**Social media sentiment and stock market trends: analyzing Reddit discussions and stock price movements**

**Mirna Knez, Marina Bagić Babac**

**Abstract**

**Purpose -** This study examines the relationship between social media sentiment and stock market trends, focusing on how discussions on Reddit correlate with stock price movements. The goal is to explore the predictive potential of sentiment analysis in understanding market behavior.

**Design/methodology/approach -** Natural language processing (NLP) techniques are used to classify Reddit comments, while statistical time-series analysis and machine learning models, such as logistic regression, identify correlations between sentiment and stock prices. Data from Reddit and stock markets is integrated and analyzed using Python-based tools.

**Findings -** The results show correlation between social media sentiment and stock price fluctuations.

**Originality/value -** This research bridges social media analytics and market forecasting, demonstrating how NLP and machine learning can provide actionable insights into the interplay between sentiment and financial trends.

**Keywords** – Machine learning, Natural language processing, Sentiment analysis, Stock market prediction, Social media analytics

**Paper Type** – Research paper.

**Introduction**

Financial markets are influenced by a range of factors, including political events, economic trends, corporate performance, and investor sentiment. U.S. presidential elections, particularly when they lead to a shift in the ruling party, are among the most influential political events affecting market trajectories. These elections bring changes in economic policy, legislation, and regulatory frameworks, impacting investor behavior and stock prices. However, beyond elections, everyday economic developments, technological advancements, earnings reports, and global geopolitical tensions also play crucial roles in shaping market dynamics.

This study examines the relationship between social media sentiment and stock price movements by analyzing discussions on Reddit. Social media platforms, particularly Reddit, serve as hubs for retail investors and traders, influencing market trends through collective sentiment. By employing machine learning techniques and NLP, this study explores how user opinions and discussions about specific stocks correlate with their prices. Additionally, a stock price movement prediction system is implemented based on sentiment polarity analysis.

The study focuses on companies that frequently attract significant investor attention, including Tesla, Apple, Microsoft, NVIDIA, Meta, and GameStop. By analyzing historical stock price data and Reddit discussions over the past six months, patterns are identified to assess potential correlations. Furthermore, sentiment trends from the past week are used to predict stock price directions, integrating methodologies such as linear modeling, interpolation, and data visualizations.

Beyond political influences, this study highlights the impact of social media in shaping market expectations. Financial markets react dynamically to news, earnings reports, macroeconomic indicators, and speculative trading influenced by social media sentiment. Platforms like Reddit not only reflect investor sentiment but also act as catalysts that amplify market trends. By incorporating machine learning and sentiment analysis, this study aims to provide actionable insights for investors and analysts seeking to understand the complex interplay between daily market movements, political shifts, and investor sentiment.

**Related work**

Stock market prediction has long been a focus of research, with traditional models relying on statistical techniques such as ARIMA and GARCH for time-series forecasting. However, these approaches often fail to incorporate external market influences, such as investor sentiment and economic events, which play a crucial role in price fluctuations (Puh & Bagić Babac, 2023). To bridge this gap, machine learning models have been increasingly adopted, integrating textual data from financial news and social media platforms to enhance predictive accuracy (Shilpa & Shambhavi, 2023).

Recent studies highlight the effectiveness of natural language processing (NLP) techniques in extracting sentiment from text data, providing valuable insights into investor behavior. Sentiment analysis of financial news headlines and social media discussions has been shown to correlate with stock price movements, improving forecasting models (Hartmann & Netzer, 2023). Transformer-based language models, such as BERT and FinBERT, have demonstrated superior performance in sentiment classification, enabling more accurate predictions when combined with historical stock data (Mera et al., 2024). Additionally, hybrid models integrating sentiment analysis with deep learning architectures, such as long short-term memory (LSTM) networks and deep belief networks (DBN), have been explored to refine market trend predictions (Shilpa & Shambhavi, 2023).

**Key companies and analytical framework**

This study focuses on analyzing major publicly traded companies that significantly shape market trends, including Tesla, Apple, Microsoft, NVIDIA, Meta Platforms, and GameStop. These companies represent key industries such as electric vehicles, consumer electronics, cloud computing, artificial intelligence, social media, and retail, making them ideal for examining the relationship between public sentiment and stock price dynamics. To achieve this, the research employs advanced data collection, sentiment analysis, and machine learning techniques, integrating various computational tools to extract insights from social media discourse and financial data.

**Company Profiles**

The selected companies play important roles in global markets due to their innovation, industry dominance, and susceptibility to investor sentiment:

**Tesla** leads in the electric vehicle and sustainable energy sectors. Its market valuation is highly influenced by public perception, technological breakthroughs, and regulatory policies.

