Reduction and Rescue of food waste in retail environments using data driven decisions

Srivasa Pranathy Mutcherla, *MSCS Graduate Student, Illinois Institute of Technology,*
Weslynne Ashton, *Professor, Stuart School of Business,* and Cynthia Hood, *Professor, College of Computing*

***Abstract*—Consumer-facing food businesses account for approximately 20% of food wasted in the US, with retailers specifically accounting for 5%. However, there is minimal public information about food waste generated from retail stores (e.g., grocery stores, department stores, supermarkets). A better understanding of what food is wasted in retail environments can create opportunities for redirecting these losses into beneficial reuse to feed humans and animals or be recycled into soil.**

This project aims to research different strategies being developed to address the limitations in data availability regarding food waste generation in retail environments. It also aims to inform the creation of a [machine learning data] model to generate insights for reducing, rescuing, and recycling wasted food in retail environments. Based on crucial food characteristics, such as perishability (shelf life) and edibility, we could map how different types of wasted food can be directed to food rescue (e.g., food banks, pantries, mutual aid), animal feed, or recycling (e.g., composting, anaerobic digestion).

***Index Terms*—machine learning, food waste, retail, shelf life, rescue, reduction.**

I. Introduction

WITH hunger levels at a rate of 47 million [1] people in the US, improving how food waste is generated and handled is imperative. In this context, it is necessary to highlight that food waste refers to consumable and edible food. This project focuses on the retail sector and grocery stores in particular. Considering that 4.45 million tons of surplus food was generated in 2023 alone, and an alarming 30.2% of it ended up in landfills, there is a vast scope for technology to work on reducing this. Machine learning algorithms can be used to understand the current trends in waste production and predict future numbers, along with solutions that can be implemented to reduce this. ReFed is one of the primary resources used here, and its database shares numbers on waste generated and solutions that can be implemented. One of the primary components of this project work is to use machine learning models to help bridge the gap between the surplus food generated and the solutions that can be invested.

II. Literature review

Along with research papers, human interviews were conducted to understand more about the food space. In a meeting with Prof Ashton and PhD student Chaoqun, it was highlighted that the Greater Chicago Food Depository was working on connecting retailers to food pantries. Their current model involved picking up food from retailers, storing it, and distributing it to food banks as needed. Some questions to ponder include if there is any incentive for retailers who donate food, if there is any data repository of sorts that keeps track of such measures by retailers, and if these do need to be presented to authorities like the state, or IRS, or GCFD for tax-based incentives. In a city like Chicago, transportation is not a significant barrier; it is logistics.

Some more interesting statistics include Chicago generates 450,000 tonnes per year, half of which comes from households. While the city is responsible for single-family to 4-unit houses, private management is responsible for condos and buildings. The problem remains that neither of them has compost support. Similarly, waste characterization is a work in progress, and to find multiple states following similar types is not possible. There are various teams working on the data modeling aspect, which include Ohio State University Professor Brian Roe, American University's RECIPES team, the ReFed team, and the Greater Chicago Food Depository.

Some parameters that could flag waste include generation quantity, quality, composition, and spatial location [2]. Their work also defines food loss as edible unconsumed food lost throughout the supply chain. It defines recoverable food as discarded or surplus safe for human consumption [3].

One of the many proposals included decentralized community composting for residential food waste [4]. Many insightful points were brought to the discussion as there are multiple constraints and points to be considered for such a solution. Some advantages include reduced travel time and ecological and economic benefits. The constraints also included data availability for such planning, complex factors like income levels, and access for everyone, with the need for benefits to be equally distributed. One of the key takeaways is how beneficial such measures can be and how this can be extended to the retail sector.

A key piece of literature reviewed in this section concerns how food bans have impacted food waste generation in different states [5]. The underlying algorithm was a synthetic control method, which helps understand a pattern's cause. The various benefits of these bans that policymakers strongly believe in include using the space for a better purpose, food rescue, and initiatives like composting to grow. Their dataset included working with Massachusetts and Vermont, focusing on restaurants, grocery stores, and other establishments. Their work is another reminder of how complex the data acquisition process is. It was done by correspondence with 43 state agencies, 12 public record requests, and much more preprocessing to get usable datasets. This preprocessing was also needed to help correlate the formats and values from the different state repositories. Critical factors like industrial

waste, municipal solid waste, and waste from imports and exports were not uniformly present. Once the dataset was prepared, the model that was generated tried to present what would have been the statistical outcome in the absence of the ban. This work helped understand that the outcomes would have been the same irrespective of the ban, except in Massachusetts. Their success was due to affordable execution, understandable parameters, and increased penalties.
Multiple organizations are working to tackle this issue on a corporate scale. Haystack Data, for example, works with clients to build solutions that involve database establishment, data analysis using machine learning algorithms, insight generation using dashboards, data architecture setup, and process automation. Another team is Crisp. Their product uses raw data from retailer and distributor portals and organizes it for storage, modeling, and reporting.

