AI-Powered Stock Market Assistant: Integrating Real-Time Data, NLP, and Sentiment Analysis for Investment Decisions

Moulesh Manikandan[1], Navtej Singh Randhawa[2], Mohammad Abed[3], Supria Deka[4], Jinal Mehta[5]
[1,2,3,4,5]Leeds School of Business, University of Colorado, Boulder
Author Emails: [1]moulesh.manikandan@colorado.edu, [2]navtej.randhawa@colorado.edu, [3]mohammad.abed@colorado.edu, [4]supria.deka@colorado.edu, [5]jinal.mehta@colorado.edu

Abstract

This project showcases the development of an AI-powered Stock Market Assistant that delivers real-time insights on stock trends, financial sentiment, and investment recommendations via an integrated web interface. Retail investors face the key challenge of navigating a constant stream of overwhelming financial information, hindering their ability to distill insights and make confident investment decisions.
The solution integrates a user-friendly Streamlit web interface, financial data APIs, and natural language processing (NLP) capabilities including LLM-powered summarization and BERT-based sentiment analysis. A key feature includes fine tuning BERT on simulated financial news data to demonstrate improvements in domain-specific sentiment prediction. The assistant ultimately offers personalized buy/hold/sell decisions with explainable reasoning. This tool democratizes financial analysis for retail investors, enhancing decision-making accessibility.

Introduction

The financial market is a dynamic ecosystem, heavily influenced by news reports, corporate earnings announcements, economic data releases, geopolitical developments, and unforeseen global events. In today's digital era, investors are inundated with an overwhelming volume of unstructured financial content across various sources, including news websites, social media platforms, earnings calls, and analyst commentaries. The speed and complexity of information flow make it increasingly difficult for individual investors, traders, and even financial advisors to stay consistently informed, assess the implications, and make timely investment decisions.

Traditional methods of tracking financial markets often rely on manual research, human interpretation, and delayed reporting, which may not suffice in the era of real-time trading and high-frequency decision-making. Recognizing this challenge, our project seeks to bridge the information gap by leveraging the power of machine learning, large language models (LLMs), and real-time data processing technologies to create an AI-driven financial assistant. The core objective is to empower investors with actionable, distilled insights — cutting through the noise and delivering focused, sentiment-informed recommendations on stocks and market movements.

At the heart of the system is an integrated multi-stage pipeline, designed to automatically ingest real-time stock data and relevant news articles, process and summarize large amounts of unstructured textual information, assess the market sentiment using advanced natural language processing (NLP) models, and ultimately generate intelligent investment recommendations. This seamless workflow ensures that users are not only provided with concise and relevant summaries of events but also with sentiment-driven interpretations that can guide trading or investment strategies.

Work Flow Chart:

