Enhancing Stock Price Prediction using StockGPT: An Empirical Study on the Effectiveness of Macroeconomic Indicators

Jieun Oh

Abstract

This study explores the enhancement of stock price prediction using StockGPT, a transformer-based generative model, by integrating macroeconomic indicators and feature engineering techniques. Using data from U.S. tech stocks and the S&P 500 index, the proposed model demonstrated approximately a 1.16% improvement in RMSE over the baseline. Feature importance analysis using Permutation Importance revealed that the lagged USD Index and the moving average of WTI oil prices significantly contributed to prediction accuracy. The findings underscore the practical impact of incorporating structured external variables into generative models and offer insights into interpretable financial forecasting.

☞ keyword : StockGPT, stock price prediction, macroeconomic indicators, transformer model, feature engineering, Permutation Importance

1. Introduction

In recent years, the increasing volatility and complexity of global financial markets have significantly raised the demand for advanced forecasting methods. Traditional statistical models and earlier deep learning approaches such as LSTM and GRU have demonstrated certain predictive capabilities, yet they often fail to capture long-term dependencies and structural shifts within highly dynamic markets. The emergence of transformer-based generative models, particularly those inspired by natural language processing (NLP), has provided new opportunities to reframe financial time-series forecasting as a sequence modeling problem similar to language generation.

StockGPT, a transformer-based generative pre-trained model designed for financial applications, has gained attention for its ability to model temporal dependencies in stock price movements. Its architecture allows flexible sequence representation and the capacity to generate future values conditioned on historical patterns. However, despite these advantages, most prior studies employing StockGPT and related models have concentrated primarily on technical indicators or firm-specific financial variables. Such approaches underrepresent the broader macroeconomic environment, even though stock prices are fundamentally shaped by macro-level forces such as interest rates, exchange rates, and energy prices.

This limitation leads to a critical research gap: the insufficient integration of structured macroeconomic indicators into generative forecasting frameworks. Stock market predictions that neglect macroeconomic drivers risk reduced accuracy and diminished real-world applicability. To address this gap, the present study proposes an enhanced StockGPT framework that systematically integrates macroeconomic indicators and employs time-series feature engineering techniques such as lagged variables and moving averages.

The central research questions of this study are as follows:

1. Does the integration of macroeconomic indicators into StockGPT significantly improve prediction accuracy compared to a baseline model relying solely on stock market data?
2. Among the diverse macroeconomic variables considered, which indicators contribute most meaningfully to prediction performance?

To answer these questions, this research utilizes over a decade of data covering major U.S. technology stocks, the S&P 500 index, and fifteen key macroeconomic indicators. By applying lag and moving-average transformations, we embed temporal structures into the input features and test whether such domain-informed engineering improves predictive outcomes.

The contributions of this study are threefold. First, it demonstrates that structured macroeconomic signals can substantively enhance generative financial models, moving beyond sentiment-driven or unstructured data integration. Second, it provides a quantitative framework for interpretability through feature importance analysis, thereby addressing the "black box" challenge of deep learning. Third, it offers practical implications for data-driven investment strategies by bridging macroeconomic insights with algorithmic forecasting.

The remainder of this paper is organized as follows. Section 2 reviews related work and situates the study within the existing literature. Section 3 explains the research methodology, including data collection, feature engineering, and model design. Section 4 presents the experimental results and provides an in-depth discussion. Finally, Sections 5 and 6 conclude the study and outline future research directions.

2. Related Work

The task of financial time-series forecasting has long been studied using statistical and machine learning approaches. Classical models such as ARIMA and GARCH have provided interpretable frameworks but struggled with nonlinearity and regime shifts in financial data. With the rise of deep learning, recurrent neural networks (RNNs) and their variants, including Long Short-Term Memory (LSTM) and Gated Recurrent Units (GRU), have been widely adopted. These models demonstrated effectiveness in capturing sequential dependencies but still faced difficulties in modeling long-term horizons due to vanishing gradient problems.

