Retrieval Augmentation Using News Articles and Financials Data

Ali Elahi[1] , **Aliasgar Zakir Merchant**[2] ,
{aelahi6, amerch23}@uic.edu,

Abstract

This study presents an approach for aiding investment decisions using a Large Language Model (LLM). Drawing from a study by Lin *et al.* [2022], which revealed widespread interest in investments among the national population, we focus on predicting investment suitability for individual stocks. Our method integrates financial reports data and recent news articles to provide contextual information to the LLM. Employing retrieval-augmented generation, we combine user queries with relevant news articles retrieved through web scraping, further enriched by public financial data to prompt LLMs. Our dataset includes news articles and financial metrics for 20 companies over three years. We evaluate various pre-trained LLMs, including LLaMA, GPT-2, and one fine-tuned model, FinGPT. Results demonstrate our approach's effectiveness in predicting market trends and aiding investment decisions. Our best models achieve average precision and recall of 55% and 60%, respectively, compared to traditional machine learning models like random forest, which achieve approximately 60% and 40% average precision and recall. Predictions were made for 1-3 months return, with improved accuracy observed with longer time horizons.

1 Intordution

A study by Lin *et al.* [2022] revealed that in 2021, more than 35% of the national population engaged in investments such as stocks, bonds, and other securities, with a notable 79% focusing on individual stocks. The pursuit of long-term financial gains emerged as a prevalent objective, highlighted by 96% of investors, although only 48% were willing to assume average financial risks for corresponding returns. Notably, investors often turn to various sources for guidance, including research tools, financial literature, and advice from peers and family members. Social media also plays a role, with approximately 60% of young investors utilizing it as an informational resource. However, despite these resources, the study identified a concerning lack of investor knowledge, reflected in an average quiz score of 4.7 out of 10.

Figure 1: Financial assistance chatbox that we aim to construct.

The main objective of our project is to employ a Large Language Model (LLM) to predict whether someone should invest in a stock. The idea is to provide financial reports data as well as recent news articles regarding a company providing contextual information to the LLM to make investment decisions as to whether someone should invest in the company.

In our approach, we employ pre-trained and fine-tuned language models and implement retrieval-augmented generation Lewis *et al.* [2020] to provide contextual information. We utilize the user's query to retrieve relevant information about a specific company from recent public news articles. This retrieval process ensures that the insights provided by the large language model are grounded in up-to-date and pertinent data. to retrieve information from vast amount of news articles, we first filtered the articles by data and company and used extractive stigmatization technique to find related paragraphs. Additionally, we enrich the retrieved information by incorporating public financial data related to the company in question, such as revenue, net income, and trading volume. For this purpose, we collect our own dataset containing news articles and financial metrics.

Predicting market trends and making informed investment decisions has remained a focal point within the realm of finance for quite some time. The performance of a stock in the future is typically influenced by numerous factors, including the company's own performance and industry dynamics. Moreover, broader political and economic shifts can significantly impact market behavior, as they may prompt certain parties within the stock market to respond in particular ways.

Every publicly traded company in the United States is mandated by the Securities and Exchange Commission (SEC) to disclose its financial information at the end of each quarter, and among these regulatory filings, the 10-K report stands out as a comprehensive document detailing the company's financial performance, risks, and operational insights, reported on a quarterly basis. To include the company's performance, we included the data available in financial reports in our analysis.

As was previously indicated Lin *et al.* [2022], over 60% of investors (ages 25-45) cite social media as a source of knowledge. News stories are the most easily accessible type of content on social media, and serve as essential bits of knowledge while choosing an investment. Additionally, news articles is one of the main source of information to capture the investors reactions and effects on the stock market. Thereofore, we provide the information from news articles to the language model as well as financial metrics.

To gather the news articles we have compiled a list of frequently occurring keywords connected to investments that can be utilized during the retrieval process. These keywords encompass company names, their corresponding tickers[1], and phrases like "stock." For instance, search queries such as "Apple [AAPL] stock Feb 2024" exemplify the approach aimed at locating pertinent news articles.

2 Literature Review

Gupta *et al.* [2022] compares sentence embedding strategies for document ranking. They chose six pretrained sentence embedding algorithms: SentenceBERT, Universal Sentence Encoder, InferSent, ELMo, XLNet, and Doc2Vec, that are best suitable for document ranking in information retrieval systems. They conduct their studies on four standard datasets: CACM, CISI, ADI, and Medline. In this experiment, they address three research questions: the optimal model for phrase embedding using transfer learning, the strengths and limitations of each model, and potential areas for further research.

