J. Aviat. Space Stud., 5 (1), 1-16, 2025

Journal of Aviation and Space Studies
Havacılık ve Uzay Çalışmaları Dergisi

E-ISSN: 2757-7317 & URL: https://dergipark.org.tr/pub/jass

Research Paper

Machine learning-based inflight food waste prediction for sustainable aviation

Duygu Aghazadeh[1]

Abstract. The study delves into the utilization of machine learning to predict and reduce inflight food waste, improving sustainability in aviation logistics. Inflight food waste, a major environmental problem, is determined by passenger choices, flight parameters, and catering services. The research presents two efficient machine learning algorithms, that are, Multiple Linear and Random Forest Regression to perform food waste prediction during the flights. The models are trained using a synthetically created dataset of 10,000 records and 15 features, which include factors such as meal type, waste weight, and passenger number. The study undertakes considerable feature engineering, including the development of new features such as "Waste per Passenger" and "Meal Efficiency" to increase forecast accuracy. A correlation analysis is also used to determine the most influential characteristics. The models' performance is assessed in a Python-based computational environment, with MLR concentrating on linear links between food waste and predictors and RFR on non-linear interactions. The results show that both models can effectively forecast inflight food waste, with RFR being more adaptable to complicated patterns. The research concludes with recommendations for airline managers to apply data-driven waste reduction techniques that correspond with overall sustainability goals in aviation logistics. The models created are a useful tool for optimizing inflight food, lowering environmental impact, and contributing to the industry's sustainability initiatives.

Keywords: Inflight food waste, multiple linear regression, sustainability, random forest regression.

Araştırma Makalesi

Sürdürülebilir havacılık için makine öğrenimine dayalı uçuş içi yemek israfı tahmini

Öz. Bu çalışma, havacılık lojistiğinde sürdürülebilirliği iyileştirmek için uçak içi gıda israfını tahmin etmek ve azaltmak amacıyla makine öğreniminin kullanımını araştırmaktadır. Büyük bir çevresel sorun olan uçak içi gıda israfı, yolcu tercihleri, uçuş parametreleri ve ikram hizmetleri tarafından belirlenmektedir. Bu araştırma, uçuşlar sırasında gıda israfı tahmini yapmak için Çoklu Doğrusal ve Rastgele Orman Regresyonu olmak üzere iki etkili makine öğrenimi algoritması sunmaktadır. Modeller, yemek türü, israf ağırlığı ve yolcu sayısı gibi faktörleri içeren 10.000 kayıt ve 15 özellikten oluşan sentetik olarak oluşturulmuş bir veri kümesi kullanılarak eğitilmiştir. Çalışma, tahmin doğruluğunu artırmak için "Yolcu Başına İsraf" ve "Yemek Verimliliği" gibi yeni özelliklerin geliştirilmesi de dahil olmak üzere önemli özellik mühendisliği üstlenmektedir. En etkili özellikleri belirlemek için bir korelasyon analizi de kullanılmaktadır. Modellerin performansı, Python tabanlı bir hesaplama ortamında değerlendirilmekte olup Çoklu Doğrusal Regresyon yiyecek israfı ve göstergeler arasındaki doğrusal bağlantılara, Rastgele Orman Regresyonu ise doğrusal olmayan etkileşimlere odaklanmaktadır. Sonuçlar, her iki modelin de uçuş sırasındaki gıda israfını etkili bir şekilde tahmin edebildiğini ve Rastgele Orman Regresyonunun karmaşık kalıplara daha iyi uyum sağladığını göstermektedir. Araştırma, havayolu yöneticilerine havacılık lojistiğindeki genel sürdürülebilirlik hedeflerine karşılık gelen veri odaklı israf azaltma tekniklerini uygulamaları yönünde önerilerle sonuçlanmaktadır. Oluşturulan modeller, uçuş sırasındaki yiyecekleri optimize etmek, çevresel etkiyi azaltmak ve sektörün sürdürülebilirlik girişimlerine katkıda bulunmak için kullanışlı bir araçtır.

Anahtar Kelimeler: Uçuş içi yemek israfı, çoklu doğrusal regresyon, sürdürülebilirlik, rastgele orman regresyonu.

1 Logistics Management Department, University of Turkish Aeronautical Association, 06790, Ankara, Türkiye; daghazadeh@thk.edu.tr (Corresponding Author)
https://doi.org/10.52995/jass.1536614
Received: August 21, 2024; Accepted: September 15, 2024; Published: February 28, 2025

1. INTRODUCTION

Aviation logistics refers to the process of planning, executing, and overseeing the movement of goods, services, and information within the aviation industry (Wu & Yang, 2021). This encompasses the administration of resources, infrastructure, and services associated with air transportation, such as aircraft, airports, and ground handling operations. Aviation logistics is important because it facilitates the transfer of goods, services, and people via air transportation. Aviation logistics is critical to international trade, emergency response, and tourism. It allows for the prompt delivery of high-value and time-sensitive items, the support of supply chains, and the ability to respond quickly to disasters and medical situations. Furthermore, air logistics helps to spur economic growth, create jobs, and link regions and nations. It has an important role in supporting numerous businesses and building international relationships (Sambo & Hlengwa, 2018).

Aviation logistics has always been popular, and it is anticipated to be imposed to an exponential growth through the next few years. In 2019, airlines globally transported around 4.5 billion passengers, with this figure expected to rise to approximately 10.0 billion by 2040 (Lohawala & Wen, 2024). This demonstrates a significant and growing demand for air travel, emphasizing the significance of tackling sustainability issues in aviation logistics in order to successfully reduce the unpleasant effects of aviation on the environment (Lohawala & Wen, 2024). The growing popularity of air logistics is having severe consequences for the environment due to the sector's large contribution to climate change and health issues. In the context of aviation logistics and food waste management, sustainability entails reducing the environmental effect of food waste, guaranteeing social responsibility in food handling and trash management, and ensuring economic viability. Airlines and catering firms may help the environment by decreasing food waste, addressing food insecurity, and promoting ethical corporate behavior. This may be done by strategies such as food waste avoidance, reuse, recycling, and energy recovery, in line with the Food Waste Hierarchy (FWH) and CSR standards.

Machine Learning (ML) is a subsection of artificial intelligence that trains computers to learn from data without requiring programming expertise. This technology is quickly altering different sectors, and its potential to solve the critical issue of food waste is enormous. Machine learning algorithms can identify patterns and trends in enormous amounts of data that humans may ignore. This skill has proven useful in the battle against food waste, from farm to fork. According to (Rodrigues, Miguéis, Freitas, & Machado, 2024), short-term demand forecasts are enhanced by utilization of machine learning methods by catering services which leads to the food waste reduction. The study shows that using machine learning to forecast future demand can result in a 14% to 52% reduction in the number of wasted meals and a 3% to 16% reduction in unmet demand when compared to baseline model projections. This can improve food service quality while reducing food waste, therefore lessening its environmental, social, and economic effects.

