Advances in Industrial Engineering, December 2026, 60(2): 471-492
DOI: 10.22059/aie.2026.413445.1980

RESEARCH PAPER

Data-Driven Hierarchical Forecasting Model for Point-of-Sale (POS) Transaction Counts

Mahdieh Tavakoli[1] and S. Ali Torabi[2*]

[1]*Ph.D. Candidate, School of Industrial Engineering, College of Engineering, University of Tehran, Tehran, Iran.*
[2]*Professor, School of Industrial Engineering, College of Engineering, University of Tehran, Tehran, Iran.*

Received: 28 April 2026, Revised: 19 May 2026, Accepted: 10 June 2026
© The Author(s) retain the copyright 2026

Abstract

Banks and other organizations in the financial sector, particularly payment service providers (PSPs), require accurate forecasts of transaction counts to optimize procurement planning for point-of-sale (POS) devices under budget constraints. To address this need, this study proposes a two-level hierarchical forecasting framework for POS transaction counts to support tactical decision-making. At the first level, a time-weighted long short-term memory (LSTM) model is developed to forecast monthly transaction counts over a one-year horizon. A set of candidate exogenous variables is initially identified and subsequently refined using the goal programming-based fuzzy best–worst method (GP-FBWM) to improve model performance. At the second level, a Random Forest (RF) model is employed to estimate the relative weights of daily transactions within each month based on intrinsic daily characteristics. A forward feature selection approach is further applied to identify the most influential daily features affecting transaction distribution. The proposed framework is validated through a real-world case study of a PSP in Iran. The results indicate that the time-weighted LSTM achieves high predictive accuracy with R^2 values of 0.90 and 0.89 on validation and test datasets for monthly forecasting, respectively. For daily transaction weights, the six-feature subset achieved a mean cross-validation R^2 of 0.9647 for the representative branch. These findings confirm the effectiveness and practical applicability of the proposed framework in supporting data-driven planning.

Keywords: Hierarchical Forecasting, Long Short-Term Memory, Payment Service Providers, Point-of-Sale Transactions, Random Forest.

1. Introduction

With the rapid expansion of digital technologies and e-commerce, financial platforms have become a fundamental component of modern economic systems. Payment systems, as a key element of these platforms, facilitate secure and efficient cashless transactions and play a vital role in supporting commercial activities (Benchaji et al. 2021). Among various payment infrastructures, point-of-sale (POS) systems are widely used in retail environments and generate a substantial volume of transaction data on a daily basis. Payment Service Providers (PSPs), acting as intermediaries between financial institutions and merchants, rely heavily on accurate transaction forecasts to maintain service quality and ensure efficient resource allocation (Siyal

* Corresponding author: (S. Ali Torabi)
Email: satorabi@ut.ac.ir

et al. 2019). Accurate forecasting of POS transaction counts is therefore essential for effective operational planning and decision-making (Ładyżyński et al. 2019). Underestimation of transaction demand may lead to system congestion, service disruptions, and customer dissatisfaction, whereas overestimation can result in unnecessary costs due to excess equipment and infrastructure. Therefore, reliable forecasting models are critical for balancing service quality and operational efficiency in PSPs (Shahabi and Faezy Razi 2019). Considering the importance of forecasting the number of transactions in an electronic payment system, in this study, we focus on POS transaction counts, a key variable in tactical planning for PSPs. The PSP managers believe that for a PSP to work effectively and efficiently, the volume of transactions of POS should be forecasted over the short and medium terms since service quality is precisely the distinguishing feature of successful and ineffective payment service providers and is considered as a scale or factor for measuring customer satisfaction (Feng et al. 2022). In other words, forecasting POS transaction counts is an important part of tactical planning for PSPs and can help improve the quality of electronic payment services (Peng et al. 2022). This means that if the volume of future transactions (i.e., the customer's demand for POS transactions) in a certain period is correctly forecasted for a PSP, that company can provide the necessary equipment and manpower to meet this level of demand which in turn would have a direct impact on the satisfaction of customers and also efficient management of the company's resources and costs (Li et al. 2023). For example, in the last days of a year or during a subsidy payment, a surge in the number of buyers may overload the payment switch and disrupt network service. When demand exceeds the capacity provisioned for the forecast, service interruptions can reduce customer satisfaction. Also, overestimating transaction demand may lead PSPs to deploy more devices than necessary, increasing total costs.

Demand forecasting is the process of estimating the potential demand for a product or service in the market based on quantitative and qualitative data (Gehlot et al. 2022). The availability of large datasets supports the use of quantitative forecasting methods, including machine learning (ML). With the large amount of data stored in PSPs related to transaction details, there is a need for more effective tools to process such available data and provide more accurate forecasting models. Because demand varies over time, the forecast horizon must be specified. Common periods for forecasting models are generally divided into three time spans including short-term, medium-term, and long-term (Dias et al. 2022). Hierarchical forecasting models can generate forecasts at multiple temporal aggregation levels. Hierarchical forecasting involves producing consistent forecasts or adjusting inconsistent ones by allowing individual time series to be forecasted separately, while maintaining the interconnections within the hierarchy. This includes approaches such as bottom-up, top-down, and middle-out forecasting. In the bottom-up hierarchical approach, forecasts are carried out at the lowest level of the hierarchy and then summed going up. In the top-down approach, forecasts are first generated for the top level and then disaggregated down the hierarchy. Finally, the middle-out approach is a combination of the two methods described above and it can only be used for strictly hierarchical time series. In this approach, forecasts are generated directly at the middle level. Then, for all the levels above the selected level, the bottom-up approach is used and for the levels below the middle one, the top-down approach is used. In recent years, data-driven and machine learning-based approaches have gained increasing attention in demand forecasting due to their ability to capture complex and nonlinear patterns in large-scale datasets.

Despite the growing adoption of machine learning techniques in demand forecasting, most existing studies have primarily focused on automated teller machine (ATM) transactions, while limited attention has been given to POS transaction forecasting. Moreover, many studies consider only a single time horizon, which does not adequately reflect real-world operational requirements. To address these gaps, this study aims to develop a data-driven hierarchical forecasting model for POS transaction counts by simultaneously considering monthly and daily

forecasting horizons. In this regard, the following research questions are investigated:
RQ 1: What are the most important indicators affecting monthly POS transaction counts?
RQ 2: What are the most important indicators affecting daily POS transaction counts?
RQ 3: How can a hierarchical forecasting model be developed to predict POS transaction counts across multiple time horizons?

1.1. Novelty and Contributions

Despite the growing body of research in transaction and demand forecasting, most existing studies have primarily focused on automated teller machine (ATM) transactions, with comparatively limited attention given to point-of-sale (POS) transaction forecasting. Compared with ATM transactions, POS transaction patterns are generally more dynamic and influenced by a wider range of behavioral and contextual factors. This gap persists despite the practical importance of POS systems and the large volume of transactional data generated by payment service providers (PSPs).

In addition, although several studies in other domains have investigated forecasting at multiple temporal horizons, limited research has explored hierarchical forecasting frameworks for POS transaction forecasting that maintain consistency between monthly and daily forecasts. In many practical forecasting applications, daily forecasts are generated independently and subsequently aggregated into monthly values. However, forecast errors may accumulate, and separately generated monthly forecasts may differ from the sum of the daily forecasts, particularly in highly dynamic transaction environments. Furthermore, monthly and daily forecasts often serve different operational purposes. Monthly forecasting mainly supports higher-level planning activities such as resource allocation, budgeting, and network capacity management, whereas daily forecasting is more closely related to operational monitoring and short-term transaction management. These challenges motivate the need for integrated hierarchical forecasting frameworks capable of simultaneously supporting both temporal levels while maintaining forecasting consistency.

Driven by these research gaps, this study develops a data-driven hierarchical forecasting framework for POS transaction counts that captures both short-term (daily) and mid-term (monthly) dynamics. The proposed approach integrates advanced machine learning techniques within a unified framework to improve forecasting accuracy and support more effective decision-making in PSP operations. The main contributions of this study can be summarized as follows:

- A novel top-down hierarchical forecasting framework is developed for POS transaction counts, enabling consistent modeling of monthly and daily transaction dynamics.
- The proposed framework combines deep learning and ensemble learning approaches by integrating a time-weighted long short-term memory (LSTM) model for monthly forecasting and a random forest (RF) model for daily transaction weight estimation.
- A structured feature selection process based on the goal programming-based fuzzy best–worst method (GP-FBWM) is incorporated to identify the most influential factors affecting transaction behavior, thereby improving both forecasting performance and model interpretability.
- The effectiveness and practical applicability of the proposed framework are demonstrated through a real-world case study of a Payment Service Provider (PSP), highlighting its potential for supporting data-driven operational planning, resource allocation, and transaction management.

The remainder of this paper is organized as follows. Section 2 presents a review of the relevant literature. Section 3 describes the research methodology. Section 4 reports the results and discusses the findings. Finally, Section 5 concludes the paper and provides directions for future research.

2. Literature Review

This section reviews the existing body of literature in two main streams: (i) forecasting of banking-related variables, including transaction volumes and other financial indicators such as sales and service usage within the banking domain, and (ii) demand forecasting in other industries, with a particular focus on studies employing machine learning (ML) techniques. The second stream is examined to identify commonly used forecasting methods as well as influential exogenous variables—such as temporal factors, holidays, and special events—that may affect transaction patterns over time. The section closes by identifying the research gaps addressed in this study.

2.1. Banking Transaction Forecasting

In the field of banking transactions, a substantial body of research has focused on developing predictive models for related applications such as fraud detection (Benchaji et al. 2021) and customer behavior analysis (Shahabi and Faezy Razi 2019). However, relatively limited attention has been given to directly forecasting transaction counts.