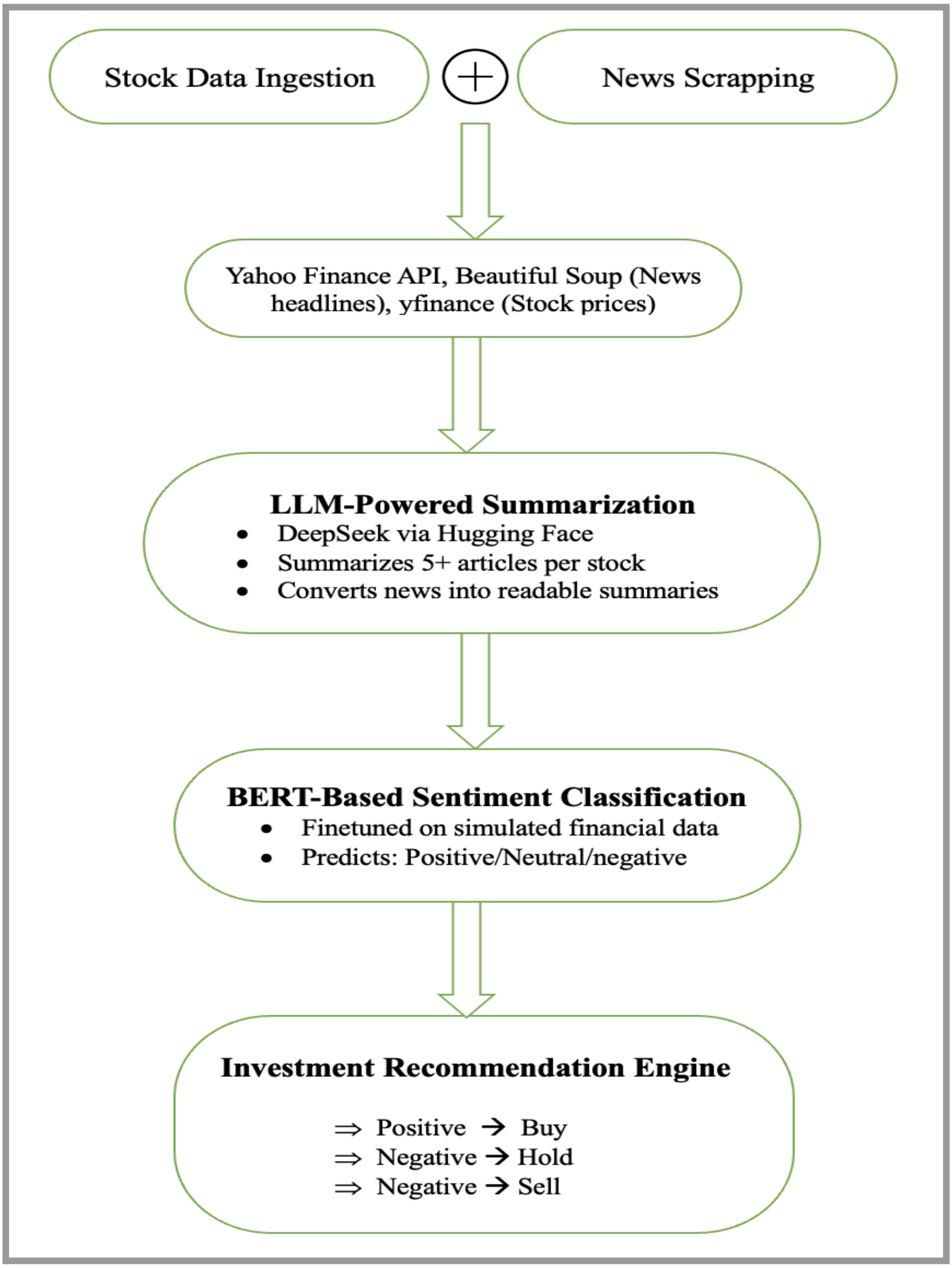

Figure 1: Workflow of the AI-powered Stock Market Assistant pipeline

Retail investors, often lacking access to advanced analytical tools, benefit from simplified, real-time insights, leveling the playing field with institutional investors. Currently, the assistant streamlines decision-making for individual investors; with further development, it could integrate with trading platforms to automate portfolio management. Our fine-tuned BERT model achieved 80% sentiment classification accuracy, enabling reliable buy/sell recommendations

Related Work

Several AI-based financial tools have been developed to assist investors, providing a benchmark for our project. For instance, Bloomberg Terminal leverages AI to deliver real-time financial data, news analytics, and sentiment analysis, primarily targeting institutional investors with comprehensive market insights [7]. While powerful, its high cost and complexity make it less accessible to retail investors, and its news analysis often focuses on structured data rather than real-time scraping of unstructured web content. Similarly, FinBERT, introduced by Araci (2019), adapts BERT for financial sentiment analysis, achieving improved accuracy on financial texts by fine-tuning on domain-specific datasets like earnings reports [2]. Unlike our approach, FinBERT focuses solely on sentiment classification and does not integrate real-time data fetching or summarization. In contrast, our Stock Market Assistant sets itself apart through its real-time news scraping using BeautifulSoup, which dynamically pulls the latest articles from Yahoo Finance, and LLM-powered summarization with DeepSeek, condensing multiple articles into concise, user-friendly insights. These features enable retail investors to access timely, actionable information without the cost barriers or technical complexity of tools like Bloomberg Terminal, while extending beyond FinBERT's scope by combining summarization, visualization, and recommendation in a single pipeline.
Devlin et al. (2018) introduced BERT for general NLP tasks, but its application to financial sentiment analysis requires domain-specific fine-tuning, as demonstrated by Araci (2019) with FinBERT, a BERT model tailored for financial texts [1], [2]).

