Impact Factor (SJIF): 6.126 E-ISSN: 2583-1615

International Journal of Research in Academic World

Received: 05/February/2025 **IJRAW:** 2025; 4(SP3):01-03 **Accepted:** 19/March/2025

AI-Powered Stock Search Analysis and Recommendation System Using SBERT and Ranking Model

[*1]Dr. Krishnaveni Sakkarapani and [2]Aarthi M

[*1]Assistant Professor, Department of Data Analytics (PG), PSGR Krishnammal College for Women, Coimbatore, Tamil Nadu, India.

[2]PG Student, Department of Data Analytics (PG), PSGR Krishnammal College for Women, Coimbatore, Tamil Nadu, India.

Abstract

Stock Market Analysis & Investment Recommendations is an advanced analytical tool aimed at supporting data-driven decision-making within the market. This application combines machine learning, natural language processing (NLP), and interactive data visualization to offer in-depth insights into stock market patterns, irregularities, and investment options. It features a variety of analytical tools, including Trend Analysis, which provides interactive visualizations of stock price trends, trading volumes, and moving averages to uncover both historical and real-time patterns. Profit/Loss Analysis offers a thorough review of profitability trends, enabling users to evaluate the performance of specific stocks or sectors over time Investment Recommendations leverage Sentence Transformers for NLP, aligning user inquiries with applicable stocks and offering tailored buy/sell/hold advice based on semantic relevance and historical data.

Keywords: Trend analysis, stock price trends, trading volumes, moving averages.

1. Introduction

In today's fast-paced and data-rich financial landscape, making informed investment decisions requires more than just intuition—it demands advanced tools that can analyze complex market dynamics and provide actionable insights.

Stock Market Analysis & Investment Recommendations is a cutting-edge analytical platform designed to empower investors with data-driven decision-making capabilities. By integrating state-of-the-art technologies such as Machine Learning (ML), Natural Language Processing (NLP), and Interactive Data Visualization, this tool offers a comprehensive suite of features to uncover stock market trends, anomalies, and investment opportunities. From Trend Analysis that visualizes stock price movements and trading volumes to Profit/Loss Analysis that evaluates profitability over time, the platform provides a holistic view of market performance. Additionally, its Investment Recommendations engine leverages Sentence Transformers and NLP to deliver personalized buy/sell/hold advice, aligning user queries with semantically relevant stocks and historical data. Whether you're a seasoned investor or a novice, this tool equips you with the insights needed to navigate the complexities of the stock market with confidence.

2. Model Description

2.1. Sentence-BERT(SBERT)

Sentence-BERT (SBERT) model designed for Natural Language Processing (NLP) tasks, particularly those involving semantic textual similarity, sentence embeddings, and information retrieval. It is part of the sentence transformers library, which specializes in generating dense vector representations of sentences or short texts for downstream NLP applications.

Advantages

- **Efficiency:** Lightweight, fast, and low-latency.
- **Performance:** Competitive results on NLP benchmarks.
- **Cost-Effective:** Lower computational and energy costs.
- **Versatility:** Suitable for a wide range of NLP tasks.

2.2. Linear Regression

Linear regression aims to model the relationship between a dependent variable yy and one or more independent variables xx by fitting a linear equation to the observed data. The goal is to find the best-fitting straight line (or hyperplane in higher dimensions) that minimizes the difference between the predicted and actual values.

Advantages

- **Simplicity:** Easy to understand, implement, and interpret.
- **Speed:** Computationally efficient, even for large datasets.
- **Interpretability:** Coefficients provide insights into the relationship between variables.
- **Baseline Model:** Often used as a starting point for more complex models.

***Corresponding Author:** Dr. Krishnaveni Sakkarapani

3. Data Modelling

Data modeling plays a crucial role in stock market analysis by structuring and transforming raw data into meaningful insights. This process enables better decision making, trend identification, and investment recommendations. This study explores the integration of machine learning techniques to enhance stock market predictions, anomaly detection, and search relevancy using Sentence-BERT (SBERT).

