International Journal of Scientific Research in Engineering and Management (IJSREM)
Volume: 10 Issue: 10 | Oct - 2026 SJIF Rating: 8.659 ISSN: 2582-3930

AI-Based Food Availability Recommendation System Using Machine Learning

Vaishnavi Ghule[1], Sarthak Lawande[2], Dr. Kavita Dhakad[3]
[1,2]M.Sc. CA, School of Information Technology, Indira University, Pune, India
[3]Associate Professor, School of Information Technology, Indira University, Pune, India
Email: ghulevaishzzz2004@gmail.com, sarthaklawande07@gmail.com

***Abstract*— Plenty of edible food gets thrown away every day, while many people nearby go without enough to eat. Part of the reason is that nobody has a quick, consistent way to decide what should happen to a surplus item. In this paper we build a small machine-learning system that looks at a retail food item and suggests one of four actions: Donate, Monitor, Discount or None. Our evidence comes from two places. We surveyed 29 food senders and 26 receivers, and we used a Kaggle inventory dataset of 1,000 records. After cleaning the data, we trained a Decision Tree, a Random Forest and a Logistic Regression model on 80% of the records and tested them on the other 200. Each model classified every test record correctly. In the Random Forest, spoilage chance, days until expiry, stock quantity and daily sales accounted for about 80% of the feature importance, and the survey respondents pointed to food safety, shelf life, quantity and distance as the details that matter most. A perfect score on a small dataset with two rare classes is not proof that the system would work in a real store, so we treat this work as a proof of concept that still needs testing on real donation records.**

***Keywords*— Food Donation, Food Waste, Recommendation System, Machine Learning, Random Forest, Decision Tree, Logistic Regression, Inventory Data.**

I. Introduction

Food waste and hunger are two sides of the same problem. A supermarket may bin bread that is still perfectly good, and a few kilometres away a shelter runs short of meals. Much of the gap comes down to coordination. Surplus food is usually passed on through phone calls and messages between a donor and an NGO, and when an item is close to expiry the person in charge often decides by gut feeling whether to donate it, put it on sale, watch it for another day or do nothing. That is a pity, because the store already holds most of what it needs to decide well: how many units are left, how many days remain before expiry, how fast the item sells and how likely it is to spoil.

We wanted to see whether those everyday inventory details could be turned into a simple recommendation. The system we describe suggests one of four actions for each item: Donate, Monitor, Discount or None. We used two kinds of data. Two short online surveys told us what senders and receivers care about when food changes hands, and a retail inventory dataset of 1,000 records from Kaggle gave us something to train on. We then built three common classifiers in Python and compared how well each one did.

This paper makes four contributions. It combines survey evidence from both sides of a donation with real-looking inventory records. It lays out a documented cleaning and feature-selection procedure. It compares three classifiers using class-wise results and not just overall accuracy. And it packages the chosen model into a pipeline that returns a suggested action for a new item.

II. Related Work

A number of recent projects try to make food donation smarter. ZeroHunger puts donation, quality checks, fraud analysis and donor–recipient matching into one web application [1]. Annapurna connects donors, receivers and volunteers and uses AI to suggest a suitable receiver [2]. Another study looks at how AI, machine learning and IoT sensors could help with allocation, delivery and prediction in donation systems [3]. Most of this work is about linking people. Few of these systems try to predict what should be done with one particular item.

Prediction research offers some useful ideas. Uen and Rodríguez compared machine-learning models that forecast food waste at county level using demographic, retail, economic and social variables [4], and Said et al. reviewed how AI and IoT are being applied to food-waste management [5]. Shelf-life research is also close to our question, since it asks how long food stays usable. Rashvand et al. surveyed that field and noted that limited data and poor generalisation keep coming up as problems [6].

Work on food recommendation adds the idea of context. Trattner and Elsweiler review food recommender systems and their open problems [7], and Min et al. describe a framework built around user models, food properties and context [8]. In a donation setting the relevant context is things like location, quantity and how urgent the situation is. It is also worth remembering that technology is only part of the picture. Hudak et al. show that federal policy has a real effect on what food banks can accept [9].

What we found missing is a lightweight approach that works on ordinary structured inventory data, without images or special sensors, and tells staff what to do with each item. Most studies handle waste prediction, quality, donation platforms and recommendation separately, and some need equipment a student project cannot get. That is the gap we try to address, and we do it with classical classifiers that are easy to run and easy to explain.

