Exploring Financial Trends through Topic Modeling and Time-Series Analysis: A Clustering Approach Using Latent Dirichlet Allocation (LDA) on Twitter Data

Deshinta Arrova Dewi[1,*], Tri Basuki Kurniawan

[1]Faculty of Science and Technology, Universitas Bina Darma, Indonesia

[2]Faculty of Data Science and Information Technology, INTI International University, Malaysia

ABSTRACT

Social media platforms, particularly Twitter, have emerged as influential arenas for financial discourse, shaping and reflecting market sentiment in real time. This study explores the thematic structure of financial discussions on Twitter, employing Latent Dirichlet Allocation (LDA) to identify key topics and their temporal dynamics. A dataset of 11,932 finance-related tweets was analyzed, revealing five distinct topics encompassing corporate earnings, macroeconomic policies, geopolitical trade issues, and market trends. By correlating tweet volumes and topic prevalence with significant financial events, the study demonstrates the utility of social media as a barometer for market activity. Unlike traditional sentiment analysis, which predominantly classifies tweets into sentiment categories such as bullish, bearish, or neutral, the application of LDA enabled the extraction of latent themes that underpin these sentiments. This nuanced approach provided deeper insights into the narratives driving market discussions, offering a more comprehensive understanding of how thematic shifts in financial discourse align with market movements. Visualization techniques, including topic-term matrices and word clouds, further elucidated the structure of these conversations, enhancing interpretability and accessibility. The findings contribute to the growing body of research on social media analytics in finance, highlighting the potential of unsupervised learning techniques for financial trend analysis. By bridging the gap between thematic exploration and temporal analysis, this study offers a methodological framework for leveraging social media data to uncover actionable insights. The implications extend beyond academic research, providing practical tools for investors, financial analysts, and policymakers to navigate the dynamic relationship between social media narratives and market behavior. Future research could expand on these insights by integrating more advanced modeling techniques, such as transformer-based models, and exploring domain-specific patterns across asset classes like stocks, commodities, and cryptocurrencies. By examining the intersection of social media, financial events, and market dynamics, this study lays the groundwork for a deeper understanding of digital narratives in financial ecosystems.

Keywords Social Media, Financial Discourse, Latent Dirichlet Allocation, Market Trends, Unsupervised Learning

Submitted 23 December 2024
Accepted 15 February 2025
Published 8 March 2025

Corresponding author
Deshinta Arrova Dewi,
deshinta.ad@newinti.edu.my

Additional Information and Declarations can be found on page 105

Introduction

The integration of social media into the digital society has revolutionized how individuals and organizations interact with information, fostering an ecosystem where influence permeates industries at an unprecedented scale. In the financial sector, social media has transitioned from a peripheral communication channel to a pivotal arena for disseminating critical insights. It empowers

How to cite this article: D.A.Dewi and T.B.Kurniawan, "Exploring Financial Trends through Topic Modeling and Time-Series Analysis: A Clustering Approach Using Latent Dirichlet Allocation (LDA) on Twitter Data ," *J. Digit. Soc.*, vol. 1, no. 1, pp. 91-108, 2025.

investors to access real-time market data, interpret emerging trends, and participate in collective financial discourse. Particularly among millennials, who dominate social media demographics, platforms like Twitter and Reddit serve not only as conduits for financial news but also as spaces for education and opinion exchange, bridging the gap between traditional financial literacy and modern market dynamics [1].

Beyond the realm of individual investors, social media has profound implications for broader economic systems. Its ability to amplify voices and democratize information reshapes traditional financial structures, fostering inclusivity that transcends geographic and socioeconomic barriers. For example, digital inclusive finance—a domain significantly driven by social media—empowers underserved populations by facilitating microloans, peer-to-peer lending, and localized investment opportunities in previously inaccessible regions [2]. This paradigm shift not only mitigates financial inequality but also cultivates resilience within marginalized communities, aligning with global development goals that prioritize economic equity and growth [2].

Interestingly, the influence of social media is not confined to the financial sector alone; its role extends seamlessly into shaping consumer behavior across multiple domains. Tourism, for instance, exemplifies a space where social media platforms wield significant power over decision-making. Travelers increasingly rely on platforms like Instagram, TripAdvisor, and TikTok to curate itineraries, driven by the immediacy of peer reviews and visual narratives. Such interactions underscore the duality of social media: while it democratizes information, it also centralizes influence in the hands of dominant content creators, thereby redefining traditional notions of authority and trust in consumer markets [3].

The intersectionality of social media and financial systems invites deeper examination of the mechanisms that govern this symbiosis. How does the immediacy of information shape investor sentiment? To what extent do platform algorithms, designed for engagement, affect the dissemination of critical financial news? These questions form the bedrock of contemporary digital finance research, as scholars aim to untangle the intricate threads linking social discourse to economic outcomes. By investigating these dynamics, this study situates itself at the confluence of technological innovation, economic inclusivity, and societal evolution, advancing our understanding of the digital society in an era of hyperconnectivity [4].

The omnipresence of social media in financial markets has redefined how information is disseminated, absorbed, and acted upon. Platforms like Twitter have emerged as powerful vectors for financial news and sentiment, functioning as conduits for both genuine insights and speculative discourse. The rapid diffusion of information on these platforms influences investor decisions, often creating a cascade effect where market trends and public sentiment become deeply intertwined. Research suggests that social media sentiment not only mirrors market sentiment but can also amplify mass media biases, producing a feedback loop that reinforces polarized perceptions among investors [5]. This intricate interplay raises profound questions about the reliability and authenticity of the information circulating in digital spaces, particularly as these narratives shape real-world financial outcomes.

The predictive power of social media sentiment further underscores its central role in financial ecosystems. Studies have demonstrated a strong correlation between sentiment analysis derived from platforms like Twitter and subsequent stock price movements, revealing the potential of these platforms as informal market indicators [6]. Unlike traditional financial analytics, which often rely on historical data and complex quantitative models, social media sentiment analysis provides a dynamic, real-time gauge of public opinion. This capability is particularly significant in volatile market conditions, where investor sentiment, shaped by a confluence of fear, hope, and speculation, can lead to abrupt market fluctuations [7], [8]. Consequently, sentiment analysis has become an indispensable tool for understanding and anticipating market behaviors in an era where digital platforms dominate the flow of information.

Beyond the realm of organic discourse, the strategic use of social media by firms adds another layer of complexity to the information landscape. Companies frequently curate their social media narratives to control the narrative surrounding their financial health. Empirical evidence reveals that firms often suppress negative financial disclosures on social media, opting instead to highlight positive developments, thereby crafting a favorable perception among stakeholders [9]. However, this selective transparency comes at a cost. When negative information does surface—either through leaks or alternate channels—it often triggers disproportionately adverse reactions among investors, compounding market instability and eroding trust in corporate communications [10]. Such practices illuminate the dual-edged nature of social media: a tool for transparency and engagement that can equally foster opacity and manipulation.