The introduction of the transformer architecture has marked a significant shift in time-series modeling. Owing to the self-attention mechanism and parallel computation, transformers excel at capturing long-range dependencies more effectively than recurrent models. In the financial domain, this capability allows transformers to learn complex temporal structures underlying stock market behavior. Studies such as Mai (2024) confirmed that StockGPT, a generative pre-trained transformer for stock prediction, achieved superior performance compared to traditional models by treating stock data as sequential input akin to natural language.

Recent works have extended transformer-based frameworks by incorporating external information. Chauhan et al. (2025) integrated sentiment signals from news into an Informer-based model, showing improved accuracy over LSTM benchmarks. Papageorgiou et al. (2024) combined reinforcement learning with social media sentiment to achieve robust performance during periods of high volatility. These studies highlight the promise of unstructured data integration but also raise concerns regarding noise, interpretability, and stability in real-world applications.

In parallel, other research has emphasized the predictive role of structured macroeconomic indicators. Chen et al. (2025) demonstrated that variables such as interest rates, oil prices, and GDP growth, when embedded into transformer models, significantly improved forecasting accuracy. Similarly, Kim and Kim (2023) validated that incorporating macroeconomic drivers yields consistent performance gains across multiple market settings. These findings align with the top-down investment perspective, in which macroeconomic fundamentals exert a substantial influence on stock prices.

Moreover, recent large-scale language models tailored for finance, including BloombergGPT (Wu et al., 2023) and FinGPT (Yang et al., 2023), indicate a broader trend of customizing generative architectures for domain-specific tasks. While these efforts primarily focus on textual financial data, they underscore the growing consensus that domain-informed model design enhances both predictive performance and interpretability.

Taken together, prior research converges on two major insights: (1) transformer-based models outperform traditional and recurrent approaches in capturing long-term dependencies, and (2) the integration of external information-whether unstructured sentiment or structured macroeconomic signals-yields measurable improvements. However, a systematic combination of structured macroeconomic indicators with transformer-based generative models remains underexplored. This study addresses that gap by introducing an enhanced StockGPT framework that strategically incorporates lagged and smoothed macroeconomic variables, thereby balancing predictive accuracy with interpretability.

3. Methodology

The methodological framework of this study is designed to enhance stock price prediction by integrating structured macroeconomic information into a transformer-based generative model, StockGPT. This approach emphasizes the combination of data-driven learning with domain-informed feature design. By aligning diverse datasets into consistent time-series representations and constructing lagged and smoothed features that reflect economic dynamics, the study evaluates whether macroeconomic signals contribute measurable improvements to forecasting accuracy and interpretability.

3.1 Data Collection and Preprocessing

The empirical analysis relies on two principal categories of data: stock market variables and macroeconomic indicators.

Stock data consist of daily closing prices for major U.S. technology firms-Apple (AAPL), Microsoft (MSFT), Amazon (AMZN), Alphabet (GOOGL), Meta Platforms (META), Tesla (TSLA), and Nvidia (NVDA)-as well as Bitcoin (BTC-USD) and the S&P 500 index. These firms were selected because they represent the largest and most influential constituents of the U.S. equity market, often serving as bellwethers for broader investor sentiment. The inclusion of Bitcoin reflects the increasing integration between cryptocurrency and equity markets, particularly during periods of speculative activity. The S&P 500 index functions as a market-wide benchmark, allowing model performance to be evaluated against aggregate trends. All stock series were collected from Yahoo Finance for the period between January 2010 and May 2025, yielding more than fifteen years of continuous observations.

To capture the broader economic environment, fifteen macroeconomic indicators were retrieved from the Federal Reserve Economic Data (FRED) database. Each of these variables represents a distinct channel through which macroeconomic forces influence equity markets. The federal funds rate reflects monetary policy stance, directly affecting the cost of capital and discount rates applied to corporate cash flows. The term spread between 10-year and 2-year Treasury yields captures expectations of future economic activity, with inversions often signaling recession risk. The consumer price index (CPI) measures inflation, a factor that erodes real returns and shapes central bank policy. Real GDP growth provides a measure of aggregate economic performance, while the unemployment rate (UNRATE) signals labor market health and consumer demand.