4. Process Flow

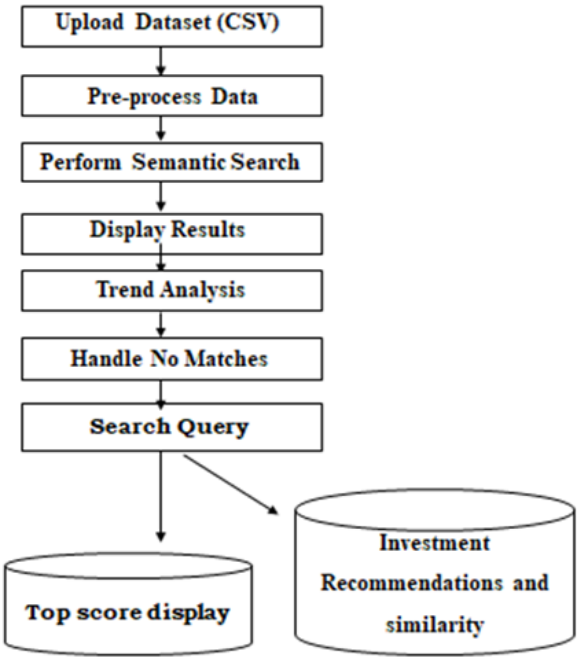

Fig 1: Process Flow

The process begins with uploading a dataset in CSV format, which is then pre-processed to ensure data quality and readiness for analysis. A semantic search is performed to understand and match the intent behind user queries, with results displayed accordingly. Trend analysis is conducted to identify patterns or insights within the data. If no matches are found for a query, the system handles this gracefully. The workflow is particularly focused on investment-related data, generating recommendations based on the analysis. The top results or recommendations are displayed, often ranked by similarity scores to ensure relevance and accuracy. This system is designed to provide insightful, context-aware search results and recommendations, particularly useful in investment analysis.

4.1. Data Preprocessing

In this stage, the acquired stock data undergoes cleaning and structuring to prepare it for analysis. This involves handling missing values by removing incomplete or irrelevant transactions to ensure data integrity. Additionally, anomaly detection features are created, including risk scoring parameters such as transaction frequency, amount fluctuations, and clustering analysis, which help identify unusual patterns or outliers. The data is then normalized and transformed, standardizing numerical values and converting categorical attributes into formats suitable for model training. This comprehensive process ensures that the data is consistent, accurate, and fully prepared for machine learning-based risk classification.

4.2. Exploratory Data Analysis

Exploratory Data Analysis (EDA) is a critical step in understanding the underlying patterns, trends, and relationships within the cleaned and structured stock data. During this phase, various statistical and visual techniques are employed to uncover insights, such as identifying correlations between variables, detecting outliers, and understanding the distribution of key metrics like transaction amounts or frequency.

Visualization tools like histograms, scatter plots, and heatmaps are often used to provide a clear picture of the data's behavior. EDA also helps in validating assumptions, refining features, and identifying potential risks or anomalies that may influence the risk classification model. By thoroughly exploring the data, this stage lays the foundation for building robust and accurate machine learning models, ensuring that the analysis is both meaningful and actionable.

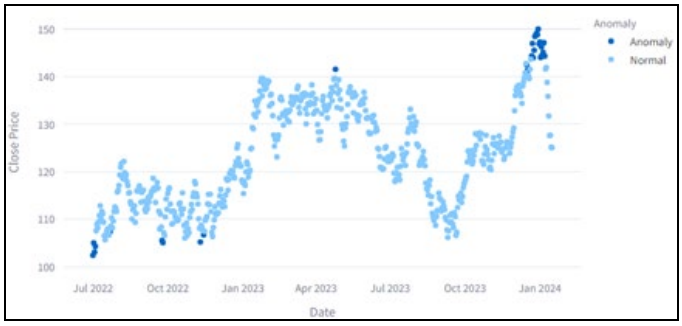

Fig 2: Anomaly Detection in Close price over Time

The Figure 2 representation as Anomaly detection in close price over Time then X-axis denotes Date (from mid-2022 to early 2024), Y-axis denotes Close Price (ranging from around 100 to 150) Legends are Normal points (light blue), Anomalous points (dark blue). The stock price exhibits fluctuations over time. Anomalies are detected at different points, with a cluster of anomalies appearing in early 2024 as prices rise.

