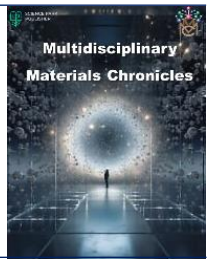
Multidisciplinary Materials Chronicles, 2025, Vol. X, Iss. X, 62-79

DOI: 10.62184/mmc.jmmc110020251

Research Article

SCIENCE PARK PUBLISHER

Multidisciplinary Materials Chronicles



Application of supervised machine learning approaches for predicting household electricity consumption in Alexandria, Egypt

Nourhan H. Farag*, Mohamed A. Abdrabo, Mohamed A. El-Iskandarani

Institute of Graduate Studies and Research - Alexandria University, 163 Horreya Avenue, Chatby, Alexandria, Egypt

Received: 10, 12, 2024; Accepted: 29, 01, 2025; Published: 06, 03, 2025

Abstract

The electricity consumption in the residential sector, which currently represents about 27% of the world's electricity consumption, has been experiencing continued growth due to economic and population growth. Therefore, proper management of future electricity provision entails predicting consumption patterns in the future to address increasing demand. For this purpose, Machine learning algorithms can support proper management of supply and demand for electricity through providing more accurate predictions of electricity consumption. However, the scarcity of data on electricity consumption patterns and their determinants, is one of the challenges that may restrict the potential of ML in predicting electricity consumption. This paper is intended to develop a machine learning based approach for predicting electricity consumption in the residential sector at the household level in Alexandria, Egypt, under data scarcity. For this purpose, the oversampling technique is applied to overcome data scarcity. It is found that bagging classifier, decision tree classifier, random forest classifier, and gradient boosting classifier have the highest performance with average accuracy exceeding 80%. This indicates that supervised machine learning algorithms that are tree-based structure gave higher accuracies for predicting seasonal household electricity consumption in the residential sector. Random forest is selected to develop an ML model for predicting electricity consumption at household level.

Keywords: Machine learning, electricity consumption determinants, oversampling, feature selection

1. Introduction

Continued technological and economic development, entails an increase in electricity demand [1]. Meeting such increasing demand implies expanding electricity generation capacities, which means burning more fossil fuels, which typically has negative environmental impacts, including different types of pollution as well as CO_2 emissions that contribute to climate change [2]. Due to the difficulty of electricity storage, proper management of the electricity sector should meet both electricity supply and demand. As a result, policymakers and economists show a great interest in predicting electricity demand among the different country sectors based on electricity consumption determinants [1].

Predicting electricity consumption through conventional techniques usually produces inaccurate results. This is due to the non-linearity between relevant determinants of electricity consumption, which limits the conventional techniques to achieve accurate prediction of electricity consumption results. Also, conventional techniques are hardly capable of handling big historical datasets on electricity consumption records and their determinants. Meanwhile, ML techniques have the ability to capture the non-linearity between electricity consumption determinants as well as processing big datasets representing time series patterns of electricity consumption. In this context, Machine Learning (ML) was commonly applied for simulating

DOI: 10.62184/mmc.jmmc110020251

the relationship between electricity consumption and its relevant determinants, and developing a persistent model for predicting electricity consumption on unseen data [3]. The applications of ML in predicting electricity consumption were recently promoted by the need for accurate prediction and availability of big data.

Varied ML models were developed for predicting electricity consumption taking into consideration community determinants. For example, random forest and support vector regression and gradient boosting algorithms were applied for daily/monthly electricity load forecasting based on gross domestic product, population growth rate, average temperature, relative humidity, and grid attributes [4]. Also, neural networks, fuzzy inductive reasoning, and random forest were applied to predict hourly electricity consumption in residential buildings based on weather conditions [5].

Similarly, a number of ML algorithms were applied for load forecasting at household level. For example, random forest, extreme gradient boosting, and Long Short-Term Memory (LSTM) algorithm were applied to forecast the load of power systems in residential buildings as a function of residential appliances [6]. A group of ML and deep learning techniques were employed to predict heating and cooling loads based on the building attributes and weather conditions [7]. Different algorithms of artificial neural networks modeled electricity consumption in residential buildings based on varied determinants such as air temperature, solar radiation, wall thickness, insulation type, the day-time, building physical design, and household size [8, 9].

It should be noted that ML has a contextual nature, where the performance of ML algorithms varies widely based on the type of the considered problem and the trained datasets. This means that there is no uniform technique that can be applied either for all problems or for specific issues [4]. Regarding data collection, big datasets are considered a prerequisite for applying ML algorithms, which are usually gathered either from database systems or through the Internet of Things technology [10]. Meanwhile, one of the challenges of applying ML in predicting electricity consumption is the data availability of electricity consumption patterns and their relevant determinants, especially at the household level [11].

Electricity consumption in residential sector contributes by about 27% of the world electricity consumption. Such a percentage has been increasing due to economic and population growth, which highlights the importance of predicting consumption pattern in the future to address increasing demand in the residential sector [12]. Predicting electricity consumption in the residential sector is dependent on varied determinants at both community and household levels. Electricity consumption is affected by a group of varied determinants at the community level, including weather conditions, the country's gross domestic products, and population size. Min, max, average temperature, and relative humidity are considered the main weather conditions parameters of the indoor cooling and heating appliances. Economic and population growth lead to increasing GDP per capita and urban population, which also maximizes the demand of electricity consumption. At household level, electricity consumption is generally determined by a number of factors. These factors can be classified into four categories: building physical characteristics, weather conditions, household appliances and their efficiency, and socio-economic factors.

This paper assesses the performance of tree-based ML algorithms to predict seasonal (summer, winter) electricity consumption patterns at household level. Such consumption patterns are evaluated in terms of three electricity consumption determinants categories which are building physical characteristics, household appliances and their efficiency, and socio-economic factors. The study intends to develop an ML persistent model that can represent and model the nonlinearity of electricity consumption determinants to predict electricity consumption at household level in Alexandria- Egypt under data scarcity. The applied tree-based ML algorithms are the Support vector classifier, K-nearest neighbor classifier, Decision tree classifier, Random Forest classifier, Bagging classifier, Multi-layer perceptron classifier, Radius neighbor classifier, and Gradient boosting classifier. Such classifiers are non-parametric algorithms that have no specific assumption of the training dataset, fit small dataset, obtain reasonable and appropriate results, and can train and model both continuous and categorical data types.

2. Case study

Egypt, as one of the developing countries that showed rapid population growth during the second half of last century, was selected to predict electricity consumption. Electricity consumption in Egypt has an increasing trend, residential sector consumes about 40% of electricity sold by electricity distribution companies, which reflects the largest share of that sector in

Multidisciplinary Materials Chronicles, 2025, Vol. X, Iss. X, 62-79
DOI: 10.62184/mmc.jmmc110020251

Research Article

electricity consumption. Electricity demand in the residential sector is expected to increase as the annual growth rate of the total number of population is expected to increase by about more than 2% to reach 116 million by 2040 [13]. The average annual growth rate of subscribers in the residential sector from 2014/2015 to 2018/2019 increased by 4% [14, 15]. Consequently, electricity demand beyond subscribers of households is expected to increase as well. Predicting electricity consumption in residential sector requires working on a case that can represent the problem and reflects its determinants.

This requires collecting data about electricity consumption at both macro and micro scales of the country. However, due to the scarcity of data availability of regions with different climatic and socio-economic conditions, Alexandria was selected for predicting electricity consumption at the household level, which specifies determinants that cannot be measured or has a marginal value at the community level. Moreover, predicting electricity consumption at the household scale is a significant case study to complement the results and conclusion to the community scale. In Alexandria, electricity consumption in residential sector increased annually from 2005 to 2018 by 3.9%. such a percentage is expected to increase due to the population growth rate **(Figure 1)**.

3. Data and methodology

To develop an ML-based prediction model for electricity consumption at household level, a methodology of three main steps is suggested **(Figure 2)**. The methodology includes determinants identification and data preparation, data preprocessing, applying supervised machine learning classification models.

3.1. Determinants identification and data preparation

Predicting electricity consumption in residential sector as a function of a number of determinants. Such determinants are related to building physical characteristics, household appliances and their efficiency, and socio-economic factors. This requires collecting historical data on these determinants. As a prerequisite for data collection, a preliminary list of determinants is developed based on the previous studies **(Table 1)**.

To predict electricity consumption at the household level, the aforementioned preliminary list of determinants **(Table 1)** was revised in accordance with the local conditions of the study area. As a result of such revision, a final list of determinants was identified, including:

- Housing conditions: (house area, story number, number of rooms, artificial lighting working hours, number of openings)
- Socio-economic conditions: (household size, householder educational level, income level, age-sex structure, social behavior)
- Household appliances: (number of appliances, appliances working hours, type of lamps).

