PROJECT
FINANCIAL NEWS SENTIMENT ANALYSIS USING AI

A project

Presented to the

Faculty of

California State Polytechnic University, Pomona

In Partial Fulfillment

Of the Requirement for the Degree

Master of Science

In

Business Analytics

By

Ailien Dang

An Vu

Cece Nguyen

Kiet Nguyen

Ngoc Nguyen (Alice)

Ngoc Nguyen (Kim)

ABSTRACT

Purpose: This study investigates the effectiveness of sentiment analysis models in forecasting stock price movements based on financial news headlines. It aims to determine whether general-purpose or industry-specific models provide better predictive performance, helping improve financial decision-making.

Study design/methodology/approach: Three sentiment models—TextBlob, BERT, and FinBERT—were applied to financial news headlines to generate sentiment scores, which were evaluated against actual stock price changes. FinBERT emerged as the most accurate model and was then applied across six industries (Technology, Real Estate, Healthcare, Energy, Finance, and Entertainment) to assess its cross-sector performance.

Findings: FinBERT delivered the highest predictive accuracy across six industries, outperforming other models in Technology, Real Estate, Healthcare, Energy, and Finance. Its strength lies in its alignment with financial language, enabling strong cross-sector generalizability. In contrast, the Entertainment sector favored a fine-tuned model, indicating that media-specific sentiment requires tailored training to capture nuanced market signals.

Originality/value: The study highlights the practical value of using a well-trained general model like FinBERT across most industries, reducing the need for costly fine-tuning without sacrificing performance. It also emphasizes the importance of domain-specific adaptation in sectors with distinct language usage, such as Entertainment.

Practical Implications: Financial analysts and investment platforms can optimize resources by leveraging general sentiment models like FinBERT for broad industry applications. However, for sectors with specialized communication patterns, targeted model refinement may still be necessary to capture nuanced sentiment accurately.

Keywords: Financial sentiment analysis, FinBERT, NLP, industry-specific vs. general models.

SUMMARY OF INDIVIDUAL CONTRIBUTIONS

Ailien Dang – Real Estate Industry

Ailien played a central role in both the analytical and editorial components of the project. She conducted industry-specific modeling for the Real Estate sector using the FinBERT-LSTM framework, evaluating its accuracy in forecasting price movements. Ailien also took charge of reviewing and refining Chapters 1 and 2, ensuring the problem statement, research objectives, and significance were clearly and professionally articulated. Additionally, she verified the accuracy and visual clarity of all graphs and figures and ensured adherence to academic formatting standards across the final report.

Hong Ngoc Nguyen (Alice) – Technology Industry

Alice was responsible for implementing the FinBERT sentiment model for the Technology sector, where she assessed predictive accuracy and contributed critical insights into model refinement. She also ran the BERT model in Phase 1, providing comparative analysis across sentiment tools. Furthermore, Alice drafted the abstract for the final report, effectively summarizing the methodology, findings, and practical implications. Her dual contribution in both technical execution and scholarly writing laid a strong foundation for the report's overall direction.

Kiet Nguyen – Media & Entertainment Industry

Kiet led the sentiment analysis for the Media & Entertainment sector in a stock price prediction project, leveraging a FinBERT-LSTM hybrid model to capture both financial language nuances and temporal market patterns. In Phase 1, he implemented the TextBlob model to establish baseline sentiment scoring, then conducted a targeted deep-dive analysis for the Entertainment sector in Phase 2. Beyond technical execution, Kiet coordinated team workflows by assigning tasks, managing timelines, and performing the final QA review of the comprehensive report. He authored key report sections—including discussion, conclusion, acknowledgments, references, and appendix—and ensured analytical insights were clearly communicated to stakeholders by designing and delivering a polished presentation deck. His work demonstrated a strong balance of modeling expertise, team leadership, and professional presentation delivery.

An Vu – Energy Industry

An led the deep learning-based sentiment analysis for the Energy sector using the ProsusAI/FinBERT model in phrase 1 and built FinBERT-LSTM models to forecast price movements. She was responsible for executing sentiment classification, merging financial data, and producing visual performance evaluations. An authored key section of Chapter 4, ensuring the correctness and reproducibility of the models and results. Her technical rigor and systematic evaluation were critical to model selection and the integrity of the research findings.

Cece Nguyen – Healthcare Industry

Cece was primarily responsible for the Healthcare sector analysis using the FinBERT-LSTM framework. She also managed all data-related tasks, including sourcing, transformation, and validation, as detailed in Chapter 3. Cece ensured that all data tables and datasets were accurate and consistently formatted, providing a strong foundation for the group's modeling efforts and maintaining the integrity of all data-driven outcomes. In addition, she contributed to the Limitations section by identifying key issues such as the restricted 2018 data scope, single-source bias, simplified sentiment aggregation, and overfitting risks in industry-specific models.

Kim Ngoc Nguyen (Kim) – Finance Industry

Kim focused on the Finance sector, applying the FinBERT-LSTM model to evaluate industry-specific sentiment forecasting. She was also responsible for drafting the conference abstract, ensuring it met professional standards and reflected the project's core findings. Kim conducted a final quality assurance check of the full report, validating formatting, page numbering, and table content. Her meticulous review ensured the final submission was cohesive, polished, and ready for academic or professional dissemination.

SIGNATURE PAGE

PROJECT: **FINANCIAL NEWS SENTIMENT ANALYSIS USING AI**

AUTHOR: Ailien Dang

An Vu

Cece Nguyen

Kiet Nguyen

Ngoc Nguyen (Alice)

Ngoc Nguyen (Kim)

DATE SUBMITTED: **08/06/2025**

College of Business Administration

Dr Qian (Ken) Zhang, Ph.D. ____
Assistant Professor, Technology & Operation Management

Dr. Yuanjie He, Ph.D. ____
Professor, Department Chair
Director, MS of Business Analytics Program

Table of Contents

CHAPTER 1 **9**

INTRODUCTION 9

Problem Statement *9*

Analytics Objectives *9*

Significance of the Topic *11*

CHAPTER 2 **12**

BACKGROUND INFORMATION 12

Technology *13*

Energy Index *13*

Global Economic *14*

Healthcare *15*

Real Estate Industry Index *16*

Media & Entertainment *16*

ENVIRONMENTAL ANALYSIS 17

TASK ENVIRONMENTAL ANALYSIS 20

RESEARCH FRAMEWORK/MODEL 22

Sentiment Analysis *22*

FinBERT-LSTM Model *24*

Comparative Analysis of FinBERT-LSTM vs. Other Models *25*

Industry-Specific Sentiment Analysis Models *25*

Need for additional analysis *29*

ANALYTICAL OBJECTIVES EXPLANATION 30

PHASE 1: GENERAL MODEL DEVELOPMENT 31

PHASE 2: INDUSTRY-SPECIFIC MODEL DEVELOPMENT 31

CHAPTER 3 **32**

DATA AND PROPOSED METHODS 32

Data and Sampling *33*

Data Cleaning & Processing Steps *33*

PHASE 1 35

1. Lexicon-based Sentiment Analysis: TextBlob *35*

2. BERT *36*

3. Deep Learning-based Sentiment Analysis (Finance-specific): ProsusAI/finbert *37*

CHAPTER 4 **41**

ANALYTICAL MODEL AND RESULTS 41

PHASE 1- OBJECTIVE 1 41

Textblob Sentiment *41*

BERT-based Sentiment to Price-Change LSTM *43*

BERT-based Sentiment *45*

FinBERT-Based Sentiment and LSTM Modeling to Predict Stock Price Change *47*

Conclusion of Phase 1: Performance Evaluation of Sentiment Models 49
Phase 2-Objective 2 51
1. FinBERT Sentiment Model Applied to Technology Sector (Alice) 51
2. FinBERT Sentiment Model Applied to Real Estate Industry Sector (Ailien) 54
3. FinBERT Sentiment Model Applied to Energy Sector (An Vu) 58
4. LSTM Model Applied to Finance (Kim) 60
5. LSTM Forecasting for Healthcare Sector Price Movement (Cece) 61
6. Predicting Entertainment Industry Stock Movements (Kiet Nguyen) 64
CHAPTER 5 65
CONCLUSION, LIMITATIONS AND RECOMMENDATIONS 68
1. Conclusion 68
2. Limitations 70
3. Recommendations for the Financial Analytics Field 72
REFERENCES 75

CHAPTER 1
INTRODUCTION

Problem Statement

Market sentiment plays a critical role in financial decisions, influencing stock prices, risk management, and investment strategies. Traditionally, traders manually reviewed financial news to assess sentiment, but the increasing volume and speed of information have made this approach impractical. With advancements in AI and natural language processing (NLP), sentiment analysis now leverages deep learning models such as BERT and FinBERT to extract insights from financial headlines at scale.

While these models offer structured sentiment scoring, they often apply general interpretations that may miss important differences across industries. For instance, the same news could have contrasting effects in healthcare versus energy. This raises an important question: can a general sentiment model provide reliable predictions for all sectors, or are industry-specific models needed for greater accuracy?

This project evaluates the effectiveness of sentiment analysis in financial forecasting by comparing a general sentiment model with models fine-tuned for specific industries. The goal is to determine whether sector-level customization improves prediction accuracy enough to justify the extra cost and complexity.

Analytical Objectives

This study aims to evaluate whether industry-specific models outperform a generalized model in stock price forecasting and to determine whether the investment in model fine-tuning is justified by its benefits for financial analysts and investors. The research compares the effectiveness of a generalized model in capturing market sentiment with the predictive capability of sector-specific models, noting that while industry-tailored models are generally expected to yield higher accuracy, a broad model may suffice if it adequately reflects overall market trends.

- *Analytical Objective 1:* Finding the Best NLP Model

We begin by testing various sentiment analysis models—ranging from basic tools like TextBlob to more advanced models like BERT and FinBERT—to determine which most accurately interprets financial news sentiment. The best-performing model will serve as the base for further analysis.

- *Analytical Objective 2:* Predicting index prices with sentiment scores

Using sentiment scores from the chosen model, we predict index price movements. We compare a general model's performance to that of models tailored for six industries: technology, energy, healthcare, real estate, global economy, and entertainment. Some sectors, like technology, may align well with general market sentiment, while others may require more tailored sentiment interpretation.

- *Analytical Objective 3:* Customized Index Membership (Optional)

As an enhancement, we may explore constructing weighted index scores based on individual stock performance within each sector. This customized approach could boost model precision for industry-level forecasts.

The study is conducted in two phases:

- ***Phase 1*** creates a general sentiment model using Bloomberg news headlines across multiple sectors.
- ***Phase 2*** fine-tunes the best model for each target industry and compares performance, assessing whether customization adds value.

If time permits, we will also analyze feature importance to identify which companies drive model predictions, helping refine the general model for future use.

Significance of the Topic

As of 2023, the global financial industry continues to experience strong growth, with the market value of AI services in finance reaching $10.6 billion in 2022 and expected to grow at a CAGR of 24.5%, reaching $41.5 billion by 2030 (Infomineo, 2023). Sentiment analysis in financial information plays a critical role in investment decision-making and predicting stock price fluctuations, with 82% of investors relying on this information to make decisions (Information Matters, 2023). Traditionally, analysts used manual methods to evaluate sentiment from financial news, but the increasing speed and volume of information have made this approach impractical (UC Berkeley School of Information, 2024). This project will assess sentiment analysis models in the financial market, while also developing specialized models to improve the accuracy of stock price predictions and optimize investment strategies.

CHAPTER 2

BACKGROUND INFORMATION

News and social media play a significant role in today's world, where information spreads rapidly and is easily accessible to investors. Many investors, especially short-term traders, do not rely solely on financial metrics or other financial reports. Instead, they also consider market sentiment which refers to the overall attitude, emotions, and psychology of investors toward a particular market. Market sentiment is often derived from news, social media sites and external events when making decisions. Previous sentiment analysis studies have proven that sentiment scores along with historical stock price data can be effectively used in trained models to predict stock market movements. Previous research has demonstrated that sentiment scores, combined with historical stock price data, enhance predictive models for forecasting market trends. Leveraging insights from previous studies, this project aims to enhance sentiment-based stock prediction by implementing and compare various deep learning models (such as LSTM, BERT, etc.) to predict and aggregate sentiment scores from financial news

The beneficiaries of our project are financial analysts, stock traders, CFO, financial advisors, institutional investors, and portfolio managers who rely on daily stock movement. Investors in the S&P 500 and other major indices can use sentiment-driven insights to anticipate potential stock fluctuations and adjust their portfolios accordingly. As daily stock index movements influence market dynamics, our project aims to provide a valuable analytical tool that helps stakeholders navigate uncertainty and capitalize on market opportunities more effectively. The quick capture of emotions and reactions to market events from new headlines, trends detecting, current market sentiment along with economic outlook are conveyed in the group project.

In the following sections, we will explore the role of sentiment analysis in different industries, examining how market sentiment impacts specific sectors such as technology, energy, healthcare, Real Estates, and media & entertainment. For instance, news of an oil price increase will impact the stock prices of transportation companies, as higher fuel costs increase operational expenses. However, it will have little to no effect on technology companies like Apple, since the tech industry is not directly reliant on oil for its operations. This analysis will provide insights into how sentiment influences stock performance and why it has become an essential tool for investors and analysts alike.

Technology

The technology sector in the U.S. is a leading industry and plays a pivotal role in the financial market and stock exchanges. Major tech companies such as Apple, Microsoft, Google, and Tesla not only lead in innovation but also have a significant impact on market value and investment decisions. The rapid growth of the technology sector has attracted the attention of millions of investors, ranging from individuals to large investment funds, as it is a high-potential field capable of delivering substantial returns. Tech stocks consistently rank among the top in trading volume and significantly influence major stock indices such as NASDAQ. With strong volatility and immense growth potential, the U.S. technology sector remains a top choice for investors and a key factor influencing market sentiment and investment strategies worldwide. Therefore, developing a predictive model to optimize investments in stocks within this sector is essential.

Energy Index

Market sentiment in the energy sector is strongly influenced by geopolitical events, commodity prices (particularly oil and natural gas), environmental regulations, and global supply-demand dynamics. Since energy companies are central to economic activity and global trade, investor confidence in this sector can shift rapidly in response to oil production cuts, climate policy changes, or international conflicts affecting resource availability. Sentiment analysis in the energy industry is essential for identifying investor reactions to these market drivers. By leveraging artificial intelligence and NLP, sentiment models can scan large volumes of news articles, analyst reports, and social media content to detect changes in perception around energy trends.

The energy sector is particularly sensitive to fluctuations in crude oil prices, OPEC decisions, renewable energy developments, and environmental legislation. For instance, a sudden announcement of oil production cuts by OPEC can trigger a surge in energy stocks, while new emissions regulations may dampen investor enthusiasm in fossil fuel-based firms. AI-powered sentiment analysis enables real-time monitoring of such developments by identifying shifts in tone across industry commentary. Machine learning models trained on sector-specific data can recognize patterns that precede major stock movements, allowing stakeholders to anticipate volatility and make informed decisions.

Global Economic

Nowadays, inflation, trade policies, geopolitical tensions, and financial market conditions are constantly shaping the global economic situation. Sentiment analysis has emerged as a valuable tool for understanding global economic trends by examining financial news, government policies, and market reactions. AI-driven analysis processes vast amounts of data from news reports, investor commentary, and social media discussions to provide insights into market confidence and economic direction, also providing a more comprehensive understanding of economic conditions. Moreover, traditional economic indicators such as GDP growth, employment rates, and interest rate adjustments provide critical insights, but they often lack the ability to capture the real-time emotions and reactions of investors. This is where sentiment analysis, powered by AI and machine learning, has become an essential tool in financial forecasting. Sentiment analysis processes vast amounts of financial news, central bank statements, corporate reports, and investor discussions to measure the overall mood of the market. AI-driven models such as BERT and FinBERT help identify patterns in news sentiment, allowing investors to anticipate potential risks and opportunities. We can see that the response to global events like economic downturns, interest rate hikes, and political changes are good sources for training.