Social media activism represents another transformative dimension of this evolving landscape, particularly during bearish market phases. The rise of movements led by collective online voices, such as those witnessed during notable short squeezes, highlights how social media can mobilize sentiment into action. Platforms like Twitter serve as hubs for activist investors to galvanize public opinion, leading to surges in trading volumes and pronounced market responses [11]. Unlike traditional market dynamics, which rely on institutional players and structured mechanisms, these grassroots-driven movements reflect the democratization of financial influence. Yet, this democratization comes with volatility, as the convergence of sentiment and action in digital spheres often leads to market swings that challenge conventional regulatory frameworks and risk assessments.

Recent advancements in data science and machine learning have significantly enhanced the ability to analyze and interpret financial data, particularly in the context of social media. For instance, studies such as [12] and [13] have demonstrated the value of unsupervised learning techniques in extracting insights from financial discourse on social media. Similarly, research like [14] and [15] highlights the role of machine learning in predicting market trends and campaign effectiveness through structured data analysis. Moreover, time-series modeling approaches, as explored in [16] and [17] illustrate the power of temporal analysis for understanding financial trends and market behaviors. These methods align with studies such as [18] and [19] which emphasize the importance of unsupervised techniques like LDA in uncovering latent patterns in unstructured data. Finally, domain-specific analyses such as [20] and [21] provide further evidence of the versatility of clustering and predictive modeling in addressing complex data-driven questions across various sectors, including finance. These contributions collectively highlight the growing reliance on machine learning and unsupervised learning techniques for extracting actionable insights from high-dimensional datasets.

The exponential growth of social media as a source of financial information has created a dual-edged paradigm for both market participants and researchers. On one hand, platforms like Twitter serve as fertile ground for disseminating financial news and gauging investor sentiment. On the other, the sheer volume and velocity of unstructured data pose significant challenges for extracting actionable insights. While sentiment analysis has been extensively explored to predict stock price movements and volatility, a critical gap remains in leveraging topic modeling and time-series analysis to understand the broader context of financial trends and their evolution over time. These methods hold the potential to uncover nuanced discussions and recurrent themes that sentiment-focused approaches often overlook, thereby expanding the toolkit for financial data mining [6].

At the heart of this study lies an exploration of financial discourse on Twitter, a platform uniquely positioned as both a reflection and influencer of market dynamics. Using Latent Dirichlet Allocation (LDA), this research identifies underlying topics within financial tweets, aiming to illuminate the diversity and depth of discussions beyond binary sentiment classifications. Temporal patterns are examined through time-series analysis, revealing how the prominence of certain topics fluctuates and correlates with real-world financial events, such as earnings reports, policy changes, or economic crises. By integrating these techniques, the study seeks to bridge the gap between unstructured social media data and structured financial analysis, providing a holistic view of how online conversations shape and respond to market realities [9].

The significance of this research extends beyond academic inquiry; it offers tangible implications for practitioners and policymakers. Investors, particularly retail participants, often lack the tools to contextualize the financial information flooding their social media feeds. This study not only aids in identifying credible and recurring themes but also maps their temporal significance, enabling investors to make informed decisions rooted in comprehensive insights. For financial analysts and firms, understanding the interplay between social media discourse and market trends opens avenues for proactive engagement with digital narratives, crafting strategies that align with evolving public sentiment and emerging financial topics [10].

Moreover, the study contributes to a broader discourse on the role of digital platforms in democratizing financial knowledge while exposing vulnerabilities inherent in such systems. By analyzing the temporal and thematic structures of financial conversations, the research underscores the power of collective discourse in shaping market behavior. Simultaneously, it sheds light on potential pitfalls, such as misinformation and herd behavior, that may arise from the unregulated flow of financial narratives. In this way, the findings serve not only to inform investment strategies but also to catalyze discussions on regulatory frameworks and digital literacy in the financial sphere, positioning this research at the intersection of innovation, inclusivity, and ethical responsibility [2].

Literature Review

Overview of Social Media in Finance

The increasing integration of social media, particularly platforms like Twitter, into financial decision-making and market prediction underscores its transformative impact on the financial landscape. As a dynamic repository of real-time opinions and sentiments, Twitter allows investors, analysts, and policymakers to gauge public perceptions that often shape market behaviors. Unlike traditional financial indicators, social media provides immediacy, offering stakeholders the ability to react quickly to emerging trends. Its capacity to aggregate diverse viewpoints democratizes access to market-relevant information, making it an invaluable resource for modern financial analysis [22], [23].

A growing body of research illustrates the correlation between social media sentiment and market dynamics. Gao et al., for example, investigated the relationship between Twitter sentiment and stock trading characteristics in the cryptocurrency domain, particularly Bitcoin. Their findings highlighted that the volume of tweets mentioning a stock strongly correlates with the trading volume of that stock, signifying that social media activity can serve as a proxy for market interest and activity. This interplay between sentiment and trading behavior demonstrates Twitter's potential as a predictive tool for financial markets, extending beyond traditional data sources to inform investment strategies [22]. Similarly, Bukovina emphasized social media's role as an early economic indicator, capable of predicting broader economic outcomes such as stock performance and consumer behavior, cementing its utility in macroeconomic analysis [24].

The predictive power of Twitter sentiment is further elucidated by Ranco et al., who examined its relationship with stock price returns and volatility in retail markets. Their analysis revealed that qualitative aspects of sentiment—beyond mere tweet volume—hold significant predictive value for future stock prices. This finding underscores the importance of sentiment analysis in financial forecasting, highlighting how nuanced interpretations of social media data can surpass quantitative measures in explaining market fluctuations. Such insights underscore the need for sophisticated analytical tools to harness the depth of information embedded in digital narratives [23].

The application of advanced machine learning techniques has added a new dimension to the analysis of social media data in financial contexts. Recent studies have explored the efficacy of bi-directional Long Short-Term Memory (LSTM) networks with self-attention mechanisms for classifying Twitter sentiment. These approaches enhance the granularity and accuracy of sentiment classification, enabling the extraction of subtle emotional undertones and thematic relevance from financial discourse [25]. By integrating these innovations, researchers can refine predictions and uncover latent patterns in social media data, reinforcing its role as a cornerstone of contemporary financial analytics.

Financial Sentiment Analysis

Sentiment analysis has emerged as a cornerstone in the study of financial markets, leveraging the dynamic nature of social media to classify investor sentiment into categories such as bullish, bearish, or neutral. This approach has shown promise in predicting market movements, yet its application is often fraught with complexities. The varying success of sentiment analysis as a predictive tool reflects the intricate interplay between market conditions, investor behavior, and the quality of data derived from platforms like Twitter. These challenges underscore the importance of refining methodologies to enhance the reliability of sentiment-based market predictions [26].