Recent advancements in financial forecasting have increasingly emphasized the application of deep learning approaches due to their ability to capture nonlinear temporal dependencies and complex behavioral patterns in large-scale financial datasets. A comprehensive review by Giantsidi and Tarantola (2025) highlights the growing dominance of deep learning architectures, particularly Long Short-Term Memory (LSTM) networks and hybrid models, in financial time series forecasting. Compared to traditional statistical methods, these models have demonstrated superior capability in modeling dynamic and nonlinear transaction behaviors.

A major stream of existing research focuses on ATM cash demand forecasting. Rajwani et al. (2017) employed regression analysis to estimate ATM cash requirements while considering holidays and special events. Similarly, Jadwal et al. (2019) applied Support Vector Machine (SVM) models to clustered ATM data in order to improve cash flow prediction accuracy. Rafi et al. (2020) proposed a VAR-MAX model with exogenous variables and demonstrated improved forecasting performance compared to conventional ARIMA models. These studies collectively indicate that incorporating external variables and machine learning techniques can substantially improve forecasting accuracy in banking transaction applications.

Several studies have also explored intelligent learning-based approaches for forecasting banking transaction volumes and profitability. Razavi et al. (2019), for example, developed an Artificial Neural Network (ANN)-based model to estimate ATM profitability by forecasting transaction counts across multiple categories, including bill payments, money transfers, and withdrawals. Using real-world data from 374 ATMs in Tehran, their proposed model achieved satisfactory predictive accuracy in estimating transaction income. Similarly, Kullaya Swamy and Sarojamma (2020) utilized Deep Belief Networks (DBN) and Neural Networks (NN) to forecast banking transaction counts and amounts, while optimizing model parameters using a hybrid Lion Algorithm–Artificial Bee Colony (L-ABC) approach. These studies demonstrate the increasing adoption of intelligent and hybrid learning methods in financial transaction forecasting.

Deep learning-based forecasting models have also received increasing attention in ATM-related studies. Asad et al. (2020) used an LSTM-based framework combined with DBSCAN clustering estimate ATM transaction counts for multiple banks in Pakistan. Rafi et al. (2021) further developed an ARIMA-based forecasting framework using transaction data from six financial institutions and compared its performance with Recurrent Neural Networks (RNNs) and Amazon DeepAR models. Their findings showed that grouping ATMs with similar cash demand characteristics improved forecasting accuracy. Likewise, Sarveswararao et al. (2023) integrated chaos theory with LSTM and Random Forest models to improve ATM withdrawal

forecasting.

More recently, hybrid deep learning architectures have been proposed to further enhance forecasting performance in transaction-related applications. For instance, Kallurkar and Chandavarkar (2024) developed a CNN-LSTM framework for forecasting Ethereum transaction fees, while Cucu and Sakar (2025) applied a hybrid CNN-LSTM model to banking transaction count prediction and reported superior performance compared with standalone machine learning and deep learning models. These findings suggest that hybrid and deep learning approaches are increasingly becoming dominant methodologies in financial transaction forecasting due to their ability to capture both spatial and temporal dependencies.

Despite these methodological advancements, several important research gaps remain. First, the majority of prior studies primarily focus on ATM transactions and cash demand forecasting, whereas POS transaction forecasting has received limited attention. Second, most existing models consider only a single forecasting horizon, such as daily, weekly, or monthly prediction, without addressing the need for integrated multi-horizon forecasting frameworks. Third, although many studies employ advanced machine learning and deep learning methods, relatively few investigate hierarchical forecasting structures capable of maintaining consistency across different temporal levels. These limitations highlight the need for more comprehensive and application-oriented forecasting frameworks tailored to POS transaction dynamics in PSP environments.

A limited number of studies have investigated transaction forecasting using POS-related data. García et al. (2021), for example, used bank card transaction data collected from POS devices and online purchases to forecast retail sales across multiple temporal horizons. Their work demonstrated the practical value of transaction-level forecasting using POS data; however, the study primarily focused on retail sales forecasting rather than operational transaction forecasting for PSPs. Similarly, Kumar et al. (2026) examined transaction behavior in digital payment systems using machine learning regression techniques and showed the effectiveness of Random Forest models in predicting transaction values. Although these studies confirm the potential of machine learning approaches in digital payment environments, they do not address hierarchical forecasting structures or joint daily and monthly transaction forecasting.

2.2. Demand Forecasting in Other Industries

Based on the literature review related to financial and banking transaction forecasting, it is evident that only a limited number of studies have directly focused on banking transaction forecasting. Therefore, this subsection reviews demand forecasting studies from other industries in order to examine commonly used forecasting approaches, evaluation metrics, and influential exogenous variables that may also be relevant to transaction forecasting problems.

Recent studies in demand forecasting have increasingly adopted hybrid machine learning and deep learning approaches to improve predictive accuracy. Sheng and Jia (2020), for example, proposed a hybrid SARIMAX-LSTM framework for short-term energy consumption forecasting. In their model, SARIMAX was employed to capture statistical patterns, while LSTM was used to reduce forecasting errors and model nonlinear dependencies. Their results demonstrated that the hybrid framework outperformed standalone models in terms of MAPE, RMSE, and R^2 metrics. Similarly, He and Yu (2020) developed a combined LightGBM-LSTM model for forecasting vegetable sales and incorporated external variables such as weather conditions, holidays, special events, and salary payment periods. Their findings indicated that the hybrid model achieved lower forecasting errors compared to standalone LightGBM and LSTM models. These studies collectively highlight the growing effectiveness of hybrid deep learning architectures in handling complex demand forecasting problems.

Several studies have also emphasized the importance of temporal and contextual variables

in improving forecasting performance. Huber and Stuckenschmidt (2020) investigated daily retail demand forecasting in German bakeries using Gradient Boosting Regression Trees (GBRTs), neural networks, and LSTM models. Their study considered several calendar-related and location-based variables, including weekdays, holidays, special events, and geographical characteristics. The results showed that LSTM models slightly outperformed the proposed GBRT-ANN framework. Likewise, Hirche et al. (2021) explored the impact of weather conditions and holidays on retail alcohol sales forecasting using SARIMAX models and showed that incorporating exogenous variables significantly improved forecasting accuracy. Falatouri et al. (2022) similarly demonstrated that integrating advertising days as external variables within a SARIMAX framework substantially reduced forecasting errors in retail supply chain applications. These findings collectively suggest that exogenous variables such as holidays, weather conditions, seasonal events, and temporal characteristics play a critical role in capturing demand fluctuations across different industries.

Deep learning methods have also been increasingly adopted in medium- and long-term demand forecasting applications. Vithitsoontorn and Chongstitvatana (2022) compared ARIMA and LSTM models for dairy product demand forecasting and reported superior performance for LSTM-based approaches. Aguiar-Pérez and Pérez-Juárez (2023) reviewed deep learning applications in smart grid demand forecasting and emphasized the importance of LSTM models in capturing temporal dependencies for short-term load prediction. Zhang and Niu (2024) proposed a hybrid CNN-LSTM framework enhanced with online review analysis for hotel demand forecasting, demonstrating that integrating textual and behavioral features can improve prediction performance. Hussain et al. (2025) further developed multivariate machine learning frameworks combining ANNs, Adaptive Neuro-Fuzzy Inference Systems (ANFIS), and econometric models for electricity demand forecasting. Their findings highlighted the importance of meteorological, seasonal, and economic variables in both short-term and long-term forecasting accuracy.

In addition to standalone forecasting approaches, hierarchical forecasting structures have also received growing attention in demand forecasting studies. Anderer and Li (2022) proposed a hierarchical forecasting framework for online shop demand prediction using N-BEATS for top-level forecasting and LightGBM for bottom-level forecasting. Their results demonstrated that hierarchical forecasting structures can outperform conventional non-hierarchical models by improving forecasting consistency across different aggregation levels. Moreover, Dai et al. (2026) provided a comprehensive review of peak demand forecasting methods and emphasized the growing role of machine learning approaches in addressing the increasing complexity and uncertainty of modern demand systems.

Overall, the reviewed studies demonstrate that hybrid machine learning and deep learning methods, particularly LSTM-based architectures, have become dominant approaches in demand forecasting due to their ability to capture nonlinear temporal patterns and complex relationships among variables. In addition, exogenous variables such as holidays, weather conditions, seasonal effects, and economic indicators consistently play an important role in improving forecasting performance. However, despite these methodological advancements, relatively limited research has explored the adaptation of such approaches to transaction forecasting problems in PSP environments, particularly within hierarchical multi-horizon forecasting frameworks that simultaneously address both monthly and daily forecasting requirements. In Table 1, the most relevant studies reviewed in this section are summarized.

2.3. Research Gaps

As Table 1 shows, although forecasting banking transactions is critical, only a limited number of studies have directly focused on banking transaction forecasting, with most existing works concentrating on ATM transactions rather than POS transactions. Although ATM and

POS forecasting both belong to the broader area of banking transaction forecasting, they differ in terms of transaction behavior and operational characteristics. ATM transactions are mainly associated with cash withdrawal patterns and cash replenishment management, whereas POS transactions are more strongly affected by consumer purchasing behavior, merchant activities, holidays, seasonal effects, and local economic conditions. Consequently, POS transaction forecasting involves more dynamic transaction patterns and requires forecasting models capable of capturing complex temporal and contextual relationships. In addition, accurate POS transaction forecasting plays an important role in supporting PSP operational planning, network management, and service quality improvement.