Data Source

The yfinance library provided daily stock price data for a 6-month period (October 2024 to March 2025), covering 50+ tickers selected based on market capitalization and trading volume. For each ticker, we retrieved fields including open, high, low, close prices, and trading volume, resulting in approximately 180 data points per ticker (6 months × 30 days). News data was scraped from Yahoo Finance using the requests and BeautifulSoup packages, collecting 5–10 articles per ticker per query, depending on availability, to ensure sufficient coverage of recent events. Preprocessing was applied to both datasets to ensure quality. For stock price data, we handled missing values by forward-filling the most recent available price, ensuring continuity for time-series visualization. News headlines were preprocessed to remove HTML tags and special characters (e.g., &, <) using regular expressions, followed by normalization (lowercasing, removing extra whitespace) and tokenization with Hugging Face's BERT tokenizer to prepare the text for sentiment analysis.

For fine-tuning BERT, we created a simulated dataset of 120 labeled financial news headlines using the DeepSeek LLM. Prompts were designed to mimic real-world financial news with clear sentiment tones, ensuring a balanced class distribution across Positive, Negative, and Neutral labels. For example, a prompt like "Generate a positive financial headline about a company's earnings" produced: "TechCorp Reports Record-Breaking Q4 Earnings, Shares Soar 10%." Similarly, a negative headline example is: "PharmaInc Faces Major Setback as Drug Trial Fails, Stock Plummets." A neutral headline example is: "AutoCorp Releases Q1 Results, Meeting Market Expectations." This approach ensured each class was equally represented, with 40 headlines per class. Exploratory Data Analysis (EDA) revealed that real-world news headlines often contain mixed sentiments, necessitating fine-tuning to distinguish subtle financial contexts. Table 1 shows the sentiment label distribution of the simulated dataset, confirming its balance.

Sentiment Class	Number of Headlines	Percentage
Positive	40	33.3%
Negative	40	33.3%
Neutral	40	33.3%

Table 1: Sentiment Label Distribution of Simulated Dataset

Methods

This project was designed with modular components that interact through a seamless end-to-end pipeline. Below are the core technologies and design choices that enable this assistant to function in real time and offer accurate recommendations:

Frontend/UI:

We used Streamlit to design an interactive web application. It allows for dynamic content display, dropdown-based stock selection, clickable news links, and visual output of model recommendations. Streamlit was selected for its rapid prototyping and interactive visualization capabilities, ideal for user-friendly financial dashboards.

Data Fetching:

Stock price history and financial statements were fetched using the yfinance API, while recent news headlines were extracted from Yahoo Finance website using requests and BeautifulSoup packages. For each ticker, we retrieved high/low/close stock prices for trend visualization and relevant financials (income statement and balance sheet).

LLM for Summarization & Ticker Resolution:

We employed DeepSeek-R1-Distill-Gwen-32B via Hugging Face's hosted inference API to summarize 5+ news articles into an easy-to-read 10-line paragraph. It also resolved company names to stock tickers via natural language prompting, enhancing usability.

Sentiment Analysis:

We initially used BERT as a zero-shot sentiment classifier on the stock news. To improve its domain alignment, we later trained a custom model based on "BertForSequenceClassification" by simulating 120 labeled stock news summaries using DeepSeek. The sentiment classes were: Positive, Negative, and Neutral. We evaluated VADER for sentiment analysis but opted for BERT due to its superior performance on contextual text.

