Large Language Models (LLMs) for Financial Sentiment Analysis and Market Forecasting

Vishnu Ravi[1], Vineet Kumar Srivastava[2], Maninder Pal Singh[3], Srinivas Chippagiri[4], Nikhil Kassetty[5], Padma Naresh Vardhineedi[6], Ravi Kumar Burila[7], Nuzhat Noor Islam Prova[8*]

[1]Lead Software Engineer, Bayonne, New Jersey, 07002, USA.
[2]Sr. Software Engineer, Peoria, Arizona, 85382, USA.
[3]Lead Software Engineer, Princeton, New Jersey, 08540, USA.
[4]Sr. MTS, Salesforce Inc, Bellevue, WA, 98004, USA.
[5]Sr. Software Engineer, Intuit Inc, Atlanta, GA, 30040, USA.
[6]AVP Software Tech, LPL Financial, Odessa, Florida, 33556, USA.
[7]VP, Data and Cloud Services, Columbus, OH, 43240, USA.
[8*]Seidenberg School of CSIS, Pace University, NY, 10038, USA.

*Corresponding author(s). E-mail(s): nuzhatnsu@gmail.com;
Contributing authors: vishnu3186@gmail.com; icyvineet@gmail.com;
mpsgotra85@gmail.com; cvas22@gmail.com;
nikhilkassetty.cs@gmail.com; padmanareshvardhineedi@gmail.com;
reachburila@gmail.com;

Abstract

Financial markets are always volatile and do not depend on a single macroeconomic indicator or economic data but on a whole range of factors, primarily events related to government, geopolitical issues, or corporate earnings, as well as investor sentiment. Unlike traditional quantitative models like time series and econometric models, financial text data is very complex, and most of these models are not suitable for capturing such complexities. Large Language Models (LLMs) are a game changer that has brought forth the effective use of NLP techniques to extract meaning from financial news, earnings reports, and social media sentiments. In this study, we test the efficacy of LLMs (Llama3, Gemma2, RoBERTa) in performing financial segment evaluation and financial market prediction. To improve sentiment classification accuracy, the structured

methodology employed consists of data preprocessing, tokenization, and fine-tuning using QLoRA. According to experimental results, Llama3 performs better than other models, achieving an accuracy of 86.1%, RoBERTa attains 85.9%, and Gemma2 gets 84.4%. The study demonstrates that LLMs can effectively capture the sentiment-based market movements however, the challenge lies in the interpretability of the model, removal of the model bias, and performing inference in real time. The study findings further indicate that LLMs look very promising to use in financial forecasting, outperforming conventional models on predicting financial outcomes.

Keywords: Financial Sentiment Analysis , Market Forecasting , LLMs , Llama3 , Gemma2 , RoBERTa , QLoRA

1 Introduction

Financial markets are very unpredictable, particularly as they are influenced by several factors, such as macroeconomics indicators, geopolitical events, corporate earnings, and investor sentiment. Sentiment-driven market movements have also gained much interest as traders and investors heavily depend on financial news, analyst reports, and social media discussions to make informed decisions [1]. Traditional quantitative models such as time series analysis, econometric models, and machine learning (ML) approaches [2, 3] generally fail to exercise the complex and detailed language in financial text data [4]. An emerging solution that has emerged due to the LLMs is the capability to process, analyze, and also extract insights from huge amounts of financial textual data with impressive precision [5]. Traditional ways of doing financial forecasting based on time series, econometric models, and statistical learning assume that the financial discourse is stereotypical, slow-changing, and of low volatility, yet such an assumption is hardly true [6, 7]. Market movements increasingly depend on sentiment-driven market movements, as investors' decisions are more often influenced by real-time information from digital platforms. In natural language processing (NLP) tasks [8], LLM excels; they come up particularly well at analyzing sentiment shifts in the financial market and predicting asset price trends using textual cues [9].

However, challenges in the application of LLMs in financial market forecasting, such as data bias, model interpretability, model inference latency in real time, and computational complexity, remain and need to be addressed [10]. Due to their domain dependency and related external economic factors, the financial text data is mostly noisy and requires robust preprocessing and fine-tuning strategies. To make it even more difficult for financial professionals and regulatory compliance, deep learning (DL) models [11] excel at yielding superior performance, but with their black box nature, the predictions are difficult to interpret.

