

Financial Forecasting in Consumer Cyclical with Economic Indicators and Tokenization

Kasper Krawczyk^{1,*}, Gary K. L. Tam¹, and Daniel Archambault²

¹School of Mathematics and Computer Science, Swansea University, Swansea, United Kingdom

²School of Computing, Newcastle University, Newcastle, United Kingdom

Email: kasper.krawczyk@gmail.com (K.K.); k.l.tam@swansea.ac.uk (G.K.L.T.); daniel.archambault@newcastle.ac.uk (D.A.)

*Corresponding author

Manuscript received September 25, 2024; revised November 5, 2024; accepted May 16, 2025; published October 23, 2025

Abstract—Machine learning in financial Time Series Forecasting (TSF) has a clear application in investment, where predicting stock price movements can inform investment strategies. The Transformer model has emerged as a powerful tool for this purpose, yet significant research gaps remain. Existing studies often focus on a set of stocks without modeling stock behavior within specific sectors. Notably, there is a lack of research on the Consumer Cyclical (CC) industry, which includes sectors such as automotive, housing, entertainment, and retail. These industries are highly sensitive to economic conditions, making them crucial for understanding broader economic impacts on stock behavior. Furthermore, we observe that many existing works neglect broader economic contexts, which is particularly important for CC analysis due to its sensitivity to economic trends. Additionally, previous studies on financial Transformer models typically use model tokens as feature vectors of multiple variates at a single timestep. This approach may not adequately capture the important relationships between dataset variates for long-term economic trends. To this end, we present the first study on financial TSF for the CC sector along with economic data. To support the research, we propose the first public benchmark dataset for the CC sector, consisting of traditional stock price time series, technical indicators, and temporal data, enriched with economic indicators. Next, we introduce an alternative tokenization approach to enhance the model's ability to capture long-term trends by preserving information about nonlinear dependencies between dataset variates. We hypothesize that this approach helps capture long-term signals more effectively. Through a comprehensive data ablation study and benchmark testing, we demonstrate that incorporating economic indicators improves the accuracy of longer-term predictions for the CC sector, and the new tokenization method enhances the performance of Transformer models. The dataset and code are made publicly available at: https://github.com/KasperKrawczyk/econtrans_dataset

Keywords—time series, forecasting, transformer, finance, stocks

I. INTRODUCTION

Accurately predicting stock market behavior offers clear benefits for maximizing returns [1]. For nearly a century, the finance industry has used statistical methods for this purpose [2]. Financial forecasting, especially Time Series Forecasting (TSF), has grown to manage risk, maximize profits, and minimize losses [3]. Consequently, stock market movement forecasting is an important research focus due to its potential to boost profitability for both institutional and private traders [4–7].

The research on financial forecasting, using both traditional statistical techniques and machine learning, is extensive and rapidly expanding. The earliest statistical

model for meteorological forecasting [2, 8] evolved into the Autoregressive Moving Average (ARMA) and Generalized Autoregressive Conditional Heteroskedasticity (GARCH) models [9]. Advances in computer hardware later enabled machine learning to dominate the field [10], with techniques like Random Forest [11], Long Short-Term Memory (LSTM) [12, 13], and Bi-directional Long Short-Term Memory (BiLSTM) networks [14]. Recently, the Transformer architecture [15] has gained attention, with efforts to enhance its forecasting abilities using specialized layers [16], recurrent network layers [17], and graph networks [18].

Despite recent progress, we observe research gaps and limitations in the literature. First, most research focuses on generalizing predictions to exchange-based indices (aggregating entire stock markets) [16, 19, 20] and general stock classification problems [21–23]. The issue of time series datasets limited to price values of one or a few stocks [23–26], or exchange-based indices [16, 20, 21], is corroborated by several surveys [27–29]. Further, many study focus on the use of price signals datasets are often insufficient to exhaustively model stocks in some sectors and industries [28, 30, 31].

Notably, to our knowledge, there has not been an attempt to model the dynamics of the luxury and retail sector, known as Consumer Cyclical (CC) [1, 28, 32]. Forecasting the CC industry is crucial for several reasons. 1) CC stocks often offer a stable yet profitable investment option, providing a degree of safety while yielding more than treasury bonds [1, 33]. 2) The performance of the CC industry can serve as a barometer for broader economic health, since consumer spending on luxury and retail goods often reflects economic confidence. 3) Although forecasting broader macroeconomic trends is challenging [28], predicting the future movements of CC industry stocks could improve policymaking for government institutions [34, 35]. Understanding future trends in this sector can help craft policies that support economic stability and growth. 4) Further insights from the CC industry can impact various downstream applications, including marketing strategies, supply chain management, and financial planning for businesses within the sector.

However, we further observe that datasets used in the research often exclude the wider economic context and use only historical stock price data, index data [14, 25], social media data [36], or price data transformed using statistical techniques [11]. The omission of economic indicators in the existing datasets makes it arguably difficult to gain insights

into CC stocks' behaviour, which tends to be influenced by the trends of the wider economy the CC companies are situated in [31]. The lack of forecasting research for CC industry, and the limited use of economic indicators prompt us to ask the question, *Could we improve CC stocks performance forecasting with Transformers by adding economic and temporal context?*

The second limitation concerns how existing Transformers conceptualize time series tokens for financial forecasting. Typically, a time series token is a vector of continuous variable observations at one timestep [16–18, 20, 21, 23–26, 37–41]. These vectors are used in forecasting models to predict future timesteps by exploring the interaction between other timesteps. However, this approach may overlook the interplay of temporal patterns in feature variates (e.g., financial price signals, technical indicators, economic indicators). Longer-term temporal patterns are often critical for financial forecasting, especially in the CC industry against economic data. This leads us to the second research question *Could we improve Transformer's performance in financial time series forecasting tasks by directly tokenising feature variates (e.g., economic indicators) instead of existing multivariate timesteps?* By using feature variates as tokens, we can model correlation between temporal feature interactions in a holistic way rather than interactions between local feature timesteps. This would reveal trends and offer better predictive capabilities that were previously overlooked.

To summarize our contributions:

To our knowledge, this is the first study to model the CC industry and consider using financial and economic data to enhance forecasting.

- (1) We create a novel economic-financial time series dataset comprises of 120 CC industry companies, including cruise operators, golf equipment manufacturers and kitchenware resellers. It includes economic and financial indicators, to provide insights from both economic context and market price signals. The dataset will be publicly accessible as a benchmark for community benefit [1, 31–32, 42–44].
- (2) We propose a simple but new feature variate tokenization for the Transformer architecture to better leverage the temporal pattern in data and model interplay between feature variates (e.g., financial, technical, economic indicators).
- (3) We compare the model's performance against baseline models and conduct an ablation study on the novel dataset through a series of experiments. The results show that the use of economic indicators and the new feature variate tokenization shows promising results.

In the following, Section II surveys relevant TSF literature. Section III discusses the concept of the Consumer Cyclical industry. Section IV outlines the creation of the experimental dataset. Section V describes our tokenization adaptation to the Transformer model. Section VI gives an overview of our experimental setup and evaluation measures. Section VII presents and discusses our findings. Section VIII concludes this paper.

II. RELATED WORK

A. Time Series Forecasting

Time Series Forecasting (TSF) methods originated in 1927 with the introduction of an auto-regressive model [2]. The ARMA method, introduced in 1931, and ARMA/GARCH, introduced in 1982, have been significant contributions to the field [8, 9]. However, these models have encountered difficulties with nonlinear time series [45]. To address this, classic ML techniques such as SVM and Adaptive Boosting have been employed [10]. These models can handle multivariate datasets, but they face challenges with out-of-sample data prediction. Deep Learning (DL) models, due to their ability to approximate nonlinear dependencies, have shown promise in this area [46]. Recent notable works based on the transformer architecture include the LogSparse Transformer, which breaks the Self-Attention quadratic memory bottleneck [47], and the Sepformer, which utilizes discrete wavelet transforms for feature extraction [48].