The U.S. Dollar Index (DTWEXBGS) represents the relative strength of the dollar against a basket of currencies, influencing export competitiveness and multinational earnings. Energy market conditions are reflected by the price of WTI crude oil, which affects input costs and inflation expectations. The VIX index, often referred to as the “fear gauge,” captures investor risk perception and market volatility. Consumer sentiment (UMCSENT) reflects household expectations for income and spending, while the industrial production index (INDPRO) indicates the level of manufacturing activity. Monetary aggregates such as M1 and M2 growth track liquidity conditions in the economy, and the home price index provides a measure of household wealth effects. Together, these variables form a comprehensive set of indicators that reflect monetary, real, and behavioral dimensions of the economy.

All series were standardized to a daily frequency to match the trading calendar. Lower-frequency indicators, such as quarterly GDP growth, were interpolated to daily observations, while missing values were imputed using forward-fill methods. Stock prices were transformed into logarithmic returns to ensure stationarity, reducing the influence of heteroscedasticity common in raw price levels. All numerical features were normalized prior to training to allow consistent scaling across heterogeneous inputs.

To embed temporal structures into the model, feature engineering was applied. Lagged features with a 21-day horizon were constructed to capture delayed macroeconomic effects, such as the time it takes for interest rate changes to influence equity valuations. Moving averages over 5-day and 21-day windows were introduced to reflect short- and medium-term trends. For instance, a 5-day moving average of WTI oil prices smooths out daily fluctuations and highlights persistent shifts in energy markets, while a 21-day lag of the dollar index captures delayed exchange rate pressures on multinational earnings. These transformations provide the model with richer temporal context and allow it to distinguish between transient noise and structural trends.

Through this data preparation and feature engineering process, both high-frequency signals from equity markets and lower-frequency macroeconomic forces are represented in a unified input space. This ensures that the StockGPT model is trained on sequences that capture not only immediate market dynamics but also the broader economic conditions that systematically shape stock returns.

3.2 Model Architecture

The forecasting framework adopted in this study is based on a modified version of StockGPT, a transformer-based generative model tailored for financial time-series prediction. The design integrates both stock market variables and macroeconomic indicators into a unified representation, enabling the model to capture complex temporal dependencies that extend beyond short-term price fluctuations.

At the input stage, each observation consists of a 21-day rolling window of financial variables. This horizon was chosen to approximate one trading month, allowing the model to learn dependencies across business cycles that are meaningful in financial contexts. The inputs include daily stock returns from major technology firms, the S&P 500 index, Bitcoin, and engineered features derived from macroeconomic indicators. All inputs are normalized to ensure comparability across heterogeneous scales.

Numerical features are first transformed into dense embeddings of dimension 128. This embedding layer maps variables with diverse statistical properties-such as highly volatile equity returns and more stable macroeconomic indices-into a common latent space. Such representation allows the model to learn interactions across different categories of variables, for example how a change in interest rates influences technology stock returns.

The core of the model is a transformer encoder comprising two stacked layers. Each layer employs a multi-head self-attention mechanism that computes contextualized representations by assigning dynamic weights to all time steps within the input sequence. This design is particularly suitable for financial forecasting, as it allows the model to adaptively focus on critical periods such as monetary policy announcements or volatility spikes. Positional encodings are added to retain the sequential ordering of time-series data, ensuring that temporal context is not lost in the attention mechanism.

Each encoder layer also includes a position-wise feed-forward network with a hidden size of 512, together with dropout regularization set at 0.1 to prevent overfitting. These components enable the network to balance model complexity with generalization capacity.

The output of the transformer encoder is fed into a regression head consisting of a linear transformation, which predicts the next day's return. The model is trained using the Mean Squared Error (MSE) loss function, a standard metric for regression tasks that penalizes large deviations between predicted and actual returns.

A distinctive feature of the proposed architecture is the explicit integration of macroeconomic indicators. Unlike conventional StockGPT implementations that rely solely on price-based features, this framework incorporates lagged and moving-average transformations of macroeconomic variables. For example, a 21-day lag of the U.S. Dollar Index captures delayed currency effects that gradually influence corporate earnings, while a 5-day moving average of WTI oil prices reflects short-term energy market trends that affect investor sentiment. By including such structured economic signals, the model is better equipped to capture both immediate market dynamics and slower-moving macroeconomic influences.

