Time Series Forecasting Using Context Rich Restaurant Sales Data in Data Science Education

Boyang Zhang, University of Turku, Finland

The European Conference on Education 2026
Official Conference Proceedings

Abstract

Restaurant sales forecasting in data science education is significant for decision making, human resource management, and inventory control. This paper proposes a decision-centric framework which includes context layer, data layer, model layer, and decision layers for the design and implementation of context-aware forecasting model. The Kaggle's Restaurant Sales Report 2024–2025 dataset is used in this research, it provides comprehensive daily sales records across multiple restaurant categories, enriched with contextual variables such as weather, promotions, special events, and pricing information. The methodology of this paper implements a structured analytical pipeline consisting of data cleaning, feature construction, time-series decomposition (trend and seasonality), and exploratory regression analysis using contextual variables such as weather, promotions, events, and prices. These steps are mapped to the proposed context, data, model, and decision layers to demonstrate how analytical outputs inform forecasting-related decisions. The findings of this research suggest that the proposed framework is suitable for forecasting the restaurant sales to practical decision making, by explicitly various layers, the framework helps to understand how data preprocessing, feature construction and modeling choices affect downstream outcomes, the prediction of the sales for the future. The framework supports the implementation of forecasting models that are transparent, reproducible and adaptable to various restaurant settings.

Keywords: data science education, time series forecasting, ensemble learning

Introduction

Nowadays, food waste in restaurants has become a global challenge, and it carries significant economic and environmental consequences. In order to understand the scale of the issue, recent studies estimate that the food service industry is responsible for approximately 26.1% of global food waste (Vardopoulos et al., 2025). Food waste brings negative influence on the daily running of restaurants; therefore, sales forecasting is vital important to inventory management, human resource management and sustainability management. Accurate sales forecasting is a challenging task, as reliable forecasting enables restaurants to optimize food purchases, staff schedules, pricing and even table allocations (Aulya, 2024; Boomija et al., 2018; Tanizaki et al., 2019). Difficulty arises because restaurant environments are dynamic, and with complex periodicities (A. Zhao & Jayadi, 2021; Tanizaki et al., 2020), the sales can fluctuate sharply by the time of year, month, week, holidays, and hours of a day (Güler et al., 2024; Žunić et al., 2020). At the same time, factors like weather, promotions, events as the various of context can also exert profound influence of sales in restaurants.

In order to make accurate forecasting, this paper presents a multi-layer decision centric process for sales prediction. It introduces a four-layer architecture which includes: context layer; data layer; model layer and decision layer. The four-layer architecture is designed to decompose data inputs and outputs, by structure data preprocessing, feature construction, modeling choices, and decision-making.

A key strength of this four-layer architecture is its universal application across most of the restaurant settings. Because this architecture can adjust its variables based on customized data, like local holidays, special events, building occupancies, and so on. It can be fitted with different business scales and dining concepts. With feature engineering of the historical dataset, the architecture can dynamically select the best model for prediction, and use the optimal weighting strategy for customized data. And with time gone by, the architecture will keep on updating and learning from historical data, and adjust the model and weight to make the most practical predictions. For example, the architecture provides its model layer with basic statistics analysis plus the latest analysis from non-linear variables. By combining predictive accuracy with computational efficiency, this four-layer architecture provides reliable, data-driven guidance for strategic business decision-making.

Theoretical Background and Proposed Framework

The concept of Information System (IS) theory is an interdisciplinary field connected with the collection, processing, storage and support for organization decision making (Davis, 1974). And Information System theory provides a diagram for understanding how organizations are able to capture complex contexts information and interpret it into organizational strategies (Hananiya et al., 2024). Especially in the food industry, organizational strategies are connected to the precise synchronization of production (Aulya, 2024), inventory management (Edmond Yeboah et al., 2022), and human resource management (Aich et al., 2026). The restaurant industry is characterized by the simultaneous production and consumption of perishable services and food materials. Inappropriate planning might leads to strategic failures, including massive food waste, suboptimal service quality, and extra labor costs (Koivunen et al., 2020; Vardopoulos et al., 2025). In order to avoid above risks, restaurant management requires an Information System that builds balance between external contexts and internal organizational strategies.

