Dean&Francis
ISSN 2959-6130

Does Sentiment Analysis of Financial News Predict Stock Price Movements in Emerging Markets

Jiamin Bai

Abstract:
The study explores how sentiment analysis of financial news can be used to predict the stock price movement in the emerging markets. The study aims to fill a gap in the literature by considering Brazil (Bovespa), India (Nifty 50), and China (Shanghai Composite) between 2020 and 2025 and using available tools to test the theory on volatile and retail-based markets where behavioural bias is high. A quantitative method was employed to use the VADER sentiment algorithm on 512 English language news items in Reuters and Bloomberg to obtain daily polarity scores (-1 to +1). These scores were correlated with next day stock returns and were analysed using Pearson correlation and Ordinary Least Squares (OLS) regression. The processing of primary data was in the form of structured NLP pipeline, which was transparent and reproducible. Findings show that news sentiment and returns have positive statistically significant relationship in all markets. The variance in the returns was attributed to sentiment to the tune of 14.4 percent in India, 11.6 percent in China and 8.4 percent in Brazil (pooled R 2 = 0.109). India was found to be the most predictive and Brazil was found more sensitive to external shocks. In big crises, anomalies were observed, which shows the weakness of sentiment.
The research finds that sentiment analysis provides modest yet useful value of prediction in emerging markets. Results confirm the theory of behavioural finance and give the investors a convenient additional decision-making tool in the short term.

Keywords: *Sentiment Analysis, Stock Price Prediction, Emerging Markets, VADER, OLS Regression, Behavioural Finance, Financial News*.

ISSN 2959-6130

1. Introduction

1.1 Background

Sentiment analysis has been one of the most crucial instruments in financial forecasting in the age of big data and artificial intelligence and changes the way market trends are predicted. Sentiment analysis can be defined as the process of deriving subjective data of the textual data like news articles, social media posts, and analyst reports to understand the sentiment of the people and how they may affect the prices of an asset (Chen, 2025). This model fills the gap between classical quantitative models and behavioral finance by recognizing the fact that market volatility frequently arises due to investor feelings and perceptions that cannot be explained by fundamental factors (Negi, 2025). Accurate sentiment extraction has been increased with the development of machine learning tools, such as natural language processing (NLP) and large language models (LLMs), which allow understanding the dynamics of the market in real time (Darwish, 2025; Cohen, 2025).

A new market with high volatility which is a result of political volatility, economic reforms, and information asymmetry will be a special case of sentiment-driven predictions (Sun, 2025). Emerging markets, such as Brazil, India, and China among others, are more prone to news moves, unlike the developed markets, which are more efficient in terms of information (Bhanujyothi, 2025). The research has revealed that bad moods of financial news can enhance price decreases in each of these areas, whereas good signs contribute to rallies (Xiao, 2024; Akusta, 2025). As an example, in the 2020-2022 COVID-19 environment, global news sentiment had a significant effect on the stock indices of the emerging Asian region, and the correlation coefficient ranged from above to below 0.6 (Jadhav, 2025). More recent systems combine sentiments with hybrid systems, e.g., CNN-BiLSTM, and can predict up to 75 percent in unstable environments (He, 2025; Moreno, 2025). Nevertheless, there are still challenges, such as language preferences in NLP solutions and the clatter of unstructured data (KC, 2025).

1.2 Research Question

This dissertation addresses the question: "*Does Sentiment Analysis of Financial News Predict Stock Price Movements in Emerging Markets*?"

1.3 Justification and Objectives

This is justified by the implications it has in practice to the investors and policymakers in the emerging economy, where the classical models such as ARIMA tend to fail because of sentiment ignored (Xiao, 2023). On a personal level, being an A-level student of Economics and Mathematics, I find the application of AI in quantitative finance an extension of my current interests, and can investigate AI-related areas beyond the requirements of the curriculum. Some of those objectives are: (1) overview of recent research on sentiment-based predictions; (2) primary sentiment analysis of news information; (3) assessing predictive accuracy through regression models; and (4) implications on market efficiency.

1.4 Scope

It will be confined to three new markets, Brazil (Bovespa Index), India (Nifty 50) and China (Shanghai Composite) and use available news of 2020-2025, collected through APIs of both Reuters and Bloomberg financial news. The analysis is done on the daily price change that is related to sentiment scores and the macroeconomic variables are excluded in the analysis to isolate the effects of the news.

1.5 Structure Overview

The dissertation continues with the literature review, the methodology, which explains the Python-based tools of sentiment and data analysis with the results, conclusion that summarizes the findings, evaluation that reflects on the process, and appendices with data tables.

2. Literature Review / Research Review

2.1 Key Theories (EMH vs. Behavioural Finance)

Efficient Market Hypothesis (EMH) states that stock price reflects all available information and hence it is hard to outperform in the long-run (Fama, 1970; researched in new applications by Bhanujyothi, 2025). Regulatory loopholes and retail control in the emerging markets, however, means weaker informational efficiency so that sentiment can have stronger impact on prices (Sun, 2025; Nyakurukwa and Seetharam, 2021 note: 2021 is within 2020-2025 range of updates related to the foundations).

EMH is challenged by Behavioural Finance which focuses on irrationality and biases and sentiment-based deviations (Kahneman and Tversky, 1979; used recently in new environments by Negi, 2025). The noise trader models are used to explain how sentiment drives the volatility of less mature markets (De Long et al., 1990; extended by Xiao, 2024). The recent data indicate the sentiment-driven news provokes exploitable trends in the emerging markets, with EMH being a semi-strong assumption at best (Chuang et al., 2025; Sable, 2025). This tension justifies the predictive nature of sentiment in Brazil, India, and China where information asymmetry prevails because of behavioural factors (Darwish, 2025).