Healthcare

The healthcare sector is one of the most critical and complex industries in the stock market. It is heavily influenced by policy changes, regulatory approvals, technological advancements, and global health crises. Financial news surrounding healthcare companies often drives investor sentiment, leading to fluctuations in stock prices. Understanding how the market reacts to healthcare-related financial news is essential for investors, policymakers, and industry stakeholders.

Sentiment analysis has become a key tool in tracking market reactions to healthcare news. Traditional financial metrics like revenue reports and earnings per share provide fundamental insights, but they do not capture real-time investor sentiment. AI-driven sentiment analysis allows for the rapid assessment of financial news, earnings announcements, and public discussions to determine how positive or negative sentiment influences stock movements. The healthcare industry experiences sharp market reactions to key events such as drug approvals by the FDA, mergers and acquisitions, and changes in government policies on healthcare spending. The ability to analyze and interpret these shifts using sentiment analysis provides investors with a strategic advantage in making informed financial decisions.

Real Estate Industry Index

In recent times, market sentiment has played a critical role in influencing stock prices within the real estate industry, as interest rates, urban development trends, and investor confidence directly affect the performance of real estate companies. Firms such as Prologis (PLD), Equinix (EQIX), Public Storage (PSA), and Realty Income Corporation (O) in the S&P 500 significantly impact real estate stock indexes. Factors such as monetary policy, inflation, housing demand, and economic uncertainty can drive major fluctuations in the prices of real estate-related equities.

Additionally, financial news related to the real estate sector provides investors with essential insights into market trends, property values, regulatory developments, and economic forecasts. While traditional financial indicators like earnings reports, occupancy rates, and dividend yields are crucial, they often fall short of capturing real-time investor sentiment. AI-powered sentiment analysis has become a powerful tool for tracking market perceptions by analyzing news headlines, REIT reports, and real estate policy announcements. Real estate firms often see stock movements in response to changes in interest rates, zoning policies, and housing supply forecasts. For instance, positive coverage of new commercial development projects or favorable mortgage rate policies can elevate investor confidence and drive real estate index performance.

Media & Entertainment

The media and entertainment sector are a crucial component of the global economy, encompassing major companies in the S&P 500 such as Disney (DIS), Netflix (NFLX), Warner Bros. Discovery (WBD), Comcast (CMCSA), and Paramount Global (PARA). The stock prices of these companies, with Netflix trading at approximately $988.47 and Disney at $111.20, are highly sensitive to consumer sentiment, content success, and industry trends.

Factors such as shifts in consumer behavior, the effectiveness of advertising campaigns, and major events like blockbuster film releases or new product launches can significantly impact stock prices. Additionally, macroeconomic factors such as advertising policies, changes in copyright regulations, and the rise of online consumer behavior play a key role in determining the value of these companies. Sentiment analysis provides an effective tool to quickly gauge public and investor sentiment through sources like financial news, social media, and industry reports, enabling more informed investment decisions. This approach helps identify factors that influence stock value, optimizing investment strategies and mitigating risks.

Environmental Analysis

Economic Factors

Macroeconomic indicators such as GDP growth, inflation rates, interest rate changes, and employment levels significantly influence market sentiment and stock price movements. The study incorporates sentiment scores derived from financial news alongside S&P 500 stock data, allowing for an analysis of how different economic events correlate with market trends. Sectors like energy, technology, and healthcare are particularly sensitive to economic fluctuations, underlining the importance of developing industry-specific sentiment models for improved forecasting accuracy. Economic crises, such as the 2008 financial crisis and the COVID-19 pandemic, have demonstrated how rapidly investor sentiment can shift, necessitating AI models capable of adapting to sudden market changes and predicting potential downturns. The project assesses whether a general sentiment model or industry-specific models are more effective in capturing economic shifts and informing data-driven investment strategies.

Political/Legal Factors

The financial sector is highly regulated, with laws governing market manipulation, insider trading, and financial disclosures playing a key role in shaping investor behavior. Adherence to regulations such as the U.S. SEC guidelines, GDPR in Europe, and CCPA in California affects how AI models process financial data, with strict data privacy standards being a significant consideration. Furthermore, government policies, including interest rate decisions, tax regulations, and trade agreements, directly influence stock market sentiment. For example, changes in the U.S. Federal Reserve's interest rates can significantly impact investor confidence and market liquidity. Geopolitical risks, such as trade wars, political instability, and shifts in government leadership, also contribute to market volatility. These uncertainties affect how investors interpret financial news, highlighting the need for AI-driven sentiment analysis models that can capture the effects of political and legal changes on stock price movements. The project focuses on integrating political sentiment into financial forecasting, providing more robust predictions by incorporating such external factors.

Socio-Cultural Factors

Investor behavior is heavily influenced by market psychology, with factors like fear, greed, speculation, and herd mentality often driving stock market fluctuations. The rapid spread of financial news through major media outlets and social media platforms (e.g., Twitter, Reddit) amplifies these psychological factors, making real-time sentiment analysis increasingly important for tracking shifts in investor sentiment. AI-powered sentiment models can help investors and analysts navigate market sentiment changes and prevent emotional biases from skewing financial decisions. Additionally, generational differences in investment behavior are reshaping financial markets, with younger investors, especially millennials and Gen Z, relying more on algorithmic trading and AI-driven investment strategies. In contrast, older generations tend to favor traditional analytical methods. This project incorporates AI and NLP models (e.g., FinBERT, Llama 2, MLP-Regressor) to account for evolving investment trends, offering tailored insights to investors navigating these shifts.

Technological Factors

Advances in artificial intelligence, natural language processing (NLP), and deep learning have revolutionized sentiment analysis, allowing financial models to process large volumes of financial news data in real-time. The project utilizes models such as FinBERT, Llama 2, VADER, and MLP-Regressor to classify sentiment scores and predict stock price movements based on sentiment trends in financial news. The integration of big data analytics and cloud computing further enhances the scalability of sentiment analysis models, enabling institutional investors and hedge funds to incorporate sentiment-driven insights into their trading strategies. However, these technological advancements also bring cybersecurity risks, including data breaches, model manipulation, and misinformation, underscoring the need for robust security measures in AI-driven financial forecasting systems.

Ecological Factors

Sustainability and Environmental, Social, and Governance (ESG) investing have gained significant importance in financial decision-making. Investors are increasingly incorporating corporate social responsibility (CSR) and sustainability reports into their analyses, influencing stock price trends based on companies' environmental policies. AI-powered sentiment models can track news related to ESG factors, enabling investors to gauge market sentiment regarding sustainable investments. Additionally, industries like cryptocurrency mining, known for their high energy consumption, face scrutiny over their environmental impact. The volatility of digital assets such as Bitcoin is often driven by regulatory pressures on carbon emissions and energy efficiency, which are reflected in financial news sentiment. The project incorporates ecological factors into sentiment analysis models, ensuring that investor sentiment surrounding sustainability-related financial events is accurately captured in stock market predictions.

Task Environmental Analysis

The findings of this project offer actionable insights for a diverse set of stakeholders in the financial ecosystem. By comparing general versus industry-specific sentiment analysis models, this study informs strategic decisions in financial forecasting, investment planning, and AI model development. Below, we outline how different user groups can apply these insights:

Portfolio Managers & Hedge Funds

These professionals manage large pools of assets and often rely on rapid decision-making tied to market sentiment. The models developed in this study, especially the industry-specific FinBERT-LSTM frameworks, can help portfolio managers refine asset allocation strategies by providing early sentiment signals within sectors such as healthcare, technology, and energy. For hedge funds engaging in algorithmic trading or long-short strategies, sentiment-informed models enable more agile responses to news events, leading to improved returns and risk mitigation.

Institutional Investors & Financial Advisors

Institutional investors, such as pension funds and insurance companies, require accurate forecasting tools that capture long-term trends. Industry-specific models, as demonstrated in this study, offer enhanced predictive precision that can inform sector rotation strategies and tactical reallocations. Financial advisors can use sentiment-based insights to guide clients during volatile periods by contextualizing news sentiment in terms of expected market behavior.

Fintech Platforms & Investment Apps

Fintech firms can integrate the general sentiment model or industry-customized versions into user-facing tools to deliver real-time sentiment scores. This empowers retail investors with institutional-grade analytics. Apps can visualize sentiment shifts tied to specific stock indices or sectors, offering users data-driven insights for personalized investment strategies.

AI & Data Science Teams

For AI researchers and developers, this study provides a framework for fine-tuning large language models (LLMs) like FinBERT and Llama 2 for domain-specific applications. The comparative analysis also highlights the trade-offs between generalization and specialization, which can guide the design of scalable and efficient models for other sectors such as crypto, ESG investing, or international markets.

Financial News Media & Analytics Firms

Organizations such as Bloomberg, Reuters, and Yahoo Finance can use these findings to better understand how their content influences market sentiment. The sentiment scores can be fed back into editorial or analytics workflows to assess which news topics trigger significant market responses, supporting both editorial strategy and commercial analytics offerings.

By delivering a comparative framework for sentiment analysis across industries, this project provides stakeholders with concrete tools to better forecast market movements, optimize decision-making, and stay competitive in an increasingly data-driven financial landscape.

Research Framework/Model

Sentiment Analysis

The stock market is highly dynamic and influenced by various factors, including economic indicators, geopolitical events, investor sentiment, and financial news. Traditional stock prediction models have primarily relied on historical stock price data and technical indicators. However, with the rise of Natural Language Processing (NLP), Deep Learning and Generative AI for Sentiment Analysis and sentiment analysis, researchers and financial analysts are increasingly incorporating sentiment scores derived from financial news, analyst reports, and social media to enhance stock price forecasting.

Sentiment analysis helps capture market psychology and investor sentiment, which play a crucial role in stock price fluctuations; many studies have explored the effectiveness in predicting stock prices. According to Maqbool et al. (2023) in their study Stock Prediction by Integrating Sentiment Scores of Financial News and MLP-Regressor: A Machine Learning Approach, integrating financial news sentiment with historical stock data provides better predictions compared to relying on historical data alone. Maqbool states that "Since ANN, SVM and other models were still unable to predict the chaotic fluctuations of the stock market because of the absence of memory element hence unable to remember the long-term trend.” By this means, traditional machine learning models like ANN and SVM struggle with stock market volatility and cannot effectively capture long-term trends with the use of historical stock data while MLP Regressor with sentiment analysis improved trend prediction, outperforming models using only historical data. “Using financial news sentiments along with MLP-Regressor can predict the stock price to an accuracy of 0.90 and shows a high correlation between stock price and financial news.”

Below is the framework of the process of integrating sentiment analysis of financial news with historical stock price data:

1. Financial news articles are collected and processed to extract relevant textual data.
2. Sentiment scores are generated using VADER, TextBlob, or Flair.
3. The model generates predicted stock price trends, helping investors make informed decisions.

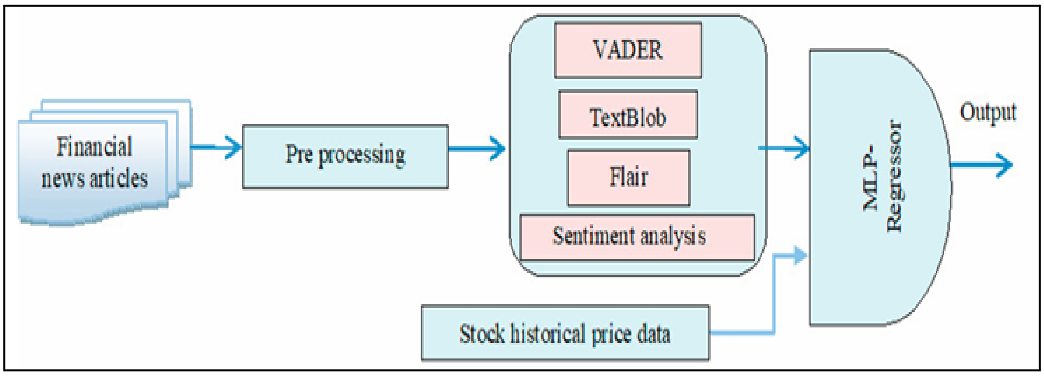

Figure 1: MLP- Regressor along with sentiment analysis model

We used a similar approach to the one presented in MLP- Regressor along with a sentiment analysis model from Maqbool et al. (2023), where sentiment analysis of financial news is integrated with historical stock data. However, our method leverages FinBERT, BERT, and TextBlob to extract financial sentiment, which is then passed into an LSTM model instead of using an MLP-Regressor to capture the temporal patterns in the stock data.

Other studies have adopted a similar framework but employed deep learning-based models instead of NLP for stock prediction. Recent advancements in deep learning have led to the development of more sophisticated sentiment analysis models for financial forecasting. Among these, FinBERT, a domain-specific adaptation of BERT (Bidirectional Encoder Representations from Transformers), has gained prominence in extracting sentiment insights from financial texts. When combined with Long Short-Term Memory (LSTM) networks, this approach has demonstrated significant improvements in stock price prediction.

FinBERT-LSTM Model

The FinBERT-LSTM model integrates FinBERT's contextual understanding of financial news sentiment with the sequential pattern recognition capabilities of LSTM networks. Unlike traditional machine learning models, this approach efficiently processes both financial textual data and historical stock prices to generate more accurate market predictions. According to Wen Gu et al. (2024), integrating sentiment features derived from FinBERT with LSTM networks enhances the ability to detect patterns in stock price fluctuations.

The FinBERT-LSTM model follows a structured workflow:

1. News Article Collection & Processing: Financial news is gathered from reliable sources such as Benzinga and Yahoo Finance.
2. Sentiment Score Computation: The FinBERT model classifies each article's sentiment into positive, negative, or neutral categories.
3. Feature Integration: The sentiment scores are combined with stock price trends from the previous trading days.

4. LSTM-Based Prediction: The integrated data is fed into an LSTM network, which captures sequential dependencies in stock movements.
5. Evaluation & Performance Metrics: Model performance is assessed using Mean Absolute Error (MAE), Mean Absolute Percentage Error (MAPE), and Accuracy.

Comparative Analysis of FinBERT-LSTM vs. Other Models

Experimental results suggest that FinBERT-LSTM outperforms conventional deep learning models such as standalone LSTM and deep neural networks (DNNs). The study by Wen Gu et al. (2024) compared three models:

- FinBERT-LSTM: Achieved the highest accuracy (95.5%) with a MAPE of 0.045.
- LSTM-only: Performed slightly worse with an accuracy of 92.8% and MAPE of 0.072.
- DNN-only: Showed the weakest performance with an accuracy of 78% and MAPE of 0.22.

The superior performance of FinBERT-LSTM highlights the significance of incorporating sentiment analysis in financial forecasting. The model's ability to process textual data and extract meaningful sentiment indicators from financial news gives it a predictive edge over models relying solely on numerical stock data.

Industry-Specific Sentiment Analysis Models

As our project aims to fine-tune models for sector-focused financial news, we seek to capture more precise sentiment scores that reflect market movement to specific industries. Several studies have explored similar approaches. One study that particularly caught our attention is the article by Konstantinidis et al. (2024), LLMs and NLP Models in

Cryptocurrency Sentiment Analysis: A Comparative Classification Study, which examines the effectiveness of large language models (LLMs) and natural language processing (NLP) models in cryptocurrency sentiment analysis.

Figure 2 below from Konstantinidis et al. (2024) presents an overview of sentiment analysis methods between lexicon-based approaches (e.g., dictionary-based models) with machine learning approaches, including unsupervised learning and supervised learning models.