In this process, multiple conversations with various stakeholders in this industry were also organized. Spatial analysis was a potential way forward to understand the mapping of food from retail stores to food banks, which would greatly benefit. Some concerns with ReFed data were that it was at a coarse grain level. However, with some exploratory data analysis, it might be possible to map it with other data sources like the census, which could help explain people's behavioral patterns towards food waste as a community.
The executive head chef at Commons, IIT, Mark Angeles, had valuable insights about how their kitchen was run efficiently to minimize food waste. Waste is classified into green waste - rinds and peels of fruits and vegetables that go into composting and red waste - food that cannot be recovered or rescued. Food is ordered in a scheduled manner to accommodate needs while reducing waste. Food not yet served is served again in appropriate sections based on shelf life. The number of pieces served on trays is also optimized for better distribution. One of the key takeaways is that plate waste is the highest contributor, which could be caused by various factors like demographics, upbringing, relationship with food growing up, et cetera.
A conversation with librarian Sean Murphy from the Paul V. Galvin Library helped identify the different avenues for finding information. Chicago Botanical Gardens, UNICEF data portal, data.gov, and data.census.gov were all shared.
An Illinois Tech-level RECIPES team meeting helped understand Azra Sungu's work on narrative and storytelling in food waste, a PhD graduate from the Institute of Design. It focused on going beyond the data problem and culture, relations, and mindsets as a qualitative work. It was tackled by creating exhibits that aimed to engage more retailers about what made a better food future by creating artifacts from the perspectives of different stakeholders in this equation.
A team shared their work on the "Too Good to Go" project in the American University's RECIPES meeting. Issues such as people getting unrealistic and unfeasible quantities of food, such as a bag of 11 egg salads, were discussed, which transfers the waste responsibility from the store to the end user. Another team was conducting user surveys on how customers perceive packaging, recycling instructions, and disposal. Prof Brian Roe's work was similar to the focus of this project as it was about "Data on grocery store chains and where consumers would prefer food waste reduction messaging (joint with industry)."

III. Machine Learning Model

After considering various ideologies, perspectives, and thought processes, it was not straightforward to make a computational choice. Various variables were involved, and the main missing point was the relevant datasets. After a discussion with Professor Shouvik Roy at the Illinois Institute of Technology, the decision was made to look into Reinforcement Learning Models. The Proximal Policy Optimization model was selected as a place to begin with, as it worked with minimal changes to the policy, which avoided huge losses while going in the right direction of the gradient. A fundamental version of the model aimed at loading the data into frames and establishing the action space, observation space, initial state, and the steps to move forward. The insights engine from ReFed was the data source, where the food waste data was used to chart the current waste scenario and predict future statistics. The food waste solutions data was used to map the options to reduce and rescue food waste while promising financial benefits and reducing CO2 emissions and water consumption. After facing major time and CPU computation issues, the reward function remained the same without any growth. This led to reworking the exploratory data analysis phase, which helped identify specific vital differences. The datasets do not have matching formats. The in-depth nature of the solutions dataset and the high-level nature of the food waste dataset cause a gap in the optimal solutions that can be implemented and their impact. The following steps were taken to update the datasets. The data frames were filtered to concentrate on the retail sector. The columns designated for aggregation are specified. Non-essential columns, such as 'sub-sector' and 'sub-sector category,' have been removed as they do not influence the results. The primary columns identified are utilized to perform the group by function. In the solutions dataset, these columns include solution group, solution priority action area, solution name, sector, and food type. The relevant columns in the food waste dataset are year, sector, and food type. After the data preprocessing was completed, the dataset became much more compact. However, this ensured the possibility of finding one or more solutions mapped to a food waste scenario to reduce and rescue the waste from ending up in landfills. During the training phase, different actions were chosen, but during the testing phase, the same action was selected. This led to a rewards graph that did not show any growth in any direction. After working with the hyperparameters and defining the evaluation to be non-deterministic, it was possible to see a change in rewards. Apart from that, exploration vs. exploitation in reinforcement learning is a tradeoff. To truly optimize the solution, the action space was also updated to work dynamically. More understanding of how the action space was defined and how it could be filtered to match the food type of the waste being considered was needed.

IV. Conclusion

This course was an enriching experience, as it allowed for exploration beyond data. With the already available numbers, it is clear that improving solutions at a faster pace is necessary. Machine learning proves to be a viable solution for identifying investment options that tackle the problem of multifold reduction of waste, emissions, and water consumption, including rescue. Various stakeholder interviews also helped bring perspectives into the picture that can help further customize the product to cater to everyone's needs. The pressing issue remains the lack of a usable dataset that could be worked upon to be scalable throughout various states and sectors. Collaborating between retailers, researchers, and policymakers is essential to formulating effective strategies that blend practical implementation with community engagement.

References

[1] Feeding America, "Food Waste in America: How You Can Help Rescue Food," Feeding America. [Online]. Available: https://www.feedingamerica.org/our-work/reduce-food-waste.

[2] N. Ai and J. Zheng, "Community-based food waste modeling and planning framework for urban regions", Journal of Agriculture, Food Systems, and Community Development, vol. 9, no. 1, pp. 39–58, 2019, doi: 10.5304/jafscd.2019.091.009.

[3] N. Ai, J. Zheng, "Chicago Food Waste Management", https://public.tableau.com/app/profile/junjun.zheng/viz/FoodWasteChicago_v031/Gens_Dash

[4] S. Pai, N. Ai, and J. Zheng, "Decentralized community composting feasibility analysis for residential food waste: A Chicago case study," Sustainable Cities and Society, vol. 50, 2019, Art. no. 101683, doi: 10.1016/j.scs.2019.101683.

[5] F. Z. Anglou, R. E. Sanders, and I. Stamatopoulos, "Of the first five US states with food waste bans, Massachusetts alone has reduced landfill waste," Science, vol. 385, pp. 1236-1240, 2024, doi: 10.1126/science.adn4216

[6] T. Simonini, Proximal Policy Optimization (PPO), https://huggingface.co/blog/deep-rl-ppo