# Airline Food Demand Prediction

Final Project Report

*Generated on: January 07, 2026*

*Prepared for: SE390 - Artificial Intelligence Projects*

## 1. Problem Statement

Airlines typically operate on thin profit margins, and in-flight catering is a significant operational cost. Over-catering leads to food waste, increased weight, and higher fuel consumption. Under-catering results in negative passenger experiences and potential compensation claims.

This project aims to solve this optimization problem by developing a Machine Learning model that predicts the exact number of meals required for a flight. By analyzing features such as flight duration, passenger demographics, and flight type, we can move away from static ratios to dynamic, data-driven predictions.

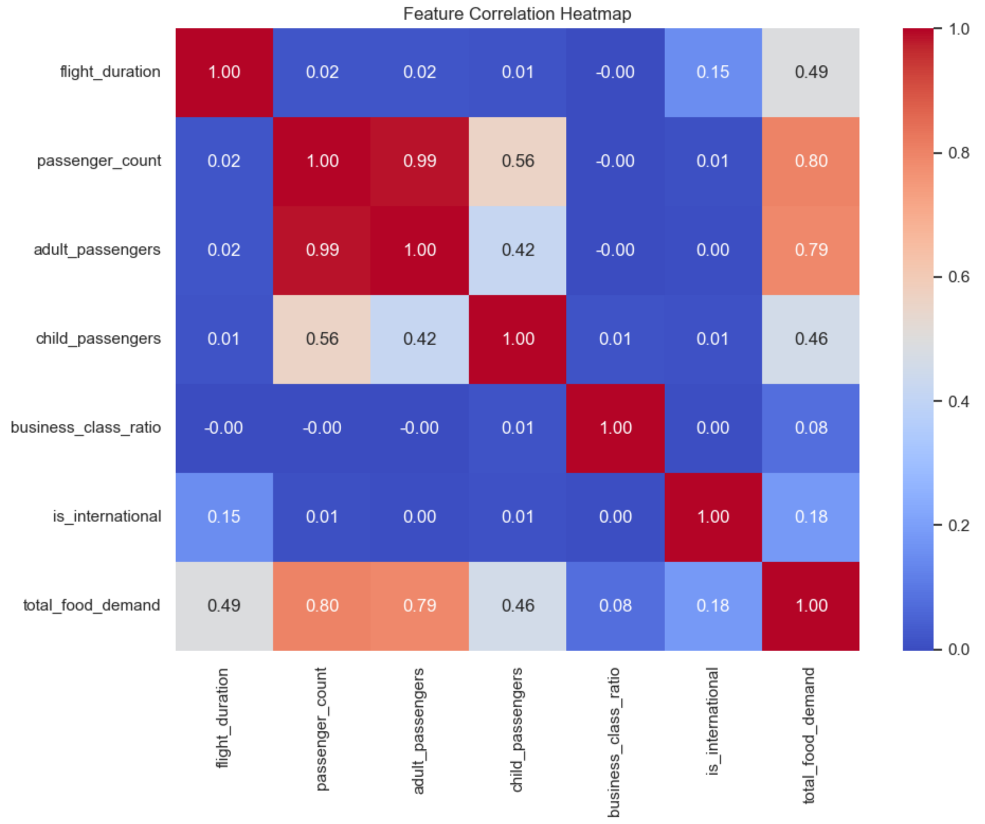
## 2. Dataset Methodology

A synthetic dataset comprising 5,500 flight records was generated to simulate realistic airline operations. The data generation process adhered to strict validation rules to ensure logical consistency:

