Artificial Intelligence Advances
https://journals.bilpubgroup.com/index.php/aia

ARTICLE

Optimizing Online Advertisement Services Predictions: A Data Analysis Approach with iTransformer and Periodicity Decoupling

Diwei Zhu[1,2], *Yunxiang Gan*[3*], *Xiaoyang Chen*[4]

[1] Hive AI, San Francisco, CA 94105, USA
[2] New York University, New York, NY 10003, USA
[3] Moloco, Redwood City, CA 94063, USA
[4] Radiawave Co., Ltd., Shen Zhen 518100, China

ABSTRACT

With the rapid development of the online advertising industry, improving the accuracy of advertising service predictions and optimizing ad placement strategies has become a critical research topic. Traditional forecasting methods often face challenges when dealing with complex and diverse advertising data, especially in handling temporal features and periodic fluctuations. To address this, this paper proposes a data analysis method based on iTransformer and a Periodic Decoupling Framework (PDF) to optimize online advertising service predictions. Without altering the Transformer network architecture, iTransformer innovatively transforms the functionality of the attention mechanism and feedforward network, treating different variables as independent tokens. This allows the model to effectively capture correlations between variables and temporal features, enhancing its ability to adapt to complex data. Meanwhile, the Periodic Decoupling Framework deeply explores periodic features in sales data, accurately separating regular variations, providing stronger support for long-term sequence forecasting. Finally, the introduction of self-supervised learning reduces reliance on labeled data, enabling the model to maintain strong generalization and performance even in data-scarce scenarios. Experimental results show that this method demonstrates superior performance in advertising service predictions, particularly in handling complex advertising data with periodic fluctuations.

*CORRESPONDING AUTHOR:
Yunxiang Gan, Moloco, Redwood City, CA 94063, USA; Email: yg281@scarletmail.rutgers.edu

ARTICLE INFO
Received: 2 September 2023 | Revised: 8 October 2023 | Accepted: 15 October 2023 | Published Online: 14 December 2023
DOI: https://doi.org/10.30564/aia.v5i1.7488

CITATION
Zhu, D., Gan, Y., Chen, X., 2023. Optimizing Online Advertisement Services Predictions: A Data Analysis Approach with iTransformer and Periodicity Decoupling. Artificial Intelligence Advances. 5(1): 49–62. DOI: https://doi.org/10.30564/aia.v5i1.7488

Keywords: iTransformer; Periodic decoupling framework; Online advertising service prediction; Self-supervised Learning; Data analysis; Time series prediction

1. Introduction

With the rapid development of the digital advertising industry, the prediction and optimization of online advertising services have become increasingly important [1]. In recent years, with the advancement of internet technology and the proliferation of big data, businesses face unprecedented opportunities and challenges in their advertising efforts [2]. Advertisers not only aim to attract target customers through precise ad placements but also hope to improve conversion rates and return on investment (ROI) [3]. Recent research on technological innovation and supply chain management shows that data-driven strategies can optimize decision-making across various sectors, including the advertising industry [4, 5]. Therefore, when formulating advertising strategies, companies must fully leverage extensive sales data, customer feedback, and market dynamics to achieve more effective advertising results. In recent years, certain technological applications have demonstrated that using data parallel acceleration methods can enhance model processing efficiency, a concept equally valuable for advertising data analysis [6]. Specifically, sales data includes quantitative indicators such as product sales, volume, and profit, as well as trends in customer purchasing behavior. By analyzing this data, businesses can identify which products perform well during specific time periods and what factors influence customer purchasing decisions. From the perspective of brand management and market feedback, shifts in customer preferences have also impacted sales models. Additionally, customer feedback provides more intuitive user experience data, including product reviews, satisfaction surveys, and discussions on social media [7, 8]. This feedback helps companies gain a deeper understanding of customer needs and preferences, enabling them to tailor ad content and delivery strategies accordingly [9, 10]. Despite having vast amounts of data, how to effectively analyze sales data in the financial domain and integrate it with customer feedback remains a pressing research challenge [11]. Similar studies in other fields have also shown that optimizing information integration is crucial for enhancing strategy effectiveness [12, 13]. In handling complex data, research outcomes such as the use of deep learning for character classification and noise reduction have also provided insights for optimizing advertising data processing [14, 15]. Many traditional data analysis methods struggle to fully uncover hidden patterns and associations within complex sales data. Additionally, cyclical characteristics within sales data, such as seasonal fluctuations and holiday effects, are often overlooked, leading to decreased accuracy in predictive models [16]. At the same time, dealing with large volumes of unlabeled data is another critical challenge in today's ad service prediction. Addressing these issues and developing a comprehensive solution that integrates sales data, customer feedback, and market dynamics will be key to improving the forecasting capabilities of online advertising services. This will not only help businesses devise more effective advertising strategies but also significantly enhance customer satisfaction and loyalty, allowing them to gain a competitive advantage in the market [17].

Current ad service forecasting methods often rely on traditional data analysis techniques, such as linear regression, decision trees, and time series analysis [18, 19]. While these methods provide basic predictive capabilities in some cases, they often fall short in extracting potential patterns and associations from complex advertising data. This is because traditional methods typically assume that relationships between data points are linear or simple, whereas real-world data is often influenced by multiple factors, including market trends, consumer behavior, and seasonal variations. As a result, these methods struggle to capture such complexities, leading to inaccurate and unreliable predictions [20]. In some fields, the effectiveness of using more complex models to capture nonlinear relationships has been validated, which is also relevant for advertising forecasting [21, 22]. Many existing methods also fail to adequately handle cyclical characteristics within sales data. Sales data often exhibits noticeable seasonal fluctuations, such as spikes in sales during holidays or changes in product demand during specific seasons. If these cyclical characteristics are not sufficiently recognized and modeled, it can directly affect the model's accuracy and stability in long-term forecasting. Traditional analysis methods often overlook these features, leading to poor model performance in predicting future changes and providing effective decision support for businesses. Another major challenge in current ad service prediction is how to handle vast amounts of unlabeled data [23, 24]. In practice, businesses often collect a large volume of user behavior data and feedback, but due to a lack of sufficient labeled data, traditional supervised learning methods struggle to be effective. Self-supervised learning, an emerging paradigm, has shown great potential by learning features from unlabeled data [25]. It can significantly improve model performance when labeled data is scarce, but effectively integrating self-supervised learning with traditional predictive models remains an important research topic. To address these issues, this paper proposes a data analysis method based on iTransformer and a periodicity decoupling framework [26, 27]. The innovative design of iTransformer treats different variables as independent tokens, enhancing its ability to model complex relationships between variables. This approach offers greater possibilities for advertising forecasting from the perspective of multidimensional data analysis, sharing similarities with data processing techniques in other fields [28, 29]. Additionally, the periodicity decoupling framework captures the cyclical characteristics within sales data, providing more robust support for long-term forecasting. By combining these two methods, this paper aims to significantly improve the prediction capabilities of online advertising services, not only enhancing model accuracy but also improving responsiveness to market dynamics, thereby helping businesses formulate more precise advertising strategies. The main objective of this study is to construct a comprehensive data analysis framework that combines the variable modeling advantages of iTransformer with the time-series feature capture capabilities of the periodicity decoupling framework. By conducting a systematic analysis of products and customer feedback, this paper aims to optimize the predictive models for online advertising, improving the match between customer needs and advertising effectiveness, ultimately enhancing customer satisfaction.

