# PREDICTION OF ON-BOARD CONSUMPTION: GENERATING SYNTHETIC DATA FOR MACHINE LEARNING TO OPTIMIZE CATERING WEIGHT

H. Kumawat, F. Giertzsch, R. God
Institute of Aircraft Cabin Systems, Hamburg University of Technology, Hein-Saß-Weg 22, 21129 Hamburg, Germany

**Abstract**

There is a great potential for reducing $CO_2$ emissions resulting from passenger aircraft cabin. In addition to further lowering the power consumption and reducing the mass of large monuments and seats, it is above all the amount of catering that can improve the $CO_2$ balance. The environmental impact of overloading, as well as the disposal and incineration of catering waste makes it necessary to organize the catering process more effectively. In our today's digital world, there is still no information about the quantities consumed and the subsequent disposal of catering on board airplanes. Real world catering on the consumption are practically non-existent and represent a major challenge for the development of machine learning (ML) -based predictive models for the actual amount of meals and beverages required. This research paper focuses on overcoming the challenges posed by unavailability of in-flight catering data and enabling on-board consumption prediction. To this end, synthetic datasets were generated based on information such as statistics on consumer age, food choices, nationality, flight details, and other catering-relevant parameters. To generate the datasets closer to realism, two regions have been chosen to differentiate passenger meals and beverage preferences: A flight from Europe to east (India) and a flight to the west (USA). Utilizing the synthetic datasets, ML models were developed to predict in-flight consumption per flight, facilitating optimized catering orders and minimizing aircraft catering overload and waste, thereby reducing $CO_2$ emissions per flight and contributing positively to environmental sustainability. The models are trained and validated using historical synthetic data from various flights journeys to ensure reliability and robustness. Based on the generated synthetic datasets, this study shows that a good prediction of catering demand that avoids overstocking of meals and beverages can be achieved. It is expected, that future access to authentic catering data on a larger scale can be used for improving and validating the synthetic data generation concept and will enhance the authenticity of synthetic datasets, resulting in precise predictions.

**Keywords**

In-flight catering, weight reduction, synthetic catering data, machine learning, predictive models

## 1. INTRODUCTION

In recent years, the growing global focus on environment protection, particularly through the reduction or avoidance of $CO_2$ emissions, has become an increasingly important concern in all industries striving to implement innovative sustainable strategies for low carbon emission. The aviation sector's recognition of the environmental impact associated with in-flight meals overloading and waste has directed the way for significant improvements in operational efficiency. In fact, the Food Waste Index Report 2021 shows that around 931 million tons of food, or 17% of all food sold to consumers in 2019, went into the waste bins of households, retailers, restaurants, and other food services galleys [1]. Overall, around 30% of all food production is wasted systemically. In the aviation industry, reducing in-flight food waste is equally compelling. According to the International Air Transport Association (IATA) Cabin Waste Handbook, 5.7 million tons of cabin waste were generated worldwide in 2017, with uneaten meals and beverages accounting about 20% of this. Recent studies have highlighted the benefits of reducing in-flight food waste within aviation industry, emphasizing its role in increasing fuel efficiency and reducing $CO_2$ emissions [2]. This goes hand in hand with the fact that minimizing food waste not only reduces the environmental impact associated with waste disposal, but also reduces the weight, volume, and cost of catering, thereby optimizing fuel consumption. For instance, as stated in [3] environmental, social and corporate governance (ESG) reporting for the full year of 2022, Southwest Airlines reports saving 33 million gallons of fuel through on-board weight reduction.

A study identified that on-board sorting and collection programs could achieve a recycling rate of 45-58 % of the total galley and cabin waste generated by the in-flight services [4]. In addition, a study from the IATA emphasizes the potential of pre-selection of In-flight meals and real-time inventory management in minimizing meals and beverage overload [5]. This research revealed the importance of identifying in-flight catering waste as a key lever of achieving significant airline cost saving and environmental benefits. The integration of artificial intelligence (AI) can be a promising solution for addressing the challenges associated with overloading of in-flight catering and $CO_2$ emission. Despite various research efforts, notable improvement in this domain have yet to be achieved. The limited availability of open-source airline catering data slows down the effective



doi: 10.25967/650481

implementation and optimization of AI-driven solutions.

To address the challenge of data unavailability in airline catering, in this paper, a methodology to generate synthetic catering data has been developed. To generate synthetic data closer to reality, two regions have been selected for variability in passengers' meals and beverages choices. The selected flight routes are from Frankfurt (FRA) to New Delhi (DEL) as well as to New York (JFK), representing long-haul flight routes starting in Europe going to the east and west. The synthetic datasets were generated using, flight routes, passenger-specific preferences and consumption patterns, which were influenced by factors such as passenger nationality and demographic information. Using this synthetic dataset, a machine learning (ML) model has been trained to predict meal and beverage consumption quantities for future flights. This predictive model enables airline caterers to better forecast the quantities of meals and beverages required for specific flights, thereby facilitating reduction in food waste, more cost savings, and environmental sustainability.

## 2. SYNTHETIC DATA GENERATION FOR PREDICTIVE MODELING

Accurately predicting the required quantities of meals and beverages before a flight journey is essential for optimizing aviation catering inventory management and minimizing waste. In aviation, airline companies and airline catering companies work closely together to provide in-flight meals and beverages to the passengers during their flight journeys. Generally, the airline provides the forecast of number of passengers for a particular flight and informs the catering company about the required food and beverages quantities in advance. The catering company takes series of process steps to fulfil the requirements of airline and load meal tray set, beverages and non-food items in catering container. These containers are brought to the aircraft on the day of delivery and the empty catering trolleys and catering waste are unloaded. These trolleys and waste are transported back to catering facility for cleaning, and the catering waste is fed into the waste management process. This provision is done manually, with the consideration that no passenger is left without a meal during the journey in order to guarantee a better passenger experience. Sometimes, that leads to overloading of meals and beverages, as well as an increase in catering waste.

