8-8-2022

Applying Artificial Intelligence to Reduce Food Waste in Small Grocery Stores

Alexandre M. Nascimento

Anna C. M. Queiroz

Vinicius V. de Melo

Fernando S. Meirelles

Applying Artificial Intelligence to Reduce Food Waste in Small Grocery Stores

Completed Research Paper

Alexandre M. Nascimento
Stanford University
alexandre.nascimento@sloan.mit.edu

Anna C. M. Queiroz
Stanford University
acmq@stanford.edu

Vinicius V. de Melo
Wawanesa Mutual Insurance
dr.vmelo@gmail.com

Fernando de Souza Meirelles
EAESP-FGV
fernando.meirelles@fgv.br

Abstract

A typical strategy for increasing grocery stores' profitability is the production and commercialization of their branded products. This strategy requires production planning, which impacts the business profitability and food waste. The purpose of this study was to develop and test an artificial intelligence (AI) model to predict the demand of a small grocery store and to use this model to support the own branded product production planning to reduce food waste. Through machine learning (ML) and feature expansion techniques, it was possible to address several existing limitations in the available data from a small grocery store in Brazil. From five distinct ML techniques created to predict the store's revenue, from which the demand is derived, the random forest algorithm was selected because of its superior accuracy. The model was tested using actual data from a small grocery store. The results demonstrate that AI techniques can enhance the production planning accuracy of food processing and promote a significant reduction in food waste while positively impacting profitability, among other fringe benefits, ultimately resulting in a more sustainable operation. The results open new possibilities for AI research to be applied to improve sustainability and reduce food waste.

Keywords

Food Waste, Artificial Intelligence, Grocery Store, Sustainability, Small Business.

Introduction

Loss and waste exist in all parts of the food supply chain. Food travels a long distance before reaching the consumer (Vieira, Carvalho, Ferraz, & Guimarães, 2017), and waste accounts for a quarter to a third of all food produced in the world, according to the Food and Agriculture Organization of the United Nations (UN) (Gustavsson, Cederberg, Sonesson, Van Otterdijk, & Meybeck, 2011). That is a multidimensional issue, including social (Parizeau, Massow, & Martin, 2015), economic (Parfitt, Barthel, & Macnaughton, 2010; Parizeau et al., 2015), and environmental (Chaboud & Daviron, 2017; Parizeau et al., 2015), among other impacts. As the world population is expected to reach 10 billion by 2050 (Vieira, Barcellos, & Matzembacher, 2018), food waste reduction has become an emerging critical issue for sustainable development (Muriana, 2017). Therefore, the UN Agenda 2030 goal is to reduce food loss and waste by 50 percent by 2030.

Economically, food waste and its generated residues account for an approximate loss of US$ 310 billion in developing countries (de Brito Nogueira et al., 2020). Initiatives have aimed to reduce these losses, such as shortening the supply chain and improving traceability. Also, food tech start-ups have emerged to explore food waste reduction using innovative technologies (Vieira et al., 2018). For example, digital business platforms can improve supply chain coordination, which can help avoid food losses (Vieira & Matzembacher, 2020). However, these initiatives can generate social benefits beyond the promising financial results since food waste reduction can help reduce starvation and improve food security (Vieira et al., 2018). Moreover, these programs generate environmental benefits since they can positively impact soil, water, energy, and agricultural resource conservation and reduce atmospheric pollution (Vieira et al., 2018).

Grocery stores play a significant role in food waste. According to Eriksson (2015), the retail stage of the supply chain contributes a significant amount of waste. Cicatiello et al. (2017) identified significant food waste in one retail store in Italy, mainly from the fresh meat and bakery departments. Lee (2018) found that food retailers in Seoul influence household food waste because of the marketing promotions and the way they may shape a household's grocery shopping patterns. In the context of small grocery stores, competitive pressures force companies to seek alternatives that result in better margins and cost reduction. A typical strategy to increase profitability is the expansion of a store's product offerings by producing and selling its own branded products.

On the one hand, this strategy can promote better control over margins and profit. On the other hand, a business must deal with many aspects of production, including planning. The production planning impacts the business profitability and ultimately the food waste, which can impact costs. Griffin et al. (2009) found that production accounted for 20% of all food waste in one US county. Therefore, food waste reduction in these operations offers an opportunity to increase the supply chain sustainability, operational efficiency, cost reduction, and profitability.

Food waste reduction can become even more challenging for grocery stores focused on natural and organic foods. Eriksson, Strid, and Hansson (2014) found that organic food products have a higher level of waste by comparing 24 organic products to their conventional counterparts. The authors found that 22 of these products had higher waste levels (from 1.5- to 29-times higher). McCarthy and Liu (2017) found that organic food consumption was linked to a higher propensity for waste food since a good deal of the food was thrown away due to spoilage, the short shelf life of fresh food, and because people forgot about food left in the refrigerator. Natural and organic grocery store customers demand fresh products (not frozen) with no food additives or preservatives, reducing the life span of the branded products produced on a small scale.

On the one hand, while a small production could reduce food waste, it can result in unfulfilled demand and lower customer satisfaction. On the other hand, the eagerness to improve sales by fulfilling the demand can result in higher food waste because these premium ingredients are usually more expensive, and food waste can erode profitability. Therefore, business managers must deal with the trade-off between food waste and lost sales.

