ISSN 1330-3651 (Print), ISSN 1848-6339 (Online) https://doi.org/10.17559/TV-20230117000232
Original scientific paper

Demand Forecasting for Food Production Using Machine Learning Algorithms: A Case Study of University Refectory

Mehmet ACI*, Derya YERGÖK

Abstract: Accurate food demand forecasting is one of the critical aspects of successfully managing restaurants, cafeterias, canteens, and refectories. This paper aims to develop demand forecasting models for a university refectory. Our study focused on the development of Machine Learning-based forecasting models which take into account the calendar effect and meal ingredients to predict the heavy demand for food within a limited timeframe (e.g., lunch) and without pre-booking. We have developed eighteen prediction models gathered under five main techniques. Three Artificial Neural Network models (i.e., Feed Forward, Function Fitting, and Cascade Forward), four Gauss Process Regression models (i.e., Rational Quadratic, Squared Exponential, Matern 5/2, and Exponential), six Support Vector Regression models (i.e., Linear, Quadratic, Cubic, Fine Gaussian, Medium Gaussian, and Coarse Gaussian), three Regression Tree models (i.e., Fine, Medium, and Coarse), two Ensemble Decision Tree (EDT) models (i.e., Boosted and Bagged) and one Linear Regression model were applied. When evaluated in terms of method diversity, prediction performance, and application area, to the best of our knowledge, this study offers a different contribution from previous studies. The EDT Boosted model obtained the best prediction performance (i.e., Mean Squared Error = 0,51, Mean Absolute Erro = 0,50, and *R* = 0,96).

Keywords: boosting; decision support systems; demand forecasting; machine learning; prediction algorithms

1 INTRODUCTION

Accurate demand forecasting plays a fundamental role in facilitating strategic organizational planning and decision-making processes. Given their crucial significance as a key determinant of business decisions, the accuracy and reliability of forecasts hold paramount importance. To estimate future demand, demand forecasting necessitates the utilization of historical demand data as well as appropriate forecasting methods. The initial phase involves the collection of relevant data about various influencing factors, while subsequent modeling and forecasting procedures require the implementation of suitable forecasting methods and models [1].

Demand forecasting plays a crucial role in providing essential insights for numerous industries [2-4]. Food demand forecasting is one of the critical issues for both businesses and sustainable development [5]. The food sector, driven by the imperative to address nutritional needs, encompasses a wide range of production activities. Food service providers such as catering companies, restaurants, cafeterias, food courts, and refectories at schools and workplaces generate substantial demand within the food industry. However, within the European Union, a staggering 88 million metric tons (Mt.) of food is wasted annually, amounting to 15 - 16% of the European Union's food value chain. More than The United Nations Environment Program reports that Turkey, in particular, discards over 7,7 Mt of food each year, positioning it as one of the foremost global contributors to food waste [6]. The incidence of food waste is not confined to specific geographic regions but is closely intertwined with a country's level of development [7]. Food waste gives rise to economic, environmental, and social challenges in numerous countries, underscoring the imperative of implementing strategies to curtail its occurrence [8]. The United Nations has incorporated Target 12,3 within its Agenda for Sustainable Development, which aims to achieve a 50% reduction in food waste by the year 2030. Likewise, the European Commission has made a commitment to combat food waste and has integrated Target 12.3 into its European Circular Economic Action Plan [9, 10].

Accurately estimating product and service volumes is the most efficient method for preventing food waste. Demand forecasting encompasses two primary approaches: qualitative and quantitative. The qualitative approach relies on the examination of past performance, expert opinions, and judgments, whereas the quantitative approach leverages historical data to project future outcomes using mathematical models [11]. Predictive Machine Learning Algorithms (MLAs) serve as invaluable tools in the demand prediction process. These methods autonomously analyze data relationships and trends, enabling future predictions based on present observations [12]. Unlike the particularities of individual sales managers, MLA methods exhibit adaptability, positioning them more effective in responding to fluctuations in data.

The goal of this study is to develop prediction models for food demand based on MLAs, such as Artificial Neural Network (ANN), Gaussian Process Regression (GPR), Support Vector Regression (SVR), Regression Tree, and Ensemble Decision Tree (EDT). Specifically, three ANN models, namely Feed Forward Neural Network (FFNN), Function Fitting Neural Network (FITNET), and Cascade Forward Neural Network (CFNN), were employed. Additionally, four GPR models (Rational Quadratic (RQ), Squared Exponential (SE), Matern 5/2, and Exponential), six SVR models (Linear, Quadratic, Cubic, Fine Gaussian, Medium Gaussian, and Coarse Gaussian), three Regression Tree models (Fine Tree, Medium Tree, and Coarse Tree), two EDT models (Boosted and Bagged), and one Linear Regression (LinR) model were designed and implemented. These models were trained and evaluated using a dataset obtained from a university refectory.

The results were compared using Mean Absolute Error (*MAE*), Mean Square Error (*MSE*), and Multiple Correlation Coefficient (*R*), commonly used metrics for prediction problems.

The contributions of this paper can be summarized as follows:

- The main purpose of the study is to provide a decision support tool to refectory managers and nutritionists to predict future menu demands. In this way, users will be able to create daily, weekly or monthly plans.

- Analyzing and forecasting the sales demand can improve short to medium-term production planning and decrease food waste. Our study presented more comprehensive research and better results than previous studies by analyzing the papers published in the last few years.

- Although food demand forecasting is a widely studied subject in the literature when the studies are evaluated in terms of method variety, forecast performance, and application area, as far as we know, such a study has not been proposed before.

- It is a difficult task to predict a large amount of food demand in a limited time without reservation, and in this study, MLA-based forecasting models that work with high accuracy are proposed, taking into account the calendar effect as well as the food ingredients.

Parameter selection and cross-validation in MLAs are essential in developing a generalized model and avoiding the memorization problem. The MLA models used in previous studies suffered from an overfitting problem, and therefore, different parametric versions of the MLAs were also included in our study.

