# ChatGPT Informed Graph Neural Network for Stock Movement Prediction

Zihan Chen
Stevens Institute of Technology
Hoboken, New Jersery, USA

Lei (Nico) Zheng
Stevens Institute of Technology
Hoboken, New Jersery, USA

Cheng Lu
Stevens Institute of Technology
Hoboken, New Jersery, USA

Jialu Yuan
University of California, Los Angeles
Los Angeles, California, USA

Di Zhu*
Stevens Institute of Technology
Hoboken, New Jersery, USA

## ABSTRACT

ChatGPT has demonstrated remarkable capabilities across various natural language processing (NLP) tasks. However, its potential for inferring dynamic network structures from temporal textual data, specifically financial news, remains an unexplored frontier. In this research, we introduce a novel framework that leverages ChatGPT's graph inference capabilities to enhance Graph Neural Networks (GNN). Our framework adeptly extracts evolving network structures from textual data, and incorporates these networks into graph neural networks for subsequent predictive tasks. The experimental results from stock movement forecasting indicate our model has consistently outperformed the state-of-the-art Deep Learning-based benchmarks. Furthermore, the portfolios constructed based on our model's outputs demonstrate higher annualized cumulative returns, alongside reduced volatility and maximum drawdown. This superior performance highlights the potential of ChatGPT for text-based network inferences and underscores its promising implications for the financial sector.

## CCS CONCEPTS

• **Computing methodologies → Natural language processing**; • **Applied computing → Economics**.

## KEYWORDS

large language models, graph neural networks, quantitative finance

**ACM Reference Format:**
Zihan Chen, Lei (Nico) Zheng, Cheng Lu, Jialu Yuan, and Di Zhu. 2023. ChatGPT Informed Graph Neural Network for Stock Movement Prediction. In *Proceedings of Workshop on Robust NLP for Finance (SIGKDD '23).* ACM, New York, NY, USA, 4 pages. https://doi.org/XXXXXXX.XXXXXXX

## 1 INTRODUCTION

The prediction of stock movements stands as one of the most intricate and elusive challenges in modern times, with accurate forecasts promising significant investment returns [4]. While stock prices are assumed to encapsulate all relevant market information [6], the process of distinguishing genuine signals from noise can critically impact forecasting efficacy. The academic community has responded to this challenge by formulating a wide array of statistical and machine learning models for forecasting purposes [5, 8, 10]. However, these approaches often overlook the underlying interdependencies among distinct equities, thus curtailing their predictive potential. Additionally, external events can influence relevant companies at different rates, leading to a phenomenon known as the "lead-lag effect" [12]. The existence of both interdependencies and lead-lag effects highlight the limitations of current methods in capturing dynamic relationships and modeling the evolution of markets.

Large Language Models (LLMs), such as ChatGPT, have demonstrated their prowess across a wide spectrum of Natural Language Processing (NLP) tasks [1]. However, their applications in the financial economics domain are still in a nascent stage. In the study, we present a novel approach that exploits ChatGPT to predict stock price movement. Our method commences with employing ChatGPT to identify and extract latent inter-dependencies among equities, thereby yielding an evolving graph that undergoes daily updates. Following this, a Graph Neural Network (GNN) is employed to generate embeddings for the target companies. These resultant embeddings are eventually integrated with a Long Short-Term Memory (LSTM) model to forecast stock movements for the upcoming trading day.

We evaluate the performance of our proposed model using a real-world dataset, setting the Dow Industrial Average 30 Companies (DOW 30) as primary targets. When applied to stock movement forecasting, our model outperforms all baseline models in terms of weighted F1, Micro F1, and Macro F1 metrics, registering a minimum improvement of 1.8%. Furthermore, we utilize the model's output to build portfolios based on both long-only and long-short strategies. The evaluation of portfolio performance reveals that our model consistently surpasses benchmarks regarding cumulative returns during the out-of-sample period, while also demonstrating lower annualized volatility and decreased maximum drawdown. These findings in both stock movement forecasting and portfolio performance evaluation underscore the effectiveness of our ChatGPT-informed GNN model, and highlight the potential of LLMs in processing financial data.