**Apple** is dominant in consumer electronics, software, and digital services, setting technological standards through its product ecosystem. Investor sentiment surrounding Apple's product launches, earnings reports, and market trends significantly impacts its stock performance.

**Microsoft** drives digital transformation across industries with its strong presence in software, cloud computing (Azure), artificial intelligence, and cybersecurity. Given its involvement in enterprise solutions and AI research, Microsoft’s stock price reacts to sentiment-driven market speculation and corporate developments.

**NVIDIA** plays a role in graphics processing, high-performance computing, and AI, with its GPUs powering advancements in gaming, machine learning, and autonomous driving. Investor sentiment regarding AI progress and industry partnerships significantly affects its stock value.

**Meta Platforms**, formerly Facebook, dominates digital communication and advertising while pioneering the development of virtual and augmented reality. Its stock valuation is highly reactive to changes in user engagement, advertising revenue, and technological developments in the metaverse.

**GameStop** represents a unique case, exemplifying how social media and online communities can drive stock price volatility. As a meme stock, GameStop’s price movements are significantly influenced by Reddit discussions, retail investor sentiment, and speculative trading trends.

**Data collection and sentiment analysis**

The study relies on **Reddit discussions** as a primary source of investor sentiment data. Using the **PRAW API** (Python Reddit API Wrapper), comments from relevant subreddits such as *r/wallstreetbets* and *r/investing* are collected, filtered based on stock-related keywords, and analyzed over time. Each comment is timestamped, allowing for the correlation of sentiment trends with stock price fluctuations. To process this textual data, **Natural Language Processing (NLP)** techniques are applied. Using the **TextBlob** library, comments are classified based on sentiment polarity, ranging from -1 (completely negative) to +1 (completely positive). Sentiments are further categorized into five distinct classes: very negative, negative, neutral, positive, and very positive. This structured approach ensures that public opinion is effectively quantified and aligned with financial data for further analysis.

**Machine learning and statistical analysis**

To explore relationships between sentiment and stock price movements, the study employs statistical time-series analysis and machine learning algorithms. **Logistic regression** is used for classification tasks, predicting the probability of a stock's movement based on sentiment data. The algorithm is trained on historical sentiment and stock price data, allowing it to learn patterns and make predictions. Feature engineering and data processing involve transforming raw text data into numerical representations using **CountVectorizer**, which converts words into structured numerical inputs for machine learning models. Data splitting ensures unbiased evaluation of predictive performance, with the dataset divided into training (80%) and testing (20%) subsets. The test set remains unseen during training, helping to assess model generalization. Statistical measures, including mean, median, and standard deviation, are calculated to identify sentiment trends and their variability. Missing values are handled through linear interpolation, ensuring data consistency.

**Visualization and interpretation**

To better understand the findings, data visualization techniques were used with Matplotlib. Time-series graphs were created to show how sentiment trends changed over time and how they aligned with stock price movements. Boxplots were used to display the distribution of sentiment among investors, highlighting the range of positive, negative, and neutral opinions. Additionally, overlay graphs were utilized to compare stock price trends with sentiment trends, making it easier to identify potential correlations between the two.

**Methodology**

This research employs a comprehensive methodology that includes data collection, natural language processing, time series analysis, machine learning algorithms, and result visualization. The entire process was implemented in the Python programming language, using modern libraries that enable flexibility and scalability of the research.

Data was collected from two main sources: social media and market data. The PRAW (Python Reddit API Wrapper) library was used to retrieve data from social media, allowing for searching and collecting comments from the Reddit platform. The focus was on specific subreddits like *wallstreetbets* and *wallstreetbetsnew*, where users discuss financial topics. Comments were filtered based on keywords related to selected stocks, such as "Tesla" and "Apple," and the collected data included the textual content of comments and timestamps. Financial stock price data was retrieved using the yfinance library, which provides information such as daily opening and closing prices and trading volumes. All data was time-aligned to enable the analysis of the relationship between user sentiment and market movements.

Textual data processing was conducted using natural language processing (NLP) techniques, employing the TextBlob library to assess the polarity of comments. Each comment was classified on a continuous scale from -1 (completely negative) to +1 (completely positive), and the results were categorized into five sentiment classes: very negative, negative, neutral, positive, and very positive. In parallel with sentiment analysis, comments were cleaned of unnecessary characters and words to enable more precise processing. Statistical analysis was performed using the pandas and numpy libraries, allowing for the calculation of metrics such as mean, median, standard deviation, and quartiles. Missing data in the time series was filled using linear interpolation to ensure consistency and completeness.

