# SYNTHETIC TIME SERIES DATA IN RESTAURANTS SUPPLY CHAIN PLANNING

**Simeon Monov, Zlatomila Mincheva, Nikolay Pavlov**

***Abstract.*** *Almost 50% of over 1 billion tonnes of food is wasted annually worldwide. The restaurant industry, with 290 million tonnes, is a major contributor to food wastage. Promising machine learning (ML) models have been developed to optimize raw material usage in restaurants, but obtaining data for training these models is a difficult task because the data is considered sensitive for intellectual property and financial reasons. Recently, large language models (LLMs) have shown promising results in generating synthetic data. In this paper, we use an LLM to generate synthetic data for restaurant dish sales and product supply and investigate how this data can be used to train different ML models. We conduct a comprehensive study and analysis of the performance of these models with our data.*

**Key words:** Synthetic Data, Time Series, LLMs, Restaurants, Supply Chain.

## Introduction

The growing problem of food waste requires effective solutions, especially in the restaurant sector, where significant losses are observed annually. Optimizing the supply chain and operational processes in restaurants can significantly help to address this issue. Machine learning has the potential to greatly improve these processes, but the challenge lies in collecting quality data for training the models. To overcome these challenges, we use large language models (LLMs), such as GPT-4 and Llama, which can generate synthetic data that effectively mimic real scenarios. This study examines the use of LLMs to create synthetic datasets that can help ML models optimize restaurant operations and reduce food waste.

## Challenges in Collecting Real Data

One of the main challenges in improving the efficiency of restaurants is the lack of quality data. Real data on restaurant operations are often difficult to collect due to reasons related to confidentiality and intellectual property, as well as the fragmented and context-specific nature of the information. Additionally, data reflecting specific aspects of the restaurant industry, such as seasonal and weather-related sales patterns, are rarely available in sufficient volume for training effective ML models. This is confirmed by the research of Grigoras and Leon, who highlight the need for synthetic time series to improve decision-making processes [1].

## Research Objectives

The primary objective of this study is to generate synthetic data that can support improving the efficiency of restaurant supply chain operations. The generated synthetic data should adequately reflect the complex nature of restaurant operations, including menu variations, customer demand fluctuations, and environmental influences. The data generated by LLM should be suitable for training different ML models capable of making accurate predictions and optimizing restaurant operations, thereby reducing food waste and improving efficiency.

## Experiment Setup

The experiment setup involved three main stages: synthetic data generation, ML model training, and performance evaluation. The synthetic data were generated using the GPT-4 and Llama language models through prompt engineering, covering restaurant characteristics, weather forecasts, and sales data. The main prompts were formulated as follows:

1. Restaurant: Included key parameters such as name, type, size, location, and menu. The generated data contained information about orders and the culinary profile of the restaurants.

2. Weather Forecasts: Generated realistic weather data for specific locations, including temperature, humidity, and general atmospheric conditions, aiming to simulate the impact of weather on restaurant operations.

3. Restaurant Sales Data: Included daily orders, accounting for the influence of time of day, weather conditions, and restaurant size on sales volume.

The trained ML models (LSTM and CNN) used the generated data for forecasting and optimizing operations in the restaurant industry. The LSTM model demonstrated high effectiveness in handling complex dependencies, while the CNN model proved more suitable for short-term forecasts requiring high speed and less complexity.

## Results

The conducted tests, using synthetic data generated by large language models (LLM), demonstrated significant advantages of the ML models LSTM and CNN in their predictive capabilities. The tests with the LSTM model showed high accuracy in long-term forecasting, successfully capturing seasonal and temporal dependencies in restaurant sales data. Meanwhile, the CNN model demonstrated excellent results in short-term forecasting, requiring quick processing and analysis of large data volumes. These tests proved that synthetic data could support ML models in effectively forecasting and optimizing resources in the restaurant industry, while also reducing food waste. The combination of both approaches can lead to even better results in planning and managing restaurant supply chains [6].