This paper presents two major contributions. First, to the best of our knowledge, this is the first study that investigates ChatGPT's capacity to infer network structures from textual data in the financial

*Corresponding author. Contact email: dzhu1@stevens.edu


*SIGKDD '23, August 7-8, 2023, Long Beach, California, USA*

ACM ISBN 978-1-4503-XXXX-X/18/06…$15.00
https://doi.org/XXXXXXX.XXXXXXX

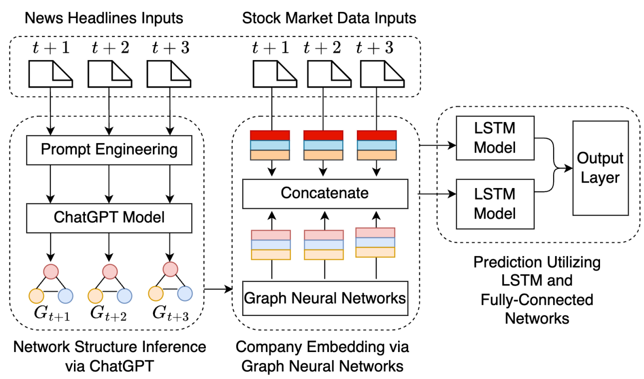


**Figure 1: Overview of the Proposed Model: Combining Graph Neural Network and ChatGPT to predict stock movements.**

economics area. While ChatGPT's impressive proficiency across diverse NLP tasks is well documented, our work distinguishes itself by pioneering the connection between time-series textual data and dynamic network structures. Second, the experiment provides compelling evidence of our model's superior performance in stock movement forecasting. Given that even minor improvements in stock prediction accuracy can yield substantial profit increments [3, 4], the enhanced performance of our model highlights its significant practical implications within the broader financial arena.

## 2 RELATED WORK

As the ability to predict stock movement plays a crucial role in shaping investment decisions, managing financial risks, and formulating trading strategies, scholars have employed a plethora of forecasting methods encompassing statistical approaches (e.g., MA, ARIMA, GARCH), machine learning techniques (e.g., KNN, SVM), as well as deep learning methods [5, 8]. Beyond modeling relationships with historical prices, researchers also integrate alternative data sources such as news articles, social media data, and financial reports, to enhance predictive capabilities [9]. However, these techniques often fall short in capturing the complex interdependencies among stocks, where the fluctuation of one stock can significantly influence the movements of other related stocks.

To capture the intricate interconnections among equities, recent researchers propose the use of GNN to better consolidate market information. GNN represents a novel branch of deep learning methods grounded in graph theory, in which companies act as nodes and relationships between companies are denoted as edges [14]. By disseminating information throughout the network, GNN empowers each company to take into account the information from its neighboring entities, thereby enhancing the learning and representation of market data. This advantage of GNN is corroborated by studies like [3] that demonstrates the superior performance of GNN in comparison to non-graph-based techniques.

Nonetheless, current GNN studies primarily construct these networks based on predefined graph structures, like financial knowledge graphs [7] or domain-specific knowledge [2]. These predefined knowledge graphs, constructed using manually-engineered features, often struggle to update in a timely manner and capture emergent relationships as the market continually evolves. In this study, we propose using LLMs, particularly ChatGPT, to respond to this limitation. While LLMs have shown remarkable capabilities across a broad array of NLP tasks [1], recent studies also illuminate their potential within the financial sector [11, 13]. Despite these advancements, there remains a lack of investigation into the effectiveness of ChatGPT's inference ability in identifying interdependencies among market equities. Addressing this research gap, our study leverages ChatGPT's capability to extract relationships among financial entities from textual inputs. Subsequently, we construct a GNN model using the graph structures inferred by ChatGPT to forecast stock movement.