Recent progress in artificial intelligence (AI), particularly machine learning (ML), enables new automation, prediction, and optimization possibilities in many fields (Nascimento, Cunha, Meirelles, Scornavacca, & Melo, 2018; Nascimento & Bellini, 2018). These methods can offer promising tools for retail operations. However, these technologies are not yet widely available to small grocery stores. Since both predictable and unpredictable components can impact food waste (Muriana, 2017), which is also impacted by management competence (Mena, Adenso-Diaz, & Yurt, 2011), AI can potentially be used for food waste reduction (Li, Lu, Ren, & He, 2013). However, to the best of our knowledge, no study has proposed and tested an AI-based approach to address the food waste of a small grocery store focused on natural and organic foods. In this context, this study aims to develop and test an AI model to enhance the production planning in a small grocery store focused on natural and organic food in Brazil to reduce the store's food waste.

Methods and Materials

A protocol similar to a previous research (Nascimento, Melo, Queiroz, Brashear-Alejandro, & Meirelles, 2021) was adopted. A research question was posed to guide the application of predictive machine learning models to assist managers of small grocery businesses in enhancing the accuracy of their planning. The defined research question was "How can data easily obtained from small grocery stores, such as daily revenue, be used to reduce food waste?". Figure 1 illustrates the protocol steps used in this work.

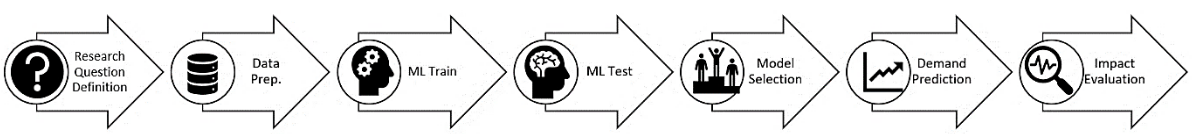

Figure 1: Research protocol.

Data Preparation

The available data were the daily sales of one retail food store over one year, beginning on the first day of August 2016 and ending on the last day of July 2017. Several procedures were used to prepare the data, as shown in Figure 2.

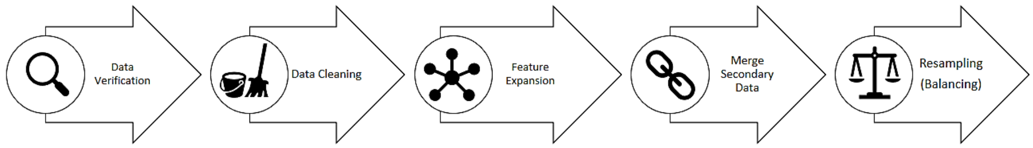

Figure 2: Dataset preparation protocol.

The first step, data verification, encompassed all the work of formatting data and addressing missing values. Since the raw dataset had only two features (date and sales revenue), it was necessary to treat the data to derive new features. The feature expansion was performed manually to improve the model learning process. Initially, a new attribute–the day of the week–was generated and added. This attribute was added in numeric format. Then, the data was partitioned by the day of the month, the month, and the year so that the original attribute was replaced with three new numeric attributes. By doing this, we could maintain the important weekly, monthly, and yearly seasonal information. Finally, to ensure that sequence information over time was not lost, a sequential numerical field was added by storing the sequence information for each day throughout the year. All data points were numeric. Subsequently, the data were supplemented with secondary climatic data. Using a historical database of climate information (https://www.wunderground.com), we extracted and added the following attributes to the store data: daily temperature information (maximum and minimum), the relative humidity, and the occurrence of a climatic event (presence of wind, thunderstorm, rain, etc.). Right after, a flag to indicate a holiday was added as another feature. Because holidays affect sales, it is usually helpful for predictive models to have a flag to distinguish whether the information is related to a holiday or an ordinary day. After some exploratory tests using regressions, it was clear it would be difficult to obtain a model that would explain the daily revenue numerically with adequate precision due to the small amount of data. For this reason, a new attribute––daily revenue level––was defined to replace the daily revenue value by partitioning the daily revenue information into five levels (classes), so discrete daily revenue intervals were used as the target variable of the model. The revenue level classes are: VL - very low (1120, 2.750]; L - low (2750, 4380]; M - medium (4380, 6010]; H - high (6010, 7640]; and VH - very high (7640, 9280]. Table 1 shows the features of the dataset prepared to be used in the ML models.

Feature	Description
SEQUENCE	Sequential number to preserve the information of the order of each day
DAYOFWEEK	Day of the week in numeric format from Sunday (1) to Saturday (7)
HOLIDAY	0 indicated that the day is not a holiday, and 1 indicated that it is a holiday
TEMPMAX	Maximum temperature of the day in Celsius
TEMPMIN	Minimum temperature of the day in Celsius
HUMIDITY	Percentage of relative humidity
CLIMATEEVENT	0 indicates no climate event that day, such as rain, thunderstorm, etc. 1 indicates one or more climate events that day
DAYOFMONTH	The day of the month (1-31)
MONTH	The month of the year (1-12)
REVENUELEVEL	Level of revenue (1, 2, 3, 4, 5)

Table 1. Dataset features after the preparation.

