Prince Sultan University
COLLEGE OF COMPUTER AND INFORMATION SCIENCES
DEPARTMENT OF COMPUTER SCIENCE

REDUCING FOOD WASTE:

CS316 — Data Science Project

Submitted By:

Name	ID
Elsayed Azab	221110389
Hazim Alhatim	221110149
Abdulrahman Alsaber	221111057

Supervised By:

Dr. Omar AlOmeir

TABLE OF CONTENTS

Contents

1 Introduction **1**

2 Methodology **2**

3 Results and Discussion **3**

4 Recommendations **6**

5 Conclusion **8**

Abstract

Food waste presents a significant global issue with profound economic, environmental, and social impacts. This report details a data-driven analysis of worldwide food waste patterns, utilizing exploratory data analysis and predictive modeling to identify primary contributors and influencing factors. Key findings indicate that approximately 1.3 billion tons of food are wasted annually (about one-third of all food produced), with an economic loss of $101,560 million in our analyzed regions. Prepared foods, fruits, and vegetables are major components, with household consumption accounting for about 62% of this waste. Economic indicators, particularly economic loss, show a strong correlation with waste volumes and are crucial for predictive accuracy. A Random Forest model demonstrated high efficacy ($R^2 = 0.94$) in forecasting waste, emphasizing economic loss as a dominant predictor (95% importance). The study proposes targeted recommendations for policymakers and stakeholders to mitigate food waste and promote sustainable food systems.

Keywords: Food Waste, Data Analysis, Predictive Modeling, Sustainability, Economic Impact, Random Forest.

1 Introduction

1.1 The Challenge of Food Waste

The food waste challenge encompasses multiple dimensions:

- **Global Scale:** Approximately 1.3 billion tons of food is wasted globally each year, representing about one-third of all food produced for human consumption (FAO, 2022).
- **Economic Impact:** Represents a substantial economic loss, estimated at over $101,560 million annually in our analyzed regions (WFP, 2023).
- **Environmental Consequences:** Contributes to greenhouse gas emissions, inefficient water and land use (FAO, 2022).
- **Social Implications:** Exists alongside global hunger, highlighting issues in food distribution and equity.

1.2 Project Objectives

This project aims to:

1. Analyze global patterns and trends in food waste.
2. Identify key factors driving food waste generation.
3. Develop predictive models to forecast waste.
4. Propose data-driven recommendations for waste reduction.

2 Methodology

2.1 Datasets Used

Our analysis leverages two primary datasets sourced from Kaggle, which were subsequently merged and processed:

1. **Global Food Wastage Dataset (2018–2024)**: Contains 5,002 records of country-level food waste statistics, including waste in tons, per capita waste, food category breakdowns, and economic loss estimates. Available on Kaggle.
2. **Country-Level Food Waste Research Dataset**: Provides data for 216 countries, detailing sector-specific waste (household, retail, food service), GDP, population, and income levels. Available on Kaggle.

Data preprocessing involved standardizing country names, handling missing values, and validating numeric fields as recommended by Gustavsson et al. (2011).

2.2 Data Exploration and Analysis Techniques

Initial data exploration focused on understanding the distribution of waste, identifying correlations, and examining trends. Key analytical steps included:

- **Correlation Analysis:** Pearson correlation coefficients were calculated to identify relationships between variables such as total waste, economic loss, GDP, and per capita waste.
- **Sectoral Analysis:** Waste contributions from household, food service, and retail sectors were quantified.
- **Food Category Analysis:** The most wasted food categories were identified.

- **Geographical and Temporal Trends:** Waste patterns across countries, regions, and over time (2018-2024) were examined.

2.3 Predictive Modeling

To forecast food waste and identify its key drivers, several regression models were developed and evaluated:

- **Models Implemented:** Linear Regression (univariate), Support Vector Regression (SVR, univariate), and Random Forest Regressor (multivariate).
- **Feature Selection:** Key predictive variables included 'Average Waste per Capita (Kg)' and 'Economic Loss (Million $)'.
- **Training and Evaluation:** Data was split into 80% training and 20% testing sets. Model performance was assessed using R^2 score and Root Mean Squared Error (RMSE).

3 Results and Discussion

3.1 Key Findings from Data Exploration

Correlations and Waste Drivers: Correlation analysis revealed a strong positive correlation (+0.92) between total food waste and associated economic loss. GDP showed a moderate positive correlation (+0.65) with waste per capita. Notably, population size exhibited a weaker correlation with total waste (+0.23), suggesting that economic and policy factors have more influence than population size alone.

