Wealth Wizardy: Crafting Automated Financial Insights

Deeptanshu
deeptanshu20371@iiitd.ac.in
Indraprastha Institute of Information Technology
Delhi, India

ABSTRACT

This work presents a novel stock prediction and insights generation system, leveraging state-of-the-art deep learning methodologies such as Recurrent Neural Networks (RNNs) and Long Short-Term Memory networks (LSTMs). The system's enhancements have led to significant improvements in prediction accuracy and its ability to discern complex temporal patterns within Nifty data.

To enhance accessibility and comprehension for laymen investors, the prediction model is integrated with Gemini LLM, employing a RAG based system. Notably, earnings call transcripts, company financial data and news articles are integrated into the predictive framework, providing a more comprehensive understanding of market sentiment and fundamental factors influencing stock movements.This integration enables the presentation of prediction insights in an easily understandable manner, facilitating informed decision-making for a broader audience.

The findings of this project demonstrate the efficacy of combining deep learning and natural language processing techniques for stock market prediction. The integration of earnings call transcripts analysis enhances the predictive power of the model, while the utilization of Gemini LLM ensures that the insights generated are accessible and actionable for both seasoned investors and novices alike.

KEYWORDS

RAG, SEC, LLM, RNN, LSTM, NIFTY

ACM Reference Format:
Deeptanshu. 2018. Wealth Wizardy: Crafting Automated Financial Insights. In *Proceedings of ACM Conference (Conference'17).* ACM, New York, NY, USA, 5 pages. https://doi.org/XXXXXXX.XXXXXXX

1 INTRODUCTION

Stock market analysis remains a profoundly challenging field due to its volatile and dynamic characteristics. This paper introduces an innovative approach to stock market insight generation by leveraging state-of-the-art language models and retrieval-augmented generation techniques. We integrate the capabilities of long short-term memory networks (LSTMs) with advanced language models such as Llama3 8B and Mistral7B, which are further enhanced by the Gemini RAG (Retrieval-Augmented Generation) system. Our system excels not only in enhancing the accuracy of financial forecasts but also in uncovering complex temporal patterns within the Nifty index, a crucial benchmark for the Indian stock market.

Additionally, our approach enriches the analysis by incorporating diverse data streams, including earnings call transcripts, comprehensive financial datasets, and timely news articles. This integration provides a richer context for understanding movements in stock prices and offers nuanced insights into market trends. The RAG system, with its sophisticated retrieval mechanisms, plays a pivotal role in contextualizing financial data in a format that is accessible and informative, even for individuals without a technical background.

Through our research, we demonstrate the significant benefits of using advanced machine learning models and multi-source data integration to improve the depth and interpretability of stock market analyses.

2 RELATED WORK

2.1 DP-LSTM: Differential Privacy-inspired LSTM for Stock Prediction Using Financial News

[Li et al. 2019] proposed a novel deep learning model, DP-LSTM, for stock price prediction12. The model leverages financial news articles as hidden information and integrates different news sources through a mechanism inspired by differential privacy12. The DP-LSTM model is built upon three key components: LSTM, VADER model, and a differential privacy (DP) mechanism12.

The authors first formulated a sentiment-ARMA model, which is an extension of the autoregressive moving average model (ARMA), by incorporating the information from financial news articles12. The LSTM-based deep neural network was then designed to reduce prediction errors and increase robustness12.

Extensive experiments were conducted on SP 500 stocks, and the results demonstrated that the proposed DP-LSTM model achieved a 0.32% improvement in mean MPA of prediction result12. Furthermore, for the prediction of the market index SP 500, the model achieved up to a 65.79% improvement in MSE12.

This work is particularly relevant to our project as it demonstrates the potential of incorporating news articles and differential privacy mechanisms into LSTM models for improved stock price prediction. We take a lot of inspiration for out project from this work.

The state of the art DP-LSTM model achieved and MPA (Mean Prediction Accuracy) of 0.9815. MPA is a metric used to evaluate the accuracy of a predictive model. It measures the average accuracy of predictions made by the model across multiple data points.

Conference'17, July 2017, Washington, DC, USA

ACM ISBN 978-1-4503-XXXX-X/18/06...$15.00
https://doi.org/XXXXXXX.XXXXXXX

The formula for calculating MPA is given by:

$$\text{MPA}_t = 1 - \frac{1}{L} \sum_{t'=1}^{L} \frac{|X_{t'} - \hat{X}_{t'}|}{X_{t'}}$$

where:

- MPA_t represents the mean prediction accuracy at time t.
- L is the total number of data points.
- $X_{t'}$ is the actual value of the target variable at time t'.
- $\hat{X}_{t'}$ is the predicted value of the target variable at time t'.