A notable study by Gao et al. examines the sentiment of Bitcoin-related tweets, identifying bullish sentiment as a significant predictor of increased returns and volatility within short timeframes. Their findings demonstrate that financial sentiment, when measured accurately, can serve as a barometer for market behavior, particularly in speculative assets such as cryptocurrencies. Similarly, Checkley et al. establish a causal relationship between sentiment and stock market performance, showing that independent metrics for bullish and bearish sentiment can effectively forecast stock price directions, volatility, and trading volumes. These studies collectively highlight the potential of sentiment analysis to offer actionable insights into market trends while underscoring its dependence on robust data interpretation frameworks [22].

However, the predictive power of sentiment analysis is not universal, as evidenced by research in emerging markets. Messaoud et al. reveal that while investor sentiment positively correlates with market liquidity, its effectiveness as a predictive tool varies across different market conditions. This variability underscores the contextual limitations of sentiment analysis, suggesting that its utility may be contingent upon factors such as market maturity and regulatory environments. Furthermore, Frydman et al. delve into the psychological dimensions of sentiment, illustrating how investor expectations, shaped by market sentiment, influence their forecasts of stock returns. These findings reveal a nuanced relationship between sentiment and market behavior, emphasizing the multifaceted nature of financial decision-making [27], [28].

Beyond basic sentiment classification, Ji and Han advocate for a more granular approach, employing structural topic modeling to uncover thematic variations in investor sentiment. By analyzing millions of tweets, their study demonstrates that sentiment is heterogeneous, influenced by diverse finance-specific factors such as sector performance, macroeconomic trends, and geopolitical events. This thematic heterogeneity suggests that integrating sentiment analysis with advanced topic modeling techniques can enhance its predictive accuracy. The work of Banchit et al. further reinforces this notion, acknowledging the inherent challenges in quantifying sentiment but affirming its profound impact on market behavior. Together, these studies illuminate both the potential and the complexity of sentiment analysis as a tool for understanding financial markets [29], [30].

Topic Modeling in Social Media Analysis

Topic modeling, exemplified by Latent Dirichlet Allocation (LDA), has emerged as a transformative tool for uncovering hidden patterns and themes within vast datasets. By identifying latent topics within unstructured text data, LDA offers researchers a mechanism to distill complexity into comprehensible insights, bridging the gap between raw data and actionable knowledge. While its applications span diverse domains, from healthcare to marketing, its potential in analyzing financial Twitter data remains largely untapped. This oversight is significant, given the platform's role as a repository of real-time financial discourse, where discussions ranging from speculative chatter to strategic insights unfold continuously, shaping market narratives [31], [32].

The utility of LDA in social media analysis lies in its capacity to parse unstructured text into coherent themes, revealing the dynamics of public sentiment and concerns. Wang et al., in their examination of public discourse during the COVID-19 pandemic, showcased how LDA could identify dominant themes that reflected collective anxieties, such as health, policy, and economic stability. These findings underscore the versatility of topic modeling in navigating large-scale datasets to provide a snapshot of societal priorities. Applied to financial Twitter data, similar techniques could extract the underlying themes driving investor conversations, shedding light on how specific events or sentiments influence market behavior [33].

In the financial realm, integrating LDA with sentiment analysis could advance our understanding of the interplay between public discourse and market movements. Kudalkar posited that sentiment extraction techniques, while initially developed for general social media analysis, hold considerable promise for financial applications. This adaptability is particularly pertinent when analyzing Twitter data, where financial discussions often reflect a mixture of speculation, strategy, and reaction to market events. By employing LDA to uncover themes within these conversations, researchers could move beyond binary sentiment classifications to explore the nuanced narratives that drive market trends [34].

Despite its potential, the application of LDA in financial Twitter analysis remains nascent. Much of the existing literature has focused on sentiment classification, which, while valuable, offers only a partial understanding of public discourse. Jurek et al.'s work on lexicon-based sentiment analysis suggests that combining sentiment and topic modeling could significantly enhance the accuracy of interpreting public opinion, a notion particularly applicable to financial contexts. Extracting topics related to specific stocks or market conditions could provide a more granular understanding of investor sentiment, enabling financial analysts to anticipate market shifts with greater precision. However, such integration remains rare, leaving a gap in methodologies that this research seeks to address [35].

Time-Series Analysis of Social Media Data

Time-series analysis has emerged as a powerful tool for examining fluctuations in public opinion, particularly as captured on platforms like Twitter. However, its application in correlating social media activity with specific financial events remains underexplored, creating a fertile avenue for research. Social media platforms serve as real-time mirrors of public sentiment, but understanding their influence on market behavior necessitates a deeper exploration of temporal patterns and their alignment with key financial occurrences. This intersection offers the potential to bridge the gap between unstructured online discourse and structured financial outcomes [36].

A pivotal study by Cwynar et al. investigated the dynamics of social media as an informational asset within financial markets, revealing a temporal lag between sentiment expression and observable market reactions. Their findings suggest that institutional investors, often wary of the volatility inherent in social media-derived insights, tend to adopt a cautious approach. This latency reflects the need for corroborative evidence before integrating social media signals into decision-making, which in turn complicates direct correlations between social media activity and market events. Such nuances underscore the importance of refining time-series methodologies to capture delayed yet significant market responses effectively [36].

In contrast, Ji and Han emphasized the necessity of differentiating between transient financial news and persistent investment-related dialogues. By disentangling these overlapping narratives, their study illustrated how the heterogeneity of investor sentiment enriches the analytical depth of time-series models. This distinction is crucial for understanding how distinct financial events, such as earnings releases or regulatory announcements, influence public discourse and, subsequently, market movements. Such a nuanced approach extends the analytical potential of time-series analysis beyond surface-level correlations, paving the way for more sophisticated insights into the interplay between social media sentiment and financial phenomena [29].

Further elucidating this connection, Valle-Cruz et al. explored the impact of social media sentiment on stock market decisions during pandemics, with a comparative analysis of the H1N1 and COVID-19 crises. Their study demonstrated a pronounced alignment between Twitter sentiment and stock price fluctuations, revealing that social media activity often serves as an anticipatory indicator of market behavior. Similarly, Ranco et al. highlighted the dual significance of tweet volume and sentiment polarity in predicting stock returns, suggesting that integrating these metrics into time-series analysis could enhance the accuracy of financial forecasting models. Together, these studies illuminate the transformative potential of time-series analysis when applied to social media data, yet they also underscore the methodological complexities that must be addressed for robust financial applications [37].

Despite these advancements, the field remains nascent, with significant gaps in understanding the mechanisms through which social media sentiment influences market reactions. Existing research points to promising correlations, but systematic methodologies that incorporate the nuances of event-specific social media dynamics are scarce. Bridging this gap requires integrating time-series analysis with advanced sentiment extraction techniques, enabling researchers to capture the temporal and thematic dimensions of market responses to financial news with greater precision. Such efforts hold the potential to transform the predictive power of social media analysis in finance, offering both practical tools for investors and deeper theoretical insights into market behavior.