In order to improve catering management, passenger services, and waste management, a machine learning (ML) -based approach for predictive modeling was chosen for this study. In the context of ML, the training of a model requires the utilization of a training dataset specifically designated for training purposes and different testing data to validate the predictions. The unavailability of a real-world in-flight catering dataset has pushed to build a model which can generate the synthetic flight catering data. This synthetic data can be used to investigate and to evaluate the functionality and performance of predictive ML models. In the future, once real-world flight catering data becomes accessible, existing ML models can be used and adapted for even more accurate catering prediction only with minor modifications, depending on the number of features and structure of dataset.

In this paper, the synthetic dataset was generated to simulate passenger and in-flight catering consumption. The dataset consists of passenger-level features such as gender, age, flight details, nationality, meal and beverage choices, and aggregated journey-level features including total passenger count, gender and age group distribution, nationality profile, class distribution, and total consumption of meals and beverages. Dataset columns like passenger age, meals and beverage quantities, were generated using parametric statistical models, including truncated normal and uniform distributions, while categorical variables, such as gender, class, nationality, and meal choice, were simulated using probability distributions. Timestamp was also included, that shows the flight departure and arrival date and time. The synthetic catering dataset was generated using Python libraries, including NumPy, Faker and Pandas. Python libraries like Scikit-learn and Plotly were implemented to visualize data, ML model training and evaluation for catering prediction.

The model generates the in-flight meals and beverages consumption dataset for daily long-haul flight over a five-year period from 2015 to 2019. For Europe to the east (India), the simulated flight from Frankfurt (FRA) to New Delhi (DEL) has been selected, as there is some educational guess for the passenger choice of meals and beverages available based on location. The flight was operated by Vistara airline and the aircraft type was Boeing 787-9 Dreamliner with 30 Business Class, 21 Premium Economy and 248 Economy Class seats available [6]. For Europe to the west (USA), the simulated flight from Frankfurt (FRA) to New York City (JFK) has been selected. The flight is operated by Lufthansa airline and the aircraft type is a Boeing 747-400 which can accommodate 67 passengers in Business Class, 32 in Premium Economy, and 272 in Economy Class [7]. These two aircraft are very similar in terms of travel times and traveling passenger counts. To generate dataset for two types of flight routes (flight from Europe to the east and flight from Europe to the west), several steps have been followed. The process begins with creating data specific to passengers traveling in economy class. After that, route-specific data generated using passenger specific data for both directions, where each row represents one individual flight journey.

These two flight journeys are long-haul flights and take around 9 hours to reach their destinations. So, it was assumed that the number of in-flight meal-serving times are similar. To generate synthetic dataset for two different flight routes, several steps have been followed in Python programming as mentioned below.

1. Import relevant libraries in Python required for generating synthetic flight catering dataset for example: NumPy, Pandas, Faker.
2. Assign passenger nationalities using probabilistic model based on the travel destination like (FRA to DEL; IND: 70%, DE: 20%, Other: 10%) and (FRA to JFK; USA: 55, DE: 35%, Other: 10%).
3. Selection of time span from 01.01.2015 to 31.12.2019 (i.e. before the COVID pandemics).
4. Define a function to generate passenger profiles with following characteristics:
   - Gender ratio: 65 % male, 35 % female,

- Name: Dummy names based on gender and nationality,
- Age: age sampled with truncated normal distribution with mean(μ) of 45 and standard deviation(σ) of 10.

5. Define categories of meals (Meal_Veg and Meal_ NonVeg), beverages (Beer, whisky, etc.).
6. Define four passenger age groups like: A (0-17 years), B (18-35 years), C (36-60 years) and D (above 60 years).
7. Provide different mass/weight for assigning number of consumed meals and beverages based of passenger age group.
8. Generate Pandas data frame with all the variable shown in Table 1.
9. Generate and save passenger specific synthetic dataset (for details cf. section 2.1.).
10. Generate and save flight journey specific dataset after applying grouping with sum() function (for details cf. section 2.2.).

## 2.1. Passenger specific data generation

In the first step of generating passenger-specific datasets for both flights, the impact of the various biases is considered. The dataset of passenger personal fictitious details, including demographics such as name, age, gender, and nationality make the dataset closer to reality. Age selection of the passenger carried out using a truncated normal distribution with the mean ($\mu$) value of 45 and standard deviation ($\sigma$) of 10. This approach ensures that age remains within a realistic and valid range via eliminating the possibility of generating negative ages. Additionally, the dataset incorporates meal and beverage choices, providing insights into passengers' dietary preferences. Flight details such as departure and arrival times, flight route, day of flying, which were helping in evaluating overall passenger behavior and consumption. Key demographic and cultural insights were also considered in the dataset generation. For example, a study [8] shows that approximately 24% of Indian are vegetarians, influencing veg meal preferences. On the other side, the research by Gallup [9] suggests that only 4% of Americans are vegetarian and 1% are vegan.

These biases help to build datasets that are both different and realistic. For the beverages, only 10% of adults in India (18.8% of men, 1.3% of women) reported drinking alcoholic drinks, and whisky is the most chosen alcoholic beverage by Indians [10]. In the USA, around 62% of adults drink alcoholic beverages, and their likely choices are beer and wine [11]. In this research, the passenger specific synthetic data generated using a Python script with utilizing relevant libraries like Faker, NumPy designed to simulate passenger choices for meals and beverages. To generate a dataset for each flight route only one run was required. For the FRA to DEL passenger-specific dataset, 413,574 data samples with a size of 44,633 Kilobyte (KB) were generated. The FRA to JFK dataset consists of 435,274 data samples with the total size of 44,633 KB. The dataset generated in this study consists of 16 variables representing passenger demographics and nationalities. The shapes of dataset for both routes has changed because of variation in number of travelling passenger counts. This comprehensive dataset serves as a valuable resource for research for developing passenger focused cabin services. The key variables are shown in Table 1.

