

Alapati Devi Anusha¹, Battepati Dharani², Punuru Divya Harshitha³, Vadlamudi Lakshmi Tejaswini⁴, Dr.P.Josephin Shermila⁵

line 2: ARTIFICIAL INTELLIGENCE AND DATA SCIENCE City, Country
alap20ad004@rmkcet.ac.in, batt20ad007@rmkcet.ac.in, punu20ad038@rmkcet.ac.in,
vadl20ad053@rmkcet.ac.in, josephinshermilaads@rmkcet.ac.in

Abstract: This paper presents the development of an innovative Equity Research News Tool leveraging Langchain, OpenAI, and FAISS. Langchain facilitates natural language processing tasks, enabling efficient extraction and summarization of relevant news articles. OpenAI enhances language understanding and generation capabilities, while FAISS optimizes similarity search for efficient data retrieval. Integrating these technologies empowers analysts to streamline equity research processes, extract actionable insights, and make informed investment decisions in real-time.

Keywords — *Equity Research, Langchain, OpenAPI, FIASIS, News Tool, Financial analysis.*

I. INTRODUCTION

This chapter gives introduction to the problem, also includes the information about the existing system and their disadvantages proposed system and its advantages along with future enhancement. The Equity Research News Tool is a comprehensive language model (LLM) project designed for equity research analysts. Leveraging the power of langchain, the OpenAI API, and FAISS, this tool allows analysts to conduct efficient and insightful news

research in the ever-evolving financial landscape. It uses fancy language models to understand recent news about a company's stocks and market trends. Just share a few recent links, and the tool will give you insights without the jargon. Ask questions about profits or losses, and get easy-to-understand answers. Our goal is to simplify investing by using cool technology from OpenAI to navigate the complex world of finance.

In the fast-paced world of equity research, staying abreast of the latest news and developments is paramount for informed decision-making. The sheer volume of available information poses significant challenges for analysts. We present an innovative solution leveraging cutting-edge technologies: Langchain, OpenAI, and FAISS. By combining the power of natural language processing, advanced language understanding, and efficient data retrieval, our Tool revolutionizes the way analysts gather, analyze, and extract insights from news articles related to equities.

II. PROBLEM STATEMENT

Traditional equity analysis struggles with manually gathering and analysing vast amounts of news data, leading to

inefficiency missing crucial information, this tool aims to automate news collection, utilize AI for sentiment analysis and deliver actionable insights for decision.

III. EXISTING SYSTEM

The current landscape of equity analysis tools features several models that share a common approach—requiring users to manually input text for analysis. In this paradigm, users copy and paste relevant content into the model, which then performs an in-depth analysis and provides actionable results. However, it's important to note that these models often come with constraints, including text limits or word count restrictions, which can impact the breadth and depth of the analysis.

IV. LITERATURE SURVEY

1. Artificial Intelligence Applied to Stock Market Trading

Year: 2021

Abstract:

The application of Artificial Intelligence (AI) to financial investment is a research area that has attracted extensive research attention since the 1990s, when there was an accelerated technological development and popularization of the personal computer. Since then, countless approaches have been proposed to deal with the problem of price prediction in the stock market. This paper presents a systematic review of the literature on Artificial Intelligence applied to investments in the stock market based on a sample of 2326 papers from the Scopus website between 1995 and 2019. These papers were divided into four categories:

portfolio optimization, stock market prediction using AI, financial sentiment analysis, and combinations involving two or more approaches. For each category, the initial introductory research to its state-of-the-art applications are described. In addition, an overview of the review leads to the conclusion that this research area is gaining continuous attention and the literature is becoming increasingly specific and thorough.

Authors: iyyapam, afraj alom.

[1] The authors created this paper that The time series forecasting system can be used for investments in a safe environment with minimized chances of loss. The Holt–Winters algorithm followed various procedures and observed the multiple factors applied to the neural network. The final module helps filter the system to predict the various factors and provides a rating for the system. This research work uses real-time dataset of fifteen stocks as input into the system and, based on the data, predicts or forecasts future stock prices of different companies belonging to different sectors. The dataset includes approximately fifteen companies from different sectors and forecasts their results based on which the user can decide whether to invest in the particular company or not; the forecasting will give an accurate result for the customer investments.

2. Forecasting Stock Market Prices Using Machine Learning and Deep Learning Models: A Systematic Review, Performance Analysis and Discussion of Implications

Authors: Deepak Sudhakar AnupKumar

[1] The authors impart the information that The financial sector has greatly impacted the monetary well-being of consumers, traders, and financial institutions. In the current era, artificial intelligence is redefining the limits of the financial markets based on state-of-the-art machine learning and deep learning algorithms. There is extensive use of these techniques in financial instrument price prediction, market trend analysis, establishing investment opportunities, portfolio optimization, etc. Investors and traders are using machine learning and deep learning models for forecasting financial instrument movements. With the widespread adoption of AI in finance, it is imperative to summarize the recent machine learning and deep learning models, which motivated us to present this comprehensive review of the practical applications of machine learning in the financial industry. This article examines algorithms such as supervised and unsupervised machine learning algorithms, ensemble algorithms, time series analysis algorithms, and deep learning algorithms for stock price prediction and solving classification problems. The contributions of this review article are as follows: (a) it provides a

description of machine learning and deep learning models used in the financial sector.

