FOODLIGENCE – PREDICTING EXPIRY DATE OF PERISHABLE FOODS TO REDUCE LOSS AND WASTE

Trewon Weerasooriya and Kishore Kumar

Faculty of Computing, Sri Lanka Institute of Information Technology Malabe, Sri Lanka

ABSTRACT

"Food loss and waste" is a growing issue for the environment, the economy, and society worldwide. It has an adverse effect on social, environmental, and economic issues. Poor planning, excessive production, and customer perception are the main causes of food loss and waste. Inefficient use of perishables before to expiration is another significant contributor to food loss and waste. The study primarily concentrates on Sri Lanka's hotels and restaurants, one of the two main sub-sectors of the hospitality and food services business. To reduce food loss and waste during the processes of the food supply chain of hotels and restaurants, the research suggests an expiry predictor for perishable food items using artificial intelligence. It also suggests a donation platform to distribute any surplus or unconsumed perishables to the needy/beneficiaries. The predictor enables hotels and restaurants to be aware of the best dates to use the perishables they have purchased, while if any are left over or spoil, they can be listed using the freshness index so that people in need can request them and make purchases for a price range the hotels and restaurants have specified. The solution aims to reduce waste in the food supply chain and control it to the greatest extent feasible.

KEYWORDS

Artificial Intelligence, Machine Learning,Random Forest, XGBoost, Food Expiry

1. INTRODUCTION

In terms of money, it is estimated that 1.3 billion metric tons of food are lost or wasted annually, costing the world economy $1 trillion [1]. According to an examination of data gathered from the three main disposal stations in Colombo, Sri Lanka wastes about 3963 metric tons of food every day [2]. In the context of this study, "food loss" refers to the loss of food along the supply chain during the various stages of cultivation, harvest, storage, processing, and transportation[3].

Reducing food loss and waste, notably in Sri Lankan hotels and restaurants, is the primary objective of the research. According to a study conducted in Sri Lanka by the Food and Agriculture Organization of the United Nations, it was discovered that out of all the major entities/organizations, a single hotel/restaurant generates waste at a rate of 2.4 tons weekly, compared to 0.79 tons and 0.44 tons at wholesale markets and supermarkets, respectively [4] as shown in Figure 1. This study is the basis for the choice of hotels and restaurants.

DOI:10.5121/cseij.2022.12615

Fig.1. Food waste chart

The main causes of food loss and waste include inadequate planning for purchases, buying too much food, and failing to use food before it expires [3]. By addressing the issue of "failing to consume perishables before it expires," the research intends to reduce or control food loss and waste in the food supply chains at hotels and restaurants. It is imperative to lower the present rate of food loss or waste since it endangers both humans and the environment. In addition to the financial strain it puts on hotels, restaurants, and the national and international economies, it also has significant effects on environmental sustainability. Food security, financial losses, and climate change are a few of the top concerns.

To reduce food loss and waste and improve hotel food supply chain procedures, a method is suggested in the research literature. The feature provides a platform for donations of excess or food that is about to spoil as well as a food expiry forecast for perishables. To create the solution, the research that was utilized to pinpoint the requirements was paired with an analysis of earlier research in the area. Foodligence focuses on supporting any company or body that is involved in food production, storage, and consumption because the solution can be constructed as a component-based through an API.

2. RELATED WORK

A machine learning technique that predicts which stock will degrade first was created by the Computer Engineering Department at PICT in Pune, Maharashtra, India, and published in the paper "Analysis of Post-Harvest Losses: An Internet of Things and Machine Learning Approach" [5]. The Authors Instead, than using the first in, first out strategy, it gives warehouse managers the information they need to create a dispatch sequence. It also talks about the factors that influence food deterioration. The number of days the stock has been in storage, the temperature and humidity of the warehouse, and the harvesting time's divergence from the ideal harvesting period (measured in days) have all been proven to affect the amount of storage loss [5]. This study employed the Elastic Net Regression model to forecast food loss in the stock and then rearranged the stock's dispatch order to reduce food loss. The ElasticNet algorithm has not been compared to other machine learning algorithms to ensure that it is the approach that yields the best results for the dataset being utilized.

