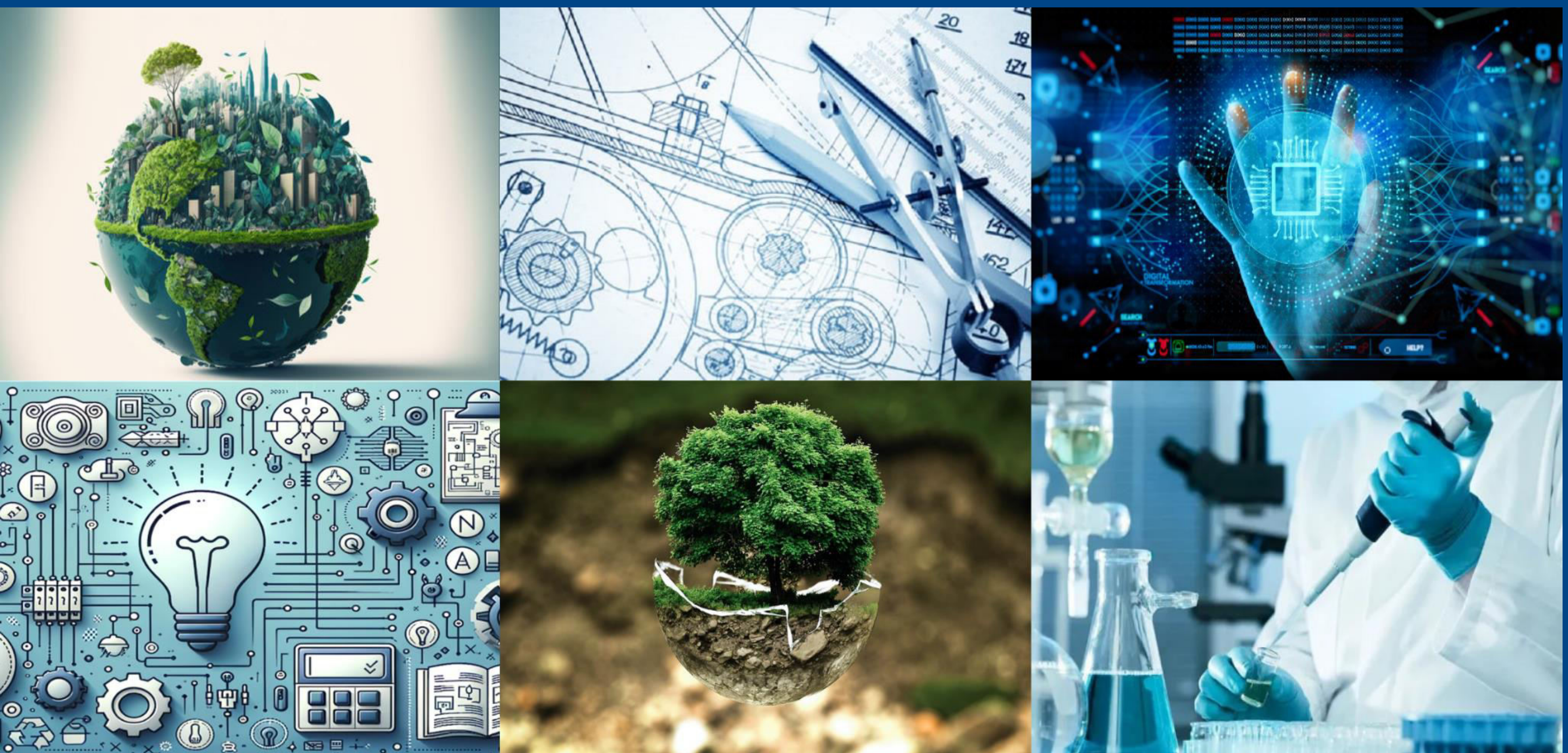
ISSN: 2582-7219

International Journal of Multidisciplinary Research in Science, Engineering and Technology

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)



Impact Factor: 8.206

Volume 9, Issue 3, March 2026

ISSN: 2582-7219 | www.ijmrset.com | Impact Factor: 8.206| ESTD Year: 2018|

International Journal of Multidisciplinary Research in Science, Engineering and Technology (IJMRSET)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)