2 RELATED WORK

The estimation of demand and potential food waste can aid food service providers in determining the appropriate quantity of ingredients and preparing an adequate number of meals. Excessive production, on one hand, can result in the unnecessary disposal of food and wastage of financial resources. Conversely, inadequate production may lead to shortages that adversely impact customer satisfaction and the overall experience of food service users [13]. A common approach to addressing these issues in food services is the implementation of demand planning strategies that rely on order reservations. Utilizing booking systems for pre-ordering and reserving meals is an effective method for demand planning in food services, as it can effectively mitigate overproduction and minimize food waste [14]. However, it is not customary to make lunch reservations in university refectories. Consequently, previous studies have mostly focused on the development of demand prediction models based on retrospective data, employing statistical and MLA techniques and tools.

According to Tsoumakas [15], MLAs possess superior computational capabilities and an enhanced capacity to handle additional variables, rendering them more effective and adaptable when compared to conventional statistical techniques for forecasting food sales. Bozkır and Sezer [16] developed a web-based tool called "ADEM" for menu forecasting using data mining techniques. ADEM serves as a decision support system, aiding cafeteria managers and nutritionists in predicting future menu demands and generating plans accordingly. The tool underwent testing with a dataset from Hacettepe University refectory spanning 44 months. The Decision Tree (DT) method was utilized during model creation, resulting in DT models achieving a Variance Accounted For (*VAF*) level of 80.78%. Xinliang and Dandan [17] introduced a prediction model for Jiao Tong University's restaurant sales, incorporating the Baidu index and meteorological factors. The model utilized a Back Propagation (BP) neural network for analysis and was compared with the time series estimation method to evaluate its performance. Kılıç et al. [18] conducted a study to predict the daily food consumption at a university's refectory. They utilized ANN, Support Vector Machines (SVM), and Logistic Regression (LR) algorithms in the WEKA tool, evaluating their performance based on Root Mean Square Error (*RMSE*). The research demonstrated that MLAs can effectively predict meal consumption, and the study also presented the performance of these methods for campuses with different profiles and scales. Pereira [19] proposed a study exploring the application of advanced data mining techniques, including Random Forest (RF), ANN, and SVM, to develop a system for predicting daily food consumption in a canteen. The study focused on estimating the number of meals needed for each meal type. The results showed that RF models performed well in predicting meat consumption, while SVR models were successful in predicting fish and vegetarian consumption. Yang and Sutrisno [20] conducted demand forecasting research for a franchise bakery in China that sells bread with a one-day storage period. The study involved collecting and analyzing data from over 10 million point-of-sale transactions. The findings revealed that both regression analysis and ANN techniques demonstrated that the number of sales in the initial hours could be utilized to predict the sales for the remaining hours of the day. Hast [21] evaluated multiple MLAs for estimating in-flight meal demand. Feature significance analysis was conducted on the dataset, and LR, SVR, XGBoost, and a Multi-Layer Perceptron (MLP) were selected as MLAs. *MAE* was used as the error metric, and the SVR model showed poorer performance in model fitting and prediction time compared to the other three models. Faezirad et al. [13] proposed an MLA-based model that incorporates students' reservations and show/no-show rates to reduce food waste at universities offering food subsidies. Cost analysis demonstrated that the model can reduce food waste volume by up to 79% and effectively control penalty and waste costs. Posch et al. [22] proposed a forecasting approach using two Bayesian generalized additive models: Negative Binomial and Normal. The models incorporate point-of-sales data from a staff canteen. The Negative Binomial model, which accounts for multi-seasonal effects and trend changes, outperformed other methods, achieving the best performance with *MAE* of 0,73. The approach was evaluated against well-established estimation methods using two datasets from a restaurant and a staff canteen. Woltmann et al. [23] proposed a forecast model based on MLAs that utilized a collection of derived features to predict the quantities and absolute counts of dish portions per day in a campus canteen. The researchers approached the problem by employing ANN, SVR, Gradient Boosting Regressor (GBR) with DT. The GBR model gave the best result with 20,2% Symmetric Mean Absolute Percentage Error (*SMAPE*).

Tab. 1 chronologically lists the previous studies given above. It becomes evident that there is a lack of comprehensive demand forecasting studies using MLAs for businesses focusing on high-volume food sales during specific meal periods, particularly lunch, becomes apparent. Such businesses include canteens, cafeterias, food courts, and school/workplace refectories. Nonetheless, it is noteworthy that the application of MLA-based models in those studies has also been restricted and limited in scope.

Table 1 Related studies on demand forecasting for food production titles

Reference	Methods	The Best Performance
Bozkır and Sezer [16]	DT	*VAF* = 80,78%
Xinliang and Dandan [17]	Time series, BP	*MAPE* = 11,68%
Kılıç et al. [18]	ANN, SVR, LR	*RMSE* = 11,96
Pereira [19]	RF, SVM, ANN	*R-squared* = 0,82
Yang and Sutrisno [20]	Regression Analysis, ANN	*MAE* = 3,5
Hast [21]	LR, SVR, XGBoost, MLP	*MAE* = 0,024
Faezirad et al. [13]	ANN	*R* = 0,75
Posch et al. [22]	Negative Binomial, Normal	*MAE* = 0,73
Woltmann et al. [23]	ANN, SVR, GBR	*SMAPE* = 20,2%

3 MATERIALS AND METHODS

This section presents the comprehensive exposition of the datasets and MLAs employed in the study.

3.1 The Dataset

The dataset used in the study consists of one-year daily menus and corresponding sales numbers obtained from the refectories of Mersin University. The dataset, including 2241 sales records collected via turnstile machines at the refectory, was employed for both the training and testing phases. Tab. 2 presents the variables and their respective attributes within the dataset. Before processing the data, textual values were transformed into numeric values. Additionally, a grouping procedure was executed to consolidate similar data, thereby enhancing the network's learning capability. Tab. 3 shows an illustrative example of the dataset after the conversion to numeric values. The dataset covers diverse features such as weekdays, main courses, side dishes, soups, and the number of individuals, among others. By training these features with various combinations and methods, the objective is to achieve the most accurate estimation of the number of individuals who will dine on a given day.

3.2 Methods

The following subsections provide a concise overview of the MLAs employed in the study.

3.2.1 ANN Models

Our study employed three ANN models, specifically FFNN, FITNET, and CFNN. The hyperparameters utilized in the study were selected empirically to identify the most effective predictive model. Tab. 4 shows an overview of the parameter details of the ANN models used. The process of adjusting the weights and biases within the network to optimize the fitting of input samples and their corresponding outputs is referred to as ANN training. This training is essential when the underlying relationship between inputs and outputs is not readily apparent. Through training, the network learns to recognize specific inputs and generate accurate outputs. The primary objective of the network training algorithm is to establish a mapping between input-output samples, facilitating the formation of a functional relationship [24].

