Restaurant Customer Flow Prediction Based on CNN-LSTM-Attention Model

Shaowei Dong

School of Economics and Management, Tongji University, Shanghai, China

Abstract: Customer flow prediction is essential for optimizing resource allocation and enhancing operational efficiency in restaurants. This paper addresses the challenge of predicting restaurant customer flow by incorporating external data, such as platform views and weather conditions, alongside feature derivation and selection to improve data quality. For the model design, we propose a hybrid prediction model based on CNN-LSTM-Attention. The CNN is used for local feature extraction, the LSTM captures the long-term dependencies of the time series, and the Attention mechanism dynamically focuses on key features. Through comparative and ablation experiments, we demonstrate that the proposed model significantly outperforms other benchmark models in metrics such as MAPE, RMSE, and R^2, indicating superior prediction performance. Additionally, SHAP-based explainability analysis further illuminates the influence of key features on the prediction results, enhancing the model's explainability and practical application value.

***Keywords*:** customer flow prediction; deep learning; feature engineering; explainability

1. Introduction

The service industry in China has experienced rapid growth in recent years, becoming one of the key drivers of economic expansion. As the economic structure transforms and consumption patterns evolve, the service sector's role in the Chinese economy has become increasingly vital, particularly with the rise of urbanization, technological innovation, and shifts in consumer demand. According to data from the National Bureau of Statistics, the service sector's share of China's GDP surpassed 50% in 2015 and continues to grow[1]. The restaurant industry is a crucial segment of this service sector. As consumer demands shift and lifestyles improve, the restaurant industry is undergoing significant transformation and upgrading. From traditional dining establishments to the emergence of food delivery platforms and consumer review platforms, the industry faces new challenges and opportunities in terms of service quality, operational efficiency, and customer experience.

Customer flow prediction is a crucial factor in restaurant operations and management, playing a significant role in improving efficiency, optimizing service processes, and reducing costs. By forecasting customer flow during different time periods, restaurants can better schedule staff shifts to avoid understaffing during peak hours or overstaffing during off-peak periods. Additionally, they can adjust purchasing volumes based on predicted customer flow, optimizing inventory management and minimizing waste or stockouts caused by surplus or shortages. Furthermore, customer flow predictions enable restaurants to implement dynamic pricing strategies — raising prices during peak hours and offering discounts during off-peak times to attract customers and increase revenue. In this way, customer flow prediction is essential for resource allocation, cost management, customer satisfaction, and marketing strategies. It helps restaurants make data-driven decisions, enhance operational effectiveness, and better navigate uncertainty and market fluctuations.

Traditional restaurant customer flow prediction methods are primarily based on time series models and machine learning algorithms, which often struggle to capture the complex dependencies between multi-dimensional data, limiting prediction performance. Additionally, these methods typically rely on internal historical data for forecasting, overlooking external factors such as platform views and ratings — indicators of customer purchasing intent — and weather conditions, which can significantly influence customer flow. While deep learning-based prediction methods offer improved accuracy, they often suffer from a lack of explainability. The internal mechanisms and decision-making processes of these models are difficult to understand, making it challenging for restaurant managers and operators to extract valuable business insights. This lack of transparency limits the practical application of these models and their ability to provide effective decision support.

In response to the limitations of current restaurant customer flow prediction methods, this paper proposes a hybrid model combining Convolutional Neural Networks (CNN), Long Short-Term Memory (LSTM) networks, and the Attention mechanism. Specifically, CNN is used to extract local features from time series data, LSTM captures long-term dependencies, and the Attention mechanism enables the model to focus on the time periods and features most influential to the prediction, thereby improving accuracy. Additionally, the paper incorporates external data sources—such as platform views and weather data — into the traditional customer flow dataset and applies feature engineering techniques. To enhance model explainability, SHAP (Shapley Additive Explanations) explainability method is introduced, helping restaurant managers identify the key factors influencing customer flow predictions. This enables more targeted optimization of operational strategies.