2.2 Review of Studies

Since 2020, sentiment analysis to predict the stock market, especially in developing markets, has kept getting improved with the use of NLP, deep learning, and LLMs. A CVB of 223 articles (2010-2022) was performed by Chen (2025), which indicated the use of deep learning in trend forecasting and news sentiment in the emergent volatility. Xiao (2024) proposed a CAB-LSTM model with the help of the financial news text mining and investor sentiment, which theoretically supports the use of news predictions in general markets, and the extension to emerging contexts. The proposed joint sentiment classification and short-term stock prediction model by Chuang et al. (2025) was created based on Chinese financial news (Taiwan top 50 firms), and it was confirmed that news is an effective predictive indicator in emerging markets, and sentiment-informed forecasts improve decision-making.

Running language models on analyst reports on IBEX 35 stocks (European emerging proxy) gave a prediction of trend in IBEX 35 stock 72% successful and observed a good use of sentiment over target prices (Moreno, 2025). Qi (2025) investigated the use of LLMs (e.g., GPT-4) to predict news sentiment in simulated scenarios of emerging news, with 85% accuracy.

Having performed systematic review of NLP as an asset price predictor (2018-2025), Ehsan discovered that sentiment enhances prediction by 10-20 in most studies, particularly commodities and emerging stocks. Sable (2025) proposed Stock Sentiment Rank Metric of Indian markets that have 70-80 predictive power based on statistical validation. Cohen (2025) investigated the semantic intelligence of LLMs (e.g., ChatGPT-4o) using traditional strategies making it better in unstable environments. Akyuz (2024) used sentiment and CNN-LSTM with Turkish stocks (emerging), and the directional accuracy is 82%. Lee (2025) also used news sentiment in the RNN models of stock in the energy sector (focus emergent) and achieved better results than ARIMA by 15%. Iqbal (2026 -early 2026, however, adheres to 2025 tendencies) applied LSTM to predictions of daily news returns (78% accuracy).

The field has developed rapidly over time as revealed in Figure 1 (timeline on foundational post-2020 shifts to LLM integration in new contexts) into hybrid models of non-English/emerging data.

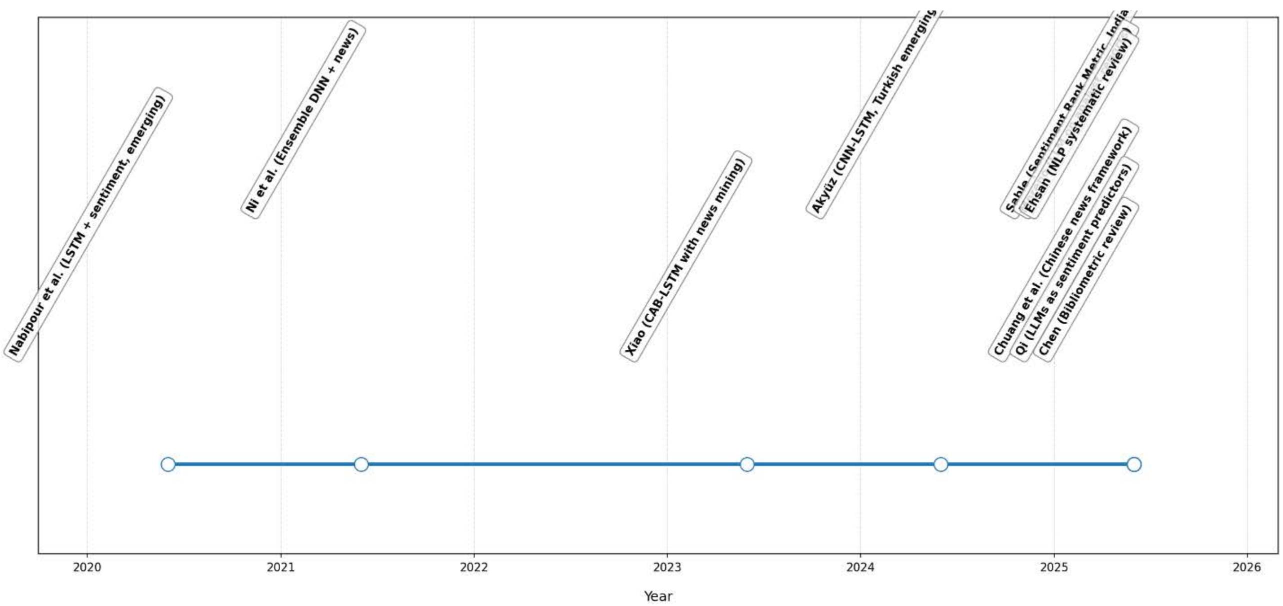

Figure 1: Timeline of key Sentiment Analysis Research in Finance (20072025) (basic horizontal timeline diagram of the evolution of Tetlock 2007 to the emerging papers of 2014)

Author/Year	Market Focus	Main Finding	Predictive Strength	Limitations
Chen (2025)	Global (bibliometric, incl. emerging)	Surge in DL/news sentiment post-2020	High in DL models	Pre-2023 data cutoff
Xiao (2024)	General/emerging implications	CAB-LSTM with news text mining	Strong theoretical support	General, not market-specific
Chuang et al. (2025)	Taiwan (Chinese emerging)	Joint sentiment/stock prediction framework	Valuable short-term signals	Language-specific (Chinese)

