Unveiling Patterns of Food Waste: Analyzing ZeroWasteEats Data for Sustainable Solutions

Albert Nguyen Nalvi Tran Maria Vega Vazquez

1 INTRODUCTION

World hunger has been on the rise since 2014, and along with the rising problem of waste, it has become a concern for the environment. The problem of food waste has become more evident with each passing year. Approximately, 13% of food is wasted in harvest and retail combined, meanwhile, 17% is wasted among households. Food waste is an issue because not only does the food go uneaten, but valuable resources–land, water, energy, and labor–are being used in vain since the products are getting discarded. Additionally, waste in landfills is a reason why greenhouse gas emissions have been on the rise, consequently contributing to global warming. In turn, global warming affects crop production as the temperature changes have the potential to decrease crop production, cause unexpected water unavailability, and affect the quality of the food grown–all of which can cause scarcity and increase prices of food. Given all the reasons previously mentioned, we aim to identify trends and understand what factors contribute to food waste, explore any potential correlations with factors that could contribute to food waste, and visualize the data.

2 RELATED WORKS

In this section, we will examine how machine learning algorithms have been utilized by researchers to explore different strategies for reducing food waste. A study by X. Wan, J. Li, L. Xie, Z. Wei, J. Wu, Y.W. Tong, X. Wang, Y. He, and J. Zhang [1] explores the use of machine learning algorithms to predict the maturity of compost derived from food waste, aiming for automation in the composting process. Four algorithms—Random Forest, Extreme Gradient Boosting (XGBoost), Light Gradient Boosting Machine (LightGBM), and Multilayer Perceptron (MLP)—were applied to forecast the seed germination index (GI) and carbon/nitrogen (C/N) ratio. The models showed high accuracy, with XGBoost performing exceptionally well. The researchers identified key factors that influence compost maturity, such as composting time, pH level, electrical conductivity (EC), and moisture content. The study further developed a prediction application to guide composting processes based on the models' outputs. The study found that when the pH and EC are not in the optimal range, it is recommended to increase pH and decrease EC, as well as keep the place well-ventilated. Although GI predictions were quite accurate, C/N ratio forecasts faced challenges due to limited data and model constraints, highlighting the need for broader research parameters to enhance prediction accuracy. Nonetheless, the research highlights the potential of machine learning in enhancing the management of food waste composting by enabling early warning and process regulation, culminating in the development of a prediction application for automated reactor composting systems.

Another study by M. Rodrigues, V. Miguéis, S. Freitas, and T. Machado [2] investigates the application of machine learning models for short-term demand forecasting in food catering services to reduce food waste. There are four models for predicting food demand, categorized into two groups. The first group includes two causal models: one utilizes the Random Forest algorithm, and the other employs the Light Gradient Boosting Machine (LightGBM) algorithm. The second group consists of two-time series models: one is based on the Long Short-Term Memory (LSTM) neural network, and the other uses a transformer neural network. The researchers collected data that included the historical demand by dish type, the daily number of students attending classes, weather-related variables (temperature and precipitation), and date-related variables such as day of the week and month. For some food catering services, additional data like menu classifications and reservations were also utilized. This variety allowed for a comprehensive approach to forecasting, taking into account various factors that could affect food demand in catering settings. The results showed that the models based on the Random Forest and LSTM neural network algorithms significantly outperformed traditional forecasting methods, reducing wasted meals by 14% to 52% and decreasing unmet demand by 3% to 16%. However, there are limitations within the study. The study focuses on specific catering settings that might not represent all food services, leading to potentially non-generalizable results. Additionally, it only considered the number of dishes served, not the quantity of food, which may not accurately suggest preparation amounts. The evaluation relied on baseline models that might not reflect actual forecasting practices due to the lack of real forecast data. Lastly, the accuracy and comprehensiveness of the collected data could impact the models' effectiveness, and the study did not account for individual customer preferences or plate waste. Despite its limitations, this approach demonstrates the potential of machine learning to improve food catering operations, enhancing service levels while mitigating food waste's environmental, social, and economic impacts.