The chief idea in this research is to predict the inflight food waste amount using effective machine learning algorithms. Hence, considering the features, their characteristic and objective of the research, two efficient and fast algorithms are suggested, that are, Multiple Linear Regression (MLR) and Random Forest Regression (RFR). As far as the author's knowledge, there is not any publicly available inflight waste dataset, hence, the developed models are applied to a synthetically generated dataset, and the results proved the accuracy and efficiency of the suggested algorithms.

The paper is organized as follows: First, the existing literature on food waste reduction, inflight food waste reduction and machine learning in food waste reduction is surveyed. Next, the materials and methods are explained. Section 4 provides the steps and equations for the developed approaches, then in Section 5 the numerical results are analyzed. In the conclusion section, useful managerial insights are drawn using the results from the analysis.

2. LITERATURE REVIEW

In this section, the existing literature on inflight food waste reduction and general food waste reduction are investigated. Available researches are analyzed based on the solution techniques they implemented and the performance that their supposed model has shown. Based on our investigation, there are a limited number of researches which consider inflight waste prediction. Not all of the related researches utilize machine learning methods for carrying out the forecasting operation.

(Blanca-Alcubilla, et al., 2019) aims at enhancing the in cabin waste management and waste production minimization through inflight waste analysis. They evaluate 145 airplanes, separated by flight time and passenger class, and categorized and described the waste into 20 distinct elements. The research seeks to offer a complete, representative, and tailored analysis for the food waste produced in aviation sector aiming at reduction of the waste and giving evidence for a greener management system. The investigation is conducted as part of the European initiative LIFE + Zero Cabin Waste. They distinguish between modified and unmanipulated materials and identified particular materials deemed trash. They make recommendations based on waste characterization, such as employing bi-compartmentalized garbage carts and removing unpopular foods from menus to decrease waste creation, as well. Another study which does not directly focus on waste amount reduction and tackles the factors effecting waste production is (van der Walt & Bean, 2022). The article assesses the benefits of integrating product substitution and meal demand uncertainty in an inventory decision-making model as a potential solution to the inflight catering industry's waste problem. The authors develop a stochastic and multi-objective decision-making model that seeks to discover the most efficient combination of meals to order for a certain trip, given the industry's two competing objectives. The model is intended to balance the relevance of limiting the number of meals loaded onboard a flight in order to save waste while simultaneously decreasing the probability of meal shortages and disgruntled passengers. The article proposes to decrease inflight food waste by including product substitution and meal demand uncertainty into the inventory decision-making model. (Hast, 2019) compares machine learning algorithms (MLAs) for forecasting consumer demand for inflight meals and identifies relevant data elements for this prediction. The study's goal is to find the best number of inflight meals to load onboard a commercial flight in order to reduce food waste upon landing. The study is closely connected to inflight food waste prediction since it aims to address the challenge that airline firms confront when selecting how many meals to carry in order to minimize waste while meeting consumer demand. The project uses MLAs and past flight data to anticipate the number of meals for each trip, decreasing food waste and enhancing customer happiness.

(Megodawickrama, 2017) looks at how passenger load factor fluctuation affects average daily flight kitchen waste in Sri Lanka's airline catering business. The study seeks to find viable techniques for reducing food waste in flight kitchens by studying the variation in passenger load factor (PLF) within 24 hours of an airline's expected time of departure (ETD) for average daily flight kitchen waste. They use historical data from 14 international airlines catered by a flight catering company. It focuses on variables such as initial passenger load factor, final passenger load factor, passenger load factor variability, and number of meals catered per day as independent variables, with production waste per meal as the dependent variable. The study's findings reveals that there is inconsistency in the client airline's initial and final passenger loads, resulting in uncertainty on the manufacturing floor. This inconsistency results in swings in the average daily kitchen waste per meal, which impacts the profit margin. The study also finds that increasing the number of meals per day reduces production waste per meal, and that demand uncertainty has a substantial impact on waste rise in the production area. Overall, the study emphasizes the need of controlling passenger load factor fluctuation in flight kitchens to reduce food wastage. It advocates creating a robust forecasting system and deploying a meal bank system to reduce production waste through smart menu planning. In (Tofalli, Loizia, & Zorpas, 2018), researchers perform a compositional analysis of garbage generated during flights, with an emphasis on food waste, paper, and plastics, as well as how passengers and airline rules influence waste output. The research finds that passenger behaviors and each airline's policy generated various types of trash during flights, particularly food waste. According to the study, customers' food selections and airline rules, such as whether they serve complimentary meals or sell food during flights, have a substantial impact on inflight food waste. For example, giving passengers the choice of reserving or not booking their meals when booking their flights might help to decrease food waste and losses.

In (Dhir, Talwar, Kaur, & Malibari, 2020), a systematic literature review (SLR) is used to critically examine the current condition of food waste in the hotel and food services sectors. This technique is implemented by searching, evaluating, and synthesizing peer-reviewed literature to identify relevant research themes and gaps in existing knowledge. The study focuses on collecting chosen articles around nine topics reflecting various elements of food waste, such as waste generating sources, residual handling, and waste reduction strategies. Furthermore, the authors conduct detailed study profiling to offer summary statistics on the selected papers' research design, data analysis methodologies, variables explored, and theoretical lens applied. The study finishes with a paradigm that combines the findings to inform future empirical research in the field.

Inflight food waste prediction and reduction can be considered from two different aspects: airline-based factors and customer-based factors. Our research is distinguishable from (Halizahari, Mohamad, Anis, & Wan, 2021), in that, they analyze the effect of customer related features such as customer meal preferences on the amount of waste produced instead of investigating the firm related factors. They implement multiple linear regression method as well, and report that among all the features waste management and food quality from the customer side are the most influential items. (Teoh, 2018) proposes a bi-objective inflight food waste reduction model aiming at reducing food waste while increasing customers' expectations of inflight catering service. The strategy entails developing a bi-objective optimization model that incorporates both the supply (airline) and demand (passengers) sides under uncertainty. The concept provides the option of offering a light meal in addition to a conventional meal, allowing customers to choose their meal type during the trip booking process and alter their meal selection before aircraft departure or during their flight. The model is intended to maximize the number of regular and light meals, taking into account both reserved and unreserved meal options. An exemplary case study is performed to show the applicability of the constructed model, utilizing data from international long-haul flights operated by Malaysia Airlines.