Although the prediction model is intended to forecast the daily demand for its branded products, the store did not provide information about the daily demand per item. Therefore, an alternative approach to derive the daily demand from the daily revenue level was adopted, and the model created from the available data forecasted the expected revenue for a specific day. This model is given by Revenuelevel = F (Sequence, Dayofweek, Holiday, Tempmax, Tempmin, Humidity, Climateevent, Dayofmonth, Month).

Another issue was that the dataset was unbalanced. The low, low, medium, high, and very high classes have 36, 109, 137, 28, and 2 observations, respectively. That issue was addressed by applying a resampling technique to balance the dataset by augmenting the number of observations of the minority classes. This approach helps the model consider the minority classes as important as the majority ones. Therefore, we used the supervised resampling filter in Weka 3.8 to balance the dataset to ensure the generation of rules would cover all classes. The configuration was a bias of 0.95 and 250% of the sample size. No other filter was used. Consequently, an augmented version of the original dataset was created.

Training the classifiers

Five classification techniques with their default configuration in the Weka machine learning framework version 3.8 were selected to be trained using the augmented version of the original dataset. Logistic regression (LR) was selected because it is a traditional and straightforward technique to predict events. Although neural networks (NN) are typically a black-box technique, they do not output a human-readable explanation that supports the classifications performed. They are usually powerful for finding complex relationships between dependent and independent variables that can result in good predictions. The type of NN selected was the multilayer perceptron (MLP), one of the most commonly used because of its easy implementation, fast operation, and reduced training set requirements (Orhan, Hekim, & Ozer, 2011). The other selected technique was the decision tree J48 (DT), another classical technique. The DT is a rule learner classifier, creating a list of human-readable decision rules. Therefore, DT is considered a white-box technique that helps managers understand the rationale for the classification decisions. The two other techniques were rule-based: PART and random forest (RF). PART is a rule learner that creates a list of decision rules. To accomplish that, PART builds partial decision trees at each iteration and turns the "best" leaf into a rule, avoiding the computational cost of performing a global optimization such as DT's. RF uses an ensemble learning method based on training multiples and the mode of the individual tree classes, which avoids overfitting the training data.

Testing and selecting the classifiers

Following the protocol, each of the selected classifiers was induced using the augmented dataset, and subsequently, the resulting classifiers were tested with the original dataset. It is noteworthy this it is not the same as using the training set for testing, which could provide a overfit. That is required because one should not use the cross-validation strategy to evaluate the classifier's performance, as an observation in the training set will likely be in the test set.

After the induced classifiers had been tested, the best classifier was selected. A paired t-test (with correction) was used to compare the accuracy measures of each model. The significance of the test was set at 5%.

Predicting the demand and evaluating the impact of the forecast

Although the model predicted the revenue level for a specific day, a forecast of the demand for its branded products was required. Ideally, the historical sales per product information would need to be available to support the induction of an ML model to predict these sales. However, as explained previously, that information was unavailable, and the following approach relying on the derivation of the demand for own branded products from the daily revenue was adopted.

In addition to the daily revenue, additional information was available: (a) the majority of own branded products were from the bakery; (b) bakery products corresponded to 25% of the daily sales; (c) the average price of those products per kg was R$ 62.00, and (d) the average cost of those products per kg was R$ 33.00. This additional information made deriving the average demand for bakery products from the daily sales forecast possible. To derive the average demand for bakery products from the daily sales forecast, for each revenue level, we calculated the average value of the range, which represents the average total sales of the day. From this value, 25% was calculated, corresponding to the daily sales of own branded products (bakery). Then, this fraction value was divided by R$ 62.00 to calculate the average demand in kg of own branded products for each revenue level.

Finally, the last activity evaluated the impact of production planning based on the model forecasts. The evaluation was performed using the created ML model to predict the demand for its branded products for each day of the dataset and compare the prediction to the actual demand. Using the predicted values to dimension the production and the real revenue data to derive the actual demand, it was possible to measure the impact of the proposed approach on the production cost, sales revenue, gross profit, food waste and lost sales. Moreover, three other performance ratios were used for performance evaluation: return on investment (ROI as revenue over production cost), food waste in production (FWP as % of the production), and fulfilled demand (FD as % fulfillment of the total demand).

Moreover, the metrics and ratios computed from the proposed approach were compared to those obtained from alternative production planning approaches. The alternative approaches were used as benchmarks to provide a more comprehensive evaluation of the proposed technique. The first approach was based on the production of the lowest demand. This approach is referred to here as "conservative"; even though there was no latent demand that could be fulfilled with higher production, this approach aims to minimize food waste. The second approach was based on the production of the highest demand. This approach is referred to here as "aggressive" and aims to fulfill all the demand

and maximize revenue, despite potential losses due to food waste. The third benchmark approach referred to as "balanced" is based on using the average demand to dimension the production of own branded products: this approach balances revenue and food waste. Of course, alternative strategies based on averages of each day of the week or moving averages could be used rather than the "balanced" approach used here. However, for this study, the average of the whole demand was adopted since the goal was to use a simple benchmark that would fall between extremes.