2. Related Works

Time series forecasting is a method used to predict future events, trends, or behaviors by analyzing patterns in past data. This approach helps address various real-world problems, including stock market analysis[2], residential load forecasting[3], traffic condition prediction[4], and influenza epidemic forecasting[5], among others.

Traditional time series modeling methods typically rely on statistical principles, with two main approaches. One approach is demand time series analysis based on the inertia principle, such as ARIMA[6], while the other uses causality-based prediction methods that consider influencing factors, like regression analysis[7]. These methods generally require the time series to be stationary and involve significant manual intervention, particularly when performing multivariate collaborative forecasting. As a result, they are more suitable for small-scale time series prediction problems. For larger datasets with multiple variables and dimensions, machine learning models—such as SVM[8], RF[9], and XGBoost[10]—often provide better forecasting performance due to their more complex structures and broader parameter considerations. Deep learning has also become an effective tool for time series forecasting, with prominent models including CNN[11], RNN[12], and attention networks[13].

In restaurant demand forecasting, existing studies have primarily used internal data, such as historical customer flow, menu prices, and promotional information, to build machine learning models for prediction[14]. Building on this, some researchers have incorporated external data, such as weather and epidemic information[15][16], or employed deep learning models to enhance forecasting accuracy[17][18].

3. Model Structure

3.1 CNN

CNN is a deep learning model designed for processing grid-like data. It extracts local features through convolution and pooling operations, offering advantages such as parameter sharing and spatial invariance. In time series forecasting, CNN efficiently captures local patterns within the data, identifying feature variations in short time windows and providing rich local feature representations for subsequent models.

3.2 LSTM

LSTM is a type of Recurrent Neural Network (RNN) designed to overcome the challenge of capturing long-range dependencies that traditional RNNs struggle with. It introduces memory cells and gate mechanisms (input, forget, and output gates) to retain important long-term information while filtering out irrelevant noise during training.

In time series forecasting, LSTM models long-term dependencies and captures the dynamic patterns of input data, making it well-suited for tasks that require temporal continuity and non-linear characteristics.

3.3 Attention

The Attention mechanism is a technique that has gained widespread use in deep learning, particularly excelling in natural language processing (NLP) and time series analysis. Its core idea is to enable the model to dynamically assign attention weights, focusing more computational resources on the input components that have the greatest impact on the output.

In time series forecasting, the Attention mechanism allows the model to automatically focus on key time periods and features while discarding less relevant information, thereby significantly improving forecasting accuracy.

3.4 CNN-LSTM-Attention

This paper combines CNN, LSTM, and the Attention mechanism to fully leverage their complementary strengths. The model structure is shown in Figure 1. First, a 1D CNN performs convolution operations along the time dimension to extract high-level features from local time windows. Then, LSTM models the dependencies between distant time steps, capturing both short-term and long-term dynamic changes. Finally, the Attention mechanism assigns weights to the features output by LSTM, dynamically focusing on key features at different time steps, thereby enhancing the model's ability to emphasize important information.

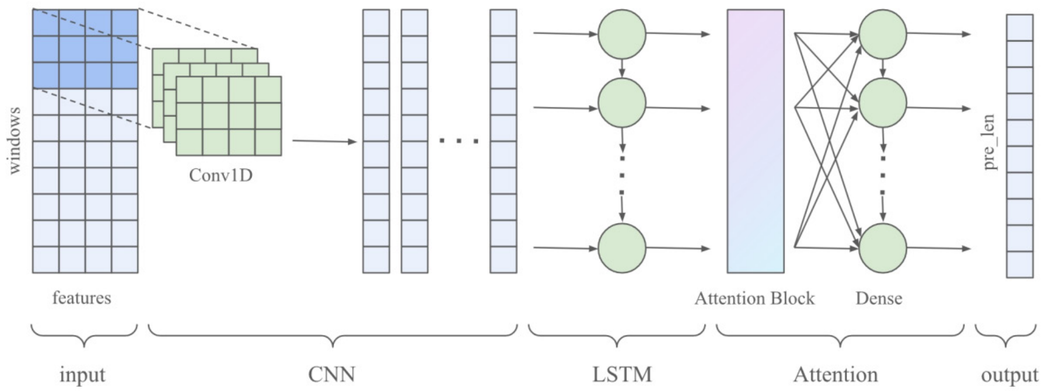

Figure 1. Model Structure

3.5 Evaluation Indicators

The model uses three evaluation metrics: Root Msean Square Error (RMSE), Mean Absolute Percentage Error (MAPE), and the coefficient of determination (R^2). RMSE reflects the absolute error between the actual and predicted values, MAPE indicates the relative error, and R^2 measures the goodness of fit.