However, there is a vast amount of research regarding food waste reduction/ prediction, only a limited part of them concentrates on inflight generated waste management. Major part of the papers investigate the food waste management/ reduction operations in catering firms. To be more illustrative, (Thamagasorn & Pharino, 2019) studies the quantity and composition of food waste created throughout the Halal food production process in the aircraft catering industry. A food waste audit, material flow analysis (MFA), and eco-efficiency analysis are among the techniques used. The food waste inspection involves separating food waste into types, taking pictures, gathering financial documents, and inventorying food purchases. The MFA is utilized to identify hotspots in the food production process, while the eco-efficiency analysis evaluated food commodities as high, medium, or low eco-efficiency, indicating low to high priority for food waste management action. (Ross, 2014) studies food waste in an airline caterer's production kitchen, with the goal of quantifying the waste, understanding why it is generated, and developing solutions to prevent, reduce, or divert it. The study uses a mixed techniques approach, which includes trash audits, observations, document analysis, and interviews. The author also performs qualitative research to better understand the reasons of food waste, an essential factor that is sometimes neglected by quantitative data alone. To identify the causes leading to waste generation, the study uses a system model and process architecture particularly tailored for the production operations of an airline-catering kitchen. The study's goal is to give insights into the airline catering sector, which has had little publicly available research, as well as to evaluate the implementation of waste avoidance techniques. (Phothisuk, 2019) intends to investigate the waste situation, waste types, waste reduction recommendations, and waste reduction outcomes from Thai airlines' inflight services. The researchers conduct qualitative research, which includes document analysis and in-depth interviews with airline officials and flight attendants from several Thai carriers. The data is collected and evaluated using content analysis techniques.

According to the author's current knowledge, literature lacks research on implementation of machine learning methods put forward in this study, namely, multiple linear regression and random forest regression, to predict the amount of food waste generated during a flight. The paper mainly aims at assessing the effect of different features on the amount of the food waste generated in the flights and predict the waste values according to the features recognized as vital.

3. MATERIALS AND METHODS

3.1. Dataset Characteristics

Due to the limited open-source dataset and researches in this field, we used synthetically generated data. The dataset used in this research is simulated based on real life data, that is, the airplane characteristics and the ratios among features are adjusted and taken into account based on the real-life cases. The dataset closely resembles real-world aircraft operations, particularly in the areas of inflight meals and trash management. To exemplify, considering flight duration hours, duration data is consistent with real flight timings, which range from small local trips to long-haul international flights. In reality, flight lengths are crucial when organizing food services since longer trips require more meals and snacks. In terms of passenger count, it reflects usual airline capacity, which varies according to aircraft type. For example, smaller regional planes carry fewer passengers than bigger aircraft such as the Boeing 777 or Airbus A380. One of the features which directly influence the amount of waste produced, is the meals prepared/served. Airlines prepare meals according to the number of passengers, level of service (economy, business, first), and flight time. The quantity of meals provided is frequently somewhat lower than the amount produced owing to a variety of circumstances such as passenger preferences, dietary restrictions, or pre-booked special meals. Uneaten meals, are potential waste production resources, which are left uneaten owing to over-preparation, changing passenger preferences, or flight delays. The data covers this feature, which is a potential source for a considerable amount of food waste in aircraft.

Overall, the dataset represents the complexities of inflight meal and trash management, which are essential considerations in sustainable airline logistics. The data shows real-world difficulties and potential for development in decreasing food waste and increasing operational efficiency.

Dataset is made up of different features related to flight data. To ensure that the size of dataset does not lead to the problem of curse of dimensionality, the generated dataset is planned in a way that it contains 10,000 rows and 15 columns which correspond to the features. Dataset contains parameters and features which affect the possible amount of the inflight food waste. These features and their types are presented in Table 1:

Table 1. Main Features and Their Types

Feature	Type
Date	Categorical - DateTime
Flight ID	Categorical-Numerical/ID
Flight Duration Hours	Numerical
Passengers	Numerical
Meal Type	Categorical - String
Time of Day	Categorical - String
Destination	Categorical - String
Meals Prepared	Numerical
Meals Served	Numerical
Uneaten Meals	Numerical
Waste Weight kg	Numerical
Waste Type	Categorical - String
Waste Cost USD	Numerical
Loading Time min	Numerical

3.1.1. Feature engineering

Feature engineering is a fundamental step in machine learning that entails converting raw data into useful features that improve model performance. By carefully choosing, changing, and generating features, data scientists may increase machine learning models' accuracy, efficiency, and interpretability. The efficacy of these models is frequently dependent on the quality of the features given, making feature engineering a critical stage in the data science pipeline. The following main strategies are frequently used to refine features and improve the prediction power of machine learning systems:

1. Feature Selection: This entails finding and picking the most relevant characteristics from the data that contribute the most to the desired variable. Common techniques include correlation analysis, recursive feature removal, and regularization methods like as Lasso.

2. Feature Transformation: This approach includes modifying existing characteristics to improve their prediction potential. Examples include:

    I. Normalization and standardization including scaling characteristics to a standard range or distribution.
    II. Logarithmic Transformations: Used to deal with skewed data distributions.
    III. Polynomial Features: Adding new features by increasing current ones to a power.
3. Feature Creation: Existing features can be combined or manipulated to create new ones. For example:
    I. Interaction Terms: Creating features that depict the interaction of two or more existing features.
    II. Aggregated features: Data summation (e.g., mean, sum, count) across groups or time periods.
4. Handling Missing Data: Dealing with missing values is an important part of feature engineering. To highlight missing values, strategies like imputation (e.g., mean, median, mode) or the creation of an indicator variable are used.
5. Encoding Categorical Variables: Converting categorical data into a numerical representation that machine learning algorithms can handle. Techniques include:

I. One-hot encoding involves creating binary columns for each category.
II. Label encoding involves assigning a unique number to each category.

Feature engineering is necessary because machine learning algorithms frequently cannot operate well with raw data. The process of developing features can:

1. Improve Model Performance: Well-engineered features can result in considerable increases in model accuracy and resilience.
2. Reduce Model Complexity: By carefully choosing and designing features, you may minimize the dimensionality of your data, resulting in simpler and more interpretable models.
3. Improve Generalization: Good feature engineering can decrease overfitting, allowing the model to generalize more effectively to new data.

