"Leveraging NLP in Financial Markets: Predicting Stock Price Trend using Sentiment Analysis of News Articles"

A dissertation report submitted in partial fulfillment of requirements for Bachelorsof Business Administration (Analytics and Big Data)

Under the guidance of

UPES
UNIVERSITY OF TOMORROW

Internal Guide:

Dr. Inder Singh

Associate Professor- Operations and

Technology Management Cluster

School of Business, UPES

Submitted by:

Name: Vilakshan Dhasmana

SAP Id: 500095789

Enrolment no.: R193221040

BBA Analytics and Big Data, 3rd year

School of Business, UPES

STUDENT DECLARATION

I, hereby declare that this submission is my work and that, to the best of my knowledge and belief, it contains no material previously published or written by another person nor material that has been accepted for the award of any other degree or diploma of the universityor other institute of higher learning, except where due acknowledgment has been made in the text.

Vilakshan Dhasmana

SAP Id: 500095789
Enrollment no.: R1932201040
BBA Analytics and Big Data, 3rd year

School of Business, UPES

ACKNOWLEDGMENT

At the very beginning, I would like to express my deepest gratitude to the almighty for giving me the strength and the composure to complete my Synopsis and prepare this report within the scheduled time.

During the period of my Synopsis Report development, I have received generous help from many quarters, which I like to put on record here with deep gratitude and great pleasure.

First and foremost, I am grateful to my mentor, Dr. Inder Singh. He allowed me to encroach upon his precious time freely right from the very beginning of this work until the completion of my work.
His guidance, encouragement, and suggestions provided me with the necessary insight into the problem and paved the way for the meaningful ending of the work in a short duration.
I have no hesitation in saying that, without his constant support and valuable advice from time to time, I would probably fail to complete the work appropriately.

I would like to express my special thanks of gratitude to my Course Coordinator Dr. Ranjeet Kumar Singh, who gave me the golden opportunity to do this wonderful project and learn so many new things.

Thank you all!

UNIVERSITY MENTOR'S CERTIFICATE

This is to certify that the dissertation report entitled **Leveraging NLP in Financial Markets: Predicting Stock Price Trend using Sentiment Analysis of News Articles** submitted by Vilakshan Dhasmana to UPES for partial fulfillment of requirements for Bachelor of Business Administration in Analytics and Big Data is a bonafide record carried out by him under my supervision and guidance. To the best of my knowledge, the content of the report, in full or parts has not been submitted to any other Institute or University for the award of any other degree.

Dr. Inder Singh
Associate Professor- Operations and Technology Management Cluster
School of Business, UPES

ABSTRACT

Financial markets are influenced by various factors, and sentiment plays a crucial role in shaping investor decisions and stock prices. This study explores the link between sentiment expressed in financial news articles and subsequent stock price changes. Using natural language processing (NLP) techniques, specifically the TextBlob library in Python, we analyze sentiment polarity and intensity across a comprehensive dataset of financial news. Our goal is to uncover patterns and correlations that benefit investors and analysts. By bridging traditional financial analysis with text analytics, our work contributes to both academia and practical strategies used by financial professionals. We collect diverse news articles from Nifty 50 companies, ensuring broad market sentiment coverage. After data cleaning and preprocessing, we assign sentiment polarity scores to each article. Comparing these scores with historical stock prices, we identify correlations between sentiment and subsequent stock price movements. These findings inform predictive models for forecasting stock prices and predicting directional movements. We create two models: a regression model for numerical stock price forecasts and a classification model to predict price direction (rise or fall). Leveraging machine learning techniques like linear regression and logistic regression, we incorporate sentiment features from news articles. Our results reveal subtle connections between sentiment polarity and stock performance, offering potential predictors of future price changes. In summary, this study advances sentiment analysis in financial markets. By combining NLP and machine learning, we provide fresh insights into investor sentiment and stock prices, benefiting investors, analysts, and policymakers.

TABLE OF CONTENTS

CH. I. INTRODUCTION

Financial markets serve as the central hubs of economic activity, where individuals, businesses, and governments engage in the buying and selling of various financial assets. These markets are instrumental in allocating capital, setting prices, and facilitating economic transactions on a global scale. Grasping the fundamentals of financial markets is essential for comprehending how money flows through economies and impacts individuals worldwide.

At their core, financial markets function as platforms where buyers and sellers meet to trade financial assets, spanning from stocks and bonds to currencies, commodities, and derivatives. Each financial market operates with its own set of rules, participants, and distinctive characteristics.

One of the primary purposes of financial markets is to provide liquidity, ensuring that investors can quickly buy or sell assets at fair prices. Liquidity is crucial as it guarantees a readily available market for assets, enabling investors to convert them into cash with ease.

Price discovery is another critical function of financial markets. Prices are determined by the interplay of supply and demand, with assets experiencing price increases when demand surpasses supply and decreases when the opposite occurs. This continuous interaction between buyers and sellers establishes market prices that reflect the collective sentiments and expectations of market participants.

Financial markets also serve as platforms for risk transfer and diversification. Investors can mitigate risk by spreading their investments across various assets and markets. For instance, diversifying a portfolio across stocks, bonds, and real estate can help reduce the impact of any single asset's performance on overall investment outcomes.

Capital formation is another vital role played by financial markets. Companies and governments often raise capital by issuing securities in primary markets, with the proceeds used to finance business expansions, infrastructure projects, or public initiatives, thereby fueling economic growth.

Financial markets are typically divided into primary and secondary markets. **Primary markets** facilitate the issuance and initial sale of new securities, while secondary markets enable the trading of existing securities among investors. Examples of primary markets include initial public offerings (IPOs) and bond auctions, while **secondary markets** encompass exchanges like the New York Stock Exchange (NYSE) and NASDAQ.

Participants in financial markets comprise individual investors, institutional investors such as mutual funds and pension funds, corporations, banks, and government entities. Each participant brings its own investment goals, strategies, and risk tolerances, shaping the dynamics of the market.

Numerous factors influence the functioning of financial markets, including economic indicators, government policies, geopolitical events, and investor sentiment. For instance, positive economic reports can boost investor confidence and drive stock prices higher, while political instability may lead to market uncertainty and volatility.