$$RMSE = \sqrt{\frac{1}{n}\sum_{t=1}^{n}\left(y_t - y_t^{'}\right)^2} \tag{1}$$

$$MAPE = \frac{1}{n}\sum_{t=1}^{n}\left|\frac{y_t - y_t^{'}}{y_t}\right| \times 100\% \tag{2}$$

$$R^2 = 1 - \frac{\sum_{t=1}^{n}\left(y_t - y_t^{'}\right)^2}{\sum_{t=1}^{n}\left(y_t - \overline{y}\right)^2} \tag{3}$$

Where y_t is the actual value, $y_t^{'}$ is the predicted value, $\overline{y}$ is the mean of the actual values, and n represents the number of samples.

3.6 Explainability

Although deep learning models excel across various tasks, their complexity often makes the prediction process difficult to explain. This lack of explainability can limit their practical application. The SHAP (Shapley Additive Explanations) method provides a unified explanation framework based on Shapley values from game theory. It treats the model's predictions as the payoff generated by the collaboration of multiple features, enabling the measurement of each feature's contribution to the model's output.

In this study, we apply SHAP for the explainability analysis of deep learning models, aiming to uncover the model's reliance on input data and highlight the importance of key features.

$$\phi_i = \sum_{S \subseteq N \setminus \{i\}} \frac{|S|!\left(|N| - |S| - 1\right)!}{|N|} \cdot \left[f\left(S \cup \{i\}\right) - f\left(S\right) \right] \tag{4}$$

ϕ_i represents the Shapley value of feature i, N denotes the full set of features, S is a subset of features excluding feature i, and $f\left(S\right)$ represents the model's prediction when only the feature subset S is used.

4. Experiment

4.1 Data Description

This study uses customer flow data from a restaurant in the Tianchi dataset[19], covering the period from June 26, 2015, to October 31, 2016, for a total of 494 days. It also incorporates browsing data from a review platform and weather information, including the highest and lowest temperatures for each corresponding date, obtained from a weather website (https://lishi.tianqi.com/). The customer flow data, shown in Figure 2, exhibits a clear weekly periodicity.

A multi-step forecasting strategy is employed to predict customer flow for the next 14 days. The data is split into training and validation sets in an 8:2 ratio, with the test set consisting of the last 14 days of data.

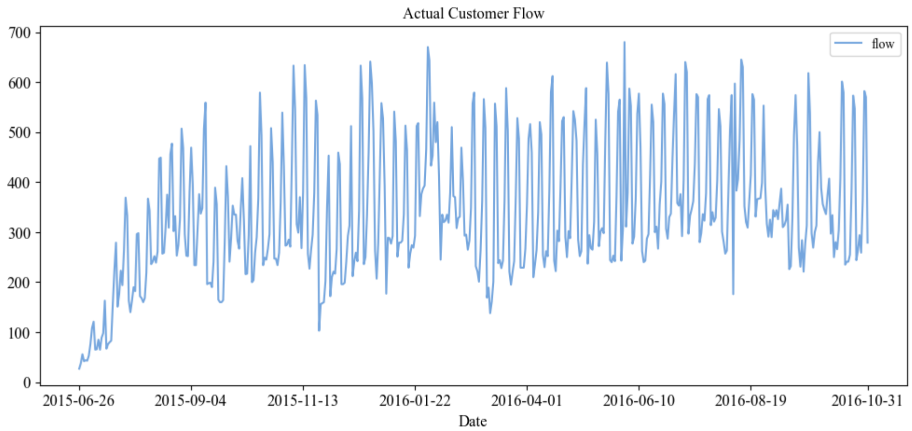

Figure 2. Actual Customer Flow

4.2 Data Preprocessing and Feature Engineering

4.2.1 Data Preprocessing

There are missing fields in the raw data, which require appropriate methods to fill in the gaps. In this study, we use techniques such as filling with nearby values, mean, mode, or specific values, based on the feature types and their meanings.