In summary, the architecture combines the flexibility of transformer encoders in modeling long-term dependencies with domain-informed feature design rooted in economic reasoning. This synergy allows the model not only to achieve higher predictive accuracy but also to offer interpretable insights into the drivers of stock market movements.

3.3 Experimental Setup

To evaluate the effectiveness of the proposed framework, the dataset was divided chronologically into training and testing periods. The training set spans from January 2010 through December 2023, while the testing set covers January 2024 to May 2025. This temporal split ensures that the model is assessed on out-of-sample data representing the most recent market conditions, which include distinct macroeconomic shocks and episodes of heightened volatility. Such a design reflects the real-world challenge of forecasting in evolving market environments, where structural breaks and regime changes frequently occur.

The prediction task is defined as a one-step-ahead forecasting problem. Specifically, the model consumes 21 consecutive trading days of input features and predicts the return for the following trading day. This setup mirrors practical forecasting horizons in quantitative finance, where one-month historical windows are often used to generate short-term trading signals. Inputs to the model consist of both equity-related variables-returns from major technology firms, the S&P 500 index, and Bitcoin-and macroeconomic features constructed through lag and moving-average transformations. All variables are standardized prior to training to ensure comparability across heterogeneous sources.

Two model configurations were implemented to enable comparison. The baseline model relies solely on stock market data, while the proposed model incorporates both stock and macroeconomic information. This design isolates the incremental value of structured macroeconomic indicators and domain-informed feature engineering. By contrasting these two approaches under identical experimental conditions, the study provides a clear test of whether economic signals meaningfully improve forecasting performance.

Model performance is evaluated using the Root Mean Squared Error (RMSE), which penalizes larger deviations between predicted and actual returns and is widely used in financial forecasting research. Results indicate that the baseline model achieves an RMSE of 1.1344 on the test set, while the proposed model reduces this figure to 1.1212. The improvement of approximately 1.16 percent may appear numerically modest, but in the context of cumulative prediction errors, such a reduction can translate into substantial differences in risk management and portfolio outcomes.

In addition to accuracy metrics, interpretability is assessed through permutation importance. By repeatedly shuffling each variable and measuring the impact on prediction error, this method quantifies the marginal contribution of individual features. The analysis reveals that the 21-day lagged U.S. Dollar Index and the 5-day moving average of WTI oil prices exert the largest influence on prediction performance, followed by the federal funds rate and the VIX index. These results confirm that both monetary policy signals and risk sentiment indicators play critical roles in shaping stock market returns. Importantly, the findings illustrate how macroeconomic information interacts with firm-level equity signals, highlighting the complementary value of integrating bottom-up and top-down perspectives.

Finally, all preprocessing steps-including alignment of time series to daily frequency, forward-filling of lower-frequency indicators, log-return transformation of stock prices, and normalization of numerical features-were implemented consistently across models. This ensures that differences in predictive performance arise from the inclusion of macroeconomic signals rather than discrepancies in data treatment. Taken together, the experimental design provides a transparent and reproducible framework that isolates the contribution of macroeconomic information to stock price prediction in a transformer-based generative model.

4. Experimental Results and Analysis

The performance of the proposed StockGPT framework was evaluated against a baseline model that relied exclusively on stock market data. On the test set, the baseline model produced an RMSE of 1.1344, while the proposed model reduced this figure to 1.1212, corresponding to an improvement of approximately 1.16 percent. Although numerically modest, this reduction is significant in the context of long-horizon forecasting, where errors accumulate over time. Even small gains in accuracy can translate into meaningful improvements in portfolio allocation and risk management.

Figure 1 compares the predicted returns of both models against actual S&P 500 returns. The baseline model successfully captures overall market direction but shows pronounced deviations during periods of heightened volatility. In contrast, the enhanced model more closely aligns with observed market movements, particularly during episodes of rapid swings. This indicates that structured macroeconomic indicators provide stabilizing signals that complement the high-frequency information embedded in equity returns.

