

FinSentix-Sentiment and Term Extraction In Financial Texts

1.A. NARESH,2.GUNIGANTI PRANA VI,3.BATTULA BHARATH REDDY,4.BUDDARTHI RUCHITHA,
5.DONDADI SRINIVAS

¹Assistant Professor, Department of AIML, SriIndu College Of Engineering & Technology, Hyderabad.
^{2,3,4,5}U.G.Scholar, Department of AIML, SriIndu College Of Engineering & Technology, Hyderabad

ABSTRACT

This project introduces a web-based platform for financial sentiment evaluation and risk assessment that combines Natural Language Processing (NLP), machine learning techniques, and quantitative finance to generate meaningful insights for investors and financial analysts. The system employs the **FinBERT** model to examine real-time financial news articles and social media discussions, categorizing stock sentiment into bullish, bearish, or neutral classes. At the same time, it gathers live stock market information and historical price trends through the **Yahoo Finance API**. A regression-driven risk analysis module calculates key financial indicators, including beta, idiosyncratic risk, market volatility, and the influence of sentiment on price fluctuations. These results are presented through an interactive dashboard that visualizes stock price movements, sentiment trends over the past ten days, and projected market outlooks.

By integrating sentiment analysis from NLP, financial data processing, and statistical risk modeling, the platform allows users to understand both behavioral and quantitative factors that influence stock performance. This combined approach helps investors interpret market sentiment alongside numerical indicators, thereby supporting informed investment strategies, effective risk management, and more intelligent financial decision-making.

Keywords - Financial Sentiment Analysis, Natural Language Processing, Machine Learning, Quantitative Finance, FinBERT Model, Stock Prediction, Yahoo Finance API, Regression Model, Risk Assessment, Beta Coefficient, Idiosyncratic Risk, Volatility Analysis, Sentiment Classification, Price Movement Forecasting, Market Outlook, Real-time Analytics, Financial Data Visualization, Predictive Modeling, Investment Decision-Making, Behavioral Finance.

INTRODUCTION

This project integrates Natural Language Processing and quantitative financial analytics to perform real-time sentiment and risk analysis in the stock market. Using the FinBERT model, it processes financial news headlines and company announcements to classify sentiment as bullish, bearish, or neutral. The system retrieves live stock data via Yahoo Finance (yFinance) and applies regression-based models to calculate beta, idiosyncratic risk, volatility, and sentiment impact on price movement. Developed with FastAPI as the backend and Ruby Sinatra as the frontend, it provides an interactive dashboard for visualizing stock sentiment trends, historical charts, and overall market outlook. By combining machine learning, financial modeling, and statistical analysis, the project offers practitioners and researchers to understand the emotional and quantitative factors influencing stock performance.

OBJECTIVE

Designed and implemented a hybrid system that fuses financial data analytics with advanced natural language processing for holistic market sentiment analysis. Employed the FinBERT model to accurately extract and classify sentiment (bullish, bearish, neutral) from real-world financial news sources. Collected live and historical stock data through the Yahoo Finance API, enabling real-time and retrospective financial analysis. Applied regression models to calculate beta values, providing insight into each stock's sensitivity to overall market movements. Estimated idiosyncratic risk by quantifying residual volatility derived from regression analyses. Developed an interactive web dashboard that visualizes live stock price trends and 10-day sentiment trajectories for user engagement.

Computed and displayed sentiment impact percentages to clearly illustrate correlations between news sentiment and stock price fluctuations. Aggregated stock-level sentiment analyses to infer and display overall market sentiment and direction. Built a user-focused web interface using Ruby Sinatra, seamlessly integrated with a FastAPI backend for efficient functionality. Enabled investors, analysts, and researchers to leverage the platform for data-driven interpretations of how public sentiment and news drive financial market activity.