Standardization is a data scaling technique that adjusts the distribution of features, bringing them to a common range. This process helps mitigate the impact of scale differences on model training and can improve computational efficiency. To prevent data leakage, standardization parameters are calculated using only the training set, and these parameters are then applied to standardize the validation and test sets.

$$x^* = \frac{x-\mu}{\sigma} \tag{35}$$

4.2.2 Feature Derivation

Due to the limited number of features in the raw data, a significant number of derived features need to be constructed to enhance the model's fitting capability. Relevant features are created based on time period and weather factors. Additionally, drawing from the business context and historical data, a series of statistical features are generated using a sliding window approach. The specific features are listed in Table 1.

Table 1. Basic Features and Derived Features

Factors	Basic Features and Derived Features
Time-related	Day of the week
	Day of the month
	Weekends and holidays
Weather	Temperature
	Weather condition (sunny/rainy/cloudy/snowy)
Historcal flow	**Flow in** $[t-T,t]$
	Max, min, mean, std, and median of flow in $[t-T,t]$
Platform views	View in $[t-T,t]$
	Max, min, mean, std, and median of view in $[t-T,t]$

4.2.3 Feature Selection

Not all features derived through feature engineering are necessarily useful. An excessive number of irrelevant features can increase model complexity and negatively impact prediction accuracy. Therefore, this paper employs variance filtering for feature selection. Variance filtering evaluates the variance of each feature to determine its informational value. If a feature's variance falls below a certain threshold, it indicates minimal variation and likely contributes little to the model's predictive performance. Consequently, such features are excluded.

4.3 Model Training and Prediction

This study implements the complete model architecture using the PyTorch library in Python. The model's input data consists of time series with the shape $(windows, features)$. After feature extraction and processing through the CNN, LSTM, and Attention layers, the data is passed through a fully connected layer, which maps it into a sequence of length 14, representing the predicted results for the next 14 days.

The model's hyperparameters are set as follows: the window length is 20. The CNN layer uses one-dimensional convolution to extract local features from the time series, with 64 convolutional kernel channels, a kernel size of 3, and padding of 1. The number of neurons in the LSTM hidden layer is set to 128. The training settings are: batch size of 16, number of epochs set to 200, loss function as MSELoss, and the optimizer is Adam.

4.4 Analysis of Results

To validate the superiority of the CNN-LSTM-Attention model proposed in this study for customer flow prediction, we conducted comparison experiments with the XGBoost, Prophet, and ARIMA models. Additionally, to assess the contribution of each module, ablation experiments were performed using the CNN-LSTM, LSTM-Attention, and LSTM models.