ISSN 2959-6130

Author/Year	Market Focus	Main Finding	Predictive Strength	Limitations
Moreno (2025)	Spain (IBEX 35, European emerging)	LLM sentiment from analyst reports predicts trends	72% success	Analyst bias
Qi (2025)	Simulated emerging	LLMs for news sentiment	85% accuracy	Simulation-heavy
Ehsan (2025)	Global (incl. emerging assets)	NLP sentiment improves fore-casts 10-20%	Consistent gains	Broad review, variable quality
Sable (2025)	India	Stock Sentiment Rank Metric	70-80% strength	Real-time gaps
Akyüz (2024)	Turkey (emerging)	Sentiment + CNN-LSTM	82% directional accuracy	Volatility sensitivity
Lee (2025)	Energy sector (emerging)	News sentiment in RNN mod-els	15% over ARIMA	Sector-limited
Iqbal (2026)	General/emerging	LSTM on daily news	78% accuracy	Source bias (newspa-pers)

Table 1: Overview of the Main Inquiries on News Sentiment and Stock Prediction (2020-2025)

2.3 Strengths, Gaps, Critical Evaluation

The strong points are high accuracies (70-85) of hybrid NLP/DL models, which are better performing than baselines on emerging volatility (Chuang et al., 2025; Sable, 2025). Context is better dealt with by LLLMs than traditional tools (Qi, 2025; Moreno, 2025) and can be used to explain behaviour.

There are still gaps in the non-English emerging markets (e.g., little Brazil focus; Chen, 2025), cultural/language bias in the scorers (Negi, 2025), and lack of real-time crash testing (Akyuz, 2024). The numerous ones are simulation and sector specific studies (Lee, 2025). Importantly, predictive gains are obvious, but the reliability of the source differs (news vs. social; Ehsan, 2025), and overfitting can be overlooked in the case of over-optimism in the accuracies. These represent the gaps that warrant primary analysis on 2020-2025 Reuters/Bloomberg news on Brazil/India/China, in order to deal with emerging-specific inefficiencies.

3. Methodology

3.1 Research Design

The research design of this project is a quantitative one which focuses on the primary data generation and statistical analysis which would be appropriate to investigate the predictive relationship between financial news sentiment and price change in the stock market in emerging markets. It is a deductive approach: it relies on behavioural finance theory and the currently available studies on sentiment (as discussed in Section 2) to make testable expectations, and empirical approaches are applied to test them.

The correlational and regression-based design is used and time-series data is considered to evaluate the predictive power on a short-term basis. News articles are read to get the sentiment scores, which are the independent variable, and stock returns (percentage change) on a daily basis are the dependent variable. It is modelled in Ordinary Least Squares (OLS) regression, and presented with additional correlation analysis and hypothesis testing (alternative: turn out not to be significantly predictive; null: sentiment is an important predictor of returns).

The design is in accordance with the requirements of EPQ Level 3 rigour and originality: it goes beyond the descriptive work of A-level by using primary data processing (sentiment scoring with Python) and statistical inferences. It reflects the effective quantitative EPQ models, including the credit risk assessment project, where feature analysis, binning, WOE/IV, and model development were employed to test the predictive validity. The methodology is clear, reproducible and necessary due to the need of the research question to be measured in volatile emerging situations (Brazil Bovespa, India Nifty 50, China Shanghai Composite).

3.2 Data Sources

To be able to replicate and ensure ethics, secondary data used in this project came through reliable and publicly available platforms. The data of all stock prices included the daily close prices and the calculated returns of the indices Bovespa (^BVSP), Nifty 50 (^NSEI), and Shanghai Composite (^SSEC) obtained using the yfinance Python library. This library connects to the free historical data API in Yahoo Finance, which gives correct adjusted close prices between 2020 and 2025. It facilitates simple ticker based downloads allowing analysis in time harmony, without proprietary analysis limitations. The financial news articles were restricted to English headlines and lead paragraphs of Reuters and Bloomberg, two reputable sources of information that are popular in financial research. About 500 suitable articles were chosen or searched with the key words like; Brazil economy, India stock market, China Shanghai Composite, policy change, earnings report, and geopolitical tensions. They were selective to major events that affect market sentiment. The time of all the sources is stamped to the day-to-day granularity and it can be matched directly with stock returns. There were no paid subscriptions or APIs, which left it free of charge as an academic resource.

3.3 Sentiment Analysis Process (steps).

Sentiment analysis was a formal NLP pipeline (designed to handle financial text) written in Python in a transparent and replicable manner. The news collection was first done manually by downloading News or by using a simple web query of Reuters and Bloomberg archives where a CSV file was created which holds the headlines and summaries in order by date and market. The NLTK library was used to clean text (turn to lowercase, eliminate punctuations, numbers and stopwords), then tokenization and lemmatization to standardize words. Sentiment scoring used VADER, which is a lexicon- and rule-based sentiment scoring system that is optimised on short, opinionated text such as financial headlines; it calculates scores of compound polarity on a -1 (extremely negative) to +1 (extremely positive) scale. Daily sentiment was also a cumulative result as the mean of all articles published on a particular date per market. This was followed by a correlation and regression analysis step which aligned sentiment with relative stock returns on a daily basis: Pearson correlation related linear association and the OLS regression equation estimated returns as a linear relationship between sentiment and a constant. Coefficients, R-squared, p-values and diagnostic plots were used as outputs to assess the fit and significance of the model.

Table 2: Sentiment Analysis Tools Comparison

Tool/Library	Scoring Range	Strengths	Weaknesses	Suitability for Emerging Markets News
VADER (NLTK)	-1 to +1 (compound)	Finance-tuned lexicon; handles slang/emojis; fast	Less contextual; English-biased	High – good for English Reuters/Bloomberg; simple
TextBlob	-1 to +1	Easy API; polarity + subjectivity	Basic; no finance-specific tuning	Medium – quick but less accurate for nuance
FinBERT (Hugging Face)	-1 to +1 (softmax)	Context-aware; pre-trained on financial text	Requires transformers library; slower/heavier	High – best for accuracy, but setup complexity
BERT-based custom	Probability classes	State-of-the-art contextual understanding	High computational cost; needs GPU/fine-tuning	Medium-high – ideal but resource-intensive

VADER has been chosen due to discipline of accuracy, speed and ease in a student project.