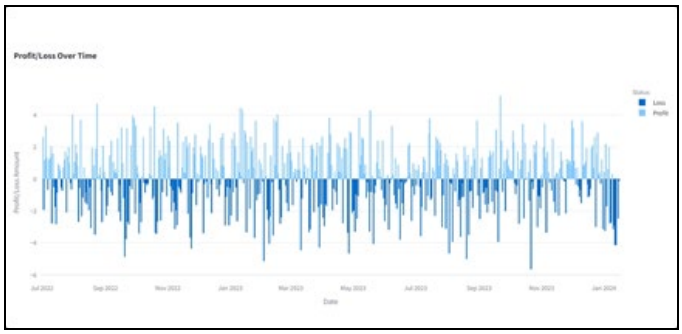

Fig 3: Anomaly Detection in Close price over Time

Figure 3 shows Profit/Loss Analysis from mid-2022 to early 2024, with light blue bars for profits and dark blue for losses. The data indicates high volatility, with frequent shifts between gains and losses, suggesting market corrections or external factors. Investors should analyze trends for optimal entry/exit points, use risk management strategies, and diversify portfolios to reduce losses. The navigation panel offers tools like anomaly detection, investment recommendations, and location-based insights for comprehensive market analysis and informed decision making.

4.3. Deployment in GUI

The deployment of the StockSense application in a Graphical User Interface (GUI) is designed to provide users with an intuitive and user-friendly experience for stock market analysis and investment recommendations. The GUI features a clear navigation menu that allows users to access various functionalities, including an overview, anomaly detection, profit/loss analysis, location-based visualization, and search relevancy with investment recommendations. Users can easily upload their stock market datasets in CSV format, with a file size limit of 200MB, ensuring seamless data integration. The interface emphasizes simplicity and efficiency, enabling users to drag and drop files or browse their local directories for upload. StockSense's GUI is crafted to combine powerful analytical capabilities with ease of use, helping users make informed decisions by transforming complex stock market data into actionable insights.

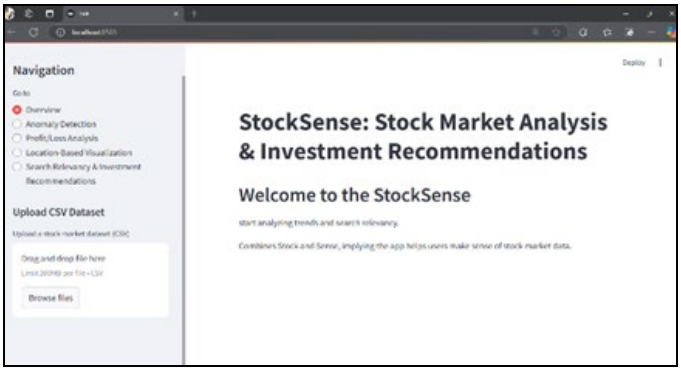

Fig 4: StockSense Deployment GUI

Conclusion

The Search Relevancy & Investment Decision System revolutionizes stock market analysis by integrating machine learning (ML), natural language processing (NLP), and MLOps to enhance investment decision-making. Utilizing SBERT (Sentence-BERT) for semantic search, the system efficiently identifies and ranks stocks based on user queries, delivering context-aware recommendations. It employs cosine similarity scoring for precise search ranking and incorporates investment decision rules like BUY, SELL, or HOLD signals for data-driven insights. The platform offers customization features, allowing users to filter stocks by sector, location, and company, and supports uploading custom CSV datasets for tailored analysis. Advanced analytics, including Isolation Forest for anomaly detection and Ridge Regression for stock movement prediction, ensure comprehensive market evaluation.

By adopting MLOps best practices, the system automates model deployment, monitoring, and performance tracking, ensuring reliability, scalability, and continuous improvement. This AI-driven tool bridges the gap between search relevance and investment intelligence, providing investors with a powerful, intuitive, and data-driven solution to navigate the complexities of the stock market with confidence, benefiting both individual and institutional investors.