## 3 METHOD

Our objective is to predict the stock movement (up, down, or neutral) for a set of target companies on the next trading day. Suppose we have a total of $N$ target companies, where $i$ denotes a specific company, $t$ represents a timestamp, and $L$ corresponds to the look-back length. Accordingly, our predictive task uses features from time $t$ to $t + L$ to forecast stock movement at time $t + L + 1$. To achieve this, we propose a novel framework that integrates ChatGPT and GNN for stock movement prediction. This framework consists of three main components: network structure inference from financial news using ChatGPT, company embedding through GNN, and stock movement prediction using sequential models and fully-connected neural networks. A comprehensive overview of our proposed framework is presented in Figure 1. We further elaborate on each component in the subsequent sections.

### 3.1 Network Structure Inference via ChatGPT

Our framework necessitates two types of time-series input features: news headlines and stock market data. The stock market data encompasses daily market information for each company. We use $\mathbf{S}_t$ to denote market data at time $t$, where $\mathbf{s}_{i,t}$ denotes the associated data of a specific company. On the other hand, news headlines are not company-specific and may cover various public companies. We thus exploit the inferential capabilities of ChatGPT to discern: 1) Which target companies could be affected by the day's news, and 2) How will these companies be affected: positively, negatively, or neutrally? To operationalize this, we design the following prompt for daily news headline input to ChatGPT:

Forget all your previous instructions. I want you to act as an experienced financial engineer. I will offer you financial news headlines in one day. Your task is to:

1. Identify which target companies will be impacted by these news headlines. Please list at least five of them.
2. Only consider companies from the target list.
3. Determine the sentiments of the affected companies: positive, negative, or neutral.
4. Only provide responses in JSON format, using the key "Affected Companies".
5. Example output: {"Affected Companies": {Company 1: "positive", Company 2: "negative"}}
6. News Headlines are separated by "\n"

The ChatGPT response provides two insightful elements: the companies being affected by the news and their corresponding sentiments. Because prior research has demonstrated a strong association between ChatGPT's sentiment on next day's stock return [11], we primarily focus on the "Affected Companies" output to construct a ChatGPT-Informed graph structure to feed GNN at the current stage. We build the graph $G_t = (V, E_t)$ at each timestamp by representing each target company as a node and building an edge between two companies if they were considered as "being affected together" by ChatGPT. For instance, if the "Affected Companies" output at $t$ is ['BA', 'AMGN', 'MSFT'], we construct edges $E_t$ among these ticker pairs: 'BA' – 'AMGN', 'BA' – 'MSFT', and 'AMGN' – 'MSFT'. After gleaning these inferred relationships from news using ChatGPT, we input these graphs sequentially into GNN to generate company embeddings.

## 3.2 Company Embedding through GNN

At this stage, we leverage GNN to transform the nodes (companies) into vector representations. As a cutting-edge deep learning model, GNN is adept at handling complex graph structures and embedding nodes into lower-dimensional vectors that encapsulate both nodes' attributes and network topology [14]. In our context, GNN integrates the features of a company and its closely interconnected companies at a given timestamp to generate embeddings. Consequently, each company's embedding through the GNN incorporates its unique features, as well as the features of relevant companies that are affected together by the news headline. Taking company $i$ and associated features at time $t$ as an example, we formally describe the GNN embeddings process as follows:

$$\mathbf{h}_{i,t}^{\text{GNN}} = \text{GNN}\left(\mathbf{h}_{i,t}; \mathbf{m}_{i,t}; \Theta_{GNN}\right) \tag{1}$$

where $\Theta_{GNN}$ symbolizes the trainable parameters in each layer of GNN, $\mathbf{h}_{i,t}$ denotes the original feature of company $i$, and $\mathbf{m}_{i,t}$ represents the aggregated information from its neighbors at time $t$. The final GNN embedding of company $i$ is denoted as $\mathbf{h}_{i,t}^{\text{GNN}}$.