Sun *et al.* [2022] address the issues of insufficient labelled training data and the large train-test gap for language modeling by proposing a framework that measures semantic similarity based on the probability of generating context similar sentences. They develop a model that is able to predict the probability of a model fitting the left and right context and then develop sentence pair similarities based on the context scores generated. They utilize their framework to generate a new dataset that has sentence pairs and their corresponding similarity scores.

Ostendorff *et al.* [2022] perform aspect based document similarity for research papers. They utilize citation information for generating similarity between two research paper pairs simply following the well known idea "if the same citation is used in two papers, then these papers are similar to one another". They make use of the section headers of the occurrence of the citation as class labels and model the task as a multi-label multi-class classification problem.

Sarthi *et al.* [2024] utilize a tree like structure to capture details in a text document. This is done by clustering chunks of text to identify similarity, summarizing those chunks and then repeating the process to generate a tree from ground up. Their objective is to utilize text summarization during the retrieval augmentation stage. Each cluster of chunks is summarized by a language model which is then converted into text embedding and clustered once again. This process repeats until when the clustering becomes unfeasible The tree is then traversed to identify the most relevant clusters.

Glass *et al.* [2022] combine neutral retrieval and re-ranking into a BART based generation task. They introduce a new reranking stage in addition to the usual ranking stage which combines document passages produced using incomparable results like ANN Index and BM25 Index. The resultant top-k documents are used in the query to produce a response. This new technique is tested on the KILT benchmark for four different tasks: slot filling, question answering, fact checking and dialog.

Jeong [2023] present an enterprise LLM based RAG architecture. They explore the various components of the RAG architecture including but not limited to Vector Databases, Embeddings, and Chunking. Additionally, they provide a LangChain based generative AI RAG model implementation.

3 Dataset

In this project, we are assembling a dataset for investment analysis. This process necessitates gathering data from two primary sources: news articles and financial reports. To achieve this, we conducted web scraping of news websites to obtain real-time updates. Additionally, we sourced financial metrics from 10-K financial reports. We collected data for 20 companies for at least the last 3 years from 2022-present. a list of companies are available in Table 1.

3.1 News Articles

We utilize the scrapper to gather articles which discuss investment in the top 20 S&P-500 companies. A basic query template given to google news to obtain relevant articles is:

$\{$*Company-Name*$\}$ stock investment $\{$*Date*$\}$ yahoo finance

We scrape the information from the generated articles and develop the dataset. A total of ten thousands news articles were extracted across 20 companies which were then used to test the retrieval algorithm used. A sample of our dataset is shown in 1. The dataset consists of 6 columns: Title, Description, Date-Time, Keywords, URL, and Content.

3.2 Financial Reports Data

The financial data is gathered from the 10-K Quarterly Filings. These reports contain operational/income statement, cash flow statement, and balance sheet, which are available on companies' websites or the SEC websites. Table 2 contains descriptions of some information found in financial reports.

4 Methodology and Results

4.1 Scrapper

The use of publicly accessible APIs and web scraping to acquire open-source information has become more popular

[1] a symbol, a unique combination of letters and numbers that represent a particular stock or security

Samples
Title: Should You Invest in Apple (AAPL) Based on Bullish Wall Street Views? **Description:** Based on the average brokerage recommendation (ABR), Apple (AAPL)... **Data-Time:** April 12, 2024 at 6:30 AM **Keywords:** Strong Buy, Zacks Rank, brokerage firms, Apple,... **Content:** Investors often turn to recommendations made by Wall Street analysts... **URL:** https://news.google.com/articles/CBMiTWh0d...
Title: Has the Sun Set on Apple's Innovation-Driven S... **Description:** The Cupertino colossus has been making only in... **Data-Time:** March 13, 2024 at 7:21 AM **Keywords:** Apple, Apple Watches, Apple Store **Content:** Once one of the strongest-growing businesses o...
Title: Bulls In A Bear Market: These 10 Stocks Clocked Gains In Excess Of 100% In 2022 - Helmerich & Payne (NYSE:HP) - Benzinga... **Description:** 2022 would go down as one of the worst years for the financial markets... **Data-Time:** December 31, 2022 at 11:11 AM **Keywords:** NA **Content:** 2022 would go down as one of the worst years for the... **URL:** https://www.benzinga.com/analyst-ratings/analyst-color/..
Companies Names That The News Are Collected About
Apple Inc. (AAPL), Amazon.com Inc. (AMZN), Alphabet Inc. (GOOGL), Microsoft Corporation (MSFT), Tesla, Inc. (TSLA), Facebook, Inc. (FB), Berkshire Hathaway Inc. (BRK.B), Johnson & Johnson (JNJ), Visa Inc. (V), JPMorgan Chase & Co. (JPM), Procter & Gamble Company (PG), Walmart Inc. (WMT), The Coca-Cola Company (KO), Netflix Inc. (NFLX), Pfizer Inc. (PFE), Walt Disney Company (DIS), Nvidia Corporation (NVDA), Alibaba Group Holding Limited (BABA), Adobe Inc. (ADBE), Mastercard Incorporated (MA)