Using feature creation technique, more informative and influential features are created from the existing data. Extra features, their type and the way they were generated are presented below in Table 2:

Table 2. Extra Features, Their Type and Formula

Extra Features	Type	Formula
Waste per Passenger	Numerical	Waste Weight kg / Passengers
Meal Efficiency	Numerical	Meals Served / Meals Prepared
Appetite Index	Numerical	Meals Served / Passengers
Uneaten Meals per Passenger	Numerical	Uneaten Meals / Passengers
Waste Cost per kg	Numerical	Waste Cost USD / Waste Weight kg

3.1.2 Feature analysis

Feature analysis is an essential component of data preparation and exploratory data analysis (EDA) in any data science or machine learning effort. It entails analyzing the properties and distribution of each feature in a dataset to better understand the underlying patterns, identify outliers, and uncover relationships between features. This technique aids in assessing the importance of each feature, directing feature selection, and offering insights for model selection and tweaking. Mean, Max, Min and standard deviation for numerical features is presented in Table 3:

Table 3. Exploratory Data Analysis Results for Numerical Features

	Mean	Max	Min	Std
Flight Duration Hours	7.97	15.00	1.00	4.02
Passengers	200.00	254.00	152.00	14.22
Meals Prepared	249.87	308.00	191.00	15.86
Meals Served	199.84	251.00	151.00	14.06
Uneaten Meals	20.02	39.00	5.00	4.48
Waste Weight kg	14.92	25.00	5.00	5.76
Waste Cost USD	176.18	300.00	50.00	72.46
Loading Time min	45.18	60.00	30.00	8.63
Storage Temp C	5.01	8.00	2.00	1.74
Waste per Passenger	0.07	0.15	0.02	0.03
Meal Efficiency	0.80	1.00	0.55	0.08
Appetite Index	1.00	1.43	0.71	0.10
Uneaten Meals per Passenger	0.10	0.20	0.03	0.02
Waste Cost per kg	14.24	59.10	2.00	9.47

Figure 1 presents the charts for different numerical features utilized in the process of prediction:

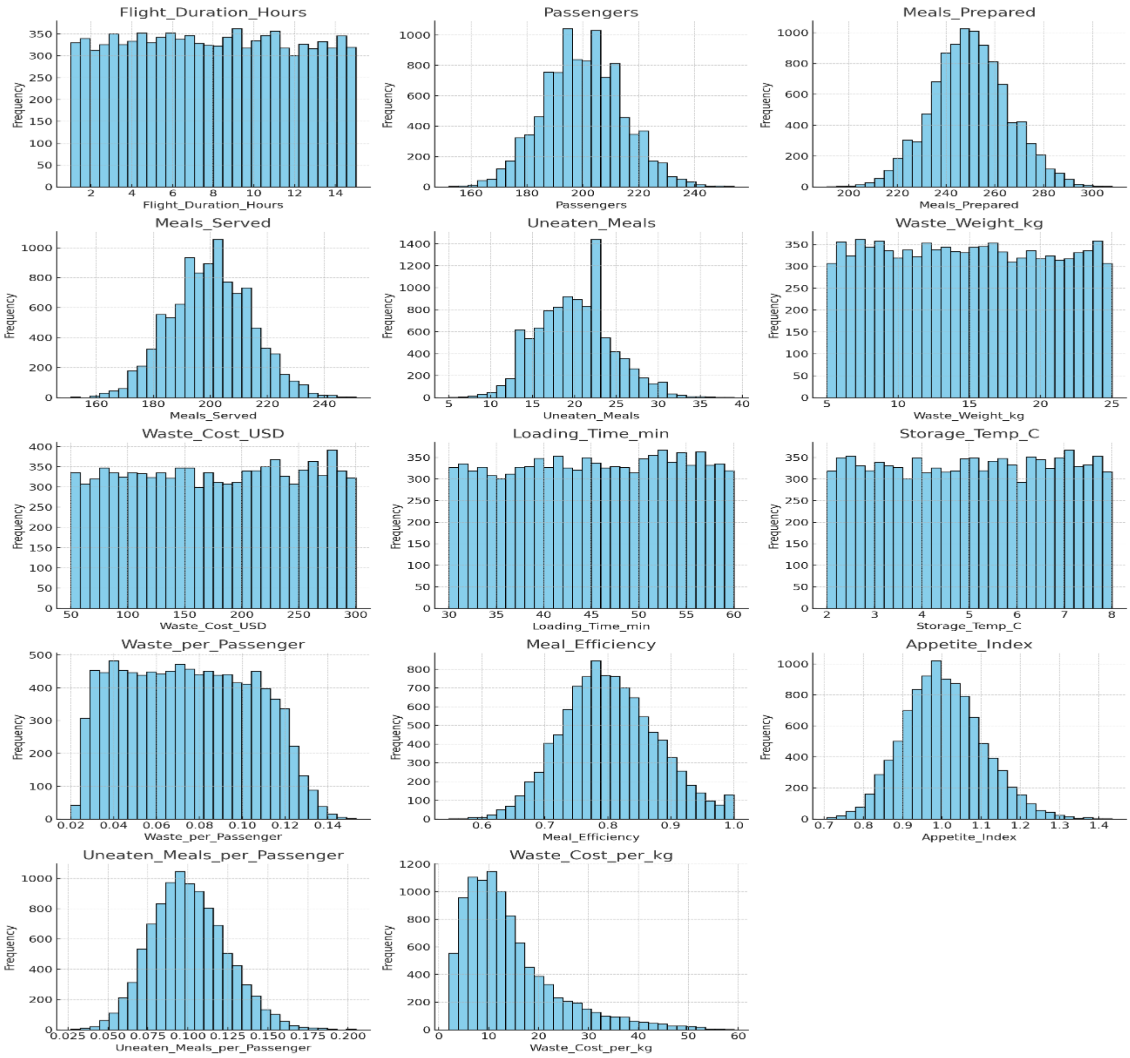

Figure 1. Histograms for Numerical Features

3.2. Correlation Analysis

This paper mainly aims at predicting the inflight food waste amount based on flight, customer and meals information. In order to yield more effective and accurate results, correlation analysis is carried out. Correlation analysis is a statistical procedure that determines and describes the degree and direction of a link between two variables. Correlation analysis determines if and how strongly two variables are connected. Using correlation coefficient values for each feature, they are ordered and selected based on their influence on the amount of waste generated. The following table illustrates the correlation coefficient for each of the features in relation to waste weight:

Table 4. Correlation Coefficient for each of the Features in Relation to Waste Weight Kilogram

Features	**Waste Weight (kg)**
Waste per Passenger	0.981
Meals Prepared	0.012
Passengers	0.004
Uneaten Meals	0.003
Loading Time min	0.001
Uneaten Meals per Passenger	0.000
Waste Cost USD	-0.002
Flight ID	-0.003

Flight Duration Hours	-0.007
Storage Temp C	-0.013
Appetite Index	-0.022
Meals Served	-0.023
Meal Efficiency	-0.025
Waste Cost per kg	-0.666

The correlation coefficient values indicate that waste per passenger is the most correlated feature with a value of 0.981. Using feature selection, which is defined as the process of identifying the most important variables (features) from a larger collection of data that contribute the most to a model's prediction performance. The primary purpose is to increase model efficiency and accuracy by removing unnecessary or redundant features, lowering the risk of overfitting, and reducing computing complexity. Four of the highly correlated features are selected for the prediction model generation:

1. Waste per Passenger,
2. Waste Cost per kg,
3. Meals Prepared
4. Passengers.

4. DEVELOPED APPROACHES

Here, two powerful machine learning methods are suggested for performing the forecasting task: Multiple linear regression and Random Forest regression. The following subsections describe the proposed techniques and give the formula.