Restaurant sales data constitute a complex time series characterized by temporal dependencies, seasonality, and multiple contextual influences (Aulya, 2024). Variables such as weather conditions, localized special events, pricing strategies, and promotional campaigns introduce are difficult as the nonlinear disruptions to standard time series forecasting (Boomija et al., 2018; Panda & Mohanty, 2023). Traditional time series models (e.g., ARIMA) are difficult to capture these high variance anomalies. Moreover, machine learning algorithms (e.g., XGBoost, LightGBM), while adept at mapping nonlinearity, are prone to overfitting noisy historical data, with limited predictive accuracy (Aydın et al., 2025). Accurate forecasting remains challenging because of the complex temporal dynamics of restaurant sales data.

In order make accurate restaurant sales forecasting, this paper proposes a decision-centric framework which includes context layer, data layer, model layer, and decision layers. It is a four-layer context aware forecasting model for mapping the data flow of Information System from the real-world context, through data acquisition and processing, to decision support.

Figure 1

Four-Layer Decision Centric Framework

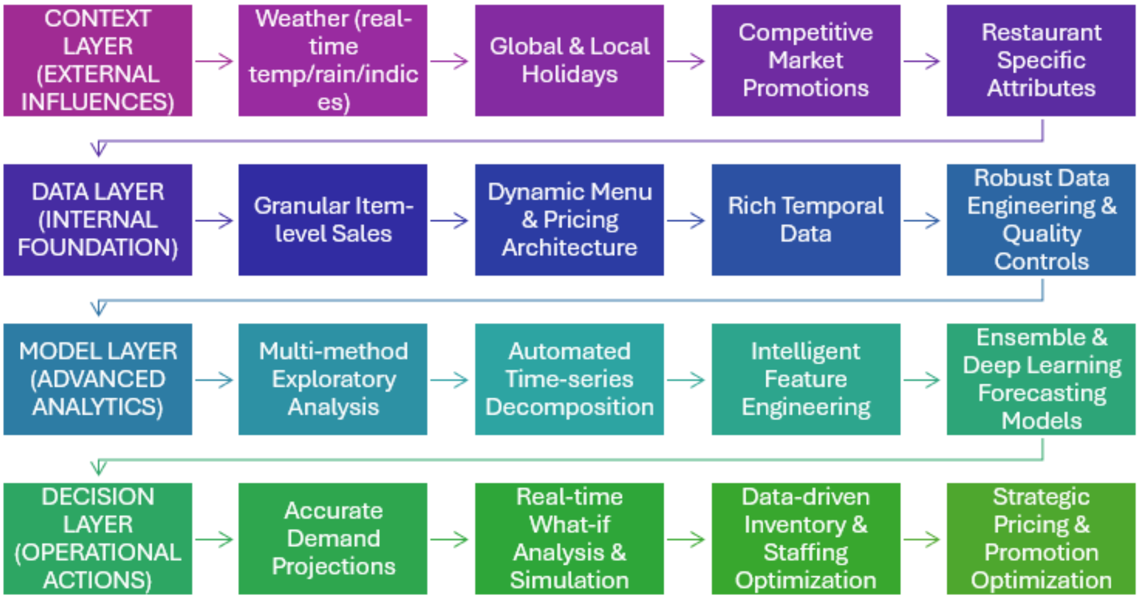


Figure 1 is the overall details of the four-layer decision centric frame. The first layer is the context layer. In IS theory, an open system must continuously interact with its external environment to remain viable and accurate (Davis, 1974; Hananiya et al., 2024). The context layer serves as the system's perceptual boundary, and this layer defines the multidimensional environment that the forecasting model needs to accommodate, encompassing variables such as real-time weather conditions which can include temperature and rain indices; global and local holidays; market promotions; and restaurant specific attributes which can include purchasing prices of product, human resource and so on. The data layer shows that system requires to collect possible external influences which might have an influence on sales.

The second layer is the data layer, it includes the storage of data and how data is cleaned. Because it is time series forecast, it needs to parse data into telemetry format. The granular item-level sales data emphasizes the use of transactional log data. This log data flows into dynamic menu and pricing architecture, such as the shifting of raw materials costs, menu updates, and promotion pricing. The relevant data sources are integrated as the rich temporal data, ensuring every transactional and financial metric is anchored with timestamp as the telemetry data. The entire second layer is about data engineer to clean the data, and make the data clean enough and get ready for next steps.

As the core of the system, the third model layer, can be regarded as the computational engine of the system. From the perspective of Information System, the computational model functions as the critical uncertainty reduction mechanism within the forecasting architecture (Taipalus et al., 2020). Multiple models are tested and evaluated based on the dataset. And a hybrid ensemble methodology could be used to make predictions based on accuracy from different models. The model layer is engineered to capture deterministic baseline patterns.