The contributions of this study are as follows: First, it proposes a novel iTransformer model that treats different variables as independent tokens, enhancing the ability to model the relationships between variables. Second, it applies the periodicity decoupling framework (PDF) to effectively capture complex periodic information within sales data, supporting long-term forecasting. Finally, it incorporates a self-supervised learning strategy, reducing reliance on labeled data and improving model performance in data-scarce scenarios. Through these innovations, this paper offers new perspectives and methods for predicting online advertising services, advancing research and application in the field.

The structure of this paper is as follows: In Section 2, we introduce related work, describing research methods in the field and discussing their advantages and disadvantages in ad service prediction and customer feedback. In Section 3, we present the main methods of this paper, such as the iTransformer architecture, the Periodicity Decoupling Framework (PDF), and the application of self-supervised learning. In Section 4, we discuss the experimental part, conduct comparative experiments, and present the results. In Section 5, we provide a discussion section, explaining the thought process behind this study and recent discussions in the field, while also pointing out the shortcomings of this approach. In Section 6, we conclude by summarizing the methodology and outlining future work.

2. Related work

Against the backdrop of the rapid development of digital marketing and advertising technology, the prediction and optimization of online advertising services have become a hot topic in research. When conducting advertising campaigns, companies typically rely on large amounts of historical data to analyze and identify potential customer demands and market trends. As the volume of data continues to increase, traditional data analysis methods face severe challenges, and there is an urgent need to develop more advanced technologies and methods to improve the accuracy and efficiency of advertising service predictions [30, 31]. Various methods have been proposed for the prediction of online advertising services.

In terms of traditional statistical methods, many early studies have adopted classical techniques such as linear regression, time series analysis, and moving averages. These methods played an important role in the initial stages of advertising service prediction, providing basic forecasting capabilities, but their limitations have gradually become apparent [32, 33]. Linear regression is one of the most basic statistical analysis methods, which predicts by establishing a linear relationship model between independent and dependent variables. The advantage of this method lies in its simplicity, ease of understanding, high computational efficiency, and suitability for small-scale datasets. Through linear regression, researchers can quickly identify key influencing factors and conduct preliminary trend analysis. However, the limitation of linear regression is that it assumes a linear relationship between independent and dependent variables, making it unable to capture complex nonlinear relationships [34]. Additionally, the model is highly sensitive to outliers, which can distort prediction results; when there is a high correlation between independent variables, the stability and predictive power of the model can also be affected. Time series analysis is another common traditional statistical method, predicting by analyzing the temporal dependencies in historical data. Common time series analysis techniques include Autoregressive Moving Average (ARMA) models and seasonal decomposition, among others [35, 36]. These methods can capture the temporal characteristics of data and perform well for data with significant periodic fluctuations. Although time series analysis fully considers the temporal order and trends in data and is suitable for both short- and long-term predictions, its limitations cannot be ignored. Choosing an appropriate time series model requires deep domain knowledge and experience, and many models assume that the data must meet certain stationarity assumptions, requiring additional preprocessing when handling non-stationary data. Moreover, when dealing with large-scale data, the computational complexity of time series analysis increases, reducing its efficiency. Moving average is a simple smoothing technique that calculates the average within a specific time window of the dataset to eliminate short-term fluctuations. This method is commonly used for preprocessing time series data to help identify long-term trends. The advantage of moving averages lies in their simplicity and effectiveness in reducing random fluctuations in data, helping to identify long-term trends. However, during the smoothing process, important short-term information may be lost, affecting prediction accuracy. At the same time, moving averages exhibit lag, which may result in a delayed response to market changes, and their adaptability is weak in rapidly changing market environments [37]. Although these traditional statistical methods provided basic analytical tools in the early stages of advertising service prediction, their linear assumptions, sensitivity to outliers, and limitations in handling complex data make them insufficient to meet the needs of companies in rapidly changing market environments. As a result, researchers urgently seek more advanced and flexible methods to improve the predictive capabilities of online advertising services.

With the development of machine learning technologies, an increasing number of studies have adopted methods such as decision trees, random forests, and support vector machines for predicting online advertising services [38]. These methods have shown better performance in handling complex data, capable of capturing nonlinear relationships within the data. For example, decision trees classify and regress by constructing a tree structure that intuitively represents the decision-making process and is suitable for handling data with multiple features. Random forests, as an important method in ensemble learning, significantly improve prediction accuracy and robustness by constructing multiple decision trees and performing voting. Support vector machines, on the other hand, classify data by finding the optimal hyperplane, which is particularly suitable for high-dimensional data. Although these models offer improvements in accuracy and flexibility, they still have some limitations. First, these models often require large amounts of labeled data for training, and their performance declines significantly in data-scarce environments. Additionally, these methods are sensitive to outliers and noisy data, which can lead to overfitting and impact the model's generalization ability. With the introduction of deep learning, new opportunities have emerged for online advertising prediction. Researchers have begun to use deep learning models such as Convolutional Neural Networks (CNNs) and Recurrent Neural Networks (RNNs) to process advertising data. CNNs excel in image recognition, and their ability to automatically extract features enables them to capture complex patterns, which is especially important when dealing with visual content in advertisements. RNNs, widely used in click-through rate prediction and user behavior analysis, are favored for their advantages in handling sequential data. However, deep learning models often require substantial computational resources and training time, which can be a bottleneck in resource-limited situations. Furthermore, deep learning's performance heavily depends on the availability of large datasets, and the model's training effectiveness may not be satisfactory in small-sample data scenarios [39]. In this context, the emergence of self-supervised learning and transfer learning provides new ideas for addressing the problem of scarce labeled data. Self-supervised learning designs pre-training tasks to allow the model to learn features from unlabeled data, thereby improving the model's generalization ability. Transfer learning enhances model performance in small-sample cases by transferring knowledge from existing models to new tasks[40]. These methods effectively alleviate the dependence on labeled data to some extent, advancing research in online advertising prediction. However, designing effective pre-training tasks and fine-tuning strategies remains a significant challenge in current research. To address these issues, researchers need to continuously explore more advanced technologies and methods to improve the predictive capabilities and performance of online advertising services.