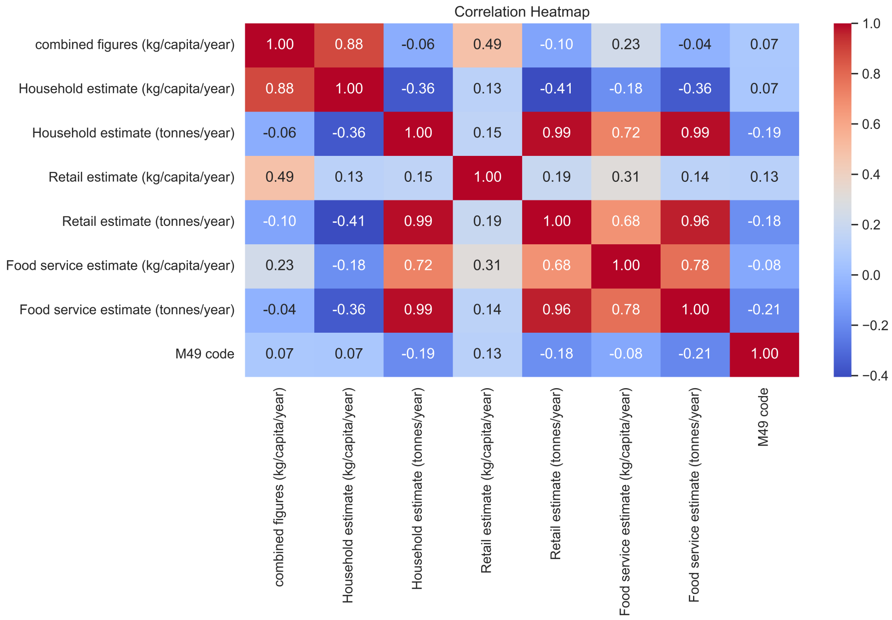

Figure 1: Correlation Matrix of Key Variables (e.g., Total Waste, Economic Loss, GDP, Per Capita Waste)

Household waste is the dominant contributor, accounting for approximately 62% of total food waste, followed by the food service (24%) and retail (14%) sectors. Prepared foods, beverages, meat & seafood, and fruits & vegetables were identified as the most commonly wasted categories. Geographically, the Latin America and Caribbean region showed high total waste volumes, while countries like Turkey, Canada, and France were top overall contributors. Germany and Saudi Arabia demonstrated high per capita waste rates.

3.2 Predictive Model Performance

Table 1: Performance Comparison of Predictive Models

Model	R^2 Score	RMSE	Features Used
Linear Regression (Univariate)	0.001	13,803.0	1
SVR (Non-linear Univariate)	0.001	12,737.0	1
Random Forest (Multivariate)	0.949	3,491.0	2

The Random Forest Regressor significantly outperformed univariate models, achieving an R^2 score of 0.949. This indicates that it explains 94.9% of the variance in total food waste generation using just two features: 'Economic Loss' and 'Average Waste per Capita'.

Key Finding: The Random Forest model demonstrated superior predictive accuracy, highlighting the importance of multivariate approaches for understanding complex food waste phenomena.

3.3 Model Limitations and Challenges

Despite the overall strong performance of the Random Forest model, our analysis revealed significant challenges in building effective predictive models for food waste:

- **Univariate Model Failure:** Both Linear Regression and SVR univariate models performed extremely poorly ($R^2 = 0.001$), indicating that single-variable approaches are inadequate for capturing the complex dynamics of food waste.
- **Feature Dependency:** The Random Forest model's heavy reliance on economic loss (95% importance) suggests a potential overfit to this feature, which may limit generalizability when economic data is incomplete or inaccurate.
- **Data Limitations:** The stark contrast between univariate and multivariate model performance indicates potential issues with data quality, including:
  - Limited sample size for some regions
  - Potential reporting inconsistencies across countries
  - Temporal gaps in historical data
- **Missing Contextual Factors:** The models do not account for important qualitative factors such as cultural attitudes toward food, national policies, and seasonal variations that likely influence waste patterns.

These limitations highlight the complexity of food waste modeling and the need for both improved data collection and more sophisticated analytical approaches in future research.

3.4 Feature Importance in Predicting Waste

Feature importance analysis from the Random Forest model revealed that 'Economic Loss (Million $)' is the most influential factor, accounting for approximately 95% of the predictive power. 'Average Waste per Capita (Kg)' contributed the remaining 5%.