2.2 Enhancing Financial Sentiment Analysis via Retrieval Augmented LLMs

[Zhang et al. 2023] presents a novel approach to financial sentiment analysis. The authors identify that financial sentiment analysis is crucial for valuation and investment decision-making. However, traditional Natural Language Processing (NLP) models are limited by their parameter size and the scope of their training datasets, which hampers their generalization capabilities and effectiveness in this field1.

Recently, Large Language Models (LLMs) pre-trained on extensive corpora have demonstrated superior performance across various NLP tasks due to their commendable zero-shot abilities. Yet, directly applying LLMs to financial sentiment analysis presents challenges: The discrepancy between the pre-training objective of LLMs and predicting the sentiment label can compromise their predictive performance. Furthermore, the succinct nature of financial news, often devoid of sufficient context, can significantly diminish the reliability of LLMs' sentiment analysis.

To address these challenges, the authors introduce a retrieval-augmented LLMs framework for financial sentiment analysis1. This framework includes an instruction-tuned LLMs module, which ensures LLMs behave as predictors of sentiment labels, and a retrieval-augmentation module which retrieves additional context from reliable external sources1. Benchmarked against traditional models and LLMs like ChatGPT and LLaMA, their approach achieves 15% to 48% performance gain in accuracy and F1 score1.

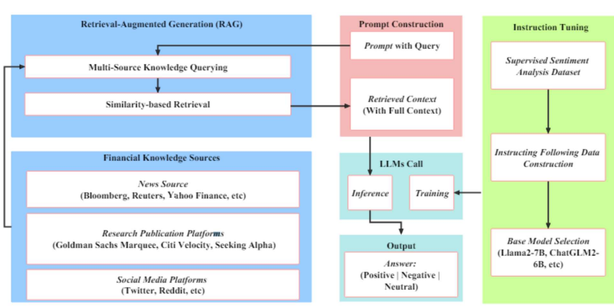

Figure 1: Framework of RAG LLM for financial sentiment analysis.

2.3 Data-Centric Financial LLM

The authors [Chu et al. 2023] propose a data-centric approach to improve the performance of Large Language Models (LLMs) in complex domains like finance1.

LLMs have shown promise for natural language tasks but struggle when applied directly to complex domains like finance. The authors identified that LLMs have difficulty reasoning about and integrating all relevant information.

To address these challenges, the authors proposed a data-centric approach to enable LLMs to better handle financial tasks. Their key insight is that rather than overloading the LLM with everything at once, it is more effective to preprocess and pre-understand the data.

They created a financial LLM (FLLM) using multitask prompt-based finetuning to achieve data pre-processing and pre-understanding. However, they faced a challenge where labeled data is scarce for each task. To overcome manual annotation costs, they employed abductive augmentation reasoning (AAR) to automatically generate training data by modifying the pseudo labels from FLLM's own outputs1.

Their experiments showed that their data-centric FLLM with AAR substantially outperforms baseline financial LLMs designed for raw text, achieving state-of-the-art on financial analysis and interpretation tasks1. They also open-sourced a new benchmark for financial analysis and interpretation.

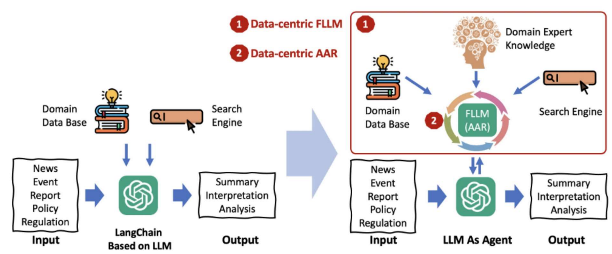

Figure 2: The framework of the financial large language model (FLLM)

3 METHODOLOGY

3.1 Data Collection

We initiated our research by collecting diverse datasets essential for stock market prediction. This involved scraping earnings call transcripts from reputable financial websites, retrieving SEC filings containing companies' financial data, and gathering news articles relevant to the stock market using keyword-based searches. The collected data was stored and processed in JSON format for further analysis.

3.2 Integration of External Data

To enhance the predictive capabilities of our model, we integrated the collected external data into our prediction framework. Earnings call transcripts provided valuable insights into companies' financial performance and future outlook, while SEC filings offered detailed information on financial statements and disclosures. News articles, retrieved based on specific keywords, were leveraged to gauge market sentiment and identify factors influencing stock movements.