3. Method

Figure 1 illustrates the overall architecture of the proposed model, which integrates iTransformer and the Periodicity Decoupling Framework (PDF) to improve the prediction accuracy of online advertising services. The architecture begins with an inverted embedding process, where input features such as sales, clicks, and time series data are embedded into independent tokens. The PDF module then decouples the input data into short-term and long-term components, effectively capturing periodic patterns. At the core of the model is the iTransformer block, which uses a multi-head attention mechanism to model complex dependencies between different variables over time. This block processes the temporal relationships in the data, allowing the model to focus on key features and interactions. After passing through the attention mechanism, a feed-forward network further refines the learned representations. Normalization layers ensure that data is standardized, reducing potential discrepancies across features. The PDF module is applied again to enhance the detection of periodic features before the output passes through the projection layer, which produces the final predictions. This architecture effectively combines the strengths of attention mechanisms and periodic analysis, allowing the model to handle complex time-series data and improve the accuracy of online advertising service predictions.

3.1 iTransformer

iTransformer introduces innovative designs based on the traditional Transformer to enhance the model's ability to capture multivariate time-series data and improve prediction accuracy for online advertising services and sales data analysis. Its overall framework is divided into three modules: input embedding, encoder, and projection layer. The algorithm architecture diagram is shown in **Figure 2**. In the input embedding module, features related to online ads and sales time-series data are fed into the iTransformer model. Unlike the traditional Transformer structure, iTransformer independently embeds the time series of each variable as a token. This design effectively preserves the relationships between various variables in advertising data, helping the model better understand customer behavior and market trends. The calculation formula for input embedding is as follows:

$$h_n^0 = Embedding(x_{(:,n)})$$

$$H^{(l+1)} = iTrBlock\left(H^l\right), l = 0, 1, \dots L-1$$

$$\hat{y}_{(:,n)} = Projection\left(h_n^L\right)$$

where $H = h_1, \dots, h_N \in R^{N\times D}$ contains N advertising feature embedding tokens of D dimensions. The superscript represents the layer index. $x_{(:,n)}$ represents the sales data and advertising features at the nth time step, and Embedding represents the embedding operation.

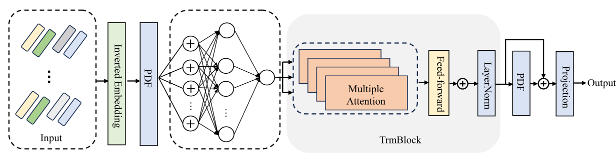

Figure 1. Overall algorithm architecture.

The encoder module utilizes multi-head attention mechanisms to compute the attention distribution of independently embedded multivariate data. This allows important features to be assigned higher weights, deeply mining the nonlinear features affecting ad performance.

$$Q^c = W^Q X$$

$$K = W^K X$$

$$V^c = W^V X$$

$$T_A(Q^c, K, V^c) = softmax\left(\frac{Q^c K^T}{\sqrt{d_k}}\right)V^c$$

$$T_M(Q^c, K, V^c) = concat\left(T_{head_1}, \dots T_{head_i}\right)W^o$$

$$T_{head_i} = T_A\left(Q_i^c, K_i, V_i^c\right)$$

Here, Q^c is the query matrix, $W^Q \in R^{d_{model} \times d_k}$, K is the key matrix, $W^K \in R^{d_{model} \times d_k}$, and V^c is the value matrix. These matrices can capture the relationships between advertising features and customer behavior. $W^V \in R^{d_{model} \times d_V}$ and $W^O \in R^{d_{model} \times d_k}$ are parameter vectors, with d_k and d_V representing the dimensions of keys and values, respectively, and d_{model} representing the input dimensions. Softmax denotes the activation function, concat represents the concatenation function, T_A denotes the attention distribution value, T_M denotes the multi-head attention distribution value, and $T_{h\ ead_i}$ denotes the ith attention distribution value.

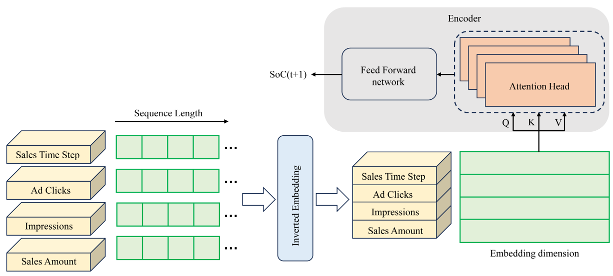

Figure 2. iTransformer model architecture diagram.

The feedforward neural network is used to reduce the overfitting effect caused by long time series, while the normalization layer is employed to standardize the data, reducing the impact of differences between various advertising features on the model.

$$T_{FFN}(x) = max(0, xW_1 + b_1)W_2 + b_2$$

$$X^{'} = T_L(X + T_{M(Q,K,V)})$$

$$X_{out} = T_L\left(X^{'} + T_{FFN}(X)\right)$$

Here, $X^{'}$ represents the normalized feature input, and x represents the elements in X. Max represents the activation function, T_L represents the normalization process, and T_{FFN} denotes the linear processing by the feedforward neural network. X_{out} represents the output matrix, W_1 and W_2 denote the model weights, and b_1 and b_2 are the bias terms. This process ensures consistency in advertising data input to the model, enhancing its prediction capability.

The projection layer module, composed of a multilayer perceptron (MLP), is responsible for performing nonlinear mapping outputs on the multivariate tokens processed by the encoder module. This module is designed to deeply represent learning for advertising features and sales data, ultimately helping the model generate accurate predictions. The iTransformer model effectively captures the relationships between advertising features and sales data by embedding each time step as an independent token. Through sequence encoding via the feedforward neural network, this paper aims to improve the model's representation learning and correlation modeling for multivariate time series, thereby enhancing the prediction accuracy of online advertising services. The innovative design of this architecture provides strong theoretical support for companies in formulating advertising strategies, promoting the development of digital marketing.

3.2 Periodicity decoupling framework

The Periodicity Decoupling Framework (PDF) is designed to optimize time-series modeling in online advertising service prediction and sales data analysis. By effectively decoupling short-term and long-term variations, PDF can improve the model's prediction accuracy and computational efficiency. The overall process can be divided into three modules: the multi-period decoupling block, the dual variation construction module, and the variation aggregation block. The architecture diagram is shown in **Figure 3**.

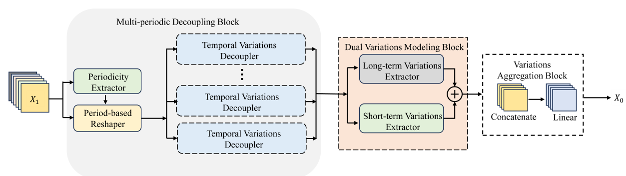

Figure 3. Periodicity Decoupling Framework model architecture diagram.