Table 1: News Article Samples

Financial Variables	**Source**	**Description**
Total Revenue	(10K) Reports	Aggregate amount of income generated from the sale of goods or services before deducting any expenses
Net Income	(10K) Reports	Also known as profit or net earnings, is the total amount of revenue earned by a company after deducting all expenses
Free Cash Flow	(10K) Reports	Represents the amount of cash generated by a company's operations
Total Assets	(10K) Reports	Combined value of all resources owned or controlled by a company
Price Momentum	Historical Pricing Data	Price Momentum 6-12 Months — Measures the relative strength and direction of a stock's price movement over the past months.
Forward Return	Historical Pricing Data	Expected or projected return on an investment or asset over a future period. **Used as target**

Table 2: Descriptions for the financial data available in financial reports extracted from Income Statements, balance sheets, and cash flow statements.

among developers. We concentrate on acquiring two categories of news articles: historical news and current news, both of which are related to the Top 20 S&P-500 firms. The foundation for developing our web scraper is provided by the Request, and Beautiful Soup libraries in Python. We crawl articles that are available on Google News and published on Yahoo Finance to gather up-to-date news articles. We will leverage the "Market News and Sentiment" API from alphavantage-API[2], which can give us access to both historical and real-time market news and sentiment data, to gather historical news items.

4.2 Retrieving Methods

Due to the high number of news articles and information about each company in a given date, we need to look for a method to extract the important information in the large number of articles. The filtering and summarizing process should be able to detect the relevant portions of the news articles and select the most relevant portions to the query.

[2]https://www.alphavantage.co/documentation/intelligence

Building upon the methodology outlined in Sarthi *et al.* [2024], we implemented a layered summarization approach. Initially, we filtered news articles based on key metadata such as title, subtitle, and publication date, searching for the company name and index ticker (each company has an symbol called ticker such as "AAPL" for apple company) within the articles. Articles containing these keywords and published before the search date were retained for further analysis.

Subsequently, we proceeded with summarizing the selected articles using. For comparison, we employed both of abstractive and extractive summarization techniques. For extractive summarization, we employed both OpenAI and Sentence-BERT embedding to encode the news articles and compute similarities with the user's query. Additionally, for abstractive summarization, we leveraged the GPT-3.5 model to generate summaries of the most relevant articles to the user's query. Figure 2 shows the summarizing processes.

To simulate user's queries we tried a set of different queries such as:

- Investing in *company-name* company in *date*.