## 3.3 Sequential Models and Output Layers

Retaining the information of the company and its neighbors, the output of the GNN is subsequently concatenated with the corresponding company's stock market data. We utilize a LSTM model as the sequential model in our framework. These combined data vectors are sequentially input into the LSTM, generating aggregated embeddings specific to each company over the lookback period. Concurrently, the stock market data undergoes a separate LSTM model to generate another set of embeddings. These two sets of embeddings are concatenated again and fed through a fully connected neural network layer to generate the final prediction for the stock movement. The process can be formalized as follows:

$$\mathbf{h}_{i,t}^{\text{COMB}} = \text{CONCAT}\left(\mathbf{h}_{i,t}^{\text{GNN}}, \mathbf{s}_{i,t}\right) \tag{2}$$

$$\mathbf{h}_{i}^{\text{COMB}} = \text{LSTM}\left(\left[\mathbf{h}_{i,t}^{\text{COMB}}, \cdots \mathbf{h}_{i,t+L}^{\text{COMB}}\right]; \Theta_{LSTM_1}\right) \tag{3}$$

$$\mathbf{h}_{i}^{\text{STOCK}} = \text{LSTM}\left(\left[\mathbf{s}_{i,t}, \cdots \mathbf{s}_{i,t+L}\right]; \Theta_{LSTM_2}\right) \tag{4}$$

$$\hat{y}_i = MLP\left(\text{CONCAT}\left(\mathbf{h}_{i}^{\text{COMB}}, \mathbf{h}_{i}^{\text{STOCK}}\right); \Theta_{MLP}\right) \tag{5}$$

where $\Theta_{MLP}$, $\Theta_{LSTM_1}$, and $\Theta_{LSTM_2}$ are the trainable parameters. Furthermore, given that we predict stock movement at $t+L+1$, this is a classification task with three categories: up, down, and neutral. Following previous literature [3], we generate the category for the ground truth based on the return ($R_i = p_{i,t}/p_{i,t-1} - 1$, where $p_i$ is the stock price) and defined thresholds ($r_{\text{up}} = 0.01$, $r_{\text{down}} = -0.01$). Specifically, the variable $y_i$ is labeled as "up" when $R_i \geq r_{\text{up}}$, "down" when $R_i \leq r_{\text{down}}$, and "neutral" otherwise. Finally, we employ cross entropy to generate the loss and backpropagate it for trainable parameters adjustment. In the following section, we assess our proposed model's performance on a real-world dataset.

# 4 EXPERIMENT

We evaluate the effectiveness of our proposed framework using a real-world dataset comprising the DOW 30 Companies as the main subjects. Since the DOW 30 composition was last updated on August 31, 2020, we opt for the period from September 1, 2020, to December 30, 2022, as our target interval. The training period extends from September 1, 2020, to September 30, 2021, consistent with the final data point integrated into ChatGPT's model training. To gather input features for both periods, we acquire daily variables of each DOW 30 company from the CRSP Databases as stock market data. For the financial news headlines, we collect 2,713,233 and 3,717,666 unique headlines for the training and test periods respectively.

In recognition of the temporal sensitivity of news and its lag effect on the stock market, we meticulously align the news timestamp with the subsequent market period. For instance, a news headline recorded before 16:00 on Day $t$ is linked with the same day's market data, and employed to predict stock movements on Day $t+1$. Conversely, if a headline is logged after 16:00, it is assigned to the succeeding day (Day $t+1$) and used to forecast stock movement on the following day (Day $t+2$). This stratagem ensures the purity of out-of-sample test results and precludes potential data leakage.

