minimal instability and consistent certain opinion, which reflected the company's dependability and versatility. On the other hand, Sony, which received the lowest possible feeling score, revealed challenges that were observed by firms that were not based in the United States.

The more far-reaching implications of this study extend to a variety of different partners. Financial backers can employ opinion analysis to identify more secure venture opportunities, and risk managers can consolidate feeling data for improved risk evaluation, particularly in high-frequency trading settings. Both of these methods can be utilized to identify more secure venture opportunities. Additionally, financial controllers may make advantage of emotional patterns in order to anticipate fundamental opportunities and prevent market upsets during the process. Nevertheless, the research also came to the conclusion that there are certain limitations. The investigation was damaged by data gaps as a result of contradicting news inclusion, and high-unpredictability stocks, such as Solidarity Advancements, required additional components in addition to feeling analysis for the purpose of producing far-reaching demonstrations. Concerns of a moral nature, such as the possibility of exerting influence over the opinions of business sector stakeholders, are also worthy of consideration.

By researching advanced opinion models such as BERT, coordinating multimodal data, and dissecting larger time spans in order to capture stable cases, further exploration may be able to address these limitations. This study demonstrates the critical competence of opinion examination in financial exchange expectations, notwithstanding the challenges that have been encountered. The discoveries support its reconciliation into relapse models as a key gadget, while also underscoring the requirement for a nuanced appreciation of its transaction with more expanded market factors and persistent strategic progressions.

Conclusion:

For the purpose of improving the degree of match for the expectation of the stock costs via the utilization of relapse models, the opinion component was utilized in the present academic work. In light of the financial news aspects that FinViz accommodated, we applied the VADER model to generate the feeling score that occurred on a daily basis. This score was based on the stocks that were selected. After that, we carried out three different types of relapse models, which were straight, quadratic, and cubic autoregressions. We discovered that the polynomial autoregressions had a higher R square pointer than the direct one, which lends credence to the findings that were presented in [14]. In addition, in order to work on the decency of attack of the autoregression, we employed the emotion score as an exogenous figure in the nonlinear autoregression. Our outcomes are trustworthy with [14,30,35], while further establishing the proof that remembering opinion examination as an exogenous variable for relapse type models can boost the decency of spasm of the models. Generally speaking, the outcomes show that utilizing the opinion score as an exogenous figure the direct autoregression raised the R coefficient from 0.189 to 0.192; hence, we can presume that coordinating the feeling factor in market relapse examination produces a superior relapse in regards to the decency of fit.

References:

1) Atkins, A.; Niranjan, M.; Gerding, E. Financial News Predicts Stock Market Volatility Better than Close Price. J. Financ. Data Sci. 2018, 4, 120–137.
2) Beautiful Soup Documentation. Available online: https://beautiful-soup-4.readthedocs.io/en/latest/ (accessed on 12 July 2022).
3) Breuss, M. Beautiful Soup: Build a Web Scraper with Python. Preuzeto 2021, 30, 2021.
4) Bustos, O.; Pomares-Quimbaya, A. Stock Market Movement Forecast: A Systematic Review. Expert Syst. Appl. 2020, 156, 113464.
5) Creamer, G.G. Can a Corporate Network and News Sentiment Improve Portfolio Optimization Using the Black–Litterman Model? Quant. Financ. 2015, 15, 1405–1416.
6) Devlin, J.; Chang, M.W.; Lee, K.; Toutanova, K. BERT: Pre-Training of Deep Bidirectional Transformers for Language Under- standing. In Proceedings of the Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Minneapolis, MA, USA, 2–7 June 2019; Volume 1, pp. 4171–4186.