Furthermore, although several studies in other domains have considered multiple forecasting horizons, limited research has investigated hierarchical forecasting frameworks for POS transaction forecasting that maintain consistency between monthly and daily forecasts. In many practical applications, daily forecasts are generated independently and then aggregated into monthly values. However, forecast errors may accumulate, and separately generated monthly forecasts may differ from the sum of the daily forecasts. To address this limitation, this research proposes a top-down hierarchical forecasting framework in which monthly transaction counts are first forecasted and daily transaction forecasts are subsequently generated through estimated daily transaction weights.

Driven by these research gaps, this study develops a hybrid machine learning framework for forecasting monthly and daily POS transaction counts for PSPs. First, the most influential exogenous indicators affecting transaction behavior are identified. Then, the GP-FBWM is applied to estimate the relative importance of the indicators and reduce the complexity and cost of long-term data collection. Based on the selected indicators, a comprehensive dataset is constructed for forecasting analysis.

The proposed framework employs a time-weighted long short-term memory (LSTM) model for monthly transaction forecasting and a random forest regressor (RFR) for estimating daily transaction weights. In the proposed top-down structure, daily transaction counts are generated by multiplying the forecasted monthly transaction count by the estimated daily transaction weight corresponding to each day. Unlike many previous studies that directly apply conventional LSTM models, the proposed time-weighted LSTM framework assigns greater importance to more recent observations in order to better capture evolving transaction patterns and changing market conditions.

While the forecasting methods used in this study are themselves well-established, their integration within a hierarchical POS transaction forecasting framework provides a novel application in the context of PSP operations. The proposed approach offers a practical and operationally interpretable framework for simultaneously supporting monthly planning and daily transaction management in PSP environments.

Table 1. Summary of the reviewed forecasting studies

Study	Forecast target	Methods	Metrics	Dataset / horizon	External variables
(Rajwani et al. 2017)	ATM transaction amounts and counts	LSTM	MSE	2.5 yr of ATM transactions *Horizon: 1 yr*	Special occasions, holidays, location
(Razavi et al. 2019)	ATM profitability	ANN	Accuracy	341 ATMs in Tehran Province, 494 d *Horizon: 30 d*	—
(Jadwal et al. 2019)	Cash needs by ATM cluster	SVM	RMSE	2 yr of ATM transactions in England *Horizon: 1 wk*	—
(Rafi et al. 2020)	Cash demand per ATM	ARMAX*	RMSE	7 ATMs, 2013–2015 *Horizon: 2 yr*	—

Study	Forecast target	Methods	Metrics	Dataset / horizon	External variables
(Asad et al. 2020)	ATM transaction amounts and counts	LSTM	RMSE	2,241 ATMs at 6 Pakistani banks, 22 mo *Horizon: 1 yr*	—
(Kullaya Swamy and Sarojamma 2020)	Bank transaction amounts and counts	DBN and NN	RMSE, MAE	203 transactions, Jun 2008–Jan 2019 *Horizon: 1 yr*	—
(Rafi et al. 2021)	Cash demand per ATM	ARIMA	MAPE, SMAPE	18 mo *Horizon: 30 d*	—
(García et al. 2021)	Retail sales	SARIMA	Bias, MAE, MAPE, RMSE	7 yr of POS transactions *Horizon: 30 d, daily/weekly/monthly*	—
(Sarveswararao et al. 2023)	ATM cash withdrawals	LSTM and random forest	SMAPE	Historical ATM withdrawals *Horizon: Daily and weekly demand*	Day of week, holidays, possibly other calendar effects
(Cucu and Sakar 2025)	Correspondent banking transaction counts	CNN–LSTM	RMSE, R^2	Historical correspondent banking transactions *Horizon: Short-term, exact horizon unspecified*	—
(Sheng and Jia 2020)	Energy consumption	Hybrid SARIMAX–LSTM	MAPE, RMSE, R^2	1 yr *Horizon: 10 d*	—
(Huber and Stuckenschmidt 2020)	Daily retail demand	GBRT and LSTM	MAE	3 yr *Horizon: —*	Holidays, special days, location
(He and Yu 2020)	Vegetable sales	LightGBM–LSTM	MAPE	2 yr *Horizon: —*	Weather, holidays, special days, salary deposit timing
(Hirche et al. 2021)	Retail alcohol sales	SARIMAX	RMSE	3 yr *Horizon: 12 mo*	Weather, holidays, location
(Falatouri et al. 2022)	Retail supply chain forecasting	SARIMAX and LSTM	MAPE, RMSE	Austrian retailer sales, 37 mo *Horizon: 30 d*	Advertising days
(Vithitsoontorn and Chongstitvatana 2022)	Dairy product forecasting	ARIMA and LSTM	RMSE, MAE, MASE	6 yr *Horizon: 9 mo*	Month, day of month, day of week, employee, holidays
(Zhang and Niu 2024)	Hotel demand	LSTM–CNN with SFL-based text features	RMSE, MAE	Hotel demand and online reviews *Horizon: Short to medium term*	Online reviews, sentiment, textual features
This study	POS transaction counts	Time-weighted LSTM, random forest regressor, GP-FBWM	R^2, MSE, RMSE, MAE	5 yr, daily and monthly *Horizon: 12 mo and 356 d†*	POS count, mean distance, temperature and income, population, gender ratio, tourist attractions, month, day number, weekday, week number, holidays, subsidy deposits

3. Materials and Methods

In the initial step, the key indicators influencing monthly POS transaction counts are identified and prioritized. Since the initial set of candidate indicators was obtained from both the literature and expert evaluation, a structured weighting method was required to incorporate expert

judgment while maintaining consistency in the decision-making process. For this purpose, the GP-FBWM method is employed. The use of GP-FBWM is justified for three main reasons. First, compared to conventional pairwise comparison methods such as AHP, the Best-Worst Method (BWM) requires fewer comparisons, thereby reducing the cognitive burden on experts and improving the consistency of judgments (Rezaei et al. 2016). Second, the goal programming formulation helps overcome the possibility of multiple optimal solutions in the original BWM model and provides a unique set of weights (Rezaei et al. 2016; Amiri and Emamat 2020). Third, the fuzzy extension enables the model to handle uncertainty and imprecision in experts' judgments, which is particularly important in business and operational decision-making environments. Therefore, this study uses the GP-FBWM developed by Rostami et al. (2023) because it provides a transparent, consistent, and uncertainty-aware mechanism for prioritizing the retained indicators before constructing the forecasting dataset.

After selecting and prioritizing the most relevant indicators, a time-weighted Long Short-Term Memory (TW-LSTM) model based on the approach proposed by Zhao et al. (2017) is developed for monthly transaction forecasting over a 12-month horizon. Monthly POS transaction counts represent a multivariate time series influenced by historical transaction behavior and exogenous variables. Traditional statistical models such as ARIMA and SARIMA are widely used for time series forecasting; however, they are mainly designed to capture linear temporal structures and often require stationarity assumptions. Moreover, they may be less effective in modeling nonlinear relationships between transaction counts and external variables (Irribarra et al. 2024). Machine learning models such as Support Vector Regression, Random Forest, and Gradient Boosting can capture nonlinear relationships among variables, but they do not inherently model long-term temporal dependencies in sequential data (Xia et al. 2024). In contrast, LSTM networks are specifically designed to learn temporal dependencies and long-term patterns in time series data. For this reason, LSTM is more appropriate for the monthly forecasting level of the proposed framework, where the objective is to capture evolving transaction trends over time (Liu et al. 2018).

In this study, a time-weighted version of LSTM is adopted because recent observations may carry more relevant information for forecasting future POS transaction counts (Baz et al. 2024; Cai et al. 2024). POS transaction behavior can change over time due to market conditions, customer behavior, seasonal effects, and operational changes. Therefore, treating all historical observations with equal importance may reduce forecasting accuracy. The TW-LSTM model assigns greater importance to more recent observations through a time-weighting function, allowing the model to better reflect the evolving nature of transaction patterns.

At the second level of the proposed hierarchical framework, a random forest regressor (RFR) is used to estimate the daily transaction weight within each month. The target variable at this level is continuous, since it represents the proportion of monthly transactions assigned to each day. Therefore, the problem is formulated as a regression task. RFR is selected because the daily transaction weight is influenced by nonlinear interactions among calendar-based and event-driven features, such as weekdays, holidays, subsidy payment periods, and month-specific effects. Random Forest is well suited for this setting because it can capture nonlinear relationships, handle high-dimensional binary features, reduce overfitting through ensemble averaging, and provide useful feature importance information. Other machine learning models, such as Support Vector Regression, Gradient Boosting, XGBoost, and neural networks, also have advantages. However, RFR was preferred in this study because it offers a suitable balance between predictive performance, robustness, interpretability, and implementation simplicity for the daily weight estimation problem. In particular, its ability to handle many binary calendar-related variables and identify influential features makes it appropriate for explaining daily transaction distribution patterns.

Additionally, since the daily-level dataset contains a relatively large number of calendar-

based binary variables, feature selection is necessary to reduce model complexity and improve interpretability. In high-dimensional feature spaces, irrelevant or redundant variables may increase computational effort and reduce forecasting performance. Therefore, a wrapper-based forward feature selection (FFS) approach is applied. FFS is selected because it evaluates candidate feature subsets based on their actual contribution to model performance, rather than relying only on statistical correlation or filter-based criteria. This makes it suitable for identifying the most informative daily features for the RFR model.

Finally, after designing the forecasting framework, the performance of the proposed model is evaluated using appropriate accuracy metrics. In this study, R^2, RMSE, and MAE are used to assess predictive performance and compare the proposed approach with benchmark models. The three metrics are defined below.