Model Fine tuning:

Training was performed using Hugging Face's Trainer API. The model was fine-tuned for one epoch on the simulated dataset with stratified label distribution. Evaluation showed substantial gains in precision, recall, and accuracy for all three sentiment classes, particularly for the underrepresented 'neutral' class. We adapted Hugging Face's open-source 'BertForSequenceClassification' implementation to suit our financial sentiment analysis task. Specifically, we fine-tuned the model on our custom dataset of 120 labeled financial news summaries, adjusting key hyperparameters to optimize performance. The learning rate was set to 2e-5 to ensure stable convergence, and we limited training to one epoch to balance accuracy gains with computational efficiency, given the dataset's size and stratified label distribution. Additionally, we modified the default Trainer API configuration by incorporating a batch size of 16 and enabling early stopping based on validation loss, which improved the model's ability to generalize to unseen financial news data.

Recommendation Logic:

Based on the model's sentiment prediction, the assistant gives AI-powered recommendations:

- Positive → **Buy**
- Negative → **Sell**
- Neutral → **Hold**

This logic aligns with traditional investor behavior, simplified for users.

Empirical Applications & Results:

Our AI-powered stock assistant was evaluated across multiple fronts—user experience, model performance, and potential impact on investor decision-making. Below are key findings from each component:

Interactive Stock Visualization:

A clean and responsive dashboard built with Streamlit and Plotly allows users to select a stock ticker and immediately explore the past six months of closing, high, and low price movements. This time-series visualization supports investors in identifying recent momentum shifts, price ranges, and technical cues—all without manual charting.

Figure 2:Stock Assistant App

LLM-Powered News Summarization:

Financial news headlines were scraped using BeautifulSoup and condensed into a single easy-to-understand summary using the DeepSeek LLM via Hugging Face’s inference API. Each summary is crafted in plain language, designed to be interpretable by non-experts. This drastically reduces information overload, helping users digest high-frequency financial content in seconds.

	2024-12-31 00:00:00	2023-12-31 00:00:00	2022-12-31 00:00:00	2021-12-31 00:00:00	2020-12-31 00:00:00
Treasury Shares Number	$24.78 Million	$16.08 Million	$1.56 Million	$1.56 Million	$nan
Ordinary Shares Number	$427.76 Million	$432.76 Million	$445.35 Million	$443.96 Million	$nan
Share Issued	$452.54 Million	$448.84 Million	$446.91 Million	$445.53 Million	$nan
Net Debt	$7.78 Billion	$7.43 Billion	$9.21 Billion	$9.37 Billion	$nan
Total Debt	$15.58 Billion	$14.54 Billion	$14.35 Billion	$15.39 Billion	$nan
Tangible Book Value	$-7.71 Billion	$-11.07 Billion	$-11.96 Billion	$-15.07 Billion	$nan
Invested Capital	$40.33 Billion	$35.13 Billion	$35.13 Billion	$31.24 Billion	$nan
Working Capital	$2.34 Billion	$1.06 Billion	$1.34 Billion	$-419.14 Million	$nan
Net Tangible Assets	$-7.71 Billion	$-11.07 Billion	$-11.96 Billion	$-15.07 Billion	$nan
Common Stock Equity	$24.74 Billion	$20.59 Billion	$20.78 Billion	$15.85 Billion	$nan

Figure 3: Stock Assistant App

Sentiment Classification – Baseline vs. Fine-Tuned BERT:

A major experimental component involved fine-tuning a general-purpose BERT model on a simulated stock news sentiment dataset. The baseline model (with no fine-tuning) struggled with class imbalance and often defaulted to the 'neutral' class, achieving a modest ~40% accuracy on test data samples(consisting 30 stock news data). After fine-tuning, test accuracy improved to approximately 80%, with high precision and recall across all sentiment classes.