**TAB 1. Variables for passenger specific synthetic dataset.**

| Variable name | Description |
|---|---|
| Gender | Gender of the passenger (dummy) |
| Given_Name | Full name of passenger (dummy) |
| Age | PAX age (dummy) |
| Class | PAX travelling class (Y-class) |
| Dep_Timestamp | Flight departure time from origin |
| Arr_Timestamp | Flight arrival time at destination |
| Year | Year of journey |
| Month | Month of journey |
| Day_of_Week | Day of journey week |
| Flight_Number | Flight number |
| Origin | Origin of flight journey |
| Destination | Destination of flight journey |
| Nationality | Nationality of passenger |
| Meal | Meal choice of PAX (Veg / NonVeg) |
| Cold_Beverage | Cold beverage choice of passenger |
| Meal_Description | Information about consumed meal |

## 2.2. Flight specific data generation

The flight-specific dataset is the second step of generating synthetic data, which is used for developing a predictive model for airline catering. This dataset is extension of the passenger specific dataset. The flight-specific dataset for the flight FRA to DEL and the flight from FRA to JFK has 1,825 rows. Both datasets include 20 columns, capturing detailed information about consumed quantities of meals and beverages in flight journeys. The two datasets have same number of samples, as the number of flight journeys was identical. The size of each dataset has been reduced to 156 KB.

The flight specific datasets include variables related to the number of passengers per flight and their consumption quantities of meals and beverages. This dataset facilitates analyzing meals and beverages choices based on passenger age and nationality. By examining the relationships between passenger demographics, passenger counts, and consumption quantities, valuable insights can be derived to enhance service offerings and operational efficiency in the aviation industry. The datasets were utilized for ML model training for both the flight route for each item available in catering. The variables that are used in creating catering consumption predictive ML models are shown in Table 2.

**TAB 2. Variables for flight specific synthetic dataset.**

| Variable Name | Description |
|---|---|
| Date | Date of flight journey |
| No_Pax | No. of passengers |
| Males | No. of male PAX |
| Females | No. of female PAX |
| Nationality_Origin | No. of PAX with nationality of origin (DE) |
| Nationality_Dest | No. of PAX with nationality of destination (IND / USA) |
| Nationality_Other | No. of PAX with other nationality |
| Eco_Pax | No. of PAX with Y-class |

| | |
|---|---|
| Age_Below_18 | No. of PAX with age below 18 years |
| Age_Between_18to35 | No. of PAX with age between 18 to 35 years |
| Age_Between_36to60 | No. of PAX with age between 36 to 60 years |
| Age_Above_60 | No. of PAX with age above 60 years |
| Meal_NonVeg | No. of non-vegetarian meals consumed during journey |
| Meal_Veg | No. of vegetarian meals consumed during journey |
| Juice | No. of juice bottles consumed during journey (1 liter) |
| Beer | No. of beer canes consumed during journey (330 ml) |
| Wine | No. of wine bottles consumed during journey (750 ml) |
| Whisky | No. of whisky bottles consumed during journey (750 ml) |
| Softdrinks | No. of soft drink cans consumed during journey (330 ml) |
| Water | No. of water bottles consumed during journey (100 ml) |

## 3. SYNTHETIC DATA: VARIATION, SCALIBILITY AND DISTRIBUTION IN IN-FLIGHT CATERING PREFERENCES

Two flight specific in-flight catering synthetic datasets were generated, incorporating various factors such as aircraft type, flight routes, passenger preferences for meals and beverages, and geographic locations. A dataset includes the number of seats booked across economy class, with biases intentionally introduced based on available information from public sources and informed assumptions. The datasets generated represent daily long-haul flight data over a five-year period from 01.01.2015 to 31.12.2019. The timespan selected till end of 2019 because of the COVID pandemics has changed the flight journey trends for more than 2 years. In this research five years of dataset were used because it includes the information about passenger count trend and catering consumption which was suitable to investigate ML predictive models.

The sample of generated datasets are shown in the Appendix A of this paper for both the flight routes. The Table A1 shows the sample of passenger specific dataset for the flight from FRA to DEL. In this table all columns with passenger details, flight details, and also the selection of meals and beverages per passenger is contained. As mentioned above, the passenger-specific dataset was modified to generate flight-specific data, in which each row in the table represents a single flight journey. The sample of flight specific dataset can be seen in Table A.3. The same method was used to create the dataset for flights from FRA to JFK. The passenger-specific data is shown in Table A.2, while the flight-related data is shown in Figure A.4. These datasets help in understanding passenger patterns and flight characteristics on international routes.

The datasets were designed to closely reflect real-world passenger behavior and demographic patterns. Bias selection and variable design were informed by publicly available data sources and refined through internal discussions within the institution. For instance, Passengers under the age of 18 are unlikely to consume alcoholic beverages, and those over 65 are tend to avoid alcohol. Additionally, female passengers likely to prefer juice or wine over stronger alcoholic beverages. In contrast, American nationals are more inclined to opt for non-vegetarian meals. Beverage quantities were determined based on serving sizes per bottle. These parameters, along with others, were carefully incorporated to create a comprehensive and realistic synthetic dataset for this research.

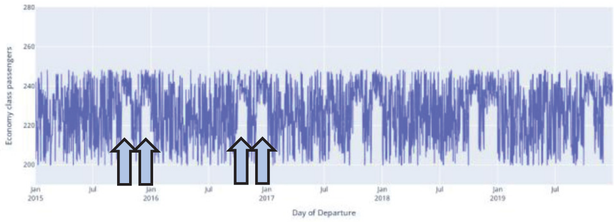

**FIG 1. Economy class PAX counts for flight FRA-DEL.**

The generated synthetic datasets also have variability in the number of travelling passengers. For example, Figure 1 shows, there is a significant increase in passenger traffic on routes such as Frankfurt to New Delhi in October and December, due to the festival seasons, reflecting seasonality in the data. On the other hand (cf. Figure 2), the passengers travelling to the USA are more in March and December because of Easter and Christmas vacations.