- Coefficient of determination: This scale-free metric compares the sum of squared prediction errors with the sum of squared deviations from the mean of the observed values. R^2 is calculated based on Eq. (1).

$$R^2 = 1 - \frac{\sum(y_i - y^{\wedge})^2}{\sum(y_i - y^{-})^2} \tag{1}$$

where y_i indicates the actual value, $y^{\wedge}$ indicates the forecasted value, and y^{-} is the mean of the observed values. The maximum possible R^2 is 1, which indicates perfect predictions. A score of 0 matches the mean-based benchmark in Eq. (1). The score can be negative when predictions perform worse than this benchmark.

- Root mean square error (RMSE) is the square root of the mean squared error, as shown in Eq. (2), and is expressed in the same units as the target variable.

$$RMSE = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(y_i - y^{\wedge})^2} \tag{2}$$

- Mean absolute error (MAE) is the mean of the absolute differences between the observed and forecast values, as shown in Eq. (3).

$$MAE = \frac{1}{N}\sum_{i=1}^{N}|y_i - y^{\wedge}| \tag{3}$$

To validate the proposed framework, a real-world case study is conducted using data obtained from a PSP in Iran. The company operates in various domains, including the provision, installation, and maintenance of POS devices, as well as offering electronic payment services. In this study, the focus is specifically on POS transaction data. The dataset consists of historical transaction records at both monthly and daily levels, which are used for model development and evaluation. The company includes 41 branches across different provinces; however, to preserve data confidentiality, only one representative branch is presented and analyzed throughout the paper. The forecasting results for the remaining branches are provided in the supplementary material. Due to confidentiality agreements, the data cannot be made publicly available. However, the data may be provided by the corresponding author upon reasonable request.

4. Results and Discussion

4.1. Selecting the Most Important Indicators

The selection of influential exogenous variables was conducted through a multi-stage

process involving literature review, expert evaluation, data feasibility analysis, and GP-FBWM-based prioritization. First, an initial pool of candidate variables affecting transaction behavior was identified from previous studies on financial transaction forecasting and demand forecasting. These variables included demographic, economic, environmental, temporal, and operational indicators that have been widely used in the literature.

Next, the candidate variables were evaluated by a panel of five experts from the PSP industry, consisting of two planning managers, one operations manager, and two data analysts. The experts were selected based on their professional experience and familiarity with transaction analysis, POS operations, and operational planning in payment service systems. During the evaluation process, the experts assessed each variable based on its relevance, interpretability, practical importance, and data availability for POS transaction forecasting.

To balance forecasting relevance, data availability, and model interpretability, only variables supported by at least three experts were retained for further analysis. Variables with limited historical availability, high redundancy, unstable behavior, or low operational relevance were excluded from the candidate set. The seven retained variables were considered potentially relevant because they were supported by both the literature and the expert panel. However, collecting complete historical data for all retained variables across 41 branches over a five-year period was highly time-consuming and costly. Therefore, the retained variables were further evaluated using the GP-FBWM method to determine their relative importance and prioritize the most influential indicators for the forecasting model. This additional weighting step helped reduce unnecessary data collection effort while preserving the most relevant variables for monthly POS transaction forecasting. Table 2 presents the initial candidate variables identified from the literature and the results of the expert screening process. Subsequently, the final set of selected variables is summarized in Table 3.

Table 2. Initial candidate variables and expert screening results

Variable	Literature support	Experts supporting inclusion	Final status	Reason for exclusion
Number of POS terminals	(Arunraj et al. 2016)	5/5	**Retained**	—
Inflation rate	(Ordu et al. 2021)	2/5	Excluded	Lack of reliable monthly branch-level data
Fuel price	(Nawaz et al. 2026)	1/5	Excluded	Low operational relevance for POS transaction counts
Average distance between POS terminals	(Rajwani et al. 2017; Huber and Stuckenschmidt 2020; Hirche et al. 2021)	4/5	**Retained**	—
Average temperature	(Roque et al. 2019; Hirche et al. 2021)	4/5	**Retained**	—
Average household income	(Sobolevsky et al. 2017)	4/5	**Retained**	—
Exchange rate	(Sheykhi et al. 2026)	2/5	Excluded	Weak direct relationship with regional POS transaction counts
Population	(Roque et al. 2019)	5/5	**Retained**	—
Advertising campaigns	(Prihatiningsih et al. 2025)	2/5	Excluded	Historical campaign data were not consistently available
Population gender ratio	(Roque et al. 2019)	3/5	**Retained**	—
Tourist attractions	(Sobolevsky et al. 2017)	4/5	**Retained**	—
Internet penetration rate	(Wirajing et al. 2025)	1/5	Excluded	Limited temporal variability during the study period

Variable	Literature support	Experts supporting inclusion	Final status	Reason for exclusion
Unemployment rate	(Moridian et al. 2025)	2/5	Excluded	Low short-term explanatory power for monthly POS transactions

Table 3. Final retained variables affecting transaction counts

Variable	Definition
(1) Number of POS terminals	Number of installed POS terminals per month.
(2) Average distance between POS terminals	Mean distance across all pairs of installed POS terminals.
(3) Average temperature	Mean temperature for each month.
(4) Average household income	Mean household income for each month.
(5) Population	Number of people older than 18 years.
(6) Population gender ratio	Percentage of people older than 18 years who are female.
(7) Tourist attractions	Indicates whether the PSP branch's region experiences high tourism-related activity in a given month.

The retained variables were then prioritized using the GP-FBWM method to determine their relative importance for monthly POS transaction forecasting. The resulting GP-FBWM weights are presented in Table 4. As shown, POS numbers, tourist attraction, and average temperature received the highest weights, respectively, indicating that these variables are the most influential indicators among the retained candidate variables.

Table 4. GP-FBWM weights for the retained variables

Variables	POS numbers	Average distance	Average temperature	Average income	Population	Population gender ratio	Tourist attraction
Weights	**0.2721518**	0.1242566	**0.1669406**	0.1033924	0.0893320	0.06135578	**0.1825709**
ξ^*= 0.4809373 CI=6.69 → CR=$\frac{0.4809373}{6.69} = 0.071889$							

4.2. Monthly Transaction Count Forecasting (First Level)

Time series data typically exhibit several underlying components, including trend, seasonality, cyclic patterns, and irregular fluctuations. The trend component reflects the long-term direction of the series, while seasonality captures recurring patterns within fixed and relatively short time intervals. Cyclic patterns represent medium- to long-term fluctuations that occur over irregular periods, and irregular components correspond to random variations caused by unforeseen events. Understanding these components is essential for selecting an appropriate forecasting model. In this study, monthly POS transaction counts are analyzed as a time series, where both trend and seasonal patterns are expected due to evolving market conditions and recurring consumer behavior. In addition, irregular fluctuations may arise from unexpected events such as economic shocks or policy changes. Formally, a time series $Y(t)$ can be represented as a combination of its underlying components, including trend $T(t)$, cycle $C(t)$, seasonality $S(t)$, and irregular component $I(t)$. Identifying and appropriately modeling these components is a critical step in improving forecasting accuracy.

For the development of the monthly forecasting model, a dataset consisting of transaction counts and selected exogenous variables was constructed based on the outcomes of the feature selection process. The dataset includes monthly records spanning 2018–2023, with 69 observations for each branch, and comprises six variables: date, branch identifier, number of active POS devices, tourism status (a binary variable indicating whether the month corresponds to a high-tourism period), average temperature, and total monthly transaction count. The number of active POS devices is available on a monthly basis for each branch, while the tourism

variable is defined using expert knowledge and historical travel patterns in each province. For example, periods such as New Year holidays are typically associated with increased travel activity, whereas months such as Ramadan may correspond to reduced tourism in certain regions. In addition, average temperature data are obtained from historical weather records. Given the presence of 41 branches in the dataset, one representative branch, namely "Tehran North-East", is selected for detailed analysis in the main text to ensure clarity and preserve data confidentiality. The same modeling procedure is applied to all branches, and the corresponding results could be reported by the authors.

Figure 1 shows the monthly transaction counts for the selected branch and their decomposition. The original series reflects both short-term fluctuations and an overall increasing trend. The trend component, extracted using the Hodrick–Prescott filter, indicates a gradual upward movement in transaction counts, particularly from 2021 onwards. The cyclical component captures medium-term variations around the trend, which are less regular than seasonal patterns and may be influenced by broader economic conditions. The seasonality component reveals a recurring annual pattern, suggesting that transaction counts systematically vary across different months of the year. Finally, the residual component represents irregular variations that cannot be explained by trend, cycle, or seasonality.