Gaps in Current Research

Despite extensive research on sentiment analysis in social media, particularly in financial contexts, a notable gap persists in the application of unsupervised learning techniques, such as Latent Dirichlet Allocation (LDA), to extract latent themes and correlate them with temporal patterns in financial data. This oversight limits the analytical scope of existing studies, as it constrains researchers to predefined sentiment categories without exploring the nuanced thematic dimensions embedded in social media discourse. Understanding these themes and their temporal evolution could provide richer insights into how online narratives shape financial markets [38], [39].

Existing studies on financial sentiment analysis predominantly focus on supervised learning approaches, leveraging labeled datasets to classify sentiments as positive, negative, or neutral. These methods, while effective in capturing immediate sentiment polarities, often fail to uncover deeper underlying patterns in the data. For instance, Mozetič et al. utilized machine learning algorithms to classify sentiments in financial tweets but did not extend their analysis to identify recurring themes or contextual drivers behind these sentiments. This limitation underscores the need for methodologies like LDA, which can explore unstructured data to reveal latent topics and offer a more comprehensive understanding of social media's impact on market dynamics [40].

The application of LDA in financial Twitter data has the potential to significantly advance the field by uncovering hidden themes that influence public sentiment over time. While studies have linked social media sentiment to stock price movements, they often fail to investigate the specific thematic content underlying these sentiments. For example, tweets about economic policies or earnings reports may evoke varying emotional responses, yet these themes remain unexplored in traditional sentiment analyses. Employing LDA could illuminate the recurring narratives that dominate discussions about specific stocks or market sectors, thereby enriching the context for understanding how thematic variations influence financial trends [41], [42].

Moreover, the lack of integration between topic modeling and sentiment analysis limits the ability to draw comprehensive conclusions about the role of social media discourse in financial markets. Existing research has primarily focused on either emotional dynamics or market performance, rarely bridging the two domains. For instance, studies on emotional sentiment in tweets related to stock indices highlight how public reactions can influence short-term market behavior, but they neglect to examine how these sentiments align with specific recurring themes over time. Integrating LDA with time-series analysis could address this gap, enabling researchers to correlate thematic trends with financial data and provide more robust models for predicting market movements. Such an approach would be particularly valuable during periods of heightened market volatility, offering investors and analysts actionable insights into the evolving nature of public sentiment and its financial implications [43], [44].

Method

The research method is a comprehensive and structured approach that involves meticulously designed steps to ensure a thorough and in-depth analysis of the subject matter at hand. This methodical process is crucial for obtaining reliable and valid results, as it systematically guides the researcher through various phases of investigation. The detailed outline of these steps is presented in Figure 1, which serves as a visual representation of the research journey.

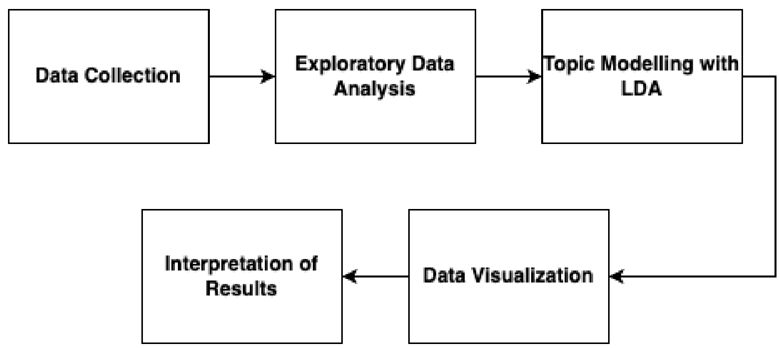

Figure 1 Research Method Flowchart

Data Collection

The foundation of this study lies in a meticulously curated dataset comprising finance-related tweets collected via the Twitter API. Social media platforms like Twitter serve as dynamic repositories for real-time sentiment and discourse, particularly in the realm of finance, where market participants express their perspectives with immediacy and granularity. To harness this potential, the dataset was compiled with a focus on tweets containing explicit financial references, such as stock symbols, company mentions, and market outlooks, thereby ensuring relevance and specificity to the study's objectives.

The dataset includes a total of 11,932 tweets, each labeled with one of three sentiment categories: bearish, bullish, or neutral. These labels are crucial for understanding the emotional and predictive dimensions of financial discourse, providing structured insights into how sentiments manifest in market-related conversations. The data was initially stored in two separate files, `sent_train.csv` and `sent_valid.csv`, each representing different subsets of the labeled tweets. Using Python's `pandas` library, the two datasets were systematically combined into a unified file, `sent_combined.csv`, to facilitate comprehensive analysis. This integration ensured uniformity in data structure while preserving the integrity of the original labels.

A preliminary inspection of the combined dataset revealed the depth and variety of the collected tweets. For example, the first five rows showcased financial statements ranging from company outlook revisions, such as Ally Financial, to sentiment-laden commentary on Deutsche Bank and Compass Point's stock evaluations. Each row comprised two essential columns: `text`, containing the tweet content, and `label`, indicating the corresponding sentiment. This structure provided a robust foundation for subsequent exploratory data analysis (EDA) and modeling efforts, aligning the data format with the requirements of advanced analytical techniques.

The decision to include labels was pivotal, as it allowed for seamless integration with machine learning workflows, particularly in the context of sentiment analysis and topic modeling. Furthermore, the inclusion of real-time discourse from diverse sources enriched the dataset, capturing a spectrum of market sentiments. The use of the Twitter API ensured that the dataset reflected authentic and contemporaneous market discussions, making it a valuable resource for exploring the intricate relationships between social media activity and financial trends. By consolidating this dataset, the study established a robust starting point for uncovering latent patterns and correlations in financial Twitter data, setting the stage for more nuanced analyses.

Exploratory Data Analysis (EDA)

EDA was performed to uncover patterns, visualize trends, and identify anomalies within the dataset. The primary objectives included examining tweet frequency over time, analyzing sentiment distribution, and preprocessing text data to prepare it for further analysis. These steps were crucial for understanding the dataset's structure and ensuring its readiness for advanced modeling techniques.

The bar chart in Figure 2 visualizes the distribution of sentiments across the dataset, categorized into Bearish, Bullish, and Neutral sentiments. Each bar represents the total count of tweets within its respective sentiment category, providing a clear comparison of sentiment proportions in the financial discussions. From the chart, it is evident that the Neutral sentiment dominates the dataset, comprising the majority of tweets with nearly 7,800 instances. This indicates that a substantial portion of the financial discourse on Twitter consists of objective or factual content that does not explicitly reflect strong positive (bullish) or negative (bearish) sentiment. These tweets likely include news, updates, or commentary that are informative but devoid of emotional bias.