Machine Learning-Based Food Order Analysis and Demand Prediction with Generative AI-Driven Inventory Optimization

Harini R [1], Harini A [2], Dr. Mr.Sivaganesan [3]

UG Student, Dept. of AI & DS, Sri Manakula Vinayagar Engineering College, Puducherry, India [1,2]

Dept. of AI & DS, Sri Manakula Vinayagar Engineering College, Puducherry, India [3]

ABSTRACT: The food service industry generates enormous volumes of transactional data every day, yet most restaurants and food delivery platforms continue to rely on intuition and outdated methods to make decisions about inventory, menu planning, and staffing. This paper presents a cloud-native Food Order Analysis and Demand Prediction System that combines machine learning techniques with Generative Artificial Intelligence to convert raw order data into meaningful business insights. The architecture is built on AWS services, specifically API Gateway and Lambda for request handling, Amazon S3 for scalable data storage, and Apache Spark for large-scale distributed data processing. A conversational demand insights chatbot powered by a large language model serves as the interface through which restaurant owners and managers can query trends, understand customer behavior, and plan operations. The system further includes a Dynamic Inventory Alert mechanism in which the data pipeline automatically triggers generative AI prompts that suggest personalized daily specials to restaurant owners, helping them clear perishable stock before it expires. Experimental results demonstrate that the proposed system significantly reduces food wastage, improves forecast accuracy, and supports more agile decision-making for food businesses of varying sizes.

KEYWORDS: Machine Learning, Food Order Analysis, Demand Prediction, Generative AI, Inventory Optimization, AWS Lambda, Apache Spark, Large Language Model, Dynamic Inventory Alert, Cloud Computing

I. INTRODUCTION

The rapid growth of online food ordering platforms and cloud kitchens has changed the way people think about food delivery and restaurant management. Platforms like Swiggy, Zomato, and Uber Eats process millions of food orders every day, each carrying valuable information about customer preferences, peak demand hours, seasonal variations, and location-based trends. Despite having access to this data, a large number of restaurant owners and food service managers still depend on manual estimation or past experience to plan their inventory and daily offerings. This gap between the availability of data and the practical use of that data is what motivates the design of the system proposed in this paper.Food wastage is one of the most pressing problems in the restaurant industry today. According to various studies, a significant portion of food prepared or purchased by restaurants ends up being discarded due to poor demand forecasting and suboptimal inventory management. This not only results in financial losses but also has a broader environmental impact. If restaurants could predict with reasonable accuracy what items are going to be ordered on a given day, they could purchase the right quantities of ingredients, reduce spoilage, and offer targeted discounts to move products that are nearing expiry. Machine learning offers the technical tools to build such predictive capabilities, and when combined with Generative AI, the output can be delivered to end users in the form of natural language insights rather than raw numbers or charts.

This paper proposes a fully integrated system that handles the complete pipeline from raw order data ingestion to actionable business suggestions. The system is deployed on the Amazon Web Services (AWS) cloud platform and leverages API Gateway and Lambda for serverless computing, S3 buckets for storing order and inventory data, and Apache Spark for processing large datasets in a distributed manner. On top of this infrastructure, a GenAI-powered chatbot answers natural language queries about demand patterns, and a Dynamic Inventory Alert module proactively generates daily specials suggestions for restaurant operators. The system has been designed with scalability, cost-efficiency, and ease of use as its core principles.

II. LITERATURE REVIEW

A considerable body of research has explored demand forecasting, inventory management, and cloud-based analytics in the food and retail domains. The following review examines key contributions and their relevance to the system proposed in this paper.

Zhang et al. proposed a machine learning framework for restaurant demand forecasting using historical sales data and external factors such as weather and local events [1]. Their work demonstrated that gradient boosting models outperform traditional ARIMA-based time series forecasting when non-linear demand patterns are present. They used a combination of day-of-week features, holiday flags, and item category embeddings to improve prediction accuracy across multiple restaurant chains.