B. Financial Time Series Forecasting Methods

Traditional ML and LSTM-based Techniques: The use of machine learning techniques in TSF has a long history.

For instance, Manojlović *et al.* [11] introduced the use of the Random Forest algorithm for classifying the movement of the Croatian Stock Exchange Index (CROBEX) index and four companies on the Zagreb Stock Exchange in 2015. This was based on 12 technical indicators derived from their Open-High-Low-Close (OHLC) and volume values. Since 2018, Long Short-Term Memory (LSTM) networks have been widely used in TSF tasks. Wu *et al.* [49] introduced the LSTM model for stock trend classification, trained on text corpus and price signals. Liang *et al.* [36] proposed an LSTM model trained on a mixed corpus of news items and technical indicators derived from 262 companies' OHLC values to classify multi-horizon market trends. Dai *et al.* [14] proposed a Bi-directional LSTM model with an attention mechanism for predicting future values of the Shanghai Securities Composite Index. Xu *et al.* [12] introduced another LSTM model with a denoising attention mechanism, trained on the StockNet dataset for multi-horizon classification. Shen *et al.* [22] proposed a Gated Recurrent Unit network, trained on a dataset of three general indices' time series, for multi-step future trend classification. Tsang *et al.* [23] proposed a new LSTM model and a dataset including two economic variates—the USD exchange rate and the interbank offered interest rate.

Transformer-Based Techniques: Recently, several Transformer architectures for financial TSF tasks have been proposed. Malibari *et al.* [25] was the first to explore this architecture in the financial TSF context, adapting the Vision Transformer by treating training sequences as flattened patches. Both Wang *et al.* [20] and Juairiah *et al.* [24] investigated the performance of the original Transformer architecture, training their models on four exchange-based indices and Microsoft Corporation's stock prices, respectively. Ramos-Perez *et al.* [16] presented the Multi-Transformer architecture, which uses bagging to pick random subsets from the training set and incorporates GARCH-based methods in the Transformer modules for next day volatility forecasting of the S&P 500. Zeng *et al.* [21] enhanced the Transformer with a Convolutional Neural Network (CNN)

module for minute-by-minute stock classification. Wang *et al.* [18] performs one-day-ahead best stock selection and incorporates a graph module and an LSTM module in addition to a Transformer module.

Despite these advancements, the original Transformer, designed for Natural Language Processing (NLP) purposes, may not be well-suited to model the time dependencies and nonlinear complex dependencies of time series [47]. As Li *et al.* [47] demonstrated, the self-attention computation often becomes context-agnostic and fails to differentiate between anomalies and seasonality. Chen *et al.* [28] observed that multiple-feature Transformer models lose some of the complex dependencies between features. Recent efforts have been made to adapt this for TSF. Notably, Cai *et al.* [50] designed the Auto-Correlation mechanism to conduct seasonality discovery as part of the learning process. Yahoo_fin [51] attempted to model periodic seasonality for traffic prediction by extracting features from timestamps. Fan *et al.* [47] proposed a Convolutional SelfAttention module to convert patches of inputs into query-key pairs. However, the loss of information about nonlinear dependencies between features is still a recognized issue with Transformers [27, 28].

We hypothesize that using feature timesteps as tokens in traditional transformers may be one of the issues. To address this, we propose using feature variate as a token. This allows the transformer to explicitly learn the nonlinear dependency of temporal patterns in a holistic way. This approach could improve forecasting, especially for CC, which is sensitive to economic and other indicators.

C. Financial Time Series Forecasting Dataset

Within the literature, the use of time series data beyond the Open-High-Low-Close (OHLC), volume and indices data has seen limited exploration. These include text corpus [49], news items [36], USD exchange rate and the interbank offered interest rate [23]. For Transformer-based models, we have observed that only two studies have used non-OHLC and volume values time series data. Li *et al.* [41] employed the StockNet dataset for a one-day-ahead price prediction task, integrating a Transformer with an LSTM network to model temporal dependencies. Lim *et al.* [17] extracted data from 31 companies from the OMI dataset and used a GRN network to reduce data noise, leveraging the recurrence relation.

Moreover, the incorporation of temporal data into input sequences has been somewhat overlooked in financial TSF. To our understanding, only Muhammad *et al.* have attempted to embed time features in their input sequences while using a Transformer model for stock price prediction tasks [26]. They used a dataset of 8 companies' stocks, confined to the finance and insurance industries, to train the model for a multi-horizon prediction regression task. The model was fed the standard OHLC and volume data, supplemented with the encoded time, representing an intriguing contribution in terms of addressing the challenge of time embedding.

To our knowledge, limited studies have incorporated information about the broader economic context or specialized their datasets to focus on specific sectors. This paper represents the first attempt to consider additional economic context for the CC sector forecasting and to

propose and curate a benchmark dataset for this purpose. Our novel tokenization approach offers a new avenue to explicitly incorporate temporal dependencies into the transformer model.

III. THE CONSUMER CYCLICALS INDUSTRY

In finance, an industry refers to a group of companies engaged in similar business activities [42]. For example, the technology sector, which involves manufacturing electronics, creating software, and providing information technology services, includes companies like Meta and Alphabet Inc. Another example is the cyclical sector [1], also known as the Consumer Cyclical (CC) sector [32], which encompasses companies that depend on personal discretionary spending. This spending involves money used by individuals on non-essential goods and services, such as household durables (e.g., appliances and furniture) [31], leisure products, and luxury goods [43, 44]. The CC industry is particularly sensitive to broader economic conditions because they directly impact the level of discretionary spending [31]. This sensitivity is referred to as cyclical, describing the extent to which an industry's output fluctuates in response to economic changes. As a result, the CC industry is expected to exhibit cyclical behavior due to regular economic shifts [1, 31].

IV. DATASET

A. Dataset Overview

In this project, we propose a dataset that consists of assembled time series data for 120 CC companies, such as Carnival Corp (cruise operator), Topgolf Callaway Brands Corp (golf equipment manufacturer) and Williams Sonoma, Inc. (kitchenware and home furnishings reseller). The discussion that follows focuses on a specific company, denoted as c , for which we aim to assemble the dataset S . Let d represent the target length of the dataset in timesteps, which are days in our study, and let $[\tau_0, \tau_d]$ denote the timeframe of the target time series. The target dataset is composed of five facets, represented as $f \in S = \{F, E, T, C, V\}$: Financial Indicators F , Economic Indicators E , Temporal features T , Stock close price values C , and Stock volume values V (see summary in Table 1).

Table 1. Summary of dataset facets

Symbol	Dataset facet description
F	26 variates (financial indicators)
E	16 variates (economic indicators)
T	6 variates (temporal features)
C	1 variate (closing price)
V	1 variate (volume values)

1) Price and volume signal (C, V)

For each of the stock, we store the stock closing price C and stock volume V . C and V contain a single variate each.

2) Financial indicators ($F_1, \dots, F_{26}$)

We source our stock data from Yahoo Finance (YF) [32], using the yahoo fin library [52]. YF divides the CC industry into 21 subindustries, which is relevant for matching a company's data with relevant economic indicators. At the time of sourcing, there are 360 publicly traded companies in

the sector whose data is available via YF [32]. Our approach is to select the K top-performing companies representing each subindustry within the CC industry or all the companies from the subindustry if there are fewer than K . This is to ensure as fair a representation for each industry as possible. Thus, each company's time series contains the OHLC and volume values for the given trading day, the adjusted close value for that day (which we discard, due to its high correlation with the close value), and the date timestamp of that trading day. Given an OHLC and volume values time series, we obtain a set of standard financial indicators temporal features $F_1, \dots, F_{26}$. Examples include Accumulation/Distribution Line, Average True Range, Moving Average Convergence Divergence (MACD) etc. The full set of features and their default parameters are summarized in Table A1.