Additionally, the generated synthetic weather data was compared with historical data to evaluate its accuracy. The analysis showed that the synthetic data successfully reflected real historical trends, capturing important temporal patterns and seasonal variations. We measured a difference of 1.4 degrees Celsius in average mean annual temperature between historic data and synthetically generated data for the same location and time. This shows that synthetic data can resemble real data and be used for training ML models, especially when access to real data is difficult.

- The LSTM (Long Short-Term Memory) model was used for multivariate forecasting, with multiple input variables such as temperature, day of the week, meal popularity, and weather conditions. This model demonstrated significant advantages in time series prediction, particularly when complex dependencies exist between input variables [1]. The results showed that the LSTM model successfully captured seasonal and temporal dependencies, such as increased orders during weekends and favorable weather conditions [2]. This is extremely important for optimizing resources and managing restaurant supply chains, as it allows better planning of inventory and staffing.

- The LSTM model demonstrated accuracy in forecasting, particularly when it comes to variables that influence sales, such as weather and day of the week. Its ability to handle long time series and retain information from previous events enables the generation of forecasts that are adaptive and reflect the dynamics of the restaurant industry [3].

- The CNN (Convolutional Neural Network) model was used for multivariate forecasting and showed good results in predicting key sales trends. The CNN model is particularly effective when working with structured data and has advantages in processing large volumes of data in a short time. This makes it suitable for short-term forecasts where processing speed is a critical factor [4]. For example, the CNN model can successfully predict fluctuations in sales at different times of the day, taking into account factors such as time of day and type of meals [5].

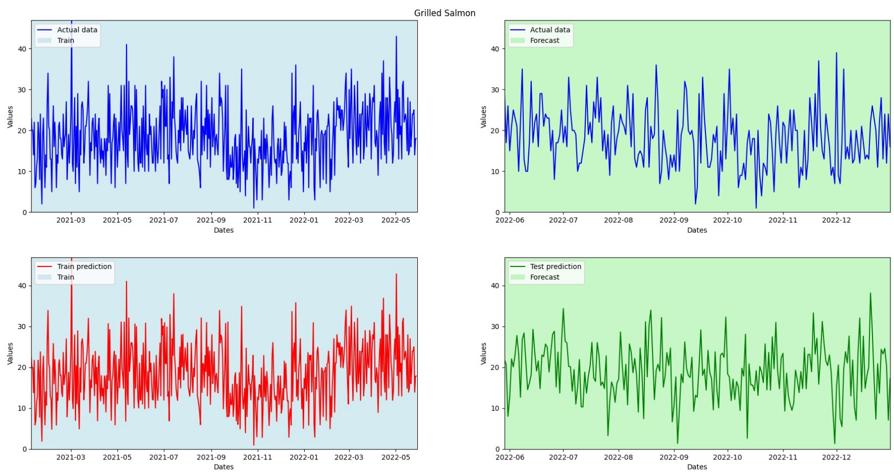


*Figure 1. Results from LSTM model forecasting 'Grilled Salmon' sales*

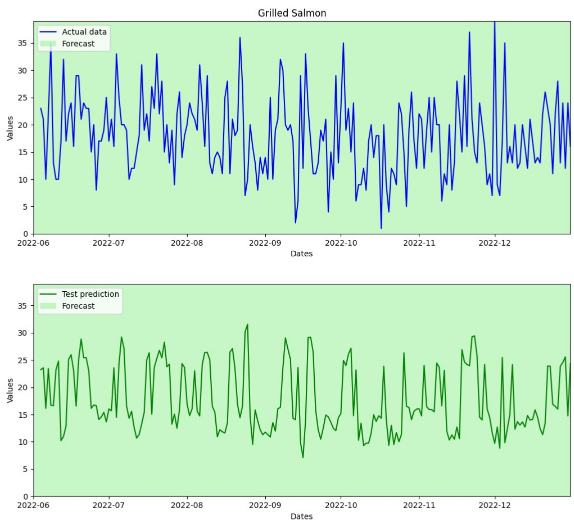


*Figure 2. Results from CNN model forecasting 'Grilled Salmon' sales*

## Analysis of Synthetic Data

The synthetic data generated using LLM successfully simulate real scenarios in the restaurant industry, including a variety of factors such as weather, types of meals, popularity of different menus, and customer preferences. This variety allows ML models to be trained on scenarios that are often missing in real data due to limitations in volume or confidentiality [2].