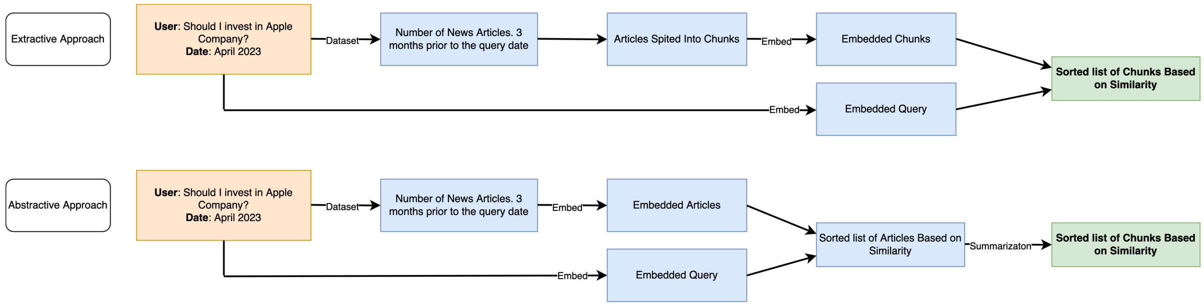

Figure 2: Summarizing flow diagram

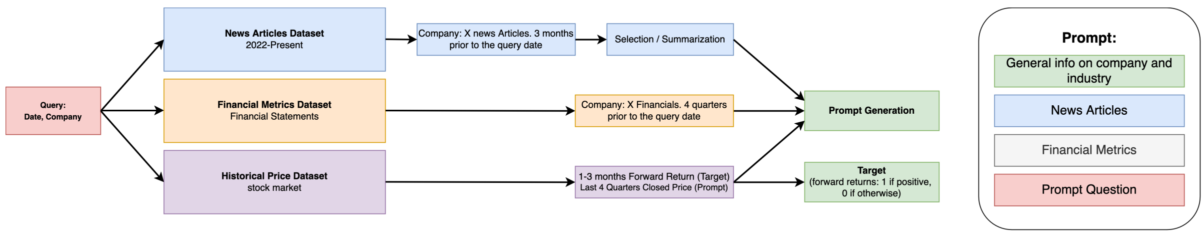

Figure 3: Prompt generation flow diagram.

- Should I invest in *company-name* in company *date*
- Is *company-name* company index going up after *date*?

Table 4 (Appendix - Last Page) shows the summaries examples. The chunks of Articles can be from one or multiple news articles. as you can see, the two OpenAI and Sentence BERT embedding can result in the same chunks. Moreover, one method can recommend chunks from the same article.

The granularity of chunks can effect information extraction. lower chunk sizes can result in abstract yes/no or Invest/Do Not Inverts summaries while larger chunk sizes can convey detailed explanations. Similarly, extractive summaries can result in limited information while abstractive summaries can be over-generalized.

4.3 Prompt Generation and Labeling

After collecting news articles and financial data, our next step involved generating prompts and their expected responses. Each prompt comprises a synopsis of the company's products and its industry, a summary of recent news articles, the first five financial metrics for the previous four quarters (as depicted in Table 2), and finally, the prompt question. The labels for each prompt would be the forward return for upcoming months after the date mentioned in the prompt question.

For instance, if the prompt question is: "Should I invest in Apple Company in July 2022?" The news articles and financial metrics will be sampled from May-June 2022 and July 2021 to June 2022, respectively. Furthermore, we will assess the results based on three targets: whether the stock in question experienced an increase (labeled as 1) or decrease (labeled as 0) in price in the months following (referred to as forward return in Table 2).

In our analysis, we examined the language model's responses across one, two, and three months of forward return to determine the optimal time period for prediction. The general prompt generation process is visualized in the 3.

Company's General Information We included a brief statement about the company's products, primary activities, and the industry it operates within. This addition serves two purposes. Firstly, stock prices are often influenced by industry trends, and we want the language model to consider this context when generating responses. Secondly, since the prompts are situated in the past (e.g., a prompt from February 2022), the target outcome has already occurred. This introduces a risk of data leakage, as the language model has already been trained on information available online. Therefore, in the prompt generation process, we replace the company's name with a placeholder like "Company X." However, to provide necessary context about the company's operations, we include a general information sentence at the beginning of the prompt.

News Articles The news articles are filtered and summarized based on the methods described in Section 4.2 but after manual analysis we decided to move forward with the extractive summarization method using OpenAI embeddings. The number of explicitely informative sentences in this method were more than the two other methods after analysing 20 examples similar to the one present in Table 4 (in appendix).

Financials The five financial metrics in Table 2 for the last four quarters prior to the date in question were given to the language model in the prompt. These financial metrics are in based on time series but to avoid longer prompts and confusion, only previous four quarters were used.

