Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

USFF: A Unified Sales Forecasting Framework with Fourier-Enhanced Decomposed Net

Qianyang Li
Xi'an Jiaotong University
Xingjun Zhang
xjzhang@xjtu.edu.cn

Xi'an Jiaotong University
Shaoxun Wang
Xi'an Jiaotong University
Jiawei Cao
Xi'an Jiaotong University
Peng Tao
Shandong New Beiyang Information Technology Co., Ltd.

Research Article

Keywords: Deep learning, Sales forecasting, Multi-model fusion, Fourier decomposition

Posted Date: January 28th, 2025

DOI: https://doi.org/10.21203/rs.3.rs-5877784/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

Additional Declarations: No competing interests reported.

USFF: A Unified Sales Forecasting Framework with Fourier-Enhanced Decomposed Net

Qianyang Li[1], Xingjun Zhang[1*], Shaoxun Wang[1], Jiawei Cao[1], Peng Tao[2]

[1*]School of Computer Science and Technology, Xi'an Jiaotong University, Xi'an 710049, China.
[2]Shandong New Beiyang Information Technology Co., Ltd., Weihai 264200, China.

*Corresponding author(s). E-mail(s): xjzhang@xjtu.edu.cn;
Contributing authors: liqianyang@stu.xjtu.edu.cn;
shaoxunwang@stu.xjtu.edu.cn; caojw123@stu.xjtu.edu.cn;
taopeng@newbeiyang.com;

Abstract

With the widespread adoption of big data science, the retail industry has seen an increasing demand for data-driven decision-making. As a critical task within the sector, product sales forecasting faces significant challenges. Traditional statistical methods are limited by their reliance on assumptions of data stationarity, while DL and ML techniques enhance accuracy but struggle to fully address complex features such as temporal dependencies, nonlinear trends, and external influences. No single method proves universally effective across all scenarios. To tackle these challenges, this paper introduces a Unified Sales Forecasting Framework (USFF) for vending machine sales prediction, which integrates statistical methods, machine learning, and deep learning techniques. By classifying data based on statistical features and employing multi-model fusion strategies, the USFF framework effectively captures the complex nature of sales data, significantly improving prediction accuracy by leveraging the strengths of various models. Furthermore, the paper presents a novel deep learning forecasting model, FEDNet (Fourier-Enhanced Decomposed Net), designed to address the dynamic dependencies and seasonal patterns inherent in long time series. FEDNet offers substantial advantages in capturing long-term dependencies and dynamic changes in time series data through the integration of reversible instance normalization, seasonal-trend decomposition, and Fourier decomposition. Experiments conducted on over 30 million real-world data points, along with five public datasets, demonstrate that

the proposed framework and deep learning model outperform baseline methods across multiple evaluation metrics, confirming their effectiveness and reliability. The proposed framework and algorithm have been successfully implemented in a retail company in China, providing valuable business decision-making support for operational strategies.

Keywords: Deep learning; Sales forecasting; Multi-model fusion; Fourier decomposition

1 Introduction

Vending machines are a popular retail option in public settings due to their 24/7 availability. They not only provide consumers with the convenience of instant purchases but also offer operators a stable source of revenue [1]. Accurate sales forecasting is essential to fully maximize the operational potential of vending machines. High-precision sales forecasting can optimize inventory management and enhance operational profitability. In the era of big data, the integration of advanced technologies such as data analysis and machine learning, combined with real-time monitoring and external data integration, has significantly improved forecasting accuracy. This enables operators to make more informed and rational business decisions.

Due to the increase in data volume and advancements in machine learning methodologies [2], sales forecasting for vending machine products, as a typical time series prediction task, has surpassed the capabilities of traditional statistical methods (e.g., ARIMA) in terms of both precision and scope [3]. The complexity of modern sales forecasting necessitates advanced algorithms to process massive datasets, uncover hidden patterns, and integrate historical and real-time information for accurate predictions. Machine learning algorithms [4], such as random forests, decision trees, and gradient boosting frameworks like LightGBM [5], have markedly enhanced prediction accuracy by adeptly capturing the complex correlations between sales and numerous influencing factors. These models offer a detailed comprehension of sales trends, seasonality, cyclical elements, and external influences, including weather conditions.

In recent years, deep learning has achieved tremendous success in the field of sales forecasting [6–8], demonstrating powerful potential. Especially in complex business environments, deep learning models provide high-accuracy predictions, making them an increasingly essential decision-making tool for many enterprises. Meanwhile, the rapid development of deep learning technologies, particularly models based on the Transformer architecture (such as Informer [9], Autoformer [10], and FEDformer [11]). These models are effective in addressing issues such as long-term dependencies, periodic fluctuations, and dynamic changes. Deep learning has advantages in time series forecasting, such as automatic feature extraction, handling long-term dependencies, and modeling complex nonlinear relationships. It is better suited for large-scale data and dynamic changes driven by multiple factors.

A major challenge in vending machine sales forecasting is the complexity of the data and the diverse scenarios involved [12]. Sales data are influenced by a variety

of factors, including weather, geographic location, promotional activities, and holidays, and exhibit significant nonlinearity and seasonality [13]. Additionally, the wide range of products in vending machines, with varying sales fluctuations across different SKUs (Stock Keeping Units), makes it difficult for traditional single models to fully address the forecasting needs of all product categories. As operational scales expand, the demand for real-time forecasting becomes increasingly urgent. To support dynamic restocking decisions, forecasting models must not only ensure high prediction accuracy but also possess fast inference capabilities. While deep learning models offer improved accuracy, they are also burdened with limitations, such as the need for large datasets, substantial computational resources, poor interpretability, and sensitivity to noise. In conclusion, no single model is capable of effectively handling all the complexities inherent in these scenarios.

This research originates from a real-world business project aimed at achieving high-accuracy daily sales forecasting at the SKU level for vending machines. Grounded in the "No Free Lunch Theorem" (NFLT) [14], this study introduces the Unified Sales Forecasting Framework (USFF) to address the issue that no single model can handle the diverse nature of the data. Given the complexity and variability of vending machine sales data, traditional single models are inadequate to meet these challenges. Therefore, the USFF integrates traditional forecasting methods, machine learning models, and deep learning algorithms, allowing it to accommodate various data characteristics and forecasting tasks. Specifically, the framework classifies sales data based on features such as sales duration, volatility, and seasonality, selecting the most suitable algorithm for each category during training. This customized algorithm selection strategy enhances forecasting accuracy across diverse product categories and improves model generalization. Empirical analysis of actual sales records from Shandong New Beiyang Corporation (SNBC) demonstrates the superior performance of the USFF framework in sales forecasting.

Moreover, existing methods still face challenges in modeling complex dynamic dependencies and non-stationarity, particularly for products with long sales cycles and seasonal fluctuations. Recently, frequency transformation (FT) [15] has gained widespread attention in deep learning models, significantly advancing the accuracy and efficiency of state-of-the-art time series analysis. This paper introduces the FEDNet (Fourier-Enhanced Decomposed Network) model, which incorporates the concept of sequence decomposition. The model first enhances data stability and stationarity through reversible normalization (RevIN) [16], then applies Fourier decomposition to capture periodic features, and finally utilizes a Multi-Layer Perceptron (MLP) network for training. Experimental results from five publicly available datasets demonstrate that the FEDNet model outperforms baseline methods.

In summary, this paper presents novel forecasting frameworks and deep learning algorithms for vending machine sales prediction, contributing to the intelligent development of the new retail sector. The primary contributions of this study are summarized as follows:

- We propose the Unified Sales Forecasting Framework (USFF) to address the inherent complexity and diversity of vending machine sales data. By classifying data

features, USFF provides a unified platform for integrating deep learning techniques, machine learning models, and traditional time series forecasting methods. This framework has been successfully applied to vending machine sales forecasting, demonstrating substantial predictive accuracy on real-world business datasets and offering robust support for informed business decision-making.
- We introduce the novel FEDNet model, specifically designed to tackle the challenges of sales forecasting for products with long cycles and seasonal fluctuations. This model leverages reversible normalization, decomposes sales trends and seasonal components, and employs Fourier decomposition to extract periodic features from the seasonal elements. This methodology significantly enhances forecasting accuracy.
- We conducted comprehensive empirical validation of the proposed framework and model. Initially, we used a two-year sales dataset from SNBC vending machines to empirically verify that the USFF framework substantially improves forecasting performance. Additionally, tests on publicly available datasets further confirmed the superiority of the FEDNet model over baseline methods, achieving outstanding prediction results.

This paper is organized as follows: Section 2 reviews the relevant literature; Section 3 elaborates on the comprehensive forecasting framework and presents the novel deep learning model; Section 4 discusses the experimental results and their analysis; and Section 5 presents the conclusions.

2 Related work

The prediction of vending machine product sales is a highly complex and multi-dimensional problem, encompassing a diverse array of product categories sold across multiple vending machines in various geographic regions. As a specialized application of time series forecasting (TSF), vending machine sales forecasting presents unique challenges. The three primary methodologies for time series prediction—traditional statistical techniques, machine learning approaches, and deep learning models—serve as the foundational approaches for addressing this problem..

(1) Traditional Statistical Techniques

Traditional statistical forecasting methods rely on analyzing historical time series data to extract parameters and construct predictive models for future trends. Fundamental methods include univariate approaches such as the exponential smoothing method [17], Autoregressive Integrated Moving Average (ARIMA) models [18], and state-space models [19]. Among these, ARIMA and SARIMA (Seasonal ARIMA) [20] are widely used methods for time series forecasting. Fattah J [21] developed an ARIMA model to analyze and predict future demand in food manufacturing, employing the Akaike Information Criterion, Schwarz Bayesian Criterion, maximum likelihood estimation, and standard error. This model provides reliable decision-making guidance for managers. Traditional statistical models are simple and easy to use; however, such methods require time series data to be stationary and can only capture linear relationships in the data, failing to account for nonlinear patterns. Additionally, these

models rely solely on historical data, and with the increasing volume of data generated from business operations and external sources, such models are often inadequate for handling forecasting tasks in many scenarios.

(2) Machine learning-based methods

Retail sales forecasting is inherently complex as it involves diverse product categories and sales terminals across various scenarios. Notably, algorithms such as XGBoost [22] and LightGBM [5] have demonstrated outstanding performance in tackling the challenges of sales forecasting. By leveraging machine learning, sales prediction has entered a new era of artificial intelligence, significantly improving both accuracy and practical applicability. Machine learning models utilize both internal and external data sources (e.g., temperature, holidays) for sales forecasting. These models can capture relationships that extend beyond the capabilities of traditional statistical techniques. By learning patterns from training datasets (i.e., historical sales data), machine learning models apply these learned relationships to new datasets. Xu[23] proposed a multi-machine learning forecasting method based on historical sales data, which significantly enhanced prediction accuracy and provided critical theoretical support for Walmart's inventory management and supply chain optimization. Luiz Augusto [24] employed XGBoost and structural change correction approaches to address rapid customer behavior changes due to external reasons. This method exceeded the Base-Lift model, a popular retail forecasting benchmark, in accuracy and reduced stockouts and inventory. Chia-Wen Kao [25] employed a combination of FFT and Light Gradient Boosting Machine (LightGBM) with a weak learning strategy to manage irregular variations in sales data, achieving remarkable results in supermarket sales forecasting. Zhou [26]combined LightGBM with conventional time series decomposition methods to effectively oversee extensive product inventories across many retail locations. This approach employed machine learning and adaptive time series models to precisely identify complex demand patterns and respond to fluctuations driven by major trends and seasonality.