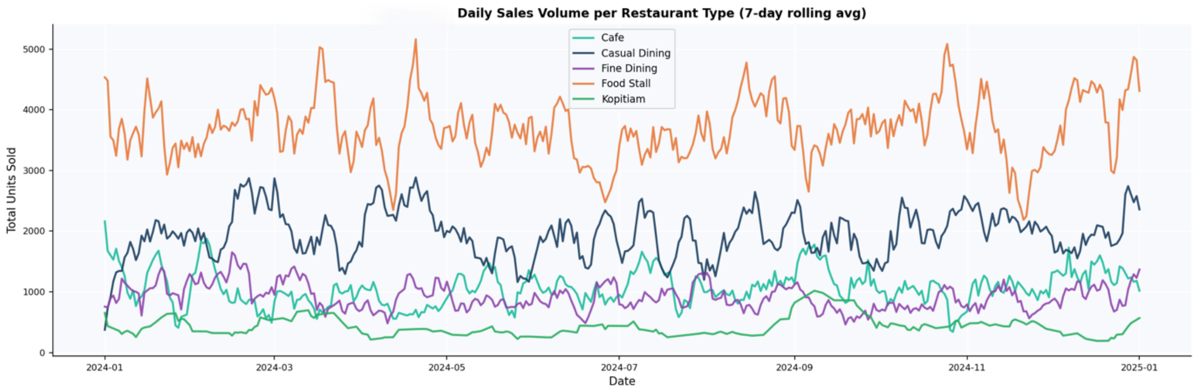
The last layer is the decision layer, it uses analytical outputs from model layer, and translates the predictions into corporate operation strategies. By directly informing critical organizational proposals, such as workforce scheduling, inventory optimization, and supply chain procurement. This layer interprets the prediction into suggestions on practical organizational decision making.

Kaggle Restaurant Dataset

In this chapter, the dataset is introduced, and decision-centric framework are discussed. The study uses the open public Kaggle Restaurant Sales Report 2024–2025 dataset, which provides an item level view of daily sales across fifty diverse restaurants over a full year, from January 2024 to January 2025. The establishments span five categories including cafes, casual dining, fine dining, food stalls, and kopitiams.

Figure 2

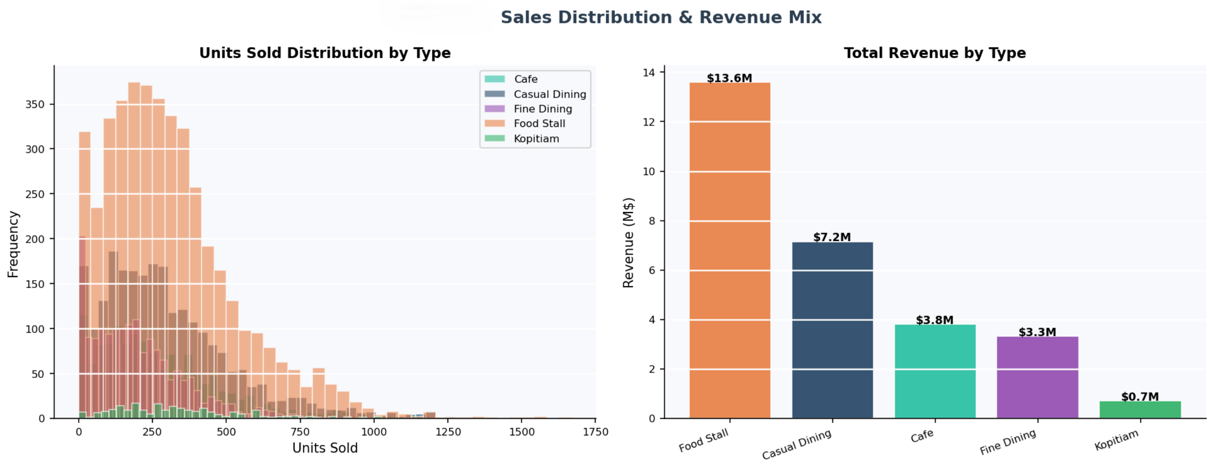
Daily Item Level Sales Volume per Restaurant Type