4.4 Models

We tried to use a viderange of pre-trained language models as well as one publicly available fine-tuned language model in financial domain. We used LLaMA2 model, 13B and 70B

Models		1 Months Forward Return					2 Months Forward Return					3 Months Forward Return				
Model	**Prompt**	NP	PP	NR	PR	ACC	NP	PP	NR	PR	ACC	NP	PP	NR	PR	ACC
LLaMA 13B	**News**	0.451	0.000	1.000	0.000	0.452	0.484	0.000	1.000	0.000	0.483	0.322	0.000	1.000	0.000	0.322
	News+Financials	0.394	0.363	0.666	0.157	0.387	0.455	0.440	0.689	0.230	0.452	0.284	0.577	0.633	0.238	0.366
LLaMA 70B	**News**	0.375	0.541	0.071	0.901	0.527	0.000	0.516	0.000	1.000	0.516	0.000	0.677	0.000	1.000	0.677
	News+Financials	0.000	0.548	0.000	1.0	0.548	0.833	0.540	0.111	0.979	0.559	0.250	0.674	0.033	0.952	0.656
GPT2	**News**	0.000	0.548	0.000	1.000	0.548	0.000	0.516	0.000	1.000	0.516	0.000	0.677	0.000	1.000	0.677
	News+Financials	0.440	0.000	0.952	0.000	0.430	0.579	0.540	0.244	0.833	0.548	0.321	0.677	0.300	0.698	0.570
FinGPT	**News**	0	0.438	0	1	0.4375	0	0.250	0.000	1.000	0.250	0.0000	0.562	0.000	1.000	0.410
	News+Financials	0.000	0.258	0.000	1.000	0.258	0.000	0.548	0.000	1.000	0.548	0.000	0.548	0.000	1.000	0.548

Table 3: The highlighted results are showing a better performance

versions as well as GPT2 model. Unfortunately we could not use models such as GPT3.5 or GPT4 due to their protection layer not to answer investment questions[3].

Finally, we used FinGPT [Wang *et al.*, 2023; Yang *et al.*, 2023], an instruction tuned model using LoRA [Hu *et al.*, 2021]. This model was tuned on instructions such as sentiment analysis in financial domain and return forecasting.

5 Experiments Results

In order to test the prediction accuracy of our financial model, we prepare a set of 100 samples. Each sample consists of the user query, the company information, a set of generated text articles, and the corresponding financial metrics. For each sample, we generate answers from our RAG models. Each answer is sent as input to chat based models which are prompted to generate (yes or no) answers to the question *Does the text suggest whether the company is financially stable or not?*. These answers are then converted to binary values and compared against a 1-3 month forward return, with 1 being a positive return and 0 being a negative return.

5.1 Prompting Process

Each sample is sent as input to our RAG models, in particular GPT-2, Llama 7B-13B-70B, and FinGPT, and answers are generated. These answers are then sent to the chat model selected. To choose our chat model, we used the Vercel AI Playground to test and select the chat model which best answered our query based on the following chat template:

- **System Prompt**: *You will answer with just 1 and 0 with 1 representing Yes and 0 representing No.*
- **User Prompt**: *Does the text suggest Company X is financially stable or not?* $\{text\}$

After manual selection of the chat-based models, we send the answers generated by our RAG model to the chat model to generate a yes or no response. The prompt was selected after carefully examining the results of the chat models for various fine-tuned prompts.

We evaluate the performance of the model by generated answers for queries with and without the company's financial metrics. These answers are converted into binary values with 1 representing a positive response and 0 representing a negative response. Each response is then compared against the ground truth and corresponding accuracy, precision and recall metrics are generated. The ground truth are obtained by retrieving the 1-3 month forward return. For each of the forward return values, we convert the forward returns to binary values with 1 being a positive return and 0 being a negative return.

5.2 Metrics

We focus on the following metrics to evaluate the performance of the RAG models:

- Precision: refers to the proportion of correctly identified positive samples divided by total number of positively identified examples.
- Recall: refers to the proportion of correctly identified positive samples divided by the total number of positive examples.
- Accuracy: refers to the proportion of correctly identified positive and negative samples divided by the total number of examples in the dataset.

5.3 Results

We began our analysis by comparing LLMs results with traditional machine learning models. We employed the Random Forest model for this classification task. Our training involved utilizing financial data and incorporating embedded news articles. With Random Forest, we achieved an average precision and recall of 60% and 40% respectively. Comparatively, LLMs demonstrate notable enhancements in performance (Table 3).

We obtained the following results upon testing the answers for our RAG model:

- GPT-2 model performed the best when it came to generating financial answers. When only news articles were provided, the recall of the positive class was one, indicating that the answers provided by the RAG models were all positive. But when the companies financial data was provided the recall reduced indicating the model was now generating negative answers as well. However, with an accuracy reduction suggests that there are false answers generated by the model.
- FinGPT model performed slightly worse as compared to GPT-2 based models. When reviewing the chat model generated binary values, it was noticed that the model returned a value of 1 for every answer indicating that

[3]The respond with answers such as I do not know or I am not able to answer to such questions.

the RAG model generated a positive response for every query sent to it. This can also be identified by the return value of 1 for the positive class recall.