The dependent variable is the electricity consumption patterns, which consumed monthly in Kwh. The electricity consumption is designed to represent the Egyptian segments which are seven patterns called segments, each of which represents a range of consumption.

To predict electricity consumption at the household level based on the prementioned determinants, data representing such determinants collected from primary source. household electricity determinants dataset was collected from a field survey. The following steps present a methodology applied for collecting household data through a field survey.

3.1.1. Designing questionnaire form

This involved designing a questionnaire form to collect data on the identified indicators. The questionnaire form was intended to acquire data on electricity consumption determinants in Alexandria. Therefore, the questionnaire involved four sections covering housing conditions, household behavior, socio-economic conditions of household, and appliance performance. “What is the number of openings?”, “What is the Family size?” and “what is the air conditioning working hours?” are examples of the questions.

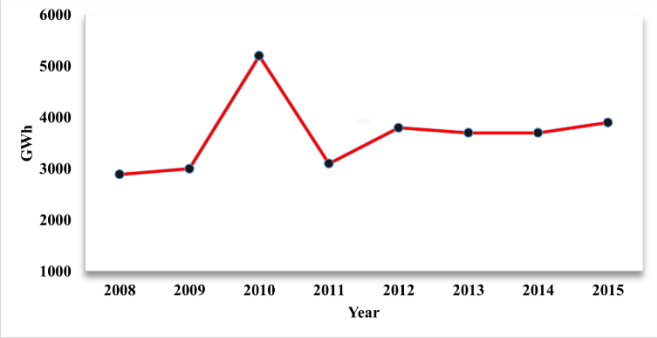


Figure 1. Annual residential electricity consumption in Alexandria [14].

Multidisciplinary Materials Chronicles, 2025, Vol. X, Iss. X, 62-79

DOI: 10.62184/mmc.jmmc110020251

Research Article

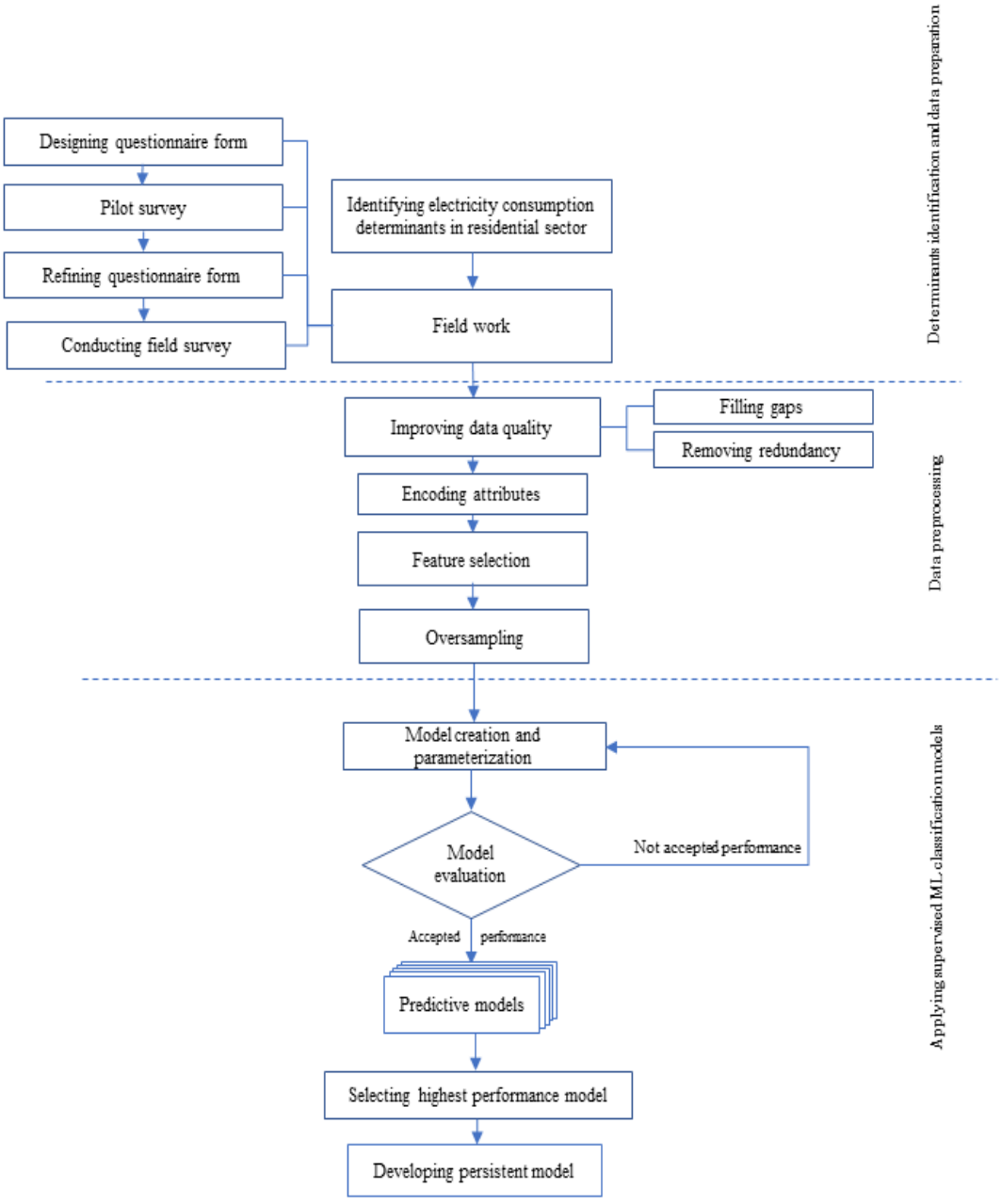


Figure 2. Suggested methodology for predicting electricity consumption.

DOI: 10.62184/mmc.jmmc110020251

Research Article

Table 1. Preliminary list of electricity consumption determinants at the household residential sector.

Category	Indicators	Expected relationship	Source
1. Housing conditions	- House area (m^2)	Positive	[16]
	- No. of windows per room on average	Negative	[16, 17]
	- Dependency on artificial lightening(hour/day)	Positive	[18, 19]
	- Insulation material	Negative	[16]
2. Socio-economic factors	- Household size	Positive	[20]
	- Age group • Householder age • Household age structure	• Has no significant association with annual summer or winter electricity consumption	[21]
		• Married households consume more energy than unmarried ones • Teenagers consume more electricity than younger one	[22]
		• Over 55 or between 19 and 35 age groups consume less energy	[16]
		Positive	[23]
	- Gender	No significant effect	[24]
		Positive	[23]
	- Income level • Householder • Household members on average	Curved line	[21, 25]
Socio-economic factors	- Educational level - Householder - Household members on average	Positive	[23, 24]
	- Number of employers - Householder - Household members on average	Positive	[23, 24]
	- Expenditure pattern	Positive	-
	- Energy conservation level	Negative	[23, 26]
	- Pet owner	Positive	[16]
3.Household appliances	- Number of appliances	Positive	[14, 16, 20, 21, 24]
	- Appliances lifetime & their working hours	Negative	[19, 26]
	- Lamps efficiency	Negative	[26]

3.1.2. Conducting a pilot survey

To assess the designed questionnaire form in terms of accessibility, clarity of the questions and quality of the answers, a pilot survey conducted online, involving twelve cases.

3.1.3. Refining questionnaire form

Based on the feedback received from the pilot survey, the questionnaire form was reviewed, and the wording of questions was edited to be clearer and focused. Therefore, about four questions were excluded as they required personal information, which prevented a number of participants from filling out the

DOI: 10.62184/mmc.jmmc110020251

survey. Two questions were excluded which are: "what are difficult questions found in the survey?", and "What are the expected questions not found in the survey?" as their answers have no significant meaning in the pilot survey. A number of questions were added including: 1. percentage of electricity bill to the householder income, 2. the electronic number of a householder electricity bill, 3. Age structure of a household, 4. Electric meter type, 5. Number of times for working appliances. Such added questions reflect the values of electricity consumption determinants.

3.1.4. Conducting a field survey

As the way of conducting a survey affects the frequency of distribution and accessibility to the target population [27], the survey was conducted through online surveys and interviews. An online survey, undertaken using google form, ensured high level of outreach to large number of people and targeted mainly people at high and intermediate educational levels. Meanwhile, interviews targeted mainly people with low education levels. The field survey was carried out over a seven-month period (from August to February 2019/2020). As a result of the field survey, 504 cases representing data on electricity consumption patterns and their relative determinants at the household level at Alexandria are collected. Such collected responses are filtered and preprocessed to be used for training and testing ML algorithms. All the collected data was anonymized to ensure data privacy and confidentiality.

3.2. Data preprocessing

Data preprocessing is an essential step as it makes data ready for training and learning. The process intended to improve data quality involves a number of tasks including removing redundancy and outliers, filling gaps and aggregating data, data transformation and feature selection [28, 29]. Such preprocessing tasks will be discussed by the following sub-sections.