In order to get an overall view of the dataset, Figure 2 offers a longitudinal analysis of unit sales across five categories from January 2024 to January 2025. By utilizing a 7-day rolling average, it shows weekly and seasonal trends across the period of time. Food Stalls (represented by the orange line) are the undisputed leaders in sales volume. It fluctuated by sharp peaks in the springtime between March to April, and August as the beginning of school, alongside a noticeable mid-summer gap around July, and another peak before December. Casual Dining (dark blue line) maintains a solid and distinct second place, with total units sold generally fluctuating between 1,500 and 3,000. While its volume is substantially lower than that of Food Stalls, Casual Dining exhibits similar cyclical demand patterns throughout the year. The remaining three categories, Cafes, Fine Dining, and Kopitiams, operate at considerably lower sales volumes, also with fluctuations.

Figure 3

Sales Distribution and Revenue Mix



As to the sales distribution and revenue mix, we can see from Figure 3. The units sold distribution by restaurant type graphic, it shows a histogram detail of the frequency of daily units sold. The "Frequency" axis captures the exact number of restaurant days that achieved specific sales volumes across the dataset, 50 specific restaurants using daily temporal variables. Food Stalls dominate the distribution with the highest occurrence rates and the widest variance, recording days with well over 500 units sold and extending into a long tail past 1,250 units. Casual Dining forms the secondary volume tier, peaking in the 100 to 400 unit range. The remaining categories, which include Cafes, Fine Dining, and Kopitiam packed at the far left of the distribution, confirming that these establishments record lower daily transaction volumes, rarely exceeding 500 units on any given restaurant-day.

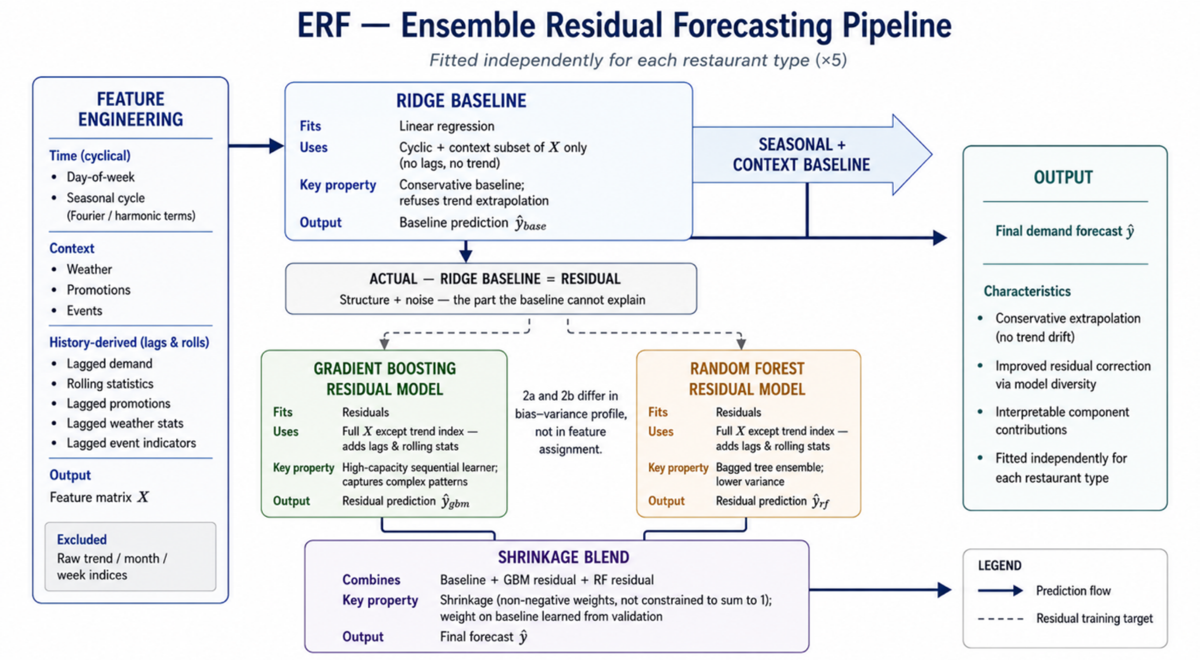
In the right graphic of Figure3, it is the analysis of "Total Revenue by Type" which shows a bar chart quantifying the cumulative financial output for each category in millions of dollars. Aligning with its sheer dominance in sales volume, the Food Stall category leads the revenue mix at $13.6M, almost doubling the earnings of the category of Casual Dining, which generated $7.2M. Cafes and Fine Dining follow with closely matched total revenues of $3.8M and $3.3M, respectively, highlighting how higher price points in fine dining can offset lower unit volumes to achieve revenue parity with higher volume cafes. Moreover, Kopitiam accounts for the smallest share of the financial mix, contributing $0.7M.