Technological advancements have revolutionized financial markets, with electronic trading platforms, algorithmic trading, and high-frequency trading enhancing efficiency and accessibility. These innovations have democratized market access, enabling a broader range of participants to engage in trading activities.

Despite their significance, financial markets carry inherent risks such as market volatility, asset price fluctuations, and the potential for investment losses. It is crucial for investors to assess their risk tolerance and investment objectives carefully before participating in financial markets.

In summary, financial markets play a pivotal role in driving economic activity, facilitating investment, and managing risk. Understanding the fundamentals of financial markets is essential for investors, policymakers, and anyone seeking to navigate the complexities of the global economy.

Let’s Dive into Secondary Markets:

Secondary markets are integral parts of the financial system, offering investors platforms to trade existing securities subsequent to their initial issuance in primary markets. These markets facilitate liquidity and price determination, enabling investors to engage in the buying and selling of securities among themselves. There exist various types of secondary markets, each serving distinct purposes and catering to different kinds of securities:

1. **Stock Exchanges**: Stock exchanges represent the most widely recognized form of secondary markets, where shares of publicly traded companies are exchanged. These exchanges provide centralized venues for stock trading, ensuring transparency, liquidity, and fair price discovery. Notable stock exchanges include the New York Stock Exchange (NYSE), NASDAQ, London Stock Exchange (LSE), and Tokyo Stock Exchange (TSE).

2. **Over-the-Counter (OTC) Markets**: OTC markets function as decentralized networks where securities are traded directly between buyers and sellers without the oversight of a central exchange. They are commonly utilized for trading stocks of smaller companies, bonds, and derivatives. Unlike stock exchanges, OTC markets offer reduced transparency and may entail higher risks due to the absence of centralized regulation. Nonetheless, they afford flexibility and accessibility to a broad array of securities.

3. **Bond Markets**: Bond markets serve as secondary markets where debt securities, such as government bonds, corporate bonds, and municipal bonds, are transacted. These markets facilitate borrowing and lending activities, enabling investors to trade fixed-income securities with varying maturities, interest rates, and credit ratings. Bond markets may operate as either exchange-traded or over-the-counter, contingent upon the type of bonds being traded and the market infrastructure.

4. **Foreign Exchange (Forex) Markets**: Forex markets act as secondary markets where currencies are traded against one another. Operational around the clock, five days a week, these markets allow participants to exchange currencies for diverse purposes, including international trade, investment, and speculation. Forex markets operate in a decentralized manner and are highly liquid, boasting trading volumes surpassing trillions of dollars daily.

5. **Commodities Markets**: Commodities markets facilitate the trading of physical goods, including agricultural products, energy resources, metals, and precious stones. These markets enable producers, consumers, and investors to buy and sell commodities for the purpose of managing price risks and speculating on future price movements. Commodities markets may be organized through exchanges such as the Chicago Mercantile Exchange (CME) and the London Metal Exchange (LME), or conducted over-the-counter.

6. **Derivatives Markets**: Derivatives markets function as secondary markets where financial instruments derived from underlying assets are traded. These instruments, termed derivatives, encompass options, futures, forwards, and swaps, providing investors with avenues to hedge against risks, speculate on price changes, and gain exposure to various asset classes. Derivatives markets may operate through exchanges or over-the-counter platforms, offering diverse trading opportunities and risk management tools.

Each type of secondary market plays a distinct role in the financial ecosystem, furnishing investors with avenues to trade a broad spectrum of securities and effectively manage their investment portfolios. These markets contribute to liquidity, price determination, and risk mitigation, thereby facilitating the efficient allocation of capital and fostering economic development.

What are Indexes?

In the stock market, an index acts as a statistical gauge representing the performance of a specific group of stocks or securities. It serves as a yardstick for investors to evaluate the overall performance of a particular market, sector, or segment. Indexes are typically constructed by calculating a weighted average of the prices or market capitalizations of the underlying assets, with periodic adjustments made to reflect changes in the constituent stocks.

The Nifty 50 index, a prominent stock market index in India, is managed by the National Stock Exchange of India (NSE). It comprises 50 of the largest and most actively traded stocks listed on the exchange. The Nifty 50 index is designed to mirror the performance of the Indian equity market and is regarded as a key indicator of the country's economic well-being.

Selection of stocks for inclusion in the Nifty 50 index is based on various criteria, including market capitalization, liquidity, trading volume, and sectoral representation. Stocks meeting these criteria are added to the index, while those failing to meet them are excluded. The composition of the Nifty 50 index undergoes periodic reviews and adjustments to ensure it accurately reflects the Indian stock market.

The Nifty 50 index is weighted by free float market capitalization, which means that the weight of each constituent stock is determined by its market capitalization adjusted for the proportion of shares available for trading. This weighting method ensures that larger companies have a greater impact on the index's performance.

The Nifty 50 index serves several purposes for investors and market participants:

1. **Benchmark**: The Nifty 50 index acts as a benchmark against which investors can measure the performance of their portfolios. Fund managers, institutional investors, and individual investors use the index as a reference point to assess their investment returns.
2. **Investment Products**: Various investment products, such as index funds and exchange-traded funds (ETFs), are designed to track the performance of the Nifty 50 index. These products offer investors an efficient way to gain exposure to the Indian equity market without having to select individual stocks.
3. **Market Sentiment**: Movements in the Nifty 50 index are closely monitored by market participants and analysts as an indicator of market sentiment and investor confidence. A rising index is typically interpreted as a positive sign, while a falling index may indicate uncertainty or negativity.
4. **Sector Representation**: The composition of the Nifty 50 index reflects the diversity of the Indian stock market across various sectors. Changes in the index's composition over time provide insights into sectoral trends and investor preferences.

In summary, the Nifty 50 index plays a vital role in the Indian stock market ecosystem, offering investors a comprehensive measure of market performance and serving as a valuable tool for investment decision-making and portfolio management.
Natural Language Processing (NLP) is a branch of artificial intelligence (AI) that focuses on enabling computers to understand, interpret, and generate human language in a way that is meaningful and contextually relevant. NLP techniques allow computers to analyze, process, and manipulate large volumes of text data, extracting valuable insights and facilitating communication between humans and machines.