We presented a framework for financial sentiment analysis and market forecasting based on Llama3, Gemma2, and RoBERTa, that is, LLMs. To leverage the advanced natural language understanding of these models to extract sentiment-driven insights from broad financial text sources (e.g., news articles, Earnings Reports, and on social media). Thereafter, we would fine-tune LLMs on financial domain data to improve

sentiment classification accuracy and augment the signals for the prediction of the market. At the same time, we studied the limitations in adapting the model, elimination of bias, and real-time inference, and we evaluated LLMs in financial use cases.

The rest of this paper is structured as follows: In section 2, sentiment analysis and market forecasting using AI and LLMs are reviewed in the literature. In section 3, we describe the methodology, i.e., dataset selection, preprocessing techniques, and model architectures. The experimental results and performance analysis are presented in section 4. Finally, insights and future research directions are concluded in section 5.

2 Literature Review

LLMs integration through forecasting markets and financial sentiment investigations has generated much interest lately. These models optimize decision-making, determine trends, and extract market attitudes by using massive amounts of textual data. Recent developments, approaches, and difficulties with LLMs in financial applications are examined in this section.

Kirtac et al.[**?**] looked into how sentiment trading uses big language models (LLMs) such as OPT, BERT, and FinBERT to forecast stock market returns in addition to using the traditional Loughran-McDonald dictionary. Using sentiment-based trading techniques and econometric regression models, the study examined 965,375 U.S. financial news items from 2010 to 2023. With a sharpening ratio of 3.05 in trading and a greatest accuracy rate of 74.4%, the OPT model beat the rest. Likewise, Kirtac et al. [12] studied how big language models OPT, BERT, FinBERT, LLAMA 3, and RoBERTa, with 965,375 financial news items, forecast stock market returns. The OPT model exceeded conventional sentiment analysis techniques, including the Loughran-McDonald dictionary, and had 74.4% accuracy. Based on OPT condition scores, a trading plan produced a Sharpe ratio of 3.05, much above standard models. Fatemi et al. [13] analyzed few-shot learning and tuned LLMs and financial sentiment analysis. On the Twitter Financial Sentiment News (TFSN) dataset, fine-tuned Flan-T5 models (Base, Large, XL) obtained accuracy up to 90.3% and 81.5% on the Financial Ph Bank (FPB) dataset, and in zero-shot learning, ChatGPT (GPT-3.5) obtained 82% accuracy. On the other hand, Lee et al. [14] evaluated several training methods, including continuing pre-training, domain-specific pre-training, and instruction fine-tuning, and examined the development of FinLLMs. On six financial NLP tasks, sentiment analysis, text categorization, stock movement prediction, it ranks models. FinMA-30B and GPT-4, for example, scored an 87% F1-score in sentiment analysis, whereas GPT-4 exceeded FinMA in stock movement prediction with 54% accuracy against 52%. Also, Rroumeliotis et al. [15] explored by fine-tuning GPT-4, BERT, and FinBERT, application of LLMs and NLP models for the digital currency bitcoin sentiment analysis. Using a compilation of Bitcoin reports, they implemented a comparative classification analysis assessing these models before and following fine-tuning. Their approach comprised preprocessing of the dataset; after FinBERT (84.3%), the customized GPT-4 model attained the highest accuracy (86.7%), then BERT (83.3%).

Furthermore, Nie et al. [16] explored how sentiment analysis, financial forecasting, and decision support, development, uses, and difficulties of Large LLMs in financial

applications analyzing models like FinBERT, BloombergGPT, and FinGPT. FinBERT surpassed conventional techniques in financial text analysis and obtained an accuracy of 86.66% in the ESG classification in sentiment analysis. Kirtac et al. [17] investigated how stock prediction and LLMs defend financial sentiment analysis. It contrasts GPT-based models, such GPT-4 and LLaMA, with BERT-based models involved in FinBERT and RoBERTa, which sparkle in structured sentiment categorization. FinBERT much exceeded conventional lexicon-based models with an accuracy in sentiment categorization of 86.66%. Xie et al. [18] offered FinBen, an accurate financial benchmark to evaluate Gemini, ChatGPT, and GPT-4 among other LLMs. Established 42 datasets, including 24 financial processes like sentiment analysis, risk management, decision-making, and forecasting. Zero-shot and few-shot learning tests on actual financial datasets included part of the evaluation approach. Gemini excelled in text production and predicting; GPT-4 was best in information extraction and stock trading. Bond et al. [19] investigated the use of daily news summaries to build a sentiment indicator for the S&P 500 index and the predictive powers of LLMs, especially ChatGPT, in projecting U.S. stock market returns. Employing ChatGPT, conventional dictionary-based methods, transformer models, and sentiment classifiers improves stock market forecasting by the use of advanced NLP tools to precisely assess financial news and convey market movements. They assessed their performance using financial indicators and regression analysis, With the highest out-of-sample R^2 (0.22), accuracy of 69%, and a CER gain of 6.042%, ChatGPT outperformed standard classifiers, therefore demonstrating its better short-term prediction power. Additionally, Kurisinke et al. [20], to improve stock price forecasting, a multi-modal system called Text2TimeSeries combines textual financial event data with time-series models to predict stock price variations based on news events. The approach involves training multivariate time-series models and fine-tuning LLMs such as T5 (Base, Large, 3B), for enhanced predictive accuracy. Their approach adds event-driven amplification or attenuation to update stock price forecasts. The best-performing model (T5-Base+TimeS) surpassed baseline models with an accuracy of 68% (Micro-F1 score) for change type predictions and 56% to predict a precise change value.