A platform was created to link food donors with the community in research on food waste titled "FoodX, a System to Reduce Food Waste" [6]. The system needs input from the donors in order to distribute the food that is to be provided [6]. The issues of food waste and starvation are both overcome with this method. The method merely seeks to link populations in need with food givers; it does not consider any alternative solutions.

Once Tan Jun Yuan, a hawker at a food stand in Singapore, observed how much food is wasted annually, he felt bad. As a result, he developed 11th hour to address this issue. A mobile app called 11th hour enables food and beverage (F&B) businesses to offer customers last-minute deals [7]. This application offers leftover and discarded food at half the original price before the restaurants close [7]. More than 100 businesses joined in less than a year, and by the end of that time, 5,000 users had downloaded and utilized the app [7].

In parallel, N. Kumar S.V., S. Balasubramaniam, S. Tharagesh R.S., P. Kumar, and B. Janavi conducted research for their article "An Autonomous Food Wastage Control Warehouse: Distributed Ledger and Machine Learning based Approach" [8]. To anticipate when perishable food would expire, a neural network model that considers a number of environmental factors has been created [8]. As a result, warehouses will be able to sell or distribute batches of goods that are anticipated to have a brief shelf life first [8]. However, neither recommendations nor solutions are offered, nor are there any alternative automatic activities based on the outcomes of the AI model. Even though the suggested method helps to determine how the food will most likely be wasted, the actions still require human engagement. The dataset that was utilized to train the model was obtained from the internet, and it also included manually generated data. A minimum training error of 0.4268[8] was attained by them.

3. METHODOLOGY

Prediction of Expiry Date

Data Collection

Data collection is the most important machine learning component. The success of the project is solely dependent on the dataset's quality. The research used data provided by Cool Beach Hotel Hikkaduwa as well as an open Egg Plant (Brinjal) dataset that is freely available online [9] to examine and apply the model. Additional synthetic data were generated to avoid overfitting the model.

Data Pre-processing

The processing methods increase accuracy while lowering dataset complexity. Due to the inconsistent, noisy, and partial nature of the used data. Using data leaning/cleansing processes, the author tried to complete missing values, smooth out noise while identifying outliers, and rectify conflicts in the data.

Training of models

The ignorance of an item's expiration date is one of the main issues we have with perishable food. And most of the time it goes unrecognized when it does. The state of the food items is therefore unknown to the storage staff until after they have expired. The application's primary focus is on predicting the expiration date, and there are several regression algorithms.

The parameters that are used to predict the expiry date of a perishable food are as follows:

- Food Type
- Days since Harvested
- Temperature of the warehouse/facility
- Humidity of the warehouse/facility
- Deflection from ideal harvest date

Instead of employing only one method, the author has chosen to use two algorithms. As a result, it offered a chance to assess and exhibit the regression algorithms' performances alongside a comparison of the two methods.

These are the two chosen algorithms:

- Random Forest Regression.
- XGBoost Regression

The hyperparameters of these two models are tweaked for improved performance after they have been trained and evaluated. The best performing model is then selected, and it is made available via a web API. On a flask server, the web API for this will be launched. The solution's summary is shown in Fig 2.

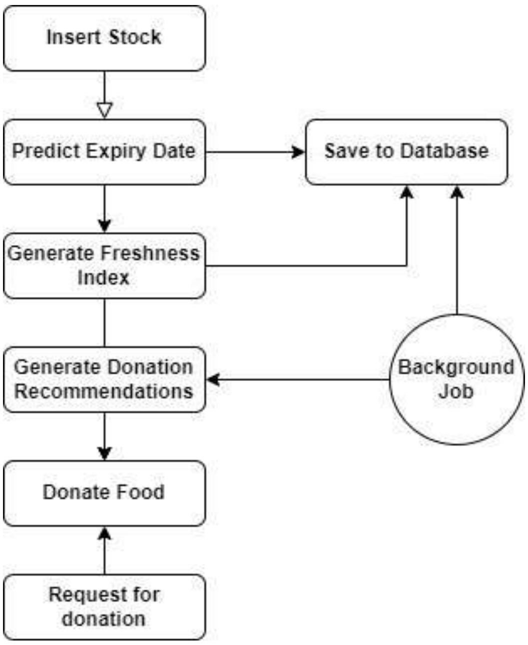

Fig.2. Solution Overview

Donation *Recommendation*

The mere anticipation of food waste does not imply that a solution has been found. A form of intervention is required to stop food loss. As a result, the system will alert the user about which products should be donated based on anticipated expiry dates.