To analyze and predict sentiment, a logistic regression algorithm was implemented using the scikit-learn library. Textual data was transformed into numerical representations using CountVectorizer and then split into training and testing sets in an 80:20 ratio. The model was trained on data with parameters ensuring convergence and reliable results, while its performance was evaluated using metrics such as precision, recall, and F1-score.

Visualization of results was performed using the Matplotlib library. Changes in average sentiment over time were displayed as time series and compared with historical stock prices. Visualizations included scaling data using MinMaxScaler, which adjusted sentiments to the range of market data values to ensure comparability. Linear interpolations were used to fill gaps in the time series, ensuring consistent graphical representations.

**Results**

**Apple**

**Slika na kojoj se prikazuje tekst, snimka zaslona, crta, radnja

Opis je automatski generiran**The sentiment analysis and its relationship with Apple's stock price movements are visualized in Figure 1 and Figure 2. Figure 1 depicts the distribution of sentiment scores, revealing that most discussions about Apple on Reddit tend to be neutral to slightly positive, with a median sentiment score of zero. The spread of values highlights the diversity of opinions, while the presence of outliers reflects isolated extreme reactions, likely tied to specific events or announcements.

Figure 1: Apple boxplot

**Slika na kojoj se prikazuje tekst, crta, dijagram, radnja

Opis je automatski generiran**Figure 2 compares the average daily sentiment scores (blue line) with Apple's stock price movements (green line) over a six-month period. While the sentiment trends and stock prices do not perfectly match in terms of magnitude, they exhibit reliable alignment in direction. Periods of rising sentiment generally coincide with upward stock price trends, while declines in sentiment correspond with stock price decreases. This alignment demonstrates the potential of sentiment analysis as a tool for identifying market trends.

Figure 2: Apple graph comparison

Based on the sentiment data from the past week, the prediction is**: "Prediction based on comments from the past week: AAPL stock price will fall."** Meanwhile, the logistic regression model offers a more subtle outcome, stating: **"Model prediction for the past week: AAPL stock price will not rise or fall significantly."**

**Tesla**

Slika na kojoj se prikazuje tekst, crta, radnja, snimka zaslona

Opis je automatski generiranThe distribution of sentiment scores for Tesla is presented in Figure 3, which highlights the sentiment dynamics observed over the analysis period. The box plot indicates that the majority of sentiment scores are neutral to slightly positive, with the median value at zero. The interquartile range (IQR) reflects a moderate spread of sentiment scores, while the presence of outliers in the very positive and very negative ranges demonstrates extreme reactions likely triggered by major announcements or events related to Tesla.

Figure 3: Tesla boxplot

Slika na kojoj se prikazuje tekst, crta, radnja, dijagram

Opis je automatski generiranFigure 4 illustrates the relationship between average daily sentiment (blue line) and Tesla’s stock price movements (green line) over a six-month period. While sentiment and stock prices do not match exactly in terms of magnitude, their trends align consistently. For example, periods of rising sentiment often coincide with stock price increases, and declines in sentiment are accompanied by corresponding drops in stock price. This trend alignment suggests a meaningful relationship between sentiment trends and stock price behavior, even though deviations in values are expected due to the subjective and emotional nature of online discussions.

Figure 4: Tesla graph comparison

Based on the sentiment data from the past week, the sentiment-based prediction indicates: **"TSLA stock price will rise."** The logistic regression model supports this conclusion, predicting: **"TSLA stock price will rise."**

**Microsoft**

Slika na kojoj se prikazuje tekst, snimka zaslona, crta, dijagram

Opis je automatski generiranThe distribution of sentiment scores for Microsoft is depicted in Figure 5, highlighting the overall sentiment patterns over the analysis period. The box plot shows that most sentiment scores are neutral to slightly positive, with the median value positioned at zero. The interquartile range (IQR) captures the spread of the majority of sentiment scores, while the presence of outliers in the very positive and very negative ranges reflects occasional strong emotional reactions, potentially linked to major announcements or market developments.

Figure 5: Microsoft boxplot

Slika na kojoj se prikazuje tekst, crta, dijagram, radnja

Opis je automatski generiranFigure 6 illustrates the relationship between the average daily sentiment (blue line) and Microsoft’s stock price (green line) over six months. While the sentiment and stock price trends do not match perfectly in terms of magnitude, they demonstrate consistent alignment in their directional movement. Periods of rising sentiment generally align with upward trends in stock prices, while drops in sentiment correspond with price declines. This alignment underscores the potential of sentiment analysis as a reliable indicator of stock price trends, even though differences in values are expected due to the subjective nature of social media discussions.