Prior to model development, the time series data were preprocessed to ensure suitability for forecasting. First, potential outliers were examined using the interquartile range (IQR) method, in which observations falling below Q1−1.5×IQR or above Q3+1.5×IQR are identified as outliers. The analysis indicated that no significant outliers were present in the dataset; therefore, no further outlier treatment was required. Subsequently, the stationarity of the time series was assessed. Although deep learning models such as LSTM do not strictly require stationary data, this step is necessary to enable a fair comparison with classical time series models, including ARIMA, SARIMA, and SARIMAX, which rely on the stationarity assumption. A stationary time series is characterized by constant statistical properties, such as mean and variance, over time. To evaluate stationarity, the Augmented Dickey–Fuller (ADF) test was employed. In this test, the null hypothesis is that the series has a unit root, and a p-value lower than the significance level (α=0.05) indicates rejection of the null hypothesis, suggesting that the series is stationary. Based on the test results, the stationarity condition was examined before proceeding to the forecasting stage. The Augmented Dickey–Fuller (ADF) test was applied to assess the stationarity of the original time series. The test statistic was 1.08 with a p-value of 0.995, indicating that the null hypothesis of non-stationarity cannot be rejected at the 5% significance level. Therefore, the series is considered non-stationary. To address this issue, two transformations were examined: seasonal differencing and rolling standardization. Seasonal differencing (lag = 12) was employed to eliminate seasonal patterns, while rolling standardization centered and scaled the series using a three-month rolling mean and standard deviation. These transformations are formulated in Eqs. (4) and (5) (Darwish et al. 2020):

$$y_t^{diff} = y_t - y_{t-12} \tag{4}$$

$$y_t^{std} = \frac{y_t - \mu_t^{(3)}}{\sigma_t^{(3)}} \tag{5}$$

Here, y_t is the transaction count in month t, and $\mu_t^{(3)}$ and $\sigma_t^{(3)}$ are the rolling mean and standard deviation computed using a three-month window, respectively. After applying these techniques, the stationarity of the transformed series was re-evaluated using the ADF test. The results showed ADF statistics of −2.58 and −0.70, with corresponding p-values of 0.037 and 0.847 for the seasonally differenced and rolling-standardized series, respectively. At the 5% level, the reported p-values indicate rejection of the unit-root null for the seasonally differenced

series only. Following the preprocessing stage, the dataset was divided into training, validation, and test sets. Specifically, the first 48 months were used for model training, the next 12 months were allocated for validation to fine-tune model parameters, and the final 9 months were reserved for testing to evaluate predictive performance on unseen data. Finally, Figure 2 shows the observed monthly counts, the validation and test forecasts, and the final one-year-ahead forecast. It should be noted that the model is applied to the original (non-stationary) time series data, as LSTM-based models do not require stationarity assumptions.

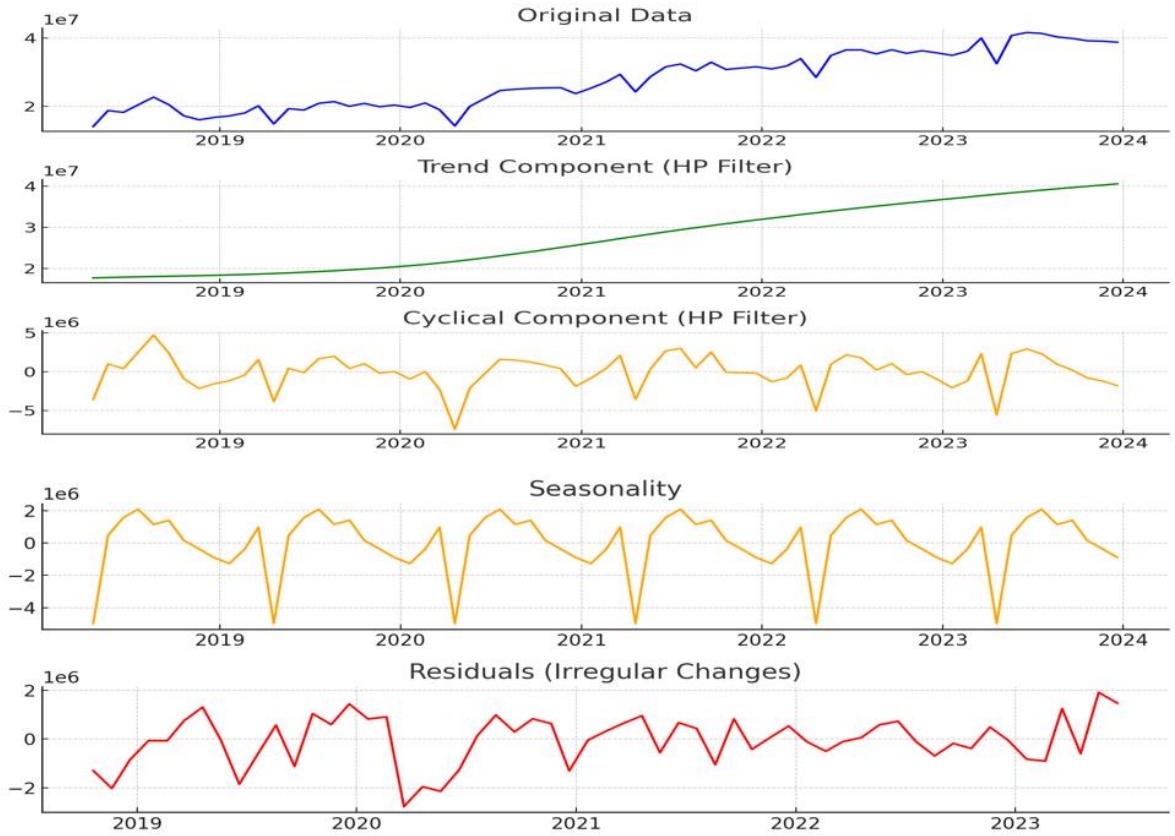

Figure 1. Decomposition of monthly POS transaction counts for the selected branch

Table 5 presents the evaluation results of the proposed forecasting framework in comparison with several benchmark time series models, including ARIMA, SARIMA, SARIMAX, and LSTM. Among these models, SARIMAX and LSTM incorporate exogenous variables, whereas ARIMA and SARIMA use historical transaction counts. The selection of these benchmark models is motivated by their widespread use in previous studies, as discussed in the literature review section.

To ensure a fair comparison, the hyperparameters of all models were carefully tuned using a grid search approach combined with a time-aware validation strategy, ensuring that the temporal ordering of the data was preserved during model evaluation. This process allowed for systematic exploration of different parameter combinations for each forecasting model. For the TW-LSTM model, the tuning process included not only architectural parameters such as the number of layers, neurons, learning rate, batch size, dropout rate, and activation function, but also different time-weighting formulations and decay parameter values. Specifically, several monotonically decreasing weighting functions, including exponential, linear, and logarithmic formulations, were empirically evaluated using different values of the decay parameter (ω). The candidate configurations were compared based on forecasting performance on the validation dataset. Among the tested alternatives, the exponential weighting function

($w(t)=e^{-\omega.t}$) with (ω=0.1) achieved the best overall forecasting performance and provided a suitable balance between emphasizing recent observations and preserving long-term temporal dependencies. The optimal hyperparameter settings for each model are reported in Table 6.

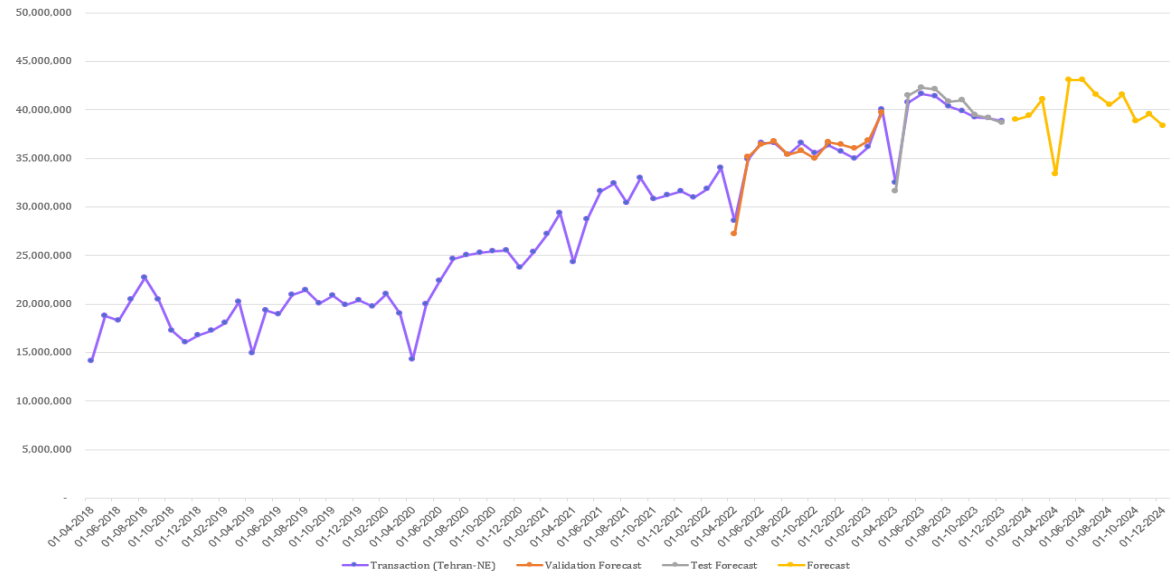

Figure 2. Observed monthly POS transaction counts and forecasts for the validation, test, and one-year-ahead periods

As shown in Table 5, the time-weighted LSTM model outperforms all benchmark models across both validation and test datasets. Specifically, it achieves the highest R^2 values (0.90 and 0.89 for validation and test sets, respectively) and the lowest error metrics (MAE and RMSE), indicating superior predictive accuracy. This improvement demonstrates the advantage of assigning greater importance to recent observations in time series forecasting, which enhances the model's ability to capture temporal dynamics more effectively than both classical statistical models and conventional deep learning approaches.