(3) Deep learning-based methods

As society advances and evolves, the volume of data accumulated by retail terminals has grown exponentially. Against this backdrop, leveraging deep learning techniques for sales forecasting has gradually emerged as a focal point of research, achieving remarkable progress in handling sequential data [27]. Given that historical sales records inherently constitute time series data, the academic community has shown increasing interest in applying deep learning methods to sales forecasting, leading to a growing body of related studies and applications.

Initial studies mostly utilized recurrent neural networks (RNNs) [28] and their derivatives, including LSTM and GRU. For instance, Kazuhi Honjo [29] proposed a model combining a pooling attention mechanism with GRU (PA-GRU), which utilizes convolutional neural network (CNN) layers to extract features and enhances predictive accuracy through two improved attention mechanisms. Similarly, Y Li [30] applied the Prophet model and an attention-based GRU model to extract both linear and nonlinear features from historical sales data, achieving more precise demand forecasts. Ahmed [31] uses several AI-based methods to forecast demand data. Moreover, Tseng et al. [32] evaluate deep learning algorithms for fashion industry sales forecasting,

focusing on new product sales for forthcoming seasons, utilizing a real-world dataset from a fashion retail company.

As deep learning architectures have advanced, new models have improved time series forecasting's efficiency and accuracy. Transformer-based models have gained a lot of traction in recent years due to their outstanding performance on a variety of AI tasks. As a result, scholars have begun utilizing Transformer techniques and their self-attention mechanisms for problems involving long-term time series forecasting (LTSF). Reformer [33] and Pyraformer [34] were among the early studies that concentrated on improving self-attention techniques. Subsequently, Autoformer [10] introduced a self-correlation mechanism to capture sequence-level temporal dependencies, while FEDformer [11]] used Fourier enhanced decomposition methods to explore time series data, shifting it from the time domain to the frequency domain, extracting deeper insights from long sequences. In sales forecasting, such models have also achieved notable breakthroughs. For example, Qianying Li [35] proposed a Transformer-based sales forecasting model utilizing an encoder-decoder architecture with multi-head attention. This model integrates particular alterations to the conventional Transformer, such as eliminating the Softmax layer in the final phase and modifying input embeddings, positional encoding, and feedforward network elements to accommodate the distinctive traits of sales data, thereby enhancing forecasting precision markedly.

Although the aforementioned methods have demonstrated promising results on their respective datasets, the inherent complexity of sales data in retail environments suggests that no single approach can achieve optimal performance across all scenarios. As a result, numerous scholars have proposed hybrid forecasting models, which have shown significant promise. For instance, Arnab Mitra [36] introduced a hybrid model (RF-XGBoost-LR) for retail chain sales forecasting by comparing five machine learning regression methods, including Extreme Gradient Boosting (XGBoost), Random Forest (RF), Adaptive Boosting (AdaBoost), Gradient Boosting, and Artificial Neural Networks (ANN). This model incorporates various factors affecting prediction accuracy, such as regional temperature and store size, which influence sales. Similarly, Supriyo Ahmed [31] proposed a Switching-Based Predicting Approach (SBFA), which combines advanced machine learning and deep learning techniques for enhanced prediction. Tran [37] developed the Lucy Hybrid model, a hybrid framework that integrates data preprocessing, model training, and evaluation metrics like MSE, RMSE, and MAE. This model evaluates and selects the best-performing models, enabling efficient storage and reuse, thus reducing training time. Experimental validation on a dataset from a large food retailer in Ecuador demonstrated the approach's effectiveness.

Overall, sales forecasting for vending machines is a challenging task. Whether using statistical forecasting, machine learning, or deep learning methods, each approach has its own distinct advantages and limitations. No single model can universally solve all forecasting problems [14]. In practical applications, it is crucial to consider the data's specific characteristics and to fully leverage the advantages of each model. To address these challenges, we propose the Unified Sales Forecasting Framework (USFF). The USFF framework classifies data based on its statistical features and employs a hybrid learning approach to optimize prediction accuracy across different SKUs, thus offering a more accurate and reliable solution for vending machine sales forecasting.

Furthermore, while Transformer-based methods have shown promising results for long-term time series forecasting, the study by DLinear [38] highlights a significant limitation of these approaches, which is the potential loss of temporal information. This raises questions about the effectiveness of novel Transformer-based methods for long-term forecasting tasks. In some instances, basic linear models may even outperform Transformer-based models. However, such models also face limitations, as their overly simplistic structure, relying on basic linear layers, struggles to capture the complex periodic patterns present in real-world datasets.

To overcome these challenges, we introduce a novel model, FEDNet (Fourier Enhanced Decomposition Network), which combines the advantages of invertible normalization techniques for handling data non-stationarity with the powerful periodic feature extraction capabilities of Fourier decomposition. By integrating these methods, FEDNet significantly improves model stability and forecasting accuracy.

3 Method

3.1 USFF

To address the limitations of single models in handling the data diversity of vending machine sales forecasting, this paper proposes a comprehensive forecasting framework—Unified Sales Forecasting Framework (USFF). This framework integrates deep learning methods, machine learning strategies, and traditional statistical techniques based on multi-model fusion strategies. As shown in Fig. 1, it consists of three stage: data preparation, model training, and model inference. By classifying data based on statistical features and matching them with appropriate models, the framework enhances the generalization performance and practical value of predictions. The objective is to enable scientific sales forecasting and operational optimization through data-driven algorithmic models.

3.1.1 Data Preparation

The system first completes the ① **data collection** process during the data preparation stage, which includes not only sales data but also external features. Based on the sales characteristics of self-service vending machines, we selected features with strong correlations, including device ID, product ID, time, sales volume, geographic location, weather, and other multidimensional information. At the same time, data cleaning, deduplication, anomaly detection, and missing value imputation are carried out to ensure data quality and consistency. Conventional sales forecasting techniques usually make predictions based on past sales data for a single device, but they frequently ignore the sales trends of comparable devices or products. The limitation of this approach lies in the failure to fully utilize the potential similarities between devices. We categorized all SKU sales data to improve the dataset and help the model collect data attributes.

Due to the inherent instability, volatility, and seasonality of vending machine sales data, the complexity of the data makes it challenging to achieve optimal predictive accuracy with a unified model. To address this challenge, the second step of USFF proposes a ② **data classification** method based on statistical features. Specifically,

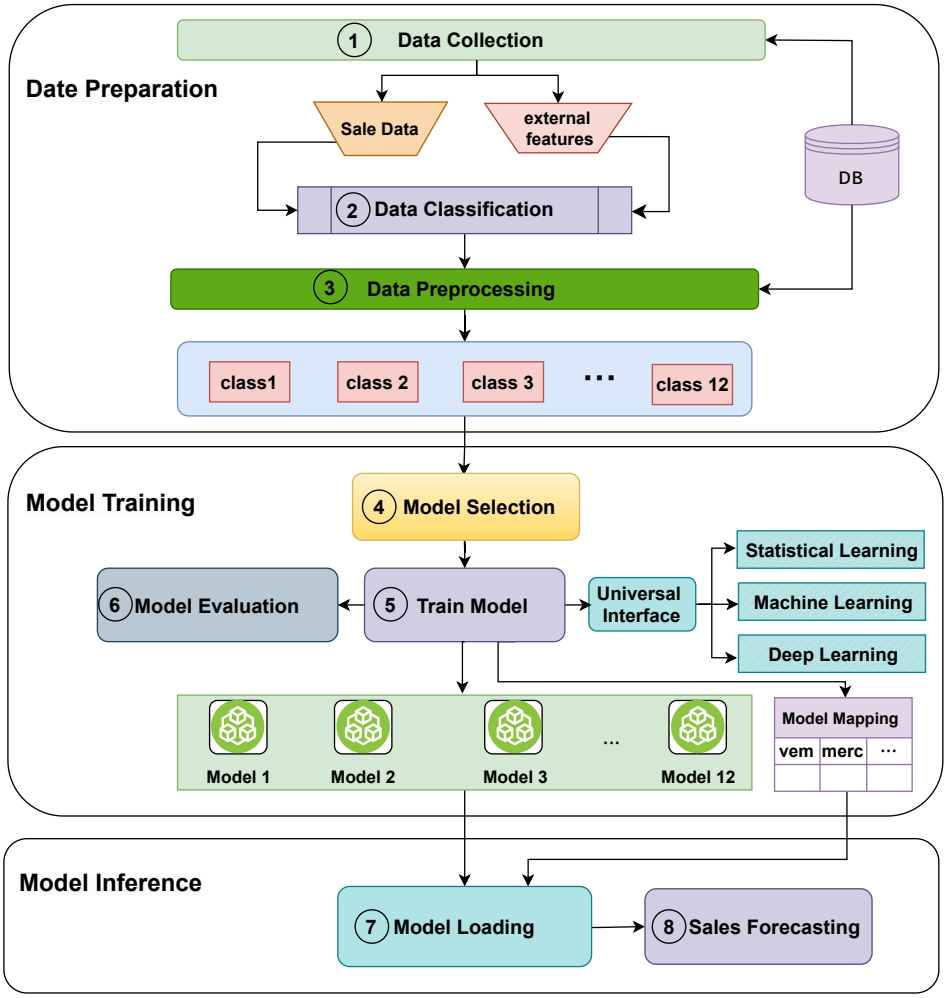


Fig. 1 Unified Sales Forecasting Framework.

the method calculates key statistical indicators such as the duration of sales data, CV coefficient, and seasonality strength to group sales data, effectively extracting classification features from the data.

The Coefficient of Variation (CV) measures dataset variability relative to the mean. It is frequently utilized for the purpose of comparing the dispersion of datasets that contain a variety of units or means.

$$CV = \frac{\sigma}{\mu} \tag{1}$$

where σ is the dataset's standard deviation and μ is its mean. CV helps compare dataset variability across scales and units. A higher CV denotes volatility, while a lower CV indicates stability and concentration around the mean.

Seasonality, a significant characteristic of vending machine sales, is defined in accordance with the approach outlined in [39]. The seasonal strength S_S is determined using the formula:

$$S_S = \max\left(0, 1 - \frac{\text{var}(R)}{\text{var}(X - T)}\right) \tag{2}$$

where X is a time series that could be divided into three parts: seasonality (S), trend (T), and the remainder (R): $X = S+T+R$ using Loess, a very flexible and strong technique for time series decomposition, to apply trend and seasonal decomposition [40].

Based on the diversity of data, we classified the daily sales data of SKUs considering statistical characteristics such as the duration of sales, volatility, and seasonality. Specifically, the duration of the sale is divided into three categories: short-term (15–60 days), medium-term (60–270 days), and long-term (greater than 270 days). For volatility, a coefficient of variation (CV) greater than or equal to 0.5 indicates high volatility, whereas a CV less than 0.5 signifies low volatility. With respect to seasonality, a seasonality strength greater than or equal to 0.5 suggests strong seasonality, while values below 0.5 indicate weak seasonality. Based on these definitions, we classified all sales data into 12 groups to fully capture combinations of different data characteristics. Sales data with durations shorter than 15 days, which have no predictive value, were excluded during the classification process.