3) Economic indicators ($E_1, \dots, E_{16}$)

The Consumer Cyclical (CC) sector is highly sensitive to economic conditions. Given that our stock data focuses on U.S.-traded companies, we are particularly interested in the country's economic context. Inspired by relevant literature [1, 32], we identified key indicators [34, 44, 53, 54] to enrich our dataset: *Consumer Confidence Index (CCI)*, *Corporate Debt as a percentage of GDP*, *Gross Domestic Product (GDP)*, *Personal Income and Outlays*, *Personal Saving Rate*, *Retail Sales*, and *Unemployment Rate*.

The CCI reflects consumer sentiment and spending intentions, significantly impacting demand in the CC sector [1, 34]. Corporate Debt as a percentage of Gross Domestic Product (GDP) provides insights into corporate financial health and investment strategies [32]. GDP serves as a broad measure of economic health, correlating with consumer spending and business investment [1, 32]. Personal Income and Outlays indicate consumer spending capacity, while the Personal Saving Rate highlights financial resilience. Retail Sales data reveals consumer spending trends, and the Unemployment Rate reflects labor market conditions affecting disposable income and confidence [1, 52, 53]. By incorporating these indicators, we aim to better analyze the relationship between economic conditions and stock performance in the CC sector.

The works indicated above enabled us to identify the institutions that collect and produce the relevant time series. Multiple US institutions maintain Application Programming Interface (API) endpoints, and we were able to access all but one indicator directly from the US sources; we accessed the Consumer Confidence Index time series through the (Organisation for Economic Co-operation and Development) OECD API. Table A2 summarizes these data sources. Each indicator for the timespan $[\tau_0, \tau_d]$ is fetched from the relevant API for further processing to finally become a variate in E . These economic indicators, denoted $E_1, \dots, E_{16}$, are summarized in Table A3.

4) Temporal features ($T_1, \dots, T_6$)

We also hypothesize that the times of year may have contributions towards the CC sectors. For example, holiday time during summer, Christmas time during winter. We also collect the relative Day, Month, Quarter of the year $T_1, \dots, T_6$ as temporal features (see Table A4).

B. Data Processing & Feature Selection

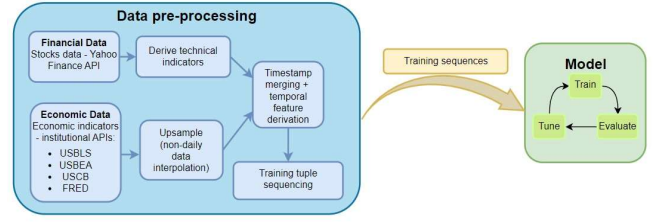


Fig. 1. Data pipeline.

Our processing pipeline (Fig. 1) sources, merges, and upsamples the relevant data facets to a format with regular intervals. Precisely, we define each feature vector x_i of the format:

$$\langle F_{1,i}, \dots, F_{26,i}, E_{1,i}, \dots, E_{16,i}, T_{1,i}, \dots, T_{6,i}, C_i, V_i \rangle$$

where i is a timestep such that $i \in [\tau_0, \tau_d]$.

C. Creating Training and Testing Pairs

Define w as the number of timesteps in the lookback window that constitute the training sequence, and p as the prediction length in timesteps. Training our predictive model involves creating pairs of input sequences (sequences of feature vectors) and corresponding ground truth sequences. Each pair, termed a “training tuple”, denoted as u , comprises an input sequence (spanning w days starting on day i) and the corresponding ground truth sequence (spanning from day $i + w + 1$ to $i + w + p + 1$). The ground truth sequence is a sequence because we are typically predicting multiple feature variates, not just a single one, and we need to compare our model's predictions against this sequence of actual outcomes to assess its performance accurately. With the sliding window approach, each new input and its corresponding predicted sequence overlap with the previous one by shifting forward one timestep. This method ensures that every sequence shares a portion of its data with the previous one, maintaining continuity across the entire time span of interest (Fig. 2).

τ_0	...	τ_{i-1}	τ_i	...	τ_{i+w}	τ_{i+w+1}	...	$\tau_{i+w+p+1}$	$\tau_{i+w+p+2}$	$\tau_{i+w+p+3}$	...	τ_d
τ_0	...	τ_{i-1}	τ_i	τ_{i+1}	...	τ_{i+w+1}	τ_{i+w+2}	...	$\tau_{i+w+p+2}$	$\tau_{i+w+p+3}$	...	τ_d
τ_0	...	τ_{i-1}	τ_i	τ_{i+1}	τ_{i+2}	...	τ_{i+w+2}	τ_{i+w+3}	...	$\tau_{i+w+p+3}$	...	τ_d

Fig. 2. The overlapping windows technique.

1) Data processing steps

Data processing and dataset assembly for a company of interest defined as c form a pipeline described below. It involves four major steps, which we outline below:

- (1) OHLC and volume values of c are sourced from Yahoo Finance [52] as a time series of 5 variates for the interval (τ_0, τ_d) .
- (2) Technical indicators are calculated for each timestep in the sourced OHLC and volume time series using the TALib library [55]. After this, Open-High-Low (OHL) values are discarded due to a high correlation with the closing price. The technical indicators form facet F , and the retained close price values and volume values form facets C and V , respectively.
- (3) Economic indicators to form variates $v \in E$ are sourced from APIs listed in Table A2.
- (4) $v \in E$ are upsampled to a daily frequency using the

2nd order polynomial interpolation to cover the interval $[\tau_0, \tau_d]$.

- (5) $\{F, C, V\}$ and $\{E\}$ are merged to form $\{F, E, C, V\}$.
- (6) Temporal features $\{T_{1,i}, \dots, T_{6,i}\}$ are derived from each timestep's timestamp and merged with $\{F, E, C, V\}$ to form $\{F, E, T, C, V\}$.
- (7) Post-merging, each item in $\{F, E, T, C, V\}$ consists of timesteps in the range $[\tau_0, \tau_d]$, enriched with 16 technical indicators, 26 economic indicators, closing prices, and timestamps. These are then sequenced as follows:
 - (i) $\{F, E, T, C, V\}$ is segmented into overlapping windows of w days, each segment starting the day after its predecessor's first day.
 - (ii) For each window starting on day i and ending on $i + w$, a training tuple u is generated as follows:

$$u = \{F, E, T, C, V\}_{i,i+w}, \{F, E, T, C, V\}_{i+w+1,i+w+p+1}$$

where p is the prediction length in days.

- (iii) The tuples are partitioned into training, validation, and test sets.

V. METHODOLOGY

The Transformer model [15] is a deep learning architecture often used for Natural Language Processing (NLP). It relies on a mechanism called self-attention to process input data. Tokens are the basic units of input data, typically words or subwords in NLP. The self-attention mechanism allows the model to weigh the importance of different tokens in a sequence, enabling it to capture relationships between tokens regardless of their position. The Transformer consists of an encoder that processes the input sequence and a decoder that generates the output sequence.

In the context of financial time series forecasting, the traditional Transformer approach treats each timestep in a time series as a separate input token. This means the model learns the temporal relationships between these timesteps. For example, Jiang *et al.* [27] and Chen *et al.* [28] follow this approach. Lim *et al.* [56] discuss the use of temporal embeddings to enhance the original tokenization approach by attaching temporal tokens representing an aggregate of temporal features per multiple time series features. Zeng *et al.* [57] demonstrate the effectiveness of using an Encoder-only Transformer for time series forecasting. Muhammad *et al.* [26] and Lim *et al.* [17] explore the idea of generating temporal features and using them as tokens. However, these existing approaches have certain limitations. Adding temporal tokens, as done by Muhammad *et al.* [26] and Lim *et al.* [17], increases the complexity of the model. It requires additional steps to generate and integrate these tokens, complicating the model architecture and training process. Additionally, while these methods capture temporal dependencies, they might not fully leverage the interactions between different feature variates in the dataset, focusing more on the temporal patterns.