3.2.1. Improving data quality

Improving data quality included data verification and enhancement to deal with missing and redundant data.

a. Removing redundancy and outliers

This step involved exploring data and removing repeated tuples to guarantee a balanced dataset and prevent overfitting. It was realized that 139 tuples were redundant, and their electricity consumption values were either missed or contained 0 KWh. So, such tuples were removed from the whole dataset and 365 representing household electricity consumption dataset were split to be 355 for training and validation and 10 tuples for testing the applied algorithms.

b. Filling gaps and aggregating data

A number of cases in the collected dataset at the household level were found to have missing values. Handling missing values can improve the information extracted by the dataset as well as the analysis and dataset learning [30].

The collected dataset contained null values concerning with the kWh consumed. So, about 97 records from household data set (365 tuples) containing null values were filled in by replacing it with the average kWh of each record [31].

It was realized that 365 tuples for the household are considered a limited dataset size for training machine learning algorithms. Limited dataset cannot achieve an acceptable accuracy for predicting electricity consumption through machine learning algorithms. Therefore, oversampling is applied for enhancing machine learning performance and attaining more precise prediction results. Applying oversampling on household dataset is presented in the oversampling section.

c. Electricity consumption determinants analysis

To ensure high performance of selected algorithms, the potential determinants are screened, revealing a subset of homogenous determinants. Accordingly, a number of these determinants are excluded from the analysis due to their limited variations, including house ownership, awareness level for energy conservation, household age-sex structure, mother educational level, income level, as well as type and number of lamps due to difficulty for evaluating their performance level. As a result, electricity consumption determinants incorporated in the analysis included eight variables reflecting housing conditions, socio-economic conditions, and household appliances **(Figure 3)**.

d. Data transformation

Data transformation is the third step to prepare dataset to be in a proper structure to be ready for training via a ML technique [32]. Data transformation includes a number of tasks, such as encoding categorical values and scaling or normalization. using proper transformation methods should be applied [33].

Regarding collected data, categorical values including storey number, number of rooms, openings number, artificial lighting working hours, householder educational level and appliances

DOI: 10.62184/mmc.jmmc110020251

working hours are converted into ordinal ones using label encoding. Also, L^2 normalization was applied to the collected data.

e. Feature selection

Feature selection, which involves the selection of the most effective determinants based on the collected dataset, can be considered as a metric function for optimizing the number of attributes to be sure that the selected attributes affect the problem with no intercorrelation between them [34, 35]. Applying feature selection to the training dataset as a preprocessing step before applying ML techniques may improve model performance and the training computational time [36].

There are almost three types of feature selection methods, which include filter method, wrapper method and embedded method that can be applied, each of which has more than one technique to be applied. Usually, filter methods, such as f-test and chi-square, use pairwise between each determinant and the target variable, which may imply disregarding interrelationships between dependent determinants. Wrapper method can take in consideration such interrelationships by making different combinations of the determinants, however, it takes more computational time based on the dimensionality of determinants. There are three techniques that can be applied for the wrapper method which are forward selection: iteratively one determinant is added and the evaluation of the predictive model evaluated to improve the model performance, backward elimination: it is the inverse step of forward selection which removes an individual determinant at a time to improve the model performance, and recursive feature elimination: applies both of the previous techniques for achieving the best subset of determinants [35, 37-39]. However, the wrapper method may lead to overfitting problem [34, 36]. The embedded method is considered an integration of the filter and the wrapper methods, it reduces overfitting that may occur by the wrapper method; however, it applies feature selection step at the training time. Thus, choosing a classifier algorithm depends on its applicability for applying the embedded method [40]. Consequently, the wrapper method was selected for applying feature selection for identifying the most electricity consumption determinants at household levels.

Wrapper method: Recursive Feature Elimination Cross validation (RFECV)

Recursive Feature Elimination Cross Validation (RFECV) is meanwhile a common wrapper method for feature selection, performs well with datasets that has nonlinear relationships as well as dataset with high dimensionality [41]. In addition, it considers the interrelationships between dependent variables through cross validation approach, which can improve the performance of machine learning algorithm and its accuracy [42]. According to RFECV, random forest classifier is considered a good base estimator for selecting the most relevant attributes to the target class especially in limited dataset size and nonlinear models [41]. Accordingly, RFECV-random forest was applied to the household electricity consumption determinants for feature selection. This was done to identify the relative importance of considered attributes to each other in terms of their contribution to the target class. RFECV was carried out using Python and scikit-learn library to identify the most relevant attributes to the target class which are assigned lower ranked, and the results are presented in RFECV results section.

f. Oversampling

Oversampling is a way of balancing instances in especially real-world problems that have shortage in data records which can represent training examples that reflect all classes in the problems such as medical and industrial applications. Such a lack in data sets is due to the nature of the problem and its complexity [43, 44].

ML algorithms performed poorly for imbalanced datasets, as such algorithms are based on balanced training datasets [45]. This can be explained by the fact that an imbalanced dataset does not represent each class, to enable the algorithm to train properly on such datasets. That is because the algorithm gives the majority class the higher attention during the classification step, leading to poor model performance [43, 44].

As a result, improving ML models for such data structure can be done by either increasing the number of instances of minority class or decreasing the number of instances of the majority one. Oversampling has been applied to the household dataset as the collected dataset did not sufficiently represent each class in the target variable. Such a solution can be achieved by creating more samples artificially and thus increasing the minority class training instances using Synthetic Minority Oversampling technique (SMOTE) [46].

Multidisciplinary Materials Chronicles, 2025, Vol. X, Iss. X, 62-79

DOI: 10.62184/mmc.jmmc110020251

Research Article

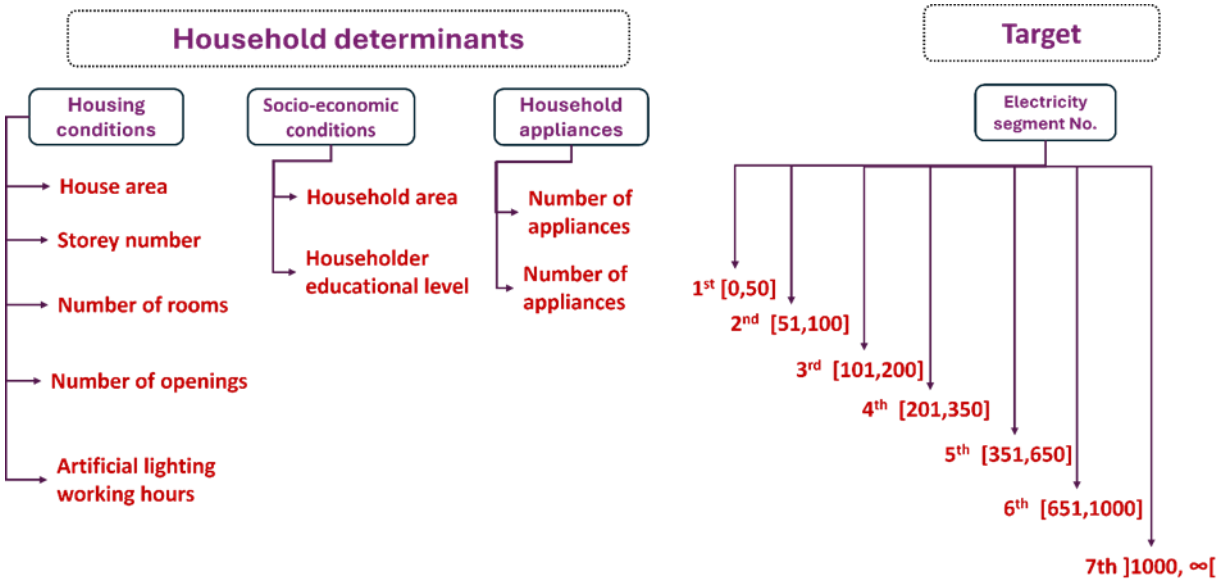


Figure 3. Household electricity consumption determinants.

According to the case study, dataset was split into two groups for winter and summer models. As consumption patterns in summer is always higher than that for summer and sub attributes related to winter model differs to that in summer model.

For example, fan working hours and Air conditioning working hours affect electricity consumption in summer but motor working hours significantly affects consumption during winter season.

The development research has been applied using Python, Jupyter Notebook lab, and Scikit learn. Python is a well-known programming language that has showed its efficiency and applicability for ML-based applications. Owing to being an open source, it has robust libraries that can be applied for data science applications [47]. Jupyter Notebook is an open-source web application which supports implementing of data, data preprocessing, data visualization, and ML techniques [48]. Regarding Scikit-learn, it is an open-source library which provides different tools to easily apply feature selection, data preprocessing, and ML as well. Moreover, it can easily be used for data analysis for different research fields. Data preprocessing is implemented using Scikit-learn including normalization, and feature selection [42].