Table 5. Performance of monthly transaction forecasting models

Model Metric	ARIMA	SARIMA	SARIMAX	LSTM	Time-weighted LSTM
Validation					
R^2	0.84	0.85	0.86	0.87	0.90
MAE	670,000	645,000	630,000	600,000	516,000
RMSE	873,000	855,000	838,000	806,000	722,000
Test					
R^2	0.82	0.84	0.85	0.86	0.89
MAE	670,000	640,000	620,000	620,000	550,000
RMSE	845,000	835,000	820,000	825,000	775,000

Table 6. Selected hyperparameters for the monthly forecasting models

ARIMA	SARIMA	SARIMAX	LSTM	Time-weighted LSTM
p = 2, d = 1, q = 2	p = 2, d = 1, q = 2 P=2, D=1, Q=2, m=12	p = 2, d = 1, q = 2 P=2, D=1, Q=2, m=12 Considering selected exogenous variables	Number of layers=3 Number of neurons= 100 learning_rate=0.001 epochs=200 batch_size=64 activation='relu' dropout=0.2, recurrent_dropout=0.2	Number of layers=3 Number of neurons= 100 learning_rate=0.001 epochs=200 batch_size=64 activation='relu' dropout=0.2, recurrent_dropout=0.2 sample_weight=weight weight =w(t)=$e^{-\omega.t}$ (ω=0.1)

As indicated in Table 5, the proposed time-weighted LSTM model achieves the best forecasting performance across all evaluation metrics. It is followed by the standard LSTM model, while SARIMAX represents the best-performing statistical model, and SARIMA and ARIMA exhibit comparatively weaker performance. Models with exogenous inputs generally performed better. The time-weighted LSTM also outperformed the standard LSTM in this comparison. This analysis was conducted separately for each of the 41 branches to capture branch-specific characteristics and ensure appropriate model implementation. As shown, when non-stationarity was detected, suitable transformations such as seasonal differencing, rolling standardization, or a combination of both were applied to meet the requirements of classical time series models (i.e., ARIMA, SARIMA, and SARIMAX). However, the time-weighted LSTM model was trained using the original (non-stationary) data, as it does not rely on stationarity assumptions. Overall, the proposed framework demonstrates robust and consistent forecasting performance across different branches, and the total transaction count of the company is obtained by aggregating the forecasted values of all branches.

4.3. Daily Transaction Weight Forecasting (Second Level)

As previously mentioned, this study develops a two-level hierarchical forecasting framework for transaction data. A key question arises as to why the second level transitions directly from monthly to daily forecasting, without considering an intermediate level such as weeks. To address this, the transaction behavior of an example month for the "Tehran North-East" branch is analyzed at both daily and weekly levels, as illustrated in Figure 3.

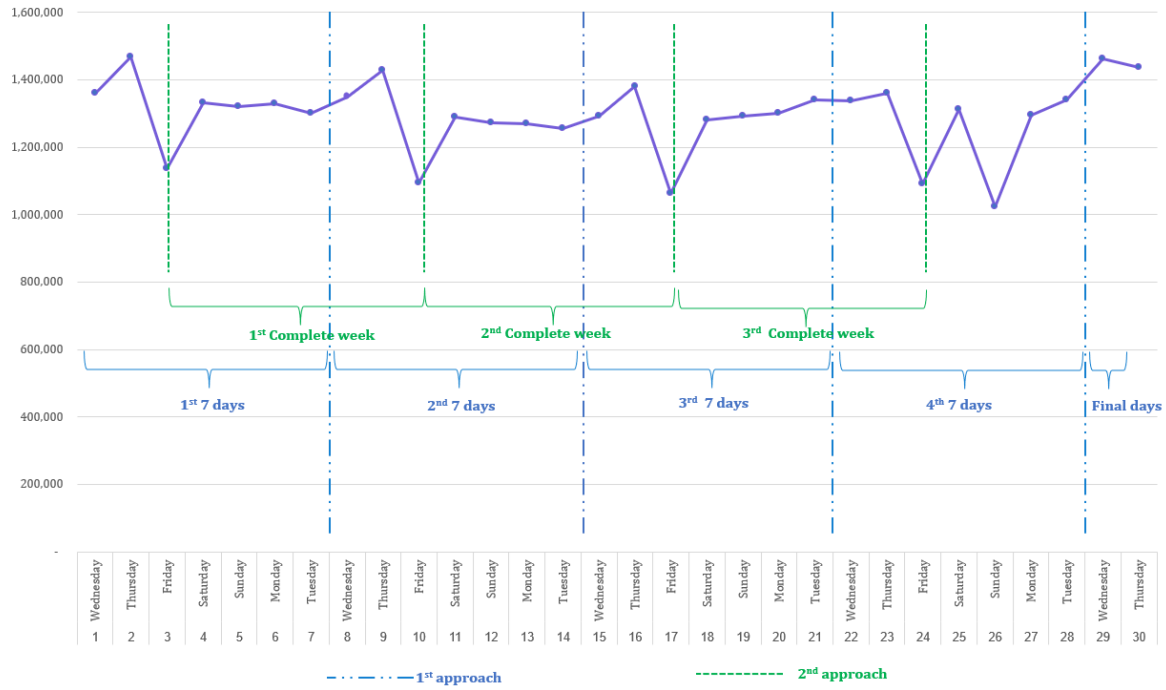

Figure 3. Daily POS transaction counts and two weekly grouping approaches for the selected branch

The primary reason for excluding the weekly level lies in the ambiguity associated with defining weeks within a month. Two common approaches exist. The first defines weeks as fixed 7-day intervals, resulting in four complete weeks and a number of remaining days that do not fit into any full week, leading to inconsistencies. Moreover, the composition of days within each 7-day interval may vary, causing fluctuations in transaction patterns. The second approach defines weeks based on the calendar (e.g., Saturday to Friday), where a month may contain three or four full weeks, again leaving some days unassigned. These inconsistencies make

weekly aggregation less reliable for modeling transaction behavior. In contrast, daily data provide a more precise representation of transaction patterns, as each day exhibits distinct characteristics influenced by behavioral, temporal, and contextual factors. Additionally, daily forecasting enables more effective monitoring of branch performance, ensuring alignment with the overall monthly plan.

Furthermore, the hierarchical structure links forecasts at the monthly and daily levels. By first modeling transactions at the monthly level, the framework captures overall transaction trends while reducing forecasting complexity. Subsequently, the monthly forecasts are disaggregated to the daily level, allowing finer-grained daily variations to be modeled without compromising consistency across temporal levels. Therefore, instead of directly forecasting daily transaction counts, the model predicts the proportion (weight) of each day within the total monthly transactions.

To ensure mathematical consistency within the hierarchical forecasting structure, the predicted daily transaction weights are normalized within each month before generating the final daily forecasts. Specifically, after the random forest regressor predicts the raw daily weights for all days of a given month, each predicted weight is divided by the sum of all predicted daily weights within the same month. Consequently, the normalized daily weights sum to 1 for each month, ensuring that the sum of the resulting daily transaction forecasts is exactly equal to the corresponding monthly forecast generated by the TW-LSTM model.

For the daily-level forecasting, a dataset comprising 2,102 daily observations was constructed for each branch. Each day is represented by a feature vector of 81 binary variables capturing its temporal and contextual characteristics, including the day of the month (31 features), month of the year (12 features), day of the week (7 features), week of the month (4 features), subsidy payment days, official Iranian holidays (25 features), and the day preceding a holiday.

The target variable is defined as the daily transaction weight, representing the proportion of total monthly transactions that occur on a given day. This formulation allows the model to capture the varying contribution of different days, rather than assuming a uniform distribution (e.g., 1/30), thereby reflecting the impact of daily characteristics on transaction intensity. Observed daily weights are calculated from historical transaction counts. Specifically, for each historical month, the daily weight is computed as the ratio of the daily transaction count to the corresponding monthly transaction count, and these historical weights serve as targets for fitting and evaluating the model. During the forecasting phase, actual future monthly transaction totals and future daily weights are not available and are therefore not used. Instead, the trained model predicts future daily transaction weights based only on known explanatory variables and calendar-related features. These predicted weights are normalized within each month and then multiplied by the forecasted monthly transaction counts generated by the TW-LSTM model in order to estimate daily transaction counts. The daily disaggregation step therefore uses forecast monthly totals and predicted weights to generate daily transaction counts.

In this framework, the daily transaction weight is considered as the dependent variable (Y), while the remaining features serve as independent variables (X). The dataset is divided into training (1,461 days), validation (365 days), and test (276 days) sets. A Random Forest Regressor (RFR) is employed to model this relationship due to its ability to capture nonlinear interactions and handle high-dimensional feature spaces effectively. To further assess the suitability of the random forest regressor, several alternative regression models were also evaluated during the model development stage, including Linear Regression (LR), Decision Tree Regressor (DTR), and Gradient Boosting Regressor (GBR). The comparison was conducted using the same training, validation, and test datasets for all models, with hyperparameters tuned by grid search. Table 7 summarizes the forecasting performance of the benchmark models on the test dataset.

Table 7. Performance of models for forecasting daily transaction weights

Model Metric	Random Forest Regressor (RFR)	Linear Regression (LR)	Decision Tree Regressor (DTR)	Gradient Boosting Regressor (GBR)
R^2	0.891	0.724	0.812	0.874
MAE	0.0025	0.0051	0.0038	0.0029
$RMSE$	0.0058	0.0094	0.0071	0.0061

The comparative analysis demonstrates that the random forest regressor provides the best overall forecasting performance among the evaluated models. Linear Regression showed limited capability in capturing the nonlinear and event-driven behavior of daily transaction patterns. Although the Decision Tree Regressor improved predictive flexibility, it was more sensitive to overfitting and produced less stable forecasting performance. Gradient Boosting Regressor achieved competitive results; however, the Random Forest Regressor provided a better balance between predictive accuracy, robustness, interpretability, and generalization capability across different branches. In addition, the ensemble structure of RFR effectively reduced model variance while maintaining strong performance in handling sparse calendar-based and event-related binary variables.

On the test set, the random forest regressor achieved an R^2 of 0.891, MAE of 0.0025, and RMSE of 0.0058, indicating that the model effectively captures the underlying patterns of daily transaction weights despite their inherent variability. The optimal configuration of RFR is determined as: *n_estimators*=100; *max_depth=10; min_samples_split=6; min_samples_leaf=3; max_features=auto, bootstrap=True*. This configuration consistently provides superior performance across most branches. Similar to the monthly-level analysis, the "Tehran North-East" branch is presented as a representative example, while the same procedure is applied to all other branches.