Now let’s understand NLP:

It has a wide range of uses across various industries and applications:

1. Sentiment Analysis: NLP techniques are commonly used to analyze the sentiment expressed in text data, such as social media posts, customer reviews, and news articles. Sentiment analysis helps businesses gauge public opinion, track brand perception, and make informed decisions based on the overall sentiment of a given text.
2. Language Translation: NLP enables automated language translation, allowing users to translate text from one language to another accurately and efficiently. Translation tools powered by NLP have become indispensable for international communication, business transactions, and global collaboration.
3. Text Summarization: NLP techniques can be used to summarize large volumes of text into concise and informative summaries. Text summarization algorithms identify key

information, extract important details, and condense lengthy documents into shorter summaries, making it easier for users to digest and understand complex information.

4. Named Entity Recognition (NER): NLP algorithms can identify and classify named entities within text data, such as names of people, organizations, locations, dates, and numerical values. NER is useful for information extraction, entity linking, and data categorization in various applications, including information retrieval and document analysis.
5. Question Answering Systems: NLP-powered question answering systems can understand natural language questions posed by users and provide relevant answers by analyzing text data. These systems utilize techniques such as semantic parsing, information retrieval, and machine learning to process queries and retrieve accurate responses from large databases or knowledge bases.
6. Chatbots and Virtual Assistants: NLP is used to develop chatbots and virtual assistants that can engage in natural language conversations with users, answer questions, provide assistance, and perform tasks autonomously. Chatbots powered by NLP algorithms are deployed in customer service, e-commerce, healthcare, and various other domains to enhance user experiences and streamline interactions.

In the context of stock markets, NLP has several applications and use cases:

1. News Sentiment Analysis: NLP techniques are employed to analyze sentiment in financial news articles, press releases, and social media posts related to stocks and companies. By quantifying sentiment polarity and strength, NLP models can assess market sentiment, predict stock price movements, and identify potential trading opportunities.
2. Market News Summarization: NLP algorithms can automatically summarize financial news articles and reports, extracting key information, market trends, and company-specific insights. Summarized news content helps investors stay informed about market developments, make timely decisions, and manage their investment portfolios more effectively.
3. Event Extraction and Impact Analysis: NLP is used to extract relevant events and announcements from textual data, such as earnings reports, mergers and acquisitions, product launches, and regulatory filings. By analyzing the impact of these events on stock prices and market volatility, NLP models provide valuable insights for traders, analysts, and investors.
4. Sentiment-based Trading Strategies: NLP-driven sentiment analysis can be integrated into algorithmic trading strategies to automate trading decisions based on market sentiment signals. By incorporating sentiment indicators derived from text data, traders can develop more sophisticated trading models, improve prediction accuracy, and enhance risk management strategies.
5. Investor Sentiment Tracking: NLP techniques enable the tracking and monitoring of investor sentiment through social media platforms, online forums, and financial news websites. By analyzing discussions, comments, and sentiment trends, NLP models help investors gauge market sentiment, investor sentiment, and market sentiment trends.

In summary, Natural Language Processing (NLP) is a powerful technology that enables computers to understand, interpret, and generate human language. NLP has a wide range of applications across industries, including sentiment analysis, language translation, text summarization, named entity recognition, question answering, and chatbots. In the context of stock markets, NLP is used for sentiment analysis of financial news, news summarization, event extraction, sentiment-based trading strategies, and investor sentiment tracking.

What is Sentiment Analysis and how is it helpful to investors?

Sentiment analysis, particularly in the context of news, is a powerful tool for investors seeking to understand market sentiment and make informed decisions. News sentiment analysis involves the use of Natural Language Processing (NLP) techniques to assess the sentiment expressed in financial news articles, press releases, and social media posts related to stocks and companies. By analyzing the tone and sentiment of these sources, investors can gain insights into market sentiment trends, identify potential market-moving events, and make timely investment decisions.

In India, where the stock market is highly influenced by news and market sentiment, investors often rely on news sources as a key factor in their decision-making process. Whether it's corporate announcements, economic indicators, or geopolitical events, news articles can have a significant impact on stock prices and market dynamics. Therefore, understanding the sentiment conveyed in news articles is crucial for investors looking to anticipate market movements and capitalize on trading opportunities.

Sentiment analysis typically categorizes sentiment into three main categories: positive, negative, and neutral. Positive sentiment indicates optimism or bullishness, negative sentiment signifies pessimism or bearishness, while neutral sentiment suggests a lack of strong sentiment one way or the other. These categories are determined by analyzing the language used in the text and assigning sentiment scores based on the presence of positive or negative words, phrases, or context.

One popular tool for sentiment analysis is the TextBlob library, a Python library for processing textual data. TextBlob offers a simple and intuitive interface for performing various NLP tasks, including sentiment analysis, part-of-speech tagging, and text classification. In the context of sentiment analysis, TextBlob utilizes pre-trained models and lexicons to analyze text and assign polarity scores ranging from -1 (very negative) to +1 (very positive). These polarity scores indicate the overall sentiment expressed in the text, allowing investors to gauge the sentiment of news articles and make informed decisions accordingly.

By leveraging sentiment analysis techniques and tools like TextBlob, investors in India can gain valuable insights into market sentiment and use this information to enhance their trading strategies. For example, positive sentiment in news articles about a particular company may indicate strong growth prospects or positive developments, leading investors to consider buying its stock. Conversely, negative sentiment may signal potential risks or challenges, prompting investors to sell or avoid investing in that stock.

Overall, sentiment analysis of news articles offers a data-driven approach to understanding market sentiment and making investment decisions. By harnessing the power of NLP and tools like TextBlob, investors can analyze sentiment trends, identify market-moving events, and adjust their investment strategies accordingly, ultimately improving their chances of success in the dynamic Indian stock market.

Imagine the potential of a software capable of predicting the impact of news on the stock market with measurable accuracy or, at the very least, providing an indication of whether the market will move up or down in response to specific news events. Such a tool could revolutionize the way investors approach decision-making in the stock market.