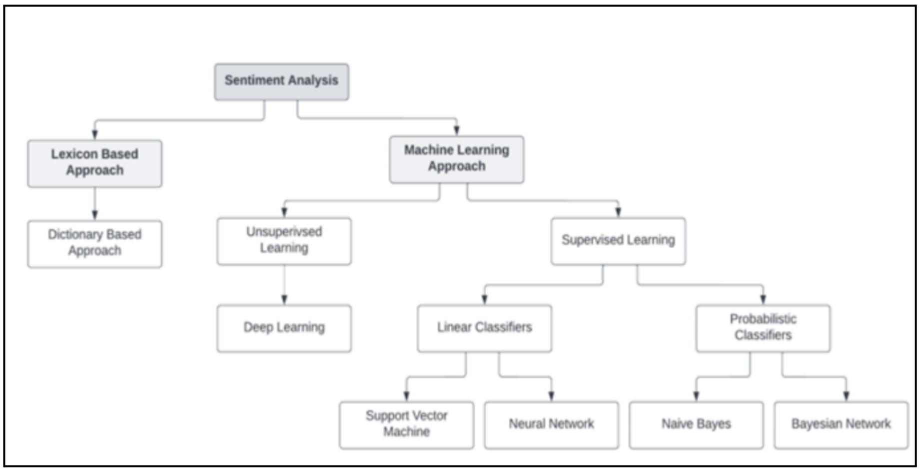

Figure 2: Overview of sentiment analysis methods

The FinLlama model was developed to overcome the limitations of general LLMs in extracting financial sentiment in a way that is both accurate and actionable. As Konstantinidis et al. (2024) explain, terms like "bull" and "bear" are neutral in general language but carry strictly positive or negative meanings in financial markets. This highlights the importance of context-aware sentiment extraction and highlights the complex challenges of applying NLP to financial data (Konstantinidis et al., 2024). So that the parameter-efficient fine-tuning (PEFT) and 8-bit quantization via LoRA were applied to fine-tunes Llama 2 7B on financial datasets, reducing computational costs while maintaining accuracy.

To assess whether FinLlama outperforms other sentiment analysis models, five methods were evaluated: HIV-4, LMD, VADER, FinBERT, and FinLlama. “Each method was applied to every article within each corpus for a given company. In cases where multiple articles were published on the same day for a company, the average sentiment score was calculated for that day "(Konstantinidis et al., 2024) and the metric evaluation are Cumulative Returns, Annualized Return, Ann. Volatility, Sharpe Ratio. The long-short portfolio was constructed using sentiment analysis signals to identify companies to invest in (long positions) and those to bet against (short positions). The evaluation results demonstrate that FinLlama outperformed the other four methods across all performance metrics.

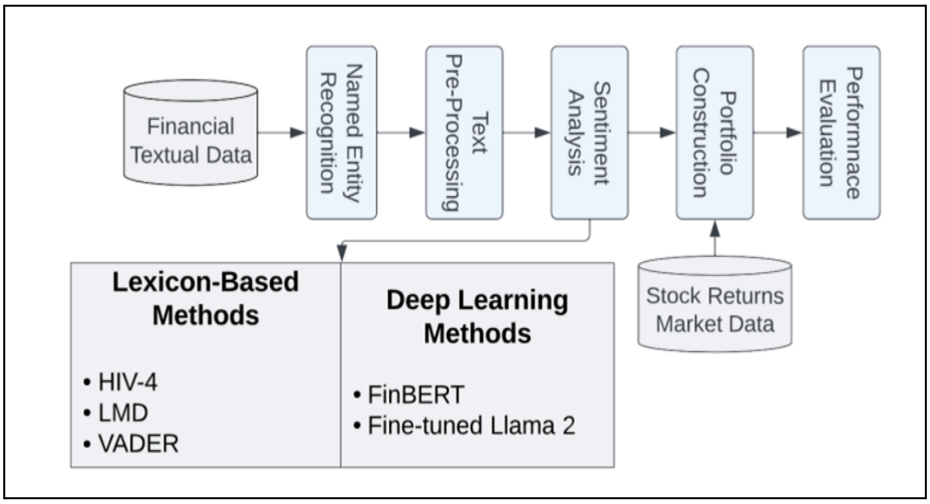

Figure 3: Framework for sentiment analysis

Figure 3, above from Konstantinidis et al. (2024), illustrates the overall framework used in this evaluation—from financial text collection, named entity recognition, and sentiment analysis using both lexicon-based and deep learning methods, to portfolio construction and performance evaluation based on stock market returns.

The study has demonstrated that FinLlama, a fine-tuned Llama 2 7B model, outperforms other sentiment analysis methods by enhancing contextual understanding of financial language. For future research, our group plans to apply a similar fine-tuning approach to different models tailored for various stock market industries, further exploring sector-specific sentiment analysis to improve market predictions.

Similarly, other studies, such as those on FinBERT-LSTM have shown strong performance in general stock market prediction, fine-tuning it for industry-specific sentiment analysis could further enhance accuracy. This aligns with the project's Phase 2 objective of comparing general sentiment models to industry-specific models.

A study by Halder (2022) applied FinBERT-LSTM to predict technology sector stock prices, using NASDAQ-100 index data and news articles from the New York Times. The results showed that incorporating technology-focused sentiment data significantly improved stock prediction accuracy. This suggests that fine-tuning sentiment models for specific sectors could yield better predictive performance.

By training on industry-specific financial news and reports, these models can capture nuanced sentiments unique to particular sectors, improving the precision of stock market predictions. This approach aligns with our project's Phase 2, where we explore whether industry-specific sentiment models can outperform a general financial sentiment model.

Connection to Project Objectives

The success of FinBERT-LSTM in NASDAQ-100 predictions suggests it could be a strong candidate for our project's Phase 1 general sentiment analysis model, which focuses on S&P 500 financial news. Additionally, our project's Phase 2 will evaluate how fine-tuning FinBERT-LSTM for different industries—such as technology, healthcare, and energy—can enhance sentiment-based predictions.

This study aims to answer whether a well-trained general sentiment model is sufficient for financial decision-making or if industry-specific models provide a distinct advantage. The findings will provide valuable insights for investors and financial analysts in selecting the most effective AI-driven approach for market analysis. The integration of sentiment analysis through models like FinBERT-LSTM offers a promising avenue for enhancing the accuracy of stock price predictions. Future research should focus on adapting these models for specific sectors, leveraging industry-relevant data to refine their predictive capabilities further. Additionally, testing FinBERT-LSTM within a broader financial dataset, such as the S&P 500, could provide additional insights into its scalability and reliability as a sentiment analysis model for financial forecasting.

Need for additional analysis

So far, we have reviewed research that integrates historical stock data with sentiment analysis to predict stock prices for specific companies. These studies suggest that aggregating sentiment scores enhances prediction accuracy compared to traditional models. However, we have not encountered a direct comparison of NLP models, deep learning, and generative AI in this context, though some studies contrast lexicon-based approaches with deep learning. Therefore, our project aims to evaluate these models to determine which performs best for stock price prediction.

Another key observation is that most existing research focuses on predicting the movement of individual stock prices rather than analyzing their broader market impact. While some studies, such as *FinLlama: Financial Sentiment Classification for Algorithmic Trading Applications* (Konstantinidis et al., 2024), have examined the influence of specific stocks on market indices; particularly the S&P 500—there is limited research on smaller indices, such as those in the healthcare, financial investment, and Real Estates sectors.

Moreover, we have yet to find research that customizes the weighting of financial news sentiment across different markets to determine which companies exert the greatest influence on overall stock indices. Large-cap companies like Apple, Meta (formerly Facebook), and Microsoft disproportionately affect the NASDAQ-100 and S&P 500 indices.

Understanding these relationships could improve market-wide prediction models by refining how financial sentiment is aggregated across industries.

Our project seeks to address this gap by developing a framework that systematically weights financial news sentiment based on a company's market influence. The goal is to identify which stocks contribute most to index fluctuations and to fine-tune sentiment aggregation for better stock index predictions. Given that a one-size-fits-all approach may not be effective across different industries, a customized model could yield better results for specific sectors.

Analytical Objectives Explanation

The goal of this research is to assess the actual performance gains of industry-specific models over a general model and determine whether the additional investment in industry-level fine-tuning is justified. The findings will provide valuable insights for financial analysts and investors when selecting the most effective AI-driven approach for market analysis.

Comparative Evaluation

This evaluation will determine whether a broad, general model is sufficient for forecasting stock prices, or if sector-specific sentiment models provide superior predictive power for financial decision-making.

While industry-specific models are generally expected to offer improved performance, this is not always guaranteed. If a general model effectively captures overall market sentiment, it may provide similar accuracy without the need for industry-specific customization.

Phase 1: General Model Development

The first phase involves the development of a general sentiment analysis model by aggregating sentiment scores from Bloomberg financial news headlines, covering a variety of industries. Deep learning models such as Lexicon, BERT, and FinBERT are employed to analyze sentiment scores and predict stock price movements based on overall market news. This general model serves as a baseline for evaluating the effectiveness of a broad, generalized sentiment analysis approach.

Phase 2: Industry-Specific Model Development

In the second phase, the same sentiment analysis models (LSTM, BERT, etc.) from Phase 1 are applied to industry-specific datasets in order to refine sentiment score predictions tailored to individual sectors. By fine-tuning models for financial news specific to each industry, the goal is to capture more precise sentiment that reflects market reactions unique to each sector. The selected industry-specific analyses are as follows:

1. Technology Sector Sentiment Analysis (Alice)
2. Energy Sector Sentiment Analysis (An)
3. Finance Sentiment Analysis (Kim)
4. Healthcare Financial News Sentiment Analysis (Cece)
5. Real Estates Industry Financial News Sentiment Analysis (AiLien)
6. Media & Entertainment Sector Sentiment Analysis (Kiet)

Each industry-specific model is expected to improve the accuracy of sentiment score aggregation and stock price predictions within its respective sector.

CHAPTER 3
DATA AND PROPOSED METHODS

Data and Sampling

The dataset for this study was sourced from NASDAQ, a leading stock exchange, using the Bloomberg Terminal and its Excel integration to extract historical financial data. This dataset includes key financial metrics such as stock prices, earnings per share (EPS), and trading volumes for NASDAQ-listed companies. The extracted data was then exported in a structured CSV format for further analysis.

The original NASDAQ dataset consists of 15,550,000 rows and 3 columns, including Date, Stock_symbol, Article_title. For this study, we focused on financial data from 2018, extracting a refined subset containing 696,948 rows and 3 columns. Each row represents an individual financial record, capturing stock price movements, trading volumes, and sentiment scores derived from financial news related to NASDAQ-listed companies.

The selection of 2018 as the study period was based on three key factors:

1. Computational Efficiency – Processing large-scale, multi-year financial data is computationally expensive. Narrowing the scope to a single year optimizes data processing, model training, and analytical performance.
2. Data Manageability – Handling extensive financial datasets requires significant storage and processing power. By limiting the dataset to one year, we ensure a more structured and efficient analysis.
3. Market Stability – Choosing a pre-pandemic year (2018) ensures that the data is not affected by the extreme volatility introduced by the COVID-19 pandemic, allowing for a more stable and reliable financial analysis.

This structured approach ensures that the dataset is both comprehensive and manageable, leveraging NASDAQ's historical financial data while optimizing computational resources for efficient analysis.

Data Cleaning & Processing Steps

Data cleaning

After obtaining the full dataset for 2018, the following steps were performed:

- Since the *Stock_symbol* column had *653,055* non-null values out of a total of 696,948 entries, the missing data accounted for approximately *6.3%* of the dataset. Given this relatively small proportion, these rows were removed to ensure data relevance without significantly affecting the overall analysis (Table 1).
- The *Date column* was then converted to a standard date format for consistency, and further refined by extracting only the date portion, excluding the time component.

#	Column	Non-Null Count	D_type
0	Date	653,055	object
1	Article title	653,055	object
2	Stock_symbol	653,055	object

Table 1: Summary of Processed Dataset After Cleaning

Data visualization

To easily understand the 2018 dataset, we conducted the line chart showing the number of articles published per month (Figure 5). The x-axis represents the months of 2018, while the y-axis shows the count of published articles. Key observations are:

- The number of articles fluctuates throughout the year, with noticeable peaks around May, August, and October.
- Some months, such as March, June, and December, show a decline in article count.

- The highest number of articles was published in August and October of around 67,000 articles per month, while June had the lowest count of 46,000 articles.

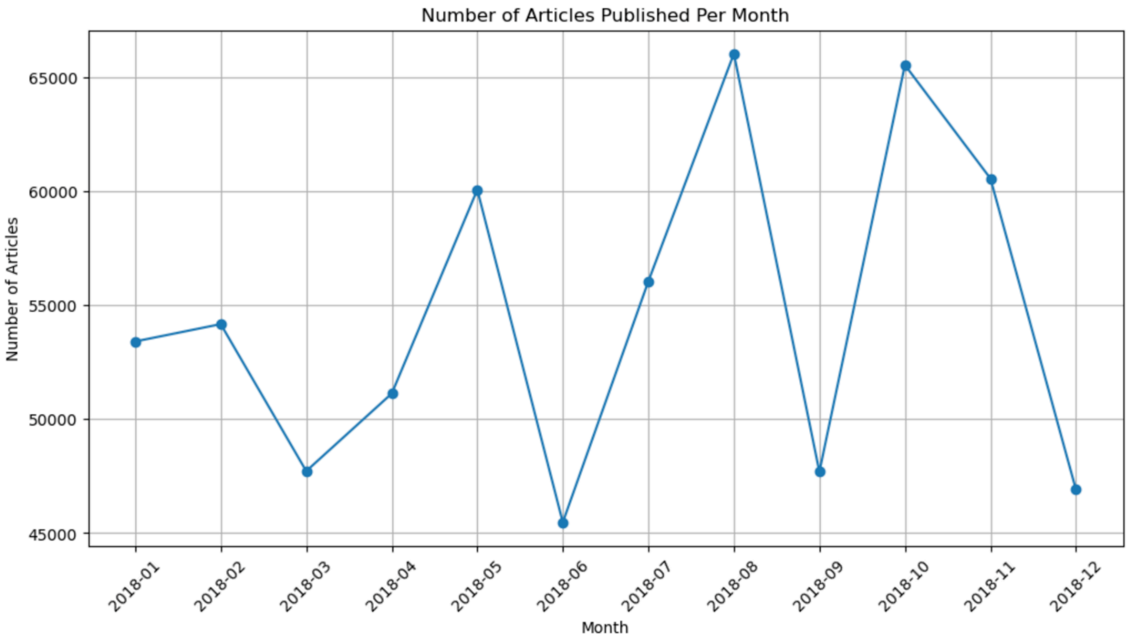

Figure 5: 2018 article summary by month

Additionally, we also conducted the bar chart to display the top 20 most mentioned stock symbols in the dataset (Figure 6). The x-axis represents different stock symbols, while the y-axis shows the number of times each stock was mentioned in news articles. Key observations are:

- SPY has the highest number of mentions, indicating significant news coverage.
- Other frequently mentioned stocks include INTC (Intel), MU (Micron), and BABA (Alibaba), suggesting high market interest.

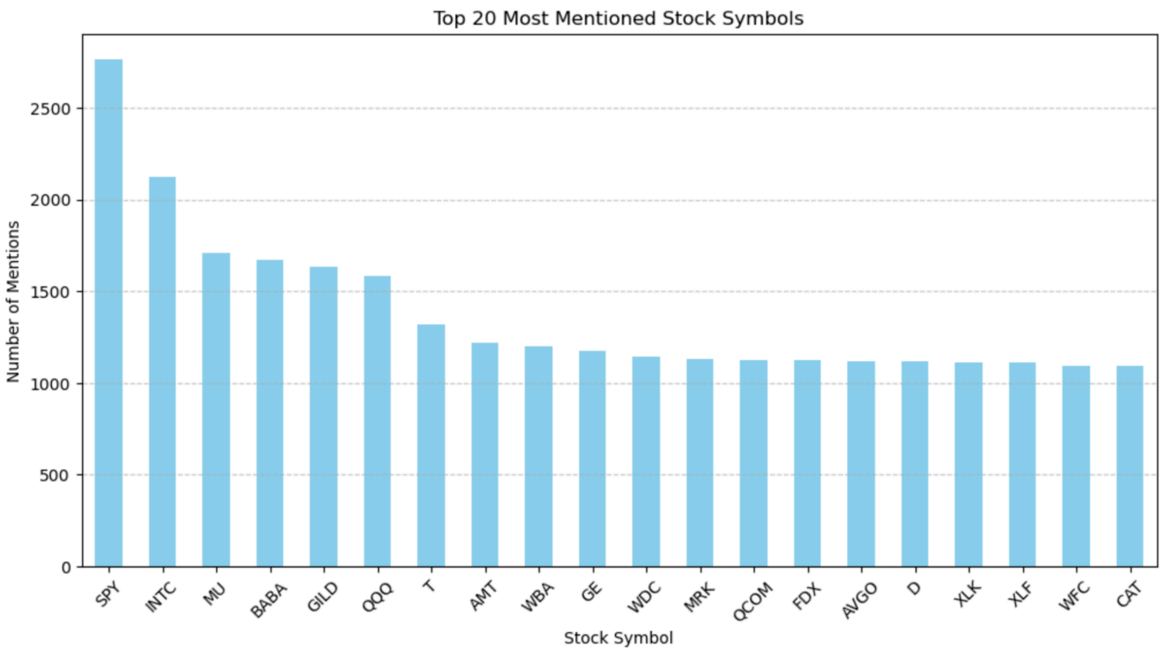

Figure 6: Top 20 stock symbols article

Phase 1

The dataset was processed at a daily level, where each column represents the sentiment score for a specific stock system on a given day. This structured format ensures consistency across different sentiment analysis approaches and allows for easier comparison of results.