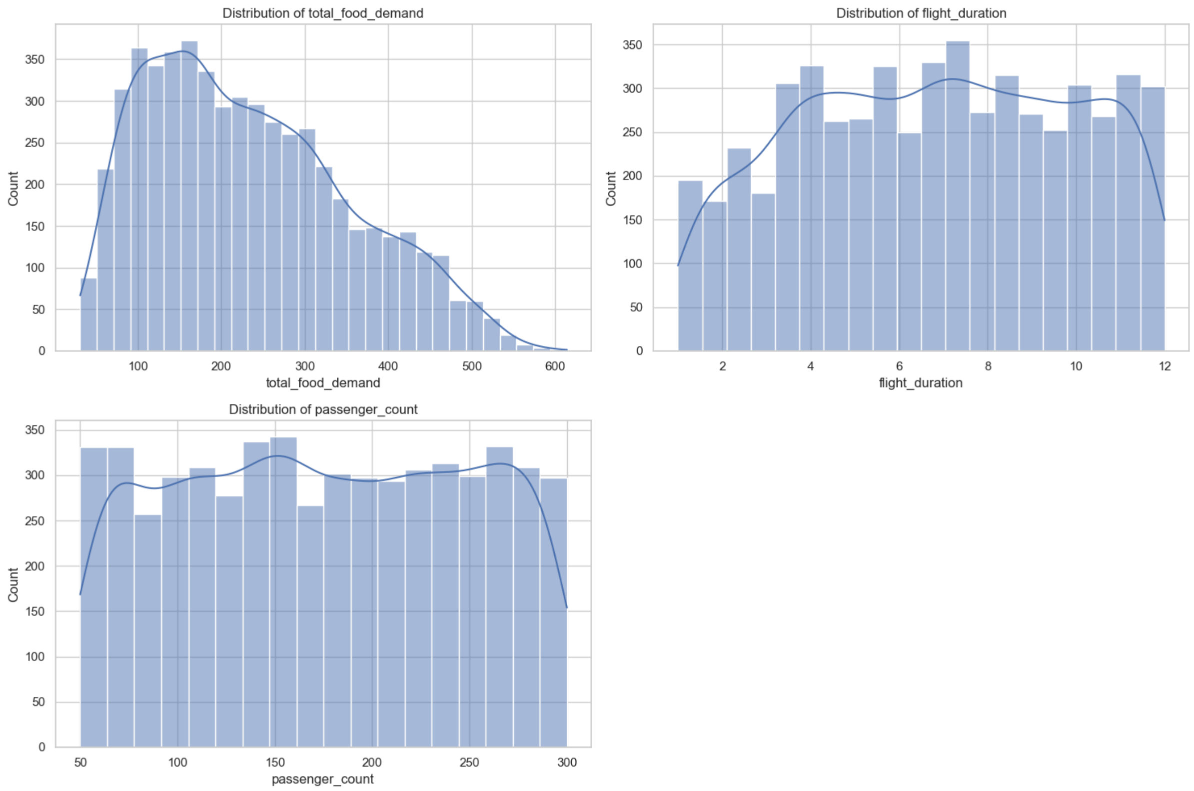
- Volume: 5,500 records (Split 80/20 for Training/Testing)
- International Flights: Minimum 3 hours duration.
- Passenger Composition: Adults + Children = Total Passengers.
- Target Variable: Derived from a weighted formula of duration, class ratios, and flight type, plus noise.

## 3. Exploratory Data Analysis (EDA)

We analyzed the relationships between features to confirm the dataset's validity and understand driving factors. The correlation heatmap below highlights strong dependencies between Passenger Count, Duration, and Food Demand.




We also examined the distribution of our target variable and its relationship with key features.



## 4. Methodology & Models

Three distinct modeling approaches were implemented to establish performance benchmarks:

**1. Baseline Model (Mean Predictor)**

A simple heuristic that predicts the average demand for every flight. This serves as the 'floor' for performance.

**2. Linear Regression**

A parametric model that assumes a linear relationship between input features and food demand. It acts as a good baseline for simple relationships.

**3. Random Forest Regressor (Selected Model)**

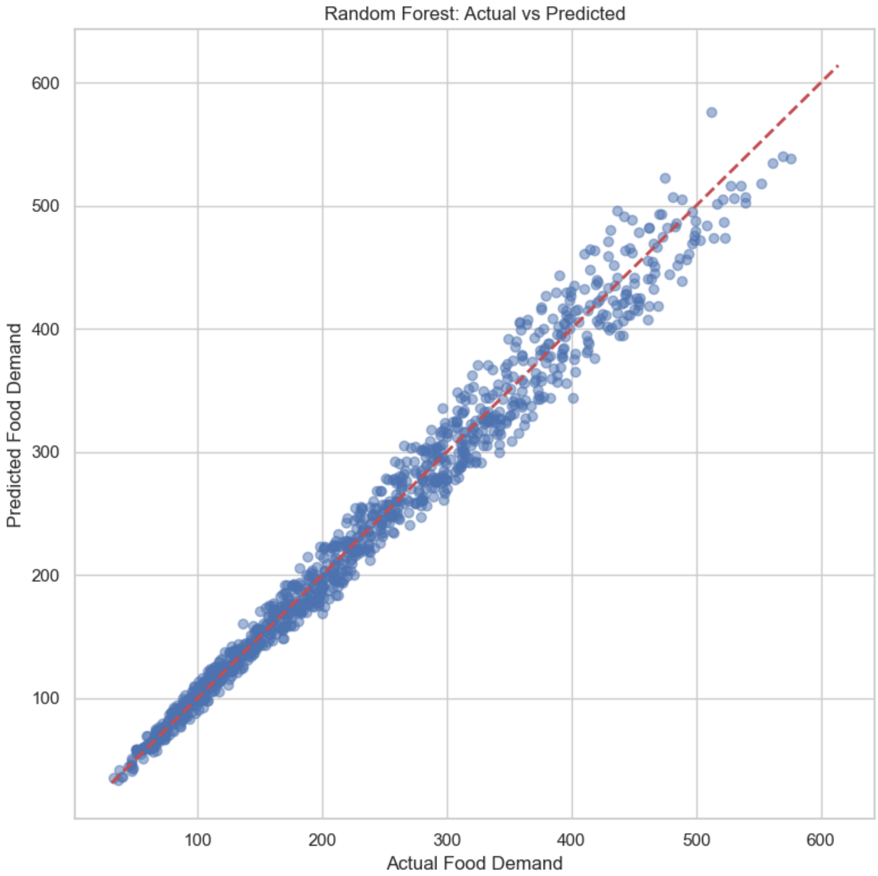
An ensemble learning method using decision trees. It was selected for its ability to capture non-linear interactions (e.g., the jump in food need for flights > 7 hours) and its robustness to outliers.