In the multi-period decoupling block, the first step is to decouple the one-dimensional time series into short-term and long-term sequences and reshape them into two-dimensional tensors based on the periods in the input sequence's frequency domain. The goal of this process is to capture the periodic fluctuations in ad click-through rates or sales data. The input to this module is the historical time series:

$$X_I = [x_1, x_2, \ldots, x_t]^T \in R^{t \times d}$$

Next, the Fast Fourier Transform (FFT) is used to analyze the time series in the frequency domain to extract frequency information:

$$A = Avg(Amp(FFT(X_I)))$$

Here, FFT represents the Fourier transform, Amp represents amplitude extraction, and Avg represents averaging over the d channels. $A \in R^t$ denotes the amplitude at each frequency. Based on the amplitude information, we select high-amplitude frequencies to capture the major components and construct the frequency set:

$$F_u = argtop - m(A)$$

$$F_{k_1} = argtop - k_1(A)$$

$$\{f_1, \ldots, f_k\} = F_{k_1} \cup top - k_2(F_u \backslash F_{k_1})$$

where F_u and F_{k_1} represent the top u and top k_1 frequencies with the highest amplitudes from A, respectively. We ensure that $u \geq k_1$. The final set of k frequencies is composed of the F_{k_1} set and the top k_2 frequencies from $F_u \backslash F_{k_1}$. Based on the selected frequencies $\{f_1, \ldots, f_k\}$ and corresponding period lengths $\{p_1, \ldots, p_k\}$(where $p_i = \frac{t}{f_i}$), the 1D input sequence $X_I \in R^t$ is reshaped into k 2D tensors. Period patching: The patch size is p and stride is s, and we segment $X^i_{2D} \in R^{f_i \times p_i}$ along dimension p_i, aggregating along dimension f_i to form a patch. Specifically, X^i_{2D} is divided into multiple patches $X^{i,j}_g \in R^{N \times P}$, where $N = \quad (p_i - p)/s \quad + 1$ is the number of patches, and each patch contains $P = f_i \times p$ time steps. $X^{i,j}_g$ denotes the j-th patch. This patching strategy condenses the complete long-term variations across all periods.

The dual variation construction module consists of a long-term variation extractor and a short-term variation extractor, using a dual-branch parallel architecture to model changes in the time series. The input to this module is the reshaped patch $X^{(i,j)}_g$. For long-term variation extraction, the patch is first projected into the latent space through linear projection:

$$X^{(i,j)}_g = linear\left(X^{(i,j)}_g\right) \in R^{N \times P}$$

Then, the patch is processed through several Transformer encoder layers:

$$\widehat{X}^{(i,j)}_g = Norm\left(X^{(i,j)}_g + MSA\left(X^{(i,j)}_g\right)\right)$$

where MSA represents the multi-head self-attention mechanism, and Norm represents the normalization operation. This process helps extract long-term variation features. In the variation aggregation block, all outputs from the dual variation construction module are merged to generate the final prediction output X_o:

$$X_o = Aggregation\left(\widehat{X}^{(1)}_g, \ldots, \widehat{X}^{(k)}_g\right)$$

This aggregation process effectively integrates the features of short-term and long-term variations, improving the model's overall prediction performance. The Periodicity Decoupling Framework, through the synergistic interaction of the multi-period decoupling block, dual variation construction module, and variation aggregation block, can effectively capture both short-term and long-term variation features in online advertising services and sales data. By performing frequency-domain analysis and structural reshaping of time series, PDF not only enhances the model's prediction accuracy but also optimizes computational efficiency. This framework provides robust theoretical support for formulating advertising strategies, helping businesses make more precise decisions in rapidly changing market environments.

3.3 Self-supervised learning

Self-supervised learning is a method that does not require manual labels, enabling feature learning by leveraging the structure and patterns in unlabeled data. When applied to online advertising service prediction and sales data analysis, self-supervised learning can effectively improve the model's generalization ability and prediction accuracy for complex data.

Self-supervised learning typically adopts a contrastive learning framework, where training is done by comparing the similarity between the original sample and its augmented counterpart. First, the sample is mapped to the feature space through an encoder f:

$$z = f(x)$$

Next, the similarity between the original sample z and the augmented sample z^\prime \) is calculated, usually using cosine similarity or inner product:

$$sim(z, z') = \frac{z \cdot z'}{\|z\|\|z'\|}$$

Self-supervised learning utilizes the structure and patterns in unlabeled data to help the model learn rich feature representations. When combined with online advertising service prediction and sales data analysis, self-supervised learning not only reduces the reliance on large amounts of labeled data but also enhances the model's generalization ability and prediction accuracy. Through effective data augmentation, contrastive learning, and fine-tuning strategies, self-supervised learning can provide strong support for optimizing advertising strategies.

4. Experiment

4.1 Experimental environment

All experiments were conducted on a machine equipped with an Intel Xeon Gold 6230 CPU running at 2.10 GHz, 256 GB of RAM, and an NVIDIA Tesla V100 GPU with 32 GB of memory. The operating system used was Ubuntu 20.04. The iTransformer and Periodic Decoupling Framework (PDF) models were implemented in Python using PyTorch (version 1.12.1), along with supporting libraries such as NumPy (version 1.21.6) and SciPy (version 1.7.3) for mathematical computations and data processing. The training and testing of the models leveraged the GPU for efficient matrix operations and parallelized data handling.

Hyperparameter tuning was performed using a grid search, exploring different configurations for learning rates, batch sizes, and attention head numbers. The final models were trained using a batch size of 512, a learning rate of 0.001, and 8 attention heads over 12 transformer layers. The models were trained over 100 epochs, with early stopping applied if no improvement was observed in the validation loss after 10 consecutive epochs.

4.2 Experimental data

- Amazon Product Reviews Dataset

The Amazon Product Reviews Dataset[41] is a large-scale collection of customer reviews on various products sold on Amazon. It includes detailed information such as product ID, customer ID, review title, review text, rating, and the timestamp of the review. This dataset is commonly used in natural language processing (NLP) tasks, sentiment analysis, and product recommendation systems to analyze customer preferences, satisfaction, and trends in product feedback.

- Amazon Sales Dataset

The Amazon Sales Dataset[42] contains comprehensive sales-related information about products sold on the Amazon platform. It includes sales metrics such as the number of units sold, revenue, customer purchase history, and time-stamped sales figures.

- Amazon Fine Food Reviews Dataset

The Amazon Fine Food Reviews Dataset[43] specifically focuses on reviews of gourmet food products available on Amazon. It contains over 500,000 reviews, including fields such as product IDs, user information, review scores, helpfulness ratings, and timestamps.