4.1. Multiple Linear Regression

Modeling the relationship between a single variable (often denoted as Y) and two or more independent variables (predictors, denoted as $X_1, X_2, X_3, \dots, X_n$) can be performed by developing a Multiple Linear Regression (MLR) model. MLR, known as a statistical method which is appropriate for this aim, is used to anticipate the value of the dependent variable based on the values of the independent factors and to comprehend how modification in the independent variables effect changes in the dependent variable. MLR is employed when the result or dependent variable is impacted by several factors. Some common uses are given below:

Economics: Predicting consumer spending based on income, age, and educational attainment.

Healthcare: Evaluating the impact of many risk variables on health outcomes, such as blood pressure, which is affected by age, food, and activity.

Marketing involves determining the influence of promotion, price, and product quality on sales.

Environmental Science: Modeling the effects of temperature, humidity, and pollution on agricultural productivity.

Considering the very basic formula for the multiple linear regression model, we have:

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \dots + \beta_n X_n + \epsilon,$$

where

1. Y is the dependent variable (the outcome you are predicting or explaining), in our case, Waste Weight kg,
2. β_0 is the intercept (the value of Y when all X_i are 0),
3. $\beta_0, \beta_1, \beta_2, \beta_3, \dots, \beta_n$ are the coefficients corresponding to each independent variable (representing the change in Y for a one-unit change in X, assuming all other variables are held constant),

4. ϵ is the error term, signifying the variation between Y's observed and expected values.

4.2. Random Forest Regression

Random Forest Regression (RFR) is a type of regression that does not make the same assumptions as linear regression. It entails creating numerous regression trees using a random subset of data from the training set and then averaging the regression tree solutions to predict the target variable with the minimum mean squared error (MSE). Unlike linear regression, RFR automatically integrates interactions between distinct predictor variables and does not need variable selection, making it simpler to accept complicated, non-linear interactions among variables. Random Forest Regression is an ensemble learning approach for regression challenges that trains many decision trees and delivers the average of their predictions. It is an extension of the Random Forest technique, which is usually used for classification but can also predict continuous values in regression.

Random Forest Regression is utilized when the connection between input features and target variables is complicated, non-linear, and contains feature interactions that are challenging to represent using typical linear models. This method works well when dealing with big amounts of data and several features. The data contains complicated patterns that are difficult to represent with a single model. Overfitting is a risk, and Random Forest tends to lessen overfitting when compared to individual decision trees.

RFR is preferred to other regression methods due to the following characteristics:

1. Reduction of Overfitting: Random Forest reduces overfitting by averaging multiple decision trees. To increase variety and lower the possibility of overfitting, each tree is trained using a random subset of features on a random part of the data.
2. Handling Non-Linear Relationships: It captures non-linear relationships and interactions between features that linear models might miss.
3. Robustness to Outliers and Noise: Because it aggregates predictions from multiple trees, it is less sensitive to outliers and noisy data.
4. Feature Importance: Random Forest provides a measure of feature importance, helping to identify which features are most predictive of the target variable.
5. Versatility: It performs well with both categorical and continuous input features and doesn't require extensive data preprocessing like normalization or scaling.
6. Parallelization: Since trees in a forest are independent of each other, the model can be parallelized, leading to faster computation times on large datasets.

RFR does not have a single clear formula, unlike linear regression. Instead, the steps are as follows:

1. Bootstrap Sampling: Select n samples at random from the dataset and replace them to construct a subset of the data for each tree.
2. Tree Building: Create a decision tree for each subgroup. Each node selects a random selection of characteristics and determines the optimum split from this subset.
3. Prediction:
   I. For every tree: Predict the output (a continuous value) using the decision rules of the tree.
   II. Final Output: The final forecast is the average of all the trees in the forest.

5. NUMERICAL RESULTS AND ANALYSIS

This section provides numerical results from implemented techniques and reveals hidden information in the available dataset. Additionally, some useful managerial insights related to inflight food waste management and reduction are drawn. The following subsections provide the technical information of the environment used for data analysis, results for MLP and RFR, correspondingly.

5.1. Technical Environment

This research is conducted using a computing environment with the following hardware and software specifications:

5.1.1. Hardware configuration

The computations are performed on a system equipped with an Intel® Core i5-1135G7 processor, operating at a clock speed of 2.40 GHz. The system had 32 GB of RAM for storage.

5.1.2. Software environment

The implementation is carried out using Python 3.10, within a PyCharm Community environment. Key libraries utilized in the analysis include pandas (version 1.2.3) for data manipulation, NumPy (version 1.19.2) for numerical computations, scikit-learn (version 0.24.1) for machine learning algorithms, statsmodels (version 0.12.2) for statistical modeling, and matplotlib (version 3.3.4) for data visualization. The research environment is managed using a Conda virtual environment configured with Python 3.8.

5.1.3. Execution details

The entire analysis, including data preprocessing, model training, and evaluation, is executed using PyCharm Community Edition. The total execution time for the project is approximately 15 minutes. The Random Forest model training was optimized using parallel processing techniques to enhance computational efficiency.

5.1.4. Special configurations

No special configurations are required beyond standard library settings, except for enabling parallel processing during the Random Forest model training to accelerate the computation.

5.2. Multiple Linear Regression Results

To implement the MLR algorithm, first the dataset is prepared and the most influential features are selected as: Waste per Passenger, Waste Cost per kg, Meals Prepared, Passengers. The target variable which is aimed to be predicted is Waste Weight kg.

Training and testing sets are generated with an 80-20 ratio. This is done by calling the train_test_split method in the sklearn.model_selection module, which ensures that 80% of the data is utilized for training and 20% for testing. A random state of 42 is used to assure repeatability. A multiple linear regression model is employed for analyzing the data. The model fitting is conducted using the training data. The statsmodels library facilitates the construction and fitting of the multiple linear regression model. The multiple linear regression model's performance is thoroughly examined using a number of statistical criteria to ensure a comprehensive grasp of its predictive capabilities. The following metrics were computed:

1. Mean Squared Error (MSE): MSE measures the average squared difference between the observed actual outcomes and the outcomes predicted by the model. It provides an indication of the model's accuracy, with lower values indicating better performance.
2. Root Mean Squared Error (RMSE): RMSE is the square root of MSE and provides an error metric in the same units as the dependent variable, making it more interpretable. Like MSE, lower RMSE values signify better model performance.
3. Mean Absolute Error (MAE): MAE represents the average absolute difference between observed and predicted values. It is a straightforward measure of prediction accuracy, with lower values indicating better performance.
4. R-squared (R^2): R^2 is a statistical measure that represents the proportion of the variance for the dependent variable that is explained by the independent variables in the model. An R^2 value closer to 1 indicates a higher explanatory power of the model.
5. Adjusted R-squared: Adjusted R^2 adjusts the R^2 value based on the number of predictors in the model, providing a more accurate measure when multiple predictors are used. It penalizes the addition of non-significant predictors, thereby preventing overfitting.
6. Mean Absolute Percentage Error (MAPE): MAPE measures the accuracy of the model's predictions as a percentage, making it easier to understand the model's performance in relative terms. Lower MAPE values indicate higher accuracy.
7. Coefficient of Determination (COD): COD, another term for R^2, reflects the goodness of fit of the model. It quantifies the proportion of the variance in the dependent variable that is predictable from the independent variables.

