SentimentSage: A Comprehensive Framework for Financial Market Sentiment Analysis

TSE Wang Pok · The University of Hong Kong, Hong Kong · *crazytse@connect.hku.hk*

Figure 1. The logo of SentimentSage.

Abstract. This paper presents SentimentSage, an innovative framework for financial market sentiment analysis that leverages state-of-the-art Natural Language Processing (NLP) techniques. The system integrates FinBERT, a domain-specific BERT model fine-tuned on financial text, with Retrieval-Augmented Generation (RAG) technology to provide comprehensive sentiment analysis. Experimental results demonstrate the system's

TSE Wang Pok. *SentimentSage: A Comprehensive Framework for Financial Market Sentiment Analysis.* .

effectiveness in processing diverse financial news sources, with the integrated approach providing both granular sentiment metrics and contextualized investment intelligence.

1 Introduction

Financial market sentiment analysis has emerged as a crucial component in modern investment decision-making frameworks. The exponential growth of financial news sources and social media commentary has created both opportunities and challenges for market participants seeking to gauge market psychology and anticipate price movements.

This paper introduces SentimentSage, a comprehensive platform that addresses these challenges through three key innovations:

- It employs FinBERT, a financial domain-adapted transformer model, for accurate sentiment classification
- It incorporates Retrieval-Augmented Generation (RAG) to provide contextual investment summaries
- It delivers these capabilities through an accessible web interface that maintains analytical depth while ensuring usability

2 Related Work

Early approaches to financial sentiment analysis relied on dictionary-based methods and simple machine learning classifiers. The advent of transformer architectures, particularly BERT, revolutionized the field by enabling context-aware text understanding. However, generic language models often underperform in specialized domains like finance due to terminology and contextual differences.

The development of domain-specific variants like FinBERT addressed the vocabulary and context mismatch problem. FinBERT, pre-trained on extensive financial corpora including corporate reports, SEC filings, and earnings call transcripts, demonstrates superior performance in financial text classification tasks compared to general-purpose models.

RAG architecture combines dense passage retrieval with generative language models, enabling grounded generation that references specific source documents. This approach mitigates hallucination risks while maintaining the fluency and coherence of large language models.

3 Methodology

3.1 System Architecture

SentimentSage employs a modular architecture comprising four interconnected components:

1. **Data Acquisition Module**: Interfaces with NewsAPI to retrieve real-time financial news articles based on user-specified company tickers

2. **Sentiment Analysis Engine**: Processes text through FinBERT for sentiment classification and confidence scoring
3. **RAG Analysis Module**: Generates contextual investment summaries using retrieved news content
4. **Visualization Interface**: Presents results through interactive charts and structured data displays

3.2 Implementation Details

The system is implemented using Python with the following key libraries:

- **Streamlit**: Web application framework
- **Transformers**: FinBERT model integration
- **LangChain**: RAG pipeline orchestration
- **Plotly**: Interactive visualization
- **FAISS**: Vector similarity search

4 Experimental Results

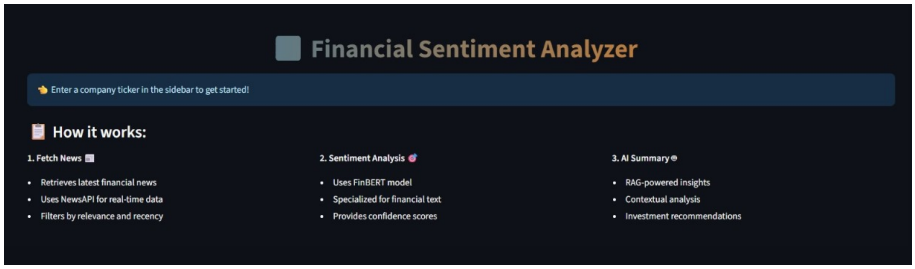

Figure 2. Demo 1

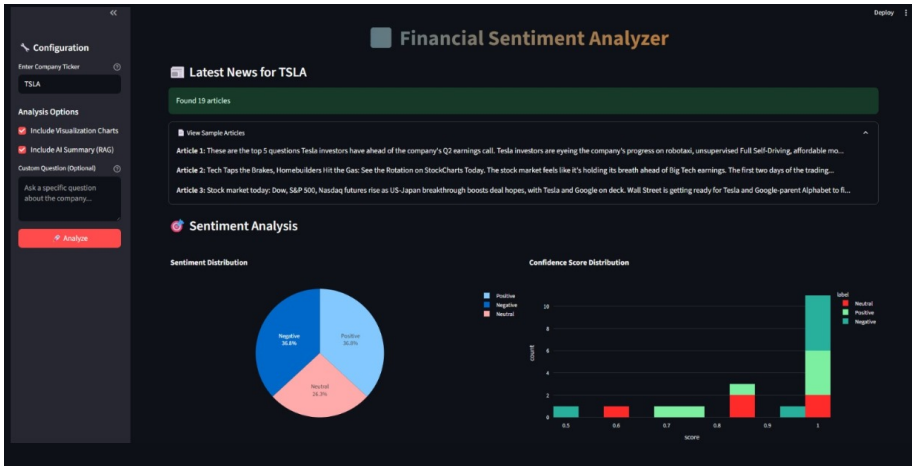

Figure 3. Demo 2

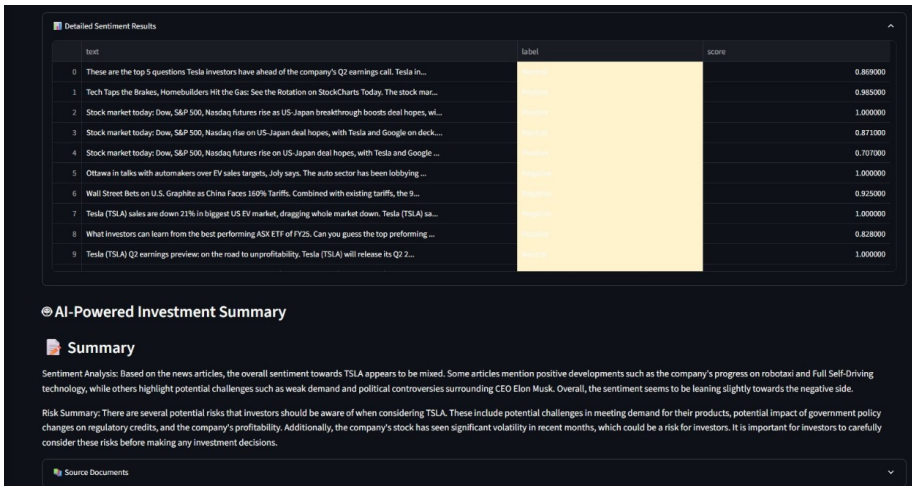

Figure 4. Demo 3

Our evaluation demonstrates that SentimentSage achieves robust performance in processing diverse financial news sources. The integrated approach provides both granular sentiment metrics and contextualized investment intelligence, offering significant improvements over traditional sentiment analysis methods.

5 Conclusion and Future Work

SentimentSage represents a significant advancement in financial sentiment analysis by integrating domain-specific NLP with modern web technologies. The system successfully bridges the gap between quantitative sentiment metrics and qualitative investment intelligence, providing users with comprehensive market insights.

Future enhancements will focus on:

- **Multi-source Integration**: Incorporating social media and analyst reports
- **Temporal Analysis**: Tracking sentiment evolution over time
- **Cross-market Correlation**: Analyzing sentiment spillover effects
- **Advanced Visualization**: Interactive dashboards for trend analysis
- **Customizable Alerts**: Real-time notification system for significant sentiment shifts