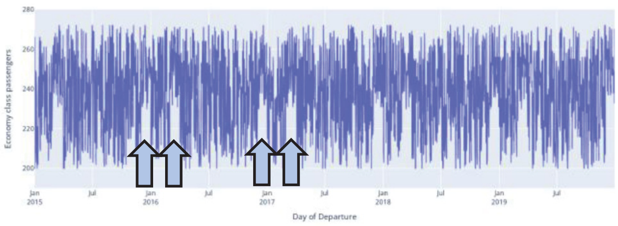

**FIG 2. Economy class PAX counts for flight FRA-JFK.**

## 4. INVESTIGATION OF MACHINE LEARNING PREDICTIVE MODELS WITH GENERATED SYNTHETIC DATA

In this study, the generated synthetic dataset which representing in-flight catering consumption was used to train and test ML models, in order to evaluate their predictive performance. Machine learning (ML) is a set of methods that enable computers to make predictions or to improve process efficiency based on data [12]. For example, to predict stock price, the ML model learns the patterns from historical stock performance.

In order to obtain a predictive catering loading list for future flights for airline and catering companies, a synthetic dataset must be created. The flight specific dataset was used to investigate various machine learning models to obtain predictions about the quantities of meals and beverages.

**Linear Regression (LR)**: A linear regression model predicts the target as a weighted sum of the feature inputs [12]. The linear regression model can be used to get prediction y based on some input features x. It follows the mathematical equation for a single instance as given below:

$$y = \beta_0 + \beta_1 x_1 + \cdots + \beta_p x_p + \epsilon \tag{1}$$

The prediction of an instance is a weighted sum of its $p$ features. The $\beta$ represents the coefficients. The $\epsilon$ is the difference between the prediction and the actual outcome.

**K-Nearest Neighbor Regression (k-NN):** k-NN regression is a non-parametric method that, in an intuitive manner, approximates the association between independent variables and the continuous outcome by averaging the observations in the same neighborhood [13]. The selection of closest point is based on calculated Euclidean distance.

$$D(p, a) = \sqrt{(p_1 - a_1)^2 + (p_2 - a_2)^2 + \cdots + (p_n - a_n)^2} \tag{2}$$

p is the new predicted point and a is the actual point. The selection of k determines the number of neighbors of observation.

**XGB Regressor**: Extreme Gradient Boosting (XGBoost) regressor is an implementation of the gradient boosting machine learning algorithm which is a type of ensemble learning method that combines the prediction for multiple weaker models to create a stronger, more accurate model [14]. It works by combining of decision trees, where each tree make prediction based on a subset of provided data.

To build the machine learning model, the flight specific dataset has been divided in 1612 training data sample and 213 testing data sample. The above mentioned three ML model have been trained on the selected input features i.e. Date, No Pax, Males, Females, Nationality_Origin, Nationality_Dest, Nationality_Other, Eco_Pax, Age_Below_18, Age_Between_18to35, Age_Between_36to60 and Age_Above_60. Each meal item and beverages consumption have different impact of the ML input features. Therefore, the ML model was trained for each target item for example: Meal_Veg, Meal_NonVeg, etc. The ML models are trained for regression not for classification so to evaluate the performance of the model, Root Mean Squared Error (RMSE) has been used. RMSE is the square root of the mean of the square of all of the error [15].

$$\mathrm{RMSE} = \sqrt{\frac{1}{n}\sum_{i=0}^{n}(S_i - O_i)^2} \tag{3}$$

where $O_i$ are the observations, $S_i$ predicted values of a variable, and $n$ the number of observations available for analysis. RMSE was used to evaluate predictive ML models because it captures the magnitude of prediction error and penalize large deviation that was needed in meals and beverages prediction. RMSE squares the errors, so the large error grows even more in the final score. This was helpful for predicting consumed meals and beverage quantities. For example: when the prediction of consumed meals quantity was 150 and actual consumption was 100, then the error is big. Use of RMSE helps to reduce the overloading of the catering waste

## 5. MACHINE LEARNING MODEL EVALUATION

This work aimed to investigate a predictive catering ML model to forecast the required quantities of meals and beverages for the future flight journeys. To investigate the performance of ML algorithms, initially the target feature for the predictive model has been selected as consumed water bottle per flight journey. That means the model will be designed to predict the required number of water bottle for the subsequent 30 days. Three regression ML models were used: LR, k-NN, and XGB Regressor. The LR was trained using the default parameters from Scikit-learn, while the k-NN model was configured with n_neighbors = 2 to detect closer relationships in the data. For XGB Regressor, parameters included n_estimators = 100, max_depth= 3, and early_stopping_rounds = 50 to reduce model complexity and prevent model from overfitting. The Figure 3 indicates that, all three ML model predictions for flight FRA to DEL and Figure 4 is for flight from FRA to JFK.

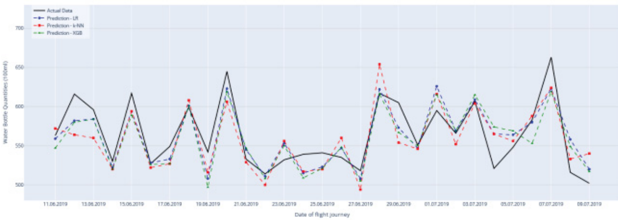

**FIG 3. Comparison of ML model prediction with actual water bottle quantities (FRA-DEL).**

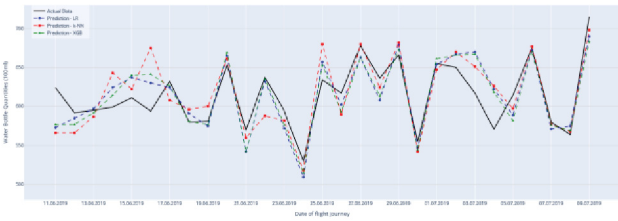

**FIG 4. Comparison of ML model predictions with actual water bottle quantities (FRA-JFK).**

The comparison of model performance had been done via employing RMSE value. The LR (RMSE 8.29) and XGB regressor (RMSE 8.49) demonstrate similar result, whereas with k-NN the RMSE (9.65) is slightly worse. The result indicates that the predicted quantities closely align with the actual quantities of water bottles. The RMSE value indicate the average deviation between predicted water bottle quantities and actual quantities. LR has lower RMSE value such as 8.29, represents the difference is about 8 bottle which was lesser than RMSE value 9.65 with the round of 9 bottles with KNN model.