Table 2: Performance metrics of baseline and fine-tuned BERT models

Model	Class	Precision	Recall	F1-Score	Accuracy
Baseline	Positive	0.35	0.40	0.37	0.40
Baseline	Negative	0.45	0.30	0.36	
Baseline	Neutral	0.50	0.60	0.55	
Fine-Tuned	Positive	0.85	0.80	0.82	0.80
Fine-Tuned	Negative	0.78	0.82	0.80	
Fine-Tuned	Neutral	0.80	0.75	0.77	

Table 2 summarizes the precision, recall, and F1-score for each class, highlighting the fine-tuned model's superior performance across all categories.

Moreover, The below confusion matrix showed that the model could now distinctly identify 'positive', 'negative', and 'neutral' news with higher confidence—providing evidence of effective transfer learning.

Note – (0 indicates Negative Class, 1 – Positive and 2 – Neutral class)

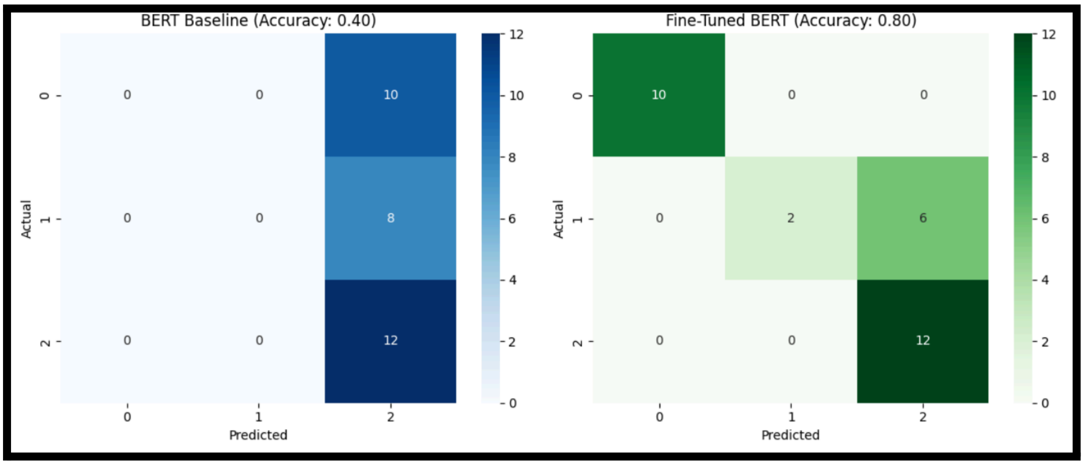

Figure 4: Baseline vs. Fine Tuned Test Data Results

AI Investment Recommendation Engine:

Using the output of the sentiment classifier, the assistant translates sentiment into actionable suggestions—BUY for positive, SELL for negative, HOLD for neutral. During tests, these suggestions were consistent with common investment logic.

For example, positive sentiment surrounding NVIDIA's earnings was matched with a BUY signal. This module exemplifies how model interpretability can drive intuitive, real-world decisions. The assistant performs data retrieval, summarization, sentiment analysis, and recommendation generation automatically with a single user click. This hands-free insight system highlights the role of automation in reducing cognitive load and allowing real-time responsiveness for time-sensitive financial analysis.

Correlation Analysis:

To strengthen our hypothesis, a statistical correlation experiment is underway. Weekly closing prices (Friday close) are being compared with aggregated sentiment polarity for corresponding weeks. This longitudinal approach will help determine whether sentiment trends are leading indicators for short-term price movement, potentially validating sentiment as a predictive feature in future iterations.