Results

As previously described, following the training and testing of the classifiers, a paired t-test with a 5% significance level was used to compare each model's accuracy and support the selection of the best classifier. All the comparisons were performed against the reference model. Table 2 shows the model accuracies and indicates significant differences.

Model	LR	MLP	DT	PART	RF
Accuracy	47.78	66.77	82.28	84.81	90.19
t-test (LR as reference)	-	>	>	>	>

Table 2. Comparison of the ML models.

The least accurate was logistic regression. That is the simplest model; therefore, its poor performance in forecasting the demand was expected. Although the MLP achieved higher accuracy than logistic regression, it surprisingly performed below expectations compared to the rule-based models (DT, PART, and RF). Those three models achieved accuracy levels higher than 80%, with the advantage of being white-box approaches. Table 3 shows the t-test results among the rule-based models, using RF as the reference model.

Model	RF	PART	DT
Accuracy	90.19	84.81	82.28
t-test (RF as reference)	-	<	<

Table 3. Comparison of the rule-based ML models.

All the models compared to RF achieved inferior accuracy according to the t-test at a 5% significance. Therefore, RF is the best model when considering accuracy. Consequently, RF was the selected classifier. The result of the revenue level prediction based on the test set using RF is shown in Table 4, which shows a good class differentiation and a reasonable degree of success.

Correctly classified Instances	281 (90%)
Incorrectly classified Instances	31 (10%)
Kappa statistic	0.83
Mean absolute error	0.08
Root mean squared error	0.18
Relative absolute error	26.27%
Root relative squared error	45.82%
Total number of instances	312

Table 4. Results of RF using the default configuration.

Predicted					
VL	**L**	**M**	**H**	**VH**	**Ground Truth**
49	1	0	0	0	**VL**
4	161	22	0	0	**L**
0	4	67	0	0	**M**
0	0	0	3	0	**H**
0	0	0	0	1	**VH**

Table 5. Confusion matrix (Green: Correct Classification; Red: Incorrect Classification).

Table 5 shows the confusion matrix. The rows indicate how the ML classifier distributed the classifications over the classes. The columns compare those classifications with the ground truth. The cells over a column represent how the instances classified into such classes should be classified over the five levels. The cells in green represent the correct classification. The cells in red indicate an incorrect classification.

The first two columns of Table 6 show the classification's precision performed for each class (revenue level). Only the underrepresented levels (high and very high) had 100% of the classifications performed correctly. This result is probably because of the bias induced by resampling those classes to force these classes to be considered essential during the ML model creation. Classes VL, L, and M (very low, low, and medium, respectively) had 98%, 86%, and 94% of the classifications performed correctly. Of 312 days, 31 days (10%) were incorrectly classified.

If the production manager adopted the conservative approach to minimize food waste, the store would produce 4.54 kg of its branded products daily; that is, the minimal demand computed by the approach illustrated in Figure 3 (left). Since the minimal revenue in the period was R$ 1124.72, and 25% of the sales, on average, are related to purchases of own-branded products, that would correspond to an average revenue of R$ 281.18 from own branded products. Considering the average price of its branded products is R$ 62.00/kg, this revenue corresponds to a demand of 4.54 kg.

Alternatively, if the production manager adopted the aggressive approach to minimize the unfulfilled demand, the store would produce 37.39 kg of its branded products daily. That is the maximal demand computed by the same approach described previously. The maximum revenue in the period would be R$ 9273.92, corresponding to an average revenue of R$ 2318.48 and an average demand of 37.39 kg from own branded products. On the other hand, the production manager could adopt the "balanced" approach to promote a better balance between food waste reduction and demand fulfillment. Since the average revenue in the period was R$ 4370.04, the revenue and demand for own branded products would be R$ 1092.51 and 17.62 kg, respectively.

Finally, the production manager could rely on an approach based on artificial intelligence to plan the production quantity for each day. In this case, the RF classifier could be used to predict the revenue level. Then, the demand for its own branded products that can be used to plan the production could be derived from the average revenue of the predicted revenue level. Table 6 shows the corresponding demand of own branded products in kg for each revenue level, which can be used to convert the predicted revenue level of a future day from the RF classifier to the quantity of own branded products that need to be produced to fulfill the expected demand for own branded products in that day.

Revenue Level	Precision	Revenue Level Average (R$)	Own-Product Average Demand (R$)	Own-Product Average Demand (kg)
VL – Very Low	98%	1935.00	483.75	7.80
L – Low	86%	3565.00	891.25	14.38
M – Medium	94%	5195.00	1298.75	20.95
H – High	100%	6825.00	1706.25	27.52
VH – Very High	100%	8460.00	2115.00	34.11

Table 6. The precision of the classification and demand of own branded products per Revenue Level.

The demand for its branded products for each day was computed from the daily revenue from the original dataset. For each of the approaches above and for each day, we calculated the following: production (kg), production cost (R$), sales (kg and R$), lost sales (kg and R$), gross profit (R$), and food waste (kg and R$). Then, we calculated the average of these variables and their total value for the 312 days (Table 7). Finally, we calculated the ROI, FWP, and FD ratios. These results are presented in Table 8.