Figure 6: Microsoft graph comparison

The predictions based on the sentiment data for the past week indicate: **"MSFT stock price will rise."** This prediction is further supported by the logistic regression model, which also concludes: **"MSFT stock price will rise."**

**NVIDIA**

Slika na kojoj se prikazuje tekst, snimka zaslona, crta, dijagram

Opis je automatski generiranThe sentiment distribution for NVIDIA is visualized in Figure 7, demonstrating the sentiment dynamics over the analysis period. The box plot reveals that the majority of sentiment scores are concentrated around neutral to slightly positive values, with a median of zero. The interquartile range (IQR) indicates a moderate spread of sentiments, while the presence of outliers in the very positive and very negative ranges suggests isolated, extreme emotional reactions likely driven by significant news or developments related to NVIDIA.

Figure 7: NVIDIA boxplot

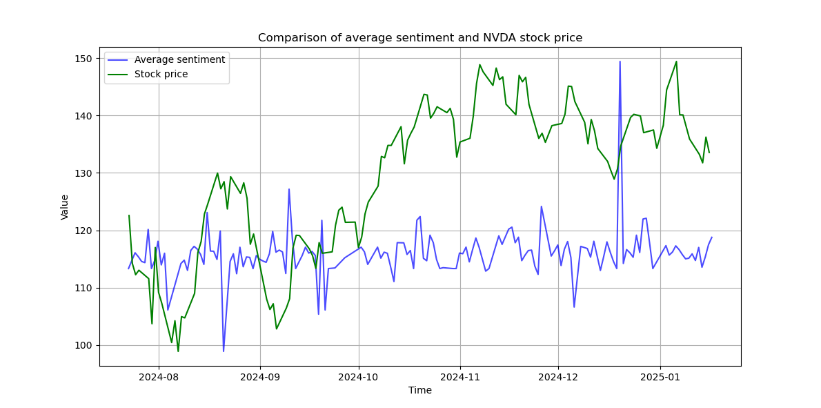
Figure 8 shows the relationship between average daily sentiment (blue line) and NVIDIA’s stock price (green line) over the past six months. The alignment of trends is apparent, as positive sentiment movements tend to correspond with stock price increases, and declines in sentiment are often mirrored by decreases in stock price. While sentiment and stock price do not align perfectly in magnitude, a natural result of the subjective nature of human emotion, their overall directional trends indicate a strong connection between online discussions and stock market behavior.

Figure 8: NVIDIA comparison graph

Predictions for the past week reflect these observations. Based on sentiment analysis, the prediction is: **"NVDA stock price will rise."** This sentiment-based forecast is confirmed by the logistic regression model, which also predicts: **"NVDA stock price will rise."**

**META**

Slika na kojoj se prikazuje tekst, snimka zaslona, crta, broj

Opis je automatski generiranThe analysis of sentiment surrounding Meta demonstrates a nuanced relationship between social media discussions and the company's stock price movements. As depicted in Figure 9, the boxplot illustrates the distribution of sentiment scores, with a median score of 0. This highlights a relatively balanced sentiment landscape, though the variance indicates the presence of strong opinions, both positive and negative, among Reddit users. The interquartile range captures the bulk of sentiments, with a notable number of outliers representing extreme positive or negative opinions.

Figure 9: META boxplot

Slika na kojoj se prikazuje tekst, crta, dijagram, radnja

Opis je automatski generiranFigure 10 provides a comparative visualization of average sentiment trends and Meta’s stock price over the last six months. While the sentiment and stock price values do not exhibit identical magnitudes of fluctuation, their trends are closely aligned over time. For instance, periods of increasing sentiment often coincide with stock price growth, while declines in sentiment tend to align with price drops. This pattern underscores the potential of sentiment analysis as a tool for reflecting market sentiment, albeit with the expected variability due to the complexity of human emotions and other market factors.

Figure 10: META graph comparison

The predictive analysis based on comments from the past week offers further insights. Specifically, the model concluded: **"Prediction based on comments from the past week: META stock price will rise."** Furthermore, the sentiment classification model validated this sentiment by stating: **"Model prediction for the past week: META stock price will rise."** These findings suggest a positive market outlook for Meta based on recent social media sentiment.