Our contribution diverges from these traditional methods by proposing a new way to handle tokenization in the Transformer model. Instead of treating each timestep in a time series as a separate input token, we suggest treating each

feature variate as a token. To this end, our method changes the research focus to learn the relationships between different feature variates (temporal patterns of indicators or signals). Using feature variates as tokens provides a holistic view and directly learns to capture interactions with long-term dependencies intrinsically supported. This is particularly beneficial for financial time series, where understanding the interplay between various indicators (e.g., technical, economic indicators) is crucial and helps the model uncover connections across different features, leading to better predictive performance.

To do so, we transpose the input data before feeding it into the model (Fig. 3). Our approach simplifies the model architecture (e.g., does not require additional steps to compute temporal tokens), making it easier to implement and train, and reducing computational overhead. Our method is also flexible and can handle both one-step-ahead and multi-step predictions, similar to generating one token at a time in Natural Language Processing (NLP).

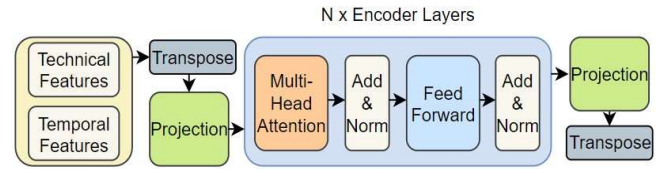


Fig. 3. Model overview.

To implement our approach, we start with data preparation. Let v represent a feature in the dataset S , and v_{num} be the total number of features. Given a sequence of feature vectors $X^{w \times v_{num}}$ (where w is the number of timesteps), we transpose it to $X^{v_{num} \times w}$. The transposed data $X^{v_{num} \times w}$ is then fed into an Encoder-only Transformer model. The input passes through a series of encoder layers. Each layer uses a multi-head self-attention mechanism to compute attention scores, highlighting the importance of each token (feature variate) in the sequence. The output of the attention mechanism is processed through a feed-forward network to create a new representation of the data. The encoder transforms the input sequence into a new representation $\hat{Y}^{v_{num} \times w}$. This representation is then transposed back to $\hat{Y}^{p \times v_{num}}$ (where p is the prediction length) to generate the forecast.

By redefining our approach to feature variate tokenization, we enable our model to uncover connections across multiple feature variates and reveal trends, offering insights and predictive capabilities that were previously overlooked. Our proposed technique looks simple but represents new thinking in using Transformers for financial time series forecasting, focusing on capturing holistic temporal feature interactions rather than interactions between local feature timesteps. Our comprehensive experiments are the first to validate this approach.

VI. EXPERIMENT SETUP AND EVALUATION MEASURES

A. Hardware

Thanks to HPC Wales a compute node was made available to us. The system configuration includes 2 Intel(R) Xeon(R) Gold 6148 CPUs running at 2.40 GHz, each with 20 cores, totaling 40 cores. It is equipped with 384 GB of RAM and an Nvidia V100 GPU. For storage, the system offers 808 TB of

usable scratch space and 231 TB of home directory space, both managed on a Lustre filesystem. We selected two Transformer models and two non-deep learning models as baseline models for benchmarking. The first Transformer model, termed the “Vanilla Transformer,” serves as the forecasting baseline. This model mirrors our proposed architecture with the exception of having its decoder module removed. It follows the encoder-only setup, similar to DLinear [57]. The second Transformer model is the Original Transformer, as introduced by Vaswani [15]. This allows us to evaluate our model’s performance against both a foundational model used for further architectural enhancements and a well-established baseline. Additionally, we included two non-deep learning models for comparison: a Support Vector Regressor (SVR) and a Random Forest (RF) model, reflecting the approaches discussed by Sapankevych and Sankar [10] and Manojlovic and Stajduhar [11], respectively.

B. Hyperparameters

The non-deep learning models, SVR and RF, were implemented using default settings from the scikitlearn library [58]. For the deep learning models, we used the Optuna framework to fine-tune hyperparameters and applied these optimized settings consistently across the baseline Transformer models. To maintain consistency, each deep learning model was trained for a maximum of 10 epochs, with early stopping managed by the ReduceLROnPlateau scheduler provided by the PyTorch library [59]. In our experiments, this scheduler typically halted training before the 10-epoch limit was reached. Detailed information about the hyperparameters can be found in Table A5.

C. Data

Our proposed dataset, detailed in Section IV, is naturally divided into five components: the financial indicators facet F, the economic indicators facet E, the temporal features T, the close price values variate C, and the volume values variate V. We conducted a data ablation study to assess the impact of various combinations of these components, resulting in the following dataset types: We tested these 7 dataset types, which are:

$\{T, C, V\}$, $\{F, C, V\}$, $\{E, C, V\}$, $\{F, E, C, V\}$, $\{E, T, C, V\}$, and $\{F, T, C, V\}$, $\{F, E, T, C, V\}$.

We tested these 7 dataset types across different prediction window lengths and lookback window lengths. Specifically, we examined lookback windows of 16, 32, and 64 days, and prediction lengths of 1, 16, 32, and 64 days. This setup resulted in 12 experiments for each dataset type (e.g., one experiment might use the $\{T, C, V\}$ dataset with a 16-day lookback window and a 1-day prediction length). Given 5 models (4 baseline models and our proposed model), we conducted a total of $12 \times 7 \times 5 = 420$ experiments. With 4 different prediction lengths, we defined 4 distinct prediction tasks. Thus, our experiments can be analyzed to determine the best-performing combinations of models and dataset types for each prediction task.

D. Evaluation Measures

To evaluate the outcomes of our experiments, we utilize Mean Absolute Error (MAE), following [57], which measures the average magnitude of errors between predicted

and actual values, without considering their direction. It is defined as:

$$\text{MAE} = \frac{1}{N} \sum_{i=1}^N |y_i - \hat{y}_i|$$

where y_i represents the actual values, $\hat{y}_i$ denotes the predicted values, and N is the total number of samples.

VII. EXPERIMENTAL RESULTS

To evaluate the proposed dataset and the updated model, we conducted a data ablation study as well as an architecture comparison study. Each experiment in our study is identified by the model, the dataset type, the task type (the prediction length), and the lookback. An experiment is assigned an average MAE score of all companies in our dataset for the given experimental setup. The scores shown in the figures are averaged for visual analysis of the experimental results, as noted in the figure captions (e.g., in Fig. 4, averages represent a model’s performance across all four lookback lengths for each prediction length).

A. Model Comparison

As the chart in Fig. 4 shows, given the task of predicting 1, 16, 32 and 64 days ahead, our model performed best across all four tasks. Interestingly, the Vanilla Transformer did not perform well in the 1-day task (only the RF model performed worse), in contrast to in the 32day forecasting task, where it performed better than both non-transformer models. The SVR model performed surprisingly well as compared to other models in the 1-day-ahead task and fell to the fifth place in the 64-day task. The Vanilla Transformer started out in place 4 in 1 day task, to degrade to the last place in 16-day task, where it remained in 32-day task. In 64-day task, the Vanilla Transformer performed better than the SVR model only. The performance differences of the three Transformer models might indicate that our new tokenization strategy yields significant performance improvement.