FFNNs are widely recognized as the most prevalent neural network models, offering the capability to establish input-output relationships by adjusting connection weights within the network. Each neuron in the network receives signals from all nodes in the preceding layer and transmits the modified signal to nodes in the subsequent layer. An illustrative instance of a neuron can be observed in Fig. 1, wherein the inputs ($x_1 - x_n$), their associated weights ($w_1 - w_n$), a bias (b), and the application of the activation function f to the weighted summation of the inputs are demonstrated. The FFNN's knowledge is encapsulated in the weights of the integration function, initially assigned as random values drawn from a normal distribution. The initial state of the network involves processing input vectors through integration and activation functions within internal neurons, subsequently producing an output signal to the final layer. Given the random selection of weights, the output of the FFNN model initially falls short when compared to observed values. To address this, a learning sample is utilized, and the weights are iteratively adjusted with the objective of minimizing the error function. Various supervised learning algorithms have been developed to facilitate the weight adjustment process, although the backpropagation approach is the predominant method employed in most FFNN applications. This iterative approach involves propagating the error backward through the network, enabling efficient weight updates and overall improvement in the FFNN's predictive capabilities [25, 26].

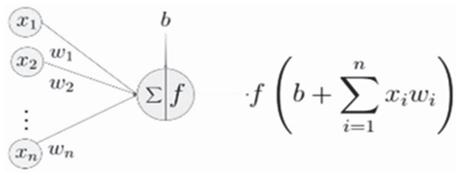


Figure 1 The schematic diagram of a neuron [25]

FITNET, a variant of FFNN, is specifically designed for the purpose of fitting input-output relationships. An FFNN with a hidden layer containing an adequate number of neurons can be employed to accurately model any finite input-output mapping problem. In FITNET, the default choice for the transfer function in the hidden layer is the tan-sigmoid function, while the output layer utilizes the pure linear transfer function [27].

The Cascade Feedforward Neural Network (CFNN) architecture consists of an input layer, one or more hidden layers, and an output layer. Within CFNN, each successive layer possesses weight and bias values. These weights are transmitted from the preceding layers to the subsequent layers until they reach the final layer, which is the output layer. Notably, unlike the FFNN, CFNN incorporates weight connections originating from the input layer and extending through the subsequent layers. This characteristic facilitates an enhanced speed of learning within the CFNN, emerging as a prominent distinction between CFNN and FFNN [28].

Table 2 The dataset attributes

Variable	Type	Usage	Description	Value
Days of the Week (A)	Continuous	Input	The name of the day ranging from Monday to Friday	From 1 to 5
Is Holiday? (B)	Boolean	Input	Is the day on the weekend/holiday or weekday?	0 or 1
Is Ramadan? (C)	Boolean	Input	Is the day on Ramadan day or not?	0 or 1
Is Exam Day? (D)	Boolean	Input	Is the day an exam day or not?	0 or 1
Is Salary Day? (E)	Boolean	Input	Is the day the salary day or not?	0 or 1
Calorie (F)	Continuous	Input	Total calorie amount of the menu	From 1 to 3 for the values between 800 and 1400
Soup (G)	Continuous	Input	Five types of soup (tomato, lentil etc.)	From 1 to 5
Main Course (H)	Continuous	Input	Beef, chicken, fish, vegetables, or legumes	From 1 to 5
Side Dish (I)	Continuous	Input	Five types of side dish (rice, pasta etc.)	From 1 to 5
Extra Dish (J)	Continuous	Input	Five types of extras (i.e. Yoghurt, fruit, salad, dessert or drink)	From 1 to 5
Number of Sales (K)	Continuous	Output	The number of sales in a lunch session.	To be predicted from 1 to 10 for the values between 0 and 10 000

Table 3 An example of a dataset converted to numeric values (column names are specified in Tab. 2)

A	B	C	D	E	F	G	H	I	J	K
2	1	0	0	0	3	3	1	1	1	1
3	0	0	1	0	3	4	2	1	2	3
4	0	0	1	0	1	5	1	2	3	4
5	0	0	0	0	2	3	1	1	1	4
1	0	0	0	0	2	3	1	1	2	7
2	0	0	0	0	3	1	2	1	1	8
3	0	0	0	0	2	1	1	2	1	9

Table 4 ANN model's hyper-parameters

Model Name	Input Layer	Hidden Layer(s)	Output Layer	Training Function	Transfer (Activation) Function
FFNN	# of layers: 1 # of neurons: 10	# of layers: 1 # of neurons: 5	# of layers: 1 # of neurons: 1	Bayesian Regularization	Hyperbolic Tangent Sigmoid Log-Sigmoid
FITNET		# of layers: 2 # of neurons: 20 / 20			
CFNN		# of layers: 3 # of neurons: 15 / 15 / 15			

3.2.2 GPR Models

Kernel-based and non-parametric GPR is a probabilistic approach that extends multivariate normal distributions to an infinite-dimensional space. Gaussian processes are utilized in statistical modeling, regression for multiple target values, and mapping analysis in higher dimensions [29].

A Gaussian Process (GP) is a collection of random variables in which any finite subset of variables follows a multivariate Gaussian distribution. Let $(X \times Y)$ denote the input and output domains, from which n independent and identically distributed pairs (x_i, y_i) are drawn. In the context of regression, assuming $y \subseteq R$, a GP over X is characterized by a mean function $\mu : X \to R$ and a covariance function $k : X \times X \to R$. The key assumption in GP regression is that the variable y is described by $y = f(x) + \xi$, where $\xi \sim N(0, \sigma^2)$ represents the distributional properties of y given the corresponding input x. The symbol "~" denotes statistical sampling. In GP regression, for each input x, there exists a corresponding random variable $f(x)$ that represents the value of the stochastic function f at that specific location. This study assumes that the observational error ξ follows a normal distribution, is independently and identically distributed, has a mean of $zero(\mu(x) = 0)$, a variance of σ^2, and $f(x)$ is drawn from the Gaussian process over X defined by the covariance function k. In other words, the relationship can be expressed as $Y = (y_1, ..., y_n) \sim N(0, K + \sigma^2 I)$, where $K_{ij} = k(x_i, x_j)$, and I denote the identity matrix.

