Democratizing Stock Market Analysis with a Sentiment Bot

Ryan Madar
College of Inovation and Technology
University of Michigan Flint
Flint, United States
rmadar@umich.edu

Cole Rutledge
College of Inovation and Technology
University of Michigan Flint
Flint, United States
dcrutled@umich.edu

Trevor Heydt
College of Inovation and Technology
University of Michigan Flint
Flint, United States
godtier@umich.edu

Jonathan Stokes
College of Inovation and Technology
University of Michigan Flint
Flint, United States
stokesj@umich.edu

Abstract*—Few people can predict the stock market and capitalize on price swings *without insider knowledge. Large Language Models (LLMs) like ChatGPT, FinGPT, and BloombergGPT provide a unique opportunity to analyze and respond to stock market news. We propose using an open-source model, fine-tuned to generate insights into stock market news. This approach enables the average person stock market news quickly and develop trading sentiments.

I. INTRODUCTION

As the old saying goes, "Time in the market beats timing the market". This adage rings true, because the stock market is a very fickle entity. Very few people have been able to predict the stock market and capitalize on price swings in specific stocks without insider knowledge. Despite this, countless have tried and countless more will try to predict the stock market in an attempt to grow their wealth exponentially. But what if you didn't have to beat the market? What if, with the help of technology, you could use worldwide sentiment of upcoming stock market trends to beat the market for you? The primary goal of this project is to understand and explore the use of large language models, and the closely related field of natural language processing, for the purpose of predicting stock trends and capitalizing on these predictions.

Large language models such as ChatGPT, Bard, and T5 have been in the public zeitgeist since 2020, with a seemingly unlimited number of uses. Companies like OpenAI and Google train these models on more data than a human could read in a lifetime, then let the models use the data to receive inputs and produce seemingly intelligent outputs. For our purposes, the data training the models we discuss will often be much narrower and more focused through a financial lens. And the output, in theory, could too be more focused, and if the models work, more accurate than even human predictions.

II. LARGE LANGUAGE MODELS

LLMs are a specific type of machine learning program that utilize transformer models to "predict" text. Transformer models a specific type of language processing architecture that uses an encoder and decoder. At the most basic level this means that LLMs that use transformer models take in text data, tokenize it so it can be readily interpreted by the computer; and then using probabilistic mathematics discovers relationships between tokens. This is how the LLM can "predict" what should come next or what the answer to a question should be. It's very similar to how the human brain recognizes patterns and then from those patterns or webs of patterns makes inferences about the world around it. Hence, why LLMs are considered neural networks, on some level they resemble the human brain and how it processes information. This is a really high-level overview of how LLMs work without getting into great detail. You could break that functionality down layers – embedding layer, recurrent layer, or the whole idea of attention in machine learning – but for now we'll stick with this summary and build on it later.

III. PREVIOUS PAPERS

Can ChatGPT Forecast Stock Price Movements? Return Predictability and Large Language Models

In this first paper, Lopez-Lira and Tang explore the potential of general unfocused large language models for predicting stock prices. They focus primarily on the advanced ChatGPT model, but also test more basic models including BERT, GPT-1, and GPT-2. They find effectiveness in ChatGPT's ability to predict stock price trends, but less so in the basic models. ChatGPT proved effective in interpreting many different inputs, but some themes, such as partnerships and reasoning about profits and sales, show much lacking holes in ChatGPT's capabilities. That being said, the study finds that ChatGPT outperforms traditional sentiment prediction methods, especially for smaller stocks experiencing negative sentiment. The basic models lack accuracy in their predictions, falling far short of traditional methods. This paper shows even a general-purpose model can be remarkably effective for our purposes, begging further study and deeper looks into just how effective domain-specific models could be.

BloombergGPT: A Large Language Model for Finance

This paper discusses the creation and use of the BloombergGPT large language model trained on financial data. BloombergGPT is a 50 billion parameter large language model trained on various financial data. Self-described as a "mixedapproach" between general-purpose models and domain-specific models, Bloomberg's goal with their model was an effective model in both purposes, general and finance specific. BloombergGPT is trained on domain-specific data, but it is such a large dataset that it is able to function more generally without losing efficiency. Just over half of the training dataset for BloombergGPT is curated as opposed to scraped from the internet and public datasets. This helps it avoid pitfalls of unreliability and toxic data. This dataset includes web sources, press releases, Bloomberg's own published articles, and more. From Bloomberg's analysis against three comparable LLMs, BloombergGPT outperforms all of them both in finance-specific and general-purpose results.