For investors, having access to a predictive software that offers insights into the potential market reaction to news could be immensely beneficial. Instead of relying solely on intuition or subjective analysis, investors could use quantifiable data and predictive analytics to inform their trading decisions. By understanding how various types of news events, such as earnings reports, economic indicators, or geopolitical developments, are likely to influence market sentiment and stock prices, investors can adjust their investment strategies accordingly.

Even if the predictions provided by such software were not entirely quantifiable, but rather offered directional indications (e.g., whether the market is expected to rise or fall in response to a specific news event), they would still be valuable for investors. Such insights could help investors anticipate market movements and take proactive measures to mitigate risks or capitalize on opportunities. Whether it's adjusting portfolio allocations, placing buy or sell orders, or hedging against potential downside, having advance knowledge of the market's likely direction can give investors a significant edge in navigating the complexities of the stock market.

In essence, a software tool that predicts the impact of news on the stock market, whether through quantifiable metrics or directional indications, has the potential to empower investors with actionable insights and enhance their decision-making capabilities. By leveraging advanced technologies such as artificial intelligence, machine learning, and natural language processing, such a tool could revolutionize the way investors analyze and respond to news events in the dynamic and fast-paced world of stock trading.

CH. II. Business Problem and Justification

In the fast-paced world of stock trading, investors are often faced with the challenge of making split-second decisions in response to news events that can impact the market. Whether it's corporate announcements, economic data releases, or geopolitical developments, news can have a profound effect on stock prices, and investors must react swiftly to capitalize on opportunities or mitigate risks. However, one of the key challenges faced by investors is the lack of quantifiable data or metrics to assess the potential impact of news on stock prices accurately. Without this information, investors may struggle to make informed decisions, leading to missed opportunities or costly mistakes.

Consider a scenario where a positive news article is published about a company, leading investors to believe that its stock price will rise. However, without knowing the extent of the potential price increase or having a quantifiable measure of the news' impact, investors may end up buying more shares than necessary, only to find that the stock price remains relatively unchanged or even decreases. Similarly, negative news events may prompt investors to sell their shares hastily, fearing a significant decline in stock prices, only to realize later that the impact was minimal or temporary.

This lack of clarity and uncertainty surrounding the potential impact of news on stock prices poses a significant challenge for investors, as it increases the risk of making misguided decisions based on incomplete or subjective information. Without reliable tools or methodologies to assess the true magnitude of the market reaction to news events, investors are left vulnerable to market volatility and may struggle to achieve their investment objectives effectively.

Justification for the Study:

Given the challenges and uncertainties associated with reacting to news events in real-time, there is a clear need for a solution that can assist investors in analyzing news sentiment and predicting the potential impact on stock prices. By leveraging advanced technologies such as natural language processing (NLP) and machine learning, a model or software could offer investors valuable insights into the sentiment of news articles and forecast the likely movement of stock prices with greater accuracy.

Such a solution would provide investors with a quantitative measure of news sentiment, enabling them to make more informed decisions about whether to buy, sell, or hold their positions in response to specific news events. By quantifying the impact of news on stock prices, investors can better assess the risk-return trade-off of their investment decisions and adjust their portfolios accordingly to optimize returns and minimize losses.

Moreover, a model or software that analyzes news sentiment in real-time and predicts stock price movements could significantly enhance investors' ability to react swiftly to market developments and capitalize on trading opportunities. By receiving timely alerts or notifications about significant news events and their potential impact on stock prices, investors can stay ahead of the curve and make proactive decisions to stay aligned with market trends and dynamics.

Overall, the development of a model or software that assists investors in analyzing news sentiment and predicting stock price movements in real-time has the potential to revolutionize the way investors approach decision-making in the stock market. By providing investors with actionable insights and predictive analytics, such a solution could empower them to navigate the complexities of the market more effectively and achieve their investment goals with greater confidence and success.

CH. III. Literature Review

- **Samuel W.K. Chan:**

    - The literature on sentiment analysis in financial texts has undergone a progressive evolution, transitioning from conventional keyword matching to more sophisticated approaches. In their notable contribution, Samuel W.K. Chan and Mickey W.C. Chong introduce a groundbreaking language-independent sentiment parser, incorporating heterogeneous context features and ensemble machine learning techniques. Additionally, their proposed sentiment assessment heuristic assigns polarity to phrases and extends to derive sentence-level polarity, offering a nuanced understanding of sentiments within different linguistic structures. The paper stands out in its applicability to financial texts, demonstrated through rigorous evaluations using a movie review dataset and financial text streams, encompassing twelve million words of financial text. Importantly, Chan and Chong's work aligns with the wisdom of the crowd concept, providing snapshots of collective sentiments from diverse textual sources. This unique approach, coupled with empirical validations, positions their work as a significant and valuable addition to the existing literature on sentiment analysis in financial contexts, addressing crucial gaps and challenges in the field.

- **Mariana Daniel:**
    - This paper focuses on gauging event popularity within a defined financial community by employing sentiment analysis on tweets spanning September 2013 to September 2015. Introducing four distinct text analysis tools, the research tackles the challenges posed by vast data volumes and noise characteristic of social networks. The study demonstrates the model's efficiency in handling large datasets and consistently detecting financial events, showcasing the influential role of the defined financial community in disseminating tweets about companies. The conclusion suggests potential extensions, aligning with existing literature. Recommendations include fortifying the financial community creation process, exploring the impact of retweets on event detection, addressing noise through supervised learning, testing diverse financial communities, and integrating sentiment analysis tools with Google Trends to enhance temporal understanding. These proposed extensions align with current trends and challenges in sentiment analysis research, contributing to the evolving discourse on effective strategies for event detection and community dynamics understanding within financial contexts.