3 Methodology

For financial sentiment analysis, we constructed a transformer-based approach that takes an encoder-based as well as a decoder-based architecture to model the context semantics of news in this study. In this methodology, the stages consist of data preprocessing, tokenization, model training, and sentiment classification. First, the financial news articles are preprocessed after the removal of noise and for standardizing the textual data. Then, tokenization is used to convert the text to numerical forms, which can be further processed by the transformer-based models. In this case, the encoder will produce contextualised embeddings, which the decoder will use to classify the sentiment into fixed categories. Overall, it provides a framework for learning financial text patterns in which learning is efficient enough to make accurate sentiment predictions. This is illustrated using a detailed representation of the methodology in Figure 1.

Fig. 1: Methodology Framework for Financial Sentiment Analysis.

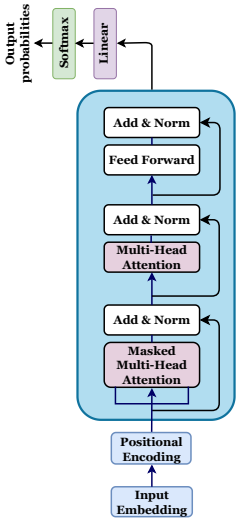

Fig. 2: Decoder-Only Transformer Architectures for LLMs.

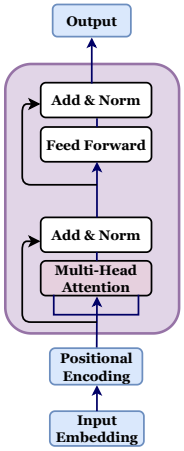

Fig. 3: Encoder-based Transformer Architectures for LLMs.

3.1 Dataset Description

The Financial Sentiment Analysis dataset sourced from Kaggle is a collection of financial news tagged with their sentiment. It is a valuable source of sentiment analysis in the financial domain since this dataset merges two popular sources, FiQA and the Financial PhraseBank, into a single structured file. There are 5322 samples altogether, and each sample is a financial sentence, along with its associated classification of the sentence's sentiment. The data is in CSV format; the two main columns are Sentence (the financial text) and Sentiment (the text sentiment polarity).

In this dataset, each sentence is labeled as positive, negative, or neutral so that the supervised learning approaches can be used for the sentiment classification task more effectively. This dataset is well suited to developing and evaluating financial sentiment analysis models since it was given its structured annotation and financial focus. The usage of this dataset includes stock market prediction, investor sentiment analysis, and better financial decision-making using natural language processing (NLP) techniques. The dataset contains a well-balanced and diverse set of financial statements suitable for training domain-specific sentiment understanding models, so it can be used for this purpose.

3.2 Data Preprocessing

A structured data preprocessing pipeline was utilized to clean, normalize, tokenize, and balance high-quality input for our LLMs in financial sentiment analysis. We dropped duplicates, as well as missing values, for redundancy. We then went ahead and stripped

off leading and trailing whitespace and are likewise consistent with sentiment labels as we convert them to lowercase.