Metrics	Conservative		Aggressive		Balanced		Artificial Intelligence	
	Avg	Total	Avg	Total	Avg	Total	Avg	Total
Production (kg)	4.54	1,414.97	37.39	11,667.19	17.62	5,497.80	15.32	4,780.79
Prod. Cost (R$)	149.66	46,694.02	1234.03	385,017.26	581.50	181,427.26	505.66	157,765.95
Demand (kg)	17.62	5,497.80	17.62	5,497.80	17.62	5,497.80	17.62	5,497.80
Demand (R$)	1092.51	340,863.34	1092.51	340,863.34	1092.51	340,863.34	1092.51	340,863.34
Sales (kg)	4.54	1,414.97	17.62	5,497.80	15.53	4,844.52	14.95	4,663.15
Sales (R$)	281.18	87,728.16	1092.51	340,863.34	962.69	300,360.29	926.65	289,115.08
Gross Profit (R$)	131.52	41,034.14	(141.52)	(44,153.92)	381.20	118,933.04	420.99	131,349.13
Lost kg sales (kg)	13.09	4,082.83	0.00	0.00	2.09	653.27	2.68	834.65
Lost sales (R$)	811.33	253,135.18	0.00	0.00	129.82	40,503.04	165.86	51,748.26
Food waste (kg)	0.00	0.00	19.77	6,169.39	2.09	653.27	0.38	117.64
Food waste (R$)	0.00	0.00	1225.97	382,502.43	129.82	40,503.04	23.38	7,293.67

Table 7. Daily averages and totals (for the period) metrics per production planning approach.

Finally, considering the balanced approach as the reference, we computed the ratio of each approach's metric to the corresponding reference's metric. Those ratios are shown in Table 9 and can provide a method to compare the performance of each metric. For example, using the proposed artificial intelligence technique would result in a production level of 87% compared to the reference, while the aggressive approach would result in a production level of 212% compared to the same reference.

Ratios	Conservative	Aggressive	Balanced	Artificial Intelligence
ROI (%)	188%	89%	166%	183%
FWP (%)	0%	53%	12%	2%
FD (%)	26%	100%	88%	85%

Table 8. Ratios per production planning approach

Metric	Balanced (Ref.)	Artificial Intelligence	Aggressive	Conservative
Production (%)	100%	87%	212%	26%
Gross Profit (%)	100%	110%	-37%	35%
Lost Sales (%)	100%	128%	0%	625%
Food Waste (%)	100%	18%	944%	0%
ROI (%)	100%	111%	53%	113%
FWP (%)	100%	21%	445%	0%
FD (%)	100%	96%	113%	29%

Table 9. Comparison ratios per production planning approach

Discussion

Testing the five classifiers demonstrated that the RF achieved the best accuracy. This result was confirmed with t-tests. Thus, this classifier's application demonstrated that the revenue level could be correctly forecasted in 90% of the cases. Therefore, the model could forecast the revenue levels in 281 of the 312 days, benefiting the small business planning-performance relationship (Brinckmann, Grichnik, & Kapsa, 2010). Beyond the superior accuracy (90%), this rule-based model has the advantage of being a white-box, i.e., it is possible to understand the underlying rationale for the predictions, which is highly desirable for managers that would need to rely on its predictions to make decisions.

The selected model provided a reasonable explanation for all revenue levels. However, it is noteworthy that this model reached superior results for the extreme levels. The accuracy in extreme cases is preferred since it improves the model utility in reducing significant food losses and carrying costs by better predicting daily and weekly inventory needs and refining production and employee schedules. Therefore, the model can become a powerful tool to increase the abilities of small business managers to recognize and mitigate extreme situations.

As mentioned before, the lack of more detailed information about each branded product demand required an alternative approach to predict the demand based on the average ratio of own branded products to daily sales and the average price and cost per kg. That was fundamental when evaluating how the ML model's ability to forecast revenue levels for production planning could impact food waste and other business metrics. Moreover, benchmark scenarios were used to support relative comparisons of the performance and impact of the selected metrics.

The comparison between the extreme scenarios (conservative and aggressive, see Table 8) draws the boundaries of the impact of two opposite strategies: food waste minimization versus sales maximization. On the one hand, guided by the conservative mindset, the production output of its own branded products would be only 4.54 kg per day, which would result in a daily cost of R$ 149.66, a daily revenue of R$ 281.18, and a daily gross profit of R$ 131.52, and no food waste. However, this approach would result in a daily sales loss of R$ 811.33 because of the unfulfilled demand. On the other hand, guided by the aggressive mindset, the production output of its own branded products would be 37.33 kg per day, which would result in a daily cost of R$ 1234.03, a daily revenue of R$ 1092.51, and a daily negative gross profit of R$ 141.52, and no daily sales loss. However, this approach would result in daily food waste of 19.77 kg. Therefore, the aggressive strategy results in considerable food waste, and although it maximizes revenue, it is not profitable.

Additionally, since this approach requires a production level of 8.24 times the conservative scenario, it demands more work, higher energy consumption, higher water volume, and, consequently, generates a higher carbon footprint (Scholz, Eriksson, & Strid, 2015), and financial. Moreover, the conservative approach uses only 12.13% of the aggressive production strategy to reach 25.74% of its sales level, resulting in a profitable operation and no food waste. Beyond the profitability and food waste, a much lower production level requires less energy, less work, and less water and produces a smaller carbon footprint (Eriksson, 2015).