The Hybrid Model: Ensemble Residual Forecasting (ERF)

The hybrid model, Ensemble Residual Forecasting (ERF) is designed based on the four decision centric layers, it fosters the development of models that are transparent, reproducible, and adaptable to diverse restaurant settings. The concept of Ensemble Residual Forecasting represents a sequential pipeline, from the statistical baseline of linear factors to the dynamical machine learning model for nonlinear variance. The residual acts as the error or variance for optimization and updating the weighting strategies of different models.

Figure 4

ERF Ensemble Residual Forecasting Pipeline



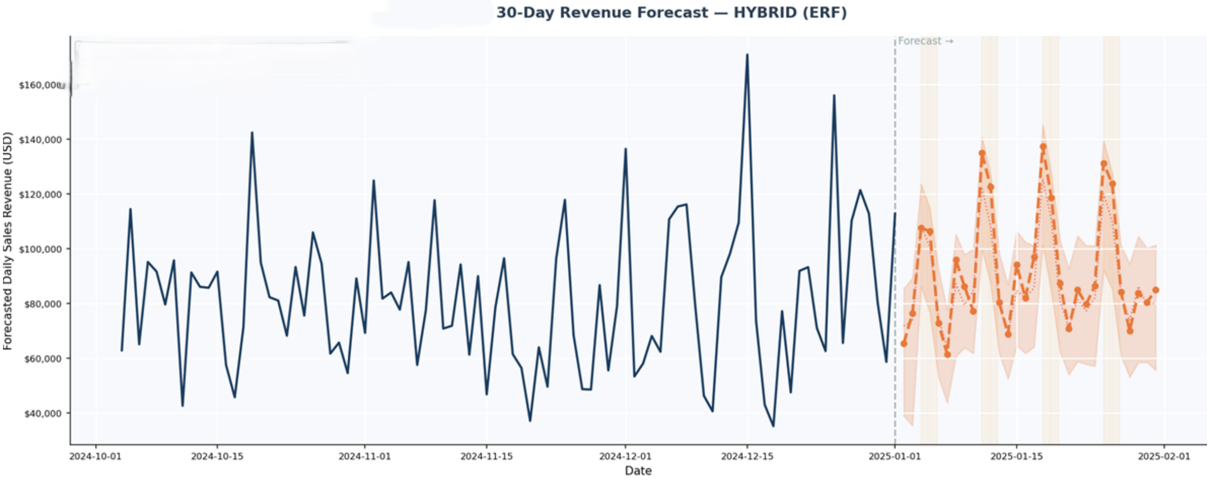
In the context layer, it begins with feature engineering of the Kaggle Restaurant Sales Report 2024–2025 dataset, which carries thirteen fields: date, restaurant_id, restaurant_type, menu_item_name, meal_type, key_ingredients_tags, typical_ingredient_cost, observed_market_price, actual_selling_price, quantity_sold, has_promotion, special_event, and weather_condition. There are multiple ways of classification of above variables, for example, three factors are engineered into daily context features: weather_condition (Sunny / Cloudy / Rainy) is encoded into (+1 / 0 / −1) and averaged per day; for has_promotion and special_event, boolean values are used. All fields in the dataset are categorized into variables for further analysis.

In the data layer, it serves as the foundational layer for building the comprehensive matrix X. It structures raw inputs into three distinct dimensions: cyclical time (day-of-week, Fourier terms) to encode stable seasonality, contextual variables (weather, promotions, events) to capture exogenous shocks, and history-derived telemetry (lagged demand, rolling statistics) to establish autoregressive momentum. Data layer is not fixed for different cases, in this research, the data layer is specifically designed for the Kaggle Restaurant Sales Dataset. It can be different in other cases, such as adding additional resources, like traffic flow variable, parking variable. In this case, for example, from the menu and pricing fields (menu_item_name, meal_type, key_ingredients_tags, typical_ingredient_cost, observed_market_price, and actual_selling_price), they are parsing into numeric financial signals. Categorical item details are mathematically encoded to represent menu composition, while continuous pricing data is transformed into elasticity metrics.