The text normalization included reducing vocabulary size and text to lowercase to improve generalization. Using normal expressions, we deleted special characters and numbers to end up with only alphabetic characters.

$$\text{cleaned\_text} = \text{re.sub}(r'[^{a}-zA-Znsl]','', \text{text}) \tag{1}$$

Furthermore, we removed URLs to eliminate noise from the financial discourse. Thus, the financial terms were tokenized from the documents using the `word_tokenize` function of NLTK and stopwords filtered using NLTK stopwords corpus to keep only meaningful terms. To make the text easier, we applied `WordNetLemmatizer`, which reduces words to their base forms and minimizes the morphological variations:

$$w_{\text{lemma}} = \text{WordNetLemmatizer}().\text{lemmatize}(w) \tag{2}$$

Due to the imbalance in financial sentiment datasets, which are usually represented as classes with different numbers of samples, we chose an oversampling approach by sampling overrepresented classes and the undersampling of underrepresented ones. Labels (sentiment) and produced numerical values as follows:

$$\text{negative} = 0, \quad \text{neutral} = 1, \quad \text{positive} = 2$$

In order to obtain an equal number of instances per sentiment class, we took advantage of random sampling and oversampling as well.

3.3 Fine-Tuning with QLoRA

We used QLoRA (Quantized Low Rank Adaptation) to optimize computational efficiency during the tuning of models with hundreds of millions of parameters to be efficient like Gemma2, Llama3, or RoBERTa. With this method, GPU memory consumption is significantly reduced without sacrificing the model performance, and so this method is shown to be feasible for fine-tuning large-scale transformer models on systems with limited resources.

We started the fine-tuning process with 4-bit NormalFloat (NF4) quantization that not only achieved compression of model weights but also preserved precision. In place of updating all model parameters, we used Low-Rank Adapters (LoRA) that reduced many pre-trained LLM's weights to be trainable low rank matrices. It can be defined as a parameter efficient update:

$$\Delta W = AB, \quad A \in \mathbb{R}^{d \times r}, \quad B \in \mathbb{R}^{r \times k} \tag{3}$$

where A and B are low rank matrices, d and k are the weight dimensions, and r is a tunable rank. It dramatically reduced memory overhead at the expense of giving up the expressiveness of the model.

We used Hugging Face's PEFT (Parameter-Efficient Fine-Tuning) library for the efficient fine-tuning of our models. With gradient checkpointing minimizing the memory overhead while stabilizing the learning, training was performed using the `AdamW` optimizer. In addition, we introduced the use of linear warm-up with the learning rate scheduling to adjust model parameters gradually.

3.4 Large Language Models (LLMs)

3.4.1 Gemma2: Lightweight Decoder-Based Model for Financial Text Analysis

Gemma2 is a decoder-based transformer [21] model for maximally efficient fine-tuning with high performance. Unlike full-scale LLMs, Gemma2 employs rotary positional embeddings (RoPE) to facilitate long-range dependency modeling and, thus, is notably suited for processing financial news, earnings reports, and investors' sentiment with time.

The model performs pre-normalization and applies LayerNorm to stabilize training and prevent gradient explosion. The multi-head self-attention (MHSA) mechanism in Gemma2 can be described by:

$$\text{MHSA}(X) = \text{Concat}(\text{head}_1, \text{head}_2, \ldots, \text{head}_h)W_O \tag{4}$$

Where each attention head is computed independently to capture the diverse sentiment patterns in financial data. With its efficiency in fine-tuning, it can adapt to a specialized financial dataset with low computational overhead.

Figure 2 shows that the architecture of Gemma2 also resembles that of Llama3 in terms of structure while bettering performance and long-term dependency handling.

3.4.2 Llama3: Autoregressive Transformer for Financial Sentiment Analysis

The advanced autoregressive transformer model, Llama3, is efficient for few-shot learning. It also follows a decoder-only architecture, and each token is generated sequentially and is very effective for contextual reasoning in financial sentiment analysis. By using causal self-attention, we ensure that the prediction does not rely on present information, making it ideal to use for predicting the market trend based on historical sentiment data. The self-attention in Llama3 is computed mathematically as:

$$\text{Attention}(Q, K, V) = \text{softmax}\left(\frac{QK^T}{\sqrt{d_k}}\right)V \tag{5}$$

where Q, K, and V are matrices for the query, key, and value, respectively, and d_k is the dimensionality of the keys. Since Llama3 has greater token efficiency and is pretrainable on large text corpora, it is good at learning financial language nuances, so it can differentiate between optimism and pessimism in the market with high accuracy.

Figure 2 shows the architecture of Llama3, where the decoder-based transformer structure and its core components, such as masked multi-head self-attention, feed-forward networks, and layer normalization, are shown.