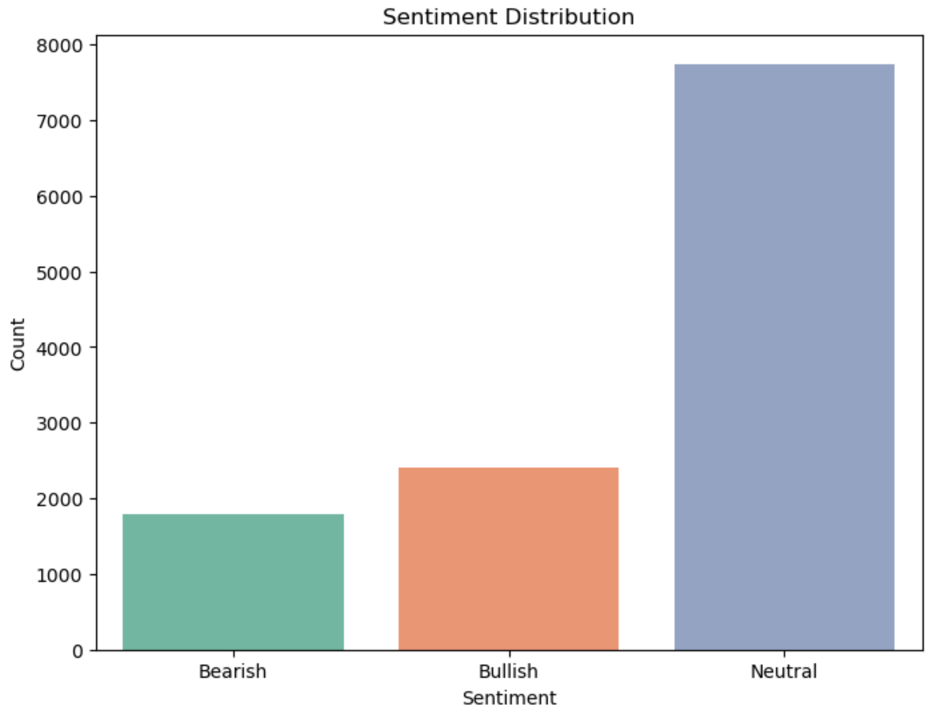

Figure 2 Sentiment Distribution

The Bullish sentiment, characterized by optimism or positive outlooks regarding market movements, appears in approximately 2,400 tweets. This suggests that a smaller segment of the dataset conveys confidence or positive predictions, such as earnings expectations, favorable stock performance, or growth forecasts. The Bearish sentiment, representing pessimism or negative outlooks, accounts for roughly 1,800 tweets. This reflects a subset of discussions expressing concerns, such as declining stock prices, poor earnings results, or broader economic challenges. Overall, the distribution highlights an imbalance in sentiment categories, with Neutral dominating over Bullish and Bearish sentiments. This trend may suggest that Twitter is primarily used as a platform for information sharing in financial contexts, while strong opinions (positive or negative) are expressed less frequently. Such an insight provides context for subsequent analyses, such as correlating sentiment proportions with financial events or market trends.

Preprocessing the textual data involved several steps, including tokenization, stopword removal, and normalization. Using the NLTK library, the `text` column was cleaned to remove non-alphanumeric characters, convert text to lowercase, and eliminate common stopwords such as "and" and "the." For example, the tweet “$ALLY - Ally Financial pulls outlook https://t.co...” was transformed into “ally ally financial pulls outlook,” stripping it of extraneous elements while retaining its semantic essence. This preprocessing step not only reduced noise but also standardized the data for subsequent analyses, such as sentiment classification and topic modeling. The resulting `processed_text` column provided a concise and structured representation of the tweets, ready for further exploration.

These EDA and preprocessing steps were instrumental in setting the stage for meaningful insights. By visualizing tweet patterns, understanding sentiment dynamics, and preparing the textual data, the analysis laid a robust foundation for uncovering latent themes and temporal correlations within financial Twitter discourse. This systematic approach ensures that the dataset is primed for advanced analytical methods, such as topic modeling and time-series analysis, to extract actionable insights from the financial conversations it encapsulates.

Tools and Algorithms

The methodological framework of this study relied on a suite of powerful Python libraries, carefully selected to address the diverse analytical needs of the research. Among these, Gensim played a central role in performing Latent Dirichlet Allocation (LDA) for topic modeling, allowing for the extraction of latent themes within the unstructured textual data. LDA's capacity to identify hidden patterns and organize them into coherent topics made it an indispensable tool for interpreting the complex narratives embedded in financial tweets. Gensim's implementation, known for its computational efficiency, provided an optimal balance between analytical depth and scalability, ensuring robust topic generation across the dataset.

Complementing Gensim, Matplotlib and Seaborn were utilized for data visualization, transforming raw numerical outputs into visually intuitive insights. These libraries enabled the creation of diverse plots, from simple line charts capturing temporal tweet frequencies to intricate count plots illustrating sentiment distributions. Such visualizations not only enhanced interpretability but also provided a foundational understanding of the dataset's structure and key characteristics. The seamless integration of these libraries within Python facilitated a dynamic interplay between raw data and graphical representation, ensuring clarity in the analysis and enabling the identification of patterns that might otherwise remain obscured.

For examining temporal trends, statsmodels was employed to conduct time-series analyses, a pivotal component of this study's exploration of financial discourse. This library's advanced functionalities, including its autoregressive integrated moving average (ARIMA) and seasonal decomposition methods, allowed for the dissection of tweet activity over time. By applying these models, the analysis uncovered periodic fluctuations and potential correlations between social media activity and significant financial events. Such temporal insights were critical in bridging the gap between online discourse and its real-world financial implications, demonstrating the library's value in capturing nuanced temporal patterns.

The integration of these tools underscored the study's commitment to leveraging cutting-edge technologies to address its research objectives. By combining Gensim's thematic exploration, Matplotlib and Seaborn's visual clarity, and statsmodels' temporal precision, the methodology achieved a harmonious balance between technical rigor and accessibility. This synergistic use of Python libraries not only enriched the analytical process but also ensured that the findings were grounded in both statistical robustness and interpretive depth, setting the stage for meaningful contributions to the discourse on financial sentiment and social media analytics.

Result and Discussion

Topic Modelling Results

The application of Latent Dirichlet Allocation (LDA) to the dataset revealed five distinct topics, each encapsulating a unique facet of financial discourse as expressed on Twitter. These topics, derived from the co-occurrence patterns of words in the tweets, provided a thematic overview of the prevailing financial narratives. Each topic was characterized by a set of high-probability terms, which, when interpreted collectively, highlighted their relevance to specific domains such as stock performance, economic policies, and market trends. This thematic extraction underscored the potential of LDA in distilling complex financial discussions into manageable and interpretable insights.