LITERATURE SURVEY

1. Early text-based sentiment in finance:

The first wave of work treated news text as bags of words and relied on finance-specific lexicons. The Loughran – McDonald dictionary corrected the bias of generic sentiment lexicons (e.g., “liability,” “executed”) that are negative in general English but neutral/technical in finance. With TF-IDF or polarity counts, classical models such as logistic regression, Naïve Bayes, and SVM achieved strong baselines for headline-level and document-level polarity classification. These approaches are transparent and fast, but they struggle with context, negation, and domain nuances (e.g., “beats expectations despite loss”).

2. Deep learning (CNN/LSTM/Attention):

As datasets grew, CNNs captured local n-grams and LSTMs/GRUs modeled longer dependencies (e.g., clause-level sentiment shifts). Attention mechanisms improved interpretability by highlighting words that drive predictions. These models clearly outperformed bag-of-words on Financial Phrase Bank, FiQA Task-1/2, and newswire corpora, but they still required heavy feature engineering (tokenization, embeddings) and lacked transfer learning across tasks (sentiment → volatility/return prediction).

3. Transformers and domain adaptation:

BERT/roBERTa introduced pretraining → fine-tuning, bringing large gains for sentence- and aspect-level finance sentiment. Domain-specific variants—FinBERT, FinBERT-tone, FinBERT-FinVocab, FinGPT-Sentiment, FinRobby—pretrain on financial corpora (SEC filings, earnings call transcripts, news, analyst reports). They consistently outperform generic BERT on FinancialPhraseBank, FiQA (headline and aspect relevance), and Twitter/StockTwits datasets, especially in handling negation, modality, and numeric context (“EPS down 5% but guidance raised”). Mixed-domain pretraining and continual learning reduce drift across market regimes.

4. Risk modeling background (beta & idiosyncratic risk):

Standard asset-pricing models (CAPM, Fama–French) decompose a stock’s return into a market component ($\beta \times$ market return) and a residual component. β captures market sensitivity; the residual variance (standard deviation of regression errors) measures idiosyncratic risk firm-specific shocks not explained by the market. Studies show firm news, tone, and linguistic uncertainty raise residual risk and sometimes forecast drawdowns, especially around earnings.

METHODOLOGY

1. Data Acquisition:

- **Market Data:** Live and historical stock prices are retrieved using the Yahoo Finance (yFinance) API.
- **Textual Data:** Financial news headlines or social media posts are input by the user to capture the sentiment that may influence stock prices.
- **Benchmark:** The S&P 500 Index (or equivalent market index) is used as a reference to calculate comparative returns.

2. Sentiment Analysis using FinBERT:

The FinBERT transformer model, specifically fine-tuned for processing financial text, analyzes each input headline or summary by classifying its sentiment as Positive (Bullish), Negative (Bearish), or Neutral. Along with the sentiment category, it provides a confidence score ranging from 0 to 1, which quantifies the degree of certainty in the classification. This sentiment score reflects how the market is likely to perceive the information, indicating either an optimistic or pessimistic impact on stock behavior.

3. Quantitative Analysis:

Return Calculation :

Daily returns for both the stock and market index are calculated as:

$$r_t = \frac{P_t - P_{t-1}}{P_{t-1}}$$

Regression Analysis:

A linear regression is performed using the model:

$$\text{Stock Return} = \alpha + \beta \times (\text{Market Return}) + \epsilon$$

α (Alpha): Represents the stock’s average return independent of the market.

ϵ (Residual): The difference between actual and predicted returns.

Idiosyncratic Risk Calculation:

The standard deviation of residuals (ϵ) represents idiosyncratic risk, i.e., the company-specific volatility not explained by market trends.

Sentiment Market Correlation:

The system analyzes the recent 10-day sentiment trend to determine how news tone correlates with stock price movements, enabling a clearer understanding of short-term market dynamics. It derives the Sentiment Impact percentage, which quantitatively measures how fluctuations in sentiment contribute to immediate stock price changes. By comparing sentiment and price movements across multiple stocks, the system infers the overall market sentiment, identifying whether the market is predominantly bullish or bearish.