[2]it provides a generic framework for stock price prediction and classification

[3] it implements an ensemble model—“Random Forest + XG-Boost + LSTM”—for forecasting TAINIWALCHM and AGROPHOS stock prices and performs a comparative analysis with popular machine learning and deep learning models.

3.A Novel AI-Based Stock Market Prediction Using Machine Learning Algorithm.

Authors: Sulthan Ahmed, Afrom Alam

The authors introduce a novel tourist destination recommendation system based on visual matching and minimal user input, Stock markets and relevant entities generate enormous amounts of data on a daily basis and are accessible from various channels such as stock exchange, economic reviews, and employer monetary reports. In recent times, machine learning techniques have proven to be very helpful in making better trading decisions. Machine learning algorithms use complex logic to observe and learn the behavior of stocks using historical data which can be

used to predict future movements of the stock. Technical indicators such as rolling mean, momentum, and exponential moving average are calculated to convert the data into meaningful information. Furthermore, this information can be used to build machine learning prediction models that learn different patterns in the data and make future predictions for accurate financial forecasting. Additional factors that are being used for stock prediction include social media influences and daily news on trading stocks. Considering these qualitative and quantitative features at the same time result in improved prediction models.

V. ALGORITHMS

We compare two types of models in FinLMEval: the Transformers encoder-only models that require fine-tuning on the labeled dataset, and decoder-only models that are prompted with zero-shot or few shot in-context instructions. Figure 1 provides an outline of evaluation methods of FinLMEval.

Encoder-only Models:

Our experiments explore the performance of various notable encoder-only models: BERT (Devlin et al., 2019), RoBERTa (Liu et al., 2019), FinBERT (Yang et al., 2020) and FLANG (Shah et al., 2022). BERT and RoBERTa are pre-trained on

general domain corpora, while FinBERT and FLANG are pre-trained on a substantial financial domain corpus. We fine-tune the language models on specific tasks. Following the fine-tuning process, inference can be performed on the fine-tuned models for specific applications.

Decoder-only Models:

We also evaluate the performance of various popular decoder-only language models: ChatGPT (Ouyang et al., 2022), GPT-4 (OpenAI, 2023), PIXIU (Xie et al., 2023), LLAMA2-7B (Touvron et al., 2023b) and Bloomberg-GPT (Wu et al., 2023). ChatGPT and GPT-4, developed by OpenAI, are two advanced LLMs that showcase exceptional language understanding and generation abilities. The models are pre-trained on a wide array of textual data and reinforced by human feed back. PIXIU is a financial LLM based on finetuning LLAMA (Touvron et al., 2023a) with in struction data. LLAMA2 is a popular open-sourced LLM pre-trained on extensive online data, and BloombergGPT is an LLM for finance trained on a wide range of financial data. As the model size of the evaluated decoder-only models is extremely large, they usually do not require fine-tuning the whole model on downstream tasks. Instead, the decoder-only models provide answers via zero-shot and few-shot in-context prompting. We conduct zero-shot prompting for all decoder only models. We manually write the prompts for every task. An example of prompts for the sentiment classification task is provided in Figure 1,

and the manual prompts for other tasks are provided in Appendix A. Furthermore, to evaluate whether few-shot in-context learning can improve the model performance, we also conduct in-context learning experiments on ChatGPT. We use two strategies to select the in-context examples for few-shot in context learning: random and similar. The former strategy refers to random selection, and the latter selects the most similar sentence regarding the query sentence. All in-context examples are selected from the training set, and one example is provided from each label class.

Streamlit: Chosen for building an interactive and intuitive user interface for the Equity Research News Tool. Streamlit provides a clean and intuitive interface, making it easy for analysts to interact with the tool. Streamlit's user-friendly syntax allows for rapid prototyping without extensive web development knowledge. It supports quick deployment and is compatible with various data science libraries. The platform's popularity stems from its ability to streamline the process of sharing and presenting data-driven insights through accessible, customizable web interfaces.

Open API:

The Integrated to generate coherent and contextually relevant text based on user queries. OpenAI API is employed to generate human-like summaries or responses, enhancing the tool's capabilities

in providing concise and informative outputs. OpenAI is known for its expertise in developing generative models, particularly Generative Pre-trained Transformers (GPT). These models excel at tasks like text generation, translation, and code completion by learning from massive datasets of text and code. OpenAI has explored applications of AI in robotics, investigating how intelligent systems can be integrated into robotic platforms to enhance their capabilities and adaptability

FAISS: Employed for efficient similarity search and clustering of news articles based on precomputed embeddings. FAISS is integrated to perform efficient similarity search on precomputed news embeddings, allowing analysts to find articles related to their queries based on content similarity. It's widely used in applications such as image and text retrieval, where fast and effective similarity search is crucial. Faiss supports both CPU and GPU implementations for enhanced performance.