Topic 0, represented by terms such as "results," "reports," "stock," and "revenue," was predominantly associated with discussions of corporate earnings and financial reporting. These tweets often highlighted quarterly earnings reports, revenue beats, and analysts' price targets, reflecting the market's reaction to performance metrics. Such conversations are critical in financial markets, where earnings announcements frequently serve as catalysts for stock price movements, influencing both institutional and retail investor decisions. The prominence of terms like "EPS" and "Q3" further emphasized the specificity of these discussions to earnings seasons, reinforcing the thematic clarity of the topic.

Topic 1, characterized by words like "earnings," "dividend," "conference," and "presentation," pointed to discourse surrounding corporate announcements and investor relations events. The inclusion of terms like "transcript" and "presentation" suggested a focus on earnings calls and shareholder meetings, where companies provide insights into their strategies and future outlooks. This topic holds particular relevance for understanding how companies shape narratives to influence investor sentiment, especially during periods of heightened market scrutiny. Additionally, the presence of terms such as "energy" and "2019" indicated a temporal and sectoral dimension, reflecting discussions tied to specific industries or past financial cycles.

Topic 2 and Topic 3 revealed broader market-focused themes. Topic 2, with terms like "stock," "funds," "hedge," and "economy," highlighted conversations around investment strategies and macroeconomic influences, including the role of hedge funds and federal policies. In contrast, Topic 3 featured terms such as "marketscreener," "economy," "billion," and "oil," suggesting a focus on economic forecasts and large-scale market trends. The overlap of terms like "economy" and "markets" across these topics underscored the interconnected nature of financial discussions, where micro-level investment decisions are often contextualized within macroeconomic narratives. This duality illuminated how social media captures both granular and holistic perspectives on market dynamics.

Finally, Topic 4, encompassing words like "stock," "market," "china," and "trade," resonated with discussions of geopolitical and international trade issues. The inclusion of terms such as "update" and "deal" pointed to ongoing narratives surrounding trade agreements and their implications for global markets. These conversations are especially relevant in understanding how external factors, such as US-China trade relations, impact investor sentiment and market performance. By visualizing these topics through word clouds and topic-term matrices, the analysis provided a clear representation of the dominant themes and their linguistic structures, offering actionable insights into the multifaceted nature of financial discourse on social media.

Interpretation of Findings

Through the identification and interpretation of these topics, the results demonstrated the efficacy of LDA in unraveling the latent structure of financial discussions. These findings not only highlighted the diversity of themes present in social media but also reinforced the importance of contextualizing sentiment within broader narrative frameworks. Such insights have the potential to enhance market analysis by providing a deeper understanding of the drivers behind investor sentiment and market movements.

The identified topics provide a compelling lens through which the interplay between social media discourse and financial markets can be understood. Each topic captures a distinct aspect of market-related conversations, ranging from corporate earnings and reports (Topic 0) to geopolitical influences such as trade agreements (Topic 4). These themes directly correlate with key drivers of market movements, where earnings announcements and economic policies often catalyze fluctuations in stock prices. For instance, Topic 0's focus on terms like "results," "reports," and "revenue" underscores the role of quarterly performance updates in shaping investor sentiment and decision-making. Similarly, the prevalence of terms such as "economy" and "trade" in Topics 2 and 4 reflects broader macroeconomic and geopolitical considerations, highlighting the diverse influences on financial discourse.

Beyond thematic relevance, the volume of tweets and the trends in topic prevalence emerge as critical indicators of market sentiment and potential forecasting tools. Periods of heightened tweet activity often coincide with significant financial events, such as earnings calls or major policy announcements, as evidenced by spikes in discussions captured within Topics 1 and 3. These observations align with the hypothesis that social media activity serves as a real-time proxy for market sentiment, offering predictive value for financial forecasting. By tracking the evolution of these topics over time, analysts could identify emerging narratives that may impact market performance, thereby gaining a competitive edge in anticipating market trends.

Comparing these findings with existing research on financial sentiment analysis reveals both alignments and distinctions. Previous studies have predominantly focused on classifying sentiments into bullish, bearish, or neutral categories, often leveraging supervised learning methods. While sentiment classification provides valuable insights into the emotional tenor of market discourse, it frequently overlooks the underlying thematic structures that drive these sentiments. In contrast, this study's application of Latent Dirichlet Allocation (LDA) demonstrates the value of uncovering latent topics, offering a more nuanced understanding of financial conversations that transcends the limitations of binary sentiment labels.

The novelty of using LDA for topic extraction lies in its ability to contextualize financial discourse within broader narrative frameworks. Unlike sentiment analysis, which focuses on immediate emotional responses, topic modeling reveals the structural and thematic underpinnings of discussions, providing deeper insights into the factors influencing market movements. This approach not only complements existing methodologies but also enriches the analytical toolkit for financial researchers and practitioners. By bridging the gap between sentiment classification and thematic exploration, the findings underscore the importance of adopting holistic analytical frameworks to fully capture the complexity of financial social media data [41]. This innovation marks a significant step forward in understanding the dynamic relationship between online discourse and market behavior, with implications for both academic inquiry and practical financial forecasting.

Conclusion

This study illuminated the thematic landscape of financial discourse on Twitter by identifying key topics through Latent Dirichlet Allocation (LDA) and analyzing their correlations with tweet volumes and financial events. The results underscored the multifaceted nature of market-related conversations, with topics ranging from corporate earnings and macroeconomic policies to geopolitical dynamics. By observing shifts in topic prevalence over time, the analysis revealed how social media activity mirrors and responds to significant market events. These findings demonstrate the relevance of Twitter as a real-time barometer of investor sentiment and highlight the value of integrating social media analytics into financial market studies.

In contributing to the field, this research advances the application of unsupervised learning techniques in understanding financial trends. Unlike traditional sentiment analysis, which often reduces discussions to polarized categories, the use of LDA enabled a more nuanced exploration of the latent structures underpinning financial conversations. This approach not only complements existing methodologies but also provides a richer context for interpreting market movements. By bridging the gap between thematic exploration and temporal analysis, the study offers a framework for leveraging social media data to gain actionable insights, thereby enriching the analytical toolkit available to financial researchers and practitioners.

The findings of this study also point to several promising avenues for future research. Advanced topic modeling techniques, such as BERT-based models, could be explored to enhance the granularity and contextual understanding of financial discourse. These models, with their ability to capture semantic relationships and contextual nuances, hold the potential to surpass traditional methods in revealing the complexities of online financial narratives. Additionally, integrating these techniques with financial forecasting tools could lead to the development of more robust predictive models that account for both the thematic and temporal dimensions of market behavior.

Further research could also investigate how specific financial domains, such as stocks, commodities, or cryptocurrencies, influence social media activity and vice versa. Examining the interplay between these domains and social media trends may uncover domain-specific patterns that could refine investment strategies and market predictions. By extending the scope of analysis to include diverse asset classes and exploring their unique characteristics, future studies can deepen the understanding of how digital narratives shape and are shaped by the financial world. These directions not only promise to build upon the insights of this study but also contribute to the evolving discourse on the integration of technology, finance, and social media.