Each data group corresponds to a category label with similar sales patterns, aiming to provide more precise input for subsequent model training and thereby improve prediction accuracy. This method not only addresses the complexity of vending machine sales data but also offers robust support for model selection and data preprocessing. Given the large number of SKUs and the multidimensional nature of vending machine data, this study avoids computationally expensive clustering algorithms (e.g., DBSCAN, KMeans) and instead employs a simplified classification method based on statistical features. This approach effectively reduces computational complexity while ensuring the adaptability of various predictive models, providing targeted data input for model selection and training.

After data classification is completed, further ③ **data preprocessing** is required to carry out feature engineering. Comprehensive treatment was applied to date features, sliding window features, lag features, and cyclical features to improve the models' capacity for prediction.

(1) Date Feature Extraction

Essential temporal attributes were derived from the date column, encompassing year, month, quarter, week number, day of the week, exact day, and holiday indicators. These attributes facilitate the identification of seasonality, periodicity, and daily trends within the time series. Year and month were employed to examine long-term patterns and seasonal fluctuations. Day of the week and holiday flags captured weekly patterns and the influence of specific holidays on sales. This level of granularity ensures the model is sensitive to both long-term and short-term temporal effects.

(2) Lag Features and Sliding Window Features

Lag and sliding window features were designed to capture dynamic patterns in time series data. Lag features shift the target variable (e.g., sales volume) to reveal

the influence of past values on the present, while sliding window features calculate rolling statistics (e.g., maximum, minimum, mean, standard deviation) over a fixed window to capture short-term volatility and trends. This method effectively analyzes variations within consecutive windows, helping the model better understand local time series patterns.

(3) Cyclical Feature Transformation

To preserve the intrinsic information of cyclical features in the time series, sine and cosine transformations were applied to periodic variables (e.g., day of the week). This transformation avoids the linear order assumption and represents cyclical features as continuous values. This encoding technique more successfully captures cyclical correlations, ensuring that the model fully utilizes the data's periodic character.

The culmination of the preprocessing phase is a classified dataset partitioned into training, validation and test sets to guarantee model efficacy and generalization ability. Through feature engineering, these datasets incorporate derived external factors, such as time features, lag features, and holiday identifiers, enhancing the overall quality of the data.

In conclusion, this structured preprocessing framework lays a solid foundation for building predictive models by providing high-quality, feature-rich datasets tailored to the complexity of vending machine sales data. By integrating domain knowledge into data preprocessing, the framework ensures the accuracy, reliability, and practicality of the forecasting results.

3.1.2 Model Training

The next phase is model training. First, based on the sales duration, volatility and seasonality characteristics of the previous step, along with the experimental results, we design a quantified model selector for automated ④ **model selection**. By quantifying the impact of each feature and considering the strengths and weaknesses of different models, this method can automatically recommend the most suitable forecasting model for various data environments.

For each statistical feature, scores are assigned to each model (e.g., SARIMA, LightGBM, and LSTM), and the total score for each model is computed. The scoring formula for each feature is given by:

$$\text{Score}_{\text{model}} = \sum_{i=1}^{n} w_i \cdot \text{score}_{i,\text{model}} \tag{3}$$

where w_i is the weight of feature i, indicating the importance of that feature in model selection; $\text{score}_{i,\text{model}}$ is the score of the model for feature i; and n is the number of features (in our case, $n = 3$ for Sales Duration, Volatility, and Seasonality). The weights w_1, w_2, w_3 and the model scores are set according to practical requirements and can be determined based on experiments or domain knowledge. For instance, in deep learning models, if sales duration is considered more critical for model selection, a greater weight may be assigned to this feature. In contrast, for SARIMA models, where seasonal factors are of paramount importance, a proportionally higher weight can be allocated to the seasonality component. Finally, the model with the highest

score is selected:

$$\text{Selected Model} = \arg\max_{\text{model}} (\text{Total Score}_{\text{model}}) \tag{4}$$

Once the appropriate model is selected, we proceed with the ⑤ **model training**. The system is designed with a unified interface that supports consistent input and output, enabling the training and evaluation of multiple algorithms to fully accommodate different data characteristics and business requirements. Specifically, the unified interface for model training is compatible with three types of methods:

Tranditional Statistical Methods: Tranditional models such as ARIMA and SARIMA are well-suited for time series data with clear seasonality or stable trends. These models effectively capture linear relationships in the data, providing reliable forecasts for datasets with relatively simple temporal patterns.

Machine Learning Methods: Algorithms like regression models, random forests, decision trees and XGBoost, are suitable for handling multi-dimensional data and nonlinear relationships between complex features. These methods have strong modeling capabilities and can uncover potential patterns from diverse datasets.

Deep Learning Methods: Advanced architectures, such as various Transformer and LSTM network structures, are particularly suitable for handling time series data with long-term dependencies. These methods can capture nonlinear trends and latent patterns, showing stronger fitting capabilities on large-scale datasets.

During the model training process, ⑥ **model evaluation** is required. Our system utilizes two performance metrics MSE and MAE to thoroughly evaluate the predictive accuracy of each model and implements cross-validation methods to guarantee the model's generalization capability. Based on this, the system selects the best-performing model for storage, providing convenient support for subsequent inference and deployment. This process not only ensures the efficiency and stability of the model in real business scenarios but also provides reliable technical support for the sales prediction tasks of the vending machines. After model training, the system outputs model files and a model mapping table, which records the corresponding model category and path for each SKU.

3.1.3 Model Inference

After completing model training, the process moves on to model inference, starting with the ⑦ **load model** step. The system loads the saved optimal model, inputs real-time data from vending machines along with external environmental factors (e.g., weather, promotions, holidays), and each SKU is matched to its respective model using the model mapping table generated during the training phase, completing the final ⑧ **sales forecasting**. This helps the operations team develop scientifically sound restocking plans and improve overall operational efficiency. In addition, the system leverages forecast results at different times and locations to provide data support for high-demand areas, enabling precise and refined operations of the vending machine network.

In engineering applications, the proposed predictive algorithm framework is integrated into big data systems, featuring high modularity, scalability, and real-time

inference capabilities. The system leverages the Ray framework for model deployment and employs Docker and Kubernetes to ensure scalability and maintainability. This integration enables parallel processing of large-scale vending machine inference requests while supporting algorithm model adjustments based on business needs. Additionally, with Continuous Integration/Continuous Deployment (CI/CD), the system supports online inference and dynamic model updates, allowing the models to adapt to changing data characteristics, thereby enhancing the system's adaptability and stability in real-world applications.

3.2 FEDNet

For products with long sales cycles and significant seasonal fluctuations, existing forecasting approaches still have space for development in terms of modeling stability and seasonality. To better capture these characteristics, this study introduces a sequence decomposition approach combined with Fourier decomposition techniques, proposing a new model—FEDNet (Fourier Enhanced Decomposition Network). The FEDNet architecture is depicted in Fig. 2.

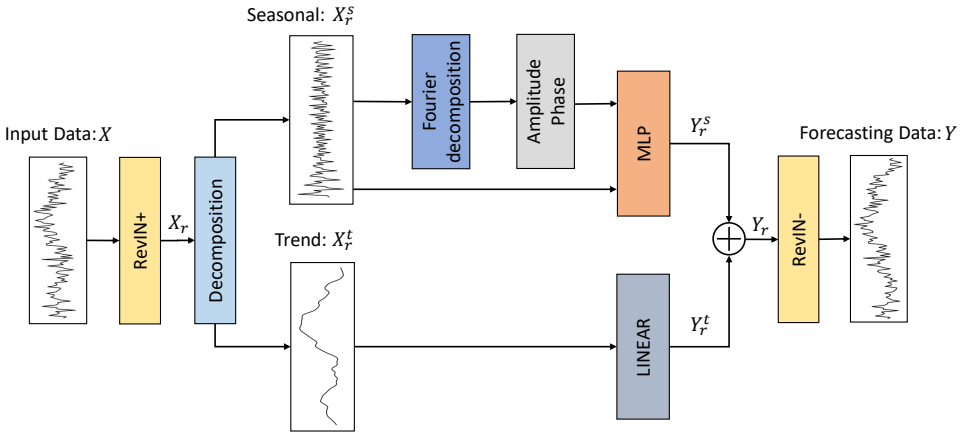


Fig. 2 The architecture of FEDNet.

The long-term sales forecasting of vending machine products is a typical application of the Long-Term Time Series Forecasting (LTSF) problem. The goal of this task is to predict future values based on historical multivariate observations, denoted as $X = (X^{(1)}, X^{(2)}, \ldots, X^{(N)}) \in \mathbb{R}^{L \times N}$ and $X^{(i)} = (x_i^{(1)}, x_i^{(2)}, \ldots, x_i^{(k)}) \in \mathbb{R}^L$, where L is the length of the lookback window, and N represents the number of feature dimensions (or channels). The objective is to forecast the values for the next H time steps, satisfying $H < L$. If the number of variables N is equal to 1, the problem simplifies to a univariate forecasting task.

Considering channel dependence, rather than assuming channel independence, can significantly improve the performance of multivariate time series forecasting, particularly on datasets with strong correlations [41]. When developing novel forecasting

methods, it is essential to fully leverage variable interactions to better capture the intrinsic structures and dynamics of the dataset. Given the relationships between our sales and feature data, we use a channel dependence method. Using channel dependencies can significantly improve model performance [39].

3.2.1 RevIN

FEDNet first employs Reversible Instance Normalization (RevIN) [16] to address the stepwise nature of the data. Zero-mean normalization is commonly used in Time Series Forecasting (TSF), but this method is incapable of eliminating non-stationary information from the data. Our model initially applies RevIn to the input forecast data, as RevIN has been shown to effectively eliminate and recover the non-stationary information mainly present in the trend. For each instance $X^{(i)} \in \mathbb{R}^L$ of the input data, the mean and standard deviation are calculated as

$$\begin{aligned} E_t\left[X^{(i)}\right] &= \frac{1}{L}\sum_{j=1}^{L} X_j^{(i)} \\ \mathrm{Var}\left[X_t^{(i)}\right] &= \frac{1}{L}\sum_{j=1}^{L}\left(X_j^{(i)} - E_t\left[X_t^{(i)}\right]\right)^2 \end{aligned} \tag{5}$$

We normalize the input data $X_t^{(i)}$ using these statistics as

$$\hat{X}_t^{(i)} = \gamma\left(\frac{X_t^{(i)} - E_t\left[X_t^{(i)}\right]}{\sqrt{\mathrm{Var}\left[X_t^{(i)}\right] + \epsilon}}\right) + \beta, \tag{6}$$

where the parameter vectors $\gamma, \beta \in \mathbb{R}^N$ are learnable. Subsequently, the model receives the normalized input data for forecasting. Finally, we apply the inverse normalization process to the model's output $\tilde{Y}_t^{(i)}$ to obtain the final forecast $Y_t^{(i)}$:

$$Y_t^{(i)} = \sqrt{\mathrm{Var}\left[X_t^{(i)}\right] + \epsilon} \cdot \left(\frac{\tilde{Y}_t^{(i)} - \beta}{\gamma}\right) + E_t\left[X_t^{(i)}\right] \tag{7}$$

In summary, by introducing reversible operations and time-step aware scaling parameters, RevIN solves the distribution shift problem the distribution shift problem that affects time series forecasting. Reversible Instance Normalization is an innovative method aimed at reducing the distribution shift phenomenon between training and testing datasets. This approach improves the model's accuracy for forthcoming data.