5. Visualization and Web Interface

The backend, built with FastAPI, is responsible for calculating and transmitting sentiment scores, beta values, idiosyncratic risk, and live stock data to the frontend interface. On the frontend, which utilizes Ruby Sinatra alongside a TradingView widget, users can access visualizations including live stock price charts, 10-day sentiment trends, market sentiment summaries, and company-specific financial metrics such as beta, idiosyncratic risk, and sentiment impact. By simply entering a stock ticker or headline, the user can instantly view these real-time analysis results in a unified dashboard, facilitating efficient and informed financial decision-making.

6. Model Improvement & Scalability

The system can be enhanced by fine-tuning the FinBERT model on localized datasets, such as those specific to the Indian stock market, to better capture regional sentiment nuances. Incorporating multiple diverse data sources like Twitter feeds, Bloomberg, or Reuters can enrich sentiment analysis and broaden market coverage. Extending the regression models to multi-factor approaches, including variables such as volatility and sector-specific trends, would improve the robustness of risk and impact assessments. Additionally, implementing rolling-window learning techniques will help the system adapt dynamically to changing market regimes and evolving financial environments, thereby maintaining predictive accuracy over time.

EXISTING SYSTEM

How the Existing System Works

- Conventional financial systems such as Bloomberg or Yahoo Finance primarily depend on numerical data like prices, volume, and indicators (MACD, RSI, etc.) to evaluate market performance.
- Sentiment analysis systems in existing setups usually employ generic NLP models or polarity-based lexicons to classify news or tweets as positive, negative, or neutral.
- These models are often trained on non-financial text (e.g., movie or product reviews), which do not capture the nuance of financial language.
- Risk metrics such as beta and idiosyncratic risk are computed independently by analysts using statistical software like Excel, R, or Python scripts.
- Thus, no single unified dashboard combines sentiment, price data, and quantitative risk analysis in real time.

Drawbacks of the Existing System

- Lack of integration: Sentiment analysis and financial modeling exist as isolated tools, requiring manual correlation between text data and stock data.
- Generic models: Existing sentiment tools rely on non-financial datasets, leading to poor understanding of domain-specific terms (e.g., “beat estimates,” “guidance lowered”).
- Limited automation: Analysts must manually download data, run regressions, and interpret results, which is time-consuming and error-prone.
- Delayed insights: Sentiment trends and market movements are not analyzed in real time, causing missed opportunities in volatile markets.
- No idiosyncratic risk estimation: Most dashboards show only price volatility or beta, ignoring company-specific risk derived from regression residuals.
- Poor visualization: Existing tools fail to combine quantitative metrics and text sentiment into a cohesive, interactive user interface.
- Cost and accessibility: Professional tools (Bloomberg Terminal, Refinitiv, RavenPack) are expensive, limiting accessibility for students and small investors.
- No cross-market comparison: Current systems rarely show overall market mood (e.g., U.S. vs. Indian market sentiment).
- No natural-language summarization: They don’t explain the results in plain English for non-technical users.

PROPOSED SYSTEM

- Hybrid Data Pipeline: The system fuses structured financial indicators (e.g., price trends, volatility, trading volume) with unstructured text data from news, reports, and social media to provide a holistic market view.
- Dynamic Sentiment Scoring: Using FinBERT’s sentiment classification, the model continuously updates sentiment scores that influence predictive models in near real-time.
- Feature Engineering & Correlation Analysis: It performs advanced feature selection to identify key variables linking investor sentiment and stock performance, improving model interpretability and accuracy.
- Predictive Dashboard & Visualization: The platform includes an interactive dashboard that visualizes sentiment trends, price predictions, and risk metrics for better decision support.
- Model Adaptability & Continuous Learning: The system retrains periodically with new data to adapt to shifting market conditions and emerging financial narratives.