- Llama based models performed the worse as compared to all other models. When reviewing the chat model generated binary values, it was noticed that the model also returned a value of 1 for every answer. When we reviewed the answers, we noticed something peculiar, all Llama models repeated the context they were provided in their answers.

We hypothesize a reason for the irregular performance of the FinGPT models to be the instructions provided to the models finr tuning. Verifying the answers for FinGPT also became difficult since a single answer generation was taking about 10 minutes on an average.

Our metrics show that models which were given a context of financial data for a company had a better accuracy in most cases than models which were not provided the financial data. Additionally, as represented in Table 3, the accuracy of the models improves as it we move towards the 3 month forward return, which indicates that the model is good at generating answers which are valid to a three month period.

References

Michael Glass, Gaetano Rossiello, Md Faisal Mahbub Chowdhury, Ankita Naik, Pengshan Cai, and Alfio Gliozzo. Re2G: Retrieve, rerank, generate. In Marine Carpuat, Marie-Catherine de Marneffe, and Ivan Vladimir Meza Ruiz, editors, *Proceedings of the 2022 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, pages 2701–2715, Seattle, United States, July 2022. Association for Computational Linguistics.

Vishal Gupta, Ashutosh Dixit, and Shilpa Sethi. A comparative analysis of sentence embedding techniques for document ranking. *J. Web Eng.*, December 2022.

Edward J Hu, Yelong Shen, Phillip Wallis, Zeyuan Allen-Zhu, Yuanzhi Li, Shean Wang, Lu Wang, and Weizhu Chen. LoRA: Low-Rank adaptation of large language models. June 2021.

Cheonsu Jeong. A study on the implementation of generative ai services using an enterprise data-based llm application architecture. *Advances in Artificial Intelligence and Machine Learning*, 03(04):1588–1618, 2023.

Patrick Lewis, Ethan Perez, Aleksandra Piktus, Fabio Petroni, Vladimir Karpukhin, Naman Goyal, Heinrich Küttler, Mike Lewis, Wen-Tau Yih, Tim Rocktäschel, Sebastian Riedel, and Douwe Kiela. Retrieval-augmented generation for knowledge-intensive NLP tasks. May 2020.

J T Lin, C Bumcrot, G Mottola, O Valdes, and G Walsh. *The Changing Landscape of Investors in the United States: A Report of the National Financial Capability Study. FINRA Investor Education Foundation*. 2022.

Malte Ostendorff, Till Blume, Terry Ruas, Bela Gipp, and Georg Rehm. Specialized document embeddings for aspect-based similarity of research papers. In *Proceedings of the 22nd ACM/IEEE Joint Conference on Digital Libraries*, New York, NY, USA, June 2022. ACM.

Parth Sarthi, Salman Abdullah, Aditi Tuli, Shubh Khanna, Anna Goldie, and Christopher D Manning. RAPTOR: Recursive abstractive processing for tree-organized retrieval. January 2024.

Xiaofei Sun, Yuxian Meng, Xiang Ao, Fei Wu, Tianwei Zhang, Jiwei Li, and Chun Fan. Sentence similarity based on contexts. *Trans. Assoc. Comput. Linguist.*, 10:573–588, May 2022.

Neng Wang, Hongyang Yang, and Christina Dan Wang. FinGPT: Instruction tuning benchmark for open-source large language models in financial datasets. October 2023.

Hongyang Yang, Xiao-Yang Liu, and Christina Dan Wang. FinGPT: Open-source financial large language models. June 2023.