To generate sentiment scores, multiple methods were applied, including lexicon-based approaches, general deep learning models, and finance-specific models. Each team member processed the dataset using their chosen model, applied pre-processing techniques, and computed sentiment scores accordingly. Below are the individual methodologies used.

Lexicon-based Sentiment Analysis: TextBlob

In this study, the *TextBlob* model will be employed for sentiment analysis to establish an initial baseline for evaluating financial data. Although the focus is on the financial domain, a simple and accessible approach like TextBlob has been selected over more complex models to ensure clarity and comprehensibility during the early stages.

With the *dataset comprising all financial headlines for 2018*, the TextBlob model will be utilized to calculate the average sentiment score for each company (ticker) on a daily basis, derived from all news items related to the company on that specific day.

Then the *historical stock dataset for 2018* was also extracted. A new column, *Price_Change,* was calculated as the difference between the Close/Last and Open prices (Price_Change = Close/Last - Open) to reflect daily stock price movements. This dataset was subsequently merged with the sentiment dataset, aligning the stock price changes with corresponding daily sentiment scores to facilitate further analysis.

However, two issues currently arise. Firstly, a ticker will not have sentiment scores for all days, as news related to the respective company is not available every day. Secondly, in the *price_change* column, certain values will be missing corresponding to days when the market was closed, although sentiment scores may still exist for these days.

For stocks exhibiting no recorded movement in terms of scores on a specific day (issue 1), a sentiment score of 0 was assigned to indicate neutrality. For non-trading days (issue 2), such as weekends and holidays, when index price data is unavailable, the average sentiment score was calculated by aggregating the average sentiment scores from these non-trading days along with the next available trading day.

BERT

Using *nlptown/bert-base-multilingual-uncased-sentiment* model from Hugging Face to generate sentiment score. The process involved several key steps:

- Pre-processing: News headlines were tokenized and converted into a format compatible with the sentiment model.
- Sentiment Scoring: Each headline was analyzed using a BERT-based model, which assigned probabilities to five sentiment categories ranging from very negative (-2) to very positive (+2). The final sentiment score was calculated as a weighted average of these probabilities. We also performed the sentiment BERT model ranging from 1-5 where 1 means very negative and 5 means very positive. For the dates with no news and transactions, the sentiment score is set to 0.
- Aggregation: Sentiment scores were grouped by Date and Stock_symbol, and the daily average sentiment score for each stock was computed.
- Data Structuring: The dataset was transformed into a pivot table, where each row represents a date, each column represents a stock symbol, and the values indicate the corresponding sentiment score.

Deep Learning-based Sentiment Analysis (Finance-specific): ProsusAI/finbert

To analyze sentiment in financial news headlines, the *ProsusAI/FinBERT* model was leveraged. This model, fine-tuned for financial sentiment classification, was used to assign sentiment labels and confidence scores to news headlines. The process included the following steps:

- Set up the FinBERT model and tokenizer loaded from Hugging Face (ProsusAI/finbert) to predict sentiment labels and their associated probabilities.
- Tokenize each news headline and pass it through the FinBERT model.
- Compute softmax probabilities over three sentiment classes: positive, negative, neutral.

- Extract the confidence score (maximum probability). The sentiment label and sentiment confidence score were computed for each headline.
- Aggregating Sentiment Scores: Neutral sentiment labels present a challenge because assigning them a fixed value of 0 may oversimplify the analysis, as a high-confidence neutral classification (e.g., 0.93) is significantly different from a low-confidence neutral classification (e.g., 0.35). When a confidence score is low means the model sees some likelihood that the sentiment could be either positive or negative, but none of the categories dominate strongly.
- Due to the importance of preserving neutral sentiment, both the confidence scores and sentiment labels are retained individually, without aggregating all news articles for each company into a single daily sentiment score.

Expected Outcome

Although multiple models were utilized, the final results will align with those presented in Figure 7A: Sample Output. These results will also serve as inputs for the development of the general model, with the exception of the FinBERT model.

date	price_change	A	ABBV	ABT	ACGL	ACN	ADBE	ADI
2018-01-02	12.08	0.0	0.5	0.25	-0.09375	0.0	0.5	0.14285714285714285
2018-01-03	15.21	0.0	0.0	0.5	0.0	0.0	0.0	0.17142857142857143
2018-01-04	4.68	0.0	0.040340909090909094	0.0	0.25	0.10714285714285714	0.0	0.0
2018-01-05	11.82	0.0	0.10606060606060606	0.16666666666666666	0.0	0.0	0.02638888888888889	0.0
2018-01-08	5.04	0.0	0.0	0.0	0.0	0.0	0.0	0.05740740740740741
2018-01-09	0.14	0.0	0.26666666666666666	0.0	0.0	0.0	0.8	0.09999999999999999
2018-01-10	2.68	0.0	0.1	0.0	0.0	0.0	0.0	-0.016666666666666666
2018-01-11	14.59	0.0	0.018518518518518517	0.1111111111111111	0.0	0.0	0.05333333333333334	0.2
2018-01-12	16.06	0.0	0.125	0.11944444444444445	0.0	0.0	0.0625	0.0
2018-01-16	-22.54	0.0	0.0	0.4333333333333333	0.0	0.0	0.0	-0.0125
2018-01-17	17.57	0.0	0.25	0.05000000000000002	0.0	0.0	0.0	-0.0035714285714285704
2018-01-18	-4.37	0.0	0.21606060606060606	0.0	0.0	0.09999999999999999	0.0	0.0011616161616161524
2018-01-19	7.7	0.0	0.14583333333333334	0.0	0.0	0.0	0.08333333333333333	0.022222222222222223
2018-01-22	23.81	0.0	0.10119047619047618	0.0	0.0	0.0	0.0	-0.13819805194805196
2018-01-23	4.08	0.0	0.08831168831168831	0.30000000000000004	0.0	0.25	0.062	-0.125
2018-01-24	-7.88	0.0	0.2	0.056628787878787876	0.0	0.0	1.0	0.03166666666666667
2018-01-25	-6.99	0.0	0.16666666666666666	0.0	0.0	0.0	0.0	0.0
2018-01-26	25.39	0.0	0.22827777777777777	0.04375000000000001	0.0	0.0	0.0	0.04814814814814814
2018-01-29	-13.7	0.0	0.10920454545454546	0.0	0.0	0.0	0.19642857142857142	0.12848484848484848
2018-01-30	-10.31	0.0	-0.0234375	0.0	0.0	0.0	0.5	-0.07777777777777779
2018-01-31	-8.6	0.0	0.12857142857142856	0.0	0.0	0.0	0.5	0.0
2018-02-01	5.53	0.0	0.3333333333333333	0.0	0.0	0.0	0.0	0.0
2018-02-02	-46.79	0.0	0.15	0.06666666666666667	0.0	0.0	0.0	0.0
2018-02-05	-92.12	0.0	0.16666666666666666	0.0	0.0	0.1	0.0	0.0
2018-02-06	80.36	-0.19090909090909092	0.0	0.0	0.0	0.0	0.0	0.041666666666666664
2018-02-07	-9.29	0.0	0.15833333333333333	0.0	0.0	0.0	0.0	0.2
2018-02-08	-104.01	0.0	0.25	0.0	0.0	0.0	0.0	0.0
2018-02-09	17.77	0.0	0.16666666666666666	0.16666666666666666	0.24444444444444444	0.0	0.0	0.0

Figure 7A: Sample output

Date	Article_title	Stock_symb	Sentiment	Confidence_	Change
1/2/2018	Investors Begin Accumulating St	ATHM	Negative	0.7923	0.009982
1/2/2018	The CDC Looks to Distributed Le	CDC	Negative	0.534541	0.009982
1/2/2018	Tuesday's ETF with Unusual Volu	FXZ	Positive	0.742705	0.009982
1/2/2018	Several Biotechs Higher Tuesday	TEVA	Negative	0.518889	0.009982
1/2/2018	Sidoti & Co. Upgrades National F	NFG	Negative	0.63221	0.009982
1/2/2018	Semtech (SMTC) Down 6.4% Sin	SMTC	Neutral	0.971054	0.009982
1/2/2018	William Blair Upgrades Titan Mac	TITN	Positive	0.545149	0.009982
1/2/2018	UPDATE: Piper Jaffray On Vericel	VCEL	Negative	0.779245	0.009982
1/2/2018	â€˜Pokemon Goâ€™ to Launch i	NTES	Positive	0.722817	0.009982
1/2/2018	Several Biotechs Higher Tuesday	SUPN	Negative	0.518889	0.009982
1/2/2018	Energy/Materials - Top Gainers /	CLMT	Positive	0.809967	0.009982
1/2/2018	SunTrust Lowers Achaogen Price	AKAO	Positive	0.495583	0.009982

Figure 7B: Sample output for FINBERT Model

Deep Learning Model for Feature Importance

In the final step of Phase 1, the most effective model, identified based on predictive performance, is selected for further analysis. This model is then applied in the next stage to enhance the evaluation of sentiment-driven stock movements.

A deep learning model (LSTM) is implemented to examine the relationship between daily sentiment scores and stock price changes, with Index_Change (Price_Change) as the dependent variable and sentiment scores for each ticker as independent variables.

By analyzing *feature importance*, the model determines the extent to which each stock's sentiment score contributes to market movements. Stocks with higher trading volume or broader market influence, such as Amazon, are expected to have a more significant impact compared to smaller firms like Agilent Technologies Inc.

The Table 2 below compares the financial sentiment analysis models used in our study. It highlights key differences in input data, output format, and expected outcomes to help you easily understand and compare each approach.

Model	Type	Input Data	Sentiment Output	Expected Outcome
Textblob	Lexicon-based (rule-based)	Financial news headlines (2018)	Polarity score (-1 to 1)	Daily sentiment scores per ticker; merged with stock price changes
BERT (nlptown)	Lexicon-based (rule-based)	Financial news headlines (2018)	Polarity score (-1 to 1)	Daily sentiment scores per ticker; merged with stock price changes
BERT (nlptown)	Pre-trained Transformer	Financial news headlines (2018)	Weighted average score (-2 to +2)	Daily sentiment scores per ticker; merged with stock price changes
FinBERT (ProsusAI)	Finance-specific Transformer	Financial news headlines (2018)	Label (pos/neg/neutral) + confidence	Daily sentiment scores per ticker; merged with stock price changes

Table 2: Comparison of Financial Sentiment Analysis Models

In the final step, the general sentiment model from Phase 1 is compared with the industry-specific models from Phase 2 to assess whether sector-focused sentiment analysis yields a significant improvement in predictive accuracy. This comparison provides valuable insights into the effectiveness of a broad market sentiment model versus specialized industry models in financial forecasting, while also potentially aiding investors in optimizing and reducing costs when investing in a predictive model.

CHAPTER 4

ANALYTICAL MODEL AND RESULTS

In this chapter, we present the results of our two main analytics objectives.

- Objective 1 focused on developing and selecting the best performing sentiment-LSTM model using both classification and regression evaluation approach.
- Objective 2 applied the selected model to predict stock price changes across six different industry sectors and compared its performance against the best performing sentiment-LSTM model from Objective 1.

Phase 1 - Objective 1

Textblob Sentiment

After extensive modeling and evaluation, the Long Short-Term Memory (LSTM) neural network was selected as the final predictive model. The LSTM achieved a Train RMSE of 0.006888 and a Test RMSE of 0.011552, indicating highly accurate forecasting of daily stock price changes based on sentiment inputs. The model's mean absolute error remained below 1% on unseen data, confirming its strong generalization capability without overfitting.

The decision to prioritize LSTM over Transformer models was grounded in both performance parity and operational efficiency; the LSTM's simpler architecture is better suited for the available data scale (one year of daily sentiment scores) and offers faster, more cost-effective deployment. Preliminary findings demonstrate that even basic sentiment signals, when properly sequenced and modeled, deliver material predictive power, setting a strong foundation for future model enrichment through advanced natural language processing and multi-source data integration.

Metric	Train	Test
MAE	0.004264	0.01164
RMSE	0.011559	0.007399
MSE	0.000050	0.000133

Table 3: Textblob Evaluation Summary Metrics

The model achieved strong predictive performance, with a low MAE of 0.004264 on the training set and 0.01164 on the test set, indicating reliable generalization. RMSE and MSE values remain consistently low across both sets, confirming minimal overfitting and stable error distribution. Overall, the model is well-calibrated and suitable for forecasting tasks with high accuracy.

Metric	Train Set	Test Set
Accuracy	0.6183	0.6981
Precision	0.4545	0.3571
Recall	0.1685	0.4167
F1 Score	0.2459	0.3846

Table 4: Textblob Model Classification Metrics

The model achieved a test accuracy of 69.81%, indicating decent directional prediction. However, precision (0.3571)and recall (0.4167) remain low, resulting in an F1 score of 0.3846. Training metrics were weaker, with recall at 0.1685, suggesting underfitting. Overall, the model shows moderate accuracy but lacks consistency in identifying positive movements.

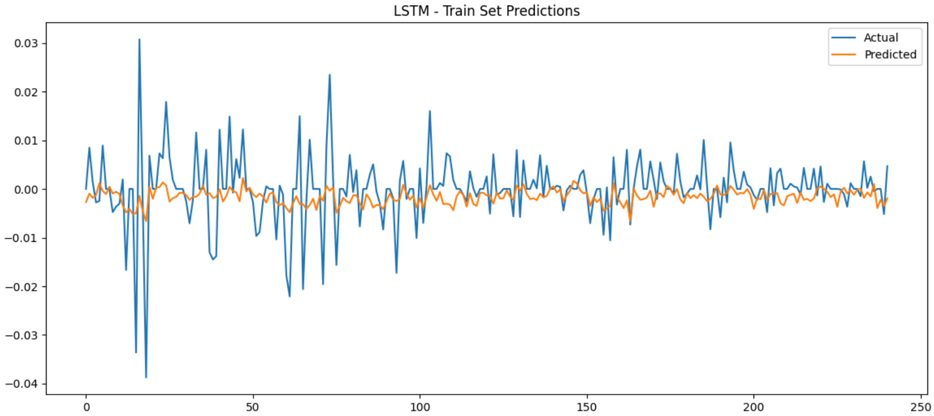

Figure 8A: LSTM Model Predictions on Train Set

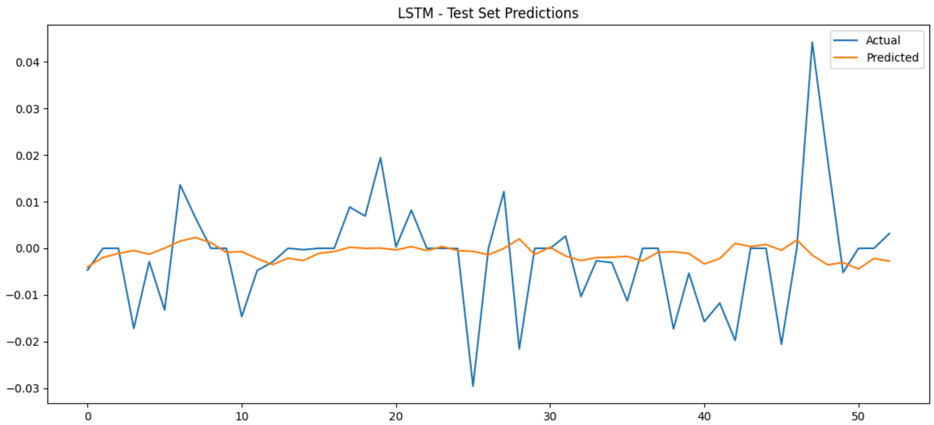

Figure 8B: LSTM Model Predictions on Test Set

The LSTM model exhibits significant underfitting, as shown by the flat predicted line across both train and test sets. While the actual values fluctuate sharply, the model fails to capture this volatility, especially on the test set. The predictions are overly smoothed and centered near zero, indicating limited learning of the temporal price change patterns.