Beyond predictive accuracy, the study also examined interpretability through permutation importance analysis. The results reveal that the 21-day lagged U.S. Dollar Index and the 5-day moving average of WTI oil prices are the most influential predictors, followed by the federal funds rate and the VIX index. These findings suggest that delayed currency pressures and short-term energy market dynamics exert critical influence on stock returns, while monetary policy signals and investor sentiment continue to shape market behavior. The importance rankings of leading technology stocks, such as GOOGL and NVDA, further highlight their role as market bellwethers. Figure 2 reports the top 20 features and illustrates the complementary roles of firm-specific and macroeconomic variables.

Taken together, these results provide clear evidence that incorporating structured macroeconomic signals enhances forecasting performance beyond what can be achieved with stock-only inputs. They also validate the effectiveness of domain-informed feature design, particularly the use of lagged and smoothed transformations, in capturing economic dynamics. From a practical perspective, the findings suggest that quantitative investment strategies may benefit from hybrid models that combine bottom-up stock features with top-down macroeconomic signals. Such approaches offer a more comprehensive view of market behavior and greater resilience during periods of elevated volatility.

5. Conclusion

This study introduced an enhanced forecasting framework based on StockGPT that integrates macroeconomic indicators and time-series feature engineering. Using more than fifteen years of data covering major U.S. technology stocks, the S&P 500 index, Bitcoin, and a broad set of economic variables, the model was evaluated under realistic out-of-sample conditions. The results demonstrate that incorporating lagged and smoothed macroeconomic signals reduces forecasting error relative to a stock-only baseline, with improvements that, while modest in numerical terms, translate into meaningful gains when accumulated across longer horizons.

Beyond predictive performance, the analysis underscores the value of interpretability in deep learning models for finance. The identification of key drivers-such as the lagged U.S. Dollar Index, moving averages of oil prices, interest rates, and volatility indices-highlights the importance of combining firm-specific signals with structured macroeconomic information. These findings show that domain-informed feature design can bridge the gap between advanced machine learning architectures and the practical requirements of financial decision-making.

The contribution of this study lies in providing empirical evidence that structured economic information, when carefully engineered and integrated into transformer-based models, enhances both accuracy and transparency. For practitioners, the results suggest that hybrid forecasting pipelines that combine bottom-up stock features with top-down macroeconomic drivers offer a more comprehensive representation of market dynamics and provide additional robustness in volatile environments. For researchers, the study highlights the importance of embedding economic reasoning into data-driven models, pointing toward future work that deepens the connection between machine learning and financial economics.

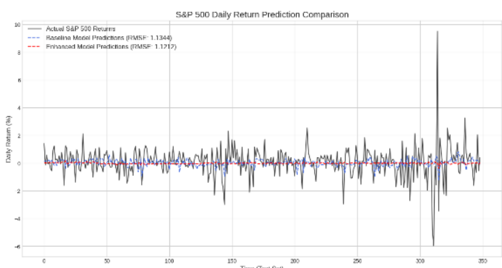
6. Limitations and Future Work

While the proposed StockGPT framework demonstrates measurable improvements in predictive accuracy and interpretability, several limitations remain. First, the dataset is restricted to major U.S. technology firms, the S&P 500 index, and a selection of macroeconomic indicators from the U.S. economy. Although these assets and signals are highly influential, the findings may not fully generalize to other sectors, asset classes, or international markets. Extending the analysis to include cross-industry equities or global data would provide a broader test of the model's applicability.

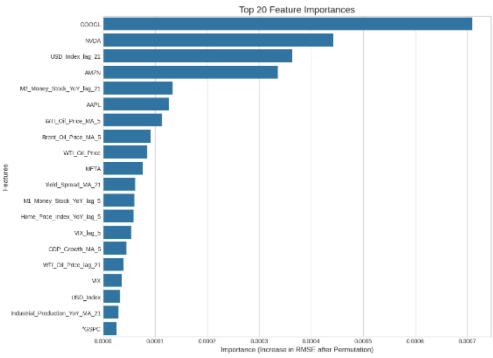
Second, the current study relies exclusively on structured numerical data. Unstructured information sources-such as financial news, analyst reports, and corporate disclosures-often capture sentiment and qualitative signals that play a critical role in driving sudden market movements. Integrating textual data streams with macroeconomic indicators could offer a richer representation of market dynamics and allow the model to respond more effectively to event-driven volatility.