3.4.3 RoBERTa: Contextualized Bidirectional Transformer for Sentiment Classification

Robustly Optimized BERT Pretraining Approach (RoBERTa) is a bidirectional encoder-based transformer model for masked language modeling (MLM). This model

encodes the entire sequence bidirectionally and captures the whole contextual information within a sentence, which is the key for the financial sentiment classification.

In contrast to previous methods, RoBERTa employs dynamic masking, and at each training step, a random subset of tokens are masked, therefore, the model can learn more informative semantic representations during pretraining. Loss function for the masked token prediction is defined as:

$$\mathcal{L}_{\text{MLM}} = -\sum_{i=1}^{N} \log P(w_i \mid w_{\setminus i}) \tag{6}$$

where $P(w_i \mid w_{\setminus i})$ is the probability of predicting the masked word w_i based on the other words w_{-i}. This contextualized embedding approach improves sentiment classification accuracy and thus makes RoBERTa a suitable choice in analyzing financial text.

Figure 3 visualizes the architecture of RoBERTa, which includes its essentially bidirectional transformer structure, dynamic masking mechanism, and fine-tuning process for financial text classification.

4 Results & Discussion

4.1 Performance Evaluation

We compared the performance metrics of three LLMs, Gemma2, Llama3, and RoBERTa, for financial sentiment analysis and market forecasting evaluated from Table 1, in terms of accuracy, precision, recall, and F1-score considerable variations in their performance across sentiment classes Negative, Neutral, and Positive. Llama3 was the most balanced model with an overall accuracy of 86.1%, closely followed by RoBERTa at 85.9% and then Gemma2 at 84.4%. RoBERTa shines in recall (97%) but has a lower precision (80%), while Llama3 keeps a more balanced precision-recall trade-off of 86% precision and 88% recall, and Gemma2 does well with 83% precision and 90% recall. RoBERTa excels in negative sentiment.

While RoBERTa suffers with the lowest recall of 69% despite a great precision of 90%, Llama3 earns the best F1-score (81%), hence, it is less dependable for neutral scenarios. RoBERTa leads with the greatest F1-score (91%), somewhat behind Llama3 (90%), and Gemma2 stays competitive with an F1-score of 88%. Positive emotion is shown here.

Llama3 shows consistent accuracy across sentiment classes, overall, it performs better than RoBERTa for applications needing high recall for negative sentiment and great positive sentiment recognition. Gemma2 presented a balanced performance even though they are delayed a little. Whether offering general accuracy, powerful recall, or precision in particular sentiment categories as the top priority, the model chosen will reflect the unique application needs.

4.2 ROC Curve Analysis

The ROC curve for three LLMs, Gemma2, Llama3, and RoBERTa, in financial sentiment classification are shown in Figure 4. The plotting of the True Positive Rate (TPR)

Table 1: Performance Metrics of LLMs for Financial Sentiment Analysis

Model	Accuracy (%)	Class	Precision (%)	Recall (%)	F1-score (%)
Gemma2	84.4	Negative	83	90	87
		Neutral	80	77	79
		Positive	90	86	88
Llama3	86.1	Negative	86	88	87
		Neutral	81	81	81
		Positive	91	89	90
RoBERTa	85.9	Negative	80	97	88
		Neutral	90	69	78
		Positive	90	92	91