III. System Workflow

The system works in six steps. First, the survey responses and the inventory file are loaded. Second, the data is cleaned: long survey headings are shortened, dates are converted, columns that are mostly empty are removed, and we check for duplicates and negative values. Third, we pick ten input features and set SuggestedAction as the target. Fourth, the three classifiers are trained. Fifth, they are compared on records they have not seen. Sixth, preprocessing and the chosen model are bundled into a single pipeline that can take a new food item and return a suggested action.

Everything was written in Python 3 on Google Colab, with pandas and NumPy for the data, scikit-learn for modelling and evaluation [10] and Matplotlib for the charts. All of these are free, so no special hardware or budget is needed. The output is a single word that store staff or NGO volunteers can read without

any technical background. We should be clear about what the system is not: it runs offline, it is not connected to a live inventory system, and it does not judge whether food is safe to eat.

IV. Datasets and Preprocessing

A. Survey data

The two online surveys ran from 26 to 28 August 2026. We received 29 responses from senders and 26 from receivers. Each survey had five multiple-choice questions about the respondent's role, the information needed to donate or request food, and the checks to make before accepting it.

B. Inventory data

The inventory data comes from the Smart Waste Inventory dataset on Kaggle [11]. It has 1,000 records and 19 columns covering 15 food items, 10 brands, five categories (Bakery, Dairy, Meat, Produce and Snacks) and five store cities (Bangalore, Chennai, Delhi, Kanpur and Mumbai). Arrival dates fall between 12 and 27 June 2025 and expiry dates between 15 June and 8 July 2025. The target column, SuggestedAction, has four classes.

C. Cleaning and feature selection

Two columns, Zone and TemperatureSupport, were empty in 656 records each (65.6%), so we dropped them and were left with 17 columns. We found no duplicate rows and no negative numbers, and every date converted without trouble. Item name, brand and store ID were not used as inputs because they label a record and say little about its condition. IsSpoiled was left out since SpoilageChance already captures the risk, and the two calendar dates were replaced with DaysUntilExpiry, which gives the time remaining as one number. The ten features we kept were Category, StockQty, DaysUntilExpiry, DailySaleAvg, SpoilageChance, StoreLocation, OnPromotion, TemperatureSensitive, DistanceToNearestStore and AvgDailySaleInNearbyStores.

D. Encoding and data split

Categorical features were one-hot encoded, and any category not seen in training is ignored at prediction time. Numeric features were passed on unscaled. We fitted the transformer on the training data only. The records were split 80/20, stratified by SuggestedAction with a random seed of 42, so the rare classes show up in both parts (Table I). There is one implementation detail worth stating plainly. TemperatureSensitive is a Boolean column, and because we picked numeric and categorical columns by data type, it ended up in neither group. The models therefore worked with nine of the ten selected features, which produced 18 processed columns.

Table I. SuggestedAction distribution and split

Action	Records	Share	Train	Test
Donate	681	68.1%	545	136
None	200	20.0%	160	40
Monitor	96	9.6%	77	19
Discount	23	2.3%	18	5
Total	1,000	100%	800	200

V. Machine-Learning Methodology

The task is multiclass classification: given a set of food attributes, predict one of four actions. A Decision Tree asks a series of yes-or-no questions about the features, such as whether days until expiry is below some value, until it reaches a decision [12]. A Random Forest grows many trees and lets them vote, which tends to be steadier than any single tree [13]. Logistic Regression estimates class probabilities from a weighted sum of the inputs, and we used it as a simple baseline [14]. The tree used default settings, the forest used 100 trees, and Logistic Regression was allowed up to 1,000 iterations, all with a seed of 42.

We judged the models on the 200 test records. Accuracy is the share of records classified correctly. Precision shows how many of the records predicted for a class really belong to it, recall shows how many of the true members were found, and the F1-score balances the two. We report weighted averages as well as class-wise values, and we used a confusion matrix to see which classes get mixed up [15]. Since Donate covers more than two thirds of the data, accuracy by itself could be misleading, which is why we looked at the other measures too.

VI. Experimental Results

A. What the data shows

Most items in the inventory are very close to expiry. The median is just 0.5 days, and at least half the items have a spoilage chance of 1.0, which helps explain why Donate is so common. Food category is closely tied to the action (Table II). Every Produce item is marked for donation, while Bakery has the lowest Donate share and makes up 81 of the 96 Monitor records. All 23 Discount records belong to a single Snacks item in Kanpur that was not on promotion. Among the numeric features, daily sales and nearby-store sales move almost together (correlation 0.972), and spoilage chance falls as days until expiry rise (−0.479).