There is an essential difference in the sustainability of both extreme strategies. Because the conservative approach results in a considerable sales loss, a business manager would try a more balanced strategy to capture more sales to increase revenue and profit. Thus, guided by the balanced strategy, the production output of own branded products would be 17.62 kg per day, which would result in a daily cost of R$ 581.50, a daily revenue of R$ 962.69, and a daily gross profit of R$ 381.20, and a daily food waste of 2.09 kg. Therefore, with almost half the aggressive strategy's production level (47.12%), the balanced approach reaches 88.12% of the maximum possible sales level, resulting in only 10.59% of food waste. Additionally, compared to the conservative strategy, the balanced approach results in almost three times its profitability (2.89%) and only 16% of its sales loss. Therefore, the balanced strategy seems to combine the desired business results with a more sustainable approach when compared to a strategy to fulfill all demands.

Finally, comparing the decisions on the production level based on AI and the balanced scenario shows the benefits of the proposed approach. Using the ML model to forecast the demand to drive the production level would result in an output of own branded products of 15.32 kg per day, which would result in a daily cost of R$ 505.66, a daily revenue of R$ 926.65, a daily gross profit of R$ 420.99, a daily sales loss of R$ 165.86, and a daily food waste of only 0.38 kg (380 g). Thus, at a production level of 86.96% of the balanced strategy, the MI approach reaches 96.26% of the balanced approach's sales level, resulting in only 18% food waste and 110.44% of its profitability. Therefore, the proposed approach brings a superior result for the business in both financial and social impact. There is a significant reduction in food waste, leading to a more sustainable operation from the perspective of resource usage combined with a superior business result. This result contributes to the efforts of several nations, including Brazil, to follow their commitment to the United Nations Sustainable Development Goal Target 12.3 (ECOSOC, 2019; Henz & Porpino, 2017). Considering the goal to reduce per capita global food waste at the retail and consumer levels by half and to reduce food losses along production and supply chains by 2030, the proposed ML model surpassed this goal by reducing food waste by almost 82% when compared to the balanced approach.

Data from Table 8 supported the impact evaluation for the entire 312-day period. The conservative approach resulted in a gross profit of R$ 41,034.14 and no food waste with a downside of R$ 253,135.18 lost sales. Conversely, the aggressive approach resulted in a negative gross profit of R$ 44,153.92, a total waste of 6,169.39 kg of food, and no sales lost. Between these extremes, the balanced approach would result in a gross profit of R$ 118,933.04, a total waste of 653.27 kg of food, and R$ 40,503.04 lost sales. Finally, the AI approach would result in a gross profit of R$ 131,349.13, a total waste of 117.64 kg of food, and R$ 51,748.26 in lost sales. Therefore, in addition to the expected differences between the extreme scenarios, using the proposed AI approach to guide the production scale would positively impact the business over the whole period, combining the highest gross profit with the lowest food waste.

Moreover, those results would be achieved with a lower operational cost. This approach also requires the use of much less food than the balanced approach (5,497.80 - 4,780.79 = 717.01), and this food could remain in the food supply chain and reduce the harmful byproducts of their production, such as carbon dioxide emissions. Additionally, operating with lower food levels would lower inventory costs and require less storage space, potentially increasing retail sales per square foot, a typical performance metric used in retail operations. Therefore, the results of the proposed approach can encourage production managers to adopt AI-based approaches to seek more sustainable and profitable operations.

Table 9 provides three performance ratios to compare and confirm the proposed approach's good performance. The ROI shows the financial performance of each operation based on distinct production strategies. The conservative production strategy could achieve the highest ROI (188%), while the aggressive production strategy would achieve the lowest ROI (89%). Although the balanced strategy provides an excellent ROI (166%), it is considerably lower than the one achieved by the proposed approach based on AI (183%). The proposed approach reaches an ROI comparable to the conservative approach. The FWP shows the food waste performance of each operation based on distinct production strategies. The best FWP could be achieved by the conservative production strategy as well (0%), while the worst FWP would be achieved by adopting the aggressive production strategy where more than half of the food production would be wasted (53%). Although the balanced strategy could provide a more reasonable FWP (12%), the proposed approach would reach a much better FWP (2%), which is also comparable to the conservative approach. Finally, the FD shows that the aggressive strategy would achieve the best performance since all the demand would be fulfilled (100%), while the worst FD performance would be achieved by the conservative approach, where only 26% of the total demand would be fulfilled. The balanced strategy would fulfill 88% of the total demand, while the proposed AI approach would fulfill 85% of the demand. Therefore, considering all the other benefits of the proposed approach, such as food waste reduction, workforce, space optimization, and profitability, not fulfilling 15% of the demand is reasonable.

