Journal of Engineering and Technology for Industrial Applications

ITEGAM-JETIA

Manaus, v10. n.50, p. 184-190. November/ December., 2024.
DOI: https://doi.org/10.5935/jetia.v10i50.1311

ISSN ONLINE: 2447-0228

RESEARCH ARTICLE OPEN ACCESS

ESTIMATION OF THE TIME OF OCCURRENCE OF THE MAXIMUM ELECTRICAL DEMAND BY SELECTING THE OPTIMAL CLASSIFICATION MODEL AND MAKING USE OF UNBALANCED DATA

César A. Yajure Ramírez[1] and Valesca M. Fuenzalida Sánchez[2]

[1] Universidad Central de Venezuela.
[2] Pontificia Universidad Católica de Chile.

[1]http://orcid.org/0000-0002-3813-7606, [2]http://orcid.org/0009-0001-6878-3852

Email: cyajure@gmail.com, vfuenzalidas@gmail.com.

ARTICLE INFO

Article History
Received: September 28, 2024
Revised: October 15, 2024
Accepted: November 10, 2024
Published: December 31, 2024

Keywords:
Evaluation metrics,
Machine learning,
Multi-criteria analysis,
Unbalanced classes,
Weighted sum.

ABSTRACT

Studies on electricity demand forecasting usually focus on the magnitude of the variable, however, the methodology used in this study also addresses the time at which the peak demand occurs, crucial for planning energy generation, smoothing the demand peaks and establishing differentiated rates. To predict the time of maximum demand, supervised machine learning algorithms were used: random forests, K nearest neighbors, support vector machine, and logistic regression. The dataset consists of hourly maximum and minimum demand data from 2021 to 2024 for a country in South America, including environmental factors such as temperature and seasonality. Since the data in the peak demand prediction variable is unbalanced, the study used oversampling techniques such as SMOTE-NC (synthetic instances of the minority classes to balance the data set). A multi-criteria decision-making approach is used to select the best classification model, considering model evaluation metrics as decision criteria. The most important conclusion drawn by the study is that the model obtained with the support vector machine algorithm turned out to be optimal, and successfully predicted the time of maximum demand on 15 of the 17 test days. The findings highlight the unbalanced nature of peak demand hours, which predominantly occur around 8 pm.

I. INTRODUCTION

Electrical demand forecasting methodologies usually focus on the magnitude of this variable, but in addition to this, determining the time in which this maximum demand occurs is very important for several reasons. For example, it is required for the purposes of determining the electrical generation that must be available to satisfy this maximum demand and thus smooth its peak [1]. Likewise, the company in charge of the electrical system could seek to reduce that maximum demand or temporarily transfer it to "flatten" the load curve, which brings us to the second reason, which is the establishment of differentiated rates for those high hours. demand [2]. Likewise, the temporal dependence between social practices and electrical energy consumption has become evident, with some of these specific practices having greater temporal dependence than others [3]. Then, knowing the time at which maximum demand occurs, the regulatory entity could implement policies for the rational and efficient use of electrical energy according to the types of users of the electrical system.

In addition, among the usual applications of forecasting the times of occurrence of maximum demand are smoothing of demand peaks, but also scheduling the charging of electric vehicles [4]. On the other hand, knowing the usual time at which the minimum power demand occurs is useful for the purposes of planning the mandatory shutdown of generation units for maintenance purposes, or even for planning the maintenance of elements of the transmission system and/or electrical distribution.

Therefore, the objective of this research is to forecast the time w3hen maximum demand will occur in a given geographic region, using data science methodology and selecting the best classification model. The supervised machine learning algorithms considered were random forests (RF), K-nearest neighbors (K-NN), support vector machine (SVC), and logistic regression (LR). To choose the best classification model, a multi-criteria technique is used in which the proposed models are the decision alternatives, and the performance evaluation metrics of the models are the decision criteria.

The topic of this research has been addressed previously, but without considering the multi-criteria approach. For example, [5] develop different models, one to predict whether the next day will be the day of maximum demand of the month, and another to predict the time of peak demand. They use machine learning algorithms to develop the models, and consider the maximum and minimum temperatures, among others, as explanatory variables of the models. The methodology was applied to the Duke energy system in North Carolina, United States. Of the 72 months of data considered, in 69 of them the models got the day of peak demand correct, and in 90% of the peak days, the actual time of peak demand was among the 2 hours with the highest probability. Likewise, [6] build classification models to predict the time of daily peak demand 24 hours in advance. They use several classification machine learning algorithms: Näive Bayes, Support Vector Machines (SVM), Random Forests, Adaptive Boosting (AdaBoost), Convolution Neural Network (CNN), LSTM neural network, and autoencoder type artificial neural network. The data used corresponds to the maximum hourly demand of the city of Ontario in Canada for the period from May 2003 to April 2008, differentiating the winter period from the summer period. They obtain that the artificial neural network is the one that has the best performance for both the winter period and the summer period. Furthermore, [7] presents