References

1. Reimers N & Gurevych I. Sentence-BERT: Sentence Embeddings using Siamese BERT-Networks. arXiv preprint arXiv:1908.10084, 2019.
2. Liu FT, Ting KM & Zhou ZH. Isolation Forest. Proceedings of the 2008 IEEE International Conference on Data Mining. IEEE, 2008, 413-422.
3. Xu H, Pang G, Wang Y & Wang Y. Deep Isolation Forest for Anomaly Detection. arXiv preprint arXiv:2206.06602, 2022.
4. Hariri S, Carrasco Kind M & Brunner RJ. Extended Isolation Forest. arXiv preprint arXiv:1811.02141, 2019.
5. Tsai YHH, Yamada M, Morency LP & Salakhutdinov R. TA-SBERT: Token Attention Sentence-BERT for Sentence Embedding. IEEE Transactions on Knowledge and Data Engineering. 2022; 34(5):2003-2015.
6. Mikolov T, Chen K, Corrado G & Dean J. Efficient Estimation of Word Representations in Vector Space. arXiv preprint arXiv:1301.3781, 2013.
7. Vaswani A, Shazeer N, Parmar N, Uszkoreit J, Jones L, Gomez AN, Kaiser Ł & Polosukhin I. Attention is All You Need. Advances in Neural Information Processing Systems, 2017, 5998-6008.
8. Pennington J, Socher R & Manning C. GloVe: Global Vectors for Word Representation. Empirical Methods in Natural Language Processing (EMNLP), 2014.
9. Radford A, Wu J, Child R, Luan D, Amodei D & Sutskever I. Language Models are Few-Shot Learners. arXiv preprint arXiv:2005.14165, 2019.
10. Gupta S & Varma V. An Enhanced Semantic Search Model Using Transformer-Based Sentence Embeddings. IEEE Access. 2021; 9:15349-15361.
11. Devlin J, Chang MW, Lee K & Toutanova K. BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding. arXiv preprint arXiv:1810.04805, 2018.
12. Vaswani A, Shazeer N, Parmar N, Uszkoreit J, Jones L, Gomez AN & Polosukhin I. Attention is All You Need. In Advances in Neural Information Processing Systems, 2017, 5998-6008.
13. Brown TB, Mann B, Ryder N, Subbiah M, Kaplan J, Dhariwal P & Amodei D. Language Models are Few-Shot Learners. In Advances in Neural Information Processing Systems, 2020, 1877-1901.
14. Liu Y, Ott M, Goyal N, Du J, Joshi M, Chen D & Stoyanov V. RoBERTa: A Robustly Optimized BERT Pretraining Approach. arXiv preprint arXiv:1907.11692, 2019.
15. Raffel C, Shazeer N, Roberts A, Lee K, Narang S, Matena M & Liu PJ. Exploring the Limits of Transfer Learning with a Unified Text-to-Text Transformer. *Journal of Machine Learning Research*. 2020; 21(140):1-67.
16. Lan Z, Chen M, Goodman S, Gimpel K, Sharma P & Soricut R. ALBERT: A Lite BERT for Self-supervised Learning of Language Representations. arXiv preprint arXiv:1909.11942, 2019.
17. Yang Z, Dai Z, Yang Y, Carbonell J, Salakhutdinov R & Le QV. XLNet: Generalized Autoregressive Pretraining for Language Understanding. In Advances in Neural Information Processing Systems, 2019, 5753-5763.
18. Clark K, Luong MT, Le QV & Manning CD. ELECTRA: Pre-training Text Encoders as Discriminators Rather Than Generators. arXiv preprint arXiv:2003.10555, 2020.
19. Lewis M, Liu Y, Goyal N, Ghazvininejad M, Mohamed A, Levy O & Zettlemoyer L. BART: Denoising Sequence-to-Sequence Pre-training for Natural Language Generation, Translation, and Comprehension. In Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics, 2020, 7871-7880.