Table II. Donate share by food category

Category	Records	Donate	Share
Bakery	209	70	33.5%
Dairy	178	140	78.7%
Meat	203	149	73.4%
Produce	133	133	100.0%
Snacks	277	189	68.2%

B. Model performance

All three classifiers put every one of the 200 test records in the right class (Table III). In the Random Forest confusion matrix every count sits on the diagonal: 5 Discount, 136 Donate, 19 Monitor and 40 None records were predicted correctly, with no mistakes. Because the models tied, we chose the Decision Tree for the final pipeline. It is the easiest to explain, and a store manager can follow its rules step by step.

Table III. Test-set performance (200 records)

Model	Acc.	Prec.	Rec.	F1
Decision Tree	1.00	1.00	1.00	1.00
Random Forest	1.00	1.00	1.00	1.00
Logistic Reg.	1.00	1.00	1.00	1.00

C. Feature importance

In the Random Forest, spoilage chance mattered most, followed by days until expiry, stock quantity and daily sales (Table IV). Those four together hold about 80% of the total importance. These numbers describe what the fitted model leans on. They do not show that any feature causes an outcome.

TABLE IV. MOST IMPORTANT FEATURES

Rank	Feature	Importance
1	SpoilageChance	0.4087
2	DaysUntilExpiry	0.1520
3	StockQty	0.1334
4	DailySaleAvg	0.1070
5	AvgDailySaleInNearbyStores	0.0477
6	Category = Bakery	0.0435

D. A new item and what the surveys say

We tried the final pipeline on a Bakery item with 100 units in stock, one day left before expiry, a spoilage chance of 0.7, average daily sales of 10 and no promotion, in a store in Pune. It recommended Monitor. Pune is not one of the five training cities, so the encoder simply ignored it, and TemperatureSensitive never reached the pipeline, so the answer rested mostly on the other inputs.

The survey answers point in the same direction as the model. Among senders, 75.9% said food name and quantity must be provided, and 65.5% picked food safety and shelf life as the information needed for safe donation. Among receivers, 80.8% would check food safety and shelf life before accepting food, 76.9% named NGO demand and required quantity, and 73.1% said distance matters because it helps find the nearest donor. At the same time, only 51.7% of senders and 65.4% of receivers described their own role correctly, and 34.5% of senders said they would throw surplus food away. That suggests people still need more awareness of how donation works. The samples are small, so these percentages are only a rough guide.

VII. DISCUSSION

The results show that ordinary classifiers can turn an inventory record into a clear action, and the inputs the models rely on (spoilage risk, time to expiry, quantity and sales) are the same things senders and receivers told us they care about. That agreement is encouraging, but it should not be pushed too far.

The perfect test scores deserve a careful look. The models learn from the SuggestedAction labels already in the dataset, and the dataset does not explain how those labels were produced. If they came from simple rules based on a few columns, such as spoilage chance and days until expiry, the classifiers may just have rediscovered those rules. That would explain both the perfect scores and the high importance of those columns, and we cannot rule it out. The test set also has only 5 Discount and 19 Monitor records, so a perfect score on those classes tells us very little. What the result really shows is that the pipeline works on this dataset, not that it will work on real donation decisions.

The model choice also deserves a comment. Since all three models tied, accuracy could not settle it, and interpretability decided the matter. A decision tree lets a staff member trace how a recommendation was reached, which matters for a tool meant to support people and not replace their judgement.

VIII. LIMITATIONS AND FUTURE WORK

This study has clear limits. The inventory data covers only about two weeks and five cities, and the surveys have just 29 and 26 responses. We took the labels as given, so any errors or biases in them carry over. TemperatureSensitive was selected but never reached the models because of the data-type filter, and Logistic Regression gave a convergence warning because the numeric features were not scaled. The new-item example used a city the model had never seen, and we evaluated on a single split without cross-validation.

There is a lot we would like to do next. We plan to test the models on real donation and inventory records from more stores and longer periods, using cross-validation and per-class reporting. We want to document how the action labels are created and have domain experts check them. On the technical side, we will encode the Boolean feature, scale the numeric inputs, tune the hyperparameters and deal with class imbalance through class weights or oversampling such as SMOTE [16]. Beyond that, we would like to add food-quality and shelf-life prediction, match surplus food to suitable NGOs using demand, quantity, distance and safety, explain individual recommendations with a method such as SHAP [17], and build a simple web app, perhaps in Streamlit, for store staff.

