Sentiments and Time Series Patterns for Improved Stock Market Predictions: A Comprehensive Study

1st Vishnu Shukla, 2nd Vedant Padole

[1,2]Student

[1]Computer Science Engineering,

[1]Shri Ramdeobaba College of Engineering and Management, Nagpur,India

Abstract - This research paper explores the intricate interplay between sentiment analysis and time series analysis in the realm of financial markets. Focusing on the dynamic relationship between information dissemination and stock market trends, we employ advanced sentiment analysis techniques, specifically leveraging the VADER model to extract sentiments from tweets and news headlines. Simultaneously, we harness the power of Long Short-Term Memory (LSTM) networks for time series analysis, aiming to discern temporal patterns and forecast future stock prices. The convergence of sentiment and time series analyses presents a holistic framework to unravel the complexities inherent in financial markets. Through a meticulous examination of these analytical approaches, we seek to shed light on how real-time sentiments, expressed on platforms such as Twitter and through traditional news outlets, intricately interact with temporal dynamics, influencing stock market movements. Our findings contribute a nuanced understanding of the symbiotic relationship between sentiment and time series analyses, offering valuable insights for market participants, analysts, and decision-makers navigating the multifaceted landscape of stock trading and investment.

Index Terms - sentiment analysis, time series analysis, financial markets, VADER model, Long Short-Term Memory (LSTM)

I. INTRODUCTION

The intricate dance between financial markets and information dissemination has long been a subject of profound interest and scrutiny. In the era of digital communication, the role of social media and traditional news sources in shaping investor sentiment and influencing stock market trends has gained unprecedented prominence. This research endeavors to unravel the symbiotic relationship between sentiment analysis and time series analysis, particularly in the context of forecasting stock market movements. By leveraging advanced sentiment analysis techniques, specifically the VADER model, we delve into the sentiments encapsulated within tweets and news headlines. Simultaneously, we employ sophisticated time series analysis methodologies, including Long Short-Term Memory (LSTM) networks, to discern temporal patterns and forecast future stock prices. This convergence of sentiment and time series analyses promises to offer a nuanced understanding of how sentiments expressed in real-time on platforms like Twitter and through traditional news outlets interplay with the temporal dynamics of financial markets. Through meticulous examination and integration of these two analytical approaches, this research seeks to contribute valuable insights for market participants, analysts, and decision-makers navigating the intricate landscape of stock trading and investment.

II. LITERATURE SURVEY

The study conducted by Ameen Abdullah Qaid Aqlan [1] on sentiment analysis serves as a cornerstone in comprehending the fundamental concepts, diverse techniques, and overarching challenges inherent in the realm of sentiment analysis. Aklan's research contributes not only theoretical frameworks but also practical implications, providing a nuanced perspective that is invaluable for those delving into sentiment analysis methodologies.

In a parallel vein, Kulkarni's [2] survey on sentiment analysis methods, applications, and challenges emerges as a key resource, particularly shedding light on the intricacies of sentiment analysis within the dynamic context of market dynamics. This survey not only outlines the various methodologies employed but also delves into the diverse applications of sentiment analysis across different industries, with a specific focus on its pivotal role in understanding market sentiments.

Fava's [3] work on time series analysis is a testament to the depth of knowledge in algorithmic values, offering a sophisticated exploration of the intricacies involved. Fava's insights into time series data bring forth a comprehensive understanding of algorithmic complexities, proving crucial for those seeking a profound comprehension of time-dependent data structures. Bose's contribution, "Introduction to Time Series Analysis and Its Applications," enriches our understanding of algorithmic applications, providing a holistic view of its diverse applications across various fields. Bose's [4] work not only introduces the foundational principles of time series analysis but also highlights its practical applications, making it an essential reference for those exploring the algorithmic aspects of this analytical domain.

Together, these seminal works provide a multifaceted foundation for researchers, practitioners, and enthusiasts in sentiment analysis and time series analysis, offering a wealth of knowledge that spans theoretical frameworks, practical applications, and the nuanced challenges associated with these dynamic fields of study.

III. APPROACH

Our primary objective was to employ sentiment analysis on selected stocks and analyse the outcomes. The following steps outline our approach in conducting this research.

(1) Sentiment Analysis

To conduct sentiment analysis on the chosen stock, it was imperative to gather relevant news articles and tweets associated with the stock. The initial phase involved preparing a dataset for further processing. For web scraping of data from business websites containing current trends related to the specified stock, we implemented Selenium. After setting up Selenium for web scraping from "fin viz" (financial visualisation), we successfully collected data encompassing the latest trends and news articles related to the selected stock. In our data collection process, we employed Beautiful Soup for parsing on the website, by utilizing Selenium and Beautiful Soup in tandem, we were able to systematically scrape recent headlines related to the selected stock from the Business Insider website.

(2) Dataset Description

Fig 1 pie chart visualizes the dataset's composition, illustrating the proportional distribution of tweets and news articles across various stocks.

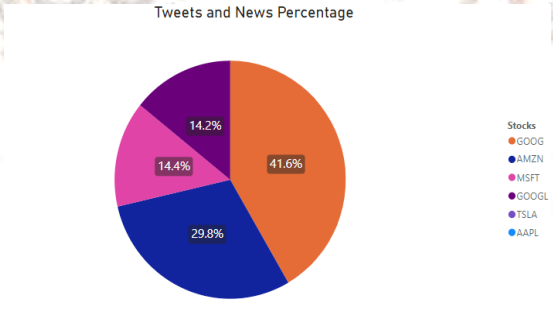

Figure 1 Tweets and News Percentage

(3) VADER (Valence Aware Dictionary and Sentiment Reasoner)

For calculating sentiments, we used VADER (Valence Aware Dictionary and Sentiment Reasoner) which is designed for processing and analysing textual data, with a particular focus on social media content. The primary strength of VADER lies in its pre-built sentiment lexicon, a dictionary containing words that are pre-assigned with sentiment scores. Each word is associated with a polarity score (positive, negative, or neutral) and an intensity score [5].