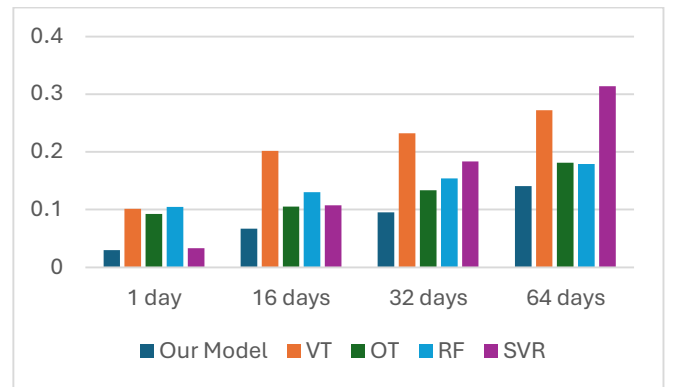


Fig. 4. Model performance per task. Each bar represents the average of all experiment results given a model and a task. The lower the MAE value, the better.

We briefly discuss the impact of lookback window length on prediction quality. Fig. 5 shows that in the Original Transformer model, longer lookback windows improve performance for 32- and 64-day tasks. In contrast, our model’s performance is less affected by lookback length,

suggesting that using feature variates as tokens allows the Transformer to rely more on inter-variate dependencies than input sequence length. Original transformer only outperforms ours when it has the largest lookback window.

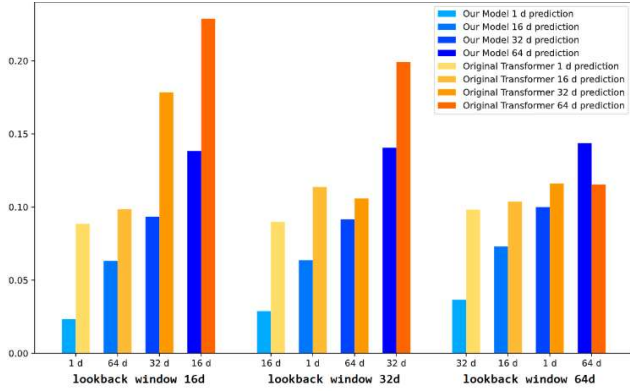


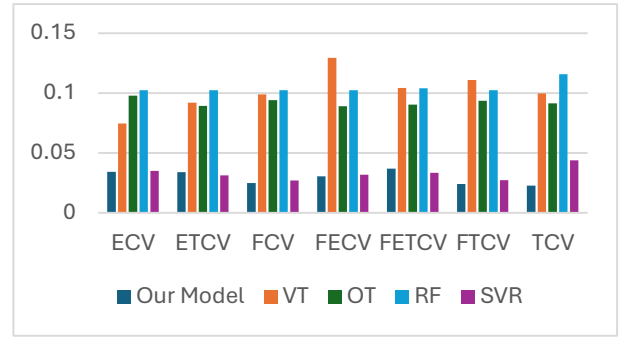
Fig. 5. Lookback window performance for each task using Our Model (OM) and the Original Transformer (OT) at different prediction lengths. Each bar represents the average of all experiments for a specific model, task, and lookback window length.

B. Dataset Ablation

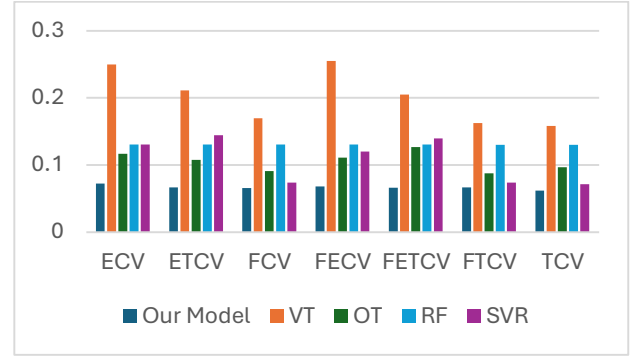
The dataset ablation study in Fig. 6 showed the changing quality of separate facets' contribution to the quality of forecasts. The $\{T, C, V\}$ dataset shows the best performance on 1, 16 and 64-day tasks, whereas the $\{F, C, V\}$ dataset does best in the 32-day task. Interestingly, the $\{F, E, T, C, V\}$ dataset, while never taking the first spot in any task, saw improvement over the 1-day, 16-day, and 32-day tasks, to degrade ever so slightly for the 64-day task. A somewhat similar behavior was exhibited by the $\{F, E, C, V\}$ dataset, whose performance improved precipitously over the 16, 32 and 64-day tasks.

This shows that the contribution of the economic facet E degrades prediction quality, and the T facet sees the opposite of this dynamic, as shown in Fig. 6. Additionally, a similar effect could be at play with the T facet. An assumption could be made that the T facet (holiday time in summer, winter Christmas time etc.) plays a more significant role in shorter-term prediction tasks. The F facet's contribution could have been hindered by keeping parameters when calculating technical indicators for particular day timesteps constant.

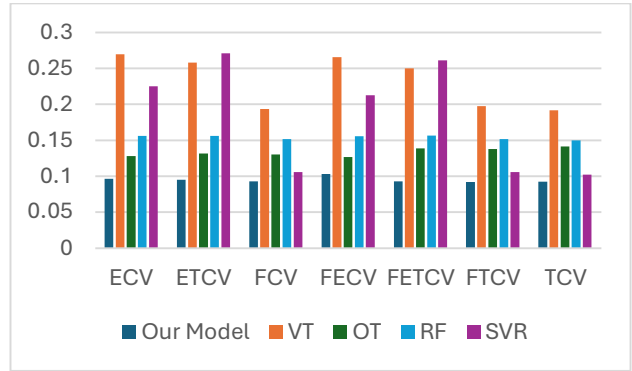
When analyzing the performance of each dataset type and model pair, on each individual task, we can also notice a relationship where the $\{E, C, V\}$ dataset tends to perform better than $\{F, E, T, C, V\}$, and both performing worse than $\{T, C, V\}$, as visualised in Fig. 7. As observed above, this could also be explained by the fact that the technical indicators are parametrized by static values (e.g., Average Directional Index (ADX) has its period set to 14 days, and Absolute Price Oscillator (APO) has the fast period set to 26 and the slow period set to 12 days; see Table A1), and not adjusted for the lookback window lengths of the respective experiments. In this case, the financial indicator features could be detrimental to the prediction quality. The inclusion of both the F and E facets improves the prediction quality in the 64-day task as compared to the $\{E, T, C, V\}$ dataset, which suggests that the idea of providing both the economic context and technical indicators for financial time series improves financial forecasting task performance.



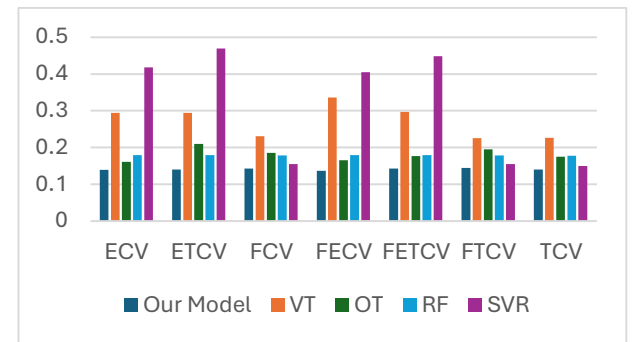
(a) 1-day prediction task



(b) 16-day prediction task



(c) 32-day prediction task



(d) 64-day prediction task

Fig. 6. Dataset performance per model, over (a) 1, (b) 16, (c) 32, and (d) 64-day prediction tasks. OT and VT stand for Original Transformer and Vanilla Transformer, respectively. A bar in each figure represents the average of all experiment results, given a dataset type, a model and a task. The lower the MAE value, the better.