There are four models available, each using a different kernel: RQ, SE, Matern 5/2, and Exponential. The RQ kernel is particularly useful for capturing data that exhibits changes at multiple scales. The SE kernel represents a function space version of a radial basis function regression model, where the inner products of fundamental functions are replaced with kernels. This approach ensures that large datasets processed in higher dimensions do not result in significant errors and enables effective handling of discontinuities. The Matern 5/2 kernel creates Fourier transforms of the Radial Basis Function kernel by incorporating spectral densities of the stationary kernel. On the other hand, the Exponential GPR differs from the SE GPR in that it does not square the Euclidean distance. Instead, slower kernels are used to replace the inner products of basic functions. The Exponential GPR performs better in handling discontinued functions, exhibiting fewer errors compared to the SE GPR [29, 30].

3.2.3 SVR Models

SVR aims to find a function that best fits the data by constructing a hyperplane in a higher-dimensional feature space. The SVR involves minimizing the empirical risk while simultaneously controlling the complexity of the model [31].

Given a training dataset consisting of input-output pairs $\{(x_1, y_1), (x_2, y_2), ..., (x_n, y_n)\}$, where x_i represents the d-dimensional input vector and y_i corresponds to the target output value, SVR aims to identify the function $f(x)$ that best approximates y. The SVR optimization problem can be mathematically formulated as follows: minimize $\frac{1}{2}\|w\|^2 + C\sum\left(\xi_i + \xi_i^*\right)$ subject to $y_i - f\left(x_i\right) \le \varepsilon + \xi_i^*$ and $f\left(x_i\right) - y_i \le \varepsilon + \xi_i$ and ξ_i , $\xi_i^* \ge 0$. In this formulation: w represents the weight vector that defines the hyperplane. C is a regularization parameter controlling the balance between the model's complexity and the deviation from the training data. ξ_i and ξ_i^* are slack variables denoting the margin violations. ε denotes the epsilon-insensitive tube, which defines the allowable margin of error in the regression task.

The objective of the optimization problem is to minimize the norm of the weight vector $\|w\|$, which corresponds to the complexity of the model, while also minimizing the sum of the slack variables subject to the constraints that ensure the deviations from the training data do not exceed ε.

Once the optimization problem is solved, the function $f(x)$ can be obtained by evaluating the input vector x with the weight vector w and applying a suitable kernel function. The effectiveness of SVR lies in the flexibility of kernel functions, which enable a wide range of solution space searches. However, selecting an appropriate kernel function is a challenge in SVR algorithms. Typically, Gaussian kernels tend to yield improved outcomes [32]. In our study, six different core SVR models were employed, including linear, quadratic, cubic, fine Gaussian, medium Gaussian, and coarse Gaussian.

3.2.4 Regression Tree Models

A regression tree is an iterative method that constructs trees for predicting categorical predictor variables (classification) or continuous dependent variables (regression). It recursively partitions subsets of the dataset by splitting them using all available predictor variables, creating two child nodes at each step. The objective is to generate subsets of the dataset that are as internally consistent as possible in relation to the target variable [33].

For a given set of input variables $X = \left\{x_1, x_2, ..., x_p\right\}$, where p is the number of input variables, and the corresponding target variable y, the regression tree aims to learn a mapping from X to y.

The regression tree algorithm recursively partitions the input space into regions, represented by the internal nodes of the tree, based on the values of the input variables. Each internal node tests a specific condition on a particular input variable to determine the splitting criterion. The splitting criterion can be expressed as: $x_i \le t$ (for binary split) or $x_i \in R$ (for multi-way split) where x_i is the value of the selected input variable, t is the threshold value, and R represents a specific range or region. The leaves of the tree represent the final output values or predictions. For a regression tree, the predicted value at each leaf node is typically the mean or median value of the target variable within that leaf node [34].

In our study, three types of Regression Trees: Fine Tree, Medium Tree, and Coarse Tree were utilized. The Fine Tree model offers a high level of interpretability and flexibility. It consists of numerous small leaves, enabling a highly flexible response function, with a minimum leaf size of 4. The Medium Tree model is also easy to interpret, with a moderate level of flexibility. It includes medium-sized leaves, resulting in a less flexible response function, and a minimum leaf size of 12. On the other hand, the Coarse Tree model maintains interpretability but has low flexibility. It comprises several large leaves, leading to a rough response function, and a minimum leaf size of 36.

3.2.5 EDT Models

EDT Bagged and EDT Boosted are recommended methods for improving the prediction capacity of decision trees. These techniques aim to enhance the predictive power of decision trees through the use of ensemble methods [35]. EDT Bagged is a combination model that merges the bagging algorithm with decision trees.

Let's assume $x = \left\{x_1, x_2, ..., x_n\right\}$ represents a vector of input variables (where n is the total number of input variables). The averaging of predictions, achieved through the bootstrap aggregation utilized in EDT Bagged, can be defined as follows [35, 36]:

$$f_{bag}\left(x\right) = \frac{1}{N}\sum_{i=1}^{N} f_i^*\left(x\right) \tag{1}$$

Eq. (1) represents a specific expression or formula, where N represents the number of diverse training datasets created through bootstrapping. The predictors $f_i^*\left(x\right)$ are defined within this context.

EDT Boosted is a regression model that follows an additive approach, where individual terms are represented by simple trees and fitted through forward, stage wise estimation. It is widely recognized as an ensemble method that combines regression trees and boosting techniques [37].

Let's consider $x = \left\{x_1, x_2, ..., x_n\right\}$ as a vector of predictors (with n being the number of predictors) and y as the response variable. The training process of the EDT Boosted algorithm is defined by the following function:

$$f\left(x\right) = \sum_k f_k\left(x\right) = \sum_k \alpha_k t\left(x, y_k\right) \tag{2}$$

In Eq. (2), α_k denotes the weights assigned to the nodes of each tree, y_k represents the split variables, and (x, y_k) represents individual DTs.

3.2.6 The LinR Model

LinR is a statistical method used to predict the outcome of a dependent variable based on a set of explanatory variables. The primary objective of LinR is to analyze and model the linear relationship between the independent variables (x) and the dependent variable (y). The equation of a linear regression line takes the form $Y = a + bX$, where X represents the explanatory variable and Y represents the dependent variable. The slope of the line is denoted by b, while a represents the intercept, indicating the value of Y

when X equals zero [38]. LinR was used in this study to prove statistical significance. The model was trained and tested using the Scikit-Learn library in the Python environment.