3.3. Applying supervised machine learning classification models

To predict electricity consumption as a function of the shortlisted determinants at the household level, the following classification algorithms are applied: Support vector classifier, K-nearest neighbor classifier, Decision tree classifier as well as Bagging classifier, Multi-layer perceptron classifier, Radius neighbor classifier, Random forest classifier, and Gradient boosting classifier. Thereafter, the results of each algorithm are validated through accuracy, precision, and recall metrics. Finally, a persistent model is developed and employed to predict electricity consumption. Consequently, the optimal values of the hyperparameters were set based on optimal performance during cross-validation and trial and error as follows.

- Support vector classifier model is created by identifying kernel function to be polynomial and regularization parameter (c) = 200 for both summer and winter models.
- K-nearest neighbor algorithm for summer and winter models are created by k=10 for summer model and k = 15 for winter model and distance as a weight function for both models.
- Decision tree algorithm is applied for summer model using gini criterion, but winter model tree is applied based on entropy.
- Random forest classifier is applied for summer and winter models such that summer model is best fitted using gini criterion while winter model is fitted by entropy criterion.
- Bagging classifier model is applied by adjusting the number of 200 base estimators. Gini criterion is applied for summer model however, entropy criterion gives better performance for winter model.

Multidisciplinary Materials Chronicles, 2025, Vol. X, Iss. X, 62-79
DOI: 10.62184/mmc.jmmc110020251

Research Article

- Multi-layer perceptron algorithm is applied to predict electricity consumption segment for winter and summer season. Logistic function best fitted household-winter model, while tanh function is appropriate for household-summer model and weight optimization function is 'lbfgs'.
- Radius neighbor classifier with default radius = 1 is applied for summer and winter models.
- Gradient boosting classifier is applied with 200 base estimators for winter model but with 500 estimators for summer model. Also, deviance loss function is applied for both models.

Such classification algorithms are applied, and their accuracy results is presented in Results and Discussion section. Consequently, based on the performance of each classification algorithm, the one with highest accuracy is selected for deploying a persistent predictive model which is discussed by the following section.

4. Results and discussion

Grouping datasets according to the related segments ensures that the collected samples reflect the whole population. As, the distribution of the collected samples almost reflect the distribution of the number of subscribers to different household usage according to the electricity segments. As it is clear, most of the sampling records collected lay in the third, and fourth segments during the summer and winter seasons **(Figure 4, Figure 5)**. The sample covers most of Alexandria's localities. It is noted that most of the records are in the third and fourth segments, during the winter season, as it is difficult for household consumption to lie in the seventh segment. This is because there are no cooling purposes during winter.

Application of feature selection methods revealed that story number, number of rooms, ventilation (number of openings), floor area, artificial lighting working hours, household size, head of household educational level, number of appliances, fan working hours, air conditioning working hours, and washing machine working hours are the most effective determinants [23, 24] according to the case study during summer. However, kettle working hours, water motor working hours, microwave working hours and heater working hours during summer season are less effective to electricity consumption. Generally, no significant seasonal difference is noted in terms of identified determinants. Similarly, story number, number of rooms, ventilation, floor area, artificial lighting working hours, household size, householder educational level, number of appliances and washing machine working hours are found to be significant determinants of electricity consumption in winter in addition to motor working hours, kettle working hours, heater working hours. It is worth mentioning that the seasonal variations in identified determinants reflect varied electricity consumption for cooling and heating purposes in the two seasons **(Figure 6)**. Generally, at household level housing conditions [17], household appliances and households' socio-economic determinants significantly affect electricity consumption, which may reflect that income level as well as the behavior of households are considered the main preliminary determinants for households' electricity demand.

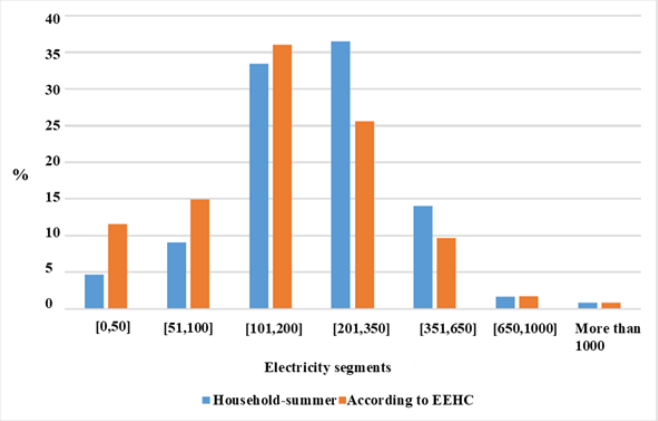


Figure 4. Household grouping during summer.

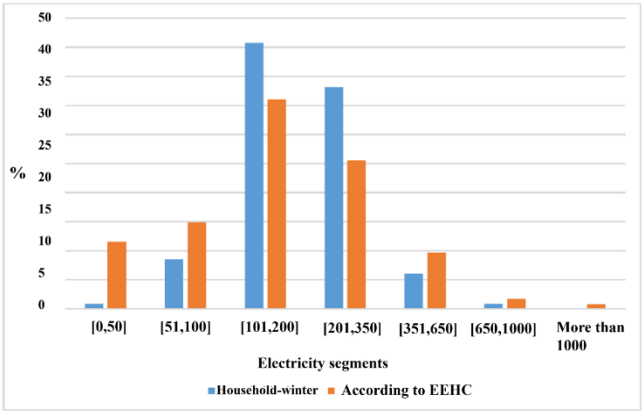


Figure 5. Household grouping during winter.

Performance of applied algorithms are evaluated based on accuracy **(Equation 1)**, precision **(Equation 2)** and recall evaluation metrics **(Equation 3)**.

Multidisciplinary Materials Chronicles, 2025, Vol. X, Iss. X, 62-79

DOI: 10.62184/mmc.jmmc110020251

Research Article

$$Accuracy(y, \hat{y}) = \frac{1}{number\ of\ samples} \sum_{i=0}^{i=number\ of\ samples-1} (\hat{y_i} - y_i) \quad (1)$$

$$Precision = \frac{true\ positive}{true\ positive + false\ positive} \quad (2)$$

$$Recall = \frac{true\ positive}{true\ positive + false\ negative} \quad (3)$$

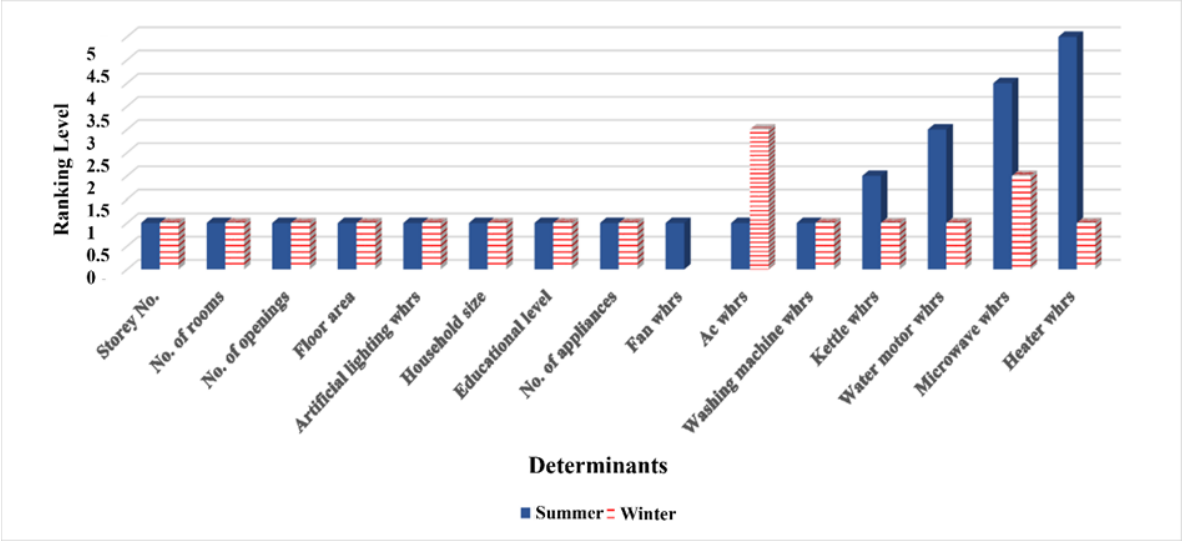


Figure 6. RFECV results for household electricity consumption determinants.

Application of the predefined classification ML algorithms on original dataset revealed low accuracy at the two seasonal models. Such low accuracy can be attributed to limited dataset size and small numbers of representative members for each class. In such a case, oversampling is usually applied in order to achieve better accuracies for the applied algorithms. Applying ML algorithms to the oversampled datasets showed a relatively high level of accuracy compared to the original datasets (**Table 2**).