The main to apply VADER is because it analyses a piece of text, it calculates the overall sentiment intensity by considering both lexical and grammatical rules. This compound score represents the normalized, weighted composite sentiment of the entire text. Due to this, the dataset containing tweets and News headlines was easily able to train on VADER because of its adept handling of social media sentiment capture. VADER provided 67.3% positive marking whereas it marked 32.7% of the content as negative.

(4) Time-Series Analysis

The initial training of the sentiment analysis model allows us to gauge the polarity of sentiments from news articles and tweets, particularly focusing on the impact of social media dynamics. Building upon this sentiment foundation, we extend our analysis to the temporal dimension through time series decomposition and correlation studies. By isolating trends, seasonality, and residual components in historical stock prices, we uncover patterns that may be influenced by sentiment changes. Through meticulous feature engineering, including lagged sentiment scores and other relevant variables, we construct a comprehensive dataset. Leveraging this enriched dataset, we employ a selection of time series forecasting models, techniques LSTM.

(4) LSTM

Long Short-Term Memory (LSTM) networks stand as a pivotal technology in the realm of time series analysis. Leveraging a specialized architecture designed for capturing temporal dependencies, LSTMs excel in modelling sequential data, making them particularly adept at forecasting and understanding complex patterns within time series datasets. Their unique ability to handle long-term dependencies sets them apart from traditional recurrent neural networks, enabling them to memorize and utilize information over extended periods. LSTMs automatically extract relevant features from input time series data, allowing them to discern intricate temporal relationships and navigate non-linear patterns.

Moreover, their adaptability to variable time steps in irregularly sampled datasets enhances their versatility in real-world scenarios. Training an LSTM involves optimizing its parameters through techniques such as backpropagation through time, enabling the network to learn and generalize from historical data. Once trained, LSTMs prove invaluable for forecasting future values in time series, demonstrating their applicability across domains like stock price prediction, energy consumption forecasting, and weather prediction. In essence, Long Short-Term Memory networks emerge as a powerful and versatile tool, revolutionizing the landscape of time series analysis through their capacity to comprehend and predict nuanced temporal dynamics.

IV. RESULT

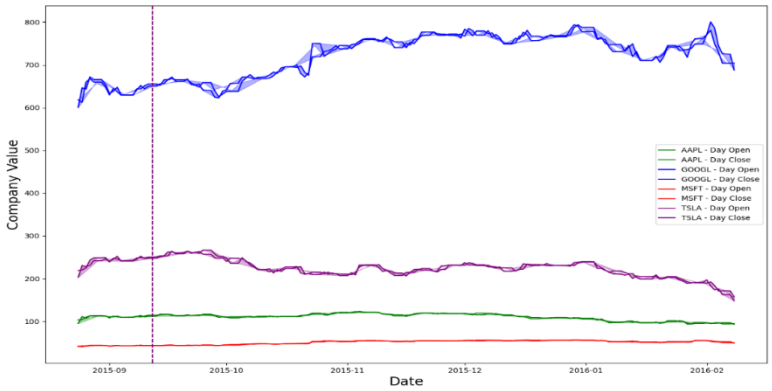

Fig 2: Sentiment Score vs Tweets

This figure presents a clear depiction of the impact of sentiment scores derived from tweets on the stock prices of two major companies, Microsoft and Tesla. The chart reveals a discernible correlation between the sentiment expressed on Twitter and the subsequent movement of stock prices. Specifically, it showcases the influence of positive or negative sentiment on the fluctuations in share prices for these two companies.

Notably, the remaining three stocks, Apple, Amazon, and Google, exhibit minimal volatility in response to tweets or news. The chart effectively illustrates that these stocks are less susceptible to the sentiments expressed on Twitter compared to Microsoft and Tesla.

V. CONCLUSIONS

It can be concluded that tweets and news headlines play a crucial and influential role in shaping stock market trends. The impact of social media, particularly Twitter, and traditional news sources on financial markets has become increasingly evident. The sentiments expressed in tweets and news headlines have the potential to sway investor perceptions, influencing their decisions to buy or sell stocks. The immediacy and widespread accessibility of social media platforms amplify the speed at which information is disseminated, allowing sentiments expressed in tweets to have swift and tangible effects on stock prices. Similarly, news headlines, often reflecting expert analyses and market trends, contribute to the broader narrative that shapes investor sentiment.The observed correlation between sentiment in tweets, news headlines, and subsequent stock market movements underscores the interconnected nature of financial markets and information dissemination. A

VI. REFERENCES

[1] Aqlan, Ameen & Bairam, Dr. Manjula & Naik, R Lakshman. (2019). A Study of Sentiment Analysis: Concepts, Techniques, and Challenges. 10.1007/978-981-13-6459-4_16..

[2] Wankhade, M., Rao, A.C.S. & Kulkarni, C. A survey on sentiment analysis methods, applications, and challenges. Artif Intell Rev 55, 5731–5780 (2022). https://doi.org/10.1007/s10462-022-10144-1

[3] Velicer, Wayne & Fava, Joseph. (2003). Time Series Analysis. 10.1002/0471264385.wei0223.

[4] Jose, Jonath. (2022). INTRODUCTION TO TIME SERIES ANALYSIS AND ITS APPLICATIONS..

[5] Hutto, C.J. & Gilbert, Eric. (2015). VADER: A Parsimonious Rule-based Model for Sentiment Analysis of Social Media Text. Proceedings of the 8th International Conference on Weblogs and Social Media, ICWSM 2014.