3.3 Performance Metrics

After training and testing all datasets, a performance evaluation was conducted to determine the best MLA model. Various commonly used statistical approaches were employed for this purpose, including the following:

$$MSE = \frac{1}{n}\left[\sum_{i=1}^{n}(O_i - P_i)^2\right] \quad (3)$$

$$MAE = \frac{1}{n}\left[\sum_{i=1}^{n}\left|(O_i - P_i)\right|\right] \quad (4)$$

$$R = \sqrt{R^2} = \sqrt{1 - \frac{\sum_{i=1}^{n}(O_i - P_i)^2}{\sum_{i=1}^{n}(O_i - P_m)^2}} \quad (5)$$

where n is the number of data used for testing, P_i is the predicted value, O_i is the observed value and O_m is the average of the observed values.

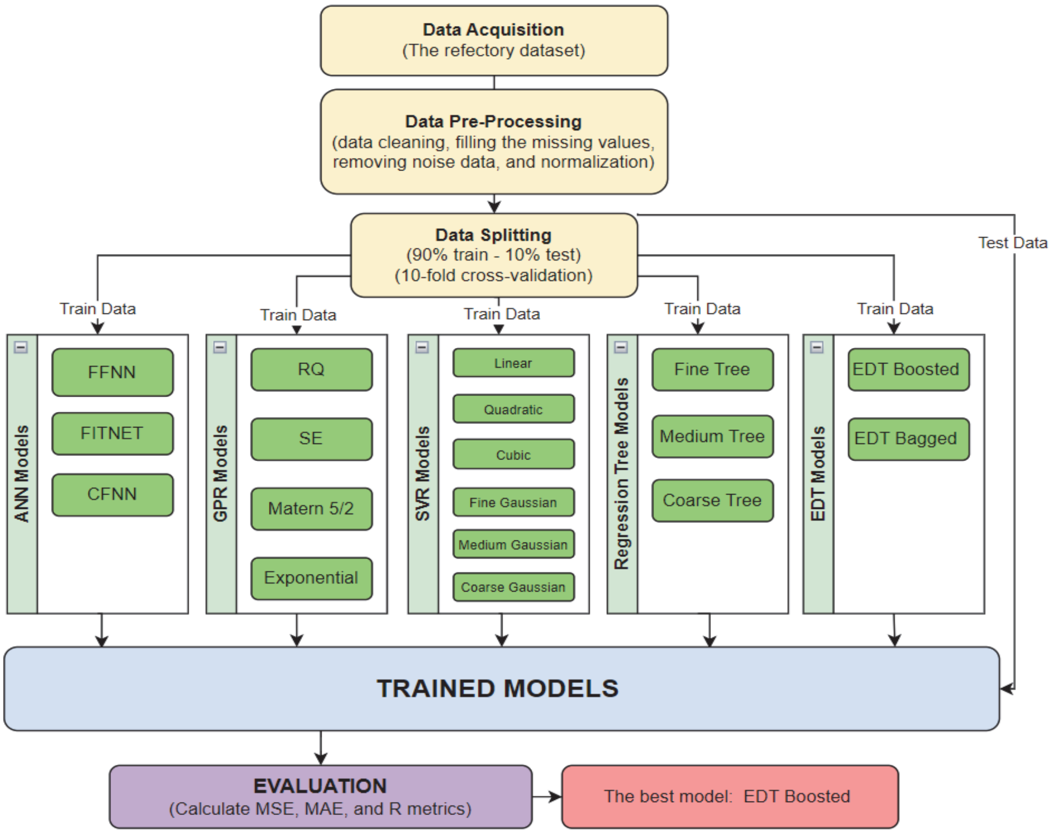


Figure 2 Block diagram of the MLA-based prediction models

MSE (Eq. (3)) is a metric that measures the average of squared differences between estimated and actual values. It quantifies the overall difference between the predicted and true values. *MAE* (Eq. (4)) calculates the average absolute difference between actual and predicted values. It is a robust metric that can handle outliers effectively, making it suitable for evaluating regression models. The correlation coefficient, denoted as *R*-value (Eq. (5)), is directly associated with the coefficient of determination, *R*-squared, in a clear manner. *R*-squared is a statistical measure that indicates the proportion of variance in the dependent variable explained by the independent variable(s) in a regression model. It measures the goodness of fit and represents the model's performance. Higher values of *R*-value indicate a better fit and performance, ranging from 0 to 1 [39]. These metrics serve as common benchmarks for evaluating the prediction performance of MLA models. They enable comparison with previous studies and provide a significant statistical magnitude for assessing the performance of developed models.

4 RESULTS AND DISCUSSIONS

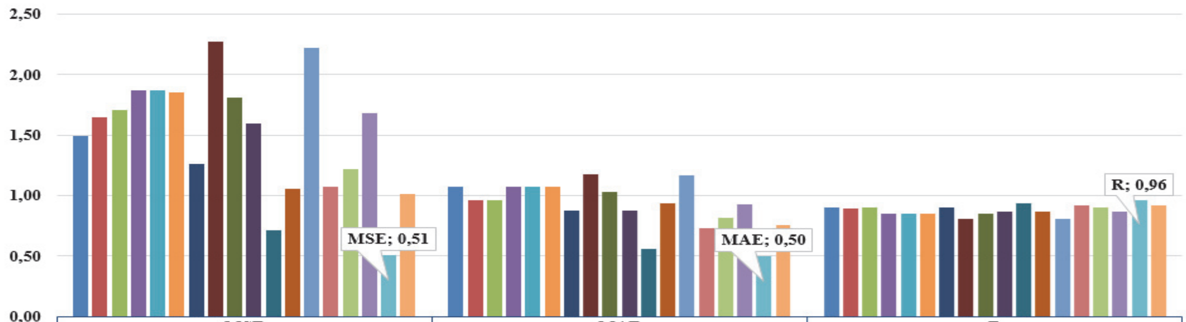
Fig. 2 shows block diagram of the MLA-based demand prediction models. The MLA models employed in the study underwent training and evaluation using a 10 - fold cross-validation technique, with 90% of the data allocated for training and 10% for testing. This process was implemented using MATLAB's Statistics and Machine Learning Toolbox, a widely used software tool in the field of data analysis [40]. In 10 - fold cross-validation, the original dataset is randomly partitioned into ten equal-sized sub-samples. From these sub-samples, one is selected as the test data while the remaining nine are designated as training data. This process is repeated ten times, with each iteration selecting a different sub-sample as the test data.