Food* Donation *Platform

Despite the steps and solutions used, food loss and waste will still wind up in landfills, but with the help of the aforementioned provider of food donation advice, perishables that are about to expire can be donated to the people in need. The hotel might agree to the system's suggestions for food donations after which the information about the food donations is made available on a public website. Anyone who is in need of food only needs to check in to the website, which will show all donated food items, and request any of the things that are listed there. In other words, it will create a platform for coordinating the distribution of food to communities in need. As seen in Figure 3, the donations will be presented on a public website.

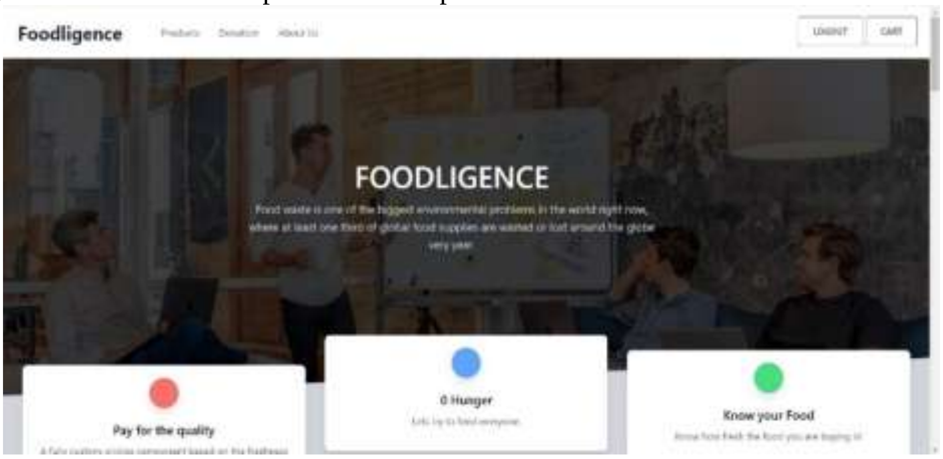

Fig.3. Donation Platform

4. Results And Discussion

A. Expiry *date* Predictor

Among all the features (independent variables) used to train the selected models, the relative importance of each feature is referred to as the feature importance. Which features are more important during model training is revealed by the concept of feature significance. In some instances, training a model just on these features will produce better outcomes. Therefore, the project team initially assessed the worth of the learned traits. Figure 4 provides an illustration of the significance of the selected features. Only one type of food was present in the dataset, hence typeoffood had a significance of 0.

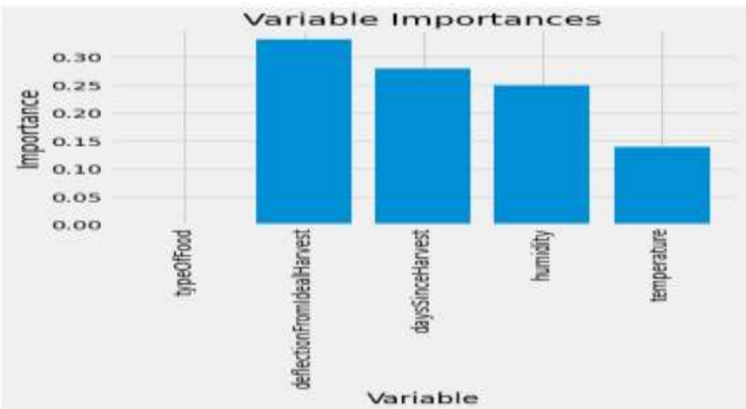

Fig.4. Feature importance

Two machine learning algorithms are developed, trained, and evaluated using the previously described open datasets that are accessible online. Two regression techniques, XGBoost and

Random Forest, are used to assess the dataset. Each instance's errors are computed and then added. The best performing model is then selected. The two main metrics we'll be using to evaluate the models are R Square and Mean Square Error. A reliable indicator of how successfully a match is made is the Mean Square Error. A statistical indicator of how closely the data resemble the fitted regression points is called R-Squared. The word "coefficient determination" is another name for it.