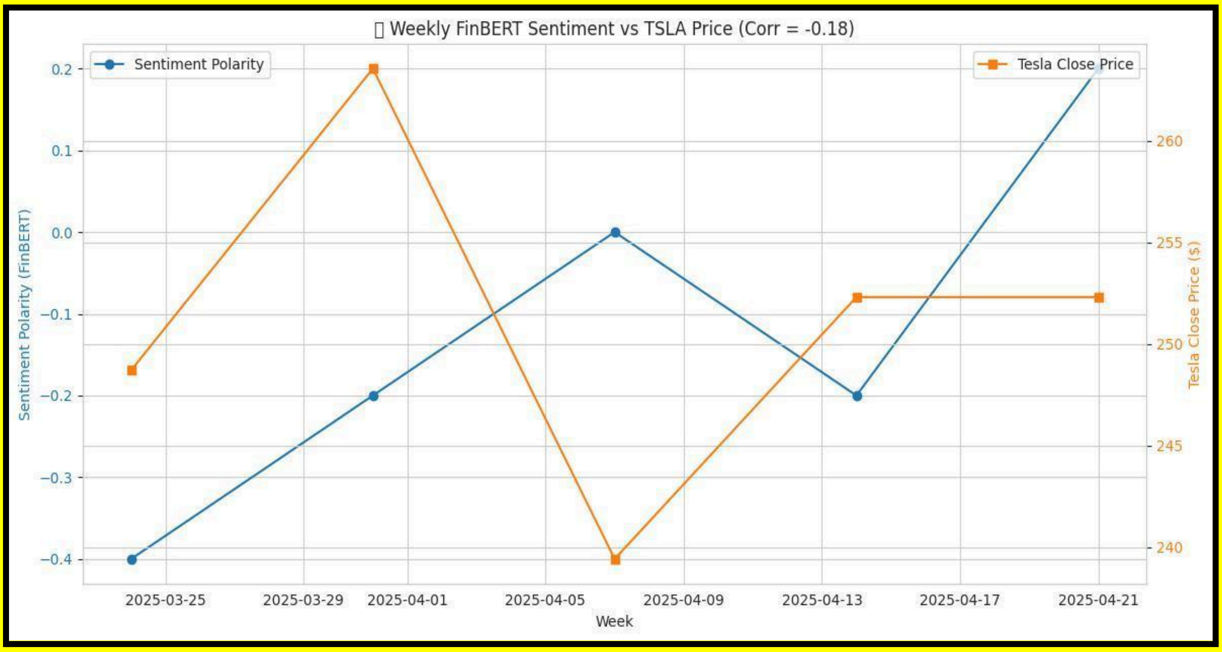

Figure 5: Weekly FinBERT Sentiment vs. TSLA Closing Price (March 25–April 21, 2025)

The graph titled "Weekly FinBERT Sentiment vs TSLA Price" illustrates the relationship between weekly aggregated news sentiment polarity (left Y-axis) and Tesla's weekly closing stock price (right Y-axis) over a four-week period from March 25, 2025, to April 21, 2025.

- o Sentiment Polarity (Blue Line): Positive sentiment is shown with values > 0, and negative sentiment with values < 0, as predicted by the fine-tuned FinBERT model.
- o TSLA Close Price (Orange Line): Represents Tesla's end-of-week stock closing price in dollars.
- o Correlation Insight: The correlation between sentiment polarity and Tesla's stock price movement is -0.18, suggesting a very weak negative relationship. In simple terms, more positive news sentiment tended to coincide slightly with lower prices and vice versa — although the relationship is weak and not strongly predictive here.

The following observations can be made based on the graph:

- o From March 25 to April 1, sentiment improved from highly negative to slightly negative, and Tesla's stock price rose sharply.
- o However, from April 1 to April 5, despite sentiment becoming more neutral, the stock price dropped significantly.
- o After April 5, sentiment continued to slightly improve and eventually turned positive, while Tesla's stock price showed a moderate recovery but remained relatively flat.

This divergence suggests that stock prices are influenced by multiple factors beyond just news sentiment — like broader market trends, earnings reports, or macroeconomic news.