3.2.2 Seasonal-trend decomposition

Subsequently, FEDNet draws inspiration from Autoformer [10] by applying seasonal-trend decomposition after each neural block, a standard practice in time series analysis that enhances the predictability of the original data [42]. Long-term sales statistics for

vending machine products usually reflect both seasonal and trend features. Seasonality relates to recurrent patterns within specified time intervals, while trends represent steady development or decline over the long term.

To accomplish this decomposition, FEDNet employs a moving average kernel to extract the trend-cycle component of the time series from the input sequence. Initial decomposition of reversible normalized data X_r into trend X_r^t and seasonal component X_r^s is done using moving averages. The X_r^s is calculated by subtracting the trend component $X_{r,t}$ from the normalized raw series X_r:

$$\begin{aligned} X_r^t &= \text{AvgPool}(\text{Padding}(X_r)_{\text{kernel}} \\ X_r^s &= X_r - X_r^t \end{aligned} \tag{8}$$

where $X_r^t \in \mathbb{R}^L$ represents the trend term and $X_r^s \in \mathbb{R}^L$ represents the seasonal component. The advantage of seasonal-trend decomposition lies in its ability to simplify the complex structure of time series data into more manageable and interpretable components, significantly enhancing the effectiveness of data analysis. This decomposition method allows for the clear extraction of cyclical and long-term trend information from the time series, providing precise and valuable signals for subsequent modeling and forecasting. This is particularly crucial for long-sequence forecasting, as it not only helps in better understanding the temporal evolution of the time series but also lays a solid foundation for future modeling and prediction efforts.

Given the stability of the trend, we modeled the trend term series $X_r^t \in \mathbb{R}^L$ with a single linear layer to predict the future H-step forecast result $\hat{X}_r^t$, as shown in the following equation:

$$Y_r^t = \text{Linear}(X_r^t) \tag{9}$$

where $Y_r^t \in \mathbb{R}^H$ and H represents the forecasting horizon.

3.2.3 Fourier decomposition

Discrete Fourier Transform (DFT) is fundamental to signal processing. In time series analysis, for a signal in the time domain with a sequence length N and uniform time intervals, the DFT is used to obtain the frequency domain representation of the signal. The DFT is defined as:

$$\mathcal{F}[x[n]] = X[k] = \sum_{n=0}^{N-1} x[n] e^{-i\frac{2\pi}{N}nk} \tag{10}$$

Where $n = 0, 1, \ldots, N-1$ and $\mathcal{F}$ represents the computation process of the Discrete Fourier Transform (DFT), Where $x(n)$ is the value of the time-domain signal at time index n, $X(k)$ represents the frequency-domain components (complex coefficients) corresponding to the frequency index k, i is the imaginary unit (i.e., $i = \sqrt{-1}$), $\frac{2\pi}{N}$ is the fundamental frequency, and N is the total number of points in the time series (signal length).

The DFT converts the time domain signal $x(n)$ to its frequency domain representation $X(k)$, with each $X(k)$ including magnitude and phase information for the

associated frequency component. In practical applications, this transformation helps in analyzing the frequency characteristics of the signal, such as identifying periodic patterns or filtering noise.

DFT inverse transformation is computed as follows:

$$\mathcal{F}^{-1}[x[k]] = x[n] = \frac{1}{N}\sum_{k=0}^{N-1} X[k]e^{i\frac{2\pi}{N}nk} \tag{11}$$

Where $n = 0, 1, \ldots, N-1$ and $\mathcal{F}^{-1}$ represent the inverse process of the Discrete Fourier Transform. The result of the Discrete Fourier Transform is a sum of a series of sine or cosine signals. Given the known values of phase, amplitude, and frequency, a sine wave can be determined. In practical code implementation, the Fourier transform results are typically stored in complex number form. Let the result for a single frequency be (a, b), which is the complex number $a+bi$, representing a specific sine wave that makes up the original time series. The corresponding amplitude A and phase ϕ are computed using the following formulas:

$$\begin{aligned} A &= \sqrt{a^2 + b^2} \\ \psi &= \arctan\frac{b}{a} \end{aligned} \tag{12}$$

The analysis indicates that the Discrete Fourier Transform (DFT) defines time series from the signal's fundamental standpoint, with its frequency-domain representation offering a more intuitive and efficient way for feature extraction. In the frequency domain, various components of the time series can be flexibly manipulated, such as filtering by removing or retaining specific frequency components. By combining frequency-domain information with deep neural networks, the inherent patterns in the time series can be fully mined, thereby enhancing the model's information modeling ability.

Algorithm 1 Pipeline of FEDNet

Input: Time series data $X \in \mathbb{R}^{L\times D}$, future horizon T.
Output: Predicted time series Y.

$X_r \leftarrow RevIN^+(X)$ // Reversible normalization
$X_r^t, X_r^s \leftarrow \text{Decompose}(X_r)$ // Trend and seasonal components
$Y_r^t \leftarrow \text{Linear}(X_r^t)$ // Trend prediction
$\{s_1, \ldots, s_5\} \leftarrow \text{Fourier}(X_r^s)$ // Amplitudes and phases
$Y_r^s \leftarrow \text{MLP}([X_r^s, s_1, \ldots, s_5])$ // Seasonal prediction
$Y_r \leftarrow Y_r^t + Y_r^s$ // Weighted summation
$Y \leftarrow RevIN^-(Y_r)$ // Inverse reversible normalization

In summary, the entire workflow of FEDNet is shown in Algorithm 1. In FEDNet, the seasonal component from the decomposition undergoes Fourier decomposition to

extract low-frequency components (the first 5 coefficients). These are combined with the original seasonal signal X_r^s and fed into an MLP network for training, resulting in the predicted seasonal term Y_r^s. The periodic features, including the amplitude and phase of the first five frequency components obtained via Fourier Transform, are fused with the original seasonal component to enhance the model's ability to capture cyclical variations. This approach improves the model's capacity to predict periodic patterns, significantly boosting forecasting accuracy and reliability. Finally, the predicted trend term Y_r^t and seasonal term Y_r^s are added together, and the inverse normalization operation $RevIN^-$ from Equation 7 is applied to obtain the final forecast result Y.

4 Experimental results and analysis

This section presents an evaluation of the proposed algorithmic framework (USFF) by analyzing its performance in forecasting the actual sales of vending machines. Through the analysis of experimental results on the SNBC sales dataset, we verify the superiority of the classifier-based hybrid model in daily SKU-level sales forecasting for multi-terminal and multi-product scenarios. Furthermore, we also demonstrate through experiments on public datasets that our proposed FEDNet outperforms the baseline models. In addition, ablation experiments are conducted to validate the effectiveness of the proposed model in improving prediction accuracy for long-term forecasts. These studies provide significant insights for retail businesses dealing with vending machine products.

4.1 Experimental setup

4.1.1 Datasets

To evaluate the performance of the USFF framework, we conducted experiments using real-world transaction data from SNBC, a Chinese vending machine operator, and compared it with nine benchmark models. The dataset of SNBC covers sales records from 3,000 vending machines across China between January 1, 2023, and September 12, 2024. It includes data for 5,000 distinct products and more than 30 million transaction entries. To avoid the impact of COVID-19 on sales data, we specifically selected data after 2023. The Table 1 shows the characteristics of the data set. Table 2 shows the classification details of the 12 categories based on the features of sales duration, volatility, and seasonality as described in Section 3.1.1.

To better simulate real-world forecasting circumstances, the data was organized chronologically rather than randomly. To evaluate of model performance and facilitate fine-tuning in time series forecasting, the dataset was partitioned into training, validation, and testing sets using a 70:10:20 split. This approach ensured a more accurate evaluation of the model's operational performance.

In evaluating FEDNet, we referenced tasks from other types of LSTF. We selected five real-world datasets for extensive experimentation to assess the proposed model. These datasets are characterized by strong periodicity and long time spans [38], covering fields such as energy, weather, and transportation.

Table 1 Data characteristics of SNBC sales dataset

Attribute	Feature	Detail	Description
Time Span	-	2023/1/1 to 2024/09/12	Time span of the dataset
Data Volume	-	3,023 vending machines, sales data, and product sales volume	Covers vending machines, sales data, and product sales volume
Features	Vending Info	Machine ID	The ID corresponding to each vending machine
		Scenario	The scenario where the vending machine is located
		City Location	The city where the vending machine is located
	Sales Info	Sales Time	The timestamp of each sales record
		Product ID	The ID corresponding to each product
		Machine ID	The ID of the vending machine selling the product
		Daily Sales	The daily sales volume
		Product Price	The price of product
		Product Brand	The brand of product
		Product Category	The category of product
	Weather Info	Max Temperature	The daily maximum temperature
		Min Temperature	The daily minimum temperature
		Wind Speed	The daily wind speed
		City Name	City location information
	Holiday Info	Holiday Category	The type of holiday
		Is Holiday	Whether it is a holiday (non-working day)

ETT dataset: This dataset comprises two years of data collected from two distinct provinces in China. ETTm1 and ETTm2 collect data every 15 minutes.

Electricity dataset: This dataset, taken from the UCI Machine Learning Repository, consists of 321 customers' hourly electricity consumption records from 2012 through 2014.

Traffic dataset: This dataset offers information on road occupancy rates and is made up of hourly data from San Francisco Bay Area freeways.

Table 2 Classification details based on statistical features

Sales Duration	Volatility	Seasonal Strength	Category
Short-term	Low	Weak	1
Short-term	Low	Strong	2
Short-term	High	Weak	3
Short-term	High	Strong	4
Medium-term	Low	Weak	5
Medium-term	Low	Strong	6
Medium-term	High	Weak	7
Medium-term	High	Strong	8
Long-term	Low	Weak	9
Long-term	Low	Strong	10
Long-term	High	Weak	11
Long-term	High	Strong	12

Weather dataset: This dataset includes climate data from 1600 locations across the United States, sourced from the National Environmental Information Center.

The data characteristics of the five public datasets listed above are displayed in Table 3.

Table 3 The data characteristics of the five public datasets

DateSet	Length	Channel	Frequency	Start-time	End-time
Etthm1	17420	7	15min	2016-07-01	2018-06-26
Etthm2	69680	7	15min	2016-07-01	2018-06-26
Electricity	26304	321	1h	2012-01-01	2014-12-31
Traffic	17544	862	1h	2016-07-01	2018-07-02
Weather	52696	21	10min	2020-07-01	2021-01-01

4.1.2 Baselines

In evaluating the USFF, we selected three classic statistical methods: SARIMA [20], Prophet, and Holt-Winters; three widely used machine learning methods: LR (Linear regression), XGBoost [22], and LightGBM [5]; and four deep learning methods: FEDNet, DLinear [38], LSTM [28], and FEDformer [11].