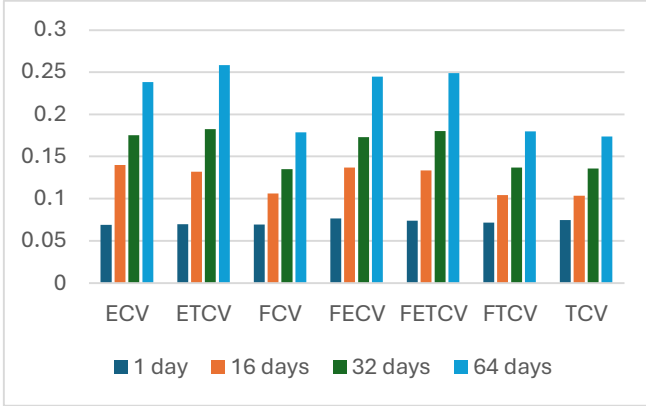


Fig. 7. Dataset performance per task. Each bar represents the average of all experiment results given a dataset type and a task. The lower the MAE value, the better.

C. Complexity and Timing

Transformer models generally require significant training time due to their quadratic complexity. Our model introduces transpose layers before and after the Transformer, maintaining a complexity similar to other models. On a V100 GPU, it converges in approximately 8 to 12 h, depending on the parameters for day-ahead forecasting and the lookback window.

VIII. CONCLUSION

In summary, our examination of financial time series forecasting reveals a critical gap in research, particularly in modeling individual business sectors like the CC sector and incorporating economic indicators for deeper contextual insights. To bridge this gap, we introduce a novel dataset and an innovative Transformer tokenization technique that better captures the intricate interplays of financial features. By transposing input time series as tokens, our approach was tested against various Transformer architectures and baseline models (SVR, RF). The findings demonstrate that integrating economic context and tokenizing variates, rather than relying on common timesteps, significantly boosts forecasting accuracy. Despite its simplicity, this tokenization method opens new pathways for exploring and integrating temporal patterns in financial time series and beyond, leading to more effective forecasting.

Our dataset is publicly accessible, and future research will explore its application with other Transformer variants and the effects of input transposition on self-attention mechanisms. Specifically, alternative Transformer architectures like the Triformer [60], Sepformer [61], SpringNet [19], and Multi-Transformer [16] could enhance our model's ability to capture the non-linear and non-stationary dynamics of stock price movements. The Triformer's long-range dependency management and the Sepformer's focus on sequential information may provide insights when combined with our use of multi-variate data as tokens that emphasize temporal patterns. Additionally, integrating multi-scale temporal tokens could improve our model's ability to capture variations in stock prices across different time horizons.

Further, various domains have made attempts to leverage the Transformer architecture, including traffic management [50], weather and energy consumption forecasting [51]. We are interested to integrate the new input

transposition technique to enhance temporal pattern modeling in these forecasting areas, thereby improving the models' ability to capture dynamic changes over time.

APPENDIX A

Table A1. Financial indicator feature calculation

Feature	Indicator Description	Parameters
F_1	Accumulation/Distribution Line	-
F_2	On-Balance Volume (OBV)	-
F_3	Average True Range (ATR)	period = 7
F_4	Average True Range (ATR)	period = 14
F_5	Normalized Average True Range (NATR)	period = 7
F_6	Normalized Average True Range (NATR)	period = 14
F_7	True Range (TRANGE)	-
F_8	Standard Deviation (STDDEV)	period = 5
F_9	Standard Deviation (STDDEV)	period = 10
F_{10}	Standard Deviation (STDDEV)	period = 14
F_{11}	Time Series Forecast (TSF)	period = 5
F_{12}	Time Series Forecast (TSF)	period = 10
F_{13}	Time Series Forecast (TSF)	period = 14
F_{14}	Average Directional Index (ADX)	period = 14
F_{15}	Average Directional Index Rating (ADX)	period = 14
F_{16}	Absolute Price Oscillator (APO)	fastperiod = 12, slowperiod = 26
F_{17}	Aroon Indicator (AROON Down)	period = 14
F_{18}	Aroon Indicator (AROON Up)	period = 14
F_{19}	Aroon Oscillator (AROONOSC)	period = 14
F_{20}	Balance of Power (BOP)	-
F_{21}	Commodity Channel Index (CCI)	period = 14
F_{22}	Chande Momentum Oscillator (CMO)	period = 14
F_{23}	Directional Movement Index (DX)	period = 14
F_{24}	Moving Average Convergence Divergence (MACD)	fastperiod = 12, slowperiod = 26, signalperiod = 9
F_{25}	MACD Signal Line	-
F_{26}	MACD Histogram	-

Table A2. Economic data sources

Indicator Name	Releasing Institution	Frequency	Composite
Consumer Confidence Index	Bureau of Economic Analysis	Monthly	No
Corporate Debt as % of GDP	Federal Reserve	Quarterly	Yes
Gross Domestic Product	Bureau of Economic Analysis	Quarterly	No
Personal Income and Outlays	Bureau of Economic Analysis	Monthly	No
Personal Saving Rate	Bureau of Economic Analysis	Monthly	No
Retail Sales	U.S. Census Bureau	Monthly	Yes
Unemployment Rate	Bureau of Labor Statistics	Monthly	No

Table A3. Economic indicator feature calculation

Feature Type	Feature	Description
Consumer Confidence & Financial Obligations	E_1	Consumer Confidence Index (CCI)
	E_2	Consumer Debt Service Payments (CDSP)
	E_3	Exports
Gross Domestic Product (GDP)	E_4	Government consumption expenditures and gross investment
	E_5	Gross domestic product
	E_6	Gross domestic product, current dollars
	E_7	Gross private domestic investment
	E_8	Imports
	E_9	Personal consumption expenditures

Personal Income and Outlays	E_{10}	Equals: Personal saving
	E_{11}	Personal consumption expenditures
	E_{12}	Personal income
	E_{13}	Personal saving as a percentage of disposable personal income
	E_{14}	Personal Savings Rate
	E_{15}	Retail Sales
	E_{16}	Unemployment Rate

Table A4. Temporal features

Feature	Calculation Description
T_1	Quarter of the year, zero-indexed
T_2	Month of the year, zero-indexed
T_3	Day of the year, zero-indexed, adjusted for leap years
T_4	Day within the current quarter, zero-indexed
T_5	Day of the month, zero-indexed
T_6	Day of the week, starting from zero (Monday)

Table A5. Proposed hyperparameters

Value	Hyperparameter
8	Number of heads
128	Feed-forward network dimensions
128	Model dimensions
1	Number of encoder layers
0.25	Encoder dropout probability
16	Batch size
10	Number of epochs
mse	Loss function type
Adam	Optimizer type
0.001	Optimizer learning rate
0.0	Optimizer weight decay
ReduceLROnPlateau	Scheduler type
0.9	Scheduler factor
7	Scheduler patience
min	Scheduler mode
1×10^{-5}	Scheduler min. Learning rate

CONFLICT OF INTEREST

The authors declare no conflict of interest.

AUTHOR CONTRIBUTIONS

Kasper identified the research direction, sources, and developed all datasets and techniques. Kasper is also responsible for drafting and revising the paper. Gary and Daniel provided supervision, contributed to writing, revising, and commenting on the paper, and both lead the research project; all authors had approved the final version.

ACKNOWLEDGEMENTS

We gratefully acknowledge the support of HPCWales and Swansea University for the GPU cluster and computing facilities.