- **David Valle-Cruz:**
    - This research investigates the impact of Twitter posts' polarity on world financial indices during pandemics, posing a specific research question on the influence of Twitter-generated sentiment on financial behaviors. Notably, the findings suggest that Twitter posts significantly influenced financial indices during both the H1N1 and COVID-19 pandemics, with a more pronounced effect observed during the latter. The contributions of this study lie in the utilization of a lexicon-based approach, enhanced by a shifted correlation analysis, revealing latent correlations in the data. Additionally, the research underscores the importance of sentiments on Twitter in affecting financial indices, with effects observable a few days after the posts. The superiority of SenticNet over other lexicons is confirmed, and a proposed correlation matrix explores the relationship between Twitter sentiment and financial market behavior, considering time shifts. The study also highlights the impact of social media propagation, particularly the increase in Twitter accounts, on financial indices' behavior over 11 years. Furthermore, the research delves into the non-normal context of pandemics, offering additional value compared to studies conducted under traditional conditions. The conclusion suggests avenues for future research, including the analysis of other indices, markets, or products, exploration of sophisticated techniques like semantic computing, and the real-time analysis of data from other social media platforms such as Facebook and WhatsApp. This study contributes to the evolving literature at the intersection of sentiment analysis, social media, and financial markets, offering insights into the dynamics of investor reactions during pandemic situations.

- **Ayman E. Khedr:**

    - The proposed model in this study addresses the complex task of understanding stock market behavior by simultaneously analyzing different types of news alongside historical numeric attributes. The research builds on three categories of news data: market-relevant news, company-specific news, and financial reports authored by experts in the financial domain. The two-stage model incorporates a sentiment analysis phase using the Naïve Bayes algorithm to determine news polarities (positive or negative) and a subsequent prediction phase that integrates these polarities with processed historical numeric data attributes, employing the K-NN algorithm to forecast future stock trends. The results demonstrate notable improvements in prediction accuracy, with the sentiment analysis stage achieving up to 86.21% accuracy, and the overall model achieving an accuracy of 89.80% in predicting future stock market behavior. The study underscores the importance of considering different values of numeric attributes, revealing superior accuracy compared to prior research. The model's utilization of both Naïve Bayes and K-NN methods contributes to its robust performance, aligning with existing literature emphasizing the strong relationship between stock news and price changes. Future enhancements to the model could involve incorporating technical analysis indicators and recognizing emotional sentences for improved sentiment analysis. Additionally, the study suggests the potential incorporation of social media news for a more comprehensive understanding of stock market dynamics, opening avenues for further exploration at the intersection of sentiment analysis, numerical attributes, and financial forecasting.

- **Tushar Rao:**

    - Rao and Srivastava (Year) explore the relationship between Twitter sentiment analysis and stock market performance, building upon prior research indicating Twitter's influence on market sentiment (Bollen, Mao, & Zeng, 2011; Zhang, Fuehres, & Gloor, 2011). They address limitations in existing approaches by developing a scalable model that captures mass sentiment toward specific indices or companies. Their study demonstrates strong correlation between Twitter sentiment and stock returns, achieving up to 91% directional accuracy. Overall, their research contributes to understanding the impact of social media sentiment on stock market dynamics.

- **Rajendra N. Paramanika:**
    - Recent literature explores asymmetric GARCH models to understand the impact of positive and negative investor sentiments on financial markets. This study introduces news-based sentiment analysis specific to the Indian stock market context, revealing the dominance of negative sentiment and the prevalence of noise traders. The findings underscore the need for regulatory interventions to mitigate market volatility driven by irrational exuberance. Future research could focus on sector-specific analyses within the Indian stock market to deepen our understanding of market dynamics and inform targeted investment strategies.

CH. IV. Research Gap

While there is a significant body of research demonstrating the impact of news on stock prices, particularly evidenced by real-time market reactions to news events, there remains a notable gap in the literature concerning the development of a functional model capable of predicting the immediate impact of news on stock prices. Although instances such as the recent statement by the SEBI governor regarding concerns over inflated prices in small and mid-cap stocks resulting in a subsequent market downturn provide anecdotal evidence of the influence of news on market movements, there is limited scholarly inquiry into creating a predictive model that operates in real-time to forecast such impacts, especially within the context of the Indian stock market.

Existing research has predominantly focused on examining the relationship between news sentiment and stock prices, often employing sentiment analysis techniques to quantify the tone and relevance of news articles. However, the majority of these studies have been retrospective in nature, analyzing historical data to draw conclusions about the broader impact of news on market behavior. While these studies offer valuable insights into the general patterns and trends linking news events to stock price movements, they fall short of providing actionable real-time predictions that investors can leverage to make informed trading decisions.

Moreover, the limited research that does exist in the domain of predictive modeling for stock price impact tends to concentrate on markets outside of India, leaving a significant gap in understanding how news affects stock prices specifically within the Indian context. Given the unique dynamics and characteristics of the Indian stock market, including factors such as regulatory environment, investor sentiment, and market structure, it is essential to conduct research tailored to this specific market to capture its nuances accurately.

This study seeks to address these gaps in the literature by developing a working model capable of predicting the immediate impact of news on stock prices in real-time, with a focus on the Indian stock market. By leveraging classification models to predict the direction of price movements (i.e., whether prices will rise or fall) and regression models to forecast actual price changes, this research aims to provide investors with actionable insights that can inform their trading decisions promptly.

The dataset used for this study comprises news articles related to the top constituents of the Nifty 50 index, representing a diverse array of sectors within the Indian economy. By focusing on news directly relevant to these key companies, this research aims to capture the most impactful news events that are likely to influence overall market sentiment and stock price movements.

In summary, while prior research has demonstrated the influence of news on stock prices, there remains a gap in the literature regarding the development of predictive models that operate in real-time and specifically cater to the nuances of the Indian stock market. This study aims to bridge this gap by creating a robust predictive model that can assist investors in making informed trading decisions based on timely and accurate assessments of news-related market dynamics.

CH. V. Research Design & Methodology

The research design encompasses the meticulous collection of a diverse and extensive dataset, focusing on news articles related to companies within the Nifty 50 sectors. Specifically, the top constituents with the highest weightage in each sector, totaling approximately 10 companies per sector, will be selected. This meticulous selection process yields a dataset comprising news articles associated with around 140 companies

Objectives:

1. To uncover and analyze the intricate relationship between sentiment in financial news articles and subsequent stock price movements, aiming to provide a comprehensive understanding of how these sentiments dynamically shape and influence the broader dynamics of stock markets over time.