Linear Regression is a well-established machine learning algorithm that can be applied on small-scale datasets and it require relatively less computational power to run. This makes it suitable for scenarios with limited computational resources. XGBoost regressor is gradient boosting algorithm, that requires GPU to run and it is better to use when there is high volume of in-flight catering data available. In this study, linear regression has been selected as the primary model for acquiring the prediction due to its less complexity and resource requirement. In addition, XGBoost regressor will be used in situations where large-scale real world in-flight catering data is available as well as high computational power for handling complex dataset.

## 6. RESULT AND DISCUSSION

To evaluate inventory efficiency and forecasting accuracy, actual, predicted, and loaded quantities for meals and beverages were converted into weight-based metrics. Each item's unit weight was standardized and converted to grams for consistency. In this study, all the meals and beverages items were assigned to their respective mass. The loaded quantities were considered with the number of passengers travelling. The item names, their masses and the loaded quantities for each catering item for 30 days were presented in Table 3. The quantities were based on the number of travelling passenger and their food choices.

**TAB 3. Inflight catering items, mass and quantities.**

| Catering item | Mass per item (gm) | Loaded Quantity (FRA-DEL) | Loaded Quantity (FRA-JFK) |
|---|---|---|---|
| Meal_NonVeg | 450 | 100 | 180 |
| Meal_Veg | 400 | 100 | 50 |
| Juice | 1030 | 100 | 100 |
| Beer | 350 | 100 | 150 |
| Wine | 1600 | 6 | 8 |
| Whisky | 1200 | 6 | 8 |
| Softdrinks | 358 | 100 | 150 |
| Water | 100 | 650 | 750 |

The study shows how much catering mass can be reduced with the catering predictive LR model. In Figure 5, the actual loaded mass or water bottles, actual consumption and LR prediction have been shown for the fight from FRA to DEL. The plot below shows the around 10 to 15 kg mass can be reduced if water bottles are loaded in aircraft considering ML model predictions. The water bottle mass for flight journey from FRA to JFK can be reduced by 20 to 25 kg that is show in Figure 6.

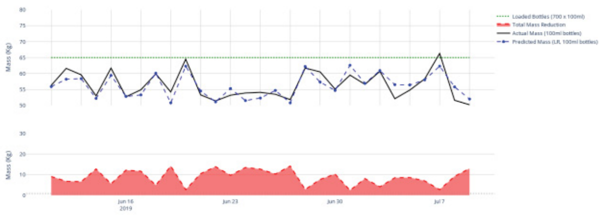

**FIG 5. Loaded water bottle mass, predicted mass and mass reduction (FRA-DEL).**

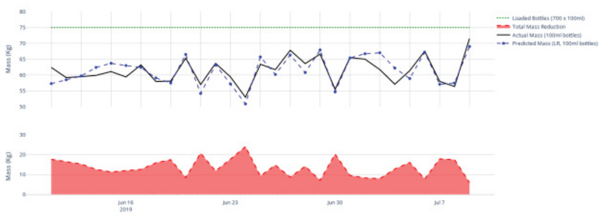

**FIG 6. Loaded water bottle mass, predicted mass and mass reduction (FRA-JFK).**

The main aim of this research was to evaluate how much total catering mass can be reduced for the flight journey that can help to not overload and also passenger get sufficient food and beverage items in their journey. The predictive ML model are built for each item that helps to generate prediction for a specific item. The quantities of meals and beverages for both the flights for 30 days are collected via executing LR model. Afterwards, the predicted quantities are multiplied with their respective mass to calculate total predictive mass for the flight. The Figure 7 shows the total loaded mass by airline, total actual consumed mass for flight and the total predicted mass from the predictive model for FRA to DEL flight. The area below shows, how much potential mass can be reduced if the predictive model is used for loading the meals and beverages. For flight FRA to DEL total mass of 70 to 100 kg can be reduced. The Figure 8 shows the overloading of 100 to 130 kg can be avoided in flight from FRA to JFK with the implementation of ML catering predictive model.

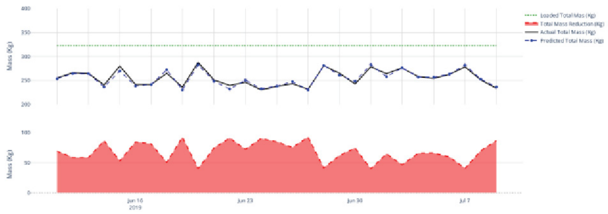

**FIG 7. Loaded total mass, predicted mass and mass reduction for flight (FRA-DEL).**

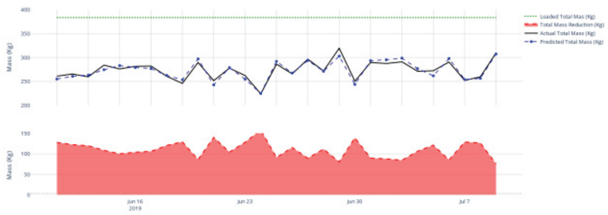

**FIG 8. Loaded total mass, predicted mass and mass reduction for flight (FRA-JFK).**

## 7. CONCLUSION AND OUTLOOK

This study investigated the potential of generated synthetic data and predictive machine learning models for in-flight catering to reduce overloading of meals and beverages that can result in decreasing $CO_2$ emissions, as well as better inventory management. Based on the generated synthetic datasets, this study shows that a good prediction of catering demand that avoids overstocking of meals and beverages. The results show that with the use of generated synthetic data, catering weight reduction of 70 to 100 kg per flight is plausible. The generation of synthetic datasets for flight from FRA to DEL and FRA to JFK is done by considering the flight routes and passenger food choices to make them close to reality. Machine learning algorithms, including LR, k-NN, and XGB regressor are built with synthetic data and selected based on collected RMSE values.