The model is trained on 80% of the dataset and tested on the remaining 20%. Model results and values for evaluation metrics are summarized in Table 5 and Table 6, respectively, providing a comprehensive overview of the model's predictive performance. These metrics collectively indicate the robustness of the model and its suitability for predicting Waste Weight kg based on the selected features.

Table 5. Model Results for MLR

	Coefficients	P-Value	CI 2.5%	CI 97.5%
Const	-14.410	0.000	-14.624	-14.196
Waste per Passenger	196.924	0.000	196.491	197.357
Waste Cost per kg	-0.004	0.000	-0.005	-0.002
Meals Prepared	0.000	0.708	-0.001	0.000
Passengers	0.073	0.000	0.073	0.074

Table 6. MLR Model Performance Evaluation Metrics

Metric	Value
Mean Squared Error (MSE)	0.177
Root Mean Squared Error (RMSE)	0.420
Mean Absolute Error (MAE)	0.291
R-squared	0.995
Adjusted R-squared	0.994
Mean Absolute Percentage Error (MAPE)	2.576
Coefficient of Determination (COD)	0.995

The following chart shows the results from MLR in terms of actual and predicted waste amount:

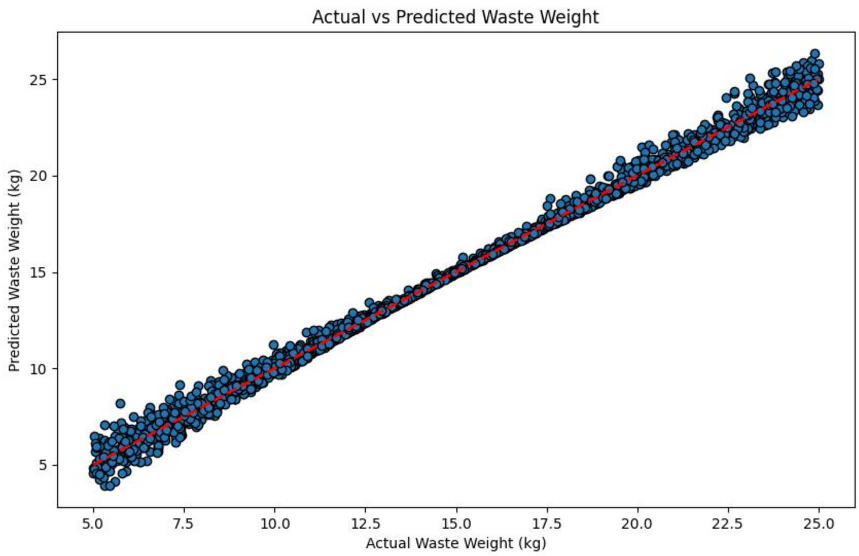

Figure 2. Actual Waste vs Predicted Waste Weight

The following sections discuss the model findings and the performance metrics.

5.2.1. Interpretation of the model findings

Table 5 summarizes the multiple linear regression analysis findings. The model has four predictor variables: waste per passenger, waste cost per kg, meals prepared, and passengers. The intercept (constant) term is -14.41, which represents the dependent variable's baseline level when all predictors are zero. The coefficient for Waste per Passenger is 196.92, and the p-value is less than 0.001, indicating a significant positive link and statistical significance. Waste Cost per Kg has a negative coefficient of -0.0037 and a highly significant p-value, showing a minor but significant inverse association with the dependent variable. In contrast, Meals Prepared has a coefficient of -0.00011 and a high p-value (0.708), indicating no significant influence. Passengers have a positive and substantial influence, with a coefficient of 0.073. The confidence intervals supplied for each coefficient enhances its dependability.

5.2.2. Interpretation of model performance evaluation results

The model evaluation results indicate an excellent performance of the multiple linear regression model in predicting Waste Weight kg. The key metrics are interpreted as follows:

1. Mean Squared Error (MSE): An MSE of 0.177 suggests that on average, the squared errors of the predictions are quite low, indicating high accuracy of the model.
2. Root Mean Squared Error (RMSE): An RMSE of 0.420 implies that the model's predictions deviate from the actual values by approximately 0.42 units of Waste Weight kg on average. This relatively low RMSE further supports the high accuracy of the model.
3. Mean Absolute Error (MAE): An MAE of 0.291 indicates that, on average, the model's predictions are off by about 0.291 units of Waste Weight kg, which is a small error margin, signifying precise predictions.
4. R-squared (R^2): An R^2 of 0.995 means that 99.5% of the variance in Waste Weight kg is explained by the model. This extremely high R^2 indicates an excellent fit of the model to the data.
5. Adjusted R-squared: An adjusted R^2 of 0.994, very close to the R^2 value, confirms that the model's predictors are highly significant and that the model is not overfitting.
6. Mean Absolute Percentage Error (MAPE): A MAPE of 2.576 indicates that the model's predictions deviate from the actual values by an average of 2.576%, which is a very low percentage error, suggesting high predictive accuracy.
7. Coefficient of Determination (COD): COD, synonymous with R^2, also shows that 99.5% of the variance in Waste Weight kg is predictable from the independent variables, reinforcing the robustness of the model.

5.2.3. Insights on reducing food waste based on model results

The multiple linear regression model gives valuable insights into the factors that influence food waste, allowing us to develop meaningful waste-reduction initiatives. Here are some important findings drawn from the model's performance and feature significance:

1. Optimize the Meals Prepared:
   - I. Feature Insight: Meals Prepared is a predictor in the model, showing a direct link to Waste Weight kg.
   - II. Actionable Strategy: By studying the link between the quantity of meals cooked and the waste generated, operations may be changed to better match meal preparation to passenger demand. Implementing real-time data analytics and demand predictions can assist in preparing the appropriate number of meals, eliminating overproduction and waste.
2. Manage Waste per Passenger:
   - I. Feature Insight: Waste per Passenger is another key predictor. This shows that individual passenger behavior and waste creation are important considerations.
   - II. Actionable Strategies: Initiatives to educate passengers about waste reduction, such as giving reduced portion sizes or allowing them to modify meal quantities, can help reduce waste per passenger. Encouraging passengers to pre-select meals can also help to better match culinary tastes, resulting in less wasted food.
3. Control waste costs per kilogram:
   - I. Feature Insight: Waste Cost per kg influences total waste weight, suggesting a relationship between trash disposal costs and waste generation.
   - II. Actionable Strategies: Recycling and composting can help to reduce trash disposal costs. Furthermore, rewarding waste reduction through cost-saving strategies might encourage

employees and passengers to reduce waste. Implementing a strong waste management system that monitors trash disposal expenses and discovers cost-cutting alternatives can be useful.