According to their average accuracy, the applied algorithms can be classified into three main categories (**Figure 7**):

- Highly performance algorithms: this category involves those algorithms that have high accuracy exceeding 80% including: bagging classifier (base estimator extra trees classifier), decision tree classifier, random forest classifier and gradient boosting classifier.
- Intermediate performance algorithms: the accuracy of algorithms of this category ranged between 40-80%. These algorithms include Support vector classifier, k-nearest neighbor classifier, Multi-layer perceptron.
- Low performance algorithms: this category is represented in Radius neighbor classifier, which has low accuracy of less than 40%.

The highest accuracy algorithms are Bagging classifier, Random Forest, and Gradient boosting. Radius neighbor classifier revealed the lowest accuracy, which may be attributed to improper radius of the selected area that is irrelevant to data distribution [49]. It is noted that the accuracy of Radius neighbor classifier decreased by oversampling. That is why this algorithm is from the k-nearest neighbor algorithms family, which revealed low performance with large datasets. This is due to the cost complexity for calculating the radius distances between the existing points and the new predicted ones. Also, k-nearest neighbors classifier has low performance level with high dimension datasets. This is due to the difficulty for calculating the radius distance in multi-dimensional space [50, 51].

Multidisciplinary Materials Chronicles, 2025, Vol. X, Iss. X, 62-79

DOI: 10.62184/mmc.jmmc110020251

Research Article

Table 2. Accuracy of applied classification machine learning algorithm.

Classification algorithm	Accuracy of summer models			Accuracy of winter models		
	Original data	Oversampled data	Accuracy change	Original data	Oversampled data	Accuracy change
Bagging classifier Base classifier: extra trees classifier	0.40	0.84	+0.44	0.46	0.88	+0.42
Random forest classifier	0.42	0.84	+0.42	0.48	0.87	+0.39
Gradient boosting classifier	0.36	0.84	+0.48	0.44	0.87	+0.43
Decision tree classifier	0.32	0.81	+0.49	0.40	0.85	+0.45
k-nearest neighbor	0.37	0.78	+0.41	0.44	0.78	+0.34
Support vector classifier	0.41	0.56	+0.15	0.47	0.63	+0.16
Multi-layer perceptron	0.36	0.49	+0.13	0.44	0.58	+0.14
Radius neighbor classifier	0.37	0.14	-0.23	0.46	0.16	-0.3

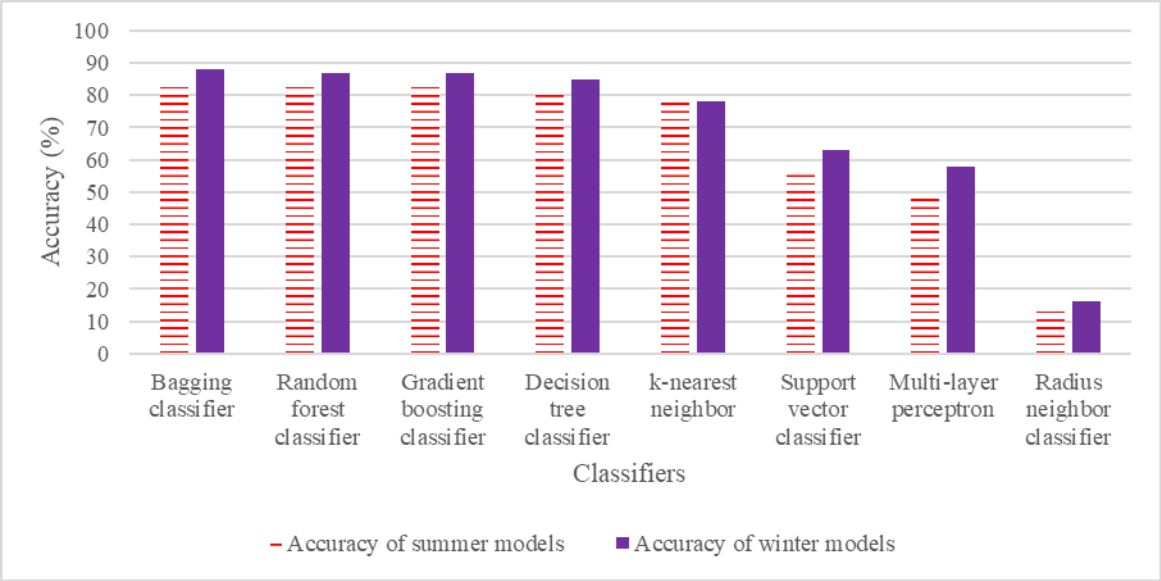


Figure 7. Performance of applied algorithms at household level.

Support vector classifier showed relatively higher performance compared to Multi-layer perceptron classifier in predicting hourly electricity consumption in residential sector, [52]. As Feedforward Neural Networks showed lower performance with categorical inputs.

K-nearest neighbor and Decision tree, is preferred for categorical and ordinal target variables [53], and are good classification algorithms for non-parametric dataset as they do not assume any functional form to the trained dataset [54]. Therefore, they showed higher performance than other applied classifiers for predicting electricity segments. Such high-performance

Multidisciplinary Materials Chronicles, 2025, Vol. X, Iss. X, 62-79

DOI: 10.62184/mmc.jmmc110020251

Research Article

algorithms, particularly decision tree classifier, are preferred for categorical and ordinal target variables [53].

Based on precision metrics results, mean precision is calculated for the 5-folds, then the average of the 5-fold of each class of a specific algorithm is compared to the overall mean precision of each class. It is found that Bagging classifier, Gradient boosting classifier, Random Forest classifier and Decision tree classifier showed higher precision performance (**Table 3**). The precision of the Support vector classifier, Multi-layer perceptron, and Radius neighbor classifier is less than mean precision value for summer and winter models (**Figure 8, Figure 9)**. Therefore, such classifiers are excluded to be a persistent model, and only Bagging classifier, Gradient boosting classifier, Random forest decision tree, and k-nearest neighbor classifiers revealed on average higher performance for predicting electricity segments at the household level.

Similarly, the dataset is evaluated based on recall metrics (**Table 4),** and it is realized that only Bagging classifier, Gradient boosting classifier, Random Forest, and Decision tree classifier showed better performance. Recall results of k-nearest neighbor classifier, Multi-layer perceptron classifier, Support vector classifier and Radius neighbor classifier are not appropriate for predicting electricity consumption segments. As, their recall average percentage are less than the average value of all segments (**Figure 10, Figure 11**).

Table 3. Precision performance of applied algorithms in summer and winter models.

Algorithm	Mean Precision													
	First segment		Second segment		Third segment		Fourth segment		Fifth segment		Sixth segment		Seventh segment	
	Summer model	Winter model	Summer model	Winter model	Summer model	Winter model	Summer model	Winter model	Summer model	Winter model	Summer model	Winter model	Summer model	Winter model
Bagging classifier	0.96	1	0.90	0.90	0.58	0.70	0.56	0.64	0.99	0.96	0.98	0.99	1	-
Gradient boosting Classifier	0.93	1	0.90	0.90	0.61	0.69	0.52	0.64	0.96	0.96	0.96	1	0.99	-
Random forest Classifier	0.91	1	0.89	0.92	0.63	0.67	0.52	0.63	0.97	0.97	0.97	0.99	1	-
Decision tree classifier	0.88	0.99	0.83	0.85	0.57	0.66	0.52	0.63	0.95	0.92	0.95	0.98	0.95	-
k-nearest neighbor	0.71	0.93	0.74	0.73	0.66	0.63	0.65	0.71	0.92	0.71	0.92	0.93	0.91	-
Multi-layer Perceptron	0.44	0.87	0.44	0.43	0.22	0.34	0.27	0.35	0.85	0.46	0.85	0.846	0.59	-
Support vector classifier	0.44	0.91	0.44	0.50	0.24	0.41	0.25	0.34	0.85	0.59	0.85	0.88	0.78	-
Radius neighbor classifier	0.14	0.03	0	0.03	0	0.03	0	0.03	0	0.03	0	0	0	-
Mean precision	0.68	0.84	0.64	0.67	0.44	0.52	0.41	0.50	0.81	0.70	0.81	0.83	0.78	-

Multidisciplinary Materials Chronicles, 2025, Vol. X, Iss. X, 62-79
DOI: 10.62184/mmc.jmmc110020251

Research Article

Bagging classifier with decision tree as a base estimator, Random Forest, and Gradient boosting classifiers almost showed higher and similar accuracies. Random forest is similar to bagging classifier; the only difference between the two classifiers is that the former considers, at each decision tree, a random subset of the dataset features [55, 56]. However, Gradient boosting classifier minimizes the loss function value consecutively based on the previous results of the created tree, and this may lead to overfitting in case of noisy dataset [55, 57, 58].