IX. CONCLUSION

We set out to see whether survey evidence and retail inventory data could be combined to recommend an action for each food item. Three classifiers were trained on 800 records and tested on 200, and all of them reached 100% accuracy, precision, recall and F1-score. We kept the Decision Tree as the final model because it is the easiest to interpret. Spoilage chance, days until expiry, stock quantity and daily sales were the most influential features, and the surveys showed that food safety, shelf life, quantity and distance are what donors and receivers value most. Because the scores are perfect, the data is small and some classes are very rare, the system needs to be validated on real data before anyone relies on it. It is meant as an aid for decision-making, and judgements about food safety must stay with trained people.

REFERENCES

[1] V. Suresh, A. N. Jonnalagadda, D. Jagdheesh and M. Balasubramanian, “ZeroHunger: A digital solution for reducing food waste and hunger alleviation,” 2025, doi: 10.2991/978-94-6463-866-0_67.

[2] V. Dhanokar *et al.*, “Annapurna: An intelligent food donation system with AI-based receiver recommendation,” 2026, doi: 10.51583/IJLTEMAS.2026.150500008.

[3] B. Bhargavi *et al.*, “AI based food donation and distribution system,” 2026, doi: 10.17148/IJARCCE.2026.154313.

[4] T.-S. Uen and L. F. Rodríguez, “Geospatial analytics and machine learning for forecasting county-level food waste in U.S. retail markets,” *Resources, Conservation and Recycling*, vol. 231, Art. no. 108906, 2026, doi: 10.1016/j.resconrec.2026.108906.

[5] Z. Said *et al.*, “Intelligent approaches for sustainable management and valorisation of food waste,” *Bioresource Technology*, vol. 377, Art. no. 128952, 2023, doi: 10.1016/j.biortech.2023.128952.

[6] M. Rashvand, Y. Ren, D.-W. Sun, J. Senge, C. Krupitzer *et al.*, “Artificial intelligence for prediction of shelf-life of various food products: Recent advances and ongoing challenges,” *Trends in Food Science & Technology*, vol. 159, Art. no. 104989, 2025, doi: 10.1016/j.tifs.2025.104989.

[7] C. Trattner and D. Elsweiler, “Food recommender systems: Important contributions, challenges and future research directions,” 2017.

[8] W. Min, S. Jiang and R. Jain, “Food recommendation: Framework, existing solutions and challenges,” 2019.

[9] K. M. Hudak, E. Friedman, J. Johnson and S. E. Benjamin-Neelon, “Food bank donations in the United States: A landscape review of federal policies,” *Nutrients*, vol. 12, no. 12, Art. no. 3764, 2020, doi: 10.3390/nu12123764.

[10] F. Pedregosa, G. Varoquaux, A. Gramfort *et al.*, “Scikit-learn: Machine learning in Python,” *J. Mach. Learn. Res.*, vol. 12, pp. 2825–2830, 2011.

[11] Kaggle, “Smart Waste Inventory Dataset (wildlearner).” [Online]. Available: https://www.kaggle.com/datasets/wildlearner/smart-waste-inventory-dataset

[12] J. R. Quinlan, “Induction of decision trees,” *Machine Learning*, vol. 1, no. 1, pp. 81–106, 1986, doi: 10.1007/BF00116251.

[13] L. Breiman, “Random forests,” *Machine Learning*, vol. 45, no. 1, pp. 5–32, 2001, doi: 10.1023/A:1010933404324.

[14] J. Tolles and W. J. Meurer, "Logistic regression: Relating patient characteristics to outcomes," *JAMA*, vol. 316, no. 5, pp. 533–534, 2016, doi: 10.1001/jama.2016.7653.

[15] M. Sokolova and G. Lapalme, "A systematic analysis of performance measures for classification tasks," *Information Processing & Management*, vol. 45, no. 4, pp. 427–437, 2009, doi: 10.1016/j.ipm.2009.03.002.

[16] N. V. Chawla, K. W. Bowyer, L. O. Hall and W. P. Kegelmeyer, "SMOTE: Synthetic minority over-sampling technique," *J. Artif. Intell. Res.*, vol. 16, pp. 321–357, 2002, doi: 10.1613/jair.953.

[17] S. M. Lundberg and S.-I. Lee, "A unified approach to interpreting model predictions," in *Advances in Neural Information Processing Systems (NeurIPS)*, vol. 30, 2017, pp. 4765–4774.