ANNUAL REPORT ANALYSIS WITH ADVANCED LANGUAGE MODELS: A STOCK INVESTMENT STRATEGY ENHANCEMENT

Ishan Desai[1], Rebanta Daadhiich[2], Rushi Savani[3], Sonali Jadhav[4]

[1]Department of Information Technology, Dwarkadas J. Sanghvi College of Engineering, Maharashtra, India
[2]Department of Computer Engineering, Thadomal Shahani Engineering College, Maharashtra, India
[3]Department of Information Technology, Thadomal Shahani Engineering College, Maharashtra, India
[4]Professor, Dept. of Computer Engineering, Thadomal Shahani Engineering College, Maharashtra, India

---***---

Abstract - *Publicly traded firms' annual reports include important details about their financial situation that can be used to estimate possible effects on the company's stock price. These reports are quite extensive, often surpassing 100 pages in length. Even for a single company, analysing these data is a laborious task; imagine how much more so for all the companies in existence. Financial specialists have mastered the art of swiftly and efficiently obtaining important information from these documents over time. But years of expertise and practice are needed for this. By utilizing Large Language Models' (LLMs) capabilities, this article seeks to streamline the evaluation of each company's annual report. The LLM's observations are combined with historical stock price data and organized in a Quant-styled dataset. In terms of S&P 500 returns, the walk forward test findings indicate a promising outperformance.*

Key Words: **Chat-GPT, LLM, Stocks, Investing, Quantitative Finance**

1.INTRODUCTION

The American stock market provides the stocks that were taken into consideration for this analysis. The top 1500 US companies are represented by the large cap index, midcap index, and small-cap index. These records are referred to as 10-K filings. Using the company's 10-K filing, investors assess a company's banking statements along with balance sheet. This essay uses the terms 10-K and annual report interchangeably.

Along with the company's financial statements, the 10-K includes the statement of cash flows, statement of assets, and statement of income. However, it also provides additional useful information not shown by financial measures and ratios. The fact that these components cannot be summed up into a single number makes them challenging to evaluate.

Large Language Models (LLMs), such as GPT-3.5 (sometimes referred to as Chat-GPT), have recently become effective tools for improving comprehension and analysis of lengthy documents, including tasks like document summarization [1]. According to A. Lopez-Lira et al. [2], LLMs can be used to accurately anticipate stock prices in the context of financial applications. For our use case, we investigate whether LLMs could be used to respond to complicated queries financial analysts could have about the business and which could be answered using the data from yearly reports. "Is there a clear growth and innovation strategy in place for the business?" is one example of a question like this. Are there any current initiatives or strategic partnerships?

2. Context and Associated Works

The study makes use of Open-AI's GPT-3.5 version, which Chat-GPT is currently using as well. In order to predict how the stock price will change over the course of the following day, A. Lopez-Lira et al. [2] create a prompt with a news headline that is specific to the firm and ask the LLM to assign the right mood to it. Their study's findings demonstrated the statistically significant predictive power of Chat-GPT-powered sentiment analysis-based stock picking. Although their research provides insightful information about the use of LLMs in finance, there are several difficulties. They first provide the LLM instructions on how to respond to a prompt in binary form, which is then utilized to decide whether to buy or sell the company's shares. Their study's findings demonstrated the statistically significant predictive power of Chat-GPT powered sentiment analysis-based stock picking. Although their research provides insightful information about the use of LLMs in finance, there are several difficulties. They first provide the LLM instructions on how to respond to a prompt in binary form, which is then utilized to decide whether to buy or sell the company's shares. They show in their paper that the strategy's net return is significantly affected by the inclusion of transaction costs.

This work introduces the idea of obtaining insights from thorough company annual reports, drawing inspiration from the work of A. Lopez-Lira et al. [2] to apply LLMs in Investment decision making while also taking the drawbacks of short-term trading into consideration. Since these reports are released yearly, the signals they produce have a longer duration, and selecting stocks based on these signals would only result in a small annual transaction cost. Furthermore, the best performing stocks for the next year are predicted in this paper by combining the output of the LLM with a Machine Learning model. The machine learning model has the ability to identify important features and ignore less important ones. It can also capture complex relationships between the features, which will ultimately result in predictions that are more accurate.

3. DATA

The historical 10-K filings made by businesses are available in the SEC's Edgar database. These reports, together with the corresponding filing dates, are retrieved and saved locally. In order to complete this exercise, 24,200 10-K filings from the years 2002 to 2023 were retrieved; these files take up about 85 GB of disk space.

Assigning corresponding target values is just as important as creating features for the machine learning model. The return at intermediate intervals, such as the percentile of time. Additionally, the stock's maximum and minimum returns for this time frame are calculated. In Section 4.4, the finer points of figuring out target values for the ML model (from raw returns) are discussed. In addition to the returns unique to each stock, index returns are also computed at the beginning and ending periods. This helps to compare the returns on the stock portfolio when the ML model is applied to the data set.