The highest performance algorithms were evaluated to select the most appropriate classifier for predicting electricity consumption at the household level. Firstly, to avoid potential overfitting in case of noise datasets Gradient boosting classifier was excluded. Moreover, despite their similarity, Random Forest classifier gives usually better performance compared to Bagging classifier as at each created decision tree, randomly select a subset of the dataset features. Accordingly, it was decided to employ Random Forest classifier as a persistent model for predicting electricity consumption.

Testing model performance was carried out by using ten records of the collected dataset and it is found that the model showed a noticeable high accuracy in winter (70%) compared to summer model (10%). This is due to varied residents' consumption behaviors in summer.

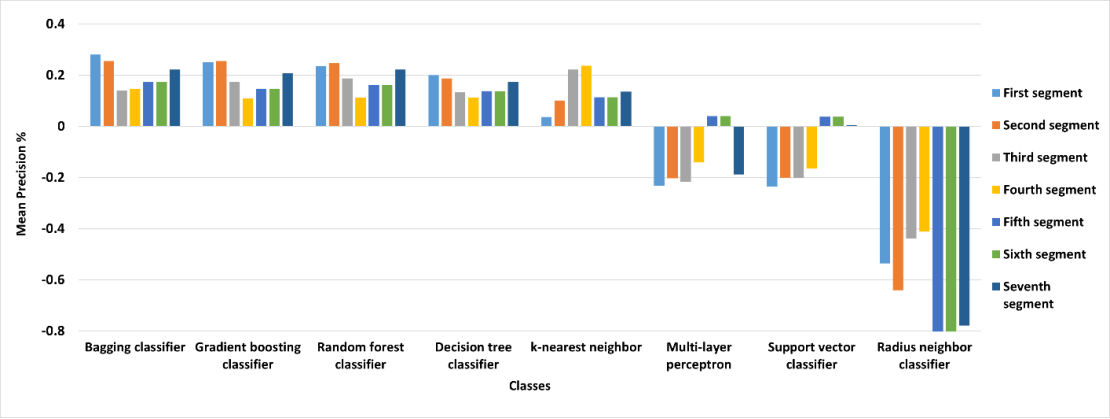


Figure 8. Difference between precision of each class and the average deviation class precision in summer model.

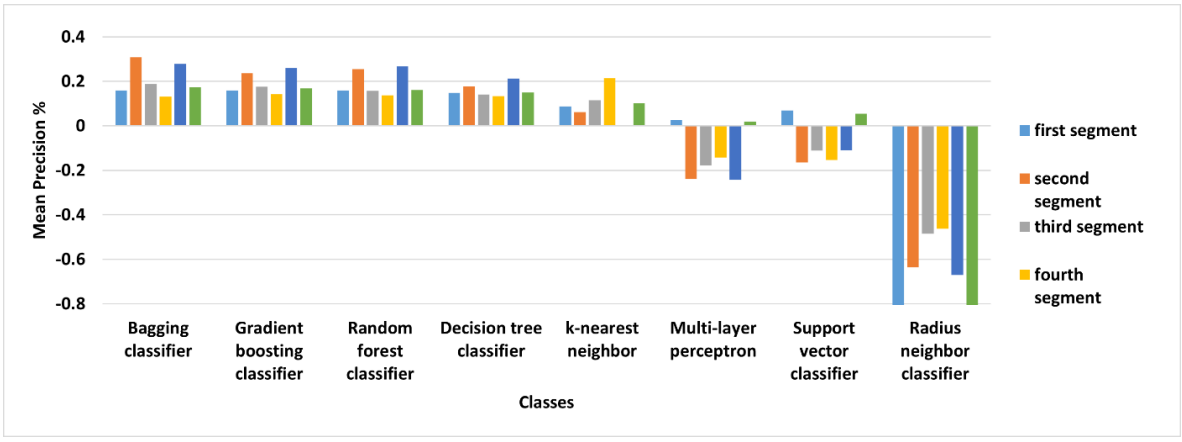


Figure 9. Difference between precision of each class and the average deviation class precision in winter model.

Multidisciplinary Materials Chronicles, 2025, Vol. X, Iss. X, 62-79

DOI: 10.62184/mmc.jmmc110020251

Research Article

Table 4. Recall performance of applied algorithms in summer and winter models.

Algorithm	Mean Recall													
	First segment		Second segment		Third segment		Fourth segment		Fifth segment		Sixth segment		Seventh segment	
	Summer model	Winter model	Summer model	Winter model	Summer model	Winter model	Summer model	Winter model	Summer model	Winter model	Summer model	Winter model	Summer model	Winter model
Bagging classifier	1	1	1	1	0.59	0.57	0.31	0.68	0.99	1	1	1	1	-
Gradient boosting Classifier	1	1	1	1	0.56	0.51	0.33	0.69	0.98	1	1	1	1	-
Random forest Classifier	1	1	1	1	0.57	0.52	0.33	0.67	0.99	1	1	1	1	-
Decision tree classifier	1	1	1	1	0.43	0.41	0.27	0.67	0.98	1	1	1	1	-
k-nearest neighbor	1	1	1	1	0.37	0.16	0.12	0.53	0.99	1	1	1	1	-
Multi-layer Perceptron	0.47	1	0.51	0.52	0.18	0.23	0.17	0.21	0.41	0.5	1	1	0.70	-
Support vector classifier	0.60	1	0.53	0.58	0.15	0.28	0.11	0.19	0.52	0.73	1	1	1	-
Radius neighbor classifier	1	0.2	0	0.2	0	0.2	0	0.2	0	0.2	0	0	0	-
Mean precision	0.89	0.9	0.75	0.79	0.36	0.36	0.20	0.48	0.73	0.80	0.88	0.88	0.84	-

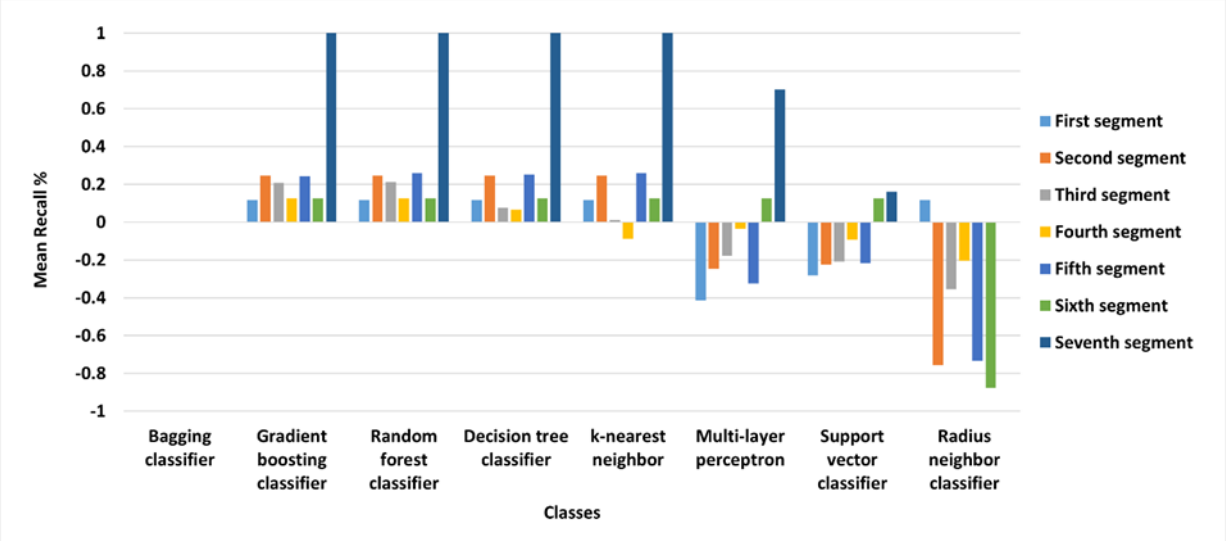


Figure 10. Difference between recall of each class and the average deviation class recall in summer model.