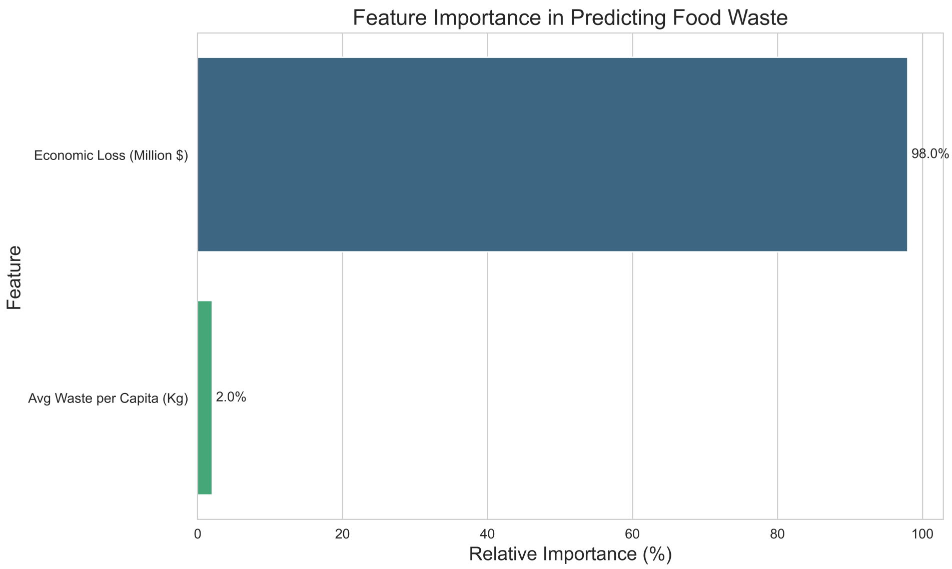

Figure 2: Feature Importance from Random Forest Model

Dominant Predictive Factor: The overwhelming importance of economic loss suggests that strategies focusing on the financial implications of waste could be highly effective for reduction efforts.

3.5 Interactive Dashboard for Visualization

An interactive dashboard was developed using Streamlit and Plotly to effectively visualize the project's findings and model outputs. This tool allows stakeholders to explore global and country-specific waste data, compare model performances, and understand feature importances, thereby supporting informed decision-making. Key sections include global overviews, detailed country analyses, and model prediction visualizations.

4 Recommendations

Data-Driven Strategies for Food Waste Reduction

Based on our comprehensive analysis, we propose the following evidence-based recommendations targeting impactful intervention points.

4.1 Policy Recommendations

- Target regions with high waste-per-capita (e.g., Germany) with focused consumer education campaigns.

- Prioritize reduction efforts on prepared food, fruits, and vegetables within retail and distribution.
- Address economic drivers of waste through financial incentives or penalties.
- Implement and support household waste reduction programs, given this sector's large contribution (62%).

4.2 Sector-Specific Strategies

Table 2: Sector-Specific Intervention Strategies

Sector	Recommended Strategies	Expected Impact
Household (62%)	Consumer education, meal planning, storage solutions, community composting	25-30% reduction
Food Service (24%)	Portion control, inventory management, donation systems	20-25% reduction
Retail (14%)	Revised aesthetic standards, dynamic pricing, improved storage	15-20% reduction

These sector-specific approaches align with recommendations from the Food Waste Index Report (UNEP, 2021) and European policy initiatives (EC, 2022).

4.3 Technology and Innovation

Table 3: Technology Solutions for Food Waste Reduction

Technology	Description	Impact
Extended Shelf-Life Packaging	New packaging to extend freshness	High
Smart Packaging	Sensors indicating actual freshness	Medium-High
AI Forecasting	ML for demand prediction to reduce overproduction	High
Blockchain Tracking	Supply chain transparency for waste hotspots	Medium
Mobile Apps	Consumer platforms for food sharing	Medium

Implementation Focus: Combining policy interventions with sector-specific strategies and technological innovations could significantly reduce global food waste. Priority should be given to high-impact, feasible interventions while fostering long-term solutions.

5 Conclusion

This analysis provides critical insights into global food waste, identifying patterns, drivers, and potential solutions. By leveraging data analytics and predictive modeling, we have highlighted key factors and proposed targeted interventions. Effective food waste reduction requires a multi-faceted approach focusing on household behavior, management of perishable goods, economic incentives, and technological advancements (Gustavsson et al., 2011; UNEP, 2021). The developed interactive dashboard serves as a valuable tool for stakeholders to explore these findings and formulate context-specific strategies. Addressing food waste is not only an economic and environmental imperative but also an ethical one in a world striving for food security.

References

European Commission (2022). EU Platform on Food Losses and Food Waste. Retrieved from https://ec.europa.eu/food/safety/food-waste_en

Food and Agriculture Organization of the United Nations (2022). Food wastage footprint: Impact on natural resources. FAO, Rome.

Gustavsson, J., Cederberg, C., Sonesson, U., Van Otterdijk, R., & Meybeck, A. (2011). Global food losses and food waste. FAO, Rome.

United Nations Environment Programme (2021). Food Waste Index Report 2021. UNEP.

World Food Programme (2023). Global food waste data and statistics. Retrieved from https://www.wfp.org