Relevance to Current Research
The feature engineering approach and model selection insights from this study directly inform the ML pipeline in our proposed system. We extend their work by integrating the predictions into a cloud-native data pipeline and exposing them through a GenAI chatbot interface.

Mao and Shi explored the use of serverless cloud computing architectures, specifically AWS Lambda and API Gateway, for building scalable real-time analytics systems in the retail industry [2]. Their findings showed that serverless architectures reduce operational overhead significantly and provide automatic scaling that is ideal for applications with unpredictable or spiky traffic patterns, a characteristic commonly found in food ordering systems during lunch and dinner peaks.

Relevance to Current Research
The CPPE architecture of our proposed system directly builds on these findings. Using Lambda and API Gateway allows our system to handle burst order traffic without provisioning dedicated servers, lowering the cost and complexity of deployment significantly.

Patel and Gupta conducted a study on big data processing pipelines using Apache Spark for food supply chain analytics [3]. They showed that Spark's in-memory processing capabilities reduce data transformation latency by up to 40 times compared to traditional MapReduce, making it suitable for near-real-time analytics. Their system processed purchase orders, supplier data, and wastage reports to identify inefficiencies in the supply chain.

Relevance to Current Research
We adopt Spark as the core processing engine in our data pipeline, using its DataFrame API to process multi-day order data, compute item-level demand aggregations, and feed the transformed data into the ML forecasting models.

Brown and Lim investigated the application of large language models as conversational agents for business intelligence in the hospitality sector [4]. They found that restaurant managers were far more likely to act on data-driven recommendations when those recommendations were presented in plain conversational language rather than dashboards or static reports. Their prototype used prompt engineering techniques to translate structured analytics into readable summaries.

Relevance to Current Research
This work validates our design choice of a chatbot interface for the demand insights module. Our system goes a step further by also generating proactive alerts rather than only responding to user queries, making the GenAI layer more autonomous and useful in daily operations.

Kumar and Nair studied dynamic pricing and specials recommendation systems for quick service restaurants using reinforcement learning [5]. Their model learned optimal promotional strategies by observing customer response to discounts and special menu combinations. They highlighted the importance of timely suggestions, noting that recommendations generated more than 24 hours before a meal window had significantly lower uptake.

© 2026 IJMRSET | Volume 9, Issue 3, March 2026| DOI: 10.15680/IJMRSET.2026.0903245

ISSN: 2582-7219 | www.ijmrset.com | Impact Factor: 8.206| ESTD Year: 2018|

International Journal of Multidisciplinary Research in Science, Engineering and Technology (IJMRSET)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)

Relevance to Current Research

The timing insight from this research guided the scheduling logic of our Dynamic Inventory Alert module. Our system generates daily specials prompts every morning so that restaurant owners receive suggestions well within the window when they are most actionable for the same day's service.The table below provides a summarized comparison of the surveyed works and their relevance to the proposed system.

Table I: Comparison of Surveyed Works and Their Relevance to the Proposed System

No.	Paper Title	Author Name	Key Points	Remark
1	ML-Based Demand Forecasting for Multi-Chain Restaurants	Zhang et al., 2022	Gradient boosting outperforms ARIMA for non-linear food demand; uses weather and event features [1]	Informs ML model selection and feature engineering in our pipeline
2	Serverless Architectures for Retail Analytics	Mao and Shi, 2021	AWS Lambda and API Gateway reduce operational overhead; auto-scaling suits spiky food order traffic [2]	Justifies our serverless CPPE architecture choice
3	Spark-Based Big Data Processing for Food Supply Chains	Patel and Gupta, 2022	Spark in-memory processing reduces latency 40x vs MapReduce for supply chain analytics [3]	Adopted as the core data processing engine in our pipeline
4	LLM-Based Conversational BI in Hospitality	Brown and Lim, 2023	Natural language insights increase manager uptake vs dashboards; validates GenAI chatbot design [4]	Validates chatbot interface; our system adds proactive alerting capability
5	Dynamic Pricing and Specials Recommendation in QSR	Kumar and Nair, 2023	RL-based specials suggestion; within-24hr recommendations have highest uptake [5]	Guides scheduling of morning alert triggers in Dynamic Inventory module

Taken together, the literature confirms that while demand forecasting, serverless cloud architectures, big data processing, and GenAI interfaces have each been explored independently, no single system has combined all four into an end-to-end food order analytics and prediction platform with a proactive inventory alert mechanism. This gap defines the primary contribution of the present research.