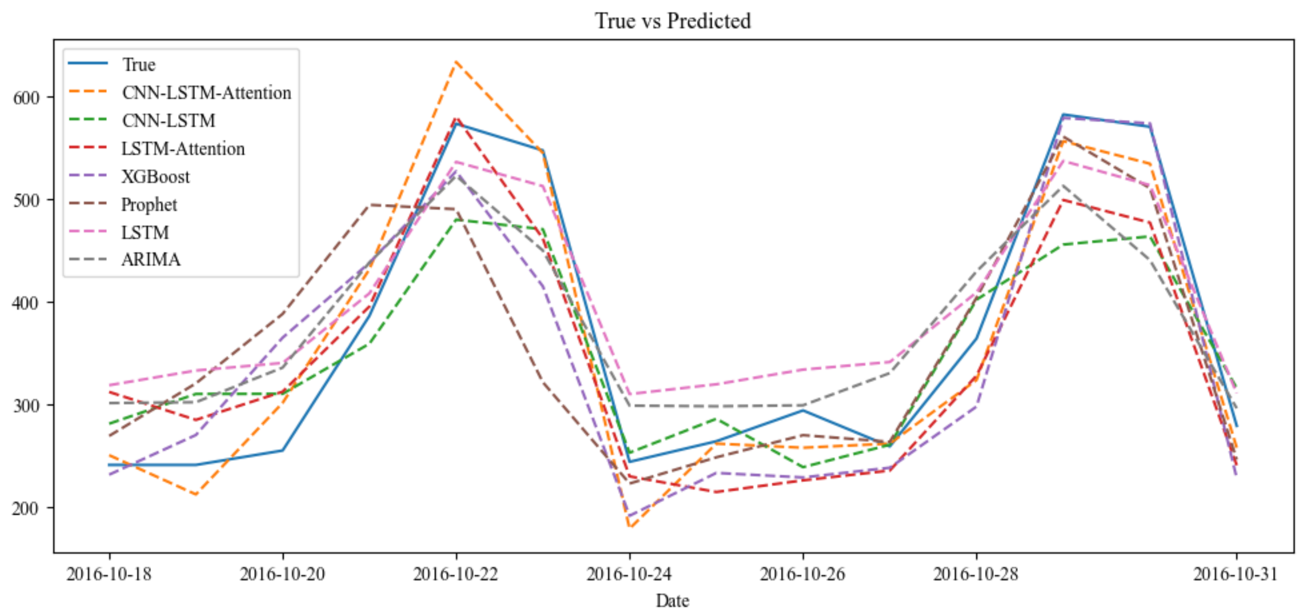

Figure 3. Prediction Results for Each Model

The prediction results of each model are shown in Figure 3. While almost all models can capture the periodicity in the time series, there are noticeable differences in prediction accuracy. The CNN-LSTM-Attention model, in particular, produces a prediction curve that closely aligns with the actual data curve, indicating the highest accuracy overall.

Table 2 presents the performance metrics for each model. It can be observed that the CNN-LSTM-Attention model achieves a relative error of 9.08%, an absolute error of 36.19, and a goodness of fit of 0.93, outperforming all other models across all evaluation metrics, demonstrating its excellent predictive capability.

When the Attention mechanism is removed, the CNN-LSTM model's ability to focus on key time points and features is insufficient, leading to an overall decline in performance. The LSTM-Attention model can focus on important time steps and features, but without CNN's local feature extraction capability, it struggles to capture fine-grained patterns. The LSTM model, lacking both local feature extraction and the support of the attention mechanism, performs worse than more complex combined models.

As a traditional tree-based model, XGBoost can handle nonlinear relationships but is less suited to model temporal dependencies in time series data, resulting in poorer performance compared to deep learning models. Prophet and ARIMA, which rely on strong assumptions, are unable to handle nonlinearity and multivariate features, leading to poor prediction accuracy.

Table 2. Model Performance

Model	MAPE/%	RMSE	R^2
CNN-LSTM-Attention	9.08	36.19	0.93
CNN-LSTM	14.26	64.66	0.77
LSTM-Attention	14.31	56.01	0.82
LSTM	18.35	59.05	0.81
XGBoost	14.68	60.15	0.80
Prophet	17.27	85.44	0.60
ARIMA	38.44	159.10	0.42

4.5 Explanation

Through SHAP value results shown in Figure 4, it was found that platform browsing volume has a significant impact on the model's predictions, with higher values generally having a positive effect on the predicted customer flow. This suggests that restaurant operators should focus on increasing exposure and click-through rates on review platforms. By promoting online activities (such as special promotions or featured positions), they can boost page views and, consequently, customer flow.

The influence of weekends on the model's output is also substantial, indicating that customer flow tends to be higher during weekends. As a result, restaurants can implement appropriate meal prep strategies and increase staffing levels to meet the peak demand on weekends. The impact of different weekdays on customer flow varies, suggesting that operational strategies should be adjusted according to the unique customer flow patterns of each day.

Weather factors, such as rainy days, typically lead to a reduction in customer flow, but their effect can fluctuate. Additionally, both current and historical customer flow statistics play a role, indicating that customer flow exhibits periodic and trending characteristics.