Those comparisons can be visualized in Figure 3 (left), where the ratios from each distinct approach are plotted on a radar chart. Finally, the data from Table 9 were plotted using a radar chart, shown in Figure 3 (right), to provide a visual and comparative analysis of the distinct production strategies. The balanced strategy was considered the reference. Thus, the values for all ratios are 100%. The ratios for all other approaches were calculated using the reference approach. Using this chart, it is possible to visually conclude that the proposed approach combines the best results, such as a moderate production level, FD, and gross profit compared to the balanced approach, and food waste, FWP, and ROI comparable to the conservative approach. However, although the gross profit is more similar to the reference than any other approach, the gross profit using the proposed method exceeds the reference gross profit by 10%, which is a remarkable result considering its production level is 87% of the reference.

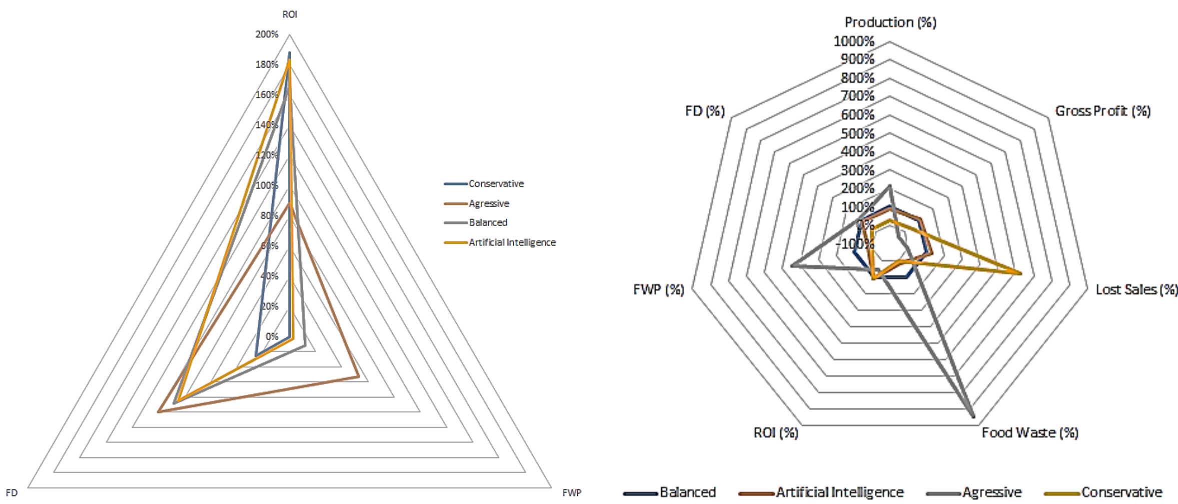

Figure 3: Absolute (left) and relative (right) ratios of each production planning strategy

The analyses show that the proposed AI approach can offer a promising alternative to improve sustainable small business operations. This approach can result in superior profitability and ROI with lower operational costs, production levels, and considerably lower food waste. Therefore, the proposed approach can reduce energy and water consumption, the required storage space used for raw ingredients and produced food, and, consequently, the operation's carbon footprint. Ultimately, adopting the proposed approach on a large scale could offer a promising way to reduce food waste significantly.

Conclusions

This study proposed an innovative use of a new AI-based method to reduce food waste in a small grocery store. This method relies on the prediction of its branded product demand to support the production planning using an exclusive dataset from a small grocery store focused on natural and organic products. The impact of limited data scarcity could be reduced by combining external datasets (secondary data, such as weather), performing manual feature expansion engineering, and resampling the data to balance the dataset.

Among five distinct ML techniques, RF was selected because of its superior accuracy. This technique correctly forecasted revenue levels for 90% of the days in the test period (312 days total). An alternative approach was used to convert the forecasted revenue levels into demand (in kg) of own branded products for daily production planning.

The results provide evidence that the proposed technique is promising. This technique achieved a low food waste level (380 g/day) and the highest gross profit (R$ 131,349.13) compared to the result achieved by the benchmarks. Additionally, the proposed method reduced the production cost while maintaining an outstanding level of demand fulfillment (85%) and ROI (183%). Moreover, the proposed approach reduced the total food waste for all 312 days from 6,169.39 kg (aggressive production strategy) and 653.27 kg (balanced production strategy, the best benchmark) to only 117.64 kg. Considering that the UN Sustainable Development Goal Target 12.3 aims to reduce food waste by half by 2030, the proposed approach achieved an impressive result of almost 82% food waste reduction compared to the best performing benchmark (balanced approach).

Data science techniques such as feature expansion and data augmentation addressed the dataset size and imbalance limitations. Furthermore, only a few attributes were included in the original database, which is an opportunity for future research. Hence, further research is recommended to evaluate alternative paths for addressing the described limitations. For example, merging with secondary data sources could increase the number of dataset features. Finally, future research with other food business types and sizes is recommended to deepen the available research on this topic.

References

Brinckmann, J., Grichnik, D., & Kapsa, D. (2010). Should entrepreneurs plan or just storm the castle? A meta-analysis on contextual factors impacting the business planning--performance relationship in small firms. *Journal of Business Venturing*, *25*(1), 24–40.

Chaboud, G., & Daviron, B. (2017). Food losses and waste: navigating the inconsistencies. *Global Food Security*, *12*, 1–7.