Query Info	**Query:** Should I invest in Visa company Feb 2024? **Count:** 3 **Chunk Size:** 3
Extractive: OpenAI	**Chunk 1:** similarity: 0.660 **Article Title:** Visa vs. Mastercard: Which Stock Is the Better Buy Today? **Text:**If you want to buy into blue chip businesses with long runways for sustainable growth, **consider buying both.** Should you invest $1,000 in Visa right now? Before you buy stock in Visa, consider this: The Motley Fool Stock Advisor analyst team just identified what **they believe are the best stocks for investors to buy now... and Visa wasn't one of them.** **Chunk 2:** similarity: 0.659 **Article Title:** Want $1 Million in Retirement? Invest $50,000 in These 3 Stocks and Wait a Decade. **Text:** Should you invest $1,000 in Visa right now? Before you buy stock in Visa, consider this: The Motley Fool Stock Advisor analyst team just identified what they believe are the **best stocks for investors to buy now... and Visa wasn't one of them.** The 10 stocks that made the cut could produce monster returns in the coming years. **Chunk 3:** similarity: 0.651 **Article Title:** The Zacks Analyst Blog Highlights NVIDIA, Visa, Amgen, Chubb and PACCAR. **Text:** The company's strategic acquisitions and alliances are fostering long-term growth and consistently driving its revenues. It expects net revenues to increase in the low double digits for fiscal 2024. **Visa, fueled by increased payments and sustained investments in technology, is witnessing bottom-line growth.** The ongoing shift to digital payments is advantageous for Visa, with strong domestic volumes supporting its overall performance.
Extractive:S-BERT	**Chunk 1:** similarity: 0.724 **Article Title:** Want $1 Million in Retirement? Invest $50,000 in These 3 Stocks and Wait a Decade. **Text:** Should you invest $1,000 in Visa right now? Before you buy stock in Visa, consider this: The Motley Fool Stock Advisor analyst team just identified what **they believe are the best stocks for investors to buy now... and Visa wasn't one of them.** The 10 stocks that made the cut could produce monster returns in the coming years. **Chunk 2:** similarity: 0.672 **Article Title:** The Zacks Analyst Blog Highlights NVIDIA, Visa, Amgen, Chubb and PACCAR **Text:** The company's strategic acquisitions and alliances are fostering long-term growth and consistently driving its revenues. It expects net revenues to increase in the low double digits for fiscal 2024. Visa, fueled by increased payments and sustained investments in technology, is witnessing bottom-line growth. The ongoing shift to digital payments is advantageous for Visa, with strong domestic volumes supporting its overall performance. **Chunk 3:** similarity: 0.628 **Article Title:** Want $1 Million in Retirement? Invest $50,000 in These 3 Stocks and Wait a Decade. **Text:** **At any rate, Visa's no-nonsense business model has delivered exceptional shareholder gains.** Over the last decade, Visa's stock has generated an annualized total return (price appreciation plus dividend payments) of 18.4%. That means $50,000 invested in Visa 10 years ago would be worth $271,000 today – not too shabby.
Abstractive: GPT-3.5	**Chunk 1:** similarity: 0.676 **Article Title:** Why Visa (V) is a Top Stock for the Long-Term. **Text:** **Warren Buffett's Berkshire Hathaway holds a massive portfolio of stocks, including American Express, Visa, and Mastercard.** These three companies dominate the payments landscape and have strong economic moats, making them attractive investments for long-term investors. While their valuations may seem high, their profitable business models and network effects make them unlikely to be disrupted. Despite not being top picks for Buffett currently, these stocks should be on investors' radar for potential long-term growth. **Chunk 2:** similarity: 0.627 **Article Title:** Visa (V) Enhances Offerings With Virtual Corporate Cards. **Text:** The article discusses the potential for cryptocurrency, specifically Bitcoin, to become the 12th sector of the S&P 500 in the next few years. This is due to factors such as widespread adoption, the stability of stablecoins, and the success of companies like MicroStrategy who have heavily invested in Bitcoin. The recent approval of Bitcoin ETFs by the U.S. Securities and Exchange Commission has also opened the door for institutional adoption. The article suggests that with a current market cap of over $2 trillion, the cryptocurrency market has a chance to become a major player in the stock market. **Chunk 3:** similarity: 0.584 **Article Title:** Visa vs. Mastercard: Which Stock Is the Better Buy Today? **Text:** Building an investment portfolio from scratch can be overwhelming, but **focusing on stocks that are set to outperform the market, such as Visa (V), can lead to success.** V has been on the Zacks Focus List since 2017 and has seen a 191.34% increase in shares since then. With analysts expecting earnings to grow 13% for the current fiscal year, buying stocks with rising earnings estimates like V can be profitable.

Table 4: Another Example – Comparison on summarization methods. **The purple sentences** are the ones that seem explicitly informative in user's eyes. The similarity values are referring to cosine similarity between chunks and query.