The linear regression performs the best among the three of them, and for each catering item such as Meal_Veg, Meal_NonVeg, and water a LR model has been created. These models' predictions for each item combined together to obtain the complete list of predicted catering item quantities for specific flight journey. Finally, the prediction for the loading meals and beverage quantities is acquired, and the total mass is calculated. The total mass has been compared to the actual loaded mass to see the potential of reducing total catering mass for both flight journeys. This research opens the door to introduce machine learning in airline catering and its capability to decrease the catering waste and $CO_2$ emission. As of now the unavailability of real-world catering data problem is approached by using synthetic data. Once real-world flight catering data is available, implementing and applying this approach in practice can optimize the loading of catering, thus reducing waste, fuel consumption and consequently, $CO_2$ emissions.

## ACKNOWLEDGEMENTS

This work was supported by the LuFo VI-1 project “Lernendes Galley-Catering-System” (engl.: Learning Galley Catering System, LGCS, project number 20D1928C) funded by the Federal Ministry for Economic Affairs and Climate Action based on the decision of the German Bundestag.

**Contact address:**

hiteshkumar.kumawat@tuhh.de

## REFERENCES

[1] H. Forbes, T. Quested and C. O’Connor. *UNEP Food Waste Index Report 2021*. United Nations Environment Programme, Nairobi, 2021. https://wedocs.unep.org/bitstream/handle/20.500.11822/35280/FoodWaste.pdf (accessed 2025/19/09).

[2] International Air Transport Association (IATA). IATA Annual Review, 2023. https://www.iata.org/contentassets/c81222d96c9a4e0bb4ff6ced0126f0bb/annual-review-2023.pdf (accessed 2025/19/09).

[3] Southwest Airlines Co. *2022 One Report.* https://www.southwest.com/swa-resources/pdfs/communications/one-reports/Southwest-Airlines-2022-One-Report.pdf (accessed 2025/19/09).

[4] X. D. Li, C. S. Poon, S. C. Lee, S. S. Chung and F. Luk. *Waste reduction and recycling strategies for the in-flight services in the airline industry – Resources, Conservation and Recycling*. Vol. 37, p. 87–99, 2002. https://doi.org/10.1016/S0921-3449(02)00074-5 (accessed 2025/19/09).

[5] E. Morris, M. Roberts, K. Patterson and N. Sweet. *IATA Cabin Waste Handbook*, 2019. https://www.iata.org/contentassets/821b593dd8cd4f4aa33b63ab9e35368b/iata-cabin-waste-handbook.pdf (accessed 2025/19/09).

[6] UK Boeing 787-9 Type 1. *AeroLOPA Detailed aircraft seat*. https://www.aerolopa.com/uk-789-1 (accessed 2025/19/09).

[7] Boeing 747-400, Lufthansa AG. *Seat maps Boeing B747-400.* https://www.lufthansa.com/us/en/74e (accessed 2025/19/09).

[8] N. Fitzgerald. *Veganism Statistics Indian* in *2022 - How many vegans are there in India?* Truly Experiences Blog, 27 Oct. 2023. https://trulyexperiencesblog.com/veganism-statistics-india/ (accessed 2025/19/09).

[9] J. M. Jones. *In U.S., 4% Identify as Vegetarian, 1% as Vegan*. Gallup.com, 24 Aug. 2023. https://news.gallup.com/poll/510038/identify-vegetarian-vegan.aspx (accessed 2025/19/09).

[10] R. Rana. *Alcohol Consumption in India 2024*. Know Noida City, 4 July 2024. https://www.noidabusinessguide.com/alcohol-consumption-in-india/ (accessed 2025/19/09).

[11] M. Brenan. *More Than Six in 10 Americans Drink Alcohol.* Gallup.com, 14 August 2023. https://news.gallup.com/poll/509501/six-americans-drink-alcohol.aspx (accessed 2025/19/09).

[12] C. Molnar. *Interpretable Machine Learning: A Guide for Making Black Box Models Explainable*. https://christophm.github.io/interpretable-ml-book/ (accessed 2025/19/09).

[13] A. Teixeira-Pinto. *2 K-Nearest Neighbours Regression – Machine Learning for biostatistics*, https://bookdown.org/tpinto_home/Regression-and-Classification/k-nearest-neighbours-regression.html (2022).

[14] J. D. Hedengren, *Machine Learning for Engineers-XGBoost Regressor*, https://apmonitor.com/pds/index.php/Main/XGBoostRegressor (accessed 2025/19/09).

[15] T. O. Hodson. *Root-mean-square error (RMSE) or mean absolute error (MAE): when to use them or not*. Geoscientific model development. Vol. 15, https://doi.org/10.5194/gmd-15-5481-2022 (2022).