Our benchmark selection is rooted in the two types of input features we utilize. We deploy LSTM and ARIMA model for stock market data, and state-of-the-art sentence transformers (News-Embed model) to leverage the financial news headlines. Furthermore, corroborating previous research that affirms the predictive power of ChatGPT's direct outputs [11], we incorporate the sentiment judgment from ChatGPT on stocks as a benchmark.

**Table 1: Model Performance of Stock Movement Prediction**

| Model | Weighted F1 | Micro F1 | Macro F1 |
|---|---|---|---|
| ChatGPT | 0.3970 | 0.4607 | 0.3085 |
| News-Embed | 0.4059 | 0.4318 | 0.3425 |
| Stock-LSTM | 0.4036 | 0.4132 | 0.3455 |
| **Our Model** | 0.4133 | 0.4423 | 0.3529 |

We assess our proposed model on two tasks: First, we scrutinize its performance in the classification of stock movement predcition. The evaluation metrics included weighted F1, Macro F1, and Micro F1 scores. Second, we construct portfolios based on the model outputs, and evaluate their performance in terms of accumulated return, volatility, and maximum drawdown. Detailed results from these experiments will be elucidated in the subsequent section.

# 5 RESULTS

The experimental results of stock movement forecasting are presented in Table 1, with two primary observations. First, our proposed model persistently outperforms both the stock-LSTM and News-Embed models in all three metrics, recording a minimum enhancement of 1.8%. Notably, our model distinguishes itself from stock-LSTM by employing dynamic graph structures that ChatGPT generates from daily financial news. This suggests the potency of ChatGPT's zero-shot learning capability in inferring networks from text. Since earlier research has demonstrated a 0.005 increase in the Micro F1 score can result in a profit increase of 12% [3, 4], our model offers considerable practical implications within the financial field.

Second, though past studies have emphasized the strong correlation between the sentiment outputs from ChatGPT and stock movements [11], our findings indicate that amalgamating these outputs with GNN amplifies performance. Despite ChatGPT delivering commendable Micro F1 scores, this is largely due to an inherent data imbalance during the testing phase, as the 58.5% of stock movements were neutral. ChatGPT's predictive prowess falters when forecasting stock downtrends, with a score of 10.88%, compared to our model's 19.46% in this category. This pattern echoes in time-series models like ARIMA, which predominantly predict all movements as neutral. The enhanced ability of our model to forecast both upward and downward movements is instrumental in aiding investors to limit losses and maintain portfolio stability.

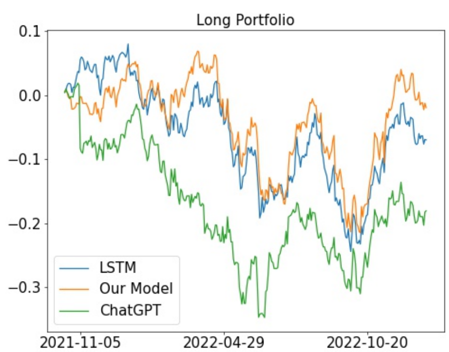


**Figure 2: Portfolio Performance During the Test Period**

We also evaluate the economic implications of our model by constructing portfolios grounded in the model's outputs. Given that each stock's predicted outcome is either upward, neutral, or downward for the next trading day, we construct portfolios with both long-only strategy and long-short strategy. We then conduct the backtest on these portfolios during the out-of-sample period (October 1, 2021, to December 30, 2022). The cumulative returns for the Long Portfolio are depicted in Figure 2. As seen from the figure, our proposed model consistently outperforms both the LSTM and ChatGPT model in terms of cumulative returns. This persistent superiority signifies the effectiveness of our model in predicting positive stock returns. Moreover, the portfolio derived from ChatGPT outputs exhibits significantly higher annualized volatility (23.61%) compared to our model (14.06%). The maximum drawdown of the ChatGPT model (0.2112) also substantially exceeds that of ours (0.1242). As previously noted, this discrepancy is primarily due to ChatGPT's limitations in predicting negative returns, thereby rendering it prone to higher volatility.