SYSTEM REQUIREMENTS

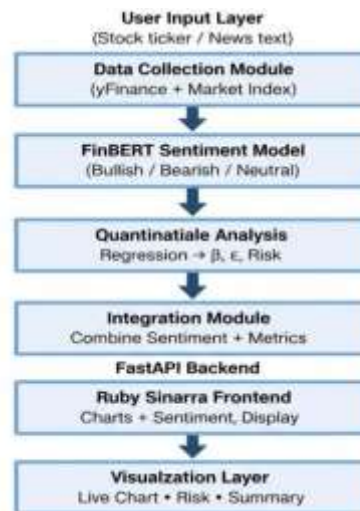
HARDWARE SPECIFICATIONS

- *RAM:* 8–16 GB or higher for efficient model execution and data processing.
- *Processor:* Intel i5/i7 or AMD Ryzen 5/7 series for fast computation and performance.
- *Storage:* Minimum 10 GB free space; 20 GB SSD recommended for faster data handling and caching.
- *GPU:* Optional, but NVIDIA GPU recommended for FinBERT model acceleration and faster inference.
- *Internet:* Stable connection (minimum 5 Mbps; 20 Mbps or higher recommended) for real-time data retrieval and updates.

SOFTWARE SPECIFICATIONS

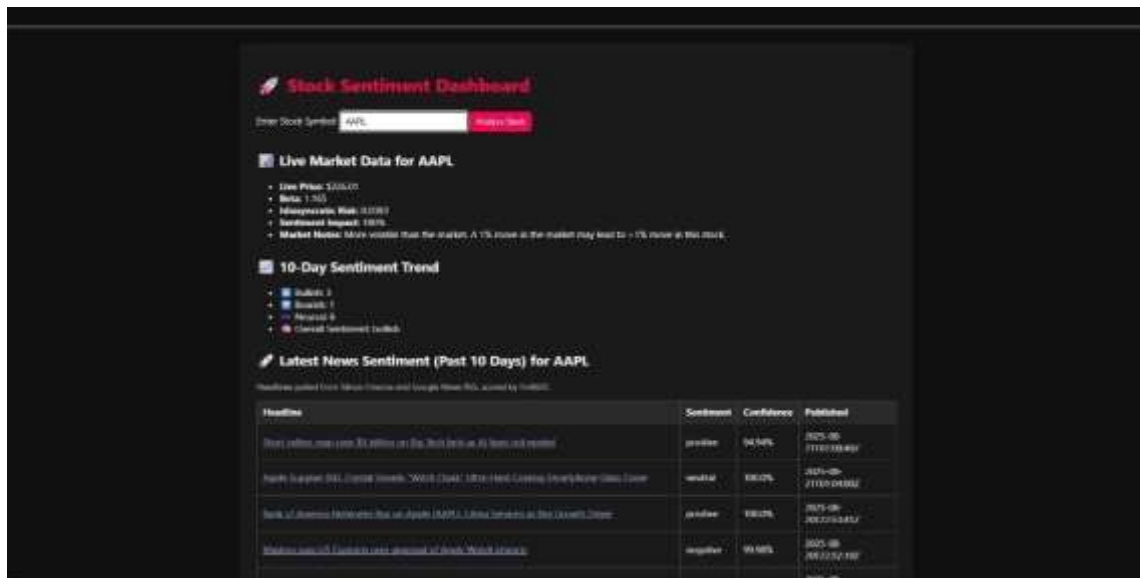
- *Operating System:* Windows 10, Ubuntu 20.04+, or macOS for cross-platform compatibility.
- *Backend Framework:* Python 3.10+ with FastAPI for building high-performance APIs.
- *Frontend Framework:* Ruby 3.0+ with Sinatra for lightweight and efficient web interface development.
- *Machine Learning Libraries:* Transformers (Hugging Face), NumPy, Pandas, and Scikit-learn for model training, data analysis, and prediction tasks.
- *Financial Data Source:* yFinance API for fetching real-time and historical stock market data.
- *Visualization Tools:* TradingView widget for live market charts and Matplotlib/Plotly for analytical data visualization.
- *Web Server:* Uvicorn or Gunicorn for serving FastAPI applications efficiently.
- *Browser Support:* Compatible with latest versions of Chrome, Edge, and Firefox.