1) Random Forest Regression:

The ensemble learning approach is used by Random Forest Regression. A Decision Tree is a fundamental model, and a Random Forest model's output is the total of all Decision Trees' outputs. A technique for merging many models to enhance forecast accuracy is called model ensembling. In Random Forests, each base model is independently constructed using a different subsample of data[10]. The categorical features are where the random forest model shines. As a result, the dataset worked well with this model. The aforementioned dataset was then used to train a Random Forest model.

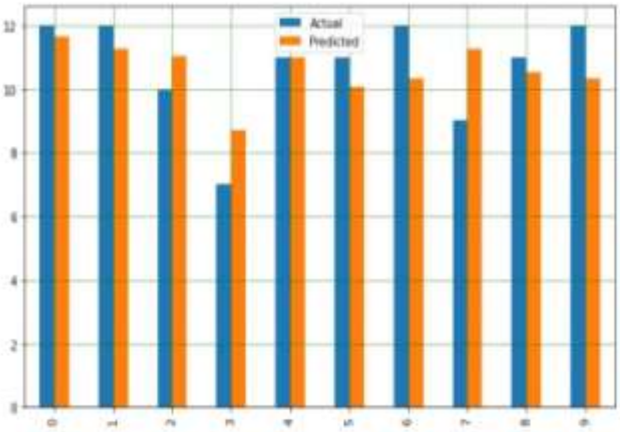

Fig.5. Random Forest Performance

In Fig. 5, the model's predicted values are displayed next to the actual values. The blue line displays the actual data from the inventory, while the orange line shows the values that the model predicts. As seen in Fig. 5, the model performs brilliantly when applied to the test data.

Dataset	R Square	MSE	MAE
Test Data	0.36088	1.5402	1.0656
Train Data	0.5110	1.3527	0.8774

Table 1. Random Forest Performance Metrics

Based on the measures that have selected, Table 1 shows how well the trained models have performed. The root mean squared in the image above is 1.2471732. This indicates that the model was accurate in predicting the expiration date to within 1.2 days.

2) XGBoost Regression*:*

A distributed, scalable gradient-boosted decision tree (GBDT) machine learning framework is called Extreme Gradient Boosting (XGBoost). One of the best machine learning models for regression, it provides parallel tree boosting. Similar to Random Forest, XGBoost also performed well when applied to the test data set.

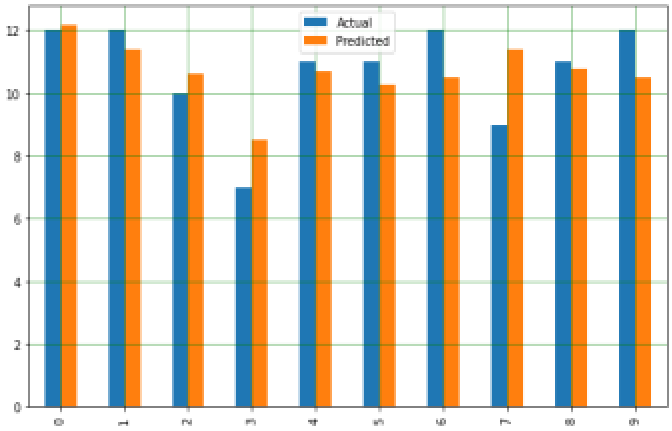

Fig.6. XGBoost Performance

The blue line in Fig. 6 represents the inventory's actual data, and the orange line represents the model's anticipated values. In this instance, the Xgboost model outperforms the previous model by a wide margin. The model has obviously performed well.