Declarations

Author Contributions

Conceptualization: D.A.D.; Methodology: T.B.K.; Software: D.A.D.; Validation:

D.A.D.; Formal Analysis: T.B.K.; Investigation: T.B.K.; Resources: T.B.K.; Data Curation: D.A.D.; Writing Original Draft Preparation: T.B.; Writing Review and Editing: T.B.K.; Visualization: D.A.D.; All authors have read and agreed to the published version of the manuscript.

Data Availability Statement

The data presented in this study are available on request from the corresponding author.

Funding

The authors received no financial support for the research, authorship, and/or publication of this article.

Institutional Review Board Statement

Not applicable.

Informed Consent Statement

Not applicable.

Declaration of Competing Interest

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

References

[1] H. Hasanudin, "The Role of Social Media in Influencing Investment Decisions in the Millennial Generation," *J. Multidisiplin Sahombu*, vol. 3, no. 01, pp. 124–130, 2023, doi: 10.58471/jms.v3i01.1967.

[2] C. He, A. Li, D. Li, and J. Yu, "Does Digital Inclusive Finance Mitigate the Negative Effect of Climate Variation on Rural Residents' Income Growth in China?," *Int. J. Environ. Res. Public. Health*, vol. 19, no. 14, p. 8280, 2022, doi: 10.3390/ijerph19148280.

[3] V. Gupta, "The Influencing Role of Social Media in the Consumer's Hotel Decision-Making Process," *Worldw. Hosp. Tour. Themes*, vol. 11, no. 4, pp. 378–391, 2019, doi: 10.1108/whatt-04-2019-0019.

[4] D. H. N. Barus, "How Social Media Marketing Affecting Brand Equity and Student Engagement in Higher Education Marketing," *Econ. Manag. Inf.*, pp. 1–11, 2023, doi: 10.58195/emi.v2i2.82.

[5] J. Ren, H. Dong, B. Padmanabhan, and J. V. Nickerson, "How Does Social Media Sentiment Impact Mass Media Sentiment? A Study of News in the Financial Markets," *J. Assoc. Inf. Sci. Technol.*, vol. 72, no. 9, pp. 1183–1197, 2021, doi: 10.1002/asi.24477.

[6] Q. Liu, W.-S. Lee, M. Huang, and Q. Wu, "Synergy Between Stock Prices and Investor Sentiment in Social Media," *Borsa Istanb. Rev.*, vol. 23, no. 1, pp. 76–92, 2023, doi: 10.1016/j.bir.2022.09.006.

[7] Q. Liu, "Data Selection and Collection for Constructing Investor Sentiment From Social Media," *Humanit. Soc. Sci. Commun.*, vol. 11, no. 1, 2024, doi: 10.1057/s41599-024-03316-7.

[8] L. M. Dunham and J. García, "Measuring the Effect of Investor Sentiment on Financial Distress," *Manag. Finance*, vol. 47, no. 12, pp. 1834–1852, 2021, doi: 10.1108/mf-02-2021-0056.

[9] M. J. Jung, J. P. Naughton, A. Tahoun, and C. Wang, "Do Firms Strategically Disseminate? Evidence From Corporate Use of Social Media," *Account. Rev.*, vol. 93, no. 4, pp. 225–252, 2017, doi: 10.2308/accr-51906.

[10] W. Elliott, S. M. Grant, and F. D. Hodge, “Negative News and Investor Trust: The Role of $Firm and #CEO Twitter Use,” *J. Account. Res.*, vol. 56, no. 5, pp. 1483–1519, 2018, doi: 10.1111/1475-679x.12217.
[11] P. Carrasco and G. Michelon, “The Power of Stakeholders’ Voice: The Effects of Social Media Activism on Stock Markets,” *Bus. Strategy Environ.*, vol. 26, no. 6, pp. 855–872, 2017, doi: 10.1002/bse.1973.
[12] T. Wahyuningsih and S. C. Chen, “Analyzing Sentiment Trends and Patterns in Bitcoin-Related Tweets Using TF-IDF Vectorization and K-Means Clustering,” *J. Curr. Res. Blockchain*, vol. 1, no. 1, Art. no. 1, Jun. 2024, doi: 10.47738/jcrb.v1i1.11.
[13] T. Hariguna and A. S. M. Al-Rawahna, “Unsupervised Anomaly Detection in Digital Currency Trading: A Clustering and Density-Based Approach Using Bitcoin Data,” *J. Curr. Res. Blockchain*, vol. 1, no. 1, Art. no. 1, Jun. 2024, doi: 10.47738/jcrb.v1i1.12.
[14] N. M. Nhat, “Applied Random Forest Algorithm for News and Article Features on The Stock Price Movement: An Empirical Study of The Banking Sector in Vietnam,” *J. Appl. Data Sci.*, vol. 5, no. 3, pp. 1311–1324, Sep. 2024, doi: 10.47738/jads.v5i3.338.
[15] B. Berlilana, A. M. Wahid, D. Fortuna, A. N. A. Saputra, and G. Bagaskoro, “Exploring the Impact of Discount Strategies on Consumer Ratings: An Analytical Study of Amazon Product Reviews,” *J. Appl. Data Sci.*, vol. 5, no. 1, Art. no. 1, Jan. 2024, doi: 10.47738/jads.v5i1.163.
[16] Berlilana and A. M. Wahid, “Time Series Analysis of Bitcoin Prices Using ARIMA and LSTM for Trend Prediction,” *J. Digit. Mark. Digit. Curr.*, vol. 1, no. 1, Art. no. 1, May 2024, doi: 10.47738/jdmdc.v1i1.1.
[17] S. A. Ghaffar and W. C. Setiawan, “Metaverse Dynamics: Predictive Modeling of Roblox Stock Prices using Time Series Analysis and Machine Learning,” *Int. J. Res. Metaverese*, vol. 1, no. 1, Art. no. 1, Jun. 2024, doi: 10.47738/ijrm.v1i1.6.
[18] S. Yadav and A. R. Hananto, “Comprehensive Analysis of Twitter Conversations Provides Insights into Dynamic Metaverse Discourse Trends,” *Int. J. Res. Metaverese*, vol. 1, no. 1, Art. no. 1, Jun. 2024, doi: 10.47738/ijrm.v1i1.2.
[19] A. Paramita and A. Ramadhan, “An Unsupervised Learning and EDA Approach for Specialized High School Admissions,” *J. Appl. Data Sci.*, no. Query date: 2024-10-12 10:59:38, 2024, [Online]. Available: http://www.bright-journal.org/Journal/index.php/JADS/article/view/178
[20] Y. Durachman and A. W. B. A. Rahman, “Clustering Student Behavioral Patterns: A Data Mining Approach Using K-Means for Analyzing Study Hours, Attendance, and Tutoring Sessions in Educational Achievement,” *Artif. Intell. Learn.*, vol. 1, no. 1, Art. no. 1, Nov. 2024.
[21] T. Sangsawang, “Predicting Student Achievement Using Socioeconomic and School-Level Factors,” *Artif. Intell. Learn.*, vol. 1, no. 1, Art. no. 1, Nov. 2024.
[22] X. Gao, W. Huang, and W. Hua, “Financial Twitter Sentiment on Bitcoin Return and High-Frequency Volatility,” *Virtual Econ.*, vol. 4, no. 1, pp. 7–18, 2021, doi: 10.34021/ve.2021.04.01(1).
[23] G. Ranco, D. Aleksovski, G. Caldarelli, M. Grčar, and I. Mozetič, “The Effects of Twitter Sentiment on Stock Price Returns,” *Plos One*, vol. 10, no. 9, p. e0138441, 2015, doi: 10.1371/journal.pone.0138441.
[24] J. Bukovina, “Social Media Big Data and Capital Markets—An Overview,” *J. Behav. Exp. Finance*, vol. 11, pp. 18–26, 2016, doi: 10.1016/j.jbef.2016.06.002.
[25] “A Prediction of Twitter Sentiment Class Using Bi-Directional Long Short Term Memory With Self-Attention Layer Mechanism,” *Int. J. Adv. Trends Comput. Sci. Eng.*, vol. 9, no. 5, pp. 6892–6899, 2020, doi: 10.30534/ijatcse/2020/04952020.
[26] M. Checkley, D. A. Hign, and H. Alles, “The Hasty Wisdom of the Mob: How Market Sentiment Predicts Stock Market Behavior,” *Expert Syst. Appl.*, vol. 77, pp. 256–263, 2017, doi: 10.1016/j.eswa.2017.01.029.
[27] D. Messaoud, A. B. Amar, and Y. Boujelbène, “Investor Sentiment and Liquidity in Emerging Stock Markets,” *J. Econ. Adm. Sci.*, vol. 39, no. 4, pp. 867–891, 2021,