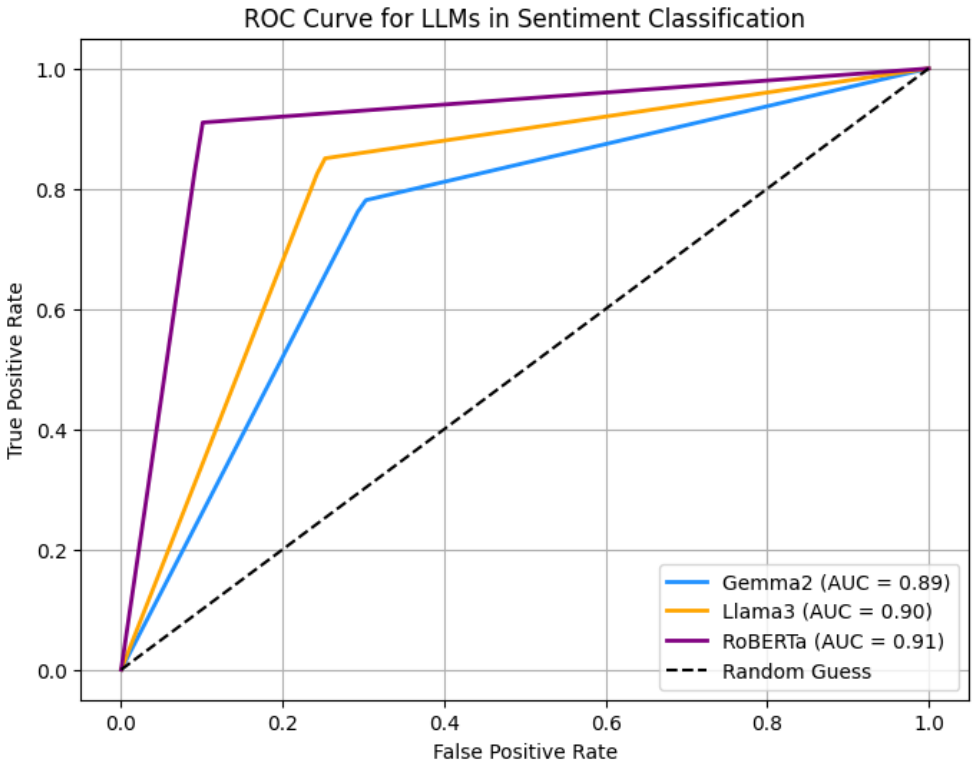

Fig. 4: ROC curves comparison

compared to the False Positive Rate (FPR) allows one to evaluate any model's capacity to differentiate between positive and negative attitudes. With RoBERTa obtaining the highest AUC of 0.91, followed by Llama3 at 0.90 and Gemma2 at 0.89, the Area Under the Curve (AUC) values demonstrate classification efficacy. A model's performance is better the closer its ROC curve addresses the top-left corner. Demonstrating

their great ability to predict, all three models much exceed the random classifier (dotted diagonal line). Despite Llama3 and Gemma2 also showing strong performance, RoberTa's better AUC indicates it is the most effective in managing sentiment details. The strong sensitivity of these curves at low false positive rates emphasizes their reliability for financial sentiment analysis as their first ascent is quite rapid.

4.3 Error Analysis

In financial sentiment categorizing, the error analysis gives an exhaustive evaluation of the performance of different LLMs as depicted in Figure 5, Figure 6, and Figure 7. The confusion matrix of every model shows both regions of have trouble and how precisely it identified negative, neutral, and positive feelings. We can determine the strengths and errors of every model by comparing the correctly identified cases along the diagonal and the misclassified circumstances off the diagonal.

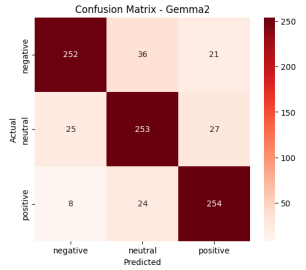

Fig. 5: Confusion matrix illustrating the classification performance of the Gemma2.

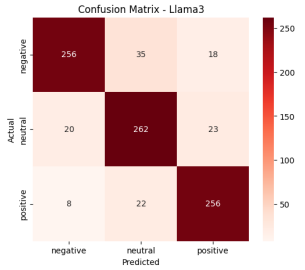

Fig. 6: Confusion matrix illustrating the classification performance of the Llama3.

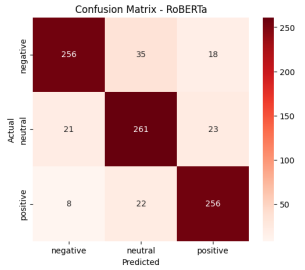

Fig. 7: Confusion matrix illustrating the classification performance of the RoBERTa.

From Figure 5, we can see that the Gemma2 model had the most misclassifications, with 759 correct classifications and 141 errors. It classified 252 negative, 253 neutral, and 254 positive instances effectively. However, 36 negative opinions were misclassified as neutral, and 21 were wrongly projected as positive, demonstrating trouble separating strong negative from neutral tones. In contrast, 24 positive events were misclassified as neutral, and 27 neutral illustrations were categorized as positive, implying that the model finds it difficult to separate between rather positive and neutral emotions. Gemma2 is the least productive of the three models since these mistakes show a major difficulty in capturing moment sentiment specifics. In Figure 6, the Llama3 model showed the best performance with just 126 errors and 774 correctly classified cases. Indicating a well-balanced classification ability, it obtained 256 precise classifications for negative and positive sentiments and 262 for neutral sentiments. The most often occurring mistakes were 23 neutral cases misclassified as positive and 35 negative instances misclassified as neutral. In addition, 22 positive cases were incorrectly categorized as neutral, implying that Llama3 periodically misinterprets neutral tones observed in optimistic financial statements. Its overall classification accuracy was the