- Amazon Electronics Dataset

The Amazon Electronics Dataset[44] includes customer reviews and metadata for electronic products sold on Amazon. It covers items such as smartphones, laptops, and home appliances, providing data on customer ratings, review text, product descriptions, and sales rankings.

4.3 Evaluation metrics

To comprehensively evaluate the performance of the proposed iTransformer model and the Periodicity Decoupling Framework (PDF) in optimizing online advertisement service predictions, several key metrics are utilized: Mean Squared Error (MSE), Mean Absolute Error (MAE), Accuracy (ACC), and F1 Score. These metrics provide a detailed assessment of the model's performance in both regression and classification tasks.

- MSE

MSE measures the average squared difference between predicted and actual values for continuous variables such as sales amounts or ad impressions. It assigns a higher penalty to larger errors, making it useful for evaluating prediction precision in continuous data. The formula is:

$$MSE = \frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2$$

where y_i represents actual values, $\hat{y}_i$ represents predicted values, and n is the total number of data points. A lower MSE indicates that the model predictions are closely aligned with the actual values.

- MAE

MAE measures the average absolute difference between predicted and actual values. It offers a direct view of prediction error by focusing on the magnitude of errors without considering their direction. MAE is particularly useful when the goal is to understand how much the predictions deviate from actual outcomes. The formula is:

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

Lower MAE values suggest that the model provides more accurate predictions across the dataset.

- ACC

Accuracy evaluates the proportion of correct predictions out of all predictions made in classification tasks, such as predicting whether an advertisement leads to a click. It is defined as:

$$ACC = \frac{TP+TN}{TP+TN+FP+FN}$$

where TP is the number of true positives, TN is the number of true negatives, FP is the number of false positives, and FN is the number of false negatives. Higher accuracy reflects better overall classification performance.

- F1 Score

The F1 Score combines both precision and recall into a single metric, providing a balanced evaluation of a classification model's performance. It is particularly useful when dealing with imbalanced datasets, where both false positives and false negatives need to be minimized. The F1 Score is defined as:

$$F1 \text{ Score} = 2 \times \frac{Precision \times Recall}{Precision + Recall}$$

A higher F1 Score indicates that the model achieves a good balance between precision and recall, which is crucial for predicting critical events like ad clicks or conversions.

4.4 Experimental comparison and analysis

In order to evaluate the effectiveness of the proposed model, we conducted a series of experiments comparing its performance to several state-of-the-art models across four different Amazon datasets. The evaluation metrics used include Mean Squared Error (MSE), Mean Absolute Error (MAE), Accuracy (ACC), and F1 Score, providing a comprehensive understanding of the predictive and classification capabilities of each model.

Table 1 demonstrates the performance of several models across four Amazon datasets: Product Reviews, Sales, Fine Food Reviews, and Electronics. The proposed model (Ours) consistently outperforms other models on all datasets, achieving the lowest MSE and MAE while also delivering the highest ACC and F1 Score. For instance, in the Amazon Product Reviews Dataset, our model achieves an ACC of 0.942 and an F1 Score of 0.934, which are higher than the closest competitor, demonstrating better classification capability. Similarly, in the Amazon Sales Dataset, our model records an impressive ACC of 0.956 and an F1 Score of 0.941, reflecting its effectiveness in accurately predicting sales trends. The trend continues in the Amazon Fine Food Reviews and Amazon Electronics datasets, where our model maintains the highest ACC (0.937 and 0.941) and F1 Scores (0.946 and 0.954), highlighting its superior ability to capture intricate patterns in the data and provide accurate, balanced predictions. This demonstrates that the combination of iTransformer and the Periodicity Decoupling Framework (PDF) significantly enhances both prediction accuracy and classification performance across diverse datasets. **Figure 4** shows a visual comparison of model performance on four datasets.

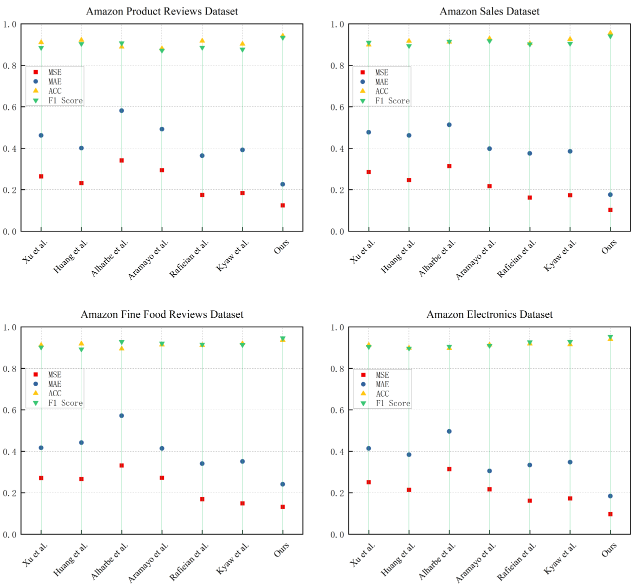

Figure 4. Visualization Comparison of Model Performance Across Four Amazon Datasets.

To further assess the computational efficiency of the proposed model, we compare the FLOPs, Inference Time, and Training Time across four datasets. These metrics provide insights into the model's resource consumption and processing speed during both the training and inference phases. The following table summarizes the comparison of training metrics for various models on the Amazon datasets, highlighting the computational advantages of our proposed model.

Table 2 highlights the performance of various models across four datasets by comparing three key indicators. The proposed model (Ours) demonstrates significant improvements in all three aspects. First, in terms of FLOPs, our model consistently shows the lowest values across all datasets, such as 3.53G on the Amazon Product Reviews Dataset, compared to higher values from models like Xu et al. with 6.06G and Aramayo et al. with 7.58G. This suggests that our model is more computationally efficient, requiring fewer operations for processing. Second, when looking at Inference Time, our model again leads with the fastest results, for instance, 216.74 ms on the Amazon Product Reviews Dataset, outperforming other models like Xu et al. (281.80 ms) and Kyaw et al. (234.17 ms). This demonstrates that our model is capable of making quicker predictions, an essential factor in real-time applications. Lastly, for Training Time,

Table 1. Performance Comparison of Different Models on Four Amazon Datasets.