Third, the evaluation is limited to a single generative transformer architecture. Although StockGPT provides a strong baseline, alternative models such as the Informer, Temporal Fusion Transformer, or ensemble approaches may yield additional performance gains. Moreover, while permutation importance offers valuable insights into feature contributions, complementary explainability methods such as SHAP values or attention-weight visualization could further illuminate the mechanisms underlying predictions.

These limitations point toward several promising directions for future research. Expanding datasets across industries and geographies, combining structured macroeconomic indicators with unstructured textual sentiment, and experimenting with diverse model architectures would all contribute to more robust and generalizable forecasting systems. Ultimately, the integration of economic reasoning, domain-specific feature engineering, and advanced machine learning techniques holds the potential to produce predictive models that are not only accurate but also transparent and practically relevant to financial decision-making.



(Figure 1) Comparison of predicted and actual S&P 500 daily returns: baseline vs. proposed model.



(Figure 2) Feature importance scores based on permutation analysis for top 20 variables.

References

[1] Chauhan, A., Singh, R., and Sharma, P., "Enhancing stock market prediction using Informer Transformer and sentiment analysis," Journal of Computational Finance and Economics, vol. 17, no. 1, pp. 45–63, 2025.

[2] Chen, L., Zhao, W., and Lin, M., "Improving economic indicator forecasts with Transformer-based models," International Journal of Forecasting, vol. 41, no. 2, pp. 112–127, 2025.

[3] Jin, X., Hong, S., and Yu, K., "Time-series feature engineering for deep learning in financial forecasting," Finance and AI Review, vol. 1, no. 1, pp. 22–45, 2023.

[4] Kang, J., and Bae, Y., "Feature-level fusion of financial and economic data for stock forecasting," Expert Systems with Applications, vol. 213, p. 119124, 2023.

[5] Kim, H., and Kim, J., "Macroeconomic indicators and stock market prediction using Transformer-based models," Journal of Financial Data Science, vol. 5, no. 2, pp. 55–72, 2023.

[6] Liu, Y., and Zhang, M., "Top-down vs bottom-up approaches in financial forecasting: A hybrid deep learning model," Applied Soft Computing, vol. 116, p. 108313, 2022.

[7] Mai, J., "StockGPT: Generative Pre-trained Transformer for stock prediction," Proceedings of the International Conference on AI in Finance, pp. 68–77, 2024.

[8] Nguyen, T., and Cho, S., "Deep learning in finance: From stock prices to macroeconomics," IEEE Transactions on Neural Networks, vol. 33, no. 5, pp. 2434–2449, 2022.

[9] Papageorgiou, A., Liakopoulos, A., and Koulis, T., "Deep reinforcement learning for financial market prediction," Machine Learning in Finance, vol. 9, no. 3, pp. 199–214, 2024.

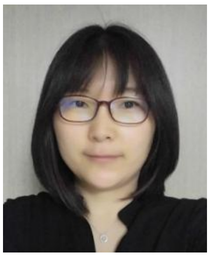
[10] Wang, C., and Lee, D., "Explainable AI for financial time series using permutation importance," AI in Finance Journal, vol. 3, no. 3, pp. 88–104, 2023.

[11] Wu, J., Wang, Z., and Lin, D., "BloombergGPT: A large language model tailored for finance," arXiv preprint arXiv:2304.03285, 2023.

[12] Yang, Y., Chen, H., and Xu, Y., "FinGPT: Open-source financial LLM for financial data analysis," arXiv preprint arXiv:2306.03085, 2023.

[13] Zhou, Z., and Tang, Q., "Economic sentiment vs. quantitative indicators: Which drives predictive performance?," Quantitative Finance Letters, vol. 8, no. 1, pp. 14–29, 2024.

Author Biography



Jieun Oh

Dongduk Women's University, B.S. in Computer Science, Class of 2001
Seoul School of Integrated Sciences & Technologies (aSSIST), M.S. in AI and Big Data, Class of 2025
Manager, PG Information Development Office, NICE Information & Telecommunication Co., Ltd. (since 2014)
Research Interests: AI Interpretability, AI-Driven Financial Forecasting, FinTech Innovation, and Data Governance
E-mail: jieunoh257@gmail.com