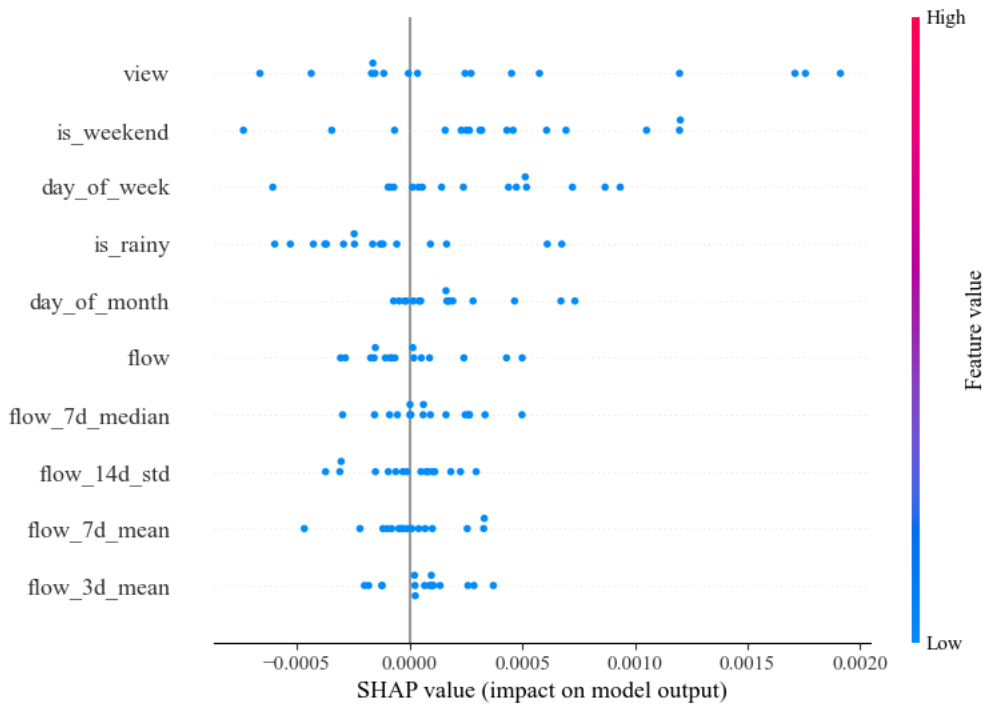

Figure 4. Feature Importance

5. Conclusion

This paper addresses restaurant customer flow prediction and proposes a hybrid model based on CNN-LSTM-Attention, which leverages both the characteristics of the data and external influencing factors. In the research, we first introduced external data sources, such as browsing volume from review platforms and weather data, and performed feature engineering and selection to enhance the quality and information density of the model inputs. In terms of model design, the CNN-LSTM-Attention model effectively integrates CNN's local feature extraction capabilities, LSTM's ability to capture long-term dependencies, and the dynamic focus provided by the Attention mechanism.

Through a series of comparative and ablation experiments, this paper demonstrates that the proposed model significantly outperforms traditional time series models (such as ARIMA and Prophet) as well as machine learning models (such as XGBoost) in the task of restaurant customer flow prediction. Moreover, it achieves optimal performance among deep learning ensemble models. Specifically, the proposed model yields the best results on key metrics, such as MAPE, RMSE, and R^2, highlighting its ability to capture complex time series patterns and the impact of external factors on customer flow. Additionally, the SHAP-based explainability analysis uncovers the key features and influencing factors behind the model's predictions, further enhancing its explainability and practical value.

In the future, the performance of the proposed model will be further evaluated using customer flow data from a variety of restaurant types and datasets from different scenarios, to assess its generalizability and robustness across broader contexts. Additionally, we plan to incorporate more diverse data sources, such as user text reviews, rating information from review platforms, and restaurant location attributes (e.g., business districts, foot traffic density, and nearby transportation facilities).

References

[1] National Bureau of Statistics of China (2024). *China Statistical Yearbook*. Beijing, China: China Statistics Press.