Model	Amazon product reviews dataset				Amazon sales dataset			
	MSE	MAE	ACC	F1 score	MSE	MAE	ACC	F1 score
Xu et al. [45]	0.264	0.462	0.911	0.885	0.286	0.477	0.899	0.91
Huang et al. [46]	0.232	0.401	0.922	0.904	0.247	0.462	0.917	0.894
Alharbe et al. [47]	0.341	0.581	0.889	0.907	0.314	0.513	0.913	0.915
Aramayo et al. [48]	0.294	0.492	0.881	0.872	0.217	0.398	0.929	0.918
Rafieian et al. [49]	0.175	0.364	0.918	0.886	0.162	0.375	0.906	0.902
Kyaw et al. [50]	0.184	0.392	0.903	0.877	0.173	0.385	0.926	0.905
Ours	0.124	0.226	0.942	0.934	0.103	0.176	0.956	0.941

Model	Amazon fine food reviews dataset				Amazon electronics dataset			
	MSE	MAE	ACC	F1 score	MSE	MAE	ACC	F1 score
Xu et al. [45]	0.271	0.417	0.912	0.902	0.251	0.414	0.913	0.904
Huang et al. [46]	0.266	0.442	0.919	0.893	0.214	0.384	0.899	0.897
Alharbe et al. [47]	0.332	0.572	0.895	0.928	0.314	0.496	0.897	0.906
Aramayo et al. [48]	0.272	0.414	0.914	0.921	0.217	0.305	0.914	0.91
Rafieian et al. [49]	0.169	0.341	0.912	0.916	0.162	0.334	0.919	0.927
Kyaw et al. [50]	0.149	0.351	0.92	0.914	0.173	0.348	0.915	0.929
Ours	0.132	0.241	0.937	0.946	0.097	0.184	0.941	0.954

our model shows the shortest duration, such as 284.46 s on the Amazon Electronics Dataset, significantly faster than Rafieian et al. (389.41 s) and Alharbe et al. (357.08 s). This indicates that our model not only trains more efficiently but also can be iterated more quickly in practice. Meanwhile, **Figure 5** provides a visual comparison of these results.

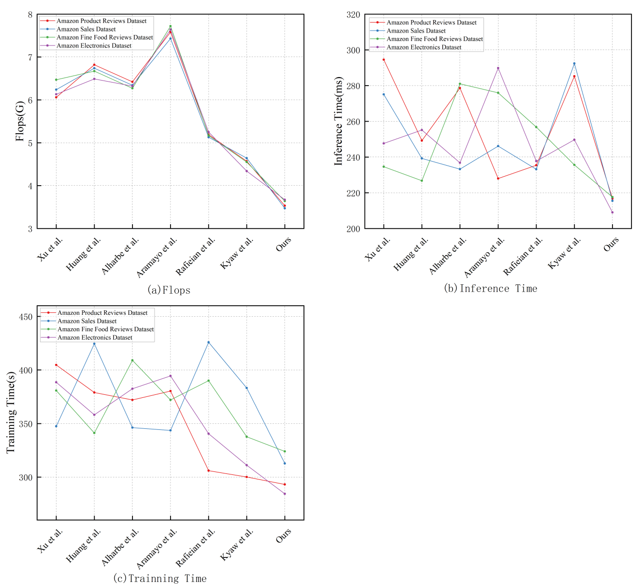

Figure 5. Visual Comparison of Training Metrics.

To better understand the contribution of each component to the model's overall performance, we conducted ablation studies on four datasets. These experiments evaluate the impact of the iTransformer and the Periodicity Decoupling Framework (PDF), both individually and in combination, by comparing the model's performance across various metrics. The following table presents the results of these ablation studies, highlighting how each addition improves the model's accuracy and reduces error rates.

Table 3 demonstrates the results of ablation experiments, showing how each component contributes to the model's performance. The baseline model shows the highest MSE and MAE, along with lower ACC and F1 Scores, indicating weaker performance. The addition of the iTransformer improves performance, reducing errors and increasing accuracy. Introducing the PDF (Periodicity Decoupling Framework) further boosts performance, particularly in reducing MSE and enhancing F1 Scores by capturing periodic trends in the data. The combination of iTransformer and PDF delivers the best results across all datasets, with the most optimal performance metrics: for example, in the Amazon Product Reviews Dataset, it achieves the lowest MSE (0.124) and MAE (0.226), along with the highest ACC (0.942) and F1 Score (0.934). Similarly, on the Amazon Sales Dataset, it attains an MSE of 0.103, MAE of 0.176, ACC of 0.956, and F1 Score of 0.941, showcasing its superior predictive and classification capabilities. **Figure 6** shows the trend of each metric.

Table 2. Comparative Analysis of Training Metrics Across Four Amazon Datasets.

Model	Amazon product reviews dataset			Amazon sales dataset		
	Flops (G)	**Inference time (ms)**	**Trainning time (s)**	**Flops (G)**	**Inference Time (ms)**	**Trainning time (s)**
Xu et al. [45]	6.06	281.80	361.23	6.24	277.24	425.12
Huang et al. [46]	6.82	251.24	315.62	6.74	291.47	372.49
Alharbe et al. [47]	6.42	267.44	393.83	6.34	268.66	386.35
Aramayo et al. [48]	7.58	220.17	336.85	7.43	279.39	357.77
Rafieian et al. [49]	5.19	258.00	315.05	5.13	289.48	425.82
Kyaw et al. [50]	4.57	234.17	353.34	4.64	295.59	375.06
Ours	3.53	216.74	293.14	3.47	215.46	312.73
Model	**Amazon Fine Food Reviews Dataset**			**Amazon Electronics Dataset**		
	Flops (G)	**Inference time (ms)**	**Trainning time (s)**	**Flops (G)**	**Inference time (ms)**	**Trainning time (s)**
Xu et al. [45]	6.47	224.38	373.41	6.13	270.86	302.62
Huang et al. [47]	6.67	254.75	393.19	6.49	220.04	306.74
Alharbe et al. [47]	6.27	234.67	367.86	6.32	246.06	357.08
Aramayo et al. [48]	7.72	228.12	395.51	7.64	270.68	305.65
Rafieian et al. [49]	5.17	245.19	396.55	5.25	268.09	389.41
Kyaw et al. [50]	4.55	286.02	365.32	4.34	222.94	334.38
Ours	3.64	217.56	324.06	3.67	208.97	284.46

Table 3. Results of Ablation Studies on Four Amazon Datasets.