To further enhance model performance and reduce unnecessary complexity, a feature selection process was implemented using a forward feature selection (FFS) approach combined with fourfold cross-validation. In this method, features are progressively added to the model based on their contribution to improving predictive performance, and the optimal subset is selected according to validation results. This approach ensures that only the most informative features are retained, leading to a more efficient, interpretable, and robust model while reducing the risk of overfitting. The R^2 values reported in Table 8 represent the model performance across the validation folds during the feature selection process, demonstrating how the progressive inclusion of features enhances predictive accuracy until the highest-scoring subset among those evaluated is reached.

Table 8. Forward feature selection for forecasting daily transaction weights

Row	Feature names	CV scores	Avg. score
1	Week day_Friday	0.4862, 0.6021, 0.5328, 0.4892	0.5276
2	Week day_Friday, Before vacation	0.6174, 0.6152, 0.6399, 0.6941	0.6417
3	Week day_Friday, Before vacation, Month_Esfand	0.6895, 0.6714, 0.7016, 0.7156	0.6945
4	Week day_Friday, Before vacation, Month_Esfand, Subsidy payment	0.7862, 0.6954, 0.7395, 0.7010	0.7305
5	Week day_Friday, Before vacation, Month_Esfand, Subsidy payment, Holiday_Norooz	0.9211, 0.9846, 0.9657, 0.9388	0.9526
6	Week day_Friday, Before vacation, Month_Esfand, Subsidy payment, Holiday_Norooz, Holiday_Oil day	0.9511, 0.9785, 0.9835, 0.9455	0.9647

As shown in Table 8, the forward feature selection (FFS) results reveal an overall improvement in model performance as calendar features are added. With Friday as the sole feature, the model achieved a mean cross-validation R^2 of 0.5276. However, by adding features such as the day before holidays, the month of Esfand, subsidy payment periods, and major holidays such as Nowruz and Oil Day, the mean cross-validation R^2 increased to 0.9647 with all six features. This substantial increase highlights the critical role of calendar-based and event-driven variables in explaining variations in daily transaction patterns. Importantly, similar patterns were observed across all branches, where this specific group of features consistently led to the best performance. This indicates that these variables capture general behavioral and temporal effects in transaction data, making them highly generalizable across different regions. In particular, the final selected feature set—including Friday (weekend effect), the day before holidays, the month of Esfand, subsidy payment periods, Nowruz holidays, and Oil Day—captures key temporal and event-driven factors influencing transaction behavior. These variables reflect significant shifts in customer activity patterns, such as increased transactions during weekends, pre-holiday periods, major national events, and end-of-year seasonal peaks. Finally, once the daily transaction weights are forecasted, daily transaction counts can be forecast by multiplying the normalized predicted weights by the corresponding monthly transaction forecasts. This approach ensures consistency between the two hierarchical levels. Therefore, by having access to the calendar-based features for the upcoming year, along with a reliable 12-month-ahead forecast at the monthly level, the framework can generate daily transaction count forecasts for the entire year.

5. Conclusions

This study proposed a two-level hierarchical forecasting framework for predicting POS transaction counts in the electronic payment industry. At the monthly level, a set of relevant indicators was first identified based on the literature and expert knowledge. To reduce data collection complexity and focus on the most impactful variables, a goal programming-based fuzzy best–worst method (GP-FBWM) was applied to prioritize and select key indicators. The results revealed that the number of POS devices, tourist attraction, and average temperature are the most influential factors affecting transaction counts.

Based on the selected features, a time-weighted LSTM model was developed for monthly transaction forecasting and compared with commonly used time-series models, including ARIMA, SARIMA, SARIMAX, and standard LSTM. The results demonstrated that the proposed model outperformed the benchmark models across all evaluation metrics, achieving an average R^2 of approximately 0.83 and 0.81 on the validation and test sets across all branches, with peak values reaching 0.90 and 0.89, respectively. These findings confirm the effectiveness of the time-weighted LSTM model in capturing temporal patterns and improving forecasting accuracy.

At the second level, a random forest regressor (RFR) was employed to estimate daily transaction weights based on calendar-related features. The model achieved strong predictive performance, with an average R^2 of approximately 0.87–0.88 across all branches, ranging from 0.80 to 0.94. Using a forward feature selection (FFS) approach, a consistent set of key variables—including Friday, the day before holidays, the month of Esfand, subsidy payment periods, Nowruz holidays, and Oil Day—was identified as the most influential features. These variables capture significant behavioral and event-driven patterns in transaction data and were found to be generalizable across different branches.

The proposed hierarchical structure enables the generation of daily transaction forecasts by combining monthly predictions with daily weights. In this approach, the normalized forecast daily weights are multiplied by the corresponding monthly transaction forecasts, ensuring

consistency across different temporal levels. By leveraging known calendar information for the upcoming year along with reliable monthly forecasts, the framework provides a procedure for generating daily transaction count forecasts that sum to the corresponding monthly forecast.

From a theoretical perspective, this study contributes to the literature by integrating feature selection, deep learning, and ensemble learning methods within a unified hierarchical forecasting framework. Unlike conventional approaches that focus on a single forecasting horizon, the proposed model captures both monthly and daily dynamics, providing a more comprehensive understanding of transaction behavior.

From a managerial perspective, the findings offer valuable insights for payment service providers (PSPs) by providing monthly and daily transaction forecasts. The identification of key drivers such as POS availability, tourism patterns, temperature, and calendar-based events allows managers to better anticipate demand fluctuations and optimize resource allocation. This can lead to improved service quality, reduced operational costs, and enhanced customer satisfaction. Furthermore, the proposed framework can be adapted to other organizations or industries with similar transaction-based data, provided that model parameters are recalibrated accordingly.

Future research may explore the inclusion of additional external variables, the application of alternative machine learning or deep learning models, or the extension of the proposed framework to other transaction channels and forecasting contexts.

References

Aguiar-Pérez JM, Pérez-Juárez MÁ (2023) An insight of deep learning-based demand forecasting in smart grids. Sensors 23:1467. https://doi.org/10.3390/s23031467

Amiri M, Emamat MSMM (2020) A goal programming model for BWM. Informatica 31:21–34. https://doi.org/10.15388/20-INFOR389

Anderer M, Li F (2022) Hierarchical forecasting with a top-down alignment of independent-level forecasts. Int J Forecast 38:1405–1414. https://doi.org/10.1016/j.ijforecast.2021.12.015

Arunraj NS, Ahrens D, Fernandes M (2016) Application of SARIMAX model to forecast daily sales in food retail industry. Int J Oper Res Inf Syst 7:1–21. https://doi.org/10.4018/IJORIS.2016040101

Asad M, Shahzaib M, Abbasi Y, Rafi M (2020) A long-short-term-memory based model for predicting ATM replenishment amount. In: 2020 21st International Arab Conference on Information Technology (ACIT). IEEE, pp 1–6. https://doi.org/10.1109/ACIT50332.2020.9300115

Baz A, Logeshwaran J, Natarajan Y, Patel SK (2024) Enhancing mobility management in 5G networks using deep residual LSTM model. Appl Soft Comput 165:112103. https://doi.org/10.1016/j.asoc.2024.112103

Benchaji I, Douzi S, El Ouahidi B (2021) Credit card fraud detection model based on LSTM recurrent neural networks. J Adv Inf Technol 12:113–118. https://doi.org/10.12720/jait.12.2.113-118

Cai S, Gao H, Zhang J, Peng M (2024) A self-attention-LSTM method for dam deformation prediction based on CEEMDAN optimization. Appl Soft Comput 159:111615. https://doi.org/10.1016/j.asoc.2024.111615

Cucu FC, Sakar CO (2025) Transaction count prediction in correspondent banking using a hybrid CNN-LSTM neural network. In: Pattern Recognition and Artificial Intelligence: Selected papers from the 6th Mediterranean Conference on Pattern Recognition and Artificial Intelligence (MedPRAI24). Springer, pp 1255–1274. https://doi.org/10.1007/978-3-031-90893-4_84

Dai S, Meng F, Dai H, et al (2026) Machine learning in peak demand forecasting: foundations, trends, and insights. Renew Sustain Energy Rev 227:116500. https://doi.org/10.1016/j.rser.2025.116500

Darwish A, Rahhal Y, Jafar A (2020) A comparative study on predicting influenza outbreaks using different feature spaces: application of influenza-like illness data from Early Warning Alert and Response System in Syria. BMC Res Notes 13:33. https://doi.org/10.1186/s13104-020-4889-5

Dias SV, Al Mamun A, Alam MK, Zainol NR (2022) Predicting the adoption of mobile banking practices among Bangladeshi millennial's. In: Proceedings of International Conference on Emerging Technologies and Intelligent Systems: ICETIS 2021 (Volume 1). Springer, pp 429–441. https://doi.org/10.1007/978-3-030-82616-1_37

Falatouri T, Darbanian F, Brandtner P, Udokwu C (2022) Predictive analytics for demand forecasting – a comparison of SARIMA and LSTM in retail SCM. Procedia Comput Sci 200:993–1003. https://doi.org/10.1016/j.procs.2022.01.298