REFERENCES

- [1] Z. Bodie, A. Kane, and A. J. Marcus, *Investments*, 10th ed. New York, NY, US: McGraw-Hill Education, 1999.
- [2] G. U. Yule, "On a method of investigating periodicities in disturbed series, with special reference to wolfer's sunspot numbers," *Philosophical Transactions of the Royal Society of London. Series A, Containing Papers of a Mathematical or Physical Character*, vol. 226, pp. 267–298, 1927.
- [3] Y. Qi and J. Xiao, "Fintech: AI powers financial services to improve people's lives," *Communications of the ACM*, vol. 61, no. 11, pp. 65–69, Oct. 2018. doi: 10.1145/3239550
- [4] A. Bahrammirzaee, "A comparative survey of artificial intelligence applications in finance: Artificial neural networks, expert system and hybrid intelligent systems," *International Journal of Neural Computing and Application*, vol. 19, no. 8, pp. 1165–1195, 2010.
- [5] M. Usmani, S. H. Adil, K. Raza, and S. S. A. Ali, "Stock market prediction using machine learning techniques," in *Proc. 2016 3rd International Conference on Computer and Information Sciences (ICCOINS)*, Aug. 2016, pp. 322–327. doi: 10.1109/ICCOINS.2016.7783235
- [6] X. Zhong and D. Enke, "A comprehensive cluster and classification mining procedure for daily stock market return forecasting," *Neurocomputing*, vol. 267, pp. 152–168, Dec. 2017. https://doi.org/10.1016/j.neucom.2017.06.010
- [7] E. Alomari, R. Mehmood, and I. Katib, "Sentiment analysis of arabic tweets for road traffic congestion and event detection," *Smart Infrastructure and Applications: Foundations for Smarter Cities and Societies*, pp. 37–54, 2020. https://doi.org/10.1007/978-3-030-13705-2_2
- [8] G. T. Walker, "On periodicity in series of related terms," in *Proc. of the Royal Society of London. Series A, Containing Papers of a Mathematical and Physical Character*, vol. 131, no. 818, Jun. 1931, pp. 518–532. https://doi.org/10.1098/rspa.1931.0069
- [9] R. F. Engle, "Autoregressive conditional heteroscedasticity with estimates of the variance of United Kingdom inflation," *Econometrica: Journal of the Econometric Society*, vol. 50, no. 4, pp. 987–1007, 1982. https://doi.org/10.2307/1912773
- [10] N. I. Sapankevych and R. Sankar, "Time series prediction using support vector machines: A survey," *IEEE Computational Intelligence Magazine*, vol. 4, no. 2, pp. 24–38, May 2009. doi: 10.1109/MCI.2009.932254
- [11] T. Manojlović and I. Štajduhar, "Predicting stock market trends using random forests: A sample of the Zagreb stock exchange," in *Proc. 2015 38th International Convention on Information and Communication Technology, Electronics and Microelectronics (MIPRO)*, May 2015, pp. 1189–1193. doi: 10.1109/MIPRO.2015.7160456
- [12] H. Xu, L. Chai, Z. Luo, and S. Li, "Stock movement predictive network via incorporative attention mechanisms based on tweet and historical prices," *Neurocomputing*, vol. 418, pp. 326–339, Aug. 2020.
- [13] A. Moghar and M. Hamiche, "Stock market prediction using LSTM recurrent neural network," *Procedia Computer Science*, vol. 170, pp. 1168–1173, Jan. 2020. https://doi.org/10.1016/j.procs.2020.03.049
- [14] H. Dai, W. Wang, J. Cao, and H. Wu, "A deep neural network for stock price prediction," *Journal of Physics: Conference Series*, vol. 1994, no. 1, p. 012029, Aug. 2021. https://dx.doi.org/10.1088/1742-6596/1994/1/012029
- [15] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, Kaiser, and I. Polosukhin, "Attention is all you need," in *Advances in Neural Information Processing Systems*, Curran Associates, Inc., 2017, vol. 30.
- [16] E. Ramos-Pérez, P. J. Alonso-González, and J. J. Núñez-Velázquez, "Multi-transformer: A new neural network-based architecture for forecasting S&P volatility," *Mathematics*, vol. 9, no. 15, 1794, Jan. 2021. https://doi.org/10.3390/math9151794
- [17] B. Lim, S. Ö. Arık, N. Loeff, and T. Pfister, "Temporal fusion transformers for interpretable multi-horizon time series forecasting," *International Journal of Forecasting*, vol. 37, no. 4, pp. 1748–1764, Oct. 2021. https://doi.org/10.1016/j.ijforecast.2021.03.012
- [18] H. Wang, T. Wang, S. Li, J. Zheng, S. Guan, and W. Chen, "Adaptive long-short pattern transformer for stock investment selection," in *Proc. of the Thirty-First International Joint Conference on Artificial Intelligence*, Jul. 2022, pp. 3970–3977. https://doi.org/10.24963/ijcai.2022/551
- [19] Y. Lin, I. Koprinska, and M. Rana, "SpringNet: Transformer and spring DTW for time series forecasting," in *Proc. Neural Information Processing: 27th International Conference, ICONIP 2020*, 2020, pp. 616–628. https://doi.org/10.1007/978-3-030-63836-8_51
- [20] C. Wang, Y. Chen, S. Zhang, and Q. Zhang, "Stock market index prediction using deep Transformer model," *Expert Systems with Applications*, vol. 208, 118128, Dec. 2022. https://doi.org/10.1016/j.eswa.2022.118128
- [21] Z. Zeng, R. Kaur, S. Siddagangappa, S. Rahimi, T. Balch, and M. Veloso, "Financial time series forecasting using cnn and transformer," arXiv Preprint, arXiv: 2304.04912, Apr. 2023.
- [22] G. Shen, Q. Tan, H. Zhang, P. Zeng, and J. Xu, "Deep learning with gated recurrent unit networks for financial sequence predictions," *Procedia Computer Science*, vol. 131, pp. 895–903, Jan. 2018. https://doi.org/10.1016/j.procs.2018.04.298
- [23] G. Tsang, J. Deng, and X. Xie, "Recurrent neural networks for financial time-series modelling," in *Proc. 2018 24th International Conference*