The purpose is to assess the model's performance by using each sub-sample as validation data once [41].

The dataset was pre-processed using well-known techniques such as data cleaning, filling the missing values, removing noise data, and normalization before applying MLA-based prediction models. The performance results of the models are presented in Fig. 3.

When we analyzed the results, it was noted that the EDT Boosted model gave the best prediction performance (i.e., *MSE* = 0,51, *MAE* = 0,50, *R* = 0,96). The Fine Gaussian (i.e., *MSE* = 0,71, *MAE* = 0,56, *R* = 0,94) can be considered as the second-best and EDT Bagged (i.e., *MSE* = 1,01, *MAE* = 0,76, *R* = 0,92) and Fine Tree (i.e., *MSE* = 1,07, *MAE* = 0,73, *R* = 0,92) both can be considered as the third-best models. The results they yielded are very close to each other.

We can rank other models according to the prediction successes; FFNN (i.e., *MSE* = 1,49, *MAE* = 1,07, *R* = 0,90), CFNN (i.e., *MSE* = 1,71, *MAE* = 0,96, *R* = 0,90), Exponential (i.e., *MSE* = 1,26, *MAE* = 0,88, *R* = 0,90), Medium Tree (i.e., *MSE* = 1,22, *MAE* = 0,82, *R* = 0,90), FITNET (i.e., *MSE* = 1,65, *MAE* = 0,96, *R* = 0,89), Cubic (i.e., *MSE* = 1,60, *MAE* = 0,88, *R* = 0,87), Medium Gaussian (i.e., *MSE* = 1,06, *MAE* = 0,94, *R* = 0,87), Coarse Tree (i.e., *MSE* = 1,68, *MAE* = 0,93, *R* = 0,87), RQ (i.e., *MSE* = 1,87, *MAE* = 1,07, *R* = 0,85), SE (i.e., *MSE* = 1,87, *MAE* = 1,07, *R* = 0,85), Matern 5/2 (i.e., *MSE* = 1,85, *MAE* = 1,07, *R* = 0,85), Quadratic (i.e., *MSE* = 1,81, *MAE* = 1,03, *R* = 0,85), Linear (i.e., *MSE* = 2,27, *MAE* = 1,18, *R* = 0,81), Coarse Gaussian (i.e., *MSE* = 2,22, *MAE* = 1,17, *R* = 0,81), Linear Regression (i.e., *MSE* = 6,20, *MAE* = 4,88, *R* = 0,45).



	MSE	MAE	R
FFNN	1,49	1,07	0,90
FITNET	1,65	0,96	0,89
CFNN	1,71	0,96	0,90
RQ	1,87	1,07	0,85
SE	1,87	1,07	0,85
Matern 5/2	1,85	1,07	0,85
Exponential	1,26	0,88	0,90
Linear	2,27	1,18	0,81
Quadratic	1,81	1,03	0,85
Cubic	1,60	0,88	0,87
Fine Gaussian	0,71	0,56	0,94
Medium Gaussian	1,06	0,94	0,87
Coarse Gaussian	2,22	1,17	0,81
Fine Tree	1,07	0,73	0,92
Medium Tree	1,22	0,82	0,90
Coarse Tree	1,68	0,93	0,87
EDT Boosted	0,51	0,50	0,96
EDT Bagged	1,01	0,76	0,92

Figure 3 The performance results of the prediction models

4.1 Comparisons with Previous Studies

Although there are many studies on food demand forecasting, the research conducted for dining areas, including refectories and cafeterias, within establishments facing considerable food demand within short time periods, such as during lunch breaks and without reservation, is very limited.

Bozkır and Sezer [16] used *VAF* as a performance metric in their study. *VAF* is a metric employed to assess the efficacy of a regression model by quantifying the extent to which the variance in the predicted values contributes to explaining the total variance in the actual values. Since it is a metric in the same category as *R*-squared and *R*-value, it can be said that the estimation performance of this study remained at 80,78%. The error rates observed in the studies conducted by Xinliang and Dandan [17], Kılıç et al. [18], Yang and Sutrisno [20], and Woltmann et al. [23] are notably elevated compared to the model errors observed in our study. Although Hast [21] study shows an exceptionally low *MAE* value, it is important to note that a significant proportion of the estimated flight meal requests in this study are made through pre-booking. This particular factor directly affects the prediction accuracy of the model. While Hast's study has a remarkably low *MAE*, we find that the majority of the estimated flight meal requests in this study are pre-booked. This is a factor that directly affects the prediction accuracy of the model. None of the training data in our study includes pre-booking.

In relation to prediction performance and model design, the studies of Faezirad et al. [13] and Posch et al. [22] closely resemble this study. A comparative evaluation can be made between Faezirad et al. [13]'s study in terms of the *R*-value, Posch et al. [22] study in terms of *MAE*, and the best performance results of our study. Fig. 4 clearly illustrates that the proposed study outperformed the existing literature in terms of both metrics. Moreover, our study encompasses a more comprehensive approach, incorporating model diversity and employing MLAs with different architectures (ANN, trees, etc.).

The linear regression model, which is the only statistical method among the models, gave the lowest performance results, as expected. These results prove the superiority of MLA-based algorithms over traditional statistics-based algorithms. The primary finding of this study lies in the validation of the efficacy of DT-based models for regression problems. Despite SVR and ANN-based algorithms being commonly employed in regression problems within the existing literature, they failed to surpass the performance of DT-based algorithms in this study.

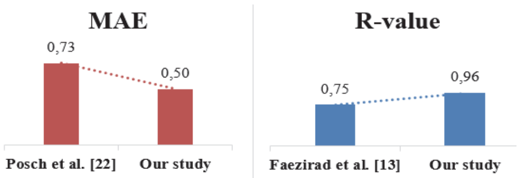


Figure 4 Comparison with the previous studies

5 CONCLUSIONS

In this study, MLA-based models that estimate the daily demand for mass food places using university refectory data are presented. The main purpose of the study is to compare the methods not used in the literature and to reach the best result by using the real data of the university refectory and more than one MLA. In the study, it was tried to obtain meaningful data by applying feature extraction and feature selection processes to the menu taken from the university refectory. Multiple measures and metrics were applied to the models to assess their applicability and performance. In this context, five types of prediction models (i.e., ANN, GPR, SVR, Regression Tree, and EDT) with different hyper-parameters have been analyzed. The conclusion follows that all four evaluated prediction models except EDT performed with results close to each other in terms of prediction accuracy and with no significant difference against each other except the Fine Gaussian model. However, the EDT Boosted model achieved two times more successful results than most of the models according to *MSE* and *MAE* metrics.