7) Ferreira, F.G.D.C.; Gandomi, A.H.; Cardoso, R.T.N. Artificial Intelligence Applied to Stock Market Trading: A Review. IEEE Access 2021, 9, 30898–30917.
8) Feuerriegel, S.; Gordon, J. Long-Term Stock Index Forecasting Based on Text Mining of Regulatory Disclosures. Decis. Support Syst. 2018, 112, 88–97.
9) Giudici, P.; Mezzetti, M.; Muliere, P. Mixtures of Products of Dirichlet Process for Variable Selection in Survival Analysis. J. Stat. Plan. Inference 2003, 111, 101–115.
10) Hochreiter, R. Computing Trading Strategies Based on Financial Sentiment Data Using Evolutionary Optimization. In Advances in Intelligent Systems and Computing; Springer International Publishing: New York, NY, USA, 2015; pp. 181–191. ISBN 9783319198248.
11) Hu, Y.; Liu, K.; Zhang, X.; Su, L.; Ngai, E.W.T.; Liu, M. Application of Evolutionary Computation for Rule Discovery in Stock Algorithmic Trading: A Literature Review. Appl. Soft Comput. 2015, 36, 534–551.
12) Hutto, C.J.; Gilbert, E. VADER: A Parsimonious Rule-Based Model for sentiment analysis of social media text. In Proceedings of the international AAAI Conference on Web and Social Media, Catalonia, Spain, 17–21 July 2014; p. 18.
13) Jothimani, D.; Shankar, R.; Yadav, S.S. A Big Data Analytical Framework for Portfolio Optimization. Cell 2014, 3, 1–15.
14) Kumbure, M.M.; Lohrmann, C.; Luukka, P.; Porras, J. Machine Learning Techniques and Data for Stock Market Forecasting: A Literature Review. Expert Syst. Appl. 2022, 197, 116659.
15) Leow, E.K.W.; Nguyen, B.P.; Chua, M.C.H. Robo-Advisor Using Genetic Algorithm and BERT Sentiments from Tweets for Hybrid Portfolio Optimisation. Expert Syst. Appl. 2021, 179, 115060.
16) Nagy, J. Behavioral Economics and the Effects of Psychology on the Stock Market; State University of New York College at Buffalo—Buffalo State College: Buffalo, NY, USA, 2017.
17) Nardo, M.; Petracco-Giudici, M.; Naltsidis, M. Walking down wall street with a tablet: A survey of stock market predictions using the web. J. Econ. Surv. 2016, 30, 356–369.
18) Nasukawa, T.; Yi, J. Sentiment Analysis. In Proceedings of the International Conference on Knowledge Capture—K-CAP '03; ACM Press: New York, NY, USA, 2003; p. 70.
19) Natural Language Toolkit. Available online: https://www.nltk.org/ (accessed on 12 July 2022).
20) Nti, I.K.; Adekoya, A.F.; Weyori, B.A. A Systematic Review of Fundamental and Technical Analysis of Stock Market Predictions; Springer: Dordrecht, The Netherlands, 2020; Volume 53, ISBN 0123456789.
21) Ravikumar, S.; Saraf, P. Prediction of Stock Prices Using Machine Learning (Regression, Classification) Algorithms. In Proceedings of the 2020 International Conference for Emerging Technology (INCET), Belgaum, India, 5–7 June 2020; pp. 1–5.
22) Rouf, N.; Malik, M.B.; Arif, T.; Sharma, S.; Singh, S.; Aich, S.; Kim, H.C. Stock Market Prediction Using Machine Learning Techniques: A Decade Survey on Methodologies, Recent Developments, and Future Directions. Electronics 2021, 10, 2717.
23) Sadorsky, P. A Random Forests Approach to Predicting Clean Energy Stock Prices. J. Risk Financ. Manag. 2021, 14, 48.
24) Schumaker, R.P.; Chen, H. Textual Analysis of Stock Market Prediction Using Breaking Financial News.

ACM Trans. Inf. Syst. 2009, 27, 1–19.

25) Shah, D.; Isah, H.; Zulkernine, F. Stock Market Analysis: A Review and Taxonomy of Prediction Techniques. Int. J. Financ. Stud. 2019, 7, 26.
26) Shynkevich, Y.; McGinnity, T.M.; Coleman, S.A.; Belatreche, A. Forecasting Movements of Health-Care Stock Prices Based on Different Categories of News Articles Using Multiple Kernel Learning. Decis. Support Syst. 2016, 85, 74–83.
27) Siganos, A.; Vagenas-Nanos, E.; Verwijmeren, P. Divergence of Sentiment and Stock Market Trading. J. Bank. Financ. 2017, 78, 130–141.
28) Theodorou, T.I.; Zamichos, A.; Skoumperdis, M.; Kougioumtzidou, A.; Tsolaki, K.; Papadopoulos, D.; Patsios, T.; Papanikolaou, G.; Konstantinidis, A.; Drosou, A.; et al. An AI-Enabled Stock Prediction Platform Combining News and Social Sensing with Financial Statements. Futur. Internet 2021, 13, 138.
29) Tung, L. Programming Language Pythons' Popularity Ahead of Java for the First Time but Still Trailing c. Available online: https:// www.zdnet.com/article/programming-language-pythons-popularity-ahead-of-java-for-first-time-but-still-trailing-c/ (accessed on 12 July 2022).
30) Unity Software Inc. Financials. Available online: https://www.crunchbase.com/organization/unity-technologies/company_financials (accessed on 12 July 2022).
31) Usmani, S.; Shamsi, J.A. News Sensitive Stock Market Prediction: Literature Review and Suggestions. PeerJ Comput. Sci. 2021, 7, e490.
32) Wei, Y.-C.; Lu, Y.-C.; Chen, J.-N.; Hsu, Y.-J. Informativeness of the Market News Sentiment in the Taiwan Stock Market. North Am. J. Econ. Financ. 2017, 39, 158–181.
33) Weng, B.; Ahmed, M.A.; Megahed, F.M. Stock Market One-Day Ahead Movement Prediction Using Disparate Data Sources. Expert Syst. Appl. 2017, 79, 153–163.
34) Yıldırım, D.C.; Toroslu, I.H.; Fiore, U. Forecasting Directional Movement of Forex Data Using LSTM with Technical and Macroeconomic Indicators. Financ. Innov. 2021, 7, 1–36.
35) Zhang, X.; Zhang, Y.; Wang, S.; Yao, Y.; Fang, B.; Yu, P.S. Improving Stock Market Prediction via Heterogeneous Information Fusion. Knowl.-Based Syst. 2018, 143, 236–247.
36) Zhou, Z.; Gao, M.; Liu, Q.; Xiao, H. Forecasting Stock Price Movements with Multiple Data Sources: Evidence from Stock Market in China. Phys. A Stat. Mech. Its Appl. 2020, 542, 123389.