Cicatiello, C., Franco, S., Pancino, B., Blasi, E., & Falasconi, L. (2017). The dark side of retail food waste: Evidences from in-store data. *Resources, Conservation and Recycling*, *125*, 273–281. https://doi.org/https://doi.org/10.1016/j.resconrec.2017.06.010

de Brito Nogueira, T. B., da Silva, T. P. M., de Araújo Luiz, D., de Andrade, C. J., de Andrade, L. M., Ferreira, M. S. L., & Fai, A. E. C. (2020). Fruits and vegetable-processing waste: a case study in two markets at Rio de Janeiro, RJ, Brazil. *Environmental Science and Pollution Research*, *27*(15), 18530–18540. https://doi.org/10.1007/s11356-020-08244-y

ECOSOC, U. N. (2019). Special Edition: Progress towards the Sustainable Development Goals Report of the Secretary-General. *Advanced Unedited Version. New York (US): United Nations*.

Eriksson, M. (2015). *Supermarket food waste* (Vol. 2015).

Eriksson, M., Strid, I., & Hansson, P.-A. (2014). Waste of organic and conventional meat and dairy products—A case study from Swedish retail. *Resources, Conservation and Recycling*, *83*, 44–52. https://doi.org/https://doi.org/10.1016/j.resconrec.2013.11.011

Griffin, M., Sobal, J., & Lyson, T. A. (2009). An analysis of a community food waste stream. *Agriculture and Human Values*, *26*(1–2), 67–81.

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

Henz, G. P., & Porpino, G. (2017). Food losses and waste: how Brazil is facing this global challenge? *Horticultura Brasileira*, *35*(4), 472–482.

Lee, K. C. L. (2018). Grocery shopping, food waste, and the retail landscape of cities: The case of Seoul. *Journal of Cleaner Production*, *172*, 325–334. https://doi.org/https://doi.org/10.1016/j.jclepro.2017.10.085

Li, Z., Lu, H., Ren, L., & He, L. (2013). Experimental and modeling approaches for food waste composting: A review. *Chemosphere*, *93*(7), 1247–1257. https://doi.org/https://doi.org/10.1016/j.chemosphere.2013.06.064

McCarthy, B., & Liu, H. B. (2017). Food waste and the 'green' consumer. *Australasian Marketing Journal (AMJ)*, *25*(2), 126–132. https://doi.org/https://doi.org/10.1016/j.ausmj.2017.04.007

Mena, C., Adenso-Diaz, B., & Yurt, O. (2011). The causes of food waste in the supplier–retailer interface: Evidences from the UK and Spain. *Resources, Conservation and Recycling*, *55*(6), 648–658. https://doi.org/https://doi.org/10.1016/j.resconrec.2010.09.006

Muriana, C. (2017). A focus on the state of the art of food waste/losses issue and suggestions for future researches. *Waste Management*, *68*, 557–570. https://doi.org/https://doi.org/10.1016/j.wasman.2017.06.047

Nascimento, A.M., Cunha, M. A. V. ., Meirelles, F. S., Scornavacca, E., & Melo, V. V. (2018). A Literature Analysis of Research on Artificial Intelligence in Management Information System (MIS).

Nascimento, Alexandre Moreira, & Bellini, C. G. P. (2018). Artificial intelligence and industry 4.0: The next frontier in organizations. *BAR - Brazilian Administration Review*, *15*. Retrieved from http://www.scielo.br/scielo.php?script=sci_arttext&pid=S1807-76922018000400100&nrm=iso

Nascimento, Alexandre Moreira, de Melo, V. V., Queiroz, A. C. M., Brashear-Alejandro, T., & de Souza Meirelles, F. (2021). Artificial intelligence applied to small businesses: the use of automatic feature engineering and machine learning for more accurate planning. *Revista de Contabilidade e Organizações*, *15*, 1–15.

Orhan, U., Hekim, M., & Ozer, M. (2011). EEG signals classification using the K-means clustering and a multilayer perceptron neural network model. *Expert Systems with Applications*, *38*(10), 13475–13481.

Parfitt, J., Barthel, M., & Macnaughton, S. (2010). Food waste within food supply chains: quantification and potential for change to 2050. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *365*(1554), 3065–3081.

Parizeau, K., Massow, M. von, & Martin, R. (2015). Household-level dynamics of food waste production and related beliefs, attitudes, and behaviours in Guelph, Ontario. *Waste Management*, *35*, 207–217. https://doi.org/https://doi.org/10.1016/j.wasman.2014.09.019

Scholz, K., Eriksson, M., & Strid, I. (2015). Carbon footprint of supermarket food waste. *Resources, Conservation and Recycling*, *94*, 56–65. https://doi.org/https://doi.org/10.1016/j.resconrec.2014.11.016

Vieira, L. M., Barcellos, M. D. de, & Matzembacher, D. E. (2018). De um limão, uma limonada.

Vieira, L. M., Carvalho, Í. C. S. de, Ferraz, R. L., & Guimarães, C. M. C. (2017). Ações para redução de perda e desperd{\'\i}cio de alimentos na cadeia de hortifrutigranjeiros em São Paulo.

Vieira, L. M., & Matzembacher, D. E. (2020). How Digital Business Platforms Can Reduce Food Losses and Waste? In *Food Supply Chains in Cities* (pp. 201–231). Springer.