SARIMA is a classical model for time series forecasting, integrating autoregressive (AR), differencing (I), moving average (MA) components, and seasonal components.

It is well-suited for modeling data with trends and periodic features. Its parameter selection requires manual tuning based on ACF and PACF plots, making it ideal for handling non-stationary data with clear time dependencies.

Prophet is an open-source tool for time series forecasting designed by Facebook in 2017, which is specifically designed for handling time series data characterized by strong seasonality and pronounced trend variations. Based on an additive model, Prophet automatically handles missing values, holiday effects, and outliers, and allows users to easily adjust model parameters and assumptions.

Holt-Winters is a weighted moving average model used to capture trends and seasonal components in time series data. It provides simple smoothing predictions through either an additive or multiplicative model, but it performs weakly when dealing with complex nonlinear trend data.

Linear Regression(LR) is a straightforward statistical model that represents the relationship between the target variable and predictors through a linear function. While computationally efficient, it lacks the ability to model nonlinear features or seasonal patterns commonly found in time series data.

XGBoost is a powerful ensemble learning algorithm built on the principle of gradient boosting, renowned for its computational efficiency, scalability, and high accuracy. It can handle nonlinear features and various types of data, but for time series problems, feature engineering is necessary to fully leverage its potential.

LightGBM is another gradient boosting model known for its fast training speed and low memory usage. It is well-suited for large-scale data and can efficiently handle both categorical and continuous features. However, when applied to time series forecasting, feature engineering is necessary to capture time dependencies effectively.

FEDformer introduces the Frequency Enhanced Decomposition Transformer (FEDformer) method, which combines Transformer architecture with frequency domain analysis to address the computational expense and difficulty in capturing global views in long-term time series forecasting.

DLinear explores a Transformer-based Long-Term Series Forecasting (LTSF) approach and questions the effectiveness of this research direction. The authors present a straightforward one-layer linear model, named LTSF-Linear, for comparative analysis. Experimental data indicate that LTSF-Linear substantially surpasses current intricate Transformer-based Long-Term Sequence Forecasting models applied on nine real-world datasets.

LSTM was developed to overcome the long-term dependency limitations of traditional recurrent neural networks (RNNs) by incorporating gating mechanisms. These mechanisms enable LSTM to effectively capture and model long-term sequential relationships in time series data. Renowned for its versatility, the model has achieved outstanding performance across various tasks, establishing itself as a foundational tool in sequence modeling.

To assess the performance of FEDNet, we conducted a comparison with various advanced Transformer-based models. These models include Pyraformer [33], FEDformer , Informer [9], and Autoformer [10]. Additionally, we included the recent linear model DLinear [38] and the classical LSTM model as the baseline for comparison. FEDformer, DLinear, and LSTM are described in the previous paragraph.

Pyraformer uses a multiresolution representation and pyramidal attention module, offering high accuracy and efficiency in both short and long-range forecasting.

Autoformer introduces a novel decomposition architecture called Autoformer, which incorporates an Auto-Correlation mechanism to address long-term time series forecasting issues. By combining time series decomposition with internal deep models, Autoformer is capable of stepwise decomposition for complex time series.

Informer is an effective Transformer-based model designed specifically to address long-sequence time series forecasting issues. It integrates the ProbSparse self-attention mechanism and employs self-attention distillation to enhance the processing of input sequences. Additionally, the model features a generative decoder that accelerates long-sequence prediction.

4.1.3 Evaluation metrics

In evaluating the predictive performance of the models, this paper employs two primary metrics: Mean Absolute Error (MAE) and Mean Squared Error (MSE). These estimations reflect the deviation between predicted and actual values to varying degrees, as well as the dispersion between them. For both MAE and MSE, lower values indicate minimal difference between expected and actual values, indicating reduced dispersion. The formulas for MAE and MSE are as follows:

$$\begin{aligned} \text{MAE} &= \frac{1}{n}\sum_{i=1}^{n}|y-\hat{y}| \\ \text{MSE} &= \frac{1}{n}\sum_{i=1}^{n}(y-\hat{y})^2 \end{aligned} \tag{13}$$

where y is the actual value and $\hat{y}$ is the predicted value for each observation i, with n being the total number of observations.

4.1.4 Parameter Settings

In the experiment of USFF, the main hyperparameter settings of the 10 models are shown in Table 4.

In the FEDNet experiment, which is based on the conventional LTSF task, all models adhere to identical experimental parameters and are utilized for four distinct future prediction durations $T \in \{96, 192, 336, 720\}$. For Transformer-based models, the default lookback window $L = 96$ is used, while for DLinear and FEDNet, $L = 336$ is used. In our experiment, We adopt the Adam optimizer, using a batch size of 32 and an initial learning rate of 0.0001.

Each experiment is conducted five times to ensure reliability. All methods are implemented on a system equipped with dual Intel(R) Xeon(R) Gold 5218 CPUs and a single NVIDIA 4090 GPU, running on Ubuntu 20.04.4.

Table 4 The hyperparameters in USFF

Method	Algorithm	Hyperparameter
Statistical method	SARIMA	autoregressive terms p = 1, order of differencing d = 1 seasonal period = 7
	Prophet	weekly_seasonality = True
	Holt-Winters	seasonal_periods = 7 seasonal = 'mul' trend = 'add'
Meachine Learning	Linear Regression	fit_intercept = True
	LightGBM	boosting_type = gbdt objective = tweedie num_leaves = 256 subsample = 0.65 learning_rate = 0.015
	XGBoost	boosting_type = gbdt objective = regression num_leaves = 256 max_depth = 10 subsample = 0.65 learning_rate = 0.015
Deep Learning	FEDNet / DLinear	batch_size =128 lookback_window=49 prediction_durations=14
	FEDformer	num_heads = 8 d_model = 256 hidden_size = 1024 batch_size = 128 num_layer = 4 lookback window=49 prediction_durations=14
	LSTM	num_layer = 3 hidden_size = 128 learning_rate = 0.005

4.2 Experiment Result

This subsection presents USFF's experimental results on SNBC's real sales dataset and FEDNet's experimental results on public datasets.

4.2.1 USFF Result

For the classification of the aforementioned 12 dataset categories, we conducted a comparison using 10 methods from three major types of models. The experimental results are shown in Table 5 and 6:

Table 5 The experimental results of USFF from Category 1 to Category 6

		Statistical methods			**Machine learning**			**Deep learning**			
Category	Metric	SARIMA	Prophet	Holt-Winters	LR	LightGBM	XGBoost	FEDNet	Dlinear	LSTM	FEDformer
Category1	MSE	**1.1338**	1.3363	1.3887	1.6257	1.5353	1.8127	1.7728	1.7321	1.8248	1.6521
	MAE	**0.7142**	1.0714	0.9286	1.2143	1.1581	1.4286	1.4286	1.2857	1.6286	1.3429
Category2	MSE	**1.2817**	1.6036	1.2536	1.6036	1.3887	1.8127	1.3628	1.3363	1.7571	1.6355
	MAE	**0.9286**	1.0271	1.0213	1.1429	1.0714	1.4286	1.1429	1.0714	1.5412	1.2571
Category3	MSE	5.1824	4.3916	4.062	2.1557	**1.8112**	1.8676	2.1318	2.0043	1.9495	1.7806
	MAE	4.1429	3.4286	3.5001	1.8166	**1.6281**	1.6250	1.9174	1.8259	1.7569	1.6229
Category4	MSE	5.8919	6.2163	5.1475	1.7271	**1.5529**	1.6914	1.8393	2.0817	1.7857	1.8610
	MAE	4.4286	4.7857	4.4286	1.3455	**1.2199**	1.3141	1.6429	1.8442	1.6119	1.6428
Category5	MSE	3.7702	3.151	3.7607	1.6430	**1.2028**	1.4230	1.9329	1.8311	1.7559	1.7466
	MAE	3.2143	2.7857	3.1429	1.4083	**1.0921**	1.3141	1.8259	1.7344	1.7121	1.6000
Category6	MSE	7.746	8.0178	7.1714	2.0129	2.1104	**1.8159**	2.1217	1.9787	2.0271	1.8503
	MAE	6.5714	7.6152	6.1429	1.6596	1.9046	**1.5340**	1.9393	1.7710	1.8954	1.7308

The optimal result is highlighted in bold.

Table 6 The experimental results of USFF from Category 7 to Category 12

		Statistical methods			**Machine learning**			**Deep learning**			
Category	Metric	SARIMA	Prophet	Holt-Winters	LR	LightGBM	XGBoost	FEDNet	Dlinear	LSTM	FEDformer
Category7	MSE	7.3193	5.3012	5.2222	2.1207	**1.5352**	1.6699	1.8690	1.7885	2.0907	1.9053
	MAE	6.5781	4.6429	4.2857	1.7915	**1.3769**	1.5025	1.6978	1.6429	1.8843	1.6978
Category8	MSE	6.3358	5.757	5.5997	1.7672	**1.5703**	1.8159	2.1948	1.8899	2.2137	1.9053
	MAE	5.6912	4.7143	4.5321	1.5842	**1.4397**	1.5340	2.0052	1.7893	1.9906	1.6978
Category9	MSE	7.6017	7.3387	7.1614	1.6498	**1.4448**	1.4694	1.9743	1.9451	2.2070	2.1727
	MAE	7.0714	5.4286	6.5714	1.3455	**1.1256**	1.2199	1.8259	1.7930	1.9906	1.9723
Category10	MSE	4.4159	4.7585	4.6675	2.2000	1.9264	2.1228	**1.4895**	1.6113	2.2372	1.6936
	MAE	3.6429	4.0714	3.9286	1.8794	1.6596	1.7538	**1.2416**	1.3891	1.6286	1.4143
Category11	MSE	3.4949	5.2801	2.7903	1.2847	**1.2354**	1.3046	1.9524	2.1309	2.5000	2.0020
	MAE	2.6429	4.7143	2.2143	1.1759	**1.0628**	1.0942	1.8259	1.9686	2.1404	1.8726
Category12	MSE	3.4017	5.5613	3.4538	1.7870	1.6376	1.7623	**1.5019**	1.6791	2.5806	1.7903
	MAE	2.9071	4.6429	2.7857	1.4720	1.4397	1.5025	**1.1307**	1.2681	1.7143	1.4714

The optimal result is highlighted in bold.

According to the experimental results, the performance of different models varies significantly across different categories of data. Statistical learning methods (such as SARIMA, Prophet, and Holt-Winters) exhibit certain advantages when applied to relatively short, low-volatility time series, especially SARIMA, which demonstrates the lowest MAE and MSE values in Category 1 and Category 2. These methods are well-suited for short-term and clearly periodic scenarios. However, as data complexity increases, the effectiveness of these methods notably decreases, particularly for non-stationary data (e.g., Category 7 and Category 8), where their performance is inadequate.

Conversely, machine learning techniques (like LightGBM, XGBoost, and LR) exhibit significant adaptability and resilience, particularly in more complex categories like Category 3 to Category 9, where their nonlinear modeling capabilities significantly enhance prediction accuracy. Among these, LightGBM and XGBoost perform excellently in most scenarios, making them suitable for data with rich features, noise, or nonlinear relationships.