doi: 10.1108/jeas-11-2020-0198.
[28] R. Frydman, N. Mangee, and J. Stillwagon, “How Market Sentiment Drives Forecasts of Stock Returns,” *J. Behav. Finance*, vol. 22, no. 4, pp. 351–367, 2020, doi: 10.1080/15427560.2020.1774769.
[29] R. Ji and Q. Han, “Understanding Heterogeneity of Investor Sentiment on Social Media: A Structural Topic Modeling Approach,” *Front. Artif. Intell.*, vol. 5, 2022, doi: 10.3389/frai.2022.884699.
[30] A. Banchit, S. Abidin, S. Lim, and F. Morni, “Investor Sentiment, Portfolio Returns, and Macroeconomic Variables,” *J. Risk Financ. Manag.*, vol. 13, no. 11, p. 259, 2020, doi: 10.3390/jrfm13110259.
[31] T. Wang, L. Ke, K.-P. Chow, and Q. Zhu, “COVID-19 Sensing: Negative Sentiment Analysis on Social Media in China via BERT Model,” *Ieee Access*, vol. 8, pp. 138162–138169, 2020, doi: 10.1109/access.2020.3012595.
[32] D. Kudalkar, “Social Media Sentiment Analysis,” *Int. J. Res. Appl. Sci. Eng. Technol.*, vol. 11, no. 10, pp. 1696–1708, 2023, doi: 10.22214/ijraset.2023.56290.
[33] C. Á. Iglesias and A. Moreno, “Sentiment Analysis for Social Media,” *Appl. Sci.*, vol. 9, no. 23, p. 5037, 2019, doi: 10.3390/app9235037.
[34] A. Jurek, M. Mulvenna, and Y. Bi, “Improved Lexicon-Based Sentiment Analysis for Social Media Analytics,” *Secur. Inform.*, vol. 4, no. 1, 2015, doi: 10.1186/s13388-015-0024-x.
[35] “Sentiment Classification F Rom Social Media F or Stock Prediction W Ith Data Mining,” *Int. J. Innov. Technol. Explor. Eng.*, vol. 9, no. 5, pp. 155–161, 2020, doi: 10.35940/ijitee.e2226.039520.
[36] A. Cwynar, W. Cwynar, R. Pater, and K. Filipek, “Social Media as an Information Source in Finance: Evidence From the Community of Financial Market Professionals in Poland,” *Ijdar*, pp. 29–58, 2019, doi: 10.4192/1577-8517-v19_2.
[37] D. Valle-Cruz, V. Fernandez-Cortez, A. López-Chau, and R. Sandoval-Almazán, “Does Twitter Affect Stock Market Decisions? Financial Sentiment Analysis During Pandemics: A Comparative Study of the H1N1 and the COVID-19 Periods,” *Cogn. Comput.*, vol. 14, no. 1, pp. 372–387, 2021, doi: 10.1007/s12559-021-09819-8.
[38] I. Mozetič, M. Grčar, and J. Smailović, “Multilingual Twitter Sentiment Classification: The Role of Human Annotators,” *Plos One*, vol. 11, no. 5, p. e0155036, 2016, doi: 10.1371/journal.pone.0155036.
[39] Q. Li, Z. Wang, P. Li, L. Liu, Q. Gong, and Y. Chen, “The Effect of News and Public Mood on Stock Movements,” *Inf. Sci.*, vol. 278, pp. 826–840, 2014, doi: 10.1016/j.ins.2014.03.096.
[40] I. M. Haile and Y. Qu, “Mitigating Risk in Financial Industry by Analyzing Social-Media With Machine Learning Technology,” *Eur. J. Electr. Eng. Comput. Sci.*, vol. 6, no. 2, pp. 33–37, 2022, doi: 10.24018/ejece.2022.6.2.428.
[41] A. M. Priyatno and F. I. Firmananda, “N-Gram Feature for Comparison of Machine Learning Methods on Sentiment in Financial News Headlines,” *Riggs J. Artif. Intell. Digit. Bus.*, vol. 1, no. 1, pp. 01–06, 2022, doi: 10.31004/riggs.v1i1.4.
[42] F. H. Khan, U. Qamar, and S. Bashir, “Senti-CS: Building a Lexical Resource for Sentiment Analysis Using Subjective Feature Selection and Normalized Chi-Square-based Feature Weight Generation,” *Expert Syst.*, vol. 33, no. 5, pp. 489–500, 2016, doi: 10.1111/exsy.12161.
[43] A. M. Mostafa, “Enhanced Sentiment Analysis Algorithms for Multi-Weight Polarity Selection on Twitter Dataset,” *Intell. Autom. Soft Comput.*, vol. 35, no. 1, pp. 1015–1034, 2023, doi: 10.32604/iasc.2023.028041.
[44] E. Kotelnikov and E. Mitiagina, “Usage of Text Sentiment Analysis Models and Methods in the Sociological Studies,” *Destech Trans. Comput. Sci. Eng.*, no. smce, 2017, doi: 10.12783/dtcse/smce2017/12410.