Model	Amazon product reviews dataset				Amazon sales dataset			
	MSE	**MAE**	**ACC**	**F1 score**	**MSE**	**MAE**	**ACC**	**F1 score**
baseline	0.355	0.574	0.814	0.806	0.346	0.554	0.823	0.817
+iTransformer	0.284	0.468	0.869	0.873	0.267	0.473	0.884	0.862
+PDF	0.217	0.376	0.893	0.909	0.187	0.352	0.906	0.912
+iTransformer-PDF	0.124	0.226	0.942	0.934	0.103	0.176	0.956	0.941
Model	**Amazon fine food reviews dataset**				**Amazon electronics dataset**			
	MSE	**MAE**	**ACC**	**F1 score**	**MSE**	**MAE**	**ACC**	**F1 score**
baseline	0.324	0.527	0.811	0.821	0.331	0.539	0.832	0.822
+iTransformer	0.226	0.416	0.864	0.879	0.271	0.424	0.867	0.855
+PDF	0.159	0.311	0.882	0.914	0.171	0.317	0.902	0.916
+iTransformer-PDF	0.132	0.241	0.937	0.946	0.097	0.184	0.941	0.954

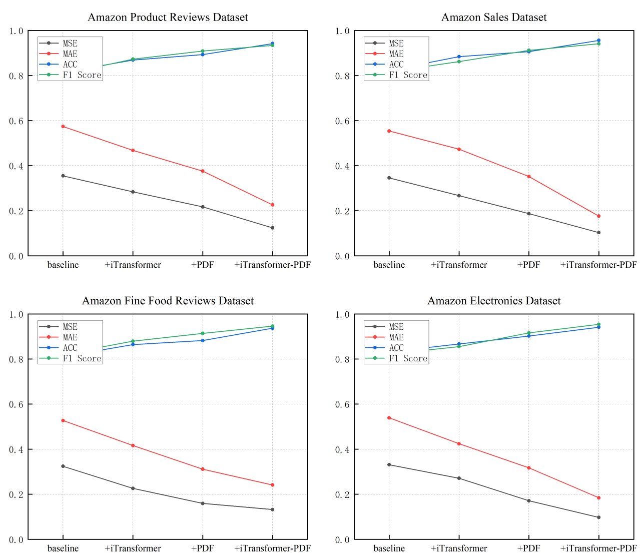

Figure 6. Visual Comparison of Ablation Experiments on Four Amazon Datasets.

5. Conclusion

In this paper, we present a novel model that combines the iTransformer with the Periodicity Decoupling Framework (PDF), enhanced by Self-Supervised Learning, to improve the accuracy of online advertising service prediction and sales data analysis. The iTransformer captures complex relationships and temporal features by treating different variables as independent tokens, while the PDF decouples short-term and long-term periodic patterns, providing robust support for accurate forecasting. Self-Supervised Learning, introduced into the model, helps reduce reliance on labeled data by leveraging the inherent structure of the data, allowing the model to generalize better in data-scarce environments. Experimental results across four Amazon datasets—Product Reviews, Sales, Fine Food Reviews, and Electronics—demonstrate that the proposed model significantly outperforms existing state-of-the-art methods in terms of MSE, MAE, Accuracy (ACC), and F1 Score. Notably, in the Amazon Sales Dataset, the model achieves an MSE of 0.103, MAE of 0.176, ACC of 0.956, and F1 Score of 0.941, showcasing its superior predictive accuracy. The combination of iTransformer, PDF, and Self-Supervised Learning proves to be highly effective in capturing both complex data patterns and periodic trends, as validated by the ablation experiments. Furthermore, the model demonstrates improved computational efficiency, with lower FLOPs, faster inference times, and shorter training times, making it ideal for real-time applications such as advertising optimization and sales forecasting. In the future, expanding the use of Self-Supervised Learning, integrating external data sources, and enabling real-time data processing will further enhance the model's predictive power and applicability in dynamic business environments.

References

[1] Haleem, A., Javaid, M., Qadri, M.A., et al., 2022. Artificial intelligence (AI) applications for marketing: A literature-based study. International Journal of Intelligent Networks. 3, 119–132.

[2] Losheniuk, I., Kabanova, O., Berger, A., et al., 2023. The future of virtual reality in marketing and advertising: benefits and challenges for business. Futurity Economics & Law. 3(3), 176–189.

[3] Bekh, A., 2020. Advertising-based revenue model in digital media market. 33(2), 547–559.

[4] Lei, J., 2022. Green Supply Chain Management Optimization Based on Chemical Industrial Clusters. Innovations in Applied Engineering and Technology. 1–17.

[5] Lei, J., Nisar, A., 2023. Investigating the Influence of Green Technology Innovations on Energy Consumption and Corporate Value: Empirical Evidence from Chemical Industries of China. Innovations in Applied Engineering and Technology. 1–16.

[6] Xiong, S., Zhang, H., Wang, M., et al., 2022. Distributed Data Parallel Acceleration-Based Generative Adversarial Network for Fingerprint Generation. Innovations in Applied Engineering and Technology. pp. 1–12.

[7] Elalem, Y.K., Maier, S., Seifert, R.W., 2023. A machine learning-based framework for forecasting sales of new products with short life cycles using deep neural networks. International Journal of Forecasting. 39(4), 1874–1894.

[8] Li, C., Tang, Y., 2023. The Factors of Brand Reputation in Chinese Luxury Fashion Brands. Journal of Integrated Social Sciences and Humanities. pp. 1–14.

[9] Zhao, Z., Ren, P., Yang, Q., 2023. Student Self-Management, Academic Achievement: Exploring the Mediating Role of Self-Efficacy and the Moderating Influence of Gender—Insights From a Survey Conducted in 3 Universities in America. Journal of Integrated Social Sciences and Humanities. pp. 1–12.

[10] Li, J., Yu, L., Gao, H., et al., 2011. Grouping-enhanced resilient probabilistic en-route filtering of injected false data in WSNs. IEEE transactions on parallel and distributed systems. 23(5), 881–889.

[11] Adigwe, C.S., Abalaka, A.I., Olaniyi, O.O., et al., 2023. Critical analysis of innovative leadership through effective data analytics: Exploring trends in business analysis, finance, marketing, and information technology. 23(22), 460–479.

[12] Chen, X., Zhang, H., 2023. Performance Enhancement of AlGaN-based Deep Ultraviolet Light-emitting Diodes with AlxGa1-xN Linear Descending Layers. Innovations in Applied Engineering and Technology. 1–10.

[13] Yu, L., Li, J., Cheng, S., et al., 2011. Secure continuous aggregation via sampling-based verification in wireless sensor networks. In Proceedings of The 2011 IEEE INFOCOM: IEEE. pp. 1763–1771.

[14] Xiong, S., Chen, X., Zhang, H., 2023. Deep Learning-Based Multifunctional End-to-End Model for Optical Character Classification and Denoising. Journal of Computational Methods in Engineering Applications. 1–13.

[15] Xiong, S., Zhang, H., Wang, M., 2022. Ensemble Model of Attention Mechanism-Based DCGAN and Autoencoder for Noised OCR Classification. Journal of Electronic & Information Systems. 4(1), 33–41.

[16] Cheng, C.C., Shiu, E.C., 2023. The relative values of big data analytics versus traditional marketing analytics to firm innovation: An empirical study. Information & Management. 60(7), 103839.