Furthermore, deep learning methods (such as FEDNet, Dlinear, LSTM, and FEDformer) gradually demonstrate their superiority on long-term dependencies or complex pattern data. Particularly, FEDNet demonstrates the best performance in Category 10 and Category 12, reflecting its ability to effectively model long-duration and highly periodic time series.

Overall, model selection should be tailored to the data's characteristics. For short sales periods with stable and regular time series, such as Categories 1 and 2, statistical learning methods (e.g., SARIMA) deliver the best results. For data with more nonlinear or complex features, such as Categories 3 through 9, machine learning methods (e.g., LightGBM) outperform others with superior computational efficiency. For data with long-term dependencies and strong periodicity, such as Categories 10 and 12, deep learning models (e.g., FEDNet) demonstrate optimal performance, making them ideal for handling multidimensional time series or capturing global trends. Based on the above experimental results, we made corresponding adjustments to the feature weights in the model selector and the model scores to ensure that the selected model is optimal. Ultimately, the above three models were chosen and applied to the SNBC retail operation system, providing data support for business decision-making.

4.2.2 FEDNet Result

On five public datasets, our FEDNet was compared with six deep learning methods: Dlinear, FEDformer, Autoformer, Informer, Pyraformer, and LSTM models. The experimental results of Transformer-based and DLinear models are sourced from DLinear [38]. A lower MSE or MAE indicates better performance. The comparison results across different forecasting horizons (96, 192, 336, 720) on five public datasets are shown in Table 7. IMP represents the percentage improvement of the best result over the second-best result.

Table 7 Forecasting results on five datasets. The best results are highlighted in bold, while the second-best results are underlined.

Method		**IMP**		**FEDNet** (ours)		**Dlinear** (2023)		**FEDformer** (2022)		**Autoformer** (2022)		**Informer** (2022)		**Pyraformer** (2022)		**LSTM** (1997)	
Metric		MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE
Electricity	96	6.43 %	5.06 %	**0.131**	**0.225**	0.140	0.237	0.193	0.308	0.201	0.317	0.274	0.368	0.386	0.449	0.375	0.437
	192	7.19 %	4.42 %	**0.142**	**0.238**	0.153	0.249	0.201	0.315	0.222	0.334	0.296	0.386	0.386	0.443	0.442	0.473
	336	8.28 %	4.49 %	**0.155**	**0.255**	0.169	0.267	0.214	0.329	0.231	0.338	0.300	0.394	0.378	0.443	0.439	0.473
	720	3.45 %	3.99 %	**0.196**	**0.289**	0.203	0.301	0.246	0.355	0.254	0.361	0.317	0.449	0.376	0.445	0.980	0.814
Traffic	96	1.95 %	5.67 %	**0.402**	**0.266**	0.410	0.282	0.587	0.366	0.613	0.388	0.719	0.391	2.085	0.468	0.559	0.549
	192	4.26 %	5.57 %	**0.405**	**0.271**	0.423	0.287	0.604	0.373	0.616	0.382	0.696	0.379	0.867	0.467	0.843	0.453
	336	5.73 %	5.07 %	**0.411**	**0.281**	0.436	0.296	0.621	0.383	0.622	0.337	0.777	0.420	0.869	0.469	0.847	0.453
	720	7.51 %	4.13 %	**0.431**	**0.302**	0.466	0.315	0.626	0.382	0.660	0.408	0.864	0.472	0.881	0.473	0.853	0.455
Weather	96	6.25 %	7.59 %	**0.165**	**0.219**	0.176	0.237	0.217	0.296	0.266	0.336	0.300	0.384	0.896	0.556	0.369	0.406
	192	7.73 %	3.19 %	**0.203**	**0.273**	0.220	0.282	0.276	0.336	0.307	0.367	0.598	0.544	0.622	0.624	0.416	0.435
	336	8.68 %	3.45 %	**0.242**	**0.308**	0.265	0.319	0.339	0.380	0.395	0.395	0.578	0.739	0.376	0.753	0.455	0.454
	720	4.64 %	8.29 %	**0.308**	**0.333**	0.323	0.362	0.403	0.428	0.419	0.428	1.059	0.741	1.004	0.934	0.535	0.520
ETTm1	96	5.69 %	7.01 %	**0.282**	**0.319**	0.299	0.343	0.379	0.419	0.505	0.475	0.672	0.571	0.543	0.510	0.863	0.664
	192	5.07 %	3.56 %	**0.318**	**0.352**	0.335	0.365	0.426	0.441	0.553	0.496	0.795	0.669	0.557	0.537	1.113	0.776
	336	2.98 %	4.66 %	**0.358**	**0.368**	0.369	0.386	0.445	0.459	0.621	0.537	1.212	0.871	0.754	0.655	1.267	0.832
	720	3.29 %	4.51 %	**0.411**	**0.402**	0.425	0.421	0.543	0.490	0.671	0.561	1.166	0.823	0.908	0.724	1.324	0.858
ETTm2	96	10.78 %	6.92 %	**0.149**	**0.242**	0.167	0.260	0.203	0.287	0.255	0.339	0.365	0.453	0.435	0.507	2.041	1.073
	192	3.13 %	7.26 %	**0.217**	**0.281**	0.224	0.303	0.269	0.328	0.281	0.340	0.533	0.563	0.730	0.673	2.249	1.112
	336	5.34 %	2.92 %	**0.266**	**0.332**	0.281	0.342	0.325	0.366	0.339	0.372	1.363	0.887	1.201	0.845	2.568	1.238
	720	6.55 %	7.84 %	**0.371**	**0.388**	0.397	0.421	0.421	0.415	0.433	0.432	3.379	1.338	3.625	1.451	2.720	1.287

According to the quantitative analysis in Table 7, SDHNet obtains the following average reductions in MSE compared to the best results across all benchmarks: 6.34 % on the Electricity dataset, 4.86 % on the Traffic dataset, 6.82 % on the Weather dataset, 4.25 % on the ETTm1 dataset, and 6.45% on the ETTm2 dataset. Similarly, the average MAE reductions are as follows: 4.49 % on the Electricity dataset, 5.11 % on the Traffic dataset, 5.63 % on the Weather dataset, 4.93 % on the ETTh2 dataset, and 6.32 % on the ETTm2 dataset. As a result, FEDNet achieved the best performance across all benchmark tests for every forecasting horizon setting. In terms of quantitative performance, FEDNet consistently outperformed the best results in all benchmark tests. Notably, compared to other specialized Transformer-based models, LSTM performed the worst. The design of FEDNet enables better modeling of periodicity, allowing the model to effectively utilize longer historical windows, which contain more information for making accurate predictions.

4.2.3 Ablation Experiment

The FEDNet proposed in this study consists of three key components: reversible normalization(RevIn), trend-seasonal decomposition(T-S Decomp), and Fourier decomposition(F Decomp). To illustrate the beneficial effects of these elements, we carry out extensive experiments in this section. The table 8 displays the experimental findings, with "w/" denoting the module's presence, "w/o" denoting its absence, and bolding the best results.

4.2.3.1 Reversible Normalization Module

An ablation experiment initially investigated the impact of the reversible normalization module on time series forecasting models. Table 8 demonstrates that this module is crucial for improving the model's predicted accuracy and stability. Specifically, when the reversible normalization module was removed, the model's prediction errors increased across multiple datasets. Theoretically, the reversible normalization operation unifies the scale characteristics of the data, reducing the negative impact of outliers and inconsistencies in feature dimensions on model training. The value of reversible normalization is reflected not only in the improvement of prediction accuracy but also in enhancing the model's robustness and stability. Overall, the experiment provides strong evidence for the critical role of reversible normalization in time series forecasting tasks and offers important design insights for optimizing time series modeling.

Table 8 The ablation experiment results of the three key components. The optimal result is highlighted in bold.

RevIN	T-S Decomp	F Decomp	Horizon	Electricity		Traffic		Weather		Etthm2	
				MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE
w/	w/	w/	96	**0.131**	**0.225**	**0.402**	**0.266**	**0.165**	**0.219**	**0.149**	**0.242**
			192	**0.142**	**0.238**	**0.405**	**0.271**	**0.203**	**0.273**	**0.217**	**0.281**
			336	**0.155**	**0.255**	**0.411**	**0.281**	**0.242**	**0.308**	**0.266**	**0.332**
			720	**0.196**	**0.289**	**0.431**	**0.302**	**0.308**	**0.333**	**0.371**	**0.388**
w/o	w/	w/	96	0.138	0.237	0.410	0.281	0.173	0.241	0.163	0.251
			192	0.157	0.249	0.418	0.299	0.226	0.273	0.223	0.303
			336	0.169	0.272	0.436	0.316	0.265	0.302	0.283	0.342
			720	0.203	0.304	0.463	0.323	0.319	0.364	0.396	0.421
w/	w/o	w/	96	0.166	0.298	0.412	0.322	0.265	0.294	0.231	0.294
			192	0.192	0.366	0.423	0.396	0.312	0.364	0.296	0.383
			336	0.263	0.406	0.496	0.462	0.378	0.412	0.401	0.422
			720	0.436	0.506	0.763	0.606	0.409	0.453	0.514	0.533
w/	w/	w/o	96	0.139	0.231	0.408	0.279	0.175	0.236	0.156	0.261
			192	0.151	0.248	0.426	0.284	0.215	0.282	0.221	0.313
			336	0.167	0.267	0.431	0.291	0.262	0.319	0.279	0.342
			720	0.208	0.301	0.462	0.317	0.315	0.353	0.407	0.412

4.2.3.2 Seasonal-trend decomposition Module

Through the ablation study of the seasonal-trend decomposition module, we evaluated the effect of removing this crucial component from the FEDNet, specifically by modeling directly on the unprocessed raw time series data without any decomposition operations. Table 8 shows the changes in model performance after the seasonal-trend decomposition module was removed. The seasonal-trend decomposition module improves forecasting performance, as seen by the model's considerable decline in predicted accuracy across several datasets. This module separates trend and seasonal components from the raw time series signal, allowing the neural network to learn long-term trends and periodic patterns in the time series data, thereby enhancing future prediction accuracy. In contrast, when the raw signal was input directly into the network without decomposition, complex patterns at different time scales intertwined, increasing the difficulty of model training and imposing higher demands on the model's generalization ability. Consequently, seasonal-trend decomposition markedly improved its capacity to elucidate the inherent structure of the time series, rendering it an essential element for augmenting forecasting precision.

4.2.3.3 Fourier decomposition Module

Finally, an ablation experiment was carried out to investigate the function of the Fourier decomposition module in the FEDNet. Specifically, by removing this module and directly using the seasonal components for prediction, the impact of its absence on model performance was assessed. The experimental results in Table 8 clearly show that Fourier decomposition is crucial for enhancing the prediction accuracy of the model. Once the module was removed, the model's performance deteriorated significantly across multiple datasets, indicating that Fourier decomposition effectively extracts periodic features from the raw time series data and aids the model in capturing the periodic patterns and their variations through frequency domain decomposition. In conclusion, Fourier decomposition is essential in FEDNet, enhancing the model's capability to handle complex time series data by improving prediction accuracy and feature extraction.