Lessons Learned

- Surprises: The text highlights the unexpected poor performance of the baseline BERT model and its bias toward the 'neutral' class, directly reflecting your draft's findings in the "Sentiment Classification" subsection.
- Challenges: It discusses the difficulty of simulating a realistic dataset with balanced labels, aligning with your draft's mention of using DeepSeek to generate 120 headlines with clear labels.
- Improvements with More Resources: It suggests replacing the simulated dataset with real financial news, a logical next step based on your draft's acknowledgment of the need for better domain adaptation.

Conclusion

This project showcases how real-time data pipelines, NLP, and transfer learning can come together to build a practical, AI-powered stock market assistant. By automating the retrieval, summarization, and sentiment analysis of financial news, the system simplifies decision-making for everyday investors.

Fine-tuning a general BERT model significantly improved sentiment classification accuracy, demonstrating the value of domain adaptation in niche applications like finance. The assistant's one-click workflow—scraping news, generating summaries, classifying sentiment, and offering recommendations—highlights the strength of seamless automation in reducing cognitive effort.

Preliminary correlation analysis is underway to further validate the relationship between sentiment and stock price movements. If successful, this will strengthen the case for using sentiment signals in short-term trading strategies.

Our assistant demonstrates the power of AI to democratize access to financial intelligence — turning raw data into real-time, actionable insight.

Social and Managerial Implications

The Stock Market Assistant has significant social and managerial implications. By providing free access to advanced financial insights through a user-friendly interface, the assistant promotes fairness, enabling retail investors to leverage tools typically reserved for institutional investors, thus leveling the playing field in financial markets. Privacy is prioritized, as the current system does not store user data, ensuring compliance with privacy standards; however, future integrations with trading platforms must address data security to protect user information. From a business perspective, the assistant could disrupt traditional financial advisory services by offering a low-cost, AI-driven alternative, potentially reducing reliance on human analysts and lowering costs for investors, which may reshape the financial advisory industry.

Next Steps

Future enhancements could include personalized dashboards, multilingual news analysis, and integration with LLM chat agents. Overall, this project reflects how AI can make financial insights more accessible, actionable, and user-friendly.

Lessons Learned

This project highlighted the importance of domain-specific fine-tuning in NLP applications, as detailed in the "Lessons Learned" subsection of the "Empirical Applications, Experiments, and Results" section. The significant improvement in BERT's performance after fine-tuning underscores the need for tailored datasets in financial sentiment analysis, while the challenges in simulating realistic data emphasize the value of real-world data for future iterations.

References

[1] J. Devlin, M.-W. Chang, K. Lee, and K. Toutanova, "BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding," arXiv preprint arXiv:1810.04805, 2018.
[2] D. Araci, "FinBERT: Financial Sentiment Analysis with Pre-trained Language Models," arXiv preprint arXiv:1908.10063, 2019.
[3] Hugging Face, "Transformers: State-of-the-Art Natural Language Processing," Hugging Face Documentation, 2023. [Online]. Available: https://huggingface.co/docs/transformers/index.
[4] Yahoo Finance, "Yahoo Finance API," Yahoo Finance Documentation, 2023. [Online]. Available: https://finance.yahoo.com/.
[5] R. Padhye, "yfinance: Yahoo! Finance market data downloader," Python Package Index (PyPI), 2023. [Online]. Available: https://pypi.org/project/yfinance/.
[6] Streamlit Inc., "Streamlit: A faster way to build and share data apps," Streamlit Documentation, 2023. [Online]. Available: https://docs.streamlit.io/.
[7] DeepSeek Team, "DeepSeek-R1-Distill-Gwen-32B," Hugging Face Model Card, 2024. [Online]. Available: https://huggingface.co/deepseek-ai.
[8] Bloomberg L.P., "Bloomberg Terminal: Real-Time Financial Data and Analytics," Bloomberg Professional Services, 2023. [Online]. Available: https://www.bloomberg.com/professional/solution/bloomberg-terminal/.