WORKFLOW DIAGRAM

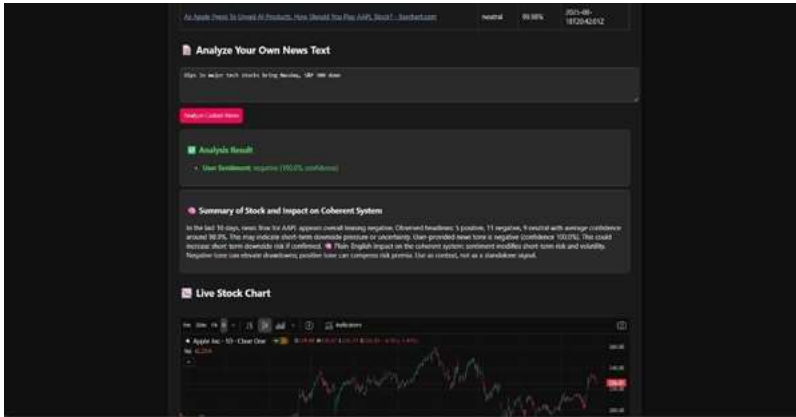


OUTPUT:

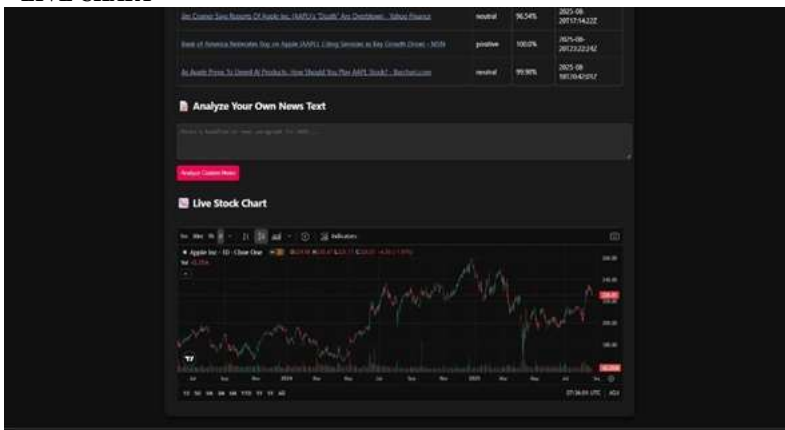
MAIN INTERFACE



STOCK SUMMARY



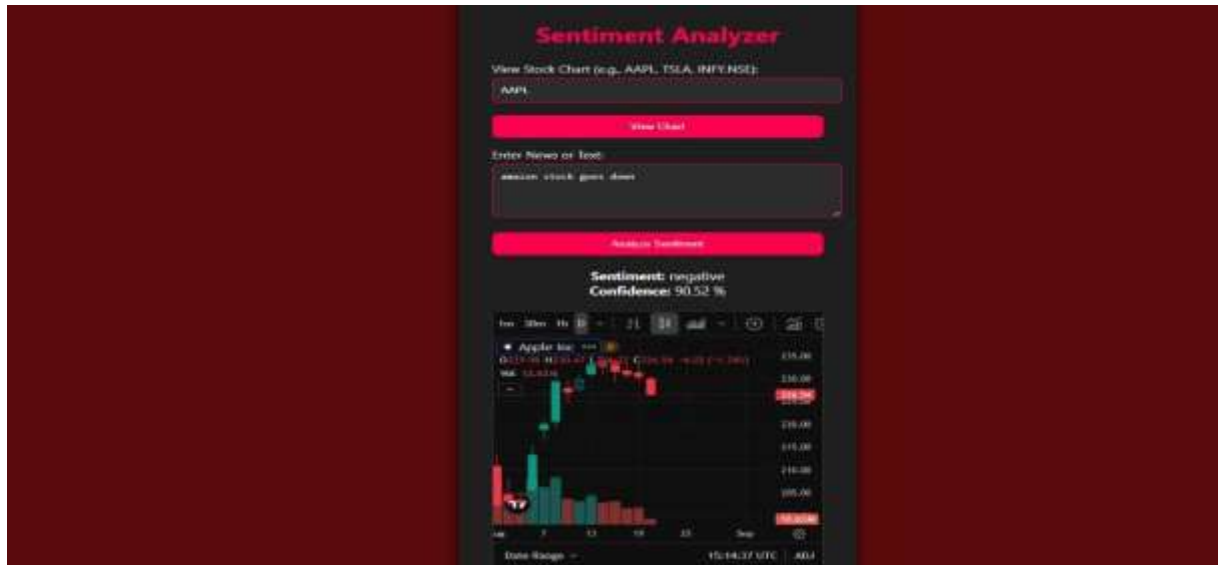
LIVE CHART



PAST 10 DAYS ARTICLE/NEWS

Article/News Title	Sentiment	Confidence	Date
Apple's New iPhone 15 Pro Max... (truncated)	neutral	76.54%	2025-03-20T17:14:22Z
Bank of America Technol... (truncated)	positive	100.0%	2025-03-20T17:22:24Z
Apple's New iPhone 15 Pro Max... (truncated)	neutral	79.98%	2025-03-20T16:42:07Z
Apple's New iPhone 15 Pro Max... (truncated)	neutral	76.54%	2025-03-20T17:14:22Z
Bank of America Technol... (truncated)	positive	100.0%	2025-03-20T17:22:24Z
Apple's New iPhone 15 Pro Max... (truncated)	neutral	79.98%	2025-03-20T16:42:07Z
Apple's New iPhone 15 Pro Max... (truncated)	neutral	76.54%	2025-03-20T17:14:22Z
Bank of America Technol... (truncated)	positive	100.0%	2025-03-20T17:22:24Z
Apple's New iPhone 15 Pro Max... (truncated)	neutral	79.98%	2025-03-20T16:42:07Z
Apple's New iPhone 15 Pro Max... (truncated)	neutral	76.54%	2025-03-20T17:14:22Z
Bank of America Technol... (truncated)	positive	100.0%	2025-03-20T17:22:24Z
Apple's New iPhone 15 Pro Max... (truncated)	neutral	79.98%	2025-03-20T16:42:07Z

MOBILE VIEW



CONCLUSION

This project effectively integrates *AI-driven sentiment analysis* with *quantitative financial modeling*, resulting in a unified system capable of analyzing market trends in real time. By leveraging *FinBERT* for context-aware sentiment extraction from financial texts and combining it with regression-based models for evaluating *beta* and *idiosyncratic risk*, the system bridges the gap between market psychology and numerical finance. This dual-layered approach provides a more comprehensive understanding of stock behavior.