4. Monitor Passenger Numbers:
   I. Feature Insight: The number of passengers is a strong predictor of waste weight.
   II. Actionable Strategy: Accurate passenger forecasts can aid in improved meal planning and waste control. Implementing dynamic inventory management systems that change meal preparation based on real-time passenger data will help guarantee that food supply and demand match more closely, decreasing waste.

Considering the insights drawn from the analysis, firms may considerably minimize food waste by concentrating on the regression model's essential parameters, which include meals provided, waste per passenger, waste cost per kilogram, and passenger numbers. Using data-driven initiatives and developing a sustainable culture may result in more efficient operations, financial savings, and a beneficial environmental effect.

5.3. Random Forest Regression Results

This research aims at implementing a Random Forest Regression model to estimate inflight waste amounts based on several key features: waste per passenger, waste cost per kilogram, the number of meals prepared, and the number of passengers. Random Forest Regression, a robust ensemble learning method, is particularly effective for handling complex datasets with multiple features and non-linear relationships. It operates by constructing a multitude of decision trees during training and outputs the average prediction of the individual trees, thereby improving predictive accuracy and controlling overfitting.

The model is developed using the Python programming language within a PyCharm environment. The dataset is split into training (80%) and testing (20%) subsets to evaluate the model's performance. The Random Forest Regressor from the scikit-learn library is used with 100 estimators, ensuring a sufficient number of trees to capture the intricacies of the data. The "random_state" parameter is set to 42 to ensure the reproducibility of results.

5.3.1. Model evaluation

The model's performance is evaluated using several metrics as tabulated in Table 7:

Table 7. Performance Evaluation for Random Forest Regression Method

Metric	Value
Mean Absolute Error (MAE)	0.041
Root Mean Squared Error (RMSE)	0.070
R-squared	1.000
Explained Variance	1.000
Mean Absolute Percentage Error (MAPE)	0.003

These metrics demonstrate the model's high accuracy, with an R-squared value very close to 1, indicating that the model explains nearly all the variability in the response data. The low MAE and RMSE values further suggest minimal errors in the predictions. The Explained Variance Score reflects the proportion of the variance that the model accounts for, and the MAPE provides an intuitive measure of the error in percentage terms. The predicted and actual waste weights are illustrated in Figure 3:

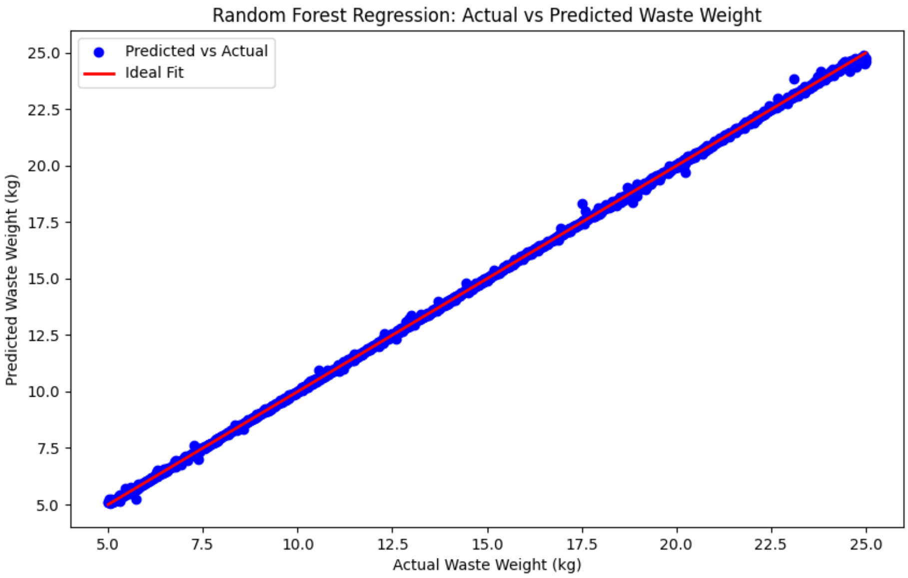

Figure 3. Random Forest Regression: Actual vs. Predicted Waste Weight

The Random Forest model also offers insights into how important each feature is in predicting inflight waste weight. The Waste per Passenger, Waste Cost per kg, Meals prepared, and Passengers are identified as significant contributors, with their importance ranking derived from the mean decrease in impurity criterion used within each decision tree.

To summarize, the Random Forest Regression model developed in this study effectively predicts inflight waste amounts with high accuracy, providing valuable insights for optimizing waste management strategies in aviation logistics. By integrating these predictions with operational decision-making, airlines can enhance sustainability, reduce unnecessary waste, and improve overall efficiency.

The use of a Random Forest Regression model to analyze inflight food waste reveals crucial factors—such as meal preparation, passenger count, waste per passenger, and waste disposal costs—that have a substantial impact on waste creation. Based on these findings, airlines may improve meal forecasting by more precisely matching meal portions with passenger data, avoiding unnecessary food production and waste. Additionally, integrating dynamic pricing for trash management and providing passengers with configurable meal options might help to further reduce waste. These efforts will not only lower the environmental effect of inflight operations, but will also result in financial savings, increased passenger pleasure, and a greater commitment to corporate social responsibility.

5.4. Comparison of the Performance for Suggested Solution Techniques

When comparing the results for inflight food waste management, the Random Forest regression model significantly outperforms the multiple regression method across various key performance metrics:

1. Error Metrics:
   I. The Mean Absolute Error (MAE) for Random Forest is significantly lower (0.041) compared to the multiple regression method (0.291). Similarly, the Root Mean Squared Error (RMSE) for Random Forest (0.070) is much lower than that of multiple regression (0.420). These lower error values indicate that the Random Forest model's predictions are closer to the actual waste amounts, making it a more accurate predictor.
   II. The Mean Absolute Percentage Error (MAPE) also favors Random Forest, with an exceptionally low value of 0.003, compared to 2.576 for multiple regression. A lower MAPE indicates that the Random Forest model's predictions have less relative error, making it more reliable for practical decision-making.
2. Goodness of Fit:

I. The R-squared and Adjusted R-squared values for multiple regression are 0.995 and 0.994, respectively, indicating a very good fit. However, the R-squared and Explained Variance for the Random Forest model are both 1.000, suggesting a near-perfect fit to the data.
II. The Coefficient of Determination (COD), which is equivalent to the R-squared value for multiple regression, further supports this comparison, where Random Forest again demonstrates superior predictive performance.