Overall, the ablation experiments in this section further validate the effectiveness and necessity of the three key components in FEDNet: the normalization operation, trend and seasonal component decomposition and Fourier decomposition module. The experimental results confirm that these design choices are reasonable and essential, collectively ensuring the model's efficient forecasting capability.

4.3 Visual Analysis

In the USFF experiment, we selected the two-week sales forecast results of three SKU products from three different categories for visual comparison and analysis, as illustrated in Figs. 3, 4, and 5. The predicted sales were rounded up to the nearest integer. Through this process, we were able to validate the reasonableness of the models we selected.

The SKU sales forecasting results depicted in Fig. 3 indicate that the SARIMA model exhibits superior performance in Category 1. where the sales cycles are short and the data exhibits low volatility. In Category 5, the LightGBM model achieves

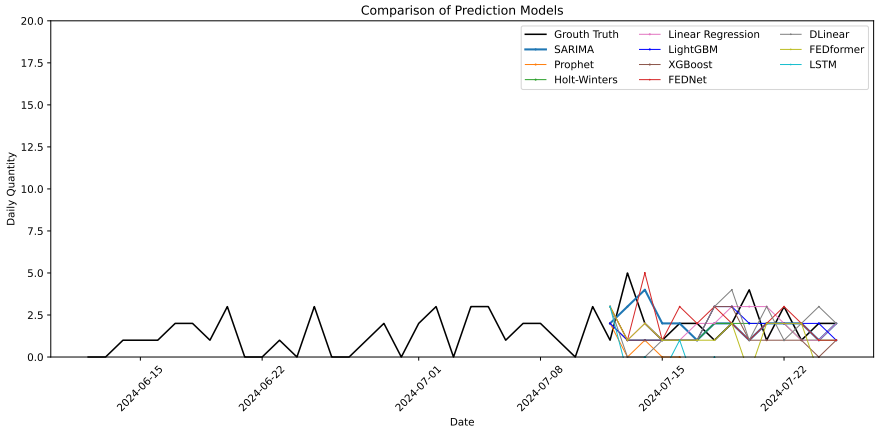


Fig. 3 Visualization comparison of the forecasting results for the sales of a single SKU in Category 1 across 10 models.

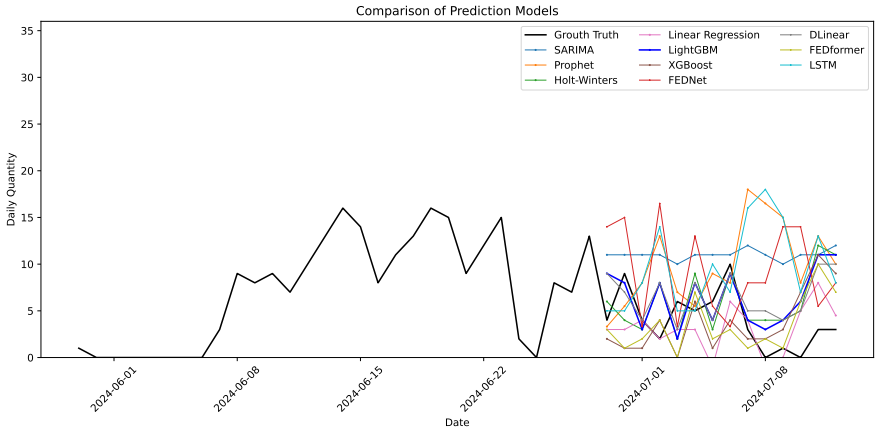


Fig. 4 Visualization comparison of the forecasting results for the sales of a single SKU in Category 5 across 10 models.

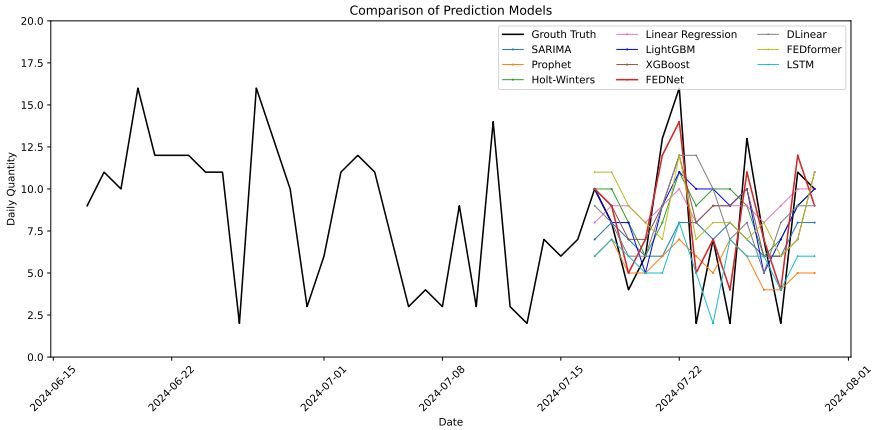
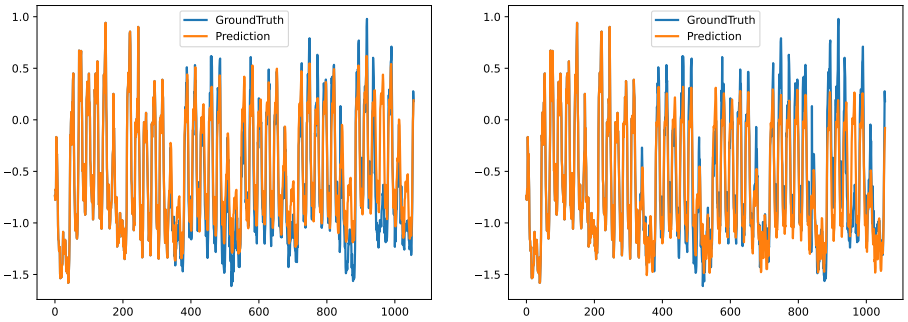


Fig. 5 Visualization comparison of the forecasting results for the sales of a single SKU in Category 10 across 10 models.

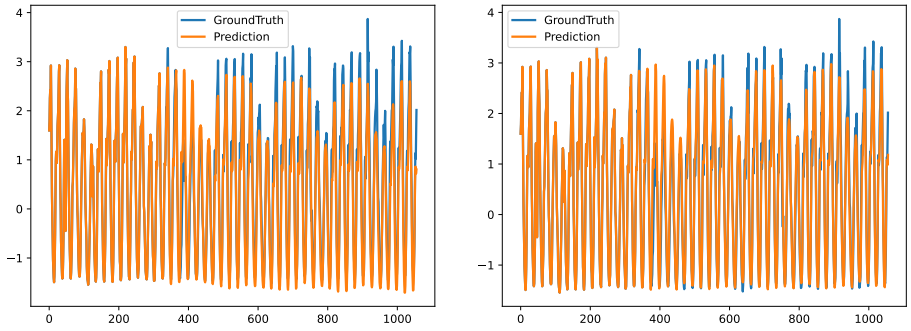
the best results, as this type of data features moderate sales cycles, high volatility, and strong periodicity. In Category 10, the FEDNet model delivers the best performance, as the sales cycles are longer and the periodicity of the data is more evident. Overall, after the model classification in USFF, the integration of multiple models better adapts to various complex datasets, improving prediction accuracy and model generalization. The visualized results further validate the rationality of the proposed algorithm framework.

For the FEDNet experiments, we conducted a visual comparison between the FEDNet and DLinear models, as shown in Figs. 6, 7, and 8. For comparison, we plotted the expected outcomes of the final channel of the test sets from three datasets. The prediction horizon was set to 720 time steps for each experiment. The visualization results show that FEDNet outperforms the baseline model DLinear across all three datasets. FEDNet is better able to capture the periodic variations in the different datasets.



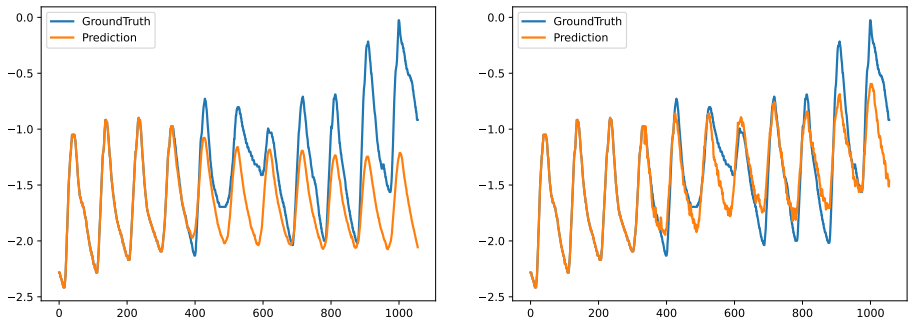
(a) input-720 under Dlinear (b) input-720 under FEDNet

Fig. 6 Visualization comparison of Dlinear and FEDNet on the Electricity dataset



(a) input-720 under Dlinear (b) input-720 under FEDNet

Fig. 7 Visualization comparison of Dlinear and FEDNet on the Traffic dataset



(a) input-720 under Dlinear

(b) input-720 under FEDNet

Fig. 8 Visualization comparison of Dlinear and FEDNet on the ETTm2 dataset

5 Conclusions

This study presents a novel predictive algorithmic framework, named the Unified Sales Forecasting Framework (USFF), designed to address the challenge that a single model cannot effectively handle the diverse nature of data in vending machine product sales forecasting. The framework integrates statistical classification, data preprocessing, classifier construction, and hybrid modeling strategies, enabling accurate prediction of vending machine product sales across various time spans, volatility, and seasonality. By combining the strengths of statistical learning, machine learning, and deep learning techniques, the USFF framework effectively accommodates the diverse characteristics of sales data. Additionally, this study introduces FEDNet, an innovative deep learning algorithm aimed at overcoming prediction accuracy challenges caused by complex dynamic dependencies and non-stationarity in long-term sales forecasting. FEDNet integrates reversible instance normalization, seasonal-trend decomposition, and Fourier decomposition to address these issues. Experimental results demonstrate that FEDNet outperforms traditional models in terms of metrics such as MSE and MAE, showcasing its efficiency and generalizability in multidimensional time-series forecasting. The USFF framework and the FEDNet approach introduced in this research provide innovative algorithmic tools and technical support for vending machine sales prediction, with promising application potential.

Future research will focus on enhancing the model's predictive accuracy. While existing studies primarily rely on historical sales data and external features, they do not fully account for the potential impact of market dynamics, such as foot traffic, competitor activities, and promotions. Subsequent work will explore integrating these external factors into the model to enhance its accuracy. Additionally, as sales data volumes continue to increase, deep learning techniques will become progressively more advantageous. We propose incorporating attention mechanisms and graph neural networks to more effectively capture the complex dependencies and interactions across data channels, thereby improving prediction accuracy and ultimately providing stronger support for business decision-making.

Author Contributions Methodology and investigation: Qianyang Li, Shaoxun Wang; Software: Qianyang Li, Jiawei Cao; Writing—original draft preparation: Qianyang Li; Writing—review and editing: Qianyang Li, Shaoxun Wang; Visualization: Qianyang Li, Jiawei Cao; Validation: Shaoxun Wang, Peng Tao; Resources: Xingjun Zhang, Peng Tao; Supervision: Xingjun Zhang, Peng Tao.