IV. ANALYSIS PROCESS

Our analysis involves taking a collection of financial articles from finhub. The model then determines if the articles are more positive, negative or neutral on the company. Then it takes all of the articles for a company for a specific day and determines if it is positive, negative or neutral. This is repeated for every day for a given period of time. Then based on this information we can determine how positive or negative the news is for a specific company over a period of time and if the stock should be bought.

V. DATA COLLECTION

We collected data for 6 different stocks: Apple (AAPL), Microsoft (MSFT), Intel (INTL), Nvidia (NVDA), GE Aerospace (GE), and Hasbro (HAS). We analyzed these stocks over varying time periods to analyze a range of capabilities for our model. Microsoft, Intel and Nvidia were analyzed from August to December of 2023. GE Aerospace and Hasbro were analyzed from January to March of 2024. And finally, Apple was analyzed solely for the month of March 2024.

VI. Analysis

The outcomes we observed throughout the course of the project are somewhat vague in terms of their useability but nonetheless interesting. The two highest performers – Nvidia and Microsoft – were modeled relatively well. The model more often than not predicted that one should buy the two stocks and their positive/negative (P/N) ratio shows that both stocks did in fact have significantly more up days than down days. It should be noted that these two stocks were observed for the longest period of time. The next two stocks we analyzed were Intel and GE, neither of which were standout performers when compared to the previous two. Both had P/N ratios above one but by significantly less than Microsoft and Nvidia, the highest being Intel's at 2.91. Intel was observed for four months, GE however was observed for only three. Notably, Intel also saw the largest negative percentage change of all six stocks. The last two were Hasbro and Apple which were observed for three and one month respectively. Apple had particularly middling performance and saw metrics similar to Intel and GE as far as prediction average and P/N ratio are concerned. But Hasbro has the only negative average prediction from the model as well as a P/N ratio below one. This would indicate that not only were there more down days than up days, but the model was keen on predicting that one should sell Hasbro more often than one should purchase it. So, the model seems to be pretty good at indicating whether the day will be down or up (a buy or sell day) based on the given news but there is still room for improvement. It currently lacks more complex functionalities such as the ability to manage a currently held portfolio in any way. For example, if a user already owned a stock and wanted to know whether to hold or sell it the model is unable to indicate the correct choice. It could also be interesting to implement some way for the model to examine the best amount of a stock to buy given the stock's risk history to minimize the potential loss in case of an unexpected outcome. The model isn't a perfect tool; even so, it provides a strong foundation for a publicly available resource that could begin to democratize the stock market for an average user in a meaningful way.

References

[1] GfG. "What Is a Large Language Model (LLM)." GeeksforGeeks, GeeksforGeeks, 10 Jan. 2024, www.geeksforgeeks.org/large-language-model-llm/.

[2] Iriarte, M. (2023, June 29). Predicting the financial market with large language models. EnterpriseAI. https://www.enterpriseai.news/2023/06/29/predicting-the-financial-market-with-large-language-models/

[3] Lopez-Lira, A., & Tang, Y. (2023). Can chatgpt forecast stock price movements? return predictability and large language models. arXiv preprint arXiv:2304.07619.

[4] Naveed, H., Khan, A. U., Qiu, S., Saqib, M., Anwar, S., Usman, M., ... & Mian, A. (2023). A comprehensive overview of large language models. arXiv preprint arXiv:2307.06435.

[5] What Is a Large Language Model (LLM)?, www.cloudflare.com/learning/ai/what-is-large-language-model. Accessed 17 Mar. 2024.

[6] Wu, S., Irsoy, O., Lu, S., Dabravolski, V., Dredze, M., Gehrmann, S., ... & Mann, G. (2023). Bloomberggpt: A large language model for finance. arXiv preprint arXiv:2303.17564.