## APPENDIX A:

### TAB A.1. PAX specific synthetic dataset (FRA-DEL)

| | Dep_Timestamp | Arr_Timestamp | Given_Name | Gender | Age | Class | Water | Year | Month | Day_of_Week | Flight_Number | Origin | Destination | Nationality | Meal | Cold_Beverage | Meal_Description |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 1 | 2015-01-01 00:00:00 | 2015-01-02 00:00:00 | Amruta | F | 45 | Economy | 3 | 2015 | 01 | 4 | UK26 | FRA | DEL | IND | Meal_Veg | Softdrinks | Vegetable Pasta |
| 2 | 2015-01-01 00:00:00 | 2015-01-02 00:00:00 | Irmingard | F | 56 | Economy | 3 | 2015 | 01 | 4 | UK26 | FRA | DEL | IND | Meal_Veg | Softdrinks | Dal Makhani |
| 3 | 2015-01-01 00:00:00 | 2015-01-02 00:00:00 | Traugott | M | 38 | Economy | 1 | 2015 | 01 | 4 | UK26 | FRA | DEL | IND | Meal_Veg | Wine | Vegetable Pasta |
| 4 | 2015-01-01 00:00:00 | 2015-01-02 00:00:00 | Helmut | M | 29 | Economy | 3 | 2015 | 01 | 4 | UK26 | FRA | DEL | IND | Meal_Veg | Juice | Dal Makhani |
| 5 | 2015-01-01 00:00:00 | 2015-01-02 00:00:00 | Susan | F | 55 | Economy | 4 | 2015 | 01 | 4 | UK26 | FRA | DEL | IND | Meal_Veg | Beer | Dal Makhani |
| ... | -- | -- | -- | -- | -- | -- | -- | -- | -- | -- | -- | -- | -- | -- | -- | -- | -- |
| 414380 | 2019-12-30 00:00:00 | 2019-12-31 00:00:00 | Trudel | F | 45 | Economy | 1 | 2019 | 12 | 1 | UK26 | FRA | DEL | IND | Meal_Veg | Beer | Vegetable Pasta |
| 414381 | 2019-12-30 00:00:00 | 2019-12-31 00:00:00 | Mahmut | M | 42 | Economy | 4 | 2019 | 12 | 1 | UK26 | FRA | DEL | IND | Meal_NonVeg | Whisky | Lamb Curry |
| 414382 | 2019-12-30 00:00:00 | 2019-12-31 00:00:00 | Gebhard | M | 51 | Economy | 6 | 2019 | 12 | 1 | UK26 | FRA | DEL | IND | Meal_NonVeg | Softdrinks | Chicken Curry |
| 414383 | 2019-12-30 00:00:00 | 2019-12-31 00:00:00 | Stefanie | F | 40 | Economy | 4 | 2019 | 12 | 1 | UK26 | FRA | DEL | IND | Meal_NonVeg | Softdrinks | Chicken Curry |
| 414384 | 2019-12-30 00:00:00 | 2019-12-31 00:00:00 | Utkarsh | M | 37 | Economy | 3 | 2019 | 12 | 1 | UK26 | FRA | DEL | Other | Meal_NonVeg | Juice | Chicken Curry |

### TAB A.2. PAX specific synthetic dataset (FRA-JFK)

| | Dep_Timestamp | Arr_Timestamp | Given_Name | Gender | Age | Class | Water | Year | Month | Day_of_Week | Flight_Number | Origin | Destination | Nationality | Meal | Cold_Beverage | Meal_Description |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 1 | 2015-01-01 00:00:00 | 2015-01-02 00:00:00 | Richard | M | 58 | Economy | 2 | 2015 | 01 | 4 | LH400 | FRA | JFK | USA | Meal_NonVeg | Beer | Beef Stew |
| 2 | 2015-01-01 00:00:00 | 2015-01-02 00:00:00 | Eric | M | 39 | Economy | 3 | 2015 | 01 | 4 | LH400 | FRA | JFK | DEU | Meal_NonVeg | Softdrinks | Beef Stew |
| 3 | 2015-01-01 00:00:00 | 2015-01-02 00:00:00 | Norbert | M | 49 | Economy | 2 | 2015 | 01 | 4 | LH400 | FRA | JFK | USA | Meal_NonVeg | Juice | Beef Stew |
| 4 | 2015-01-01 00:00:00 | 2015-01-02 00:00:00 | Timo | M | 48 | Economy | 2 | 2015 | 01 | 4 | LH400 | FRA | JFK | USA | Meal_NonVeg | Beer | Chicken Rice |
| 5 | 2015-01-01 00:00:00 | 2015-01-02 00:00:00 | Werner | M | 44 | Economy | 3 | 2015 | 01 | 4 | LH400 | FRA | JFK | DEU | Meal_NonVeg | Juice | Beef Stew |
| ... | -- | -- | -- | -- | -- | -- | -- | -- | -- | -- | -- | -- | -- | -- | -- | -- | -- |
| 436247 | 2019-12-30 00:00:00 | 2019-12-31 00:00:00 | Tracy | F | 57 | Economy | 4 | 2019 | 12 | 1 | LH400 | FRA | JFK | DEU | Meal_NonVeg | Beer | Beef Stew |
| 436248 | 2019-12-30 00:00:00 | 2019-12-31 00:00:00 | Andrew | M | 51 | Economy | 2 | 2019 | 12 | 1 | LH400 | FRA | JFK | Other | Meal_NonVeg | Juice | Chicken Rice |
| 436249 | 2019-12-30 00:00:00 | 2019-12-31 00:00:00 | Erna | F | 32 | Economy | 2 | 2019 | 12 | 1 | LH400 | FRA | JFK | DEU | Meal_NonVeg | Beer | Beef Stew |
| 436250 | 2019-12-30 00:00:00 | 2019-12-31 00:00:00 | Sarah | F | 52 | Economy | 4 | 2019 | 12 | 1 | LH400 | FRA | JFK | Other | Meal_NonVeg | Juice | Chicken Rice |
| 436251 | 2019-12-30 00:00:00 | 2019-12-31 00:00:00 | Michael | M | 44 | Economy | 0 | 2019 | 12 | 1 | LH400 | FRA | JFK | Other | Meal_Veg | Beer | vergetable Curry |