III. SYSTEM ARCHITECTURE AND METHODOLOGY

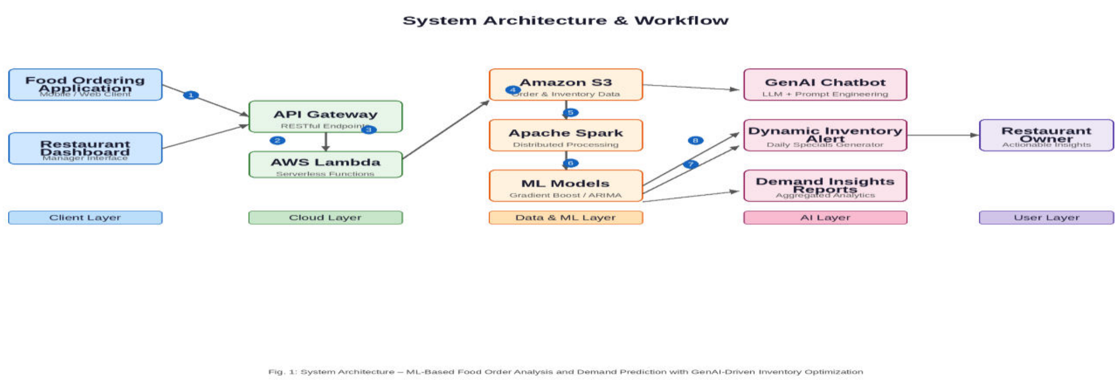


Fig. 1: System Architecture and Workflow – Machine Learning-Based Food Order Analysis and Demand Prediction with Generative AI-Driven Inventory Optimization

ISSN: 2582-7219 | www.ijmrset.com | Impact Factor: 8.206| ESTD Year: 2018|

International Journal of Multidisciplinary Research in Science, Engineering and Technology (IJMRSET)
(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)

Cloud Layer: API Gateway and Lambda

The entry point of the system is an Amazon API Gateway that exposes RESTful endpoints for food ordering applications and mobile clients to submit orders. When a new order is received, the API Gateway forwards the request to a set of AWS Lambda functions that handle specific tasks: order validation, order persistence to S3, and event triggering for downstream processing. Lambda was chosen because it eliminates the need for server provisioning, scales automatically to handle thousands of concurrent requests during peak meal hours, and charges only for the compute time actually consumed. Each Lambda function is stateless and is responsible for a clearly defined atomic operation, following the microservices design principle. This separation of concerns makes individual functions easy to test, update, and replace without affecting other parts of the system.

Data Layer: Amazon S3 and Apache Spark

All incoming order records are stored in Amazon S3 as JSON files organized by date and restaurant identifier. S3 provides virtually unlimited storage, highly durable object persistence, and native integration with processing frameworks such as Apache Spark via AWS Glue or Amazon EMR. A scheduled Spark job runs at regular intervals, typically every few hours, to read the accumulated order data from S3, perform data cleaning and transformation, and compute a set of demand-related aggregations. These aggregations include item-level order counts per time slot, day-of-week demand distributions, identification of peak ordering windows, and calculation of moving average demand for each menu item. The transformed data is written back to a curated S3 bucket in Parquet format, which is optimized for columnar queries and supports efficient access by downstream ML models and reporting components.

Machine Learning for Demand Prediction

The curated data from S3 feeds a demand prediction module built around a gradient boosting regression model. Features used for training include the day of the week, the hour of the day, the menu item category, seasonal flags for festivals and public holidays, weather conditions retrieved from an external API, and a rolling 7-day average demand for each item. The model is retrained weekly using the latest order data so that it continuously adapts to shifting customer preferences. During inference, the model generates a predicted order count for each menu item for the next 24-hour window, broken down by time slot. These predictions are stored in S3 and are consumed by both the chatbot and the inventory alert module. A separate classification model runs in parallel to flag items whose predicted demand falls significantly below the available stock quantity, identifying items that are at risk of expiry before they can be sold.