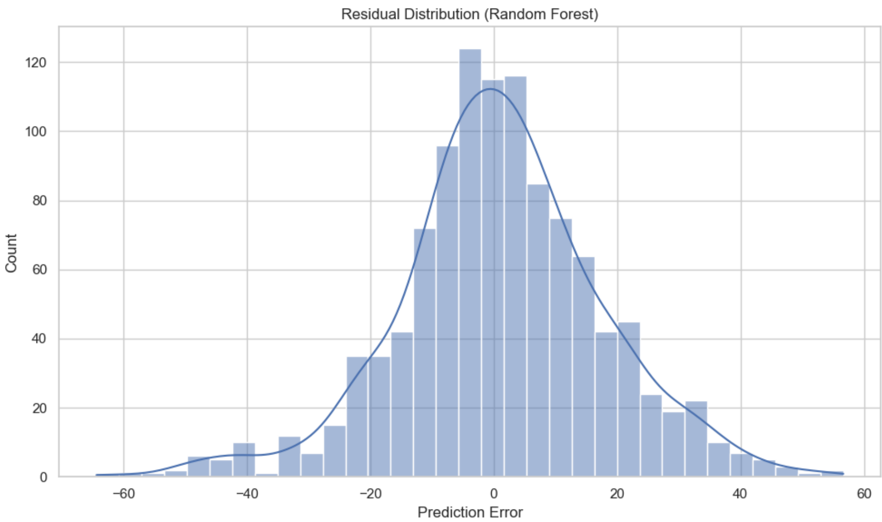
## 5. Results & Evaluation

The Random Forest model demonstrated superior performance across all metrics, significantly reducing the Mean Absolute Error (MAE) compared to the baseline.

| Model | R2 Score | MAE | RMSE |
|---|---|---|---|
| Baseline | -0.0005 | 101.644 | 121.3061 |
| Linear Regression | 0.8949 | 31.1593 | 39.3218 |
| Random Forest | 0.981 | 12.6004 | 16.7216 |

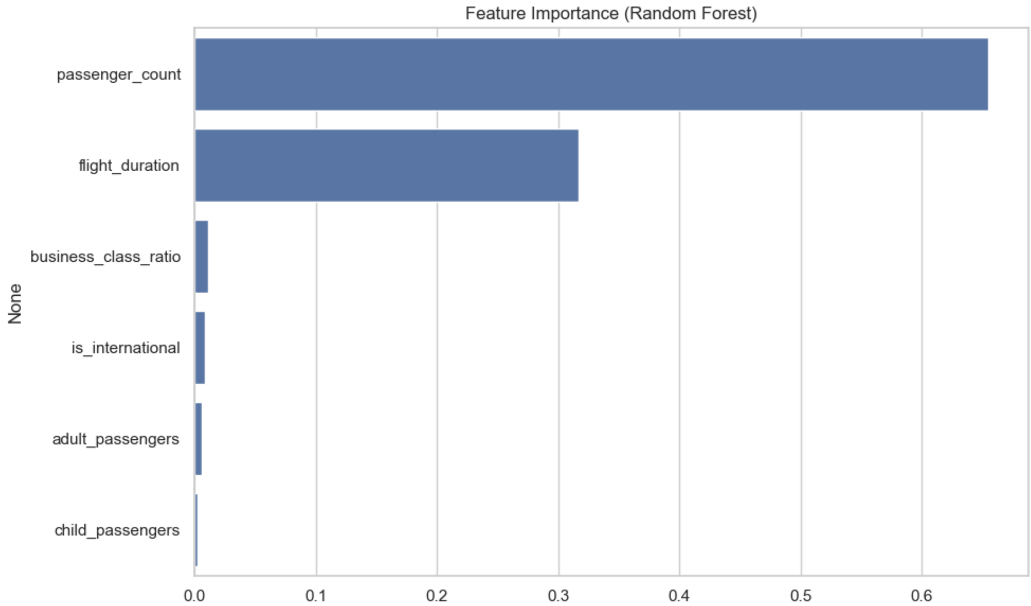
The analysis of residuals confirms that the model's errors are normally distributed and centered around zero, indicating no major systemic bias.


Random Forest: Actual vs Predicted
Predicted Food Demand
Actual Food Demand
100
200
300
400
500
600


Residual Distribution (Random Forest)
Count
Prediction Error
0
20
40
60
80
100
120
−60
−40
−20
0
20
40
60

## 6. Feature Importance Analysis

A key advantage of the Random Forest model is interpretability through feature importance. As shown below, 'Passenger Count' and 'Flight Duration' are the most critical predictors, aligning with domain knowledge.




## 7. Conclusion

The project successfully demonstrated that machine learning can accurately predict airline food demand. The Random Forest model achieved a high R2 score, making it a viable tool for deployment.

Future improvements could include adding more granular data such as passenger dietary preferences, time of day, and seasonal factors to further refine accuracy.