### TAB A.3. Flight specific synthetic dataset (FRA-DEL)

| | Date | No_Pax | Males | Females | Nationality_Origin | Nationality_Dest | Nationality_Other | Eco_Pax | Age_Below_18 | Age_Between_18to35 | Age_Between_36to60 | Age_Above_60 | Meal_NonVeg | Meal_Veg | Juice | Beer | Wine | Whisky | Softdrinks | Water |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 1 | 2015-01-01 | 243 | 160 | 83 | 48 | 166 | 29 | 243 | 1 | 40 | 191 | 11 | 100.0 | 143.0 | 96.0 | 38.0 | 3.0 | 2.0 | 67.0 | 600 |
| 2 | 2015-01-02 | 229 | 151 | 78 | 65 | 136 | 28 | 229 | 0 | 40 | 178 | 11 | 102.0 | 127.0 | 85.0 | 33.0 | 3.0 | 2.0 | 59.0 | 604 |
| 3 | 2015-01-03 | 226 | 146 | 80 | 57 | 141 | 28 | 226 | 2 | 57 | 158 | 9 | 102.0 | 124.0 | 73.0 | 31.0 | 3.0 | 2.0 | 73.0 | 587 |
| 4 | 2015-01-04 | 244 | 168 | 76 | 70 | 141 | 33 | 244 | 1 | 39 | 187 | 17 | 102.0 | 142.0 | 95.0 | 22.0 | 4.0 | 2.0 | 75.0 | 605 |
| 5 | 2015-01-05 | 247 | 173 | 74 | 62 | 157 | 28 | 247 | 2 | 46 | 189 | 10 | 103.0 | 144.0 | 93.0 | 32.0 | 3.0 | 2.0 | 73.0 | 621 |
| ... | -- | -- | -- | -- | -- | -- | -- | -- | -- | -- | -- | -- | -- | -- | -- | -- | -- | -- | -- | -- |
| 1821 | 2019-12-26 | 242 | 154 | 88 | 57 | 154 | 31 | 242 | 2 | 55 | 172 | 13 | 86.0 | 156.0 | 81.0 | 23.0 | 4.0 | 2.0 | 79.0 | 580 |
| 1822 | 2019-12-27 | 237 | 146 | 91 | 57 | 150 | 30 | 237 | 0 | 42 | 183 | 12 | 113.0 | 124.0 | 89.0 | 32.0 | 3.0 | 2.0 | 70.0 | 581 |
| 1823 | 2019-12-28 | 244 | 165 | 79 | 63 | 153 | 28 | 244 | 0 | 34 | 195 | 15 | 97.0 | 147.0 | 89.0 | 28.0 | 3.0 | 3.0 | 70.0 | 596 |
| 1824 | 2019-12-29 | 248 | 164 | 84 | 56 | 157 | 35 | 248 | 0 | 43 | 191 | 14 | 93.0 | 155.0 | 84.0 | 32.0 | 2.0 | 2.0 | 84.0 | 639 |
| 1825 | 2019-12-30 | 245 | 150 | 95 | 67 | 150 | 28 | 245 | 0 | 54 | 178 | 13 | 113.0 | 132.0 | 95.0 | 38.0 | 2.0 | 1.0 | 73.0 | 638 |

### TAB A.4. Flight specific synthetic dataset (FRA-JFK)

| | Date | No_Pax | Males | Females | Nationality_Origin | Nationality_Dest | Nationality_Other | Eco_Pax | Age_Below_18 | Age_Between_18to35 | Age_Between_36to60 | Age_Above_60 | Meal_NonVeg | Meal_Veg | Juice | Beer | Wine | Whisky | Softdrinks | Water |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 1 | 2015-01-01 | 220 | 147 | 73 | 70 | 125 | 25 | 220 | 1 | 37 | 169 | 13 | 195.0 | 25.0 | 47.0 | 94.0 | 2.0 | 3.0 | 18.0 | 558 |
| 2 | 2015-01-02 | 235 | 158 | 77 | 63 | 133 | 39 | 235 | 1 | 41 | 184 | 9 | 197.0 | 38.0 | 46.0 | 98.0 | 3.0 | 2.0 | 33.0 | 582 |
| 3 | 2015-01-03 | 254 | 165 | 89 | 61 | 156 | 37 | 254 | 1 | 41 | 200 | 12 | 220.0 | 34.0 | 49.0 | 89.0 | 5.0 | 4.0 | 29.0 | 593 |
| 4 | 2015-01-04 | 216 | 147 | 69 | 58 | 125 | 33 | 216 | 2 | 40 | 161 | 13 | 183.0 | 33.0 | 55.0 | 75.0 | 4.0 | 2.0 | 29.0 | 520 |
| 5 | 2015-01-05 | 218 | 137 | 81 | 55 | 127 | 36 | 218 | 2 | 40 | 168 | 8 | 197.0 | 21.0 | 38.0 | 79.0 | 4.0 | 3.0 | 35.0 | 570 |
| ... | -- | -- | -- | -- | -- | -- | -- | -- | -- | -- | -- | -- | -- | -- | -- | -- | -- | -- | -- | -- |
| 1821 | 2019-12-26 | 233 | 163 | 70 | 60 | 144 | 29 | 233 | 1 | 37 | 182 | 13 | 206.0 | 27.0 | 45.0 | 94.0 | 4.0 | 3.0 | 24.0 | 605 |
| 1822 | 2019-12-27 | 268 | 168 | 100 | 80 | 148 | 40 | 268 | 0 | 43 | 213 | 12 | 236.0 | 32.0 | 62.0 | 89.0 | 7.0 | 3.0 | 31.0 | 658 |
| 1823 | 2019-12-28 | 253 | 155 | 98 | 68 | 149 | 36 | 253 | 0 | 51 | 184 | 18 | 224.0 | 29.0 | 62.0 | 95.0 | 4.0 | 3.0 | 27.0 | 654 |
| 1824 | 2019-12-29 | 254 | 164 | 90 | 78 | 140 | 36 | 254 | 0 | 48 | 189 | 17 | 223.0 | 31.0 | 70.0 | 84.0 | 3.0 | 3.0 | 29.0 | 653 |
| 1825 | 2019-12-30 | 270 | 172 | 98 | 76 | 156 | 38 | 270 | 0 | 51 | 203 | 16 | 233.0 | 37.0 | 71.0 | 94.0 | 3.0 | 3.0 | 31.0 | 684 |