# 6 DISCUSSION

This study introduces a novel framework that leverages the graph inference capabilities of ChatGPT to enhance GNN forecasting performance. Our experimental results using data from DOW 30 companies underscore the superior performance of our model over all benchmarks in predicting stock movements. Furthermore, portfolios constructed based on our model's outputs yield higher cumulative returns while simultaneously exhibiting reduced volatility and drawdowns.

Despite this is the first study that integrates ChatGPT-inferred networks with GNNs, our paper is not without its limitations. First, our model only includes stock market data and news headlines as input features. Considering stock market are influenced by a complex web of interconnected factors, augmenting our model with additional input features could further boost its predictive accuracy. Second, we only utilize basic network structures in the model, which could be upgraded to more sophisticated architectures. Finally, the recent evolution of ChatGPT, with the incorporation of browsing ability and plugins, enables it to interact with the latest information. As such, enriching our model with the most current financial news and market data may further enhance its forecasting performance.

# REFERENCES

[1] Yejin Bang, Samuel Cahyawijaya, Nayeon Lee, Wenliang Dai, Dan Su, Bryan Wilie, Holy Lovenia, Ziwei Ji, Tiezheng Yu, Willy Chung, et al. 2023. A multitask, multilingual, multimodal evaluation of chatgpt on reasoning, hallucination, and interactivity. *arXiv preprint arXiv:2302.04023* (2023).

[2] Zhe Chen, Yuehan Wang, Bin Zhao, Jing Cheng, Xin Zhao, and Zongtao Duan. 2020. Knowledge graph completion: A review. *Ieee Access* 8 (2020), 192435–192456.

[3] Dawei Cheng, Fangzhou Yang, Sheng Xiang, and Jin Liu. 2022. Financial time series forecasting with multi-modality graph neural network. *Pattern Recognition* 121 (2022), 108218.

[4] Marcos Lopez De Prado. 2018. *Advances in financial machine learning*. John Wiley & Sons.

[5] Xiao Ding, Yue Zhang, Ting Liu, and Junwen Duan. 2015. Deep learning for event-driven stock prediction. In *Twenty-fourth international joint conference on artificial intelligence*.

[6] Eugene F Fama. 1965. The behavior of stock-market prices. *The journal of Business* 38, 1 (1965), 34–105.

[7] Thomas Fischer and Christopher Krauss. 2018. Deep learning with long short-term memory networks for financial market predictions. *European journal of operational research* 270, 2 (2018), 654–669.

[8] Hakan Gunduz, Yusuf Yaslan, and Zehra Cataltepe. 2017. Intraday prediction of Borsa Istanbul using convolutional neural networks and feature correlations. *Knowledge-Based Systems* 137 (2017), 138–148.

[9] Michael Hagenau, Michael Liebmann, and Dirk Neumann. 2013. Automated news reading: Stock price prediction based on financial news using context-capturing features. *Decision support systems* 55, 3 (2013), 685–697.

[10] Kyoung-jae Kim and Ingoo Han. 2000. Genetic algorithms approach to feature discretization in artificial neural networks for the prediction of stock price index. *Expert systems with Applications* 19, 2 (2000), 125–132.

[11] Alejandro Lopez-Lira and Yuehua Tang. 2023. Can ChatGPT Forecast Stock Price Movements. *Return Predictability and Large Language Models* (2023).

[12] Matthew L O'Connor. 1999. The cross-sectional relationship between trading costs and lead/lag effects in stock & option markets. *Financial Review* 34, 4 (1999), 95–117.

[13] Kai-Cheng Yang and Filippo Menczer. 2023. Large language models can rate news outlet credibility. *arXiv preprint arXiv:2304.00228* (2023).

[14] Si Zhang, Hanghang Tong, Jiejun Xu, and Ross Maciejewski. 2019. Graph convolutional networks: a comprehensive review. *Computational Social Networks* 6, 1 (2019), 1–23.