Data Collection:

- Collection of news data with the help of **NEWS API** of approximately 30 days consisting of 5067 news.
- News related to the top constituents of the **NIFTY 50 sectors**.
- Ex: HDFC, Reliance Industries, HCL Technologies etc.
- Collection of Nifty 50 Price data (Open, High, Low, Close, ADJ Close, Volume) over the same time period as the news.

Data Preprocessing:

- Cleaning of news data by removal of stop words, Stemming and tokenizing data.

Sentiment Analysis:

- Using the textblob library implementing sentimental analysis on the news.
- Categorizing News into sentiments of Positive(more than 0), Negative(less than 0) and Neutral.(Equal to 0)

Correlation Analysis:

- Performed correlation analysis between Sentiment Polarity of news and rest of the factors like Open, High, Low, Close, ADJ Close, Volume, High_Low_Difference and Close_Open_Difference.

Machine Learning Model Development:

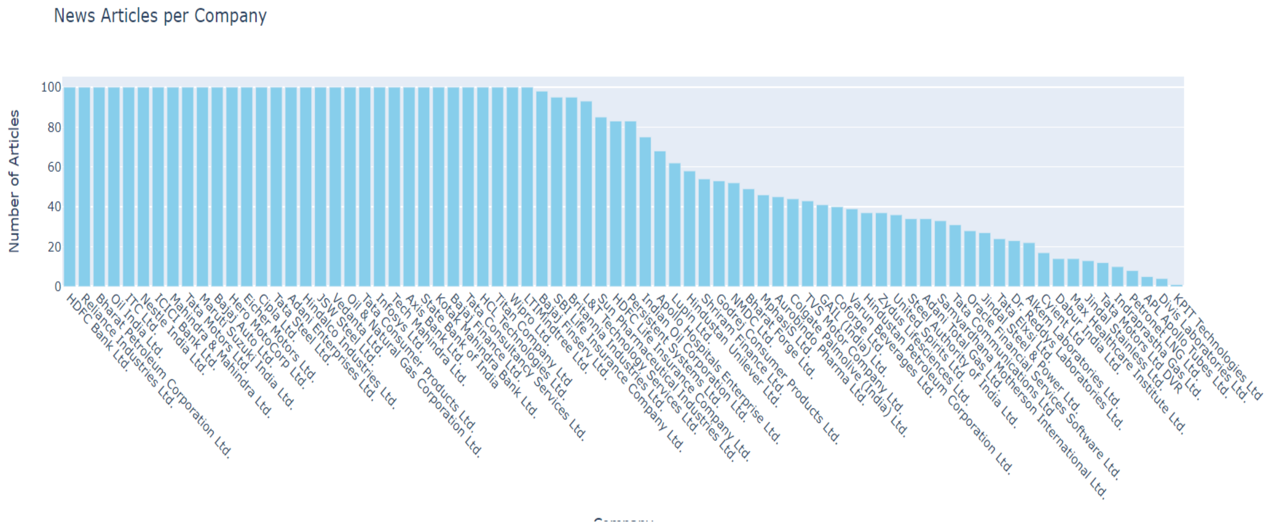
- Used Classification Model for Classifying whether the price will move up or down.

- Usage of KNN and Logistic Regression models for Classification.

CH.VII. FINDINGS AND OBSERVATION

After the extraction of data through the NEWS API we received 5067 news/datapoints of close to 140 companies.

Number of articles per company can be seen through following graphs:



Now let’s talk about number of articles by the top constituents mentioned by the nifty 50 factsheets.

Top constituents by weightage

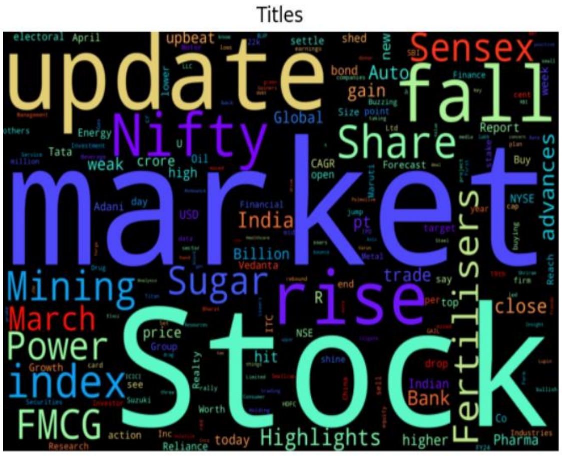
Company’s Name	Weight(%)
HDFC Bank Ltd.	11.07
Reliance Industries Ltd.	10.22
ICICI Bank Ltd.	7.80
Infosys Ltd.	5.44
Larsen & Toubro Ltd.	4.52
Tata Consultancy Services Ltd.	3.99
ITC Ltd.	3.86
Bharti Airtel Ltd.	3.25
Axis Bank Ltd.	3.02
State Bank of India	2.93

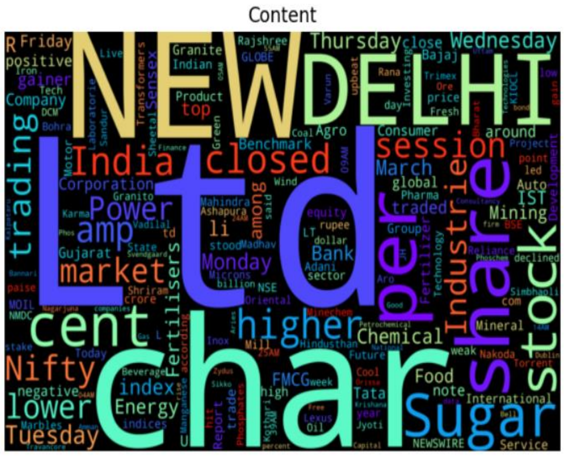
We can see approximately 100 articles per top constituents mentioned by the nifty 50 factsheet.