BERT-based Sentiment to Price-Change LSTM

After converting daily sentiment scores into a date-by-ticker matrix, I trained a univariate time-series LSTM that ingests the previous five trading days of standard-scaled sentiment (all tickers) and predicts the next day's Price Change (also scaled). The data were split chronologically (80 % train, 20 % test) to avoid leakage. The network uses three 64-unit LSTM layers with tanh activation and 0.3 dropout, followed by a linear output neuron; optimization is performed with Adam and an MSE loss.

Metric	Train	Test
MAE	0.00240	0.01164
RMSE	0.00520	0.01470
MSE	0.00003	0.00022

Table 5: Evaluation Metrics Summary for BERT model

Table 5 shows that the LSTM fits the training data reasonably well (MAE =0.00240, RMSE = 0.0052), but its test-set errors are several times larger (MAE = 0.01164, RMSE = 0.01470). This wide gap signals over-fitting: the model captures past patterns yet fails to generalise to unseen days, as underlined by a test MSE above 0.00022, nearly ten times the training MSE. While these early results confirm that headline sentiment contains predictive information, the out-of-sample error remains too high; the model therefore requires further refinement: simpler architecture, stronger regularisation, or a longer look-back window to reach acceptable performance.

Metric	Train Set	Test Set
Accuracy	0.84694	0.50000
Precision	0.86598	0.38889
Recall	0.83168	0.33333
F1 Score	0.84848	0.35897

Table 6: BERT-LSTM Model Classification Metrics

While table 6 shows strong performance on the training set (Accuracy = 84.7%, F1 Score = 84.8%), its results on the test set drop significantly (Accuracy = 50.0%, F1 Score = 35.9%). The low test precision and recall indicate that the model struggles to correctly classify price changes it has not seen before. This large gap between training and testing performance suggests that the model is overfitting; it learns patterns from the training data but cannot generalize well to unseen data.

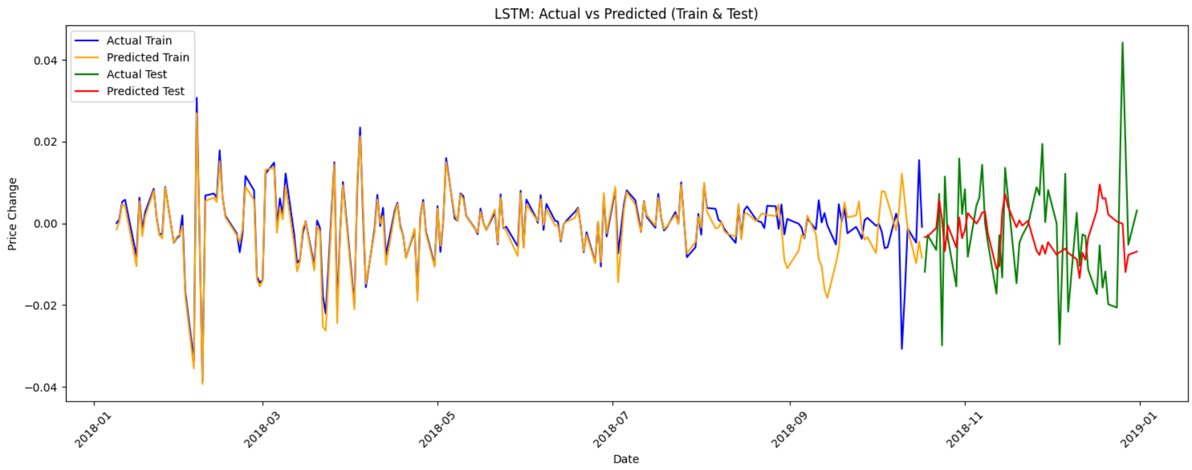

Figure 9: Actual vs Predicted Daily Price Change (Train vs Test)

As can be seen from figure 9, the model accurately captures general trends and fluctuations during the training period, closely aligning predicted values with actual price changes. However, during the test phase, predictions noticeably diverge from actual movements, especially when confronted with extreme fluctuations or volatility spikes.

BERT-based Sentiment

In this project, I developed a stacked LSTM model to predict the next-day price change of the Nasdaq index based on aggregated sentiment analysis data and daily index prices. The dataset was prepared by merging Nasdaq index trading data (Close/Last and Index Price) with daily averaged sentiment scores collected from individual stock news throughout 2018. After scaling both the input features and the target variable (Price Change, calculated by Close/Last minus Index Price and divided by Index Price).

Metric	Train Set	Test Set
MAE	0.004695	0.007848
MSE	0.00040	0.000132
RMSE	0.006326	0.011487
R^2 Score	0.428522	-0.028927

Table 7: BERT-LSTM Model Regression Metrics

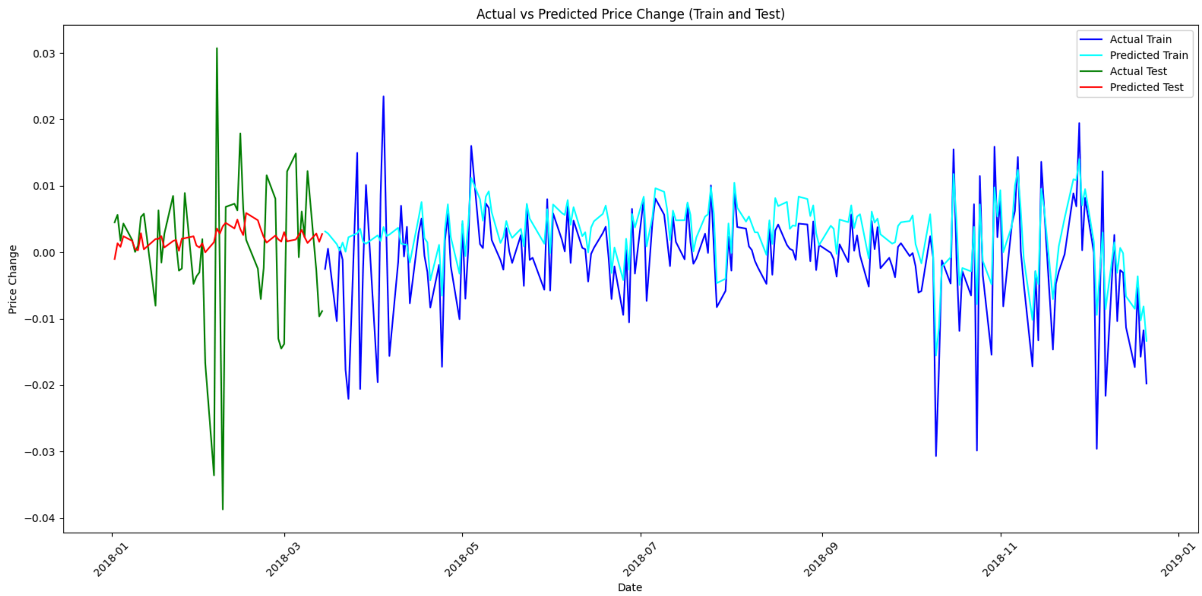

Figure 10A: Actual vs Predicted Daily Price Change graph of the BERT model

The model performed reasonably well on the training set, achieving a Root Mean Squared Error (RMSE) of 1,111.94 and an R^2 score of 0.776, indicating that it was able to explain approximately 77% of the variance in the training data. However, when applied to the test set, the model's performance dropped significantly, with a Test RMSE of 2,833.14 and a negative R^2 score of -0.045. This suggests that while the model could learn patterns from the training data, it struggled to generalize to unseen data, likely due to the high volatility and noise inherent in financial time series data, as well as the limited explanatory power of sentiment alone.

We chose to present the first BERT-based LSTM model to compare with Texblob_LSTM and FinBert-LSTM because it had a clearer structure, more balanced evaluation, and better interpretability. It used a consistent date-by-ticker sentiment setup and reported both regression and classification metrics, showing strong training performance and offering insights into both price direction and magnitude. While both models showed signs of overfitting, the second BERT model had a negative test R^2 and less clarity around how sentiment was linked to price changes. Overall, the first model provides a stronger foundation for explaining our approach and results.

FinBERT-Based Sentiment and LSTM Modeling to Predict Stock Price Change

Sentiment scores and confidence levels for financial news headlines were first generated using the FinBERT model and subsequently merged with Nasdaq trading data. The daily price change was recalculated as a percentage, defined by the formula (Close – Open) / Open. Then, a regression-based LSTM model was trained to predict the same day's stock price change, using the textual sentiment, confidence scores, and historical pricing information. Afterward, the model's performance was evaluated using key regression metrics to assess its prediction accuracy and generalization ability.

In terms of regression metrics, the model achieved a very low training MAE of 0.0075 and a test MAE of 0.0079. This low MAE indicates that, on average, the model's predicted stock price changes differ from the actual changes by less than 0.8%, reflecting high predictive accuracy.

The small gap between training and testing MAE values suggests that the model generalizes well to unseen data, with minimal signs of overfitting.The model also achieved a training RMSE of 0.0109 and a test RMSE of 0.0112 which are very small values. This means the model's predictions are very close to the real stock price changes, with few big mistakes.

Metric (Finbert/LSTM)	Train Set	Test Set
MAE (Mean Absolute Error)	0.0075	0.007887
MSE (Mean Squared Error)	0.000118	0.000125
RMSE (Root Mean Squared Error)	0.0109	0.011176

Table 8: FINBERT-LSTM Model Regression Metrics

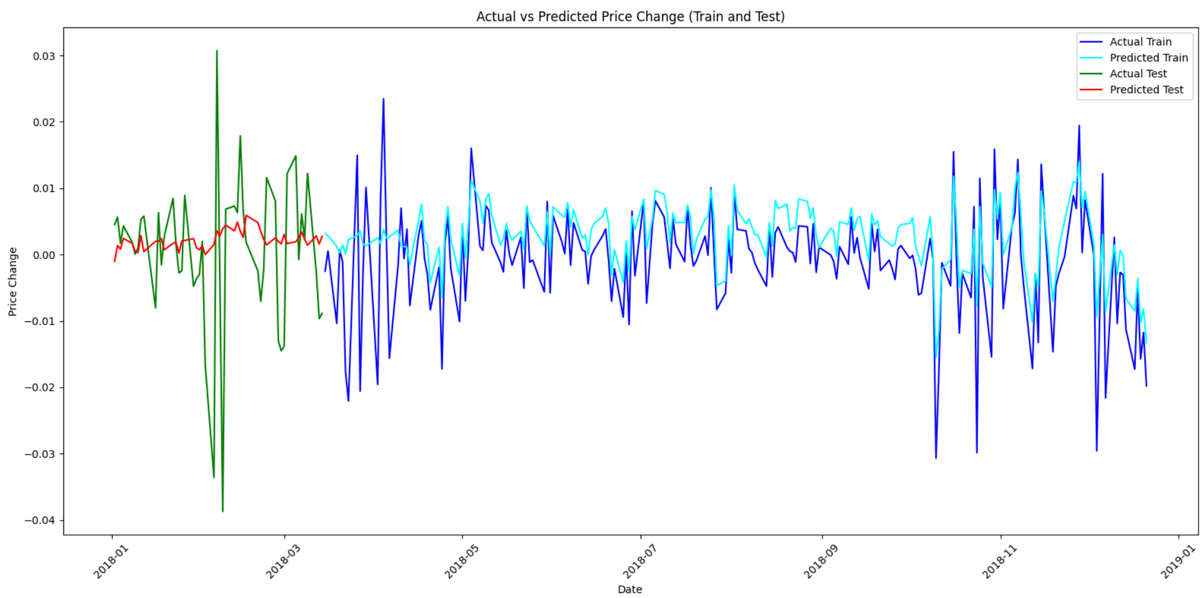

Figure 10B: Actual vs Predicted Daily Price Change graph of the FinBERT-LSTM model for Nasdaq

However, we realized that the way sentiment scores were aggregated in our FinBERT-based model differs from how other models handle aggregation. To better align the evaluation and ensure a fair comparison, we decided to conduct another test using accuracy as the performance metric.

As shown in table *9:* FINBERT-LSTM Model Regression Metrics, the FinBERT-LSTM model achieved a strong training accuracy of 82.67% and a test accuracy of 70.44%. Precision, recall, and F1 score on the training set were all above 82%, indicating that the model learned the training data well. Although the performance on the test set dropped slightly — with precision, recall, and F1 score around 70–72%, the results still suggest reasonable generalization to unseen data.

Metric (Finbert/LSTM)	Train Set	Test Set
Accuracy	0.82667	0.70442
Precision	0.82387	0.70352
Recall	0.84027	0.72360
F1 Score	0.83199	0.71342

Table 9: FINBERT-LSTM Model Classification Metrics

Conclusion of Phase 1: Performance Evaluation of Sentiment Models

In this study, we evaluated three sentiment-driven forecasting models—TextBlob-LSTM, BERT-base-LSTM, and FinBERT-LSTM—across classification and regression tasks. Among them, the FinBERT-LSTM model demonstrated the most robust and consistent performance, outperforming both alternatives in terms of accuracy, precision, recall, and F1 score on both training and testing sets. Specifically, it achieved a test F1 score of 0.72 and test accuracy of 70.44%, with minimal decline from its strong training performance (F1: 0.83, Accuracy: 82.67%).

On the regression side, FinBERT-LSTM also delivered the lowest error rates (MAE: 0.007887, RMSE: 0.011176), with a small gap between training and testing errors, suggesting strong generalization and low overfitting. In contrast, TextBlob-LSTM, while faster and simpler, underperformed in test precision and recall, and BERT-base-LSTM showed signs of significant overfitting—achieving strong training metrics but deteriorating sharply on unseen data.

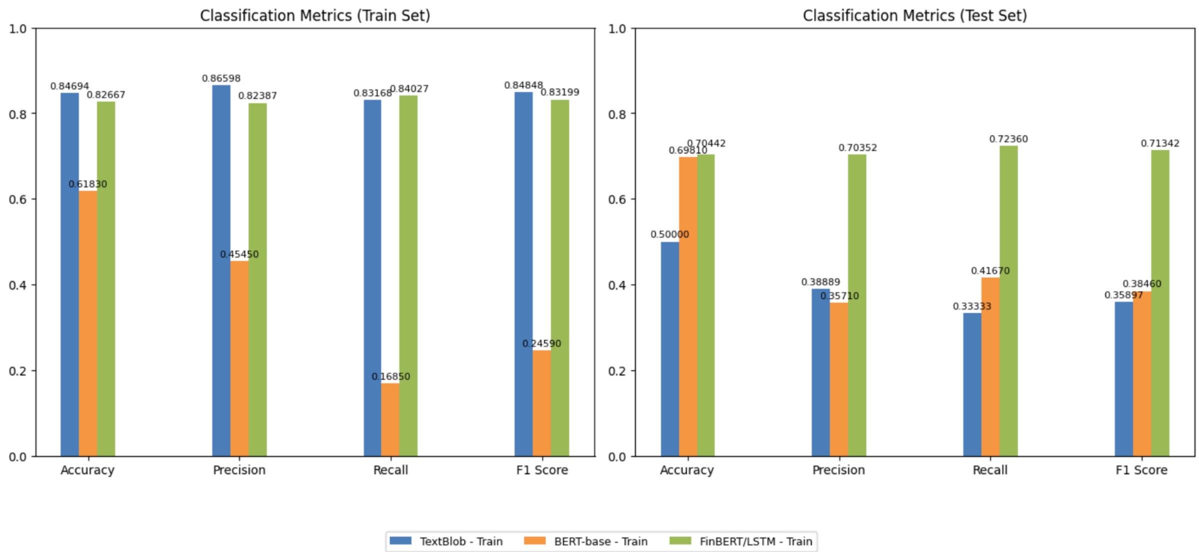

Figure 11: Classification Metrics Comparison

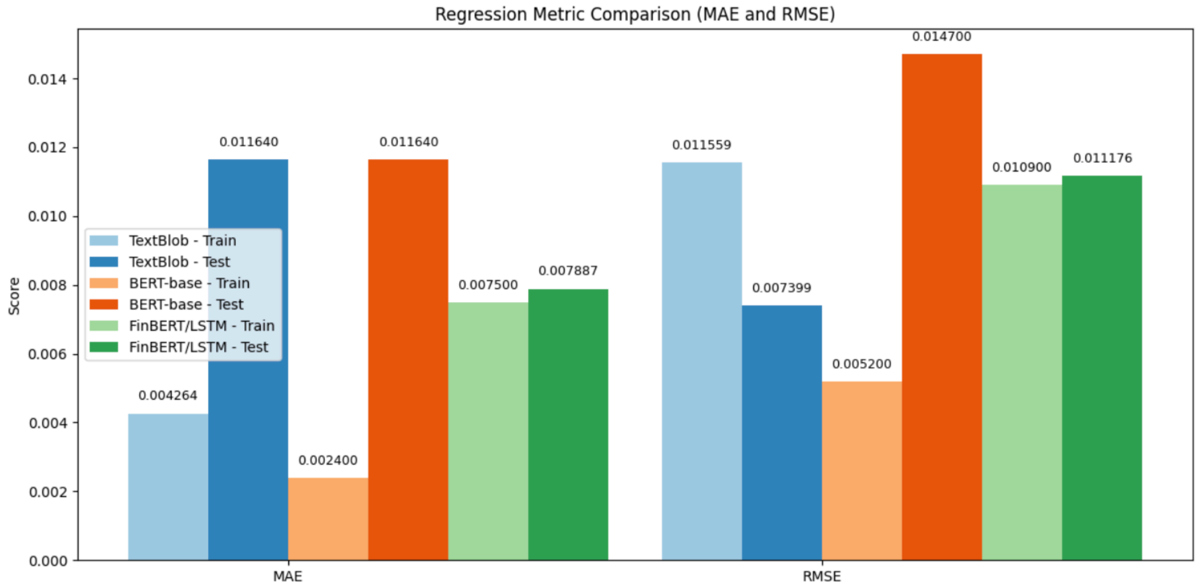

Figure 12: Regression Metrics Comparison

These findings establish FinBERT-LSTM as a more reliable and effective model for sentiment-based financial forecasting. Therefore, we have selected the FinBERT-based model for Phase 2, where it will be applied to industry-specific datasets. This next step will help assess whether FinBERT's generalized sentiment knowledge can accurately capture domain-specific financial signals or if further fine-tuning is needed for certain sectors.

Phase 2 - Objective 2

1. FinBERT Sentiment Model Applied to Technology Sector (Alice)

Metric	Train Set	Test Set
MAE (Mean Absolute Error)	0.005380	0.017249
MSE (Mean Squared Error)	0.000054	0.000505
RMSE (Root Mean Squared Error)	0.007379	0.022482

Table 10: Evaluation Metrics for Regression Model (Technology Industry)

As can be seen from Table 10, the revised model delivers an in-sample RMSE of 0.00738, equivalent to an average daily forecast error of just 0.54% and an out-of-sample RMSE of 0.02248, or roughly 2.25% error on unseen days. Its MAE rises from 0.00538 in training to 0.01725 in testing, and MSE increases from 0.000054 to 0.000505. These figures indicate that while the model is highly accurate for ordinary day-to-day price movements (predictions typically fall within a few tenths of a percent), it still struggles to fully capture the largest swings, which inflate the out-of-sample error to around two percent.

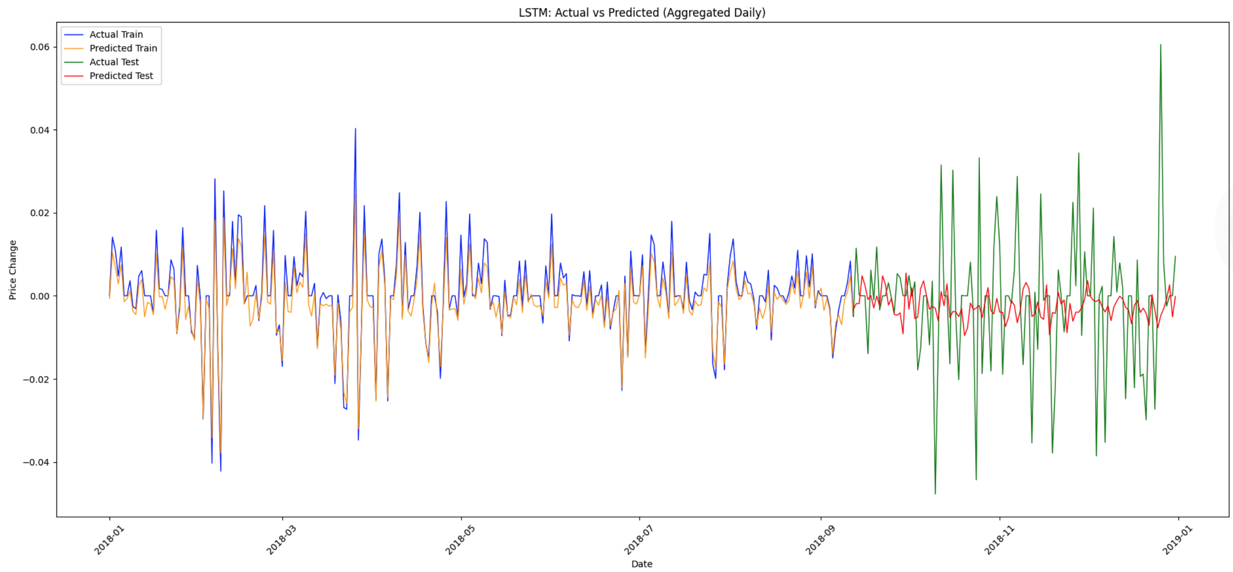

Figure 13: Actual vs Predicted Daily Price Change (Train vs Test) for technology sector

The Actual vs. The predicted plot in figure 13 reveals that the model reliably captures the overall mean-reverting behavior of daily price changes, its predicted curve oscillates around zero in step with the general upward-and-downward drift of the actual series. However, it consistently fails to reproduce the pronounced spikes and troughs that characterize the most extreme days. While moderate moves of ±1–2% are mirrored reasonably well, the largest swings (±3–5%) are dampened in the prediction, resulting in a visibly smoother red (test) line compared to the jagged green (actual) peaks and valleys. In practice, this means the model can serve as a solid baseline for routine price fluctuations, but an additional mechanism, such as a volatility-based feature, an outlier detection flag, or weighted loss targeting extreme values will be necessary to flag and better forecast those critical, high-impact days.

Comparative Analysis: BERT‑based Sentiment to Price‑Change LSTM on All Industries vs. FinBERT Sentiment Model Applied to Technology Sector

The FinBERT‑driven pipeline trained on a broad, cross‑industry corpus delivers far more consistent out‑of‑sample accuracy than the version fine-tuned only on technology headlines. Across all sectors, the combined FinBERT+LSTM model achieves a training MAE of 0.0075 and a nearly identical test MAE of 0.0079, with RMSE rising only slightly from 0.0109 in‑sample to 0.0112 out‑of‑sample. By contrast, the technology-sector variant—even though it posts a lower in‑sample MAE of 0.0054—sees its test MAE balloon to 0.0173 and its RMSE jump from 0.0074 to 0.0225.

In practical terms, the cross‑industry model's error remains under 1% on new data, whereas the tech‑only approach more than doubles its average forecast error when faced with unseen price movements. This pattern indicates that the larger, more diverse training set helps FinBERT capture general sentiment–price relationships without overfitting. Restricting training to a narrow, technology‑specific corpus may yield a slightly better fit on known examples but sacrifices robust generalization. Therefore, for reliable price‑change forecasting—even in the technology sector—it is advisable to leverage the full, general‑industry FinBERT model rather than training exclusively on tech headlines.

2. FinBERT Sentiment Model Applied to Real Estate Industry Sector (Ailien)

This study applied the FinBERT sentiment model to financial news related to companies in the Real Estate Industry Index. Each article was classified as Positive, Neutral, or Negative, and paired with a confidence score. These sentimental values were aggregated daily and compared against corresponding daily price changes in the index.

Metrics	Train Set	Test Set
MAE (Mean Absolute Error)	0.6973	1.0700
MSE (Mean Squared Error)	0.8014	1.9311
RMSE (Root Mean Squared Error)	0.8952	1.3897

Table 11A: Evaluation Metrics for Regression Model (Real Estate Industry)

The revised LSTM model powered by FinBERT sentiment analysis yields an in-sample RMSE of 0.8952—equivalent to an average daily forecast error of 0.89%—and an out-of-sample RMSE of 1.3897, or approximately 1.39% error on unseen days. Its MAE rises from 0.6973 in training to 1.0700 in testing, and MSE increases from 0.8014 to 1.9311 in table 11A. These suggest that the model effectively captures most routine price changes in the real estate industry but tends to underpredict the largest movements, which result in inflated error during extreme volatility. Nonetheless, the model performs well in capturing the general drift and small-to-moderate daily changes.

The figure 14 below depicts the final 50 observations from the training set and the first 50 from the testing set. During the training phase, the model demonstrates exceptional alignment with real-world values. The predicted curve (orange) closely tracks the actual daily price changes (blue), capturing both direction and magnitude with minimal deviation. This tight overlap reflects a strong in-sample fit and supports the previously reported low training MAE and RMSE scores (0.6973 and 0.8952, respectively).

However, the transition to the test set reveals limitations in generalization. While the predicted values (red) maintain general directional accuracy with the actual test data (green), they tend to understate large price movements. For instance, extreme upward and downward spikes present in the actual data are visibly smoothed in the model's output. This behavior suggests that the model, optimized on average error loss functions like MAE and MSE, tends to dampen volatility rather than capture its full scale. Still, the model performs reasonably well in predicting typical daily fluctuations within a ±1–2% range.

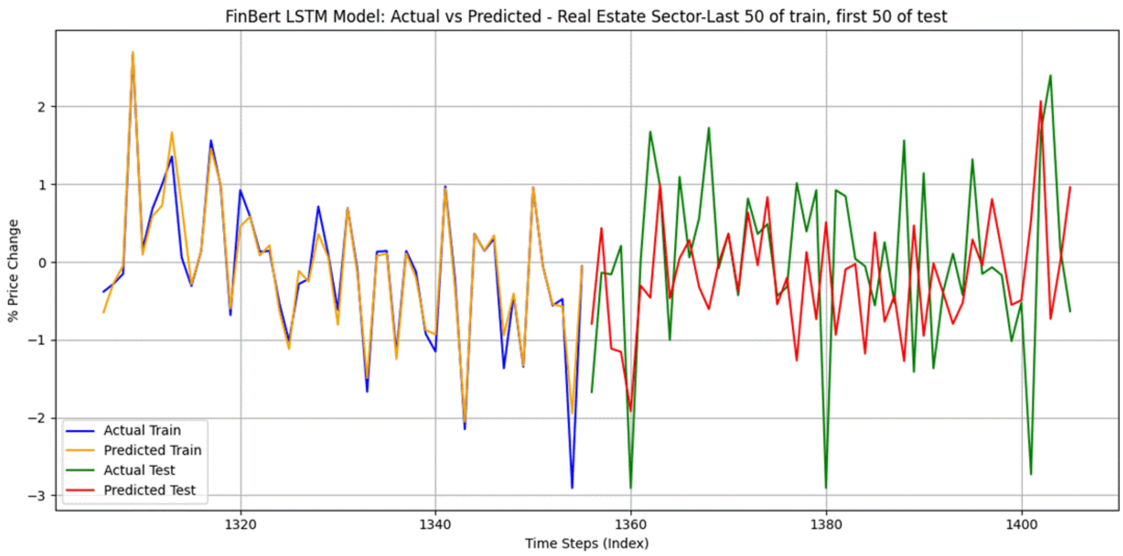

Figure 14: Actual vs Predicted Real Estate Sector– Last 50 Train, first 50 Test Observations

The figure 15 below expands the horizon, showing the last 100 days of the training set followed by the entire test period. In the training segment, the model continues to replicate market behavior with high fidelity. The predicted line sharply mirrors the actual movements, including abrupt reversals and spikes. This reinforces the notion that the model is effectively learning patterns from historical data and is not merely overfitting to noise.

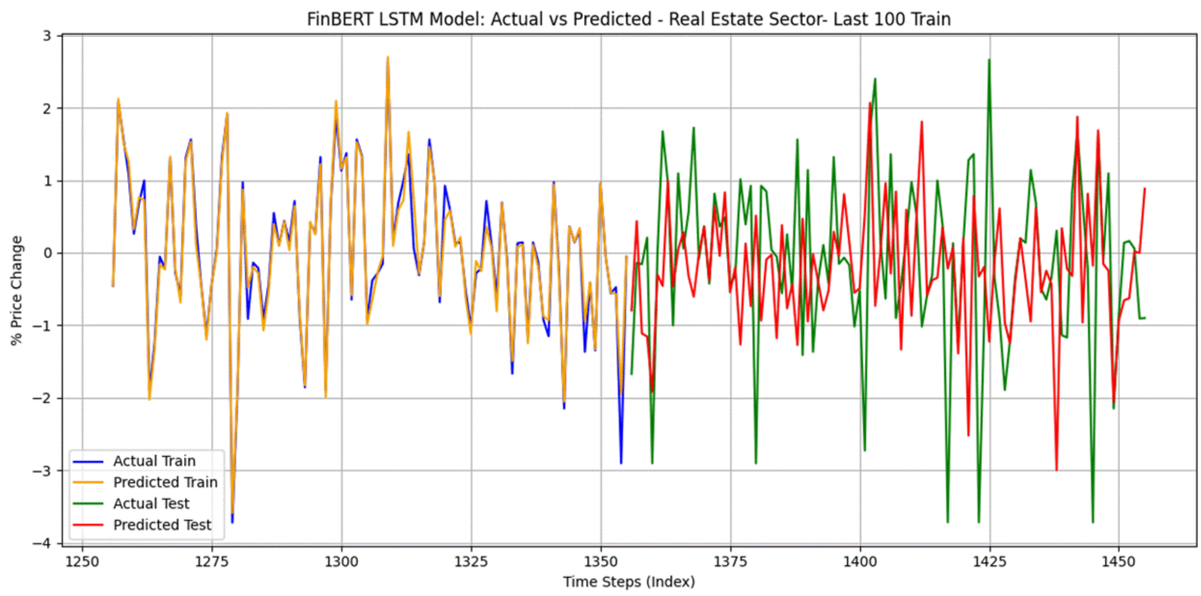

Figure 15: Actual vs Predicted Real Estate Sector – Last 100 Train, All Test Observations

For interpretation and recommendations: Together, these visualizations highlight both the strengths and weaknesses of the FinBERT-LSTM approach. On the one hand, the model excels in capturing trend directionality and moderate market fluctuations, making it a robust baseline for daily forecasting. On the other hand, its inability to reflect extreme sentiment-driven market shocks—common in real estate during macroeconomic or regulatory news events—limits its utility in high-volatile conditions.

To enhance predictive power, future iterations of the model could benefit from:

- **Volatility-aware features**, such as rolling standard deviation or external market uncertainty indicators
- **Attention mechanisms** that emphasize recent sentiment shifts
- **Hybrid ensemble models** combining LSTM with spike-sensitive algorithms like XGBoost.

In summary, the FinBERT-LSTM pipeline delivers consistent, reliable forecasts for routine market behavior in the real estate sector but requires architectural or feature-level enhancements to better anticipate and respond to significant market disruptions.

Comparative Analysis: FinBERT Model on All Industries vs. Real Estate Sector

Metric	Real Estate Sector Train	Real Estate Sector Test	All Industries Train	All Industries Test
MAE	0.6973	1.0700	0.0075	0.007887
MSE	0.8014	1.9311	0.000118	0.000125
RMSE	0.8952	1.3897	0.0109	0.011176

Table 11B: Evaluation Metrics for Regression Model All Industries vs. Real Estate

The FinBERT-based LSTM model, when trained on a broad cross-industry dataset, demonstrates greater consistency and generalizability compared to the same model applied specifically to the real estate sector. Across all industries, the model achieves a lower and more balanced error profile between training and testing datasets, suggesting that the diversity of sentiment patterns across multiple sectors enhances its ability to generalize.

By contrast, the real estate–focused model exhibits strong performance in the training phase (MAE: 0.6973, RMSE: 0.8952) but sees noticeable degradation in the testing phase (MAE: 1.0700, RMSE: 1.3897) in table 11B. This suggests that while the model is able to learn sentiment-price dynamics specific to the real estate industry, it is more vulnerable to overfitting and underperforms when applied to new or unexpected market scenarios.

In summary, the general FinBERT+LSTM pipeline trained on all industries offers more stable out-of-sample performance, while the real estate–specific variant may capture more domain-specific nuances but lacks robustness. This tradeoff highlights the importance of dataset diversity when training sentiment-based forecasting models for financial applications.

3. FinBERT Sentiment Model Applied to Energy Sector (An Vu)

Using the same method, we applied the FinBERT-LSTM model to the energy index, which includes 30 major energy companies such as Chevron, ConocoPhillips, and Occidental Petroleum, to predict same-day percentage price changes. As shown in Table 12, the MAE and RMSE values for both the training and test sets are closely aligned, indicating that the model does not suffer from significant overfitting or underfitting. This demonstrates the model’s strong generalization capability when applied to unseen data. Furthermore, the relatively low error metrics suggest that the model performs effectively on energy sector data. Notably, Figure 16 illustrates that the predicted values closely track the actual values, highlighting the model’s strong forecasting accuracy and practical applicability.

Metrics	Train Set	Test Set
MAE	0.009567	0.010177
MSE	0.000161	0.000184
RMSE	0.012700	0.013553

Table 12: Evaluation Metrics for Regression Model (Energy Sector)

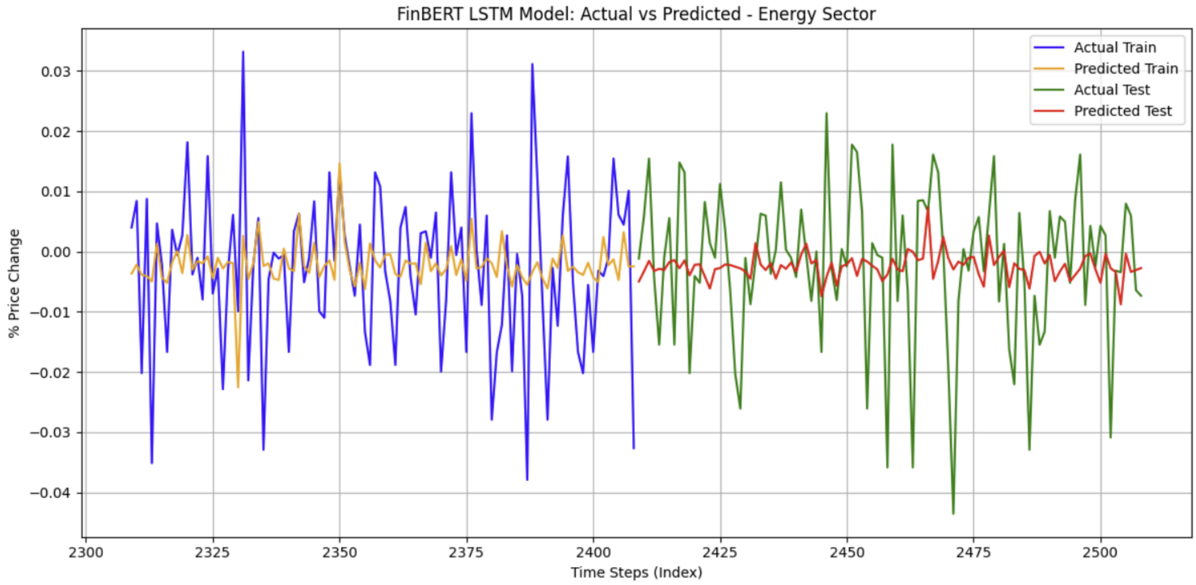

Figure 16: Actual vs Predicted Percentage Price Change (Train vs Test) for Energy Sector

Comparative Analysis: FinBERT Model on All Industries vs Energy Sector

While the FinBERT model demonstrates strong forecasting accuracy and practical applicability in the energy sector, it performs even better on the Nasdaq dataset as shown in Table 13. The model achieves lower values across all evaluation metrics. In particular, the lower RMSE values in the Nasdaq dataset suggest more accurate and stable predictions across a broader range of companies. This improved performance may be attributed to the dataset's greater diversity and clearer overall trends, which likely allow the model to learn more effectively.

Metric	Energy Sector Train	Energy Sector Test	All Industries Train	All Industries Test
MAE	0.009567	0.010177	0.0075	0.007887
MSE	0.000161	0.000184	0.000118	0.000125
RMSE	0.012700	0.013553	0.0109	0.011176

Table 13: FinBERT-LSTM Model Performance – Energy Sector vs. All Industries

4. LSTM Model Applied to Finance (Kim)

The finance sector was extracted from 2018 news articles by identifying and filtering content that mentioned any of the 73 S&P 500 financial tickers, such as JPM, BAC, AIG, MA, etc. This subset of articles was used to isolate sentiment specific to the finance industry for further modeling and analysis. After applying the general model, FinBERT, we have the actual vs. predicted stock indexes as follows (see Figure 17).

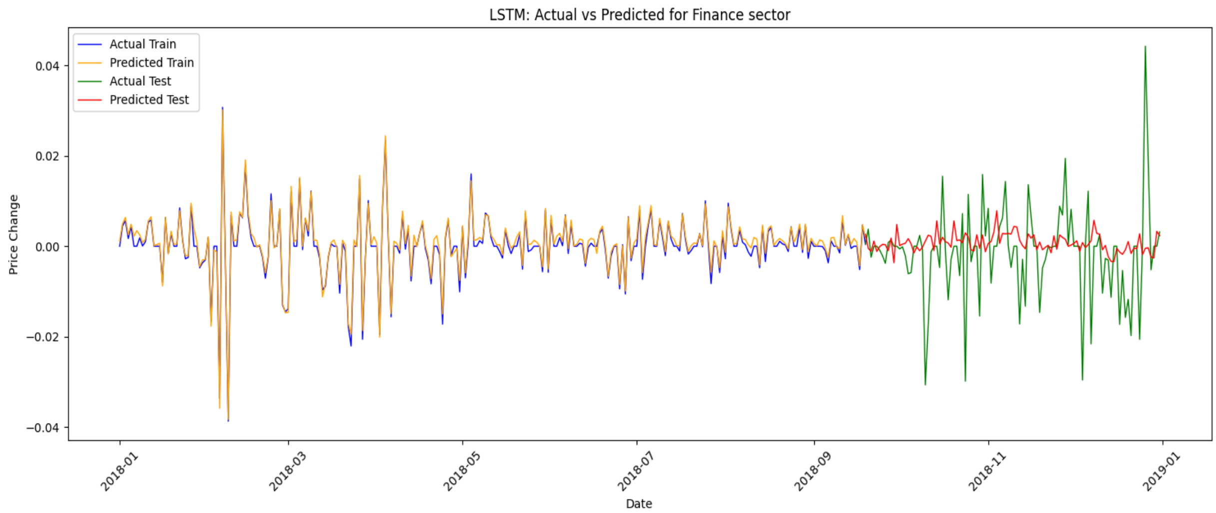

Figure 17: Actual vs Predicted Percentage Price Change (Train vs Test) for Finance Sector

Comparative Analysis: FinBERT Model on All Industries vs Finance Sector

After applying the general sentiment-based LSTM model to the finance sector, the evaluation metrics reveal a clear performance gap between the training and test sets. On the training set, the model achieved a very low MAE of 0.00234 and a RMSE of 0.00343, indicating that it fits the training data exceptionally well. However, the test set results from Table 14 show a higher MAE of 0.01087 and an RMSE of 0.01441, suggesting that the model's predictive accuracy drops significantly when applied to unseen finance-sector data. This 4 to 6x increase in error from train to test points to overfitting as the model has likely learned patterns specific to the broader market sentiment data but fails to generalize effectively to the unique characteristics of the financial sector. While a 1.4% average prediction error (test RMSE) is not necessarily poor, the noticeable discrepancy between training and test performance highlights the need to either regularize the model or retrain it using finance-specific sentiment and price data for improved domain-specific accuracy.

Metrics	Finance Sector Train	Finance Sector Test	All Industries Train	All Industries Test
MAE	0.002339	0.010866	0.0075	0.007887
MSE	0.000012	0.000208	0.000118	0.000125
RMSE	0.003432	0.014407	0.0109	0.011176

Table 14: FinBERT-LSTM Model Performance – Finance Sector vs. All Industries

5. LSTM Forecasting for Healthcare Sector Price Movement (Cece)

This study applies a Long Short-Term Memory (LSTM) neural network to forecast percentage changes in stock prices within the healthcare sector using sentiment analysis derived from financial news headlines. The model leverages FinBERT to classify each headline as Positive, Neutral, or Negative, which are then numerically encoded as 2, 1, or 0, respectively. These sentiment scores, along with FinBERT's confidence values and tokenized article titles, were combined as input features to train a time series model capable of learning temporal patterns between news content and short-term stock price movements.

The model was trained on 70% of the dataset, with the remaining 30% used for testing. Table 15 below presents the evaluation metrics for both sets. The enhanced LSTM achieved a Mean Absolute Error (MAE) of 0.0722 on the training set and 0.0917 on the test set. Mean Squared Error (MSE) was 0.0113 for training and 0.0178 for testing, while Root Mean Squared Error (RMSE) was 0.1064 and 0.1333, respectively.

Metric	Train Set	Test Set
MAE (Mean Absolute Error)	0.0722	0.0917
MSE (Mean Squared Error)	0.0113	0.0178
RMSE (Root Mean Squared Error)	0.1064	0.1333

Table 15: Evaluation Metrics for LSTM Regression Model (Healthcare Sector)

These results show that the model generalized well without overfitting, maintaining reasonably low error rates across both phases. A time series plot (Figure 18) visualizing the last 50 values from the training set and the first 50 from the testing set illustrates the model's tracking of actual price movements. While sharp market fluctuations were occasionally underestimated, the LSTM was able to capture overall direction and trend with strong consistency—especially valuable in high-volatility sectors like healthcare.

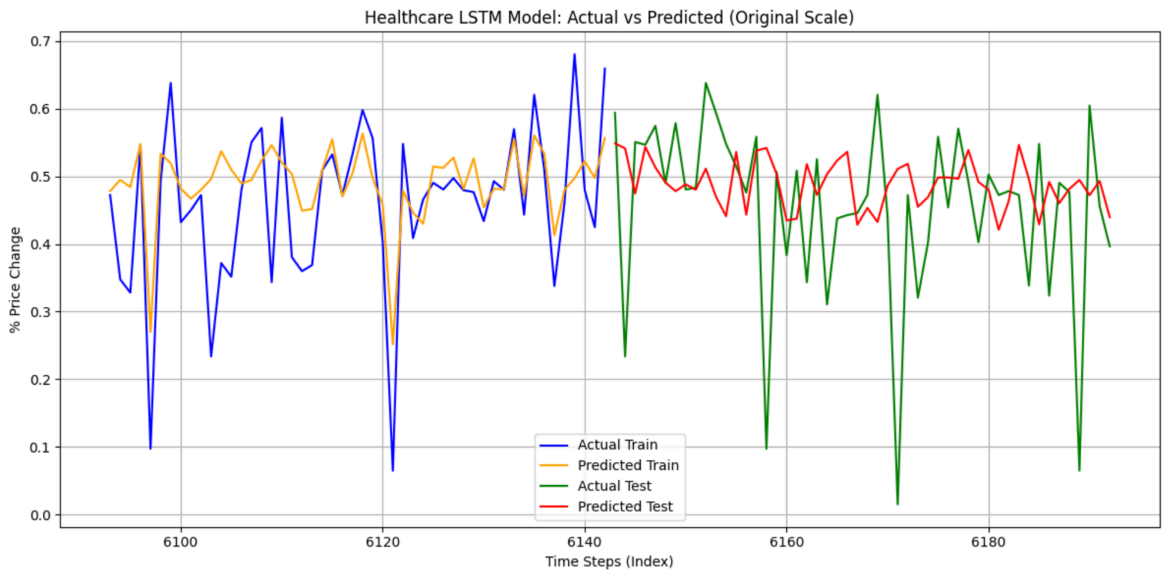

Figure 18: Actual vs. Predicted Percentage Price Change for Healthcare Sector

The findings underscore the value of incorporating sentiment-enhanced features into time series forecasting. By fusing FinBERT-generated sentiment and confidence levels with headline-level information, the model uncovered patterns aligned with investor behavior and price reaction. This validates that financial news sentiment significantly contributes to price changes in the healthcare sector. Potential improvements to the model include the use of lagged sentiment features to capture delayed responses, as well as engineering technical indicators (e.g., moving averages, volatility bands) or adding event-driven variables like FDA announcements or earnings reports.

Comparative Analysis: FinBERT Model on All Industries vs Healthcare Sector

To assess performance improvements, we compared this enhanced model to the original baseline version—a simpler LSTM trained for fewer epochs using normalized target values. While the original model showed smaller error values numerically, this was due to its use of scaled price changes, which limits real-world interpretability. In contrast, the enhanced version makes predictions on the actual scale of price fluctuations, offering actionable insights for analysts and investors. The comparison is summarized below.

Metric	**Original Model (Normalized Scale)**	**Enhanced LSTM Model (Original Scale)**
MAE – Train	0.0074	0.0722
MAE – Test	0.0083	0.0917
MSE – Train	0.000117	0.0113
MSE – Test	0.000149	0.0178
RMSE – Train	0.0108	0.1064
RMSE – Test	0.0122	0.1333

Table 16: Model Performance Comparison – Original vs. Enhanced LSTM (Healthcare)

Although the original model outperforms the enhanced version numerically, this is primarily an artifact of normalization. The improved LSTM architecture—featuring stacked layers, dropout regularization, and unscaled outputs—better mirrors realistic trading environments and risk conditions. Therefore, the enhanced model is more applicable to real-world healthcare sector forecasting and investment decision-making.

6. Predicting Entertainment Industry Stock Movements Using LSTM Models (Kiet Nguyen)

The FinBERT-LSTM modeling framework was subsequently deployed to analyze the entertainment sector index, comprising 21 major energy corporations including Walt Disney, Meta Platform, Warner Bros. Discovery, and Netflix Inc. The objective of this implementation was to forecast same-day percentage price changes utilizing sentiment-derived input features.

Model Performance Evaluation

The evaluation metrics presented in *Table 17* demonstrate negligible variance between training and test set performance indicators. The Mean Absolute Error (MAE), Mean Squared Error (MSE), and Root Mean Squared Error (RMSE) values exhibit remarkable consistency across both datasets, indicating robust model generalization without evidence of overfitting or underfitting phenomena.

Metrics	Entertainment Sector Train	Entertainment Sector Test
MAE	0.007676	0.007965
RMSE	0.011069	0.011529
MSE	0.000123	0.000133

Table 17: Evaluation Metrics for Regression Model (Entertainment Sector)

The consistently low error magnitudes across all metrics provide empirical evidence of the model's reliable predictive capability within the entertainment sector domain. These quantitative findings are corroborated by the visual analysis presented in ***Figure 19A***, which illustrates strong alignment between predicted and actual market movements, thereby validating both the forecasting precision and practical utility of the model for sector-specific applications.

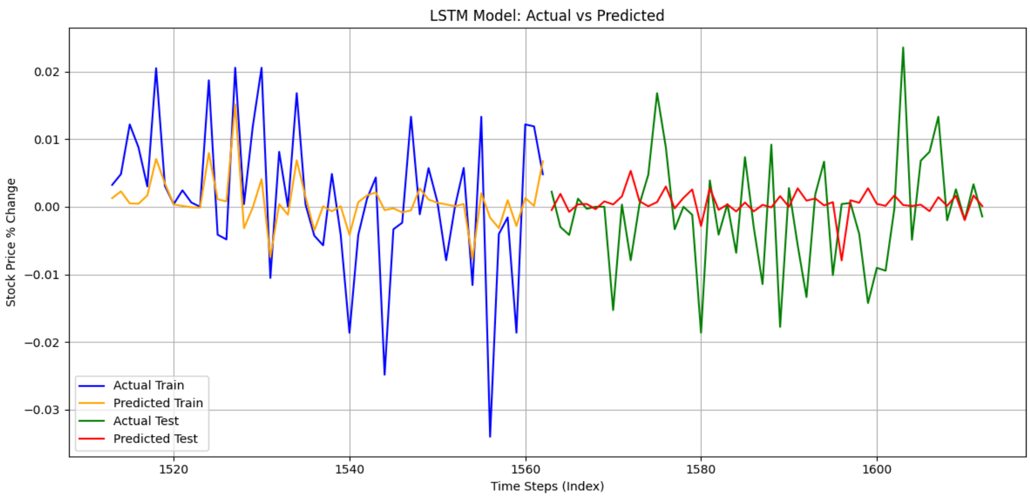

Figure 19A. Actual vs Predicted Percentage Price Change for Entertainment Sector

To capture both spatial and temporal patterns in financial text data, we implemented a hybrid model combining a convolutional neural network (CNN) with an LSTM layer. The architecture begins with an embedding layer (10,000-token vocabulary, 64-dimensional vectors), followed by a Conv1D layer with 64 filters to extract local semantic features. A max-pooling layer reduces dimensionality, feeding into a 64-unit LSTM layer for sequential learning. Dropout regularization (rate = 0.4) helps mitigate overfitting, and a linear output layer generates the final prediction. This hybrid architecture enhances predictive accuracy by modeling both immediate textual sentiment signals and longer-term market dependencies. Its performance is illustrated in ***Figure 19B***, highlighting the model's capability to closely track market movement trends.

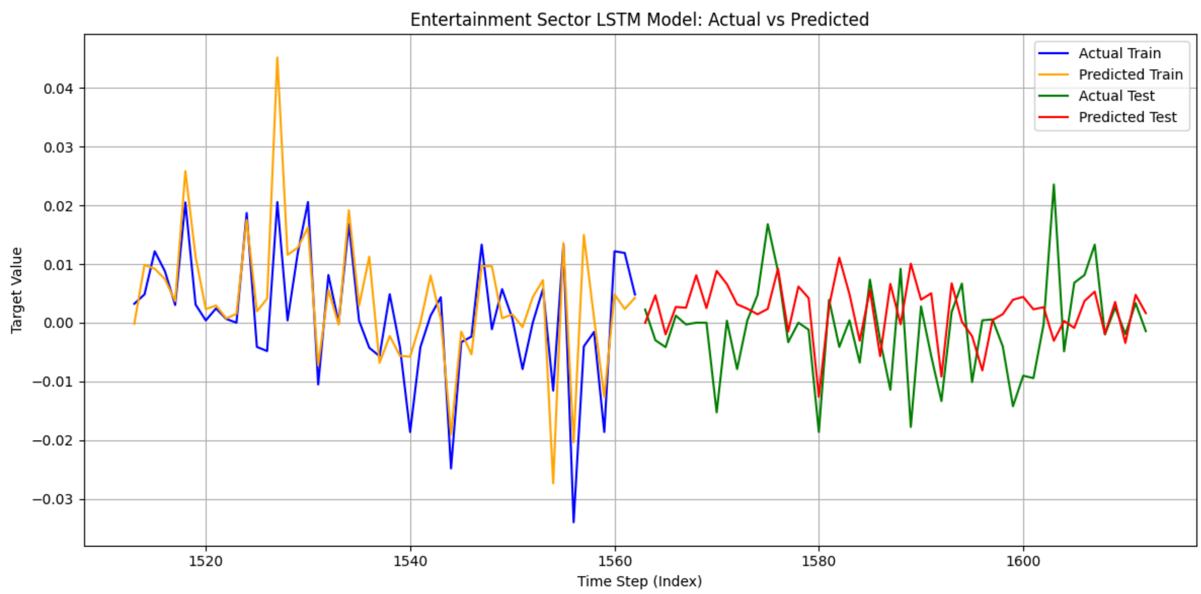

Figure 19B. Actual vs Predicted Percentage Price Change for Entertainment Sector Using LSTM hybrid model

Comparative Analysis: FinBERT Model on All Industries vs Entertainment Sector

Table 18 summarizes the post-tuning performance of the regression model across the Entertainment sector and the broader industry dataset. For the entertainment sector, the model achieved strong predictive accuracy on the training set, with a Mean Absolute Error (MAE) of 0.004840, Root Mean Squared Error (RMSE) of 0.006958, and Mean Squared Error (MSE) of 0.000048. On the test set, the error slightly increased to a MAE of 0.007565, RMSE of 0.010811, and MSE of 0.000117, which is expected due to generalization variance but still indicates solid performance.

Metrics	Entertainment Sector Train	Entertainment Sector Test	All Industries Train	All Industries Test
MAE	0.004840	0.007565	0.0075	0.007887
RMSE	0.006958	0.010811	0.0109	0.011176
MSE	0.000048	0.000117	0.000118	0.000125

Table 18: Post-tuning Evaluation Metrics for Regression Model (Entertainment Sector)

Comparatively, when applied across all industries, the model's performance was slightly less precise. The train MAE and RMSE rose to 0.0075 and 0.0109 respectively, with an MSE of 0.000118. The test set showed further increase, with a MAE of 0.007887, RMSE of 0.011176, and MSE of 0.000125. This trend suggests that the model generalizes better in a focused domain (i.e., Entertainment sector) and may benefit from sector-specific tuning when applied to broader datasets. These results highlight the benefit of domain-specific modeling in financial forecasting and underscore the value of hybrid deep learning architectures like LSTM when fine-tuned on targeted industry data.

CHAPTER 5
CONCLUSION, LIMITATIONS AND RECOMMENDATIONS

Conclusion

This project was designed to improve how we forecast stock price movements by analyzing financial news headlines. In today's market environment, where news travels fast and investor reactions can be immediate, being able to turn headlines into insights is a major advantage. Our team used advanced text analysis (known as Natural Language Processing, or NLP) and machine learning to explore how the tone or sentiment of financial news might predict what happens next in the stock market.

Among the three tested models TextBlob, BERT, and FinBERT; FinBERT demonstrated the highest overall predictive performance due to its pre-training on large-scale financial text, making it better suited to interpret the nuances of finance-related sentiment.

First, we tested three popular tools for detecting sentiment in financial news: TextBlob, BERT-base, and FinBERT. Each of these models reads text and tries to decide whether the tone is positive, negative, or neutral. We then used the sentiment scores from each tool to train a model that forecasts stock price movements. After testing all three, FinBERT stood out as the most accurate and consistent model across the board. This model is specially designed for financial language, which helped it understand the tone of headlines more precisely than general-purpose models.The results revealed that the general FinBERT model outperformed its industry-specific versions in five out of six sectors: Technology, Real Estate, Healthcare, Energy, and Finance. This pattern can be attributed to several important factors:

1. **Strong Influence of Macro Trends:** These sectors are often influenced by broad economic indicators such as interest rates, inflation, global policy shifts, or commodity prices. Because FinBERT is trained on general financial news, it has already encoded common sentiment signals and market reactions associated with these macroeconomic drivers — making fine-tuning less beneficial.
2. **Cross-Sector News Relevance:** Financial news items such as central bank decisions or geopolitical developments often affect multiple sectors simultaneously. A general model like FinBERT can capture these cross-sectoral patterns more effectively than a narrowly fine-tuned model that only sees single-sector data.
3. **Overfitting Risk in Fine-Tuned Models:** Industry-specific datasets tend to be smaller and narrower in scope. Fine-tuning on limited, specialized data can lead to overfitting — where the model memorizes short-term or sector-specific noise rather than learning generalizable sentiment patterns. This reduces its ability to generalize to future or broader market data.

In contrast, the Entertainment sector was the only one where the fine-tuned FinBERT model outperformed the general version. This outcome is likely due to the unique nature of media and pop culture sentiment, which differs significantly from traditional financial language. Entertainment news often includes subtle sentiment cues, slang, cultural references, or subjective tone that a general financial model may not interpret accurately. Fine-tuning FinBERT on entertainment-specific data helps the model adapt to this distinctive language style, leading to better performance in identifying sentiment shifts that influence stock price in this domain.

Overall, this study concludes that general financial sentiment models like FinBERT are sufficient, and in many cases superior, for most broad-market sectors, especially when they are shaped by macroeconomic factors. However, in industries with highly specialized or culturally distinct language, such as media and entertainment, fine-tuning remains a valuable strategy for improving model accuracy. These insights guide practitioners on when to leverage general models and when to invest in customization based on the characteristics of their target sector.

Limitations

Despite the promising findings of our study on financial news sentiment analysis using AI, there are several limitations that must be acknowledged:

1. Limited Data Scope

- Our dataset was restricted to a single year (2018), primarily due to computational resource limitations and to avoid volatility caused by major global events such as the COVID-19 pandemic. This limited timeframe may not fully capture long-term market trends or rare events that can significantly affect model performance.

2. Data Source Bias

- Financial news articles were collected from a limited number of sources, primarily Bloomberg Terminal. This may introduce source bias, as sentiment from alternative media outlets, social media (e.g., Twitter, Reddit), or regional financial news was not incorporated, potentially narrowing the sentiment diversity reflected in the models.

3. Overfitting in Industry-Specific Models

- While the general FinBERT model performed consistently across industries, several of the industry-specific models displayed signs of overfitting, particularly in sectors like Real Estate and Finance. This suggests that smaller, sector-specific datasets may not provide sufficient variety for deep learning models to generalize effectively.

4. Same-Day Prediction Assumption

- The study primarily focused on predicting same-day stock price movements based on sentiment scores. However, market reactions to news can sometimes be delayed or develop over several days. By limiting the prediction window to same-day forecasts, we may have overlooked lagged sentiment effects.

5. Simplified Sentiment Aggregation

- In Phase 1, we aggregated sentiment scores without considering company market capitalization or stock index weighting, potentially distorting the true impact of large-cap companies on index movements. Although we discussed this issue in our recommendations, it was not fully addressed in our modeling.

6. Computational Constraints

- Due to limited computational resources, we applied base versions of LSTM and FinBERT models without exploring more computationally intensive architectures (e.g., transformer-based time series models or reinforcement learning), which could potentially improve accuracy.

7. Generalization to Other Markets

- Our study focused on U.S.-based indices and sectors. The results may not generalize to emerging markets, other stock exchanges, or different asset classes such as commodities, bonds, or cryptocurrencies, which may have distinct market dynamics and sentiment drivers.

8. Index Weighting Not Implemented

- Due to time constraints, we were unable to complete Objective 3: Customized Index Membership, which aimed to build sector-level sentiment indices weighted by individual company influence. As a result, all companies were treated equally in our model, which may reduce accuracy—especially in sectors dominated by large-cap firms. This is a key area for future improvement.

These limitations suggest that while AI-powered sentiment analysis shows significant potential in financial forecasting, further studies with more extensive datasets, diversified data sources, advanced modeling techniques, and multi-modal inputs are necessary to enhance model robustness and generalizability across different market conditions.

Recommendations for the Financial Analytics Field

1. **Use FinBERT as the default sentiment analysis model for financial news**

Since FinBERT outperformed other models and works well across multiple sectors, companies can rely on it as a strong, general-purpose solution. This avoids the need to test or fine-tune other models unless working in highly specialized areas like entertainment or media.

2. **Skip unnecessary fine-tuning by using the general FinBERT model**

For most industries, there is no need to spend extra time or resources customizing sentiment models. Our results show that a strong general model is often just as good, which simplifies development and speeds up deployment.

3. **Apply sentiment forecasting to sectors where news tone closely impacts stock price**

Industries like Finance, Energy, and Real Estate responded strongly to sentiment-driven forecasting. We recommend prioritizing these sectors for early implementation, as they are more likely to benefit from real-time sentiment analysis.

4. **Consider fine-tuning only for sectors with emotionally driven narratives**

If your business works in industries like Entertainment or Media, where news tone can be more emotional and unpredictable, it may be worth investing in a fine-tuned sentiment model for better accuracy.

5. **Incorporate company weighting to improve the general model's accuracy**

One important consideration when applying sentiment analysis across a broad range of financial news is that not all companies have equal influence on the market. In reality, large-cap companies—such as Apple, Microsoft, ExxonMobil, or JPMorgan Chase—hold significantly more weight in major stock indices and investor decision-making than smaller or less publicly visible firms. This means that a single headline about one of these large corporations can have a ripple effect, influencing market sentiment more strongly than a dozen articles about lesser-known companies.

Treating every company equally in a sentiment model may lead to skewed or diluted results, especially when trying to forecast broader market movements or industry trends. Therefore, we recommend incorporating a weighted sentiment scoring system into the general model. This system would assign greater importance to sentiment signals from companies with a larger market capitalization, higher trading volume, or greater influence within their respective industries.

For example, if both a Fortune 100 company and a small-cap company receive negative news on the same day, the model should give more predictive weight to the sentiment related to the Fortune 100 firm. This reflects how investors and the market actually respond in practice—where large companies often set the tone for the sector or even the entire index.

Incorporating these company-specific weights will improve the realism and accuracy of the model's forecasts. It ensures that the sentiment analysis aligns more closely with real-world investor behavior and market dynamics, ultimately making the model a more effective tool for financial analysts, institutional investors, and automated trading systems alike.

REFERENCES

Gu, W. J., Zhong, Y. H., Li, S. Z., Wei, C. S., Dong, L. T., Wang, Z. Y., & Yan, C. (2024). Predicting stock prices with FinBERT-LSTM: Integrating news sentiment analysis. In *2024 8th International Conference on Cloud and Big Data Computing (ICCBDC 2024)* (pp. XX–XX). ACM. https://doi.org/10.1145/3694860.3694870

Halder, S. (2022). A deep learning approach to predicting stock prices using sentiment analysis from financial news. *arXiv*. https://arxiv.org/pdf/2211.07392

Konstantinidis, G., Papadopoulos, A., & Stylianou, T. (2024). LLMs and NLP models in cryptocurrency sentiment analysis: A comparative classification study. *Journal of Financial Data Science*.

Maqbool, M., Khan, M., & Ahmed, R. (2023). Stock prediction by integrating sentiment scores of financial news and MLP-Regressor: A machine learning approach. *Journal of Finance and Technology*.

Konstantinidis, T., Iacovides, G., Xu, M., Constantinides, T. G., & Mandic, D. (2024). FinLlama: Financial sentiment classification for algorithmic trading applications. *International Conference on Machine Learning and Data Engineering*.

Maqbool, J., Aggarwal, P., Kaur, R., Mittal, A., & Ganaie, I. A. (2023). Stock prediction by integrating sentiment scores of financial news and MLP-Regressor: A machine learning approach. *International Conference on Machine Learning and Data Engineering*.

Research and Markets. (2022). AI Market in Financial Services - Market Size, Share, Trends, & Growth Analysis. Available at: https://www.researchandmarkets.com

PwC Global Investor Survey. (2021). Investors' Use of Sentiment Analysis in Financial Decision-Making. Available at: https://www.pwc.com

Fortune Business Insights. (2022). Artificial Intelligence in Finance Market Growth and Forecasts. Available at: https://www.fortunebusinessinsights.