In the model layer, there are three stages in the process. First, a Ridge regression sets a stable, drift-free baseline using the cyclical and contextual features. Second, gradient boosting and random forest algorithms act as correctors; they are trained on the baseline's residuals, the errors, to reduce bias and variance. Finally, a shrinkage blend combines these outputs: the final forecast is built by taking the baseline prediction and adding the specific corrections from both the gradient boosting and random forest models, each scaled by its own learned weight. The weight strategies are dynamic and are updating based on the new input data as with the time gone by.

Figure 5

30-Day Revenue Forecast, the Hybrid Model



The decision layer acts as the system's final uncertainty reduction mechanism, grounding the mathematical forecasts in Information System. It interprets the hybrid outputs from the stable baseline and the specific residual corrections into actionable business intelligence. The forecast is delivered for next 30 days. By providing clear, interpretable signals with prediction, this layer empowers restaurant managers to optimize daily operations like inventory and staffing, successfully bridging the gap between predictive analytics and strategic decision-making. In Figure 5, we provide the forecast for next 30 days with a range prediction, for example one day's sale can between 85.000 to 115.000 dollar. Individual different types of restaurants can have different sales predictions based on previous sales data. Whatmore, restaurant managers can make monthly purchase of products plan. In terms of practical business deployment and model evaluation design, this study adopts a dynamic rolling strategy. To ensure that the system does not deviate from the actual sales, the framework incorporates a next day update mechanism: whenever new day's actual sales data is generated, the system automatically updates the lag features and recalculates the residual correction weights to forecast future sales. Therefore, this will offer guidance for restaurant managers who can adjust the weekly plan and monthly plan accordingly.

Discussion and Conclusion

In this chapter, a decision centric framework for restaurant sales forecasting is presented by utilizing a hybrid model, the Ensemble Residual Forecasting (ERF) model. There are four layers in the framework including context, data, model and decision layers, which decompose the data pipeline into structured process based on the Information System theory. It aims to make accurate forecasts on restaurant sales when encountering multiple linear and non-linear features.

In the evaluation, the hybrid model achieved optimal performance across key metrics. In Table1, the hybrid model's Mean Absolute Error (MAE) dropped to 126.6, meaning the daily sales forecast deviates by only about 127 units (meals/items). Combined with a Weighted Average Percentage Error (WAPE) of 11.4% and the highest explained variance ($R^2 = 0.476$), this variance is manageable for buffer stock control, proving the model's exceptional practical viability compared to other models. The evaluation demonstrates the effectiveness of the hybrid ERF model's design in applying the four-layer framework. The framework provides the data pipeline, and the ERF model provides practical solutions and interprets mathematical variables to business insights.

Table 1

Evaluation Results of the Models

Model	MAE	RMSE	R^2	WAPE (%)
Naive Baseline	302.6	359.2	1.209	27.4
GBM (standalone)	143.1	178.8	0.343	12.3
Ridge (statistical)	301.6	355.6	1.286	28.0
Stacking Ensemble	144.7	174.8	0.247	12.6
HYBRID (ERF)	126.6	162.0	0.476	11.4

While the four layer frameworks with ERF model demonstrate good performance, this study has several limitations that should be addressed. In the Kaggle Restaurant Sales Report dataset, it includes diverse operational contexts in five restaurant types, there could be more restaurant types, like holiday restaurants or institutional restaurants. Other types of restaurant data could also be tested. In order to filter out extreme daily noise, the ERF model's target variable was designed as a 7-day rolling average of units sold. While this is effective for weekly inventory procurement and staff scheduling, it masks sudden, single day sales spikes. If restaurant managers are looking for hourly or exact daily peak predictions, it would update the single day sales to hourly sales or in minutes sales, to catch up the extreme peaks for different angles of predictions. Moreover, the current pipeline forecasts total daily demand aggregated by restaurant type. It does not investigate the demand for individual menu items or specific raw ingredients, which would require handling even higher levels of data sparsity and noise.

In conclusion, this research has provided a general framework which can be applied across any restaurant settings, from local food stalls to fine dining chains. It delivers an accurate forecasting architecture that empowers restaurant managers with the early planning and decision-making capabilities based on different contexts, and it can help restaurant managers to optimize daily inventory, streamline labor scheduling, and reduce food waste. The four-layer framework is able to dynamically adjust its correction weights based on users' data. Future research should expand the ERF architecture to forecast demand for individual menu items, which would allow the system to generate direct, itemized procurement lists for kitchens. Additionally, the current ERF architecture relies on structured contextual features like weather and promotions, future iterations could integrate more non-linear variables, such as customer online feedback or social media sentiments and emotions, to better capture sudden changes in sales.