Wordcloud of the article titles and article content:

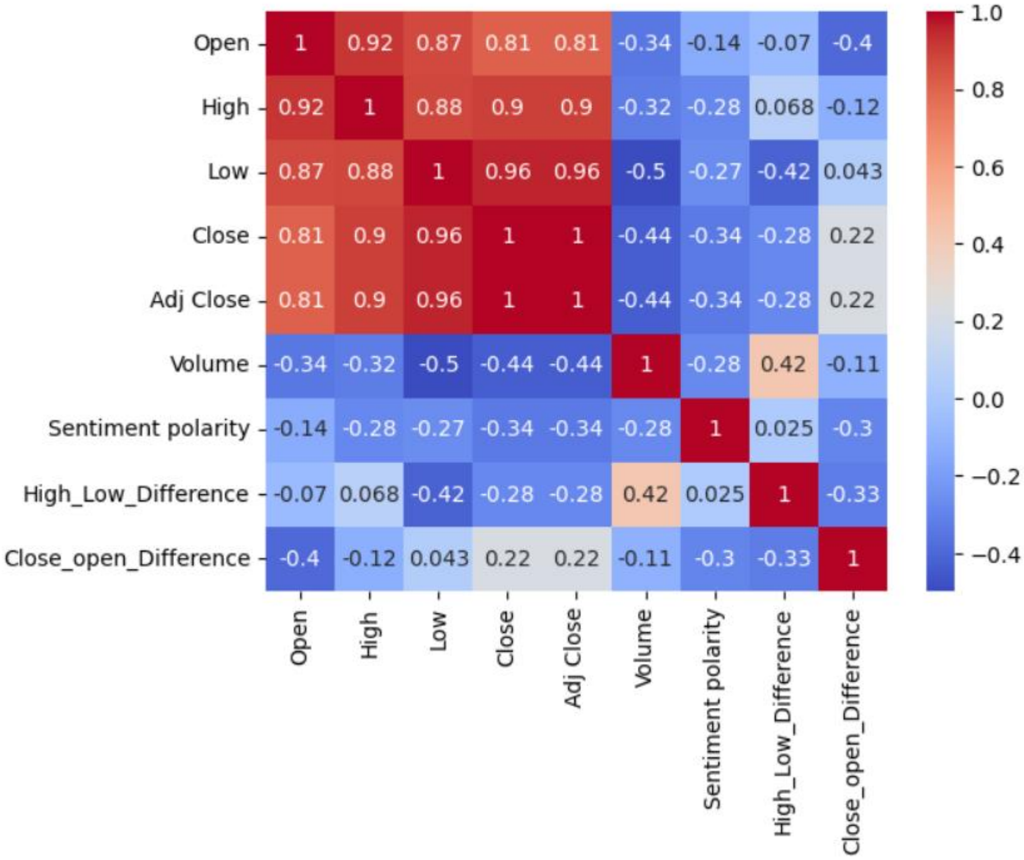
A **word cloud**, also known as a **tag cloud** or **text cloud**, visually represents text data by displaying words in varying sizes based on their frequency or importance within a given set of text. These visualizations highlight prominent terms, making it easier to identify key themes, topics, or recurring terms. Word clouds are commonly used for content analysis, keyword extraction, and labelling. However, they have limitations, such as ignoring context and providing a single snapshot of word frequencies.

So if we see the wordcloud for titles and content, then we can deduce that the articles talk a lot about terms like Update, nifty, market, stock, FMCG, sensex, share, sugar.





After calculating sentiment polarity and creating a new dataframe that contains values for open, high, low, close, AdjClose, Volume and sentiment polarity. We create 2 an additional column to it showcasing High-Low difference (diff between the values of high and low) and Close-Open difference (difference between the values of close and open).



A **heatmap** visually represents data by color-coding values. This heatmap shows correlations between different metrics. Here’s what each component means:

1. **Sentiment Polarity**:
    1. Reflects the positivity or negativity of news or social media content related to stocks.
    2. Correlations:
        - **Open, High, Low, Close, Adjusted Close,Volume**: Negatively correlated (ranging from -0.14 to -0.34) – As these prices increase, sentiment tends to be slightly more negative.
        - **High_Low_Difference**: slightly positive correlation (0.025) – As the difference is positive the sentiments also tend to be slightly positive
2. **High_Low_Difference**:
    1. Measures intra-day volatility by calculating the difference between high and low prices.
    2. Correlations:

        - **Close_Open_Difference**: Moderately positively correlated (0.33) – As the difference between closing and opening prices increases, high-low volatility tends to be higher.

3. **Close_Open_Difference**:
    1. Indicates daily return by measuring the difference between closing and opening prices.
    2. Correlations:
        - **High_Low_Difference**: Moderately positively correlated (0.33) – As intra-day volatility increases, daily return tends to be higher.

Observing the heatmap we can decode that there is a slight relationship between High_low_Difference and sentiment Polarity, based on which a model can be generated.

Results From Machine Learning models:

Classification Model

Logistic Regression: variables taken are Sentiment Polarity, Relative Change(close_open_difference) and Updown(based on relative change classification of data into up or down portraying movement of the market)

```
Accuracy: 0.67
Classification Report:
              precision    recall  f1-score   support

           0       0.00      0.00      0.00         1
           1       0.67      1.00      0.80         2

    accuracy                           0.67         3
   macro avg       0.33      0.50      0.40         3
weighted avg       0.44      0.67      0.53         3
```

Let’s break down the results from the Logistic Regression model:

1. **Accuracy**:
    - The accuracy of the model is **0.67**, which means it correctly predicts the market movement (up or down) in approximately 67% of cases.
2. **Classification Report**:
    - The classification report provides additional insights into the model’s performance:
        - **Precision**:
            - Precision for class 0 (downward movement) is **0.00**. This means that when the model predicts a downward movement, it is often incorrect.
            - Precision for class 1 (upward movement) is **0.67**. When predicting an upward movement, the model is more accurate.
        - **Recall (Sensitivity)**:

- Recall for class 0 is **0.00**. The model misses most of the actual downward movements.
- Recall for class 1 is **1.00**. It correctly identifies all actual upward movements.

- **F1-Score**:
  - The F1-score balances precision and recall. For class 0, it is **0.00**, indicating poor performance. For class 1, it is **0.80**, which is better.
- **Support**:
  - The number of instances in each class (1 for upward, 1 for downward).