This study serves as a foundation for further research for evaluating multiple regression MLAs and predicting customer demand. However, we have identified several ideas to refine our approach, stemming directly from the identified limitations within our research. Firstly, we intend to integrate a weather feature as an additional component. Secondly, we aim to assess the applicability of our approach in refectories situated within distinct contexts, such as factory canteens. Lastly, we could use various deep learning methods to further broaden the scope and potential contributions of our present study.

Acknowledgments

This research was supported by the Mersin University Scientific Research Projects Department under Grant No. 2019-1-TP2-3215.We would like to express our gratitude to the Department of Health, Culture, and Sports at Mersin University for providing the refectory dataset used in this study.

6 REFERENCES

[1] Punia, S. & Shankar, S. (2022). Predictive analytics for demand forecasting: A deep learning-based decision support system. *Knowledge-Based Systems*, *258*, 109956. https://doi.org/10.1016/J.KNOSYS.2022.109956
[2] Wang, T. (2021). A k-means group division and LSTM based method for hotel demand forecasting. *Tehnički vjesnik*, *28*(4), 1345-1352. https://doi.org/10.17559/TV-20210507172841
[3] Tozan, H., Karatas, M., & Vayvay, O. (2018). Reducing demand signal variability via a quantitative fuzzy grey regression approach. *Tehnički vjesnik*, *25*(Supplement 2), 411-419. https://doi.org/10.17559/TV-20171115130250
[4] Chang, D., Wang, Y., & Fan, R. (2022). Forecast of large earthquake emergency supplies demand based on PSO-BP neural network. *Tehnički vjesnik*, *29*(2), 561-571. https://doi.org/10.17559/TV-20211120092137
[5] Lutoslawski, K., Hernes, M., Radomska, J., Hajdas, M., Walaszczyk, E., & Kozina, A. (2021). Food demand prediction using the nonlinear autoregressive exogenous neural network. *IEEE Access*, *9*, 146123-146136. https://doi.org/10.1109/ACCESS.2021.3123255
[6] Forbes, H., Quested, T., & O'Connor, C. (2021). Food waste index report. United Nations Environment Programme. https://www.unep.org/resources/report/unep-food-waste-index-report-2021
[7] Woolley, E., Jellil, A., & Simeone, A. (2021). Wasting less food: Smart mass customisation of food provision. *Procedia CIRP*, *96*, 189-194. https://doi.org/10.1016/J.PROCIR.2021.01.073
[8] Caldeira, C., De Laurentiis, V., Ghose, A., Corrado, S., & Sala, S. (2021). Grown and thrown: Exploring approaches to estimate food waste in EU countries. *Resources, Conservation and Recycling*, *168*, 105426. https://doi.org/10.1016/J.RESCONREC.2021.105426
[9] Commission, E. (2015). Closing the loop - An EU action plan for the circular economy. https://eur-lex.europa.eu/resource.html?uri=cellar:8a8ef5e8-99a0-11e5-b3b7-01aa75ed71a1.0012.02/DOC_1&format=PDF
[10] Nations, U. (2015). Sustainable development goals. https://www.undp.org/sustainable-development-goals/life-on-land?gclid=CjwKCAjw-7OlBhB8EiwAnoOEk6WCBxBa5c0vWsQ9dfgXPlYO_3-5_HH3zHtsRlbc_yRMP7MAKDDVPBoCeagQAvD_BwE
[11] Meneghini, M., Anzanello, M., Kahmann, A., & Tortorella, G. (2018). Quantitative demand forecasting adjustment based on qualitative factors: Case study at a fast food restaurant. *Systems & Management*, *13*(1), 68-80. https://doi.org/10.20985/1980-5160.2018.v13n1.1188
[12] Van der Aalst, W. (2016). *Data science in action*. In Process Mining, 3-23. Berlin: Springer Berlin Heidelberg. https://doi.org/10.1007/978-3-662-49851-4_1
[13] Faezirad, M., Pooya, A., Naji-Azimi, Z., & Amir Haeri, M. (2021). Preventing food waste in subsidy-based university dining systems: An artificial neural network-aided model under uncertainty. *Waste Management & Research*, *39*(8), 1027-1038. https://doi.org/10.1177/0734242X211017974
[14] Steen, H., Malefors, C., Röös, E., & Eriksson, M. (2018). Identification and modelling of risk factors for food waste generation in school and pre-school catering units. *Waste Management*, *77*, 172-184. https://doi.org/10.1016/J.WASMAN.2018.05.024
[15] Tsoumakas, G. (2019). A survey of machine learning techniques for food sales prediction. *Artificial Intelligence Review*, *52*(1), 441-447. https://doi.org/10.1007/s10462-018-9637-z
[16] Bozkır, A. S. & Akçapınar, E. S. (2013). ADEM: An online decision tree based menu demand prediction tool for food

courts. *4th International Conference on Food Engineering and Biotechnology*, 1-6. https://doi.org/10.7763/IPCBEE

[17] Xinliang, L. & Dandan, S. (2017). University restaurant sales forecast based on BP neural network in Shanghai Jiao Tong University case. *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, *10386 LNCS*, 338-347. https://doi.org/10.1007/978-3-319-61833-3_36

[18] Kılıç, F., Akkaya, M. R., & Memili, N. (2018). Daily demand forecast using artificial intelligence techniques for refectory. *European Journal of Science and Technology*, *1*(13), 65-71. https://doi.org/10.31590/ejosat.397549

[19] Pereira, D. X. R. (2018). Going zero waste in canteens: Exploring food demand with data analytics. *Universidade do Porto, Porto*.