Acknowledgments We gratefully acknowledge SNBC for providing the data and technical support that were essential to this research.

Funding This work is supported by the National Natural Science Foundation of China (No. 62372366).

Data availability The data is available upon request at the following repository: https://drive.google.com/drive/folders/1ZOYpTUa82_jCcxIdTmyr0LXQfvaM9vIy.

Declarations

Conflict of interest The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

References

[1] Fildes, R., Ma, S., Kolassa, S.: Retail forecasting: Research and practice **38**(4), 1283–1318 https://doi.org/10.1016/j.ijforecast.2019.06.004

[2] Krishna, A., V, A., Aich, A., Hegde, C.: Sales-forecasting of retail stores using machine learning techniques. In: 2018 3rd International Conference on Computational Systems and Information Technology for Sustainable Solutions (CSITSS), pp. 160–166. https://doi.org/10.1109/CSITSS.2018.8768765 . https://ieeexplore.ieee.org/document/8768765/?arnumber=8768765

[3] Hasan, M.R., Kabir, M.A., Shuvro, R.A., Das, P.: A Comparative Study on Forecasting of Retail Sales (2022). https://arxiv.org/abs/2203.06848

[4] Ensafi, Y., Amin, S.H., Zhang, G., Shah, B.: Time-series forecasting of seasonal items sales using machine learning – a comparative analysis **2**(1), 100058 https://doi.org/10.1016/j.jjimei.2022.100058

[5] Ke, G., Meng, Q., Finley, T., Wang, T., Chen, W., Ma, W., Ye, Q., Liu, T.-Y.: LightGBM: A highly efficient gradient boosting decision tree. In: Advances in Neural Information Processing Systems, vol. 30. Curran Associates, Inc. https://proceedings.neurips.cc/paper/2017/hash/6449f44a102fde848669bdd9eb6b76fa-Abstract.html

[6] Efat, M.I.A., Hajek, P., Abedin, M.Z., Azad, R.U., Jaber, M.A., Aditya, S., Hassan, M.K.: Deep-learning model using hybrid adaptive trend estimated series

for modelling and forecasting sales **339**(1), 297–328 https://doi.org/10.1007/s10479-022-04838-6

[7] Li, X., Shen, J., Wang, D., Lu, W., Chen, Y.: Multi-modal transform-based fusion model for new product sales forecasting. Eng. Appl. Artif. Intell. **133**, 108606 (2024) https://doi.org/10.1016/J.ENGAPPAI.2024.108606

[8] Retail Demand Forecasting: A Multivariate Approach and Comparison of Boosting and Deep Learning Methods. International Journal on Artificial Intelligence Tools. https://www.worldscientific.com/doi/10.1142/S0218213024500015

[9] Zhou, H., Zhang, S., Peng, J., Zhang, S., Li, J., Xiong, H., Zhang, W.: Informer: Beyond efficient transformer for long sequence time-series forecasting **35**(12), 11106–11115 https://doi.org/10.1609/aaai.v35i12.17325 . Number: 12

[10] Wu, H., Xu, J., Wang, J., Long, M.: Autoformer: Decomposition transformers with auto-correlation for long-term series forecasting. In: Advances in Neural Information Processing Systems, vol. 34, pp. 22419–22430. Curran Associates, Inc. https://proceedings.neurips.cc/paper/2021/hash/bcc0d400288793e8bdcd7c19a8ac0c2b-Abstract.html

[11] FedFormer: Contextual Federation with Attention in Reinforcement Learning \textbar Proceedings of the 2023 International Conference on Autonomous Agents and Multiagent Systems. https://dl.acm.org/doi/10.5555/3545946.3598716

[12] Mehmood, U., Broderick, J., Davies, S., Bashir, A.K., Rabie, K.: Machine learning-based predictive inventory for a vending machine warehouse **7**(6), 94–100 https://doi.org/10.1109/IOTM.001.2300271

[13] Li, D., Li, X., Lin, K., Liao, J., Du, R., Lu, W., Madden, A.: A multiple long short-term model for product sales forecasting based on stage future vision with prior knowledge. Information Sciences **625**, 97–124 (2023) https://doi.org/10.1016/j.ins.2022.12.099

[14] Wolpert, D.H., Macready, W.G.: No free lunch theorems for optimization **1**(1), 67–82 https://doi.org/10.1109/4235.585893

[15] Yi, K., Zhang, Q., Cao, L., Wang, S., Long, G., Hu, L., He, H., Niu, Z., Fan, W., Xiong, H.: A Survey on Deep Learning based Time Series Analysis with Frequency Transformation. arXiv. Issue: arXiv:2302.02173. https://doi.org/10.48550/arXiv.2302.02173 . http://arxiv.org/abs/2302.02173

[16] Kim, T., Kim, J., Tae, Y., Park, C., Choi, J.-H., Choo, J.: Reversible instance normalization for accurate time-series forecasting against distribution shift. https://openreview.net/forum?id=cGDAkQo1C0p

[17] Gür Ali, Sayın, S., van Woensel, T., Fransoo, J.: Sku demand forecasting in the presence of promotions. Expert Systems with Applications **36**(10), 12340–12348 (2009) https://doi.org/10.1016/j.eswa.2009.04.052

[18] Box, G.: Box and Jenkins: Time Series Analysis, Forecasting and Control, pp. 161–215. Palgrave Macmillan UK, London (2013). https://doi.org/10.1057/9781137291264_6

[19] Durbin, J., Koopman, S.J.: Time Series Analysis by State Space Methods vol. 38. OUP Oxford, Oxford, England (2012)

[20] Siegel, A.F.: Practical Business Statistics. Academic Press, London, United Kingdom (2016)

[21] Fattah, J., Ezzine, L., Aman, Z., El Moussami, H., Lachhab, A.: Forecasting of demand using arima model. International Journal of Engineering Business Management **10**, 1847979018808673 (2018)

[22] Chen, T., Guestrin, C.: XGBoost: A scalable tree boosting system. In: Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. KDD '16, pp. 785–794. Association for Computing Machinery. https://doi.org/10.1145/2939672.2939785 . Place: New York, NY, USA

[23] Xu, R.: A method for wal-mart sales forecasting based on machine learning. In: Proceedings of the 2024 International Conference on Cloud Computing and Big Data, pp. 350–355 (2024)

[24] Andrade, L.A.C.G., Cunha, C.B.: Disaggregated retail forecasting: A gradient boosting approach. Applied Soft Computing **141**, 110283 (2023) https://doi.org/10.1016/j.asoc.2023.110283

[25] Kao, C.-W., Hung, J.-C.: Forecasting seasonal sales in supermarket chains using fft-ga and lightgbm models. In: Proceedings of the 2024 8th International Conference on Information System and Data Mining. ICISDM '24, pp. 8–14. Association for Computing Machinery, New York, NY, USA (2024). https://doi.org/10.1145/3686397.3686399

[26] Zhou, T.: Improved sales forecasting using trend and seasonality decomposition with LightGBM. In: 2023 6th International Conference on Artificial Intelligence and Big Data (ICAIBD), pp. 656–661. https://doi.org/10.1109/ICAIBD57115.2023.10206380 . http://arxiv.org/abs/2305.17201

[27] Lim, B., Zohren, S.: Time-series forecasting with deep learning: a survey **379**(2194), 20200209 https://doi.org/10.1098/rsta.2020.0209

[28] Yu, Y., Si, X., Hu, C., Zhang, J.: A review of recurrent neural networks: LSTM cells and network architectures **31**(7), 1235–1270 https://doi.org/10.1162/neco_

a_01199

[29] Honjo, K., Zhou, X., Shimizu, S.: CNN-GRU based deep learning model for demand forecast in retail industry. In: 2022 International Joint Conference on Neural Networks (IJCNN), pp. 1–8. https://doi.org/10.1109/IJCNN55064.2022.9892599 . https://ieeexplore.ieee.org/document/9892599

[30] Li, Y., Yang, Y., Zhu, K., Zhang, J.: Clothing sale forecasting by a composite gru–prophet model with an attention mechanism. IEEE Transactions on Industrial Informatics **17**(12), 8335–8344 (2021)

[31] Ahmed, S., Chakrabortty, R.K., Essam, D.L., Ding, W.: A switching based forecasting approach for forecasting sales data in supply chains. Applied Soft Computing **167**, 112419 (2024) https://doi.org/10.1016/j.asoc.2024.112419

[32] Loureiro, A.L.D., Miguéis, V.L., Silva, L.F.M.: Exploring the use of deep neural networks for sales forecasting in fashion retail **114**, 81–93 https://doi.org/10.1016/j.dss.2018.08.010

[33] Kitaev, N., Kaiser, L., Levskaya, A.: Reformer: The efficient transformer. In: 8th International Conference on Learning Representations, ICLR 2020, Addis Ababa, Ethiopia, April 26-30, 2020 (2020). https://openreview.net/forum?id=rkgNKkHtvB

[34] Liu, S., Yu, H., Liao, C., Li, J., Lin, W., Liu, A.X., Dustdar, S.: Pyraformer: Low-complexity pyramidal attention for long-range time series modeling and forecasting. In: The Tenth International Conference on Learning Representations, ICLR 2022, Virtual Event, April 25-29, 2022 (2022). https://openreview.net/forum?id=0EXmFzUn5I

[35] Li, Q., Yu, M.: Achieving sales forecasting with higher accuracy and efficiency: A new model based on modified transformer. J. Theor. Appl. Electron. Commer. Res. **18**(4), 1990–2006 (2023) https://doi.org/10.3390/JTAER18040100

[36] Mitra, A., Jain, A., Kishore, A., Kumar, P.: A comparative study of demand forecasting models for a multi-channel retail company: A novel hybrid machine learning approach **3**(4), 58 https://doi.org/10.1007/s43069-022-00166-4

[37] Tran, D.T., Huh, J.-H., Kim, J.-H.: Building a lucy hybrid model for grocery sales forecasting based on time series **79**(4), 4048–4083 https://doi.org/10.1007/s11227-022-04824-6

[38] Zeng, A., Chen, M., Zhang, L., Xu, Q.: Are transformers effective for time series forecasting? **37**(9), 11121–11128 https://doi.org/10.1609/aaai.v37i9.26317 . Number: 9

[39] Qiu, X., Hu, J., Zhou, L., Wu, X., Du, J., Zhang, B., Guo, C., Zhou, A., Jensen,

C.S., Sheng, Z., Yang, B.: TFB: Towards comprehensive and fair benchmarking of time series forecasting methods **17**(9), 2363–2377 https://doi.org/10.14778/3665844.3665863

[40] RB, C.: Stl: A seasonal-trend decomposition procedure based on loess. J Off Stat **6**, 3–73 (1990)

[41] Zeng, C., Tian, Y., Zheng, G., Gao, Y.: How Much Can Time-related Features Enhance Time Series Forecasting? arXiv. Issue: arXiv:2412.01557. https://doi.org/10.48550/arXiv.2412.01557 . http://arxiv.org/abs/2412.01557

[42] Hamilton, J.D.: Time Series Analysis. Princeton university press, New Jersey, United States (2020)