Dataset	**R Square**	**RMSE**	**MAE**
Test Data	0.4250	1.3855	0.9540
Train Data	0.5898	1.1348	0.7824

Table 2. Performance metrics of XGBoost

The RMSE of the XGBoost model versus the test dataset is 1.3855, as shown in Table 2. it is superior to the previous model in comparison.

3) **Hyper Parameter tuning**:

Given that the models' performance wasn't as strong as anticipated. The models' hyperparameters will be adjusted, it was decided. Hyperparameter tuning is the process of determining which set of hyperparameters will work best for a learning algorithm. Before learning ever begins, a hyperparameter's value is predetermined.

The performance of the XGBoost model's hyper parameters, including max depth, booster, nthread, learning rate, and min split loss, was then assessed. The aforementioned hyperparameters were then slightly altered as the results were being watched. The performance metrics of the best hyper parameters used for XGBoost are shown in Figure 12.

- n_estimators: 100
- max_depth: 3
- learning_rate: 0.1
- colsample_bytree: 0.3

Dataset	R Square	MSE	MAE
Xgboost	0.4250	1.3855	0.9540
Tuned Xgboost	0.4985	1.2084	0.9337
Improvement	0.0735	0.1771	0.0203

Table 3. Comparison of performance of XGBoost after tuning

After analyzing table 3, it can be concluded that the model's performance and accuracy slightly improved as a result of fine-tuning its hyper-parameters.

The performance of the Random Forest model's hyperparameters, including n-estimators, min samples split, min samples leaf max features, and max dept, was then assessed in order to select the optimal settings. Following several tests, the following hyper-parameters were determined to be the best.

- n estimators - 200.
- min samples split - 5
- min samples leaf - 1
- max depth – 70
- max features - sqrt

Dataset	R Square	MSE	MAE
Random Forest Model	0.36088	1.5402	1.0656
Tuned Random Forest model	0.6825	0.7651	0.7336
Improvement	0.3216	0.7751	0.332

Table 4. Comparison of performance of RF after tuning

As seen in Table 4, adjusting the Random Forest model's hyper-parameters resulted in a noticeable improvement in performance. The mean squared decreased to a minimum of 0.68. This suggests that the standard error of the model's predictions for expiration dates is less than a day.

Model	R Square	MSE	MAE
Tuned Xgboost	0.4985	1.2084	0.9337
Tuned Random Forest model	0.6825	0.7651	0.7336

Table 5. Comparison of Models

The author can infer from an analysis of Table 5 that the tweaked random forest model performed vastly better than all other models. Thus, the adjusted random forest is the ideal model for this study, it may be argued. And a mean squared error value of 0.7651 was attained. This suggests that the standard error of the model's predictions for expiration dates is less than a day.

B. Public API To Access Model

In accordance with the advice given by the industry experts, the author chose to expose the model using a flask web server. This API can be used to include this creative solution into programs like PickMe and Ubereats. The API spec is represented graphically in Figure 7.

Fig.7. Open API V3 spec for Prediction API

5. CONCLUSION & FUTURE WORK

According to the research, food loss and waste is a serious problem that hotels and restaurants are trying to address. They attempted to resolve the issue with already known solutions, but they weren't as successful. This research's major objective was to predict when perishable goods will go bad and make suggestions for preventing food waste. The best model had to be chosen, the predictor had to be put in place, the donation platform had to be made, and the provider of contribution recommendations had to be created. Before starting the actual component training operations, this research undertook several testing techniques to accurately identify the tools and technologies. Foodligence was created and developed in a way that it may be used, with only modest customization, in any sector that deals with the production, storage, and consumption of food.

One technique utilized to build an expiry predictor was XGBoost. According to the methods and results section, the XGBoost approach failed to produce the expected evaluation metric. The target was to get the mean squared error to be less than 1.0. The nature of the dataset, which prevented the model from learning from previous errors, is the main cause of this failure. Because the application would be used by well-known hotels, it was not desired to publish or end the study with the findings obtained using this methodology.