[20] Yang, C. L. & Sutrisno, H. (2018). Short-term sales forecast of perishable goods for franchise business. *10th International Conference on Knowledge and Smart Technology: Cybernetics in the Next Decades*, *KST 2018*, 101-105. https://doi.org/10.1109/KST.2018.8426091

[21] Hast, M. (2019). Evaluation of machine learning algorithms for customer demand prediction of in-flight meals. *KTH Royal Institute of Technology*.

[22] Posch, K., Truden, C., Hungerländer, P., & Pilz, J. (2022). A Bayesian approach for predicting food and beverage sales in staff canteens and restaurants. *International Journal of Forecasting*, *38*(1), 321-338. https://doi.org/10.1016/J.IJFORECAST.2021.06.001

[23] Woltmann, L., Drechsel, J., Hartmann, C., & Lehner, W. (2022). Ingredient-based forecast of sold dish portions in campus canteen kitchens. *International Conference on Data Engineering Workshops*, 111-116. https://doi.org/10.1109/ICDEW55742.2022.00023

[24] Yang, J. & Ma, J. (2019). Feed-forward neural network training using sparse representation. *Expert Systems with Applications*, *116*, 255-264. https://doi.org/10.1016/J.ESWA.2018.08.038

[25] Minemoto, T., Isokawa, T., Nishimura, H., & Matsui, N. (2017). Feed forward neural network with random quaternionic neurons. *Signal Processing*, *136*, 59-68. https://doi.org/10.1016/J.SIGPRO.2016.11.008

[26] Gerlitz, L., Conrad, O., & Böhner, J. (2015). Large-scale atmospheric forcing and topographic modification of precipitation rates over High Asia - A neural-network-based approach. *Earth System Dynamics*, *6*(1), 61-81. https://doi.org/10.5194/ESD-6-61-2015

[27] Das Gupta, A. (2011). *Useful tools for statistics and machine learning*, Springer texts in statistics. New York: Springer. https://doi.org/10.1007/978-1-4419-9634-3_20

[28] Narmandakh, D., Butscher, C., Doulati Ardejani, F., Yang, H., Nagel, T., & Taherdangkoo, R. (2023). The use of feed-forward and cascade-forward neural networks to determine swelling potential of clayey soils. *Computers and Geotechnics*, *157*, 105319. https://doi.org/10.1016/J.COMPGEO.2023.105319

[29] Zhang, N. & Leatham, K. (2018). Neurodynamics-based nonnegative matrix factorization for classification. *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, *11302 LNCS*, 519-529. https://doi.org/10.1007/978-3-030-04179-3_46

[30] Pal, M. & Deswal, S. (2010). Modelling pile capacity using Gaussian process regression. *Computers and Geotechnics*, *37*(7-8), 942-947. https://doi.org/10.1016/J.COMPGEO.2010.07.012

[31] Panahi, M., Sadhasivam, N., Pourghasemi, H. R., Rezaie, F., & Lee, S. (2020). Spatial prediction of groundwater potential mapping based on convolutional neural network (CNN) and support vector regression (SVR). *Journal of Hydrology*, *588*, 125033. https://doi.org/10.1016/J.JHYDROL.2020.125033

[32] Alipour, M., Tavallaey, S. S., Andersson, A. M., & Brandell, D. (2022). Improved battery cycle life prediction using a hybrid data-driven model incorporating linear support vector regression and Gaussian. *Chem. Phys. Chem.*, *23*(7), e202100829. https://doi.org/10.1002/CPHC.202100829

[33] Chen, W., Xie, X., Wang, J., Pradhan, B., Hong, H., Bui, D. T., Duan, Z., & Ma, J. (2017). A comparative study of logistic model tree, random forest, and classification and regression tree models for spatial prediction of landslide susceptibility. *CATENA*, *151*, 147-160. https://doi.org/10.1016/J.CATENA.2016.11.032

[34] Yang, L., Liu, S., Tsoka, S., & Papageorgiou, L. G. (2017). A regression tree approach using mathematical programming. *Expert Systems with Applications*, *78*, 347-357. https://doi.org/10.1016/J.ESWA.2017.02.013

[35] Ly, H. B., Monteiro, E., Le, T. T., Le, V. M., Dal, M., Regnier, G., & Pham, B. T. (2019). Prediction and sensitivity analysis of bubble dissolution time in 3D selective laser sintering using ensemble decision trees. *Materials*, *12*(9), 1544. https://doi.org/10.3390/MA12091544

[36] Winkler, D., Haltmeier, M., Kleidorfer, M., Rauch, W., & Tscheikner-Gratl, F. (2018). Pipe failure modelling for water distribution networks using boosted decision trees. *Structure and Infrastructure Engineering*, *14*(10), 1402-1411. https://doi.org/10.1080/15732479.2018.1443145

[37] Zięba, M., Tomczak, S. K., & Tomczak, J. M. (2016). Ensemble boosted trees with synthetic features generation in application to bankruptcy prediction. *Expert Systems with Applications*, *58*, 93-101. https://doi.org/10.1016/J.ESWA.2016.04.001

[38] Hope, T. M. H. (2020). Linear regression. *Machine Learning: Methods and Applications to Brain Disorders*. 67-81. https://doi.org/10.1016/B978-0-12-815739-8.00004-3

[39] Pessoa, C. E., Hugo, V., Silva, P., & Stefani, R. (2023). Prediction of self-healing properties of concrete modified with bacteria and fibers using Machine Learning. *Asian Journal of Civil Engineering*, *7*(25), 1-16. https://doi.org/10.21203/RS.3.RS-3133577/V1

[40] MATLAB. (2021). Statistics and machine learning toolbox. https://www.mathworks.com/help/stats/index.html?s_tid=CRUX_lftnav

[41] Karayiğit, H., İnan Acı, Ç., & Akdağlı, A. (2021). Detecting abusive Instagram comments in Turkish using convolutional Neural network and machine learning methods. *Expert Systems with Applications*, *174*, 114802. https://doi.org/10.1016/J.ESWA.2021.114802

Contact information:

Mehmet ACI, Assistant Professor, PhD
(Corresponding author)
Mersin University, Faculty of Engineering,
Department of Computer Engineering,
Çiftlikköy Campus, 33343 Mersin, Turkey
E-mail: maci@mersin.edu.tr

Derya YERGOK, Student, MsC
Mersin University, Faculty of Engineering,
Department of Computer Engineering,
Çiftlikköy Campus, 33343 Mersin, Turkey
E-mail: deryayergok@gmail.com