Given these results, the Random Forest regression model is the superior choice for inflight food waste management due to its higher accuracy, lower prediction errors, and better overall model fit. This model's potential to recognize complicated correlations and interconnections in data makes it ideal for anticipating inflight garbage, enabling more effective and accurate management tactics that can drastically cut waste and improve sustainability efforts.

6. CONCLUSION

In this research, we addressed the critical issue of in-flight food waste in the airline logistics business, with a focus on employing machine learning methodologies to anticipate and reduce loss. The study emphasizes the environmental, economic, and social consequences of inflight food waste, which aligns with the overall aims of sustainable airline logistics. The aviation industry has witnessed extraordinary expansion, and this is anticipated to continue in the future decades. This expansion, although economically advantageous, presents considerable sustainability concerns, notably in terms of managing inflight food waste. This waste has a significant environmental impact, increasing greenhouse gas emissions and other types of pollution. Furthermore, the economic impact of food waste, which includes trash disposal costs and resource loss, highlights the importance of appropriate waste management systems.

Machine learning provides an alternate perspective to this problem. By studying trends in huge datasets, machine learning algorithms can reliably forecast the amount of food waste created during flights, allowing airlines to optimize their meal supply procedures. In this study, we developed and tested two predictive machine learning models to forecast onboard food wastage using a range of information such as passenger count, meal type, and flight length. The results show that these models are successful at forecasting inflight food waste. The Multiple Linear Regression model, while basic and easy to understand, gave useful insights into the linear correlations between the selected parameters and waste generation. Considering prediction accuracy, RFR model beat the other model because of its ability to capture non-linear correlations and data interactions.

The study's key results include identifying the most relevant factors impacting food waste, such as waste per passenger and meal efficiency. These insights can help airlines make data-driven waste-reduction decisions, such as tailoring meal preparations to passenger data and flight parameters. The managerial implications of this study are important. Airlines that use machine learning-based waste prediction models can improve operational efficiency, minimize food waste costs, and contribute to environmental sustainability. The models created in this work may be linked into airline decision-making processes, generating real-time forecasts that guide meal provisioning and waste management techniques.

In conclusion, this study highlights the potential for machine learning to handle the major issue of in-flight food waste in the airline sector. Airlines may achieve more sustainable operations by utilizing data and sophisticated analytics, lowering their environmental impact while increasing profits. Future study might build on this work by combining real-world data from numerous airlines and investigating new machine learning approaches to improve forecast accuracy and operational efficiency. The continuing development and use of such models will be critical to reaching the overall objective of sustainable aviation logistics.

ETHICAL STATEMENT & GENERAL STATEMENTS

This paper meets the research and publication ethics standards.

AUTHORS' CONTRIBUTIONS

The authors have read and approved the final manuscript.

FUNDING

Not applicable.

AVAILABILITY OF DATA AND MATERIALS

Not applicable.

COMPETING INTERESTS

The authors declare that they have no competing interests.

REFERENCES

Blanca-Alcubilla, G., Roca, M., Bala, A., Sanz, N., De Castro, N., & Fullana-I-Palmer, P. (2019). Airplane cabin waste characterization: Knowing the waste for sustainable management and future recommendations. *Waste Management, 96*, pp. 57-64. doi:10.1016/j.wasman.2019.07.002

Dhir, A., Talwar, S., Kaur, P., & Malibari, A. (2020). Food waste in hospitality and food services: A systematic literature review and framework development approach. *Journal of Cleaner Production, 270*(122861). doi:10.1016/j.jclepro.2020.122861

Halizahari, M., Mohamad, M. H., Anis, W., & Wan, A. (2021). A study on in-flight catering impacts on food waste. *Solid State Technology, 64*(2), pp. 4656-4667.

Hast, M. (2019). Evaluation of machine learning algorithms for customer demand prediction of in-flight meals. Retrieved February 6, 2025, from https://www.diva-portal.org/smash/get/diva2:1337269/FULLTEXT01.pdf

Lohawala, N., & Wen, Z. P. (2024). *Navigating Sustainable Skies: Challenges and Strategies for Greener Aviation.* Retrieved February 21, 2025, from https://media.rff.org/documents/Report_24-07.pdf

Megodawickrama, P. L. (2017). Impact of Passenger Load Factor Variability on Average Daily Flight Kitchen Waste in Flight Catering Industry in Sri Lanka. Retrieved from http://dl.lib.uom.lk/bitstream/handle/123/14193/TH3665.pdf?sequence=2&isAllowed=y

Phothisuk, A. (2019). Waste reduction from the in-flight services of Airlines in Thailand. *St. Theresa Journal of Humanities and Social Sciences, 5*(2), pp. 110-119.

Rodrigues, M., Miguéis, V., Freitas, S., & Machado, T. (2024). Machine learning models for short-term demand forecasting in food catering services: A solution to reduce food waste. *Journal of Cleaner Production, 435*(140265).

Ross, J. (2014). Food waste in an airline caterer's production kitchen. University of Otago. Retrieved February 6, 2025, from https://hdl.handle.net/10523/5486

Sambo, N., & Hlengwa, D. (2018). Post-flight food waste and corporate social responsibility at South Africa Airways: Perceptions of employees at Air Chefs South Africa. *African Journal of Hospitality, Tourism and Leisure, 7*, pp. 1-17.

Teoh, L. E. (2018). A bi-objective optimization approach for inflight food waste reduction. *E3S Web of Conferences. 65*. 10.1051/e3sconf/20186504001: EDP Sciences. doi:10.1051/e3sconf/20186504001

Thamagasorn, M., & Pharino, C. (2019). An analysis of food waste from a flight catering business for sustainable food waste management: A case study of halal food production process. *Journal of Cleaner Production, 228*, pp. 845-855.

Tofalli, N., Loizia, P., & Zorpas, A. A. (2018). Passengers waste production during flights. *Environmental Science and Pollution Research, 25*(36), pp. 35764-35775. doi:10.1007/s11356-017-0800-x

van der Walt, A., & Bean, W. L. (2022). Inventory management for the in-flight catering industry: A case of uncertain demand and product substitutability. *Computers & Industrial Engineering, 165*(107914). doi:10.1016/j.cie.2021.107914

Wu, P. J., & Yang, C. K. (2021). Sustainable development in aviation logistics: Successful drivers and business strategies. *Business Strategy and the Environment, 30*(8), pp. 3763-3771.