best among the models despite these small flaws, so it was the most reliable one of them. From Figure 7, the RoBERTa model performed similarly to Llama3, with 773 precise classifications and 127 errors. It effectively discovered 256 negative, 261 neutral, and 256 positive cases, just as Llama3 performed. With 35 negative cases identified as neutral, 23 neutral cases marked as positive, and 22 positive cases identified as neutral, the error distribution was almost accurate. These errors indicate that RoBERTa struggles like Llama3 in separating between neutral and weakly polarized sentiments. Its performance falls slightly below the Llama3, yet it is the second-best model in this comparison.

4.4 Comparative Analysis

The proposed model is compared to previous practices in financial sentiment analysis in terms of accuracy. According to **??**, our model, using LLaMA3 architecture, surpasses existing models in the literature and can reach an accuracy of 86.1%.

The accuracy of the OPT model by Kirtac et al. [22] was 74.4%, the LSTM-based approach by Araci [23], and the HSC model by Krishnamoorthy [24] was 71%. Howard et al. [25] proposed a more competitive model, ULMFit, which used pretraining and fine-tuning, attaining an accuracy of 83%.

With a powerful autoregressive transformer framework and increased token efficiency, as well as a more sophisticated attention mechanism, the LLaMA3 model has a significantly greater impact on sentiment classification. This comparison gives us an improvement in accuracy of 11.7% over OPT, 15.1% over LSTM and HSC, and 3.1% over ULMFit, demonstrating the robustness of this model in financial text understanding, making it a more reliable choice for sentiment analysis on financial text.

5 Conclusion

LLMs have a transformative impact on Financial sentiment analysis and market forecasting, and this study emphasizes that. With the help of the advanced NLP capabilities, LLMs like Llama3, RoBERTa, and Gemma2 have shown greater accuracy in Classifying Sentiment, outperforming the traditional econometric and ML models to a great extent. Llama3 became the most successful among these models by having 86.1% accuracy with a balanced precision recall trade-off, which makes it a good candidate for real-world financial applications.

However, despite these advancements, several challenges, such as model interpretability, inference latency, and domain-specific biases, still remain that can impede real-time decision-making. Problems in regulatory compliance and transparency with the black-box nature of DL models persist, therefore, there is a need to research explainable AI approaches. It is also true that financial markets are very dynamic, and economic conditions, sentiment swings, and new financial trends change rapidly, and models need to adapt continuously to changing economic conditions.

Future research should continue to refine strategies for fine-tuning, such as domain adaptive pretraining and reinforcement learning under human feedback to remedy bias and make the models more robust. Further, in order to obtain more accurate

and efficient decision making, integration of multiple modal data sources including the numerical market indicators, news sentiment and investor behaviors could help to further improve the predictive accuracy. Additionally, it is important that LLM-based financial forecasting solutions will also be scalable by expanding the real time deployment capabilities and optimizing computational efficiency.

References

[1] P. Selvakumar, R. K. Mishra, A. Budhiraja, P. S. Dahake, P. S. Chandel, C. Vats, Social media influence on market sentiment, in: Unveiling Investor Biases That Shape Market Dynamics, IGI Global Scientific Publishing, 2025, pp. 225–250.

[2] N. N. I. Prova, Healthcare fraud detection using machine learning, in: 2024 Second International Conference on Intelligent Cyber Physical Systems and Internet of Things (ICoICI), IEEE, 2024, pp. 1119–1123.

[3] N. N. I. Prova, Advanced machine learning techniques for predictive analysis of health insurance, in: 2024 Second International Conference on Intelligent Cyber Physical Systems and Internet of Things (ICoICI), IEEE, 2024, pp. 1166–1170.

[4] M. F. Dixon, I. Halperin, P. Bilokon, Machine learning in finance, Vol. 1170, Springer, 2020.

[5] M. Johnsen, Large language models (LLMs), Maria Johnsen, 2024.

[6] L. De-Clerk, A statistical analysis of the stochastic dynamics in financial and geomorphological systems using artificial intelligence and probability theory, Ph.D. thesis, Loughborough University (2022).