The testing set, 500 data points (out of 6.8k) were chosen for evaluation. The terms "training set" and "testing set" will refer to corresponding sampled versions of each going forward in the paper. Your total workout will put you back by approximately $60. In addition to the financial cost, it took a long time to process. For the sampled dataset, the entire exercise involving saving the document embeddings and utilizing GPT-3.5 to process the questions took about 50 hours. The time and expense involved would rise proportionately if this exercise were expanded to include the entire dataset.

4. METHODS

A. *Getting to Annual Reports*

The top 1500 corporations by market capitalization must have access to historical 10-K filings in order to create the dataset. Wikipedia was used to compile a list of these firms along with their ticker symbols. To do this, it is possible to retrieve the URLs for all previous corporate filings using a tool like Financial Modelling Prep [4]. Note: Under the free access plan, this website may impose rate limits and is gated via API access.

B. *Document Embeddings*

To make an informed decision, a number of text embedding models are available, and their effectiveness is assessed using the Massive Text Embedding Benchmark (MTEB) [5]. The all-mpnet-base-v2 [6] stands out among these models thanks to its high MTEB score and quick processing speed. To be complete, it should be noted that text-embedding-ada-002, one of Open-AI's embedding models, is a viable substitute as well. Given the magnitude of the papers and the significant accompanying costs, it was not used in this case. The all-mpnet-base-v2 model [6] is a more useful option for this exercise because it can be done locally on a typical laptop in a reasonable amount of time.

The embeddings' storage in a vector database is an even more, modest but essential factor. ChromaDB [7] was chosen for this study because of its smooth compliance with LLama Index [8, the main LLM framework].

C. *LLM for Feature Generation*

The GPT-3.5-Turbo LLM from Open-AI was employed in this investigation [9]. Future developments in more sophisticated models are envisaged given the field's continuous advancement. Integrating LLM models from various sources is streamlined by the Llama Index architecture [8]. We've picked GPT-3.5-Turbo [9] as our preferred LLM for this exercise because of its excellent performance and simplicity of use with the Open-AI API.

D. *Label Creation*

The target max is calculated, which is the 98th percentile of the return from the filing date, and use it as a stand-in for the maximum. This figure represents the highest return that the stock could achieve in the year between two consecutive filings. Similarly, the S&P 500 index's sp500 max is also calculated.

Target 12m and Target Max are the sources of the target value for the machine learning model. In order to accomplish this, we consult Numeral's data documentation [10], which offers instructions on how to build the target variable using raw returns. The actions listed below are taken:

1. Each year's target stock values are allocated differently. This is done in order to rank stock returns within each year in a relative manner.

2. After ranking, the returns are normalized.

3. The range [0,1] contains the target values. One represents larger returns.

E. *Model for Machine Learning*

Complex feature relationships can be captured by more sophisticated techniques like Gradient Boosted Decision Trees (GBDTs), but they also require regularization and hyper-parameter tuning. Independent research into the application of GBDTs is undoubtedly feasible.

5. RESULTS

a. *Choosing the Right Number of Stocks to Purchase*

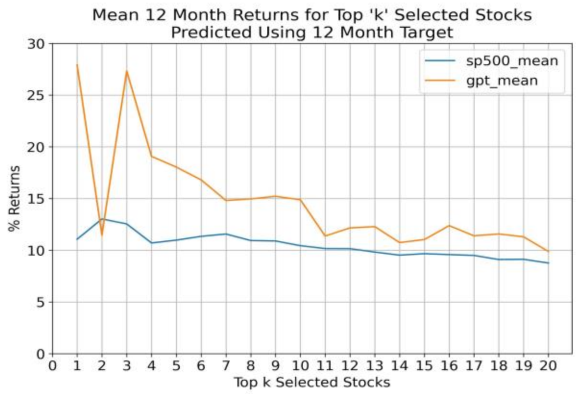


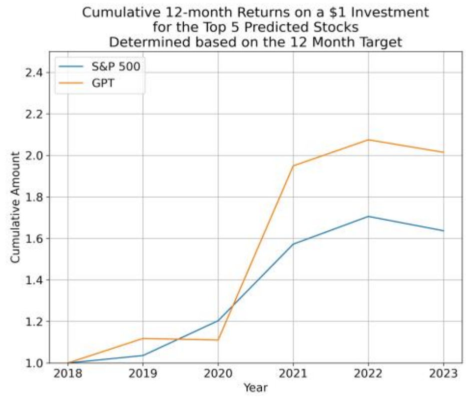
Fig. 1: Comparing the Twelve-Month Returns

Figure 1 compares the returns produced by the S&P 500 during the same time period with those produced by the GPT model using the top k stocks. The figure illustrates how the returns increase for lower k values and decrease for higher k values. This proves that, as predicted by the GPT model, higher-rated stocks provide better returns.