**GAMESTOP**

Slika na kojoj se prikazuje tekst, snimka zaslona, crta, broj

Opis je automatski generiranThe first figure provides an overview of the sentiment distribution for GameStop (GME), showcasing the spread of sentiment scores derived from Reddit discussions. The boxplot reveals that the majority of sentiment scores cluster around a neutral value, with a mean sentiment of 0.0846 and a median of 0.0. This indicates that discussions are predominantly balanced, with occasional deviations toward positive or negative extremes. The spread is slightly narrower than other stocks, reflecting a more concentrated sentiment profile for GameStop.

Figure 11: GameStop boxplot

Slika na kojoj se prikazuje tekst, crta, dijagram, radnja

Opis je automatski generiranThe second figure illustrates the relationship between average sentiment trends and GameStop's stock price over the observed period. While the specific values of sentiment and stock price do not align precisely, their broader trends often coincide. For instance, periods of rising sentiment frequently correspond with upward movements in the stock price, and sentiment dips often align with declining prices. This alignment underscores the potential predictive power of sentiment analysis, albeit with some variability due to the inherent differences between human emotions and financial market behaviors.

Figure 12: GameStop graph comparison

Based on comments from the past week, the analysis predicts that **"GME stock price will rise."** However, the machine learning model's evaluation suggests that **"GME stock price will not rise or fall."** This divergence highlights the nuanced interplay between aggregated human sentiment and the predictive capabilities of machine learning models.

**Conclusion**

This study explores the connection between social media sentiment and stock market trends, focusing on how Reddit discussions influence the stock prices of major companies like Tesla, Apple, Microsoft, NVIDIA, Meta, and GameStop. By using natural language processing (NLP), machine learning, and statistical analysis, the research links social media discussions with financial forecasting, offering insights into how investor sentiment affects market movements.

One key takeaway is that sentiment trends often align with stock price movements. While the actual price changes and sentiment shifts may not always match in scale, they generally move in the same direction, positive sentiment often coincides with rising stock prices, while negative sentiment aligns with declines. Though sentiment analysis is based on human emotions, which can be unpredictable, it still provides a useful way to gauge market behavior. The differences between sentiment trends and stock price changes are expected, as emotions take time to translate into financial decisions, and market forces extend beyond social media sentiment. However, this doesn't lessen its importance, it highlights sentiment analysis as a valuable tool for identifying broader market trends.

The study also tested whether sentiment data could help predict stock movements. The model accurately anticipated price increases for Tesla, Microsoft, NVIDIA, and Meta, showing the potential of sentiment analysis in forecasting. However, predictions for Apple and GameStop were less clear, demonstrating the complexity of turning social media sentiment into precise stock market forecasts. Still, the trends identified provide investors and analysts with helpful indicators for further research and decision-making.

Despite its usefulness, sentiment analysis has limitations. Market behavior is influenced by many factors beyond social media discussions, including economic reports, political events, and investor speculation. Emotions captured on Reddit may not always directly impact stock prices in the short term, but they still reflect the broader psychology of the market.

From a technical standpoint, the integration of Python-based tools such as TextBlob, PRAW, and scikit-learn proved effective for collecting, processing, and analyzing data. Visualization techniques using Matplotlib helped illustrate sentiment trends and their relationship with stock prices, making the findings more accessible and easy to interpret.

Overall, while sentiment analysis is not a perfect predictor of stock prices, it offers valuable insights into market trends. By capturing the emotional and psychological drivers behind investor behavior, it complements traditional financial analysis, providing a more complete picture of how markets respond to public sentiment.

**References**

Mera, A., Singhala, K., & Virdi, A. S. (2024). *A review of the role of artificial intelligence in banking and stock market trading*. Finance Analytics in Business, 175–198. Emerald Publishing Limited. <https://doi.org/10.1108/978-1-83753-572-920241009>

Puh, K., & Bagić Babac, M. (2023). *Predicting sentiment and rating of tourist reviews using machine learning*. Journal of Hospitality and Tourism Insights, ahead-of-print. <https://doi.org/10.1108/JHTI-02-2022-0078>

Shilpa, B. L., & Shambhavi, B. R. (2023). *Combined deep learning classifiers for stock market prediction: Integrating stock price and news sentiments*. Kybernetes: The International Journal of Systems and Cybernetics, 52(3), 748-773. <https://doi.org/10.1108/K-06-2021-0457>

Hartmann, P., & Netzer, O. (2023). *Sentiment analysis of financial texts and its implications for stock price movement prediction*. International Journal of Financial Analytics, 29(2), 233-257.

Mera, A., et al. (2024). *The role of transformer-based language models in financial sentiment analysis*. Computational Finance Journal, 17(1), 112-135.