- on *Pattern Recognition (ICPR)*, IEEE, 2018, pp. 892–897. doi: 10.1109/ICPR.2018.8545666
- [24] F. Juariah, M. Mahatabe, H. B. Jamal, A. Shiddika, T. R. Shawon, and N. C. Mandal, “Stock price prediction: A time series analysis,” in *Proc. 2022 25th International Conference on Computer and Information Technology (ICCIT)*, IEEE, 2022, pp. 153–158. doi: 10.1109/ICCIT57492.2022.10056009
- [25] N. Malibari, I. Katib, and R. Mehmood, “Predicting stock closing prices in emerging markets with transformer neural networks: The Saudi stock exchange case,” *International Journal of Advanced Computer Science and Applications*, vol. 12, no. 12, 2021. <http://dx.doi.org/10.14569/IJACSA.2021.01212106>
- [26] T. Muhammad, A. B. Aftab, M. M. Ahsan, M. M. Muhu, M. Ibrahim, S. I. Khan, and M. S. Alam, “Transformer-based deep learning model for stock price prediction: A case study on Bangladesh stock market,” *International Journal of Computational Intelligence and Applications*, vol. 22, no. 03, 2350013, Apr. 2023. <https://doi.org/10.1142/S146902682350013X>
- [27] Y. Jiang, Z. Pan, X. Zhang, S. Garg, A. Schneider, Y. Nevmyvaka, and D. Song, “Empowering time series analysis with large language models: A survey,” arXiv Preprint, arXiv:2402.03182, Feb. 2024.
- [28] Z. Chen, M. Ma, T. Li, H. Wang, and C. Li, “Long sequence time-series forecasting with deep learning: A survey,” *Information Fusion*, vol. 97, 101819, Sep. 2023. <https://doi.org/10.1016/j.inffus.2023.101819>
- [29] K. Benidis, S. S. Rangapuram, V. Flunkert, Y. Wang, D. Maddix, C. Turkmen, et al., “Deep learning for time series forecasting: Tutorial and literature survey,” *ACM Computing Surveys*, vol. 55, no. 6, pp. 1–36, Dec. 2022. doi: 10.1145/3533382
- [30] N. Sarantis, “Nonlinearities, cyclical behaviour and predictability in stock markets: International evidence,” *International Journal of Forecasting*, vol. 17, no. 3, pp. 459–482, Jul. 2001. [https://doi.org/10.1016/S0169-2070\(01\)00093-0](https://doi.org/10.1016/S0169-2070(01)00093-0)
- [31] B. Petersen and S. Strongin, “Why are some industries more cyclical than others?” *Journal of Business & Economic Statistics*, vol. 14, no. 2, pp. 189–198, 1996. <https://doi.org/10.2307/1392430>
- [32] Top Consumer Cyclical Sector List | Screener—Yahoo Finance. [Online]. Available: https://finance.yahoo.com/screener/predefined/ms_consumer_cyclical/
- [33] B. Deleersnyder, M. G. Dekimpe, M. Sarvary, and P. M. Parker, “Weathering tight economic times: The sales evolution of consumer durables over the business cycle,” *Quantitative Marketing and Economics*, vol. 2, no. 4, pp. 347–383, Dec. 2004. <https://doi.org/10.1007/s11129-004-0137-x>
- [34] C. J. Swanson, “Long-term financial forecasting for local governments,” *Government Finance Review*, vol. 24, no. 5, 60, 2008.
- [35] X. Chen, X. Ma, H. Wang, X. Li, and C. Zhang, “A hierarchical attention network for stock prediction based on attentive multiview news learning,” *Neurocomputing*, vol. 504, pp. 1–15, Sep. 2022. <https://doi.org/10.1016/j.neucom.2022.06.106>
- [36] Z. Liang, J. Zhang, C. Liang, H. Wang, Z. Liang, and L. Pan, “A shapelet-based framework for unsupervised multivariate time series representation learning,” in *Proc. of the VLDB Endowment*, Nov. 2023, vol. 17, no. 3, pp. 386–399. doi: 10.14778/3632093.3632103
- [37] K. Yi, Q. Zhang, L. Cao, S. Wang, G. Long, L. Hu et al., “A survey on deep learning based time series analysis with frequency transformation,” arXiv Preprint, arXiv:2302.02173, Oct. 2023.
- [38] Q. Wen, T. Zhou, C. Zhang, W. Chen, Z. Ma, J. Yan, and L. Sun, “Transformers in time series: A survey,” in *Proc. of the Thirty-Second International Joint Conference on Artificial Intelligence*, 2023, pp. 6778–6786. <https://doi.org/10.24963/ijcai.2023/759>
- [39] S. López-Ruiz, C. I. Hernández-Castellanos, and K. Rodríguez-Vázquez, “Multi-objective framework for quantile forecasting in financial time series using transformers,” in *Proc. of the Genetic and Evolutionary Computation Conference (GECCO’22)*, Jul. 2022, pp. 395–403. <https://doi.org/10.1145/3512290.3528740>
- [40] Y. Li, S. Lv, X. Liu, and Q. Zhang, “Incorporating transformers and attention networks for stock movement prediction,” *Complexity*, vol. 2022, no. 1, e7739087, Feb. 2022. <https://doi.org/10.1155/2022/7739087>
- [41] C. Langager. (May 10, 2025). Industry vs. sector: What’s the difference? *Investopedia*. [Online]. Available: <https://www.investopedia.com/ask/answers/05/industrysector.asp>
- [42] XLY: The consumer discretionary select sector SPDR fund. [Online]. Available: <https://www.ssga.com/us/en/intermediary/etfs/funds/the-consumer-discretionary-select-sector-spdr-fund-xly>
- [43] Y. Chien. (October 1, 2015). Are we really consuming more services? [Online]. Available: <https://www.stlouisfed.org/on-the-economy/2015/october/consuming-more-services-durable-nondurable-goods>
- [44] M. F. Dixon, I. Halperin, and P. Bilokon, *Machine Learning in Finance: From Theory to Practice*, Switzerland: Springer International Publishing, 2020. <https://doi.org/10.1007/978-3-030-41068-1>
- [45] R. P. Masini, M. C. Medeiros, and E. F. Mendes, “Machine learning advances for time series forecasting,” *Journal of Economic Surveys*, vol. 37, no. 1, pp. 76–111, Feb. 2023. <https://doi.org/10.1111/joes.12429>
- [46] S. Li, X. Jin, Y. Xuan, X. Zhou, W. Chen, Y.-X. Wang, and X. Yan, “Enhancing the locality and breaking the memory bottleneck of transformer on time series forecasting,” *Advances in Neural Information Processing Systems*, vol. 32, 2019. Available: https://proceedings.neurips.cc/paper_files/paper/2019/hash/6775a0635c302542da2c32aa19d86be0-Abstract.html
- [47] J. Fan, Z. Wang, D. Sun, and H. Wu, “Sepformer-based models: More efficient models for long sequence time-series forecasting,” *IEEE Transactions on Emerging Topics in Computing*, pp. 1–12, 2022. doi: 10.1109/TETC.2022.3230920
- [48] Y. Xu and S. B. Cohen, “Stock movement prediction from tweets and historical prices,” in *Proc. the 56th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, Jul. 2018, pp. 1970–1979. doi: 10.18653/v1/P18-1183
- [49] H. Wu, J. Xu, J. Wang, and M. Long, “Autoformer: Decomposition transformers with auto-correlation for long-term series forecasting,” *Advances in Neural Information Processing Systems*, vol. 34, pp. 22419–22430, 2021. Available: <https://proceedings.neurips.cc/paper/2021/hash/bcc0d400288793e8bdcd7c19a8ac0c2b-Abstract.html>
- [50] L. Cai, K. Janowicz, G. Mai, B. Yan, and R. Zhu, “Traffic transformer: Capturing the continuity and periodicity of time series for traffic forecasting,” *Transactions in GIS*, vol. 24, no. 3, pp. 736–755, Jun. 2020. <https://doi.org/10.1111/tgis.12644>
- [51] Yahoo fin documentation. [Online]. Available: https://theautomatic.net/yahoo_fin-documentation/
- [52] N. Frumkin, *Guide to Economic Indicators*, Routledge, May 2015. google-Books-ID: P6FzCQAAQBAJ.
- [53] OECD Main Economic Indicators (MEI)—OECD. [Online]. Available: <https://www.oecd.org/sdd/oecdmaineconomicindicatorsmei.htm>
- [54] TA-Lib—technical analysis library. [Online]. Available: <https://ta-lib.org/>
- [55] B. Lim and S. Zohren, “Time-series forecasting with deep learning: A survey,” *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, vol. 379, no. 2194, 20200209, Feb. 2021. <https://doi.org/10.1098/rsta.2020.0209>
- [56] A. Zeng, M. Chen, L. Zhang, and Q. Xu, “Are transformers effective for time series forecasting?” in *Proc. the AAAI Conference on Artificial Intelligence*, Jun. 2023, vol. 37, no. 9, pp. 11121–11128. <https://doi.org/10.1609/aaai.v37i9.26317>
- [57] API reference. [Online]. Available: <https://scikit-learn/stable/modules/classes.html>
- [58] PyTorch. [Online]. Available: <https://www.pytorch.org>
- [59] N. Sarantis, “Nonlinearities, cyclical behaviour and predictability in stock markets: International evidence,” *International Journal of Forecasting*, vol. 17, no. 3, pp. 459–482, Jul. 2001. doi: 10.1016/S0169-2070(01)00093-0
- [60] R. -G. Cirstea, C. Guo, B. Yang, T. Kieu, X. Dong, and S. Pan, “Triformer: Triangular, variable-specific attentions for long sequence multivariate time series forecasting”, in *Proc. the Thirty-First International Joint Conference on Artificial Intelligence*, Jul. 2022, pp. 1994–2001. doi: 10.24963/ijcai.2022/277
- [61] J. Fan, Z. Wang, D. Sun, and H. Wu, “Sepformer-based models: More efficient models for long sequence time-series forecasting”, *IEEE Transactions on Emerging Topics in Computing*, pp. 1–12, 2022. doi: 10.1109/TETC.2022.3230920.

Copyright © 2025 by the authors. This is an open access article distributed under the Creative Commons Attribution License which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited (CC BY 4.0).