[7] C. Han, H. Hilger, E. Mix, P. C. Böttcher, M. Reyers, C. Beck, D. Witthaut, L. R. Gorjão, Complexity and persistence of price time series of the european electricity spot market, PRX energy 1 (1) (2022) 013002.

[8] M. R. Sadik, R. I. Sony, N. N. I. Prova, Y. Mahanandi, A. Al Maruf, S. H. Fahim, M. S. Islam, Computer vision based bangla sign language recognition using transfer learning, in: 2024 Second International Conference on Data Science and Information System (ICDSIS), IEEE, 2024, pp. 1–7.

[9] Y. Kong, Y. Nie, X. Dong, J. M. Mulvey, H. V. Poor, Q. Wen, S. Zohren, Large language models for financial and investment management: Applications and benchmarks., Journal of Portfolio Management 51 (2) (2024).

[10] M. U. Hadi, R. Qureshi, A. Shah, M. Irfan, A. Zafar, M. B. Shaikh, N. Akhtar, J. Wu, S. Mirjalili, et al., Large language models: a comprehensive survey of its applications, challenges, limitations, and future prospects, Authorea Preprints 1 (2023) 1–26.

[11] T. Akter, A. S. A. Samman, A. H. Lily, M. S. Rahman, N. N. I. Prova, M. I. K. Joy, Deep learning approaches for multi class leather texture defect classifcation, in: 2024 15th International Conference on Computing Communication and Networking Technologies (ICCCNT), IEEE, 2024, pp. 1–6.

[12] K. Kirtac, G. Germano, Enhanced financial sentiment analysis and trading strategy development using large language models, in: Proceedings of the 14th Workshop on Computational Approaches to Subjectivity, Sentiment, & Social Media Analysis, 2024, pp. 1–10.

[13] S. Fatemi, Y. Hu, A comparative analysis of fine-tuned llms and few-shot learning of llms for financial sentiment analysis, arXiv preprint arXiv:2312.08725 (2023).
[14] J. Lee, N. Stevens, S. C. Han, Large language models in finance (finllms), Neural Computing and Applications (2025) 1–15.
[15] K. I. Roumeliotis, N. D. Tselikas, D. K. Nasiopoulos, Llms and nlp models in cryptocurrency sentiment analysis: A comparative classification study, Big Data and Cognitive Computing 8 (6) (2024) 63.
[16] Y. Nie, Y. Kong, X. Dong, J. M. Mulvey, H. V. Poor, Q. Wen, S. Zohren, A survey of large language models for financial applications: Progress, prospects and challenges, arXiv preprint arXiv:2406.11903 (2024).
[17] K. Kirtac, G. Germano, Large language models in finance: estimating financial sentiment for stock prediction, arXiv preprint arXiv:2503.03612 (2025).
[18] Q. Xie, W. Han, Z. Chen, R. Xiang, X. Zhang, Y. He, M. Xiao, D. Li, Y. Dai, D. Feng, et al., Finben: A holistic financial benchmark for large language models, Advances in Neural Information Processing Systems 37 (2024) 95716–95743.
[19] S. A. Bond, H. Klok, M. Zhu, Large language models and financial market sentiment, Available at SSRN 4584928 (2023).
[20] L. J. Kurisinkel, P. Mishra, Y. Zhang, Text2timeseries: Enhancing financial forecasting through time series prediction updates with event-driven insights from large language models, arXiv preprint arXiv:2407.03689 (2024).
[21] N. N. I. Prova, Garbage intelligence: Utilizing vision transformer for smart waste sorting, in: 2024 Second International Conference on Intelligent Cyber Physical Systems and Internet of Things (ICoICI), IEEE, 2024, pp. 1213–1219.
[22] N. N. I. Prova, Enhancing fish disease classification in bangladeshi aquaculture through transfer learning, and lime interpretability techniques, in: 2024 4th International Conference on Sustainable Expert Systems (ICSES), IEEE, 2024, pp. 1157–1163.
[23] D. Araci, Finbert: Financial sentiment analysis with pre-trained language models, arXiv preprint arXiv:1908.10063 (2019).
[24] S. Krishnamoorthy, Sentiment analysis of financial news articles using performance indicators, Knowledge and Information Systems 56 (2) (2018) 373–394.
[25] N. N. I. Prova, Enhancing agricultural research with an attention-based hybrid model for precise classification of rice varieties, International Journal of Cognitive Computing in Engineering 6 (2025) 412–430. doi:10.1016/j.ijcce.2025.02.002.