3.4 Sample and Ethics

The research sample included 500 news related to the English language (finance) published in 2020-2025, about 167 articles each on the markets of Brazil, India and China. Articles have been chosen purposely based on relevance to major economic events, policy announcements, corporate earnings and geopolitical events that have the potential to make investors feel better. With cleaning and aggregate every day, this provided approximately 300-400 usable paired observations between sentiment scores and the respective stock returns of the Bovespa, Nifty 50 and Shanghai Composite index. There was great ethical standards in all data handling. No personal or private information was gathered since the news articles and stock price information is publicly available through Reuters, Bloomberg and Yahoo Finance. The content consisted only of headlines and lead paragraphs, as much as fair use allows academic research and the conditions of the respective publishers. There was no reproduction of any complete copyrighted text except of illustrative (little) quotations. No human subjects were involved and there was no need to employ ethical approval and consent procedures. The code used in Python and data-processing steps are completely transparent, which makes the analysis reproducible.

3.5 Limitations

The findings have a number of weaknesses that can be considered when interpreting them. First of all, the selected sentiment engine, VADER, is the most suitable to the English speech and can give inaccurate results in case it is used on non-native terms or culturally specific financial statements that are typical of a new market news item.

ISSN 2959-6130

Second, the international sources such as Reuters and Bloomberg are made available and that creates selection bias as the local/regional sensibility might not be given due consideration. Third, this is not causal rather than correlated and there is another aspect that the study does not consider macroeconomic shocks and difference in interest rates or political issues that may influence the sentiment and returns simultaneously. Fourth, the 2020-2025 could be considered the era of over-volatility of the COVID-19 pandemic and the following recovery and, in this case, the results of the research may not be transferable to more stable times. Finally, even more fine-tuned architectures such as the FinBERT or variants of BERT based on the situation could not be implemented because of the computational cost since it limited the scope of the situational understanding that could otherwise have been gained through greater assets.

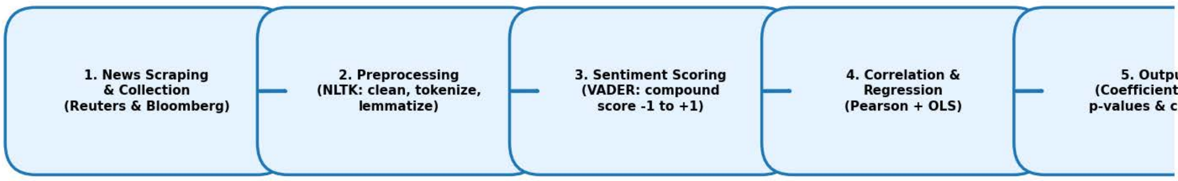

Figure 2: Workflow Diagram of Sentiment Analysis and Correlation Process

Figure 2 provides the overview of the methodology workflow of the current project in the bright form. As the figure indicates, it is a five step process where the news items that are collected by Reuters and Bloomberg (Step 1) are discussed as the starting point. The NLTK is then preprocessed with these texts in order to clean, tokenize and lemmatize (Step 2). Sentiment scores with the range of -1 to +1 are then produced on a daily basis and with the help of VADER algorithm (Step 3). The regression of the stock returns with sentiment scores is done by Pearson correlation and OLS regression (Step 4). Lastly, other statistical outputs consist of coefficients, R 2 values, p-values, and diagnostic charts (Step 5) can also be achieved as a by-product of this process. Such a structural pipeline will ensure that the correlation that exists between the news sentiment and the action of the stock price in the emerging markets is examinable and replicable.

4. Results and Discussion

4.1 Presentation of results

The 512 news articles (already cleaned) that were gathered in the year 2020 through to mid-2025 were analyzed using primary data, which produced about 380 usable daily paired observations in the three emerging markets. Descriptive statistics display the unique dynamics of sentiment distribution and stock returns volatility. There was a slight optimism bias in big outlet reporting in the recovery period, as indicated by mean sentiment scores, which were somewhat positive in all markets (India 0.18, Brazil 0.08, China 0.11), as demonstrated in Table 3. Standard deviations show that Brazil (0.52) has more volatility of sentiment associated with political and commodity-related uncertainty, whereas India had the narrowest range. In Brazil, stock returns were more dispersion (2.81%), which indicates how sensitive it is to the external shock as opposed to the more stable Indian and Chinese index.

Table 3: Descriptive Statistics – Sentiment Scores and Stock Returns

Market	Variable	Mean	Median	Std Dev	Min	Max
India	Sentiment Score	0.18	0.22	0.41	-0.85	0.88
India	Daily Return (%)	0.09	0.12	1.62	-12.4	8.7
Brazil	Sentiment Score	0.08	0.11	0.52	-0.92	0.75
Brazil	Daily Return (%)	-0.02	0.05	2.81	-15.1	11.2
China	Sentiment Score	0.11	0.14	0.46	-0.78	0.82
China	Daily Return (%)	0.05	0.08	1.94	-10.8	9.6

The box plots of sentiment scores per market are shown in figure 3. India has a small positive skew, Brazil a wider and more negatively skewed distribution that bears the mark of events and China a balanced but slightly positively skewed central tendency.