3.3 Custom RAG Implementation

To parse and retrieve relevant information from the collected data, we implemented a custom Retrieval augmented generation (RAG) architecture. This architecture utilized advanced natural language processing techniques to extract key insights from earnings call transcripts, SEC filings, and news articles. We employed the FAISS library for efficient similarity search and retrieval of relevant content, ensuring that only pertinent information was fed into the subsequent processing stages.

3.4 Integration with LLM

To derive understandable and actionable insights from complex financial data, we integrated with open source models like Llama3-8B, Mistral-7B and Gemini Language Model (LLM) into our analytical framework. Utilizing a custom Retrieval-Augmented Generation (RAG) architecture, we processed and reformatted the extracted data to be suitable for input into the LLM. This strategic integration allows the LLM to generate concise, human-readable summaries and in-depth analyses. Such capabilities bridge the technical gap between raw financial data and strategic decision-making, enhancing both accessibility and utility for users across various domains.

3.5 Stock Price Prediction using LSTM

Finally, we trained a Long Short-Term Memory (LSTM) model on historical Nifty data to predict future stock prices. The LSTM model, a type of recurrent neural network (RNN), was chosen for its ability to capture long-term dependencies in sequential data, making it well-suited for time series forecasting tasks. We evaluated the model's performance using appropriate metrics and refined our prediction system based on the obtained results. Our final accuracy came out to be around 97% (correct prediction is += 0.05 of the ground truth). The mean prediction accuracy was 0.961, compared to the 0.982 state of the art model [Li et al. 2019].

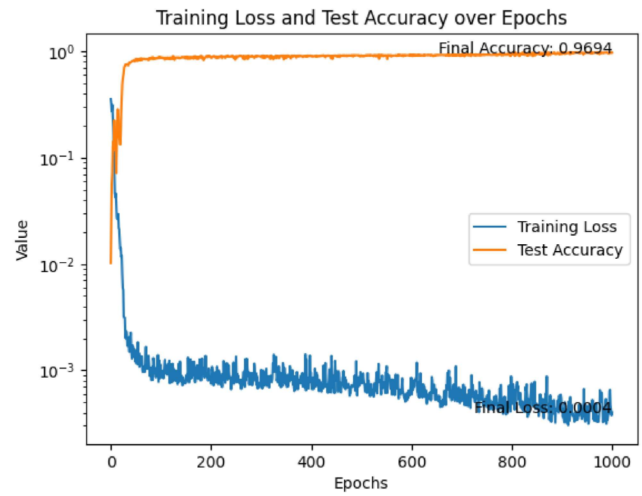

Figure 4: Loss vs Accuracy plot while training the LSTM model

4 RESULTS

Our predictive model was evaluated through a series of real-world queries to assess its analytical capabilities and the effectiveness of the integrated data sources. One such query was: "Should I invest in NVIDIA and why?" This query was chosen to test the model's

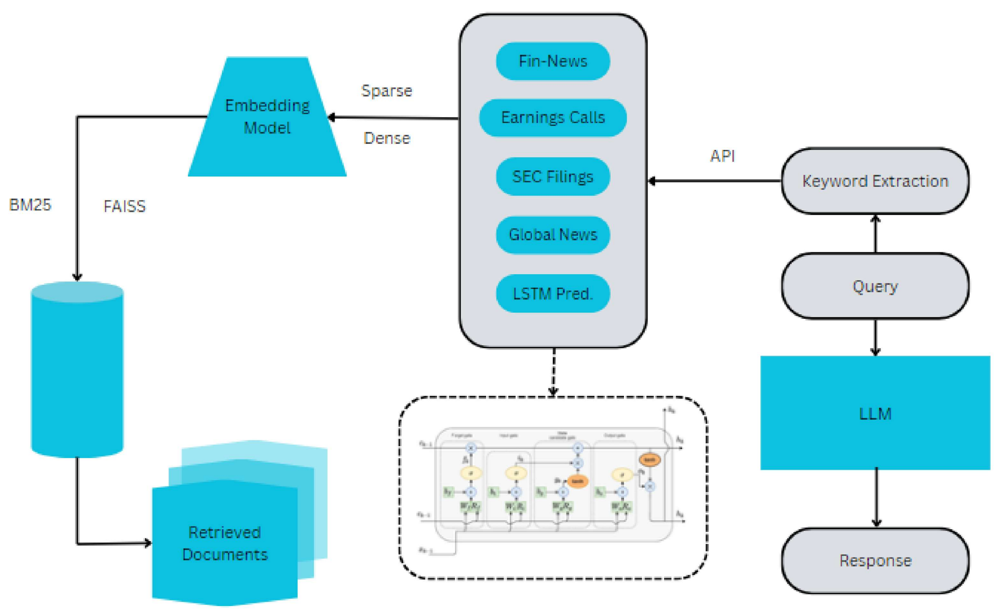

Figure 3: Pipeline for the proposed system.

ability to synthesize complex financial data and generate actionable investment insights.