The developed platform features an *interactive dashboard* that delivers key analytical components such as *stock-specific sentiment scores*, *live price and volatility monitoring*, and *market-wide sentiment trend visualization*. This enables investors, analysts, and researchers to observe both emotional and statistical influences on financial markets, facilitating better-informed decision-making and risk assessment.

Overall, this project demonstrates how *Natural Language Processing (NLP)* can enhance traditional financial analysis by transforming qualitative market sentiment into quantitative insights. It serves as a foundation for future advancements, including *multilingual sentiment models*, *deep learning-based volatility forecasting*, and *AI-driven portfolio risk management*, paving the way for more intelligent, adaptive, and emotion-aware financial systems.

REFERENCES

1. Araci, Dogu. "FinBERT: Financial Sentiment Analysis with Pre-trained Language Models." *arXiv preprint arXiv:1908.10063* (2019).
2. Devlin, Jacob, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. "BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding." *Proceedings of NAACL-HLT* (2019): 4171–4186.
3. Nassirtoussi, Alireza Khadjeh, et al. "Text Mining for Market Prediction: A Systematic Review." *Expert Systems with Applications* 41.16 (2014): 7653–7670.
4. Yang, Yu, and Xin Zhang. "Stock Movement Prediction Based on Financial News Using BERT." *IEEE Access* 9 (2021): 146028–146037.
5. Zhang, Qi, et al. "Deep Learning Models for Financial Text Mining: A Survey." *Information Processing & Management* 57.6 (2020): 102279.