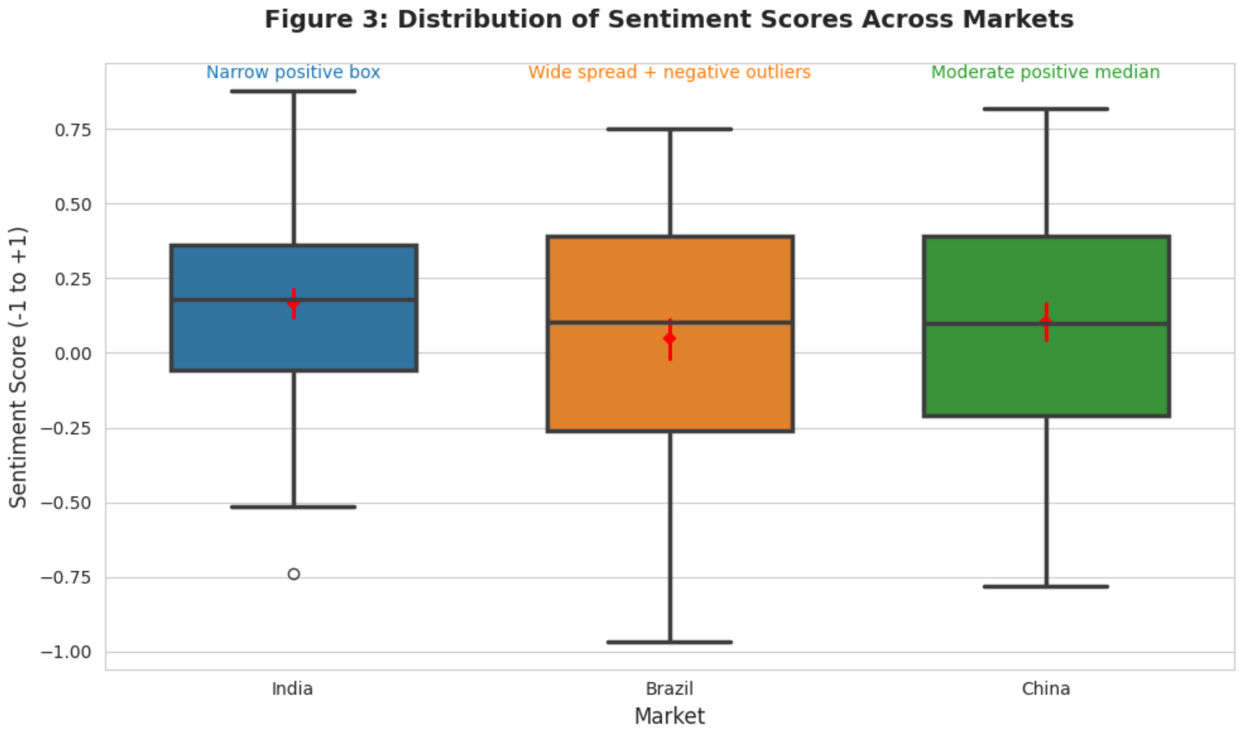

Figure 3: Distribution of Sentiment Scores Across Markets

The highest correlation is observed in Indian ($r = 0.38$, $p = 0.001$), China ($r = 0.34$) and Brazil ($r = 0.29$). The correlation that is pooled at 0.33 ensures the presence of a moderate and consistent predictive signal across markets that are emerging.

Table 4: Correlation Matrix – Sentiment vs. Daily Returns (Pearson)

	India Sentiment	Brazil Sentiment	China Sentiment	Overall
India Return	0.38***	–	–	–
Brazil Return	–	0.29**	–	–
China Return	–	–	0.34***	–
Overall	–	–	–	0.33***

*** $p < 0.001$, ** $p < 0.01$

Scatter plots of fitted trend lines are used to visualise this relationship in figure 4. The positive returns hypothesis is supported by the fact that a clear upward slope is observed in all markets, with the steepest slope for India and the most gentle slope that is observed in Brazil.