4.1 Query Analysis: NVIDIA's Investment Potential

Title: NVIDIA's Market Outlook: A Comprehensive Analysis

1. Introduction: The query initiated an analysis focused on NVIDIA's position within the competitive AI market landscape. Our model utilized data from multiple streams including financial reports, market trends, and news articles to provide a structured response.

2. Detailed Analysis:

a. Financial Health: The model pointed out the lack of comprehensive financial data in the immediate response but suggested avenues for deeper financial investigation to ensure informed decision-making. b. Market Trends: NVIDIA has shown a strong growth trajectory, reinforced by its sustained dominance in the GPU market. However, new competitive challenges, such as the formation of the UXL Foundation, underscore the need for strategic vigilance. c. Investment Risk: Identified risks include potential market share erosion due to emerging open-source technologies and the perpetual threat of technological obsolescence. d. Regulatory Impact: The analysis noted a gap in data regarding regulatory impacts, highlighting an area for future enhancement of the model's data integration.
3. Strategic Recommendations: To mitigate identified risks, the model advised:

Diversifying investment portfolios to reduce dependence on NVIDIA's market performance. Continuous monitoring of the competitive landscape, particularly developments from the UXL Foundation. Considering alternative investments within the broader AI sector to leverage overall market growth.

4. Conclusion: The decision to invest in NVIDIA should be aligned with individual financial goals and risk profiles. The model emphasized the importance of comprehensive analysis and continuous monitoring of both NVIDIA's financial health and market dynamics to make well-informed investment decisions.

Social Media Summary: "Is investing in NVIDIA a wise choice? While it remains a leader in AI, emerging competition and technological shifts could impact its market position. Stay informed and diversify your investment strategy. NVIDIAnvestment AIMarket InvestmentStrategy"

5 CONCLUSION

In this project, we have developed an advanced approach to financial insight generation that utilizes cutting-edge machine learning techniques, natural language processing, and comprehensive external data sources. By aggregating and synthesizing diverse datasets including earnings call transcripts, SEC filings, and news articles, our goal was to obtain a detailed understanding of the market dynamics.

Our implementation of a custom Retriever-Reader-Generator (RAG) architecture efficiently parsed and extracted pertinent information from the amassed data. Integration with the Gemini Language Model (LLM) was crucial in translating complex datasets into clear and actionable insights, thereby facilitating informed decision-making for financial analysts and investors.

Additionally, the application of Long Short-Term Memory (LSTM) networks underscored the capability of recurrent neural networks to identify significant temporal patterns within financial data. Although originally used for stock price forecasting, the broader application to financial trend analysis allowed us to uncover underlying market behaviors.

This project underscores the robust potential of integrating deep learning, natural language processing, and financial analytics to enhance the comprehensibility and applicability of financial data. Future research may expand upon our work by incorporating more diverse data sources, refining our model architectures, and further tailoring the synthesis of machine learning techniques with real-world financial strategies.

ACKNOWLEDGMENTS

Identification of funding sources and other support, and thanks to individuals and groups that assisted in the research and the preparation of the work should be included in an acknowledgment section, which is placed just before the reference section in your document.

REFERENCES

Zhixuan Chu, Huaiyu Guo, Xinyuan Zhou, Yijia Wang, Fei Yu, Hong Chen, Wanqing Xu, Xin Lu, Qing Cui, Longfei Li, Jun Zhou, and Sheng Li. 2023. Data-Centric Financial Large Language Models. arXiv:2310.17784 [cs.CL]

Xinyi Li, Yinchuan Li, Hongyang Yang, Liuqing Yang, and Xiao-Yang Liu. 2019. DP-LSTM: Differential Privacy-inspired LSTM for Stock Prediction Using Financial News. arXiv:1912.10806 [q-fin.ST]

Boyu Zhang, Hongyang Yang, Tianyu Zhou, Ali Babar, and Xiao-Yang Liu. 2023. Enhancing Financial Sentiment Analysis via Retrieval Augmented Large Language Models. arXiv:2310.04027 [cs.CL]