Multidisciplinary Materials Chronicles, 2025, Vol. X, Iss. X, 62-79

DOI: 10.62184/mmc.jmmc110020251

Research Article

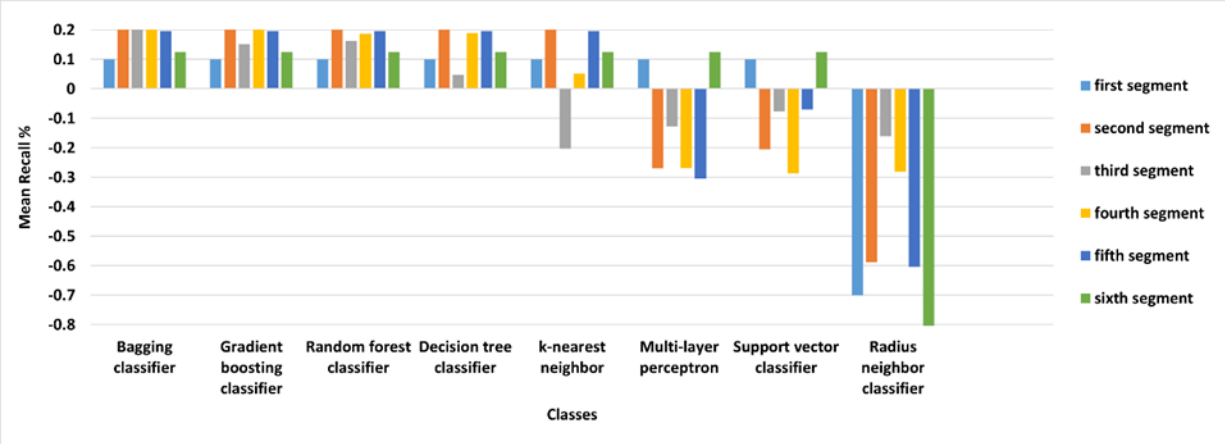


Figure 11. Difference between recall of each class and the average deviation class recall in winter model.

5. Conclusion

Developing an ML model for predicting electricity consumption under scarcity of data is considered a challenging goal. Accordingly, oversampling is applied to the training data using SMOTE to get a balanced training dataset, eight classification ML techniques are applied including: Support vector classifier, k-nearest neighbor classifier, Decision tree classifier as well as Bagging classifier, Multi-layer perceptron classifier, Radius neighbor classifier, Random forest classifier, and Gradient boosting classifier. The performance of the applied algorithms is evaluated according to accuracy, precision, and recall. Algorithms that are tree-based structures showed higher performance in terms of the evaluation metrics chosen. Random Forest is selected as a persistent model for predicting electricity consumption at household level. This is because, it considers random subset of the dataset features for each created base estimator. It should be noted that applying ML for predicting electricity consumption patterns under scarcity of data requires applying the appropriate oversampling method to ensure training of a balanced dataset.

To accurately predict electricity consumption in the residential sector in Egypt using ML models, there is a need for availability of a detailed, accurate and up to date database, which is considered challenging in developing countries, concerning for example buildings' physical characteristics, demographic, and socio-economic conditions. Such database may help in developing monthly electricity consumption profile for households, which may be taken into consideration to conserve electricity consumption. Regarding energy policy, short-term as well as long term load forecasting models are highly recommended to be developed at both community and household levels. This may positively reflect on the Integrated Sustainable Energy Strategy (ISES) to 2035 developed by the Egyptian government.

Declaration of interests

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Author contribution

Nourhan H. Farag: Data collection and analysis, Model implementation, Writing original draft.

Mohamed A. Abdrabo: Conceptualization, Methodology, reviewing and editing of the manuscript.

Mohamed A. El-Iskandarani: Supervision – Methodology, reviewing and editing the manuscript.

Author information

Corresponding author: Nourhan H. Farag*

E-mail: nourhan.hamdy@alexu.edu.eg

ORCID iD: 0000-0003-3491-4524

References

[1] T. Lorde, K. Waithe, B. Francis, The importance of electrical energy for economic growth in Barbados, Energy Econ. 32(6) (2010) 1411-1420. https://doi.org/10.1016/j.eneco.2010.05.011.

[2] I. Dokas, M. Panagiotidis, S. Papadamou, E. Spyromitros, The determinants of energy and electricity consumption in developed and developing countries: international evidence,

DOI: 10.62184/mmc.jmmc110020251

Research Article

Energies 15(7) (2022) 2558. https://doi.org/10.3390/en15072558.

[3] M.M. Forootan , I. Larki, R. Zahedi, A. Ahmadi, Machine learning and deep learning in energy systems: A Review, Sustainability 14(8) (2022) 4832. https://doi.org/10.3390/su14084832.

[4] J.V. Leme, W. Casaca, M. Colnago, M.A. Dias, Towards assessing the electricity demand in Brazil: Data-driven analysis and ensemble learning models, Energies 13(6) (2020) 1407. https://doi.org/10.3390/en13061407.

[5] S. Jurado, À. Nebot, F. Mugica, N. Avellana, Hybrid methodologies for electricity load forecasting: Entropy-based feature selection with machine learning and soft computing techniques, Energy 86 (2015) 276-291. https://doi.org/10.1016/j.energy.2015.04.039.

[6] S. Hadri, Y. Naitmalek, M. Najib, M. Bakhouya, Y. Fakhri, M. Elaroussi, A comparative study of predictive approaches for load forecasting in smart buildings. Procedia Comput Sci. 160 (2019) 173-180. https://doi.org/10.1016/j.procs.2019.09.458.

[7] A. Moradzadeh, B. Mohammadi-Ivatloo, M. Abapour, A. Anvari-Moghaddam, S.S. Roy, Heating and cooling loads forecasting for residential buildings based on hybrid machine learning applications: A Comprehensive review and comparative analysis, IEEE Access 10 (2022) 2196-2215. https://doi.org/10.1109/ACCESS.2021.3136091.

[8] S. Seyedzadeh, F.P. Rahimian, I. Glesk, M. Roper, Machine learning for estimation of building energy consumption and performance: A review, Vis.Eng. 6(1) (2018) 1-20. https://doi.org/10.1186/s40327-018-0064-7.

[9] H. Son, C. Kim, A deep learning approach to forecasting monthly demand for residential–sector electricity, Sustainability 12(8) (2020) 3103. https://doi.org/10.3390/su12083103.

[10] M. Talebkhah, A. Sali, M. Marjani, M. Gordan, S.J. Hashim, F.Z. Rokhani, IoT and big data applications in smart cities: Recent advances, challenges, and critical issues, IEEE Access 9 (2021) 55465-55484. https://doi.org/10.1109/ACCESS.2021.3070905.

[11] O. Adeoye, C. Spataru, Modelling and forecasting hourly electricity demand in West African countries, Applied Energy 242 (2019) 311-333. https://doi.org/10.1016/j.apenergy.2019.03.057.

[12] International Energy Agency (IEA), Global Energy Review 2020, Electricity, 2020. https://www.iea.org/reports/global-energy-review-2020/electricity.

[13] S.K. Ghanem, The relationship between population and the environment and its impact on sustainable development in Egypt using a multi-equation model. Environ. Dev. Sustain. 20(1) (2018) 305-342.https://doi.org/10.1007/s10668-016-9882-8.

[14] M.A.K.A. Abdrabo, M.A. Hassaan, H. Abdelraouf, Impacts of climate change on seasonal residential electricity consumption by 2050 and potential adaptation options in Alexandria Egypt, Am. J. Clim. Change 7(4) (2018) 575-585. https://doi.org/10.4236/ajcc.2018.74035.

[15] Egyptian Electricity Holding Company, Annual Report, 2019. http://www.moee.gov.eg/english_new/EEHC_Rep/2018-2019en.pdf.

[16] A. Kavousian, R. Rajagopal, M. Fischer, Determinants of residential electricity consumption: Using smart meter data to examine the effect of climate, building characteristics, appliance stock, and occupants' behavior, Energy 55 (2013) 184-194. https://doi.org/10.1016/j.energy.2013.03.086.

[17] A.S. Algendy, M.F. Anbar, Energy Efficiency in residential buildings in Egypt with special reference to windows, IJCET 7 (2017) 126-134.

[18] A.N. Ahmed, M.M. Samaan, O.M.A. Farag, A.S. El Aishy, Using simulation tools for enhancing residential buildings energy code in Egypt, Proceedings of Building Simulation 2011: 12th Conference of IBPSA, Sydney, Australia, November 2011, pp. 1141-1148. https://doi.org/10.26868/25222708.2011.1412.

[19] S. Attia, E. Gratia, A. De Herde, J.L. Hensen, Simulation-based decision support tool for early stages of zero-energy building design, Energy Build. 49 (2012) 2-15. https://doi.org/10.1016/j.enbuild.2012.01.028.

[20] O. Akihiro, Determinants of efficiency in residential electricity demand: stochastic frontier analysis on Japan, Energy Sustain. Soc. 7(1) (2017) 31. https://doi.org/10.1186/s13705-017-0135-y.

DOI: 10.62184/mmc.jmmc110020251

Research Article

[21] C. Li, Y. Song, N. Kaza, Urban form and household electricity consumption: A multilevel study. Energy Build. 158 (2018) 181-193. https://doi.org/10.1016/j.enbuild.2017.10.007.

[22] D. Brounen, N. Kok, J.M. Quigley, Residential energy use and conservation: Economics and demographics. Eur. Econ. Rev. 56(5) (2012) 931-945. https://doi.org/10.1016/j.euroecorev.2012.02.007.