The analysis results showed that using synthetic data can significantly reduce the need for real data, which are often difficult to access and expensive to collect. Synthetic data also offer the ability to control the characteristics of the generated scenarios, making them particularly useful for training ML models that need to adapt to different conditions and variables [3].

## Advantages and Limitations

One of the main advantages of using synthetic data is that they allow ML models to be trained without the need for real data, addressing issues related to confidentiality and data access [4]. Additionally, synthetic data can be generated in large volumes with a variety of scenarios, which helps improve model generalization [5].

However, there are also some limitations to using synthetic data. For example, synthetic data may not fully reflect all the details and nuances of real restaurant operations, which may lead to gaps in model training. Moreover, the quality of synthetic data depends on the quality of the LLM used and its ability to generate realistic scenarios [6].

## Conclusion

The results of the study indicate that synthetic data generated using LLM can be effectively used for training ML models that can help optimize restaurant operations. LSTM and CNN models trained with these data demonstrate high accuracy in sales forecasting and resource management, with each model having its advantages depending on the specific scenario. Combining both approaches can lead to even better results, which is a promising direction for future research.

Future research could focus on combining different ML architectures and evaluating their effectiveness under real-world conditions, using both synthetic and real data.

## Acknowledgments

This paper is partially supported by project MUPD23-FMI-009 of the Scientific Fund of the Paisii Hilendarski University of Plovdiv, Bulgaria.

## References

[1] S. Jarlöv, A. Svensson Dahl, *Restaurant Daily Revenue Prediction: Utilizing Synthetic Time Series Data for Improved Model Performance*, Uppsala University, Department of Information Technology, 2023.

[2] Z. Li, H. Zhu, Z. Lu, M. Yin, Synthetic Data Generation with Large Language Models for Text Classification: Potential and Limitations, *arXiv preprint*, arXiv:2310.07849, 2023, DOI: `10.48550/arXiv.2310.07849`.

[3] R. Ghanbari, K. Borna, Multivariate Time-Series Prediction Using LSTM Neural Networks, *IEEE*, 2021, DOI: `10.1109/CSICC52343.2021.9420543`.

[4] R. Ali, Y. Li, S. Shekhar, S. Athavale, E. Marsman, Supply and Demand Aware Synthetic Data Generation for On-demand Traffic with Real-world Characteristics, *IWCTS'17: Proc. of the 10th ACM SIGSPATIAL Workshop on Computational Transportation Science*, 2017, 36–41, DOI: `https://doi.org/10.1145/3151547.3151554`.

[5] K. Wang, K. Li, L. Zhou, Y. Hu, Z. Cheng, J. Liu, C. Chen, Multiple Convolutional Neural Networks for Multivariate Time Series Prediction, *Neurocomputing*, Vol. 360, 2019, 107–119, DOI: `https://doi.org/10.1016/j.neucom.2019.05.023`.

[6] K. Posch, C. Truden, P. Hungerländer, J. Pilz, A Bayesian Approach for Predicting Food and Beverage Sales in Staff Canteens and Restaurants, *International Journal of Forecasting*, Vol. 38, No. 1, 2022, 321–338, DOI: `https://doi.org/10.1016/j.ijforecast.2021.06.001`.

Simeon Monov[1], Zlatomila Mincheva[2], Nikolay Pavlov[3],
[1,2,3] Paisii Hilendarski University of Plovdiv,
Faculty of Mathematics and Informatics,
236 Bulgaria Blvd., 4027 Plovdiv, Bulgaria
Corresponding author: `smonov@uni-plovdiv.bg`