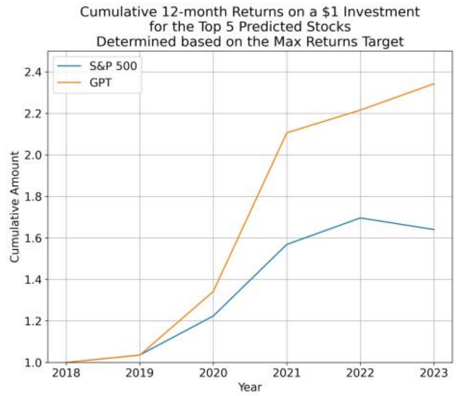
Fig. 2: Comparing the Twelve-Month Returns for K values

The mean returns over a year are shown in Figures 1 and 2, respectively. The distinction is that two different target variables were used to create these numbers. The charts above allow for some noteworthy findings. First off, having a lower value of k is helpful for maximizing returns. An appropriate choice in this situation would seem to be a k value of 5. When k is set to 5, it means that the buy strategy will be used for 5% of the available stocks. The test set consists of 500 equities that were randomly picked during a five-year period in order to provide context. Thus, choosing 5 stocks annually is equivalent to choosing 5 of 100 equities each year, or a selection rate of 5%.

b. *Examining Total Returns for Various Approaches*

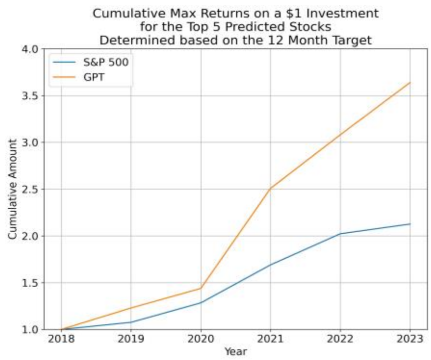


(a) Comparing with Twelve-Month Returns

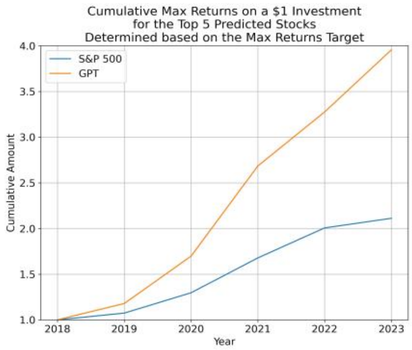


(b) Utilising Highest Return Target as a model

Fig. 3: Total Twelve-Month Returns for the Top 5 Projected Stocks on a $1 Investment



(a) Modelled with a 12-Month Returns Objective



(b) Utilising Max Returns Target as a model

Fig. 4: Cumulative Max Returns on a $1 Investment for the Top 5 Predicted Stocks

This figure presents an analysis of the highest returns technique. Returns are computed using this method starting on the date of the annual report and ending at the 98th percentile of the stock price for that year. Essentially, it is assumed that stocks are bought soon after the annual report is made public and that they are subsequently sold at or close to the year's peak price. Despite the fact that this scenario may seem overly optimistic, it's crucial to keep in mind that the S&P 500 returns are created using a similar procedure. Thus, we can directly and fairly compare the Highest Returns strategy to the index using this methodology.

6. CONCLUSION

The target (response) variable can be constructed and chosen in a flexible manner depending on different timeframes. Two different target variable types have been examined and found to be useful in this paper. However, it is crucial to acknowledge that there is still room for other approaches to defining the target variable.

This is important because a large number of actively managed strategies that are currently used to produce alpha through short-term trading strategies may not be net profitable and have high transaction costs. This paper demonstrates how investing with long-term money management (LLMs) can be advantageous without incurring high transaction costs.

REFERENCES

[1] T.T. Guang Lu, Sylvia B. Larcher, Hybrid long document summarization using c2f-far and chatgpt: A practical study. arXiv e-prints (2023). URL https://arxiv. org/abs/2306.01169. arXiv:2306.01169

[2] Gibbeum Lee, Volker Hartmann, Jongho Park, Dimitris Papailiopoulos, Kangwook Lee, Prompted LLMs as Chatbot Modules for Long Open-domain Conversation (2023). URL - https://arxiv.org/abs/2305.04533

[3] Modelling and forecasting S&P 500 stock prices using hybrid Arima-Garch Model https://iopscience.iop.org/article/10.1088/1742-6596/1366/1/012130

[4] Massive text embedding benchmark. https://huggingface.co/blog/mteb. Accessed: 2023-09-01

[5] Sentence transformers - all-mpnet-base-v2. https://huggingface.co/ sentence-transformers/all-mpnet-base-v2. Accessed: 2023-09-01

[6] J. Liu. LlamaIndex (2022). https://doi.org/10.5281/zenodo.1234. URL https: //github.com/jerryjliu/llama index

[7] Openai - gpt 3.5 turbo. https://platform.openai.com/docs/models/gpt-3-5. Accessed: 2023-09-01

[8] Numerai data documentation. https://docs.numer.ai/numerai-tournament/data. Accessed: 2023-09-01

[9] M.H. Martin Slawski, Non-negative least squares for high-dimensional linear models: consistency and sparse recovery without regularization. arXiv e-prints (2014).
URL https://arxiv.org/abs/1205.0953. arXiv:1205.0953