[23] P. Esmaeilimoakher, T. Urmee, T. Pryor, G. Baverstock, Identifying the determinants of residential electricity consumption for social housing in Perth, Western Australia, Energy Build. 133 (2016) 403-413. https://doi.org/10.1016/j.enbuild.2016.09.063.

[24] F. Taale, C. Kyeremeh, Drivers of households' electricity expenditure in Ghana, Energy Build. 205 (2019) 109546. https://doi.org/10.1016/j.enbuild.2019.109546.

[25] Z.Chun-sheng, N. Shu-Wen, Z. Xin, Effects of household energy consumption on environment and its influence factors in rural and urban areas. Energy Procedia, 14(2012)805-811. https://doi.org/10.1016/j.egypro.2011.12.1015

[26] M.Sakah, S.d.l.R. du Can, F.A. Diawuo, M.D. Sedzro, C. Kuhn, A study of appliance ownership and electricity consumption determinants in urban Ghanaian households. Sustain. Cities Soc. 44 (2019) 559-581. https://doi.org/10.1016/j.scs.2018.10.019.

[27] S. Rice, S.R. Winter, S. Doherty, M. Milner, Advantages and disadvantages of using internet-based survey methods in aviation-related research, JATE 7(1) (2017) 5. https://doi.org/10.7771/2159-6670.1160.

[28] S. Ramírez-Gallego, B. Krawczyk, S. García, M. Woźniak, F. Herrera, A survey on data preprocessing for data stream mining: Current status and future directions, Neurocomputing 239 (2017) 39-57. https://doi.org/10.1016/j.neucom.2017.01.078.

[29] S. Ramírez-Gallego, J. Luengo, J.M. Benítez, F. Herrera, Big data preprocessing: methods and prospects, Big Data Anal. 1(1) (2016) 9. https://doi.org/10.1186/s41044-016-0014-0.

[30] J. Josse, F. Husson, missMDA: a package for handling missing values in multivariate data analysis. J. Stat. Softw. 70(1) (2016) 1-31. https://doi.org/10.18637/jss.v070.i01.

[31] M.M. Islam, H. Iqbal, M.R. Haque, M.K. Hasan. Prediction of breast cancer using support vector machine and K-Nearest neighbors. 2017 IEEE Region 10 Humanitarian Technology Conference (R10-HTC), 2017, pp. 226-229. https://doi.org/10.1109/R10-HTC.2017.8288944.

[32] A. Kumar, A.A. Deshmukh, U. Dogan, D. Charles, E. Manavoglu, Data transformation insights in self-supervision with clustering tasks, Proceedings of the 37th International Conference on Machine Learning, Vienna, Austria, PMLR 108, 2020. https://doi.org/10.48550/arXiv.2002.07384.

[33] L. Saari, Detecting performance anomalies in a mobile application with unsupervised machine learning, 2019. https://kth.divaportal.org/smash/record.jsf?pid=diva2%3A1337068&dswid=253

[34] N. Deepa, B. Prabadevi, P.K. Maddikunta, T.R. Gadekallu, T. Baker, M.A. Khan, U. Tariq, An aIbased intelligent system for healthcare analysis using RidgeAdaline Stochastic Gradient Descent Classifier, J. Supercomput, 77(2) (2020) 1998-2017. https://doi.org/10.1007/s11227-020-03347-2.

[35] A.C. Pandey, D.S. Rajpoot, M. Saraswat, Feature selection method based on hybrid data transformation and binary binomial cuckoo search, J Ambient Intell. Humaniz. Comput., 11(2) (2020) 719-738. https://doi.org/10.1007/s12652-019-01330-1.

[36] C. Jie, L. Jiawei, W. Shulin, Y. Sheng, Feature selection in machine learning: A new perspective, Neurocomputing, 300 (2018) 70-79. https://doi.org/10.1016/j.neucom.2017.11.077.

[37] V. luhaniwal, Feature selection using Wrapper methods in Python, Towards Data Science 4 (2019). https://towardsdatascience.com/feature-selection-using-wrapper-methods-in-python-f0d352b346f.

[38] S. Sadeghyan, A new robust feature selection method using variance-based sensitivity analysis, arXiv preprint arXiv:1804.05092, (2018) 1-9. https://doi.org/10.48550/arXiv.1804.05092.

[39] R. Farmar, N. Han, M. McCombe, Intro to feature selection methods for data science, 2019. https://medium.com/towards-data-science/intro-to-feature-selection-methods-for-data-science-4cae2178a00a.

DOI: 10.62184/mmc.jmmc110020251

Research Article

[40] Z.M. Hira, D.F. Gillies, A review of feature selection and feature extraction methods applied on microarray data, Advance bioinform. 2015 (2015) 198363. https://doi.org/10.1155/2015/198363.

[41] B.F Darst, K.C. Malecki, C.D. Engelman, Using recursive feature elimination in random forest to account for correlated variables in high dimensional data, BMC genetics 19(1) (2018) 65. https://doi.org/10.1186/s12863-018-0633-8.

[42] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, E. Duchesnay, Scikit-learn: Machine Learning in Python, JMLR 12 (2011) 2825-2830.

[43] C. Drummond, N. Japkowicz, Manifold-based synthetic oversampling with manifold conformance estimation, Mach. Learn. 107(3) (2018) 605-637. https://doi.org/10.1007/s10994-017-5670-4.

[44] S. Delen, T. Liu, A synthetic informative minority over-sampling (SIMO) algorithm leveraging support vector machine to enhance learning from imbalanced datasets, Decis. Support Syst. 106 (2018) 15-29. https://doi.org/10.1016/j.dss.2017.11.006.

[45] M. Mayo, L. Chepulis, R.G. Paul, Glycemic-aware metrics and oversampling techniques for predicting blood glucose levels using machine learning, PLoS One 14(12) (2019) e0225613. https://doi.org/10.1371/journal.pone.0225613.

[46] S. Shahinfar, H.A. Al-Mamun, B. Park, S. Kim, C. Gondro, Prediction of marbling score and carcass traits in Korean Hanwoo beef cattle using machine learning methods and synthetic minority oversampling technique, Meat Sci. 161(2020) 107997. https://doi.org/10.1016/j.meatsci.2019.107997.

[47] S. Raschka, V. Mirjalili, Python machine learning: Machine learning and deep learning with python, scikit-Learn, and TensorFlow, Third ed., Packt publishing ltd, Birmingham, 2019. ISBN 978-1-78995-575-0.

[48] Jupyter, P. Jupyter, 2020. https://jupyter.org/.

[49] S. Pourbahrami, L.M. Khanli, A Survey of neighbourhood construction models for categorizing data points. arXiv preprint arXiv:1810.03083 (2018). https://doi.org/10.48550/arXiv.1810.03083.

[50] F. Shen, O. Hasegawa, A fast nearest neighbor classifier based on self-organizing incremental neural network, Neural Netw. 21(10) (2008) 1537-1547. https://doi.org/10.1016/j.neunet.2008.07.001.

[51] N. Kumar, Advantages and disadvantages of KNN algorithm in machine learning, The Professionals Point (2019). http://theprofessionalspoint.blogspot.com/2019/02/advantages-and-disadvantages-of-knn.html.

[52] R.E. Edwards, J. New, L.E. Parker, Predicting future hourly residential electrical consumption: A machine learning case study, Energy Build. 49 (2012) 591-603. https://doi.org/10.1016/j.enbuild.2012.03.010.

[53] P. Tsangaratos, I. Ilia, Landslide susceptibility mapping using a modified decision tree classifier in the Xanthi Perfection, Greece, Landslides 13(2016) 305-320. https://doi.org/10.1007/s10346-015-0565-6.

[54] L. Dey, S. Chakraborty, A. Biswas, B. Bose, S. Tiwari, Sentiment analysis of review datasets using naive bayes and k-nn classifier, arXiv:1610.09982 8(4) (2016) 54-62. https://doi.org/10.48550/arXiv.1610.09982.

[55] B. Sujathakumari, M. Abhishek, D. Singh, K. Aneesh, D. Rakesh, B. Mahanand. Detection of MCI from MRI using gradient boosting classifier, 1st International Conference on Advances in Information Technology (ICAIT), Chikmagalur, India, July 2019, pp. 70-75. https://doi.org/10.1109/ICAIT47043.2019.8987413.

[56] S. Glen, Decision tree vs random forest vs gradient boosting machines: Explained simply, 2019. https://www.datasciencecentral.com/decision-tree-vs-random-forest-vs-boosted-trees-explained/.

[57] M. Walker, Boosting algorithms for better predictions. 2014. https://www.datasciencecentral.com/profiles/blogs/boosting-algorithms-for-better-predictions.

[58] A. Natekin, A. Knoll, Gradient boosting machines, a tutorial. Front. neurorobot. 7 (2013) 21. https://doi.org/10.3389/fnbot.2013.00021