Faiss is widely employed for information retrieval tasks, allowing quick and accurate similarity searches in large datasets, such as text, images, or multimedia content.

V.METHODOLOGY USED

Dataset Extraction:

Utilize web scraping techniques, particularly the Scrapy Python framework, to extract data from online sources

containing information about financial sectors

Langchain: A versatile natural language processing library used for tasks such as entity recognition, sentiment analysis, and more. Langchain is utilized for various NLP tasks, such as entity recognition, sentiment analysis, and more, providing valuable insights into the content of news articles. Langchain enables applications to leverage various sources of context, such as user prompts, historical data, and external information. This allows LLMs to generate more relevant and accurate outputs.

FAISS: Employed for efficient similarity search and clustering of news articles based on precomputed embeddings. FAISS is integrated to perform efficient similarity search on precomputed news embeddings, allowing analysts to find articles related to their queries based on content similarity. It's widely used in applications such as image and text retrieval, where fast and effective similarity search is crucial. Faiss supports both CPU and GPU implementations for enhanced performance.

Faiss is widely employed for information retrieval tasks, allowing quick and accurate similarity searches in large datasets, such as text, images, or multimedia content.

VI.RESULT AND ANALYSIS

The News research tool for entity analysis using langchain provides user-friendly web application that enables equity research analysts to gather relevant news articles based on their queries, perform natural language processing (NLP) tasks using langchain, generate human-like text with OpenAI, and utilize FAISS for efficient similarity search. In ChatGPT, the zero-shot and few-shot performances are comparable in most cases. When zero-shot prompting is ineffective, adding demonstrations can improve ChatGPT's performance by clarifying the task, as the results of ESG and Headlines-PI tasks show. Demonstrations are ineffective for easy and well defined tasks, such as sentiment classifications and Headlines (PDU, PDC, PDD, AC, and FI), as the zero-shot prompts clearly instruct ChatGPT

VII.CONCLUSION

In conclusion, The Equity Research News Tool is a powerful asset for equity research analysts, offering a seamless blend of natural language processing, text generation, and similarity search. By combining langchain, the OpenAI API, and FAISS, this tool empowers analysts to stay informed and make data-driven decisions in the dynamic world of financial research. As the financial landscape evolves, so too does the Equity Research News Tool, promising continuous improvement and adaptability.

VIII.FUTURE WORK

Future enhancements to the News research tool for entity analysis could involve Real-Time Data Integration-Incorporate real-time data feeds for up-to-the-minute news analysis. User Preferences-Allow users to customize and prioritize specific entities or topics in their search results. Interactive Visualizations-Implement visualizations to enhance the understanding of news data trends and relationships .enhance the system's ability to capture nuanced preferences. Finally, conducting extensive user studies and feedback sessions will be crucial for continuously refining and optimizing the system based on user needs and preferences.

IX.REFERENCE

- [1] Julio Cesar Salinas Alvarado, Karin Verspoor, and Timothy Baldwin. 2015. [Domain adaption of named entity recognition to support credit risk assessment](#). In Proceedings of the Australasian Language Technology Association Workshop, ALTA 2015, Parramatta, Australia, December 8 - 9, 2015, pages 84–90.
- [2] Tom B. Brown, Benjamin Mann, Nick Ryder, Melanie Subbiah, Jared Kaplan, Prafulla Dhariwal, Arvind Neelakantan, Pranav Shyam, Girish Sastry, Amanda Askell, Sandhini Agarwal, Ariel Herbert-Voss, Gretchen Krueger, Tom Henighan, Rewon Child, Aditya Ramesh, Daniel M. Ziegler, Jeffrey Wu, Clemens Winter, Christopher Hesse, Mark Chen, Eric Sigler, Mateusz Litwin, Scott Gray, Benjamin Chess, Jack Clark, Christopher Berner, Sam McCandlish, Alec Radford, Ilya Sutskever, and Dario Amodei. 2020. [Language models are few-shot learners](#). In Advances in Neural Information Processing Systems 33: Annual Conference on Neural Information Processing Systems 2020, NeurIPS 2020, December 6-12, 2020, virtual.
- [3] Shijie Wu, Ozan Irsoy, Steven Lu, Vadim Dabravolski, Mark Dredze, Sebastian Gehrmann, Prabhanjan Kambadur, David S. Rosenberg, and Gideon Mann. 2023. [Bloomberggpt: A large language model for finance](#). CoRR,abs/2303.17564.
- [4] Dogu Araci. 2019. [Finbert: Financial sentiment analysis with pre-trained language models](#). CoRR, abs/1908.10063.