- The **macro average** considers equal weight for both classes and calculates the average precision, recall, and F1-score. In this case, it's **0.33** for precision, **0.50** for recall, and **0.40** for F1-score.
- The **weighted average** accounts for class imbalance (different support for each class). It's **0.44** for precision, **0.67** for recall, and **0.53** for F1-score.

KNN Model: variables taken are Sentiment Polarity, Relative Change(close_open_difference) and Updown(based on relative change classification of data into up or down portraying movement of the market)

```
Accuracy: 0.50
Classification Report:
              precision    recall  f1-score   support

           0       0.50      1.00      0.67         2
           1       0.00      0.00      0.00         2

    accuracy                           0.50         4
   macro avg       0.25      0.50      0.33         4
weighted avg       0.25      0.50      0.33         4
```

Let's interpret the results from the K-Nearest Neighbors (KNN) classifier:

1. **Accuracy**:
   - The accuracy of the KNN model is **0.50**, which means it correctly predicts the market movement (up or down) in approximately 50% of cases.
2. **Classification Report**:
   - The classification report provides additional insights into the model's performance for both classes (up and down):
     - **Precision**:
       - Precision for class 0 (downward movement) is **0.50**. When the model predicts a downward movement, it is accurate 50% of the time.
       - Precision for class 1 (upward movement) is **0.00**. The model struggles to predict upward movements.
     - **Recall (Sensitivity)**:
       - Recall for class 0 is **1.00**. It correctly identifies all actual downward movements.
       - Recall for class 1 is **0.00**. The model misses all actual upward movements.

    - **F1-Score**:
        - The F1-score balances precision and recall. For class 0, it is **0.67**, indicating reasonable performance. For class 1, it is **0.00**, reflecting poor performance.
    - **Support**:
        - The number of instances in each class (2 for both upward and downward movements).
- The **macro average** considers equal weight for both classes and calculates the average precision, recall, and F1-score. In this case, it's **0.25** for precision, **0.50** for recall, and **0.33** for F1-score.
- The **weighted average** accounts for class imbalance (different support for each class). It's **0.25** for precision, **0.50** for recall, and **0.33** for F1-score.

CH.VII. Conclusion

It is evident from the research conducted that there exists a discernible relationship between sentiments expressed in news articles and stock prices. However, this relationship is not always straightforward and can be influenced by various factors, including investor behavior, market dynamics, and the use of alternative trading strategies such as technical indicators. While sentiments captured from news articles can provide valuable insights into market sentiment, it is essential to acknowledge that investors may employ diverse strategies beyond solely relying on news sentiment to inform their trading decisions.

In some instances, the relationship between news sentiments and stock prices may exhibit a high degree of correlation, leading to predictable market movements. Conversely, there are scenarios where this relationship may not be as pronounced or may even appear contradictory. This variability underscores the complexity of market dynamics and the multifaceted nature of investor decision-making, which extends beyond the scope of news sentiment alone.

Nevertheless, despite these challenges, it is evident that classifying whether the market will move up or down based on news sentiment can yield more reliable results compared to predicting the actual magnitude of price changes. This classification approach offers a practical and actionable framework for investors to make informed decisions about their trading strategies, particularly in response to news events.

Moreover, by combining different classification models, such as logistic regression and k-nearest neighbors (KNN) classifier, it is possible to enhance the accuracy of predicting market movements. This hybrid approach leverages the strengths of each model to achieve better classification performance for both upward and downward market trends, thereby improving the overall predictive capability of the model.

Future work:

In terms of future research directions, there are several avenues that warrant exploration to further enhance the predictive capabilities of models analyzing the relationship between news sentiment and stock prices. One such avenue involves the collection of more extensive and diverse datasets to train the model effectively. While the data collected for this research was sufficient given the time constraints, a larger dataset encompassing a broader range of news sources and market conditions could provide more robust insights into market dynamics.

Additionally, future research efforts could incorporate not only price movements but also technical indicators to create a more comprehensive model for predicting market trends. By integrating technical analysis with sentiment analysis, researchers can develop a more holistic understanding of market behavior and identify patterns that may not be apparent when considering each aspect in isolation.

Furthermore, leveraging machine learning techniques to analyze classified data points where investors directly respond to news events or take contrary actions could facilitate the development of real-time predictive models. By identifying and analyzing these patterns, researchers can create models that offer actionable insights into market sentiment and provide timely recommendations for investors.

In conclusion, while the relationship between news sentiment and stock prices may be nuanced and multifaceted, there is significant potential for developing predictive models that assist investors in making informed trading decisions. By leveraging advancements in machine learning, data analytics, and sentiment analysis, researchers can continue to refine and improve these models, ultimately empowering investors to navigate the complexities of the stock market with greater confidence and success.

CH. VIII. References

- https://www.moneycontrol.com/promo/mc_interstitial_dfp.php?size=1280x540
- https://www.researchgate.net/profile/Ayman-Khedr-2/publication/318298991_Predicting_Stock_Market_Behavior_using_Data_Mining_Technique_and_News_Sentiment_Analysis/links/5a5615a50f7e9bf2a5369631/Predicting-Stock-Market-Behavior-using-Data-Mining-Technique-and-News-Sentiment-Analysis.pdf
- https://www.sciencedirect.com/science/article/pii/S0167923616301828
- https://www.sciencedirect.com/science/article/pii/S0957417416306571
- https://link.springer.com/article/10.1007/s12559-021-09819-8
- Bollen, J., Mao, H., & Zeng, X. (2011). Twitter mood predicts the stock market. *Journal of Computational Science*, 2(1), 1-8.
- Rao, T., & Srivastava, S. (Year). Analyzing Stock Market Movements Using Twitter Sentiment Analysis. *Journal Name*, Volume(Issue), Pages.
- Zhang, X., Fuehres, H., & Gloor, P. A. (2011). Predicting stock market indicators through Twitter "I hope it is not as bad as I fear". *Procedia-Social and Behavioral Sciences*, 26, 55-62.
- https://www.niftyindices.com/indices/equity/sectoral-indices