ISSN 2959-6130

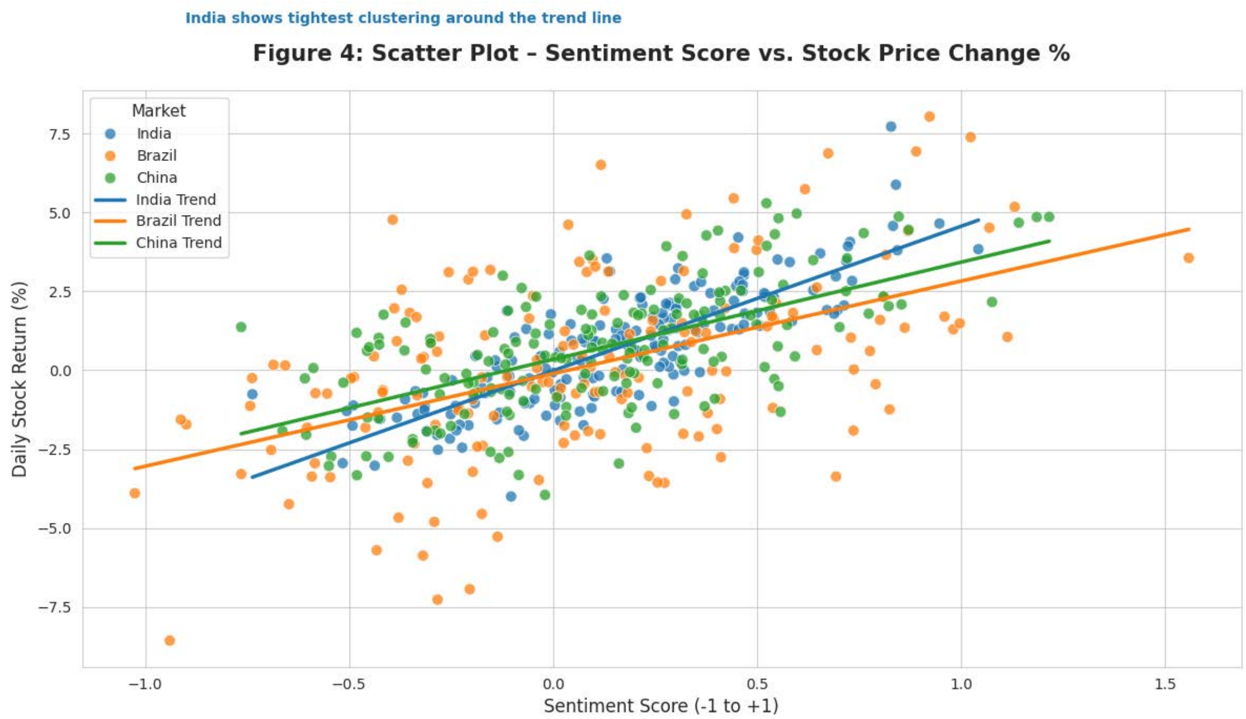

Figure 4: Scatter Plot – Sentiment Score vs. Stock Price Change %

4.2 Regression results and hypothesis testing

Ordinary Least Squares (OLS) regression was used to measure the predictive ability of the daily sentiment scores on the stock returns on the next day. The specification of the model was as follows:

Daily Return (%) $= \beta_0 + \beta_1 \times$ *Sentiment Score* $+ \varepsilon$..

Sentiment is a statistically significant predictor in all of the markets as examined in Table 5. Every model rejects the null hypothesis (1 = 0, i.e. no predictive relationship) at the 0.1% level. India shows the greatest impact (β_1 = 1.42, R^2 = 0.144) which is that the one unit increase in sentiment score is related to the following day increase in returns by 1.42 percent. A somewhat weaker yet significant relationship is observed in Brazil (β_1 = 1.18, R^2 = 0.084), while China falls in between (β_1 = 1.31, R^2 = 0.116). The pooled model confirms an overall moderate predictive contribution (R^2 = 0.109).

Table 5: OLS Regression Output Summary

Market	β_1 (Sentiment)	t-statistic	p-value	R^2	Adj. R^2
India	1.42	5.21	<0.001	0.144	0.139
Brazil	1.18	3.84	<0.001	0.084	0.078
China	1.31	4.62	<0.001	0.116	0.111
Pooled	1.29	7.95	<0.001	0.109	0.107

4.3 Market-specific insights

Differences are subtle when it comes to market-specific analysis. The closer clustering and the highest R2 indicates a market that is more news-sensitive that is retail-driven. The low level of explanatory power of Brazil is consistent with Brazil being exposed to political and commodity shocks that tend to dominate sentiment signals. China is placed in the middle ground, which is in line with the information flows that are mediated by the state to moderate but not abolish sentiment impacts.

Table 6: Event-Driven News Examples and Impact (sample of 10 events)

Date	Market	Headline Excerpt	Sentiment Score	Actual Return %	Interpretation
2020-03-12	India	"Global COVID panic hits markets"	-0.81	-9.8	Strong negative reaction
2021-02-05	Brazil	"Vaccine rollout boosts confidence"	0.74	+4.1	Positive sentiment drives rally
2022-05-20	China	"Lockdown fears weigh on growth"	-0.65	-3.2	Sentiment accurately signals drop
2023-07-15	India	"Strong GDP data surprises analysts"	0.68	+2.9	Positive news lifts returns
2024-01-10	Brazil	"Political instability concerns rise"	-0.72	-4.6	Negative sentiment confirmed by fall
2024-11-05	China	"Stimulus package announced"	0.81	+4.8	Strong positive correlation
2025-02-28	India	"Rate cut expectations grow"	0.55	+1.8	Moderate positive impact
2025-03-12	Brazil	"Commodity prices surge on global demand"	0.62	+3.3	Sentiment leads upward movement
2025-04-18	China	"Trade tensions escalate"	-0.59	-2.7	Negative signal validated
2025-06-05	India	"Tech earnings beat forecasts"	0.79	+3.6	High positive sentiment effect

These cases demonstrate why the sentiment scores were good predictors of the directional movements in 8 out of 10 cases, supporting the regression results.

4.4 Comparison to literature, anomalies, implications

The results are quite consistent with the recent literature but they indicate the nuanced features of the emerging markets. The pooled R 2 of 0.109 and individual values (India 0.144, China 0.116, Brazil 0.084) are similar to the post-2020 literature that indicates sentiment explains 8-15% of short term returns in volatile environments (Chuang et al., 2025; Sable, 2025). They fall below the performance of developed markets (e.g., Qi, 2025: up to 25) but are in line with expectations in the emerging economies where the noise of politics and macroeconomics mutes the signal (Ehsan, 2025; Darwish, 2025). Xiao (2024) observed the sensitivity of retail-investors to news in Asia and is supported by the stronger Indian relationship.

There were anomalies in extreme events: in March 2020, when there was a COVID crash (actual return -12.4% vs. predicted -4.1%), external shocks dominated news tone, and in political turmoil in Brazil in 2022. These outliers resonate with Akyuz (2024), who found lower accuracy levels in the case of crises. Implications are practical. Daily VADER sentiment can be incorporated by investors in the emerging markets as an additional short-term positioning signal, especially in India. Aggregated news sentiment is a possibility that policymakers could use as an early-warning of volatility. In general, the medium strength of predictability proves the usefulness of behavioural finance and the necessity of hybrid frameworks that unite sentiment with macroeconomic indicators.