3 DATA SET

Our analysis will use the dataset provided by the ML Olympiad ZeroWasteEats, which encompasses a variety of attributes including the country, the year of data collection, percentages and quantities of food loss, the specific activities associated with the food, and the stage of the food supply chain. We intend to use various data mining techniques to explore this dataset. By comparing the outcomes derived from different algorithms, our goal is to uncover varying insights and discern patterns that may not be immediately apparent. This comparative approach will allow us to evaluate the effectiveness of each algorithm in revealing correlations and trends within the data, ultimately providing a better understanding of the dynamics of food waste across different contexts concerning the different stages of the food supply chain. Through this analytical process, we aim to identify potential strategies for reducing food waste, thereby contributing to more sustainable food systems globally.

4 DATA PRE-PROCESSING

In preparing the ZeroWasteEats dataset for analysis, we first eliminated columns not relevant to our study's goals. To test various models, we utilized Scikit-learn's pipeline modules. This enabled us to efficiently test various predictive models by facilitating a seamless integration of data. Then, the dataset was processed for standardization–both numerical and categorical data were identified and handled separately; numerical columns were scaled for normalization, while categorical columns were transformed through one-hot encoding to convert categorical data into a format that could be effectively utilized by the algorithms. Subsequent steps involved merging the dataset with additional country-specific data, followed by a careful elimination of columns that were either redundant or not critical to our analysis, such as 'm49_code' and specific data collection method details. Furthermore, we ensured the cleanliness and integrity of our dataset by discarding any rows that lacked information on 'food_supply_stage', guaranteeing that our dataset was optimized for conducting a comprehensive analysis on food waste patterns and identifying potential reduction strategies.

5 ALGORITHMS

We have run four algorithms on our dataset: Log Regression, K-Nearest Neighbors, Random Forest Classifier, and Gradient Boosting Classifier. The models processed through the standardized pipeline yielded distinct results: Logistic Regression led with an impressive 81.72% accuracy, suggesting its strong fit for the data's linear characteristics. The Random Forest Classifier, while robust in handling various types of data, achieved a moderate 44.30% accuracy, which could indicate that the decision boundaries for this dataset are complex. Gradient Boosting Classifier surprisingly scored a low 0.81%, hinting at a possible mismatch in algorithm capability with the data's structure. Meanwhile, the K-Nearest Neighbors Classifier showed a promising result with 83.56% accuracy, showcasing its ability to effectively classify the dataset based on proximity to its nearest data points.

Note: We will write more in this section later.

6 IMPORTANCE

Food security remains a prolonged challenge for various regions worldwide, with food waste significantly exacerbating this issue. The consequences of food waste extend beyond just reducing food accessibility; they also influence food prices adversely. Additionally, as food waste escalates, the sustainability of our food systems diminishes, impacting critical resources like labor, water, land, and capital. Moreover, the disposal of unused food contributes to the accumulation of landfill waste and the production of greenhouse gases. These factors underscore the urgent need to tackle food waste and develop effective strategies to mitigate its impact.

7 RESULTS & CONCLUSION

Note: This section will be written after all algorithms are analyzed and compared.

8 UPCOMING WORK TO BE DONE

We are going to review the results that have yielded from the algorithms that we have applied so far Then, decide if there are any other techniques we can implement. Additionally, we need to get the feature importance and visualize the graphs to showcase the accuracy of our models across different algorithms. There will be charts and graphs added to the paper for further clarification on our dataset, and the attributes we are analyzing.

REFERENCES

[1] Xin Wan, Jie Li, Li Xie, Zimin Wei, Junqiu Wu, Yen Wah Tong, Xiaonan Wang, Yiliang He, and Jingxin Zhang. 2022. Machine learning framework for intelligent prediction of compost maturity towards automation of food waste composting system, Vol. 365. Bioresource Technology. https://doi.org/10.1016/j.biortech.2022.128107.

[2] Miguel Rodrigues, Vera Miguéis, Susana Freitas, and Telmo Machado. 2024. Machine learning models for short-term demand forecasting in food catering services: A solution to reduce food waste, Vol. 435. Journal of Cleaner Production. https://doi.org/10.1016/j.jclepro.2023.140265.