Feng Y, Yin Y, Wang D, Dhamotharan L (2022) A dynamic ensemble selection method for bank telemarketing

sales prediction. J Bus Res 139:368–382. https://doi.org/10.1016/j.jbusres.2021.09.067
García JR, Pacce M, Rodrigo T, et al (2021) Measuring and forecasting retail trade in real time using card transactional data. Int J Forecast 37:1235–1246. https://doi.org/10.1016/j.ijforecast.2021.02.005
Gehlot A, Ansari BK, Arora D, et al (2022) Application of neural network in the prediction models of machine learning based design. In: 2022 International Conference on Innovative Computing, Intelligent Communication and Smart Electrical Systems (ICSES). IEEE, pp 1–6. https://doi.org/10.1109/ICSES55317.2022.9914184
Giantsidi S, Tarantola C (2025) Deep learning for financial forecasting: A review of recent trends. Int Rev Econ Financ 104:104719. https://doi.org/10.1016/j.iref.2025.104719
He Z, Yu S (2020) Application of LightGBM and LSTM combined model in vegetable sales forecast. J Phys Conf Ser 1693:012110. https://doi.org/10.1088/1742-6596/1693/1/012110
Hirche M, Haensch J, Lockshin L (2021) Comparing the day temperature and holiday effects on retail sales of alcoholic beverages – a time-series analysis. Int J Wine Bus Res 33:432–455. https://doi.org/10.1108/IJWBR-07-2020-0035
Huber J, Stuckenschmidt H (2020) Daily retail demand forecasting using machine learning with emphasis on calendric special days. Int J Forecast 36:1420–1438. https://doi.org/10.1016/j.ijforecast.2020.02.005
Hussain F, Hasanuzzaman M, Abd Rahim N (2025) Multivariate machine learning algorithms for energy demand forecasting and load behavior analysis. Energy Convers Manag X 26:100903. https://doi.org/10.1016/j.ecmx.2025.100903
Irribarra N, Michell K, Bermeo C, Kristjanpoller W (2024) A multi-head attention neural network with non-linear correlation approach for time series causal discovery. Appl Soft Comput 165:112062. https://doi.org/10.1016/j.asoc.2024.112062
Jadwal PK, Jain S, Gupta U, Khanna P (2019) Clustered support vector machine for ATM cash repository prediction. In: Progress in Advanced Computing and Intelligent Engineering. Springer, pp 189–201. https://doi.org/10.1007/978-981-13-1708-8_18
Kallurkar HS, Chandavarkar BR (2024) A hybrid CNN–LSTM model for transaction fee forecasting in post EIP-1559 Ethereum. SN Comput Sci 5:638. https://doi.org/10.1007/s42979-024-02976-1
Kullaya Swamy A, Sarojamma B (2020) Bank transaction data modeling by optimized hybrid machine learning merged with ARIMA. J Manag Anal 7:624–648. https://doi.org/10.1080/23270012.2020.1726217
Kumar TKS, Thomas J, Thomas L, Menon VA (2026) Forecasting UPI transaction value in India: A machine learning regression approach to customer-initiated payments. In: Transforming E-Collaboration with AI and Emerging Digital Technologies. IGI Global Scientific Publishing, pp 33–56. https://doi.org/10.4018/979-8-3373-5747-8.ch002
Ładyżyński P, Żbikowski K, Gawrysiak P (2019) Direct marketing campaigns in retail banking with the use of deep learning and random forests. Expert Syst Appl 134:28–35. https://doi.org/10.1016/j.eswa.2019.05.020
Li X, Xu Y, Zhang X, et al (2023) Improving short-term bike sharing demand forecast through an irregular convolutional neural network. Transp Res Part C Emerg Technol 147:103984. https://doi.org/10.1016/j.trc.2022.103984
Liu S, Liao G, Ding Y (2018) Stock transaction prediction modeling and analysis based on LSTM. In: 2018 13th IEEE Conference on Industrial Electronics and Applications (ICIEA). IEEE, pp 2787–2790. https://doi.org/10.1109/ICIEA.2018.8398183
Moridian A, Radulescu M, Usman M, et al (2025) Unemployment rate and its relationship with government size, trade, inflation, urbanization, and economic growth in Romania. Heliyon 11:e41579. https://doi.org/10.1016/j.heliyon.2024.e41579
Nawaz A, Su CW, Khan SY (2026) How do climate and economic policy uncertainties relate to global fossil fuel price dynamics? North Am J Econ Financ 83:102594. https://doi.org/10.1016/j.najef.2026.102594
Ordu M, Demir E, Tofallis C, Gunal MM (2021) A novel healthcare resource allocation decision support tool: A forecasting-simulation-optimization approach. J Oper Res Soc 72:485–500. https://doi.org/10.1080/01605682.2019.1700186
Peng H, Lin Y, Wu M (2022) Bank financial risk prediction model based on big data. Sci Program 2022:3398545. https://doi.org/10.1155/2022/3398545
Prihatiningsih T, Panudju R, Prasetyo IJ (2025) Digital advertising trends and effectiveness in the modern era: a systematic literature review. Golden Ratio Mark Appl Psychol Bus 5:1–12. https://doi.org/10.52970/grmapb.v5i1.505
Rafi M, Wahab MT, Khan MB, Raza H (2020) ATM cash prediction using time series approach. In: 2020 3rd International Conference on Computing, Mathematics and Engineering Technologies (iCoMET). IEEE, pp 1–6. https://doi.org/10.1109/iCoMET48670.2020.9073937
Rafi M, Wahab MT, Khan MB, Raza H (2021) Towards optimal ATM cash replenishment using time series analysis. J Intell Fuzzy Syst 41:5915–5927. https://doi.org/10.3233/JIFS-201953
Rajwani A, Syed T, Khan B, Behlim S (2017) Regression analysis for ATM cash flow prediction. In: 2017 International Conference on Frontiers of Information Technology (FIT). IEEE, pp 212–217.

https://doi.org/10.1109/FIT.2017.00045
Razavi H, Sarabadani H, Karimisefat A, Lebraty JF (2019) Profitability prediction for ATM transactions using artificial neural networks: A data-driven analysis. In: 2019 IEEE 5th Conference on Knowledge Based Engineering and Innovation (KBEI). IEEE, pp 661–665. https://doi.org/10.1109/KBEI.2019.8735037
Rezaei J, Nispeling T, Sarkis J, Tavasszy L (2016) A supplier selection life cycle approach integrating traditional and environmental criteria using the best worst method. J Clean Prod 135:577–588. https://doi.org/10.1016/j.jclepro.2016.06.125
Roque L, Fernandes CAC, Silva T (2019) Optimal combination forecasts on retail multi-dimensional sales data. arXiv preprint arXiv:1903.09478. https://doi.org/10.48550/arXiv.1903.09478
Rostami O, Tavakoli M, Tajally A, GhanavatiNejad M (2023) A goal programming-based fuzzy best–worst method for the viable supplier selection problem: a case study. Soft Comput 27:2827–2852. https://doi.org/10.1007/s00500-022-07572-0
Sarveswararao V, Ravi V, Vivek Y (2023) ATM cash demand forecasting in an Indian bank with chaos and hybrid deep learning networks. Expert Syst Appl 211:118645. https://doi.org/10.1016/j.eswa.2022.118645
Shahabi V, Faezy Razi F (2019) Modeling the effect of electronic banking expansion on profitability using neural networks and system dynamics approach. Qual Res Financ Mark 11:197–210. https://doi.org/10.1108/QRFM-11-2018-0117
Sheng F, Jia L (2020) Short-term load forecasting based on SARIMAX-LSTM. In: 2020 5th International Conference on Power and Renewable Energy (ICPRE). IEEE, pp 90–94. https://doi.org/10.1109/ICPRE51194.2020.9233117
Sheykhi A, Taleblou R, Mohajeri P (2026) The impact of exchange rate volatility on economic growth: A meta-analysis approach. J Econ Stud 53:566–588. https://doi.org/10.1108/JES-01-2025-0023
Siyal AW, Donghong D, Umrani WA, et al (2019) Predicting mobile banking acceptance and loyalty in Chinese bank customers. SAGE Open 9:2158244019844084. https://doi.org/10.1177/2158244019844084
Sobolevsky S, Massaro E, Bojic I, et al (2017) Predicting regional economic indices using big data of individual bank card transactions. In: 2017 IEEE International Conference on Big Data (Big Data). IEEE, pp 1313–1318. https://doi.org/10.1109/BigData.2017.8258061
Vithitsoontorn C, Chongstitvatana P (2022) Demand forecasting in production planning for dairy products using machine learning and statistical method. In: 2022 International Electrical Engineering Congress (iEECON). IEEE, pp 1–4. https://doi.org/10.1109/iEECON53204.2022.9741683
Wirajing MAK, Moutie VG, Nkoa BEO, Ketchoua GS (2025) An analysis of the moderating role of Internet penetration in the banking competition-banks' profitability nexus: case of the CEMAC banking sector. Digit Financ 7:173–200. https://doi.org/10.1007/s42521-025-00129-4
Xia Y, Wang J, Zhang Z, et al (2024) A wind speed point-interval fuzzy forecasting system based on data decomposition and multiobjective optimizer. Appl Soft Comput 165:112084. https://doi.org/10.1016/j.asoc.2024.112084
Zhang D, Niu B (2024) Leveraging online reviews for hotel demand forecasting: A deep learning approach. Inf Process Manag 61:103527. https://doi.org/10.1016/j.ipm.2023.103527
Zhao Z, Rao R, Tu S, Shi J (2017) Time-weighted LSTM model with redefined labeling for stock trend prediction. In: 2017 IEEE 29th International Conference on Tools with Artificial Intelligence (ICTAI). IEEE, pp 1210–1217. https://doi.org/10.1109/ICTAI.2017.00184



This article is an open-access article distributed under the terms and conditions of the Creative Commons Attribution (CC-BY) license.