References

A. Zhao, M., & Jayadi, R. (2021). *Forecasting Daily Visitors and Menu Demands in an Indonesian Chain Restaurant using Support Vector Regression Machine*. 1–6. https://doi.org/10.1109/AIMS52415.2021.9466036

Aich, M., Kushwah, R., Bisui, R., & Tiwari, S. (2026). Strategic HR Interventions for Workforce Excellence in the Food Production Sector: A Study of Best Practices and Outcomes. *Advances in Consumer Research. Association for Consumer Research (U.S.)*, *3*, 350–359.

Aulya, N. (2024). Production Capacity Planning Strategy to Deal with Seasonal Demand in the Food and Beverage Industry. *Maroon Journal De Management*, *1*, 132–142. https://doi.org/10.37899/mjdm.v1i3.99

Aydın, B., Balcioglu, Y., & Sezen, B. (2025, March). *Food demand forecasting with machine learning methods: An application using decision tree*.

Boomija, G., Anandaraj, A., Nandhini, S., & Lavanya, S. (2018). Restaurant Visitor Time Series Forecasting Using Autoregressive Integrated Moving Average. *Journal of Computational and Theoretical Nanoscience*, *15*, 1590–1593. https://doi.org/10.1166/jctn.2018.7345

Davis, G. B. (1974). *Management Information Systems: Conceptual Foundations, Structure, and Development*. McGraw-Hill.

Edmond Yeboah, N., Opoku, R., & Kaku, G. (2022). Inventory strategies and performance of food and beverage processing industries. *International Journal of Logistics Systems and Management*, *41*, 120. https://doi.org/10.1504/IJLSM.2022.120985

Güler, A., Musa, A., Tarım, M., Saraç, O., & Gokturk, M. (2024). Forecasting Restaurant Sales with the Sensitivity of Weather Conditions and Special Days Using Facebook Prophet. *Journal of Data Applications*, 15–30. https://doi.org/10.26650/JODA.1450459

Hananiya, P., Badau, K., Olajire, F., & Yidawi, M. (2024). *Information systems in organizations*.

Koivunen, L., Rauti, S., Laato, S., Naskali, J., Nissilä, P., Ojansivu, P., Mäkilä, T., & Norrdal, M. (2020). Increasing Customer Awareness on Food Waste at University Cafeteria with a Sensor-Based Intelligent Self-Serve Lunch Line. *Proceedings of the IEEE International Conference on Smart Computing (SMARTCOMP)*. https://doi.org/10.1109/TMC49519-2030.9138571

Panda, S., & Mohanty, S. (2023). Time Series Forecasting and Modeling of Food Demand Supply Chain Based on Regressors Analysis. *IEEE Access*, *11*, 42679–42700. https://doi.org/10.1109/ACCESS.2023.3266275

Taipalus, T., Seppänen, V., & Pirhonen, M. (2020). Uncertainty in information system development: Causes, effects, and coping mechanisms. *Journal of Systems and Software*, *168*, 110655. https://doi.org/https://doi.org/10.1016/j.jss.2020.110655

Tanizaki, T., Hoshino, T., Shimmura, T., & Takenaka, T. (2019). Demand forecasting in restaurants using machine learning and statistical analysis. *Procedia CIRP*, *79*, 679–683. https://doi.org/10.1016/j.procir.2019.02.042

Tanizaki, T., Hoshino, T., Shimmura, T., & Takenaka, T. (2020). Restaurants store management based on demand forecasting. *Procedia CIRP*, *88*, 580–583. https://doi.org/https://doi.org/10.1016/j.procir.2020.05.101

Vardopoulos, I., Abeliotis, K., & Lasaridi, K. (2025). A Systematic Informetric Analysis and Literature Review of Food Waste Quantification Studies in the Food Service Industry. *Sustainability*, *17*(1). https://doi.org/10.3390/su17010103

Žunić, E., Korjenić, K., Hodžić, K., & Donko, D. (2020). Application of Facebook's Prophet Algorithm for Successful Sales Forecasting Based on Real-World Data. *International Journal of Computer Science & Information Technology (IJCSIT)*, *12*(2), 23–36. https://doi.org/10.5121/ijcsit.2020.12203