[2] Bao, W., Yue, J., & Rao, Y. (2017). A deep learning framework for financial time series using stacked autoencoders and long-short term memory. *PloS one*, *12*(7), e0180944.

[3] Kong, W., Dong, Z. Y., Jia, Y., Hill, D. J., Xu, Y., & Zhang, Y. (2017). Short-term residential load forecasting based on LSTM recurrent neural network. IEEE transactions on smart grid, 10(1), 841-851.

[4] Lv, Y., Duan, Y., Kang, W., Li, Z., & Wang, F.-Y. (2014). Traffic flow prediction with big data: A deep learning approach. Ieee transactions on intelligent transportation systems, 16(2), 865-873.

[5] Ceylan, Z. (2020). Estimation of COVID-19 prevalence in Italy, Spain, and France. *Science of The Total Environment*, *729*, 138817.

[6] Amini, M. H., Kargarian, A., & Karabasoglu, O. (2016). ARIMA-based decoupled time series forecasting of electric vehicle charging demand for stochastic power system operation. *Electric Power Systems Research*, *140*, 378-390.

[7] Cai, W., Peng, Y., & Chen, Q. (2019). Prediction of air transportation passenger volume based on multivariate regression model. *Aeronautical Computing Technique*, *49*(4), 50-53.

[8] Candelieri, A., Giordani, I., Archetti, F., Barkalov, K., Meyerov, I., Polovinkin, A., Sysoyev, A., & Zolotykh, N. (2019). Tuning hyperparameters of a SVM-based water demand forecasting system through parallel global optimization. *Computers & Operations Research*, *106*, 202-209.

[9] Qiu, X., Zhang, L., Suganthan, P. N., & Amaratunga, G. A. (2017). Oblique random forest ensemble via least square estimation for time series forecasting. *Information Sciences*, *420*, 249-262.

[10] Wang, Y., Sun, S., Chen, X., Zeng, X., Kong, Y., Chen, J., Guo, Y., & Wang, T. (2021). Short-term load forecasting of industrial customers based on SVMD and XGBoost. *International Journal of Electrical Power & Energy Systems*, *129*, 106830.

[11] Zeng, Z., Xiao, H., & Zhang, X. (2016). Self CNN-based time series stream forecasting. *Electronics Letters*, *52*(22), 1857-1858.

[12] Kim, T.-Y., & Cho, S.-B. (2019). Predicting residential energy consumption using CNN-LSTM neural networks. *Energy*, *182*, 72-81.

[13] Vaswani, A. (2017). Attention is all you need. *Advances in Neural Information Processing Systems*.

[14] Pandey, N. K., Mishra, A. K., Kumar, V., Kumar, A., Diwakar, M., & Tripathi, N. (2023). Machine Learning based Food Demand Estimation for Restaurants. 2023 6th International Conference on Information Systems and Computer Networks (ISCON).

[15] Tanizaki, T., Hoshino, T., Shimmura, T., & Takenaka, T. (2020). Restaurants store management based on demand forecasting. *Procedia CIRP*, *88*, 580-583.

[16] Yang, Y., Liu, H., & Chen, X. (2020). COVID-19 and restaurant demand: early effects of the pandemic and stay-at-home orders. *International Journal of Contemporary Hospitality Management*, *32*(12), 3809-3834.

[17] Schmidt, A., Kabir, M. W. U., & Hoque, M. T. (2022). Machine learning based restaurant sales forecasting. *Machine Learning and Knowledge Extraction*, *4*(1), 105-130.

[18] Chae, B. K., Sheu, C., & Park, E. O. (2024). The value of data, machine learning, and deep learning in restaurant de-

mand forecasting: Insights and lessons learned from a large restaurant chain. *Decision Support Systems*, *184*, 114291.
[19] Tianchi. (2018). *IJCAI-17 Customer Flow Forecasts on Koubei.com Dataset* https://tianchi.aliyun.com/dataset/dataDetail?dataId=58.

Author Bio

Shaowei Dong (2000.1—), male, Han nationality, Linquan County, Anhui Province, graduate student, research interests: machine learning Data Science.