[17] Hossain, M.A., Akter, S., Yanamandram, V., et al., 2023. Data-driven market effectiveness: the role of a sustained customer analytics capability in business operations. 194, 122745.

[18] Almeida, A., Brás, S., Sargento, S., et al., 2023. Time series big data: a survey on data stream frameworks, analysis and algorithms. 10(1), 83.

[19] Wang, C., Ma, H., He, Y., et al., 2011. Adaptive approximate data collection for wireless sensor networks. IEEE Transactions on Parallel and Distributed Systems. 23(6), 1004–1016.

[20] Zhou, L., 2020. Product advertising recommendation in e-commerce based on deep learning and distributed expression. Electronic Commerce Research. 20(2), 321–342.

[21] Lei, J., 2022. Efficient Strategies on Supply Chain Network Optimization for Industrial Carbon Emission Reduction. Journal of Computational Methods in Engineering Applications. 1–11.
[22] Ren, P., Zhao, Z., Yang, Q., 2023. Exploring the Path of Transformation and Development for Study Abroad Consultancy Firms in China. Journal of Computational Methods in Engineering Applications. pp. 1–12.
[23] Alzubaidi, L., Bai, J., Al-Sabaawi, A., et al., 2023. A survey on deep learning tools dealing with data scarcity: definitions, challenges, solutions, tips, and applications. Journal of Big Data. 10(1), 46.
[24] Feng, Z., Deqiang, C., Xiong, S., et al. Method and apparatus for file identification. Ed: Google Patents. 2019.
[25] Rani, V., Nabi, S.T., Kumar, M., et al., 2023. Self-supervised learning: A succinct review. 30(4), 2761–2775.
[26] Liu, Y., Hu, T., Zhang, H., et al., 2023. iTransformer: Inverted transformers are effective for time series forecasting. arXiv:2310.06625.
[27] Yu, L., Li, J., Cheng, S., et al., 2013. Secure continuous aggregation in wireless sensor networks. IEEE Transactions on Parallel and Distributed Systems. 25(3), 762–774.
[28] Tang, Y., Li, C., 2023. Exploring the Factors of Supply Chain Concentration in Chinese A-Share Listed Enterprises. Journal of Computational Methods in Engineering Applications. 1–17.
[29] Tang, Y., Li, C., 2023. Examining the Factors of Corporate Frauds in Chinese A-share Listed Enterprises. OAJRC Social Science. 4(3), 63–77.
[30] Zhang, J., Liu, Y., Li, Z., et al., 2023. Forecast-assisted service function chain dynamic deployment for SDN/NFV-enabled cloud management systems. 17(3), 4371–4382.
[31] Xiong, S., Yu, L., Shen, H., et al., 2012. Efficient algorithms for sensor deployment and routing in sensor networks for network-structured environment monitoring. in 2012 Proceedings IEEE INFOCOM: IEEE, pp. 1008–1016.
[32] Luzia, R., Rubio, L., Velasquez, C.E., 2023. Sensitivity analysis for forecasting Brazilian electricity demand using artificial neural networks and hybrid models based on Autoregressive Integrated Moving Average. Energy. 274, 127365.
[33] Feng, Z., Xiong, S., Cao, D., et al., 2015. Hrs: A hybrid framework for malware detection. In Proceedings of The 2015 ACM International Workshop on International Workshop on Security and Privacy Analytics. pp. 19–26.
[34] James, G., Witten, D., Hastie, T., et al., 2023. Linear regression. in An introduction to statistical learning: With applications in python: Springer. pp. 69–134.
[35] Wang, W., Yildirim, G., 2021. Applied time-series analysis in marketing. in Handbook of market research: Springer. pp. 469–513.
[36] Fang, X.-L., Gao, H., Xiong, S.-G., 2012. RPR: High-reliable low-cost geographical routing protocol in wireless sensor networks. Journal of China Institute of Communications. 33(5).
[37] Kumar, A., Shankar, R., Aljohani, N.R. 2020. A big data driven framework for demand-driven forecasting with effects of marketing-mix variables. Industrial Marketing Management. 90, 493–507.
[38] Aldelemy, A., Abd-Alhameed, R., 2023. Binary classification of customer's online purchasing behavior using Machine Learning. 5(2), 163–186.
[39] Taye, M., 2023. Understanding of machine learning with deep learning: architectures, workflow, applications and future directions. 12(5), 91.
[40] Tripuraneni, N., Jordan, M., Jin, C., 2020. On the theory of transfer learning: The importance of task diversity. 33, 7852–7862.
[41] Ni, J., Li, J., McAuley, J., 2019. Justifying recommendations using distantly-labeled reviews and fine-grained aspects. in Proceedings of the 2019 conference on empirical methods in natural language processing and the 9th international joint conference on natural language processing (EMNLP-IJCNLP). pp. 188–197.
[42] Vishwakarma, S., Garg, D., Choudhury, T., et al., 2023. Amazon Sales Sentiment Prediction and Price Forecasting Using Facebook Prophet. in International Conference on Cyber Intelligence and Information Retrieval. Springer. pp. 93–105.
[43] Yarkareddy, S., Sasikala, T., Santhanalakshmi, S., 2022. Sentiment analysis of amazon fine food reviews. In Proceedings of The 2022 4th International Conference on Smart Systems and Inventive Technology (ICSSIT): IEEE. pp. 1242–1247.
[44] Yadav, V., 2022. Sentiment Analysis of Customer Reviews on Amazon Electronics Product: Natural Language Processing Approach and Machine Learning. Dublin, National College of Ireland.
[45] Xu, Z., Li, D., Zhao, W., et al., 2021. Agile and accurate CTR prediction model training for massive-scale online advertising systems. In Proceedings of The 2021 International Conference on Management of Data. pp. 2404–2409.
[46] Huang, L., Ma, Y., Liu, Y., et al., 2020. DAN-SNR: A deep attentive network for social-aware next point-of-interest recommendation. ACM Transactions on Internet Technology (TOIT). 21(1), 1–27.
[47] Alharbe, N., Rakrouki, M.A., Aljohani, A., 2023. A collaborative filtering recommendation algorithm based on embedding representation. Expert Systems with Applications. 215, 119380.
[48] Aramayo, N., Schiappacasse, M., Goic, M., 2023. A multiarmed bandit approach for house ads recommen-

dations. SSRN Electronic Journal. 42(2), 271–292.
[49] Rafieian, O., 2023. Optimizing user engagement through adaptive ad sequencing. Marketing Science. 42(5), 910–933.
[50] Kyaw, K.S., Tepsongkroh, P., Thongkamkaew, C., et al., 2023. Business intelligent framework using sentiment analysis for smart digital marketing in the E-commerce era. 16(3), e252965.