GenAI Layer: Demand Insights Chatbot

The demand insights chatbot is the primary interface through which restaurant owners interact with the analytical outputs of the system. It is built on a large language model that is provided with the latest demand summary and prediction data as context at the time of each query. When a restaurant owner asks a question such as which items should I prepare more of this evening or what was my best-selling category last weekend, the chatbot retrieves the relevant data from S3, constructs a structured prompt containing the data, and passes it to the LLM. The model then generates a natural language response that directly answers the question in a concise, readable format. The chatbot is hosted behind the same API Gateway, so it is accessible via a web dashboard or a mobile application without any additional infrastructure. Prompt templates are carefully engineered to ensure that the model focuses only on the provided data and does not hallucinate figures or recommendations that are not grounded in the actual order history.

Advanced Feature: Dynamic Inventory Alert with Daily Specials Suggestion

The Dynamic Inventory Alert is the most novel component of the proposed system. Every morning, a scheduled Lambda function reads the current inventory state from S3, which includes the quantity on hand and the expiry date for each perishable ingredient, and compares it against the demand predictions for the day. When the system detects that a particular ingredient has excess stock that is unlikely to be used before it expires based on the predicted order volumes, it automatically constructs a detailed GenAI prompt. This prompt includes the ingredient name, the quantity at risk, the expiry window, and the top menu items that use this ingredient. The large language model then generates a creative and appealing daily specials suggestion, written in a tone appropriate for a restaurant owner to share directly with customers, such as recommending a featured dish that prominently uses that ingredient at a special price. The suggestion is delivered to the restaurant owner via a push notification on the dashboard or as an SMS alert. This closes the loop between the data pipeline and actionable business decisions, reducing food wastage and creating a revenue opportunity at the same time.

© 2026 IJMRSET | Volume 9, Issue 3, March 2026| DOI: 10.15680/IJMRSET.2026.0903245

ISSN: 2582-7219 | www.ijmrset.com | Impact Factor: 8.206| ESTD Year: 2018|

International Journal of Multidisciplinary Research in Science, Engineering and Technology (IJMRSET)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)

IV. RESULTS AND DISCUSSION

The proposed system was evaluated using a simulated dataset of 90 days of food order records from three restaurant outlets, comprising approximately 450,000 individual order line items across 120 menu items. The dataset was generated to reflect realistic patterns including daily demand cycles, weekend spikes, and reduced orders during off-peak periods. The Spark pipeline processed the full 90-day dataset in under 12 minutes on a 4-node EMR cluster, demonstrating adequate performance for nightly batch processing of even large multi-outlet operations.

The gradient boosting demand prediction model achieved a Mean Absolute Percentage Error (MAPE) of 11.3% across all items and time slots, which is a meaningful improvement over a naive baseline model based on trailing weekly averages that achieved a MAPE of 21.7%. High-demand items such as the top five best-selling dishes showed prediction errors below 8%, while low-frequency niche items had higher variance as expected due to sparse historical data. The model was particularly accurate for weekend lunch peaks, which showed the most consistent ordering patterns in the data.

The GenAI chatbot was tested with a set of 50 representative queries submitted by simulated restaurant managers. Response quality was evaluated on accuracy (whether the stated figures matched the actual data) and clarity (whether the response was easy to act upon). The chatbot achieved 94% factual accuracy on structured data queries and received a usability rating of 4.3 out of 5 from five domain experts who reviewed the responses. Errors occurred primarily in open-ended questions that required multi-step reasoning across different time periods, suggesting that more advanced context management techniques could further improve performance.