The Random Forest model, an alternate methodology used to construct the expiry predictor, produced outcomes that were on par with those of the prior model. Additionally, a mean squared error of less than 1.0 was not produced. This failure was mostly caused by the hyperparameters used in the initial testing and overfitting. A model is considered to be overfit when it develops and performs admirably on the training dataset but fails miserably on samples of test data. Underfitting is the term used to describe a model's performance when it fails to learn the problem well, performs poorly on the training dataset, and performs poorly on test samples. A model is said to have a good fit if it learns the training dataset effectively and generalizes to the original dataset successfully.

To increase the model's effectiveness and counteract overfitting, the depth of the trees was reduced. Following multiple iterations of adjusting the hyperparameters, the project team made changes while often checking the model's performance. The project team was able to pinpoint the perfect variables that led to the predicted outcome of the model.

These results were used in a pilot test, which got positive feedback from potential customers. The outstanding ratings offered by the clients on the utility of the expiry date predictor and the freshness index provided the project team with helpful information about how well the solution will perform in the market. The team also plans to get in touch with the United Nations World Food Program and work with them to minimize food loss and waste on a worldwide scale.

REFERENCES

[1] H. Agarwal, B. Ahir, P. Bide, S. Jain, and H. Barot, “Minimization of Food Waste in Retail Sector using Time-Series Analysis and Object Detection Algorithm,” 2020 International Conference for Emerging Technology (INCET), 2020.

[2] “FAO supports the Ministry of Environment on a National Roadmap to tackle the Urban Food Waste challenge in Sri Lanka — FAO in Sri Lanka — Food and Agriculture Organization of the United Nations”, Fao.org, 2021. [Online]. Available: https://www.fao.org/srilanka/news/detail-events/en/c/1430231/

[3] R. Ishangulyyev, S. Kim, and S. Lee, “Understanding Food Loss and Waste—Why Are We Losing and Wasting Food?,” Foods, vol. 8, no. 8, p. 297, 2019.

[4] “The Sustainable Development Goals Report”, United Nations, 2022. [Online]. Available: https://unstats.un.org/sdgs/report/2022/The-Sustainable-Development-Goals-Report-2022.pdf

[5] H. Purandare, N. Ketkar, S. Pansare, P. Padhye, and A. Ghotkar, "Analysis of post-harvest losses: An Internet of Things and machine learning approach," 2016 International Conference on Automatic Control and Dynamic Optimization Techniques (ICACDOT), 2016, pp. 222-226, doi: 10.1109/ICACDOT.2016.7877583. Leverenz, G. Hafner, S. Moussawel, M. Kranert, Y. Goossens and T. Schmidt, “Reducing food waste in hotel kitchens based on self-reported data”, 1st ed. Elsevier Inc., 2021, pp. 1,2,3,4,5,6,7.

[6] R. Shinta Oktaviana, D. A. Febriani, I. Yoshana and L. R. Payanta, "FoodX, a System to Reduce Food Waste," 2020 3rd International Conference on Computer and Informatics Engineering (IC2IE), 2020, pp. 361-365, doi: 10.1109/IC2IE50715.2020.9274576.

[7] Neo Chai Chin, an 11th Hour answer to cutting down on food waste, November 2016, [online] Available: http://www.todayonline.com/singapore/11th-hour-answer-cutting-down-food-waste

[8] N. Kumar S.V., S. Balasubramaniam, S. Tharagesh R.S., P. Kumar and B. Janavi, "An Autonomous Food Wastage Control Warehouse: Distributed Ledger and Machine Learning based Approach," 2020 11th International Conference on Computing, Communication and Networking Technologies (ICCCNT), 2020, pp. 1-6, doi: 10.1109/ICCCNT49239.2020.9225525

[9] “Fruits and Vegetables - Optimal Storage Conditions", Engineeringtoolbox.com,2022.[Online].Available:https://www.engineeringtoolbox.com/fruits-vegetables-storage-conditions-d_710.html

[10] Gislason PO, Benediktsson JA, Sveinsson JR (2006) Random forests for land cover classification. Pattern Recognition Letters 27 (4): 294–300.