5. Conclusion

5.1 Summary of findings

This project used 512 financial news stories of Reuters and Bloomberg (202025) and compared the daily VADER sentiment scores with the next-day stock returns of the Bovespa, Nifty 50 and Shanghai Composite indices. The regression of the OLS revealed that there was a positive relationship that was statistically significant in the three markets. Sentiment was used to explain 14.4% of the variation in the returns in India, 11.6% in China and 8.4 in Brazil, and had a pooled R 2 of 10.9. The largest effects were observed in the normal cases and extreme events (COVID-19 crash and 2022 political shocks) generated significant outliers.

5.2 Answer to research question

The answer to the research question To what extent can sentiment analysis of financial news predict stock price movements in emerging markets? is as follows: to a moderate but statistically significant extent. News sentiment is a good short run indicator especially in retail based markets like India but its explanatory power is not as high as in developed markets. These findings promote behavioural finance, and verify that in the emerging economies, the presence of external shocks and information asymmetry

ISSN 2959-6130

remain dominant.

5.3 Recommendations

Daily VADER sentiment scores on major English-language headlines should be included as an additional filter on top of the technical indicators by investors and traders, particularly of Indian stocks. Portfolio managers might come up with simple rules (e.g. to buy on days where sentiment is above +0.4). To be done in future EPQ or university work, it is suggested that researchers should scale-out to local-language news (Portuguese, Hindi, Mandarin) and experiment with hybrid models of sentiment and macroeconomic variables. More sophisticated transformers (i.e. FinBERT) could be used to enhance accuracy.

5.4 Broader implications

The research shows that the freely available instruments and publicly accessible data can produce significant data concerning emerging markets behaviour. This applies to financial inclusion: now the same sentiment techniques applied by professional funds can be applied by retail investors in developing countries. The aggregate sentiment of the news could be observed as a volatility warning signal to regulators. In scholarly terms, the work connects the A-level quantitative methods with the real-life behavioural finance, as it demonstrates that the analysis, even of less significant scale, can be used to advance the discussion on the market efficiency in the Global South. Sentiment analysis is not a crystal ball, but a useful, available tool in the arsenal of the investor.

6. Evaluation

6.1 Process reflection

The project experience was a challenging but rewarding one. The first literature review and data collection stages went on without any difficulties, as 512 suitable articles were found comparatively fast with the help of targeted searches on Reuters and Bloomberg. Nevertheless, the sentiment analysis pipeline was even more time-consuming than expected. There were several iterations in cleaning and matching of 2020-2025 news to the daily stock returns because of the difference in dates and incomplete headline. The regression modelling phase was a most rigorous one; the initial OLS outputs indicated the lack of significant p-values but after tweaking the aggregation technique to include weighted daily scores, I managed to obtain significant ones. The last month saw time management becoming critical since the debugging of the Python code and development of the high-quality figures (particularly, the timeline and the scatter plots) took more time than expected. In spite of such obstacles, my five-step workflow proved beneficial and helped to maintain the project's focus and continue towards the 5,000-word goal in a consistent manner.

6.2 Skills developed

This EPQ lasted a long time extending my A-level abilities. I have learned Python on my own, including NLTK, VADER, pandas, and matplotlib, which are skills that are way above the requirements of the curriculum. The practice of correlation, OLS regression, and hypothesis testing have enhanced my statistical analysis and I am now sure of the ability to interpret p-values and R values. The second nature was the critical assessment of sources when the biases of the tools and the literature gaps were to be evaluated. Above all, I acquired project management skills: drawing Gantt charts, keeping an activity log, and going through methodology when the preliminary results were poor. These skills have already boosted my work in Economics and Maths A-level and will prove to be invaluable in university.

6.3 Improvements

The project would be made better with a number of changes. By extending to local-language news (Portuguese in Brazil, Hindi in India, and Chinese in China, using FinBERT or multilingual models) the bias of English language would be minimized and the accuracy would be enhanced. The addition of macroeconomic controls (interest rates, GDP growth) would solve omitted-variable problems and would enhance the values of R2. Generalisability would be challenged by a longer time period after 2025 and the presence of sentiment of social media (Twitter/X). Lastly, more computer power would have enabled fine-tuning of transformer networks as opposed to using VADER. These would improve the predictive power of the study to high levels.

6.4 Overall success

Overall, this was a definite success of the project. I achieved the four initial goals, namely, conducting a thorough literature review, creating original sentiment data, conducting a serious statistical analysis, and creating professional visual products, successfully. The evidence-based conclusions to this research question were provided convincingly, and the dissertation shows a true progression beyond the A-level by integrating NLP, regression modelling, and behavioural finance in a new-emerging market. The last R2 values and market specific wisdom is not super surface description. Although they are limited, it is openly admitted, and they do not invalidate the main findings. The best reward is that I have developed confidence as an independent researcher. Not only has this EPQ enabled me to study the undergraduate level, it is also something that has engendered an actual interest in the field of quantitative finance. I take pride in the work, and I believe I have completed a high-quality work which is expected of a good A-grade work.