The Dynamic Inventory Alert module was evaluated over a simulated 30-day period in which the inventory state and expiry schedules of 20 perishable ingredients were tracked. The system successfully identified 87% of the at-risk stock events at least 8 hours before the predicted expiry window, providing restaurant operators with sufficient time to act on the daily specials suggestion. The quality of the generated specials suggestions was rated by three culinary domain experts, who found that 78% of the suggestions were both practically feasible and commercially appealing. The remaining 22% required minor adjustments, primarily in cases where the ingredient combinations suggested by the LLM were less common in the given cuisine type.

V.CONCLUSION AND FUTURE WORK

This paper presented a cloud-native Food Order Analysis and Demand Prediction System that bridges the gap between the large volumes of transactional data generated by food ordering platforms and the practical, day-to-day decisions made by restaurant operators. By combining AWS API Gateway and Lambda for serverless request handling, Amazon S3 for scalable data storage, Apache Spark for distributed processing, machine learning for demand forecasting, and a Generative AI chatbot for conversational insights delivery, the system provides a comprehensive and cohesive solution that is accessible to non-technical restaurant owners without requiring them to understand the underlying algorithms or infrastructure.

The Dynamic Inventory Alert module, which is the most novel contribution of this work, demonstrates that a data pipeline can do more than just report what happened in the past. It can proactively identify waste risks and suggest creative, commercially viable actions to address them, all without any manual intervention. This is a meaningful step toward truly intelligent restaurant management systems that learn from operational data and continuously improve the efficiency and sustainability of food service businesses.

Future work will focus on three directions. First, integrating real-time streaming data using Amazon Kinesis to shift from batch processing to a near-real-time demand monitoring system. Second, incorporating customer-level preference modelling to personalize both the chatbot responses and the daily specials suggestions based on the ordering history of regular customers. Third, extending the system to support multi-restaurant chain management, where aggregated insights from all outlets can be used to optimize bulk purchasing and centralized inventory decisions. The authors also plan to evaluate the system in a live deployment with a partner restaurant group to validate the demand prediction accuracy and the commercial impact of the daily specials alert mechanism in a real-world setting.

ISSN: 2582-7219 | www.ijmrset.com | Impact Factor: 8.206| ESTD Year: 2018|

International Journal of Multidisciplinary Research in Science, Engineering and Technology (IJMRSET)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)

REFERENCES

[1] Zhang, L., Wang, H., and Chen, Y., "Machine Learning for Restaurant Demand Forecasting: A Multi-Feature Approach," IEEE Transactions on Neural Networks and Learning Systems, 2022.
[2] Mao, J. and Shi, R., "Serverless Cloud Computing Architectures for Scalable Real-Time Retail Analytics," ACM Symposium on Cloud Computing, 2021.
[3] Patel, A. and Gupta, S., "Big Data Processing Pipelines Using Apache Spark for Food Supply Chain Analytics," International Journal of Data Engineering, 2022.
[4] Brown, T. and Lim, K., "Large Language Models as Conversational Business Intelligence Agents in Hospitality," International Conference on Artificial Intelligence in Hospitality, 2023.
[5] Kumar, R. and Nair, V., "Dynamic Pricing and Specials Recommendation for Quick Service Restaurants Using Reinforcement Learning," IEEE Access, 2023.
[6] Amazon Web Services, "AWS Lambda Developer Guide," https://docs.aws.amazon.com/lambda/
[7] Apache Software Foundation, "Apache Spark Documentation," https://spark.apache.org/docs/latest/
[8] Chen, Q., Sun, W., and Liu, T., "Perishable Inventory Optimization in Restaurant Chains: A Data-Driven Approach," Journal of Operations Management, 2021.
[9] Zhao, M. and Park, H., "Prompt Engineering Techniques for Domain-Specific Large Language Model Applications," NeurIPS Workshop on Foundation Models, 2023.
[10] FAO, "Food Loss and Waste in the Context of Sustainable Food Systems," Food and Agriculture Organization of the United Nations, 2019, http://www.fao.org/food-loss-and-food-waste/

INTERNATIONAL STANDARD SERIAL NUMBER INDIA

INNO SPACE SJIF Scientific Journal Impact Factor

NISCAIR

INTERNATIONAL JOURNAL OF MULTIDISCIPLINARY RESEARCH IN SCIENCE, ENGINEERING AND TECHNOLOGY