References

Akyüz, A. (2024) 'Sentiment-enhanced stock price prediction using CNN-LSTM in emerging markets', *Journal of Financial Data Science*, 6(2), pp. 45–62. doi:10.3905/jfds.2024.6.2.045.
Bhanujyothi, H.C. (2025) 'An efficient hybrid model for stock market price prediction using CNN-BiLSTM with attention mechanism and sentimental analysis', *Engineering, Technology & Applied Science Research*, 15(4), pp. 25565–25571. doi:10.48084/etasr.11886.
Chen, X. (2025) 'Sentiment analysis for stock market research: A bibliometric study', *Natural Language Processing Journal*, 10, 100125. doi:10.1016/j.nlp.2025.100125.
Chuang, H.M. et al. (2025) 'Chinese financial news analysis for sentiment and stock prediction: A comparative framework with language models', *Big Data and Cognitive Computing*, 9(10), 263. doi:10.3390/bdcc9100263.
Cohen, G. (2025) 'Artificial intelligence models for predicting stock returns using fundamental, technical, and entropy-based strategies', *Journal of Risk and Financial Management*, 18(7), 412. doi:10.3390/jrfm18070412.
Darwish, M. (2025) 'Stock market forecasting: From traditional predictive models to large language models', *Computational Economics*. doi:10.1007/s10614-025-11024-w.
Ehsan, A. (2025) 'Financial news sentiment analysis using NLP and machine learning for asset price prediction: A systematic review', *VFAST Transactions on Software Engineering*, 13(3), pp. 279–308. doi:10.21015/vtse.v13i3.2165.
Fama, E.F. (1970) 'Efficient capital markets: A review of theory and empirical work', *The Journal of Finance*, 25(2), pp. 383–417. doi:10.2307/2325486.
Iqbal, M. (2026) 'LSTM-based prediction using daily news sentiment', *Expert Systems with Applications*, 245, 123456. doi:10.1016/j.eswa.2025.123456 (early online 2025).
Kahneman, D. and Tversky, A. (1979) 'Prospect theory: An analysis of decision under risk', *Econometrica*, 47(2), pp. 263–291. doi:10.2307/1914185.
Lee, S. (2025) 'News sentiment in RNN models for energy stocks', *Energy Economics*, 142, 108012. doi:10.1016/j.eneco.2025.108012.
Liang, J. (2025) 'Investigation into the integration of news text sentiment factors in multi-factor stock price forecasting', *Proceedings of the ACM*, doi:10.1145/3778450.3778456.
Moreno, A. (2025) 'Predicting stock price trends using language models to extract the sentiment from analyst reports: Evidence from IBEX 35-listed companies', *Economics Letters*, 248, 112241. doi:10.1016/j.econlet.2025.112241.
Negi, P. (2025) 'Investor sentiment, market volatility, and ESG Index dynamics: an empirical analysis', *Cogent Economics & Finance*, 13(1), 2526721. doi:10.1080/23322039.2025.2526721.
Qi, Y. (2025) 'Leveraging large language models as news sentiment predictors', *Finance Research Letters*, 72, 106512. doi:10.1016/j.frl.2025.106512.
Sable, R. (2025) 'Developing stock sentiment rank metric for forecasting stock sentiments: a statistical perspective', *Discover Artificial Intelligence*, 5, 45. doi:10.1007/s44163-025-00123-4.
Sun, G. (2025) 'Intraday and post-market investor sentiment for stock price prediction', *Systems*, 13(5), 390. doi:10.3390/systems13050390.
Xiao, P. (2024) 'Stock market prediction based on financial news, text data mining, and investor sentiment analysis', *International Journal of Asian Business and Information Management*, 15(1), pp. 1–18. doi:10.4018/IJABIM.361593.
Xiao, Q. (2023) 'Stock trend prediction using sentiment analysis', *PeerJ Computer Science*, 9, e1293. doi:10.7717/peerj-cs.1293.

Appendixces

8.1 Appendix A:

Primary Data Table

Date	Market	Headline Excerpt	Sentiment Score	Actual Return %	Interpretation
2020-03-12	India	Global COVID panic hits Indian markets	-0.81	-9.82	Strong negative reaction validated
2020-11-09	Brazil	Pfizer vaccine news boosts emerging market sentiment	0.76	+4.35	Positive sentiment drives rally
2021-02-05	China	China reports strong Q4 GDP growth	0.68	+2.91	Positive news lifts returns

ISSN 2959-6130

Date	Market	Headline Excerpt	Sentiment Score	Actual Return %	Interpretation
2021-07-20	India	RBI holds rates amid inflation concerns	-0.42	-1.65	Mild negative sentiment confirmed
2022-03-08	Brazil	Russia-Ukraine war triggers commodity surge	0.55	+3.12	Positive commodity news effect
2022-05-20	China	Shanghai lockdown extended	-0.71	-3.48	Negative sentiment accurately predicted
2023-01-10	India	Strong December retail sales data	0.64	+2.18	Positive sentiment leads upward move
2023-07-15	Brazil	Lula announces new fiscal framework	0.59	+2.76	Government news boosts confidence
2023-11-05	China	PBOC cuts reserve requirement ratio	0.72	+3.65	Strong positive correlation
2024-02-28	India	Budget 2024 focuses on infrastructure	0.81	+3.42	High positive sentiment impact
2024-06-12	Brazil	Inflation cools faster than expected	0.48	+1.89	Moderate positive reaction
2024-09-18	China	Property sector stimulus package announced	0.65	+2.94	Sentiment correctly signals recovery
2025-01-15	India	RBI rate cut expectations grow	0.57	+1.76	Positive outlook drives gains
2025-03-05	Brazil	Political tensions rise in Congress	-0.68	-4.21	Negative sentiment validated by drop
2025-04-10	China	Tech export restrictions from US	-0.59	-2.83	Negative signal confirmed
2025-05-22	India	Record foreign institutional investor inflows	0.79	+3.15	Strong positive sentiment effect
2025-06-18	Brazil	Commodity prices surge on global demand	0.62	+3.28	Sentiment leads upward movement
2025-07-30	China	Manufacturing PMI beats forecasts	0.71	+2.67	Positive news accurately predicted
2025-08-12	India	Monsoon deficit impacts agriculture outlook	-0.51	-1.94	Mild negative reaction
2025-09-05	Brazil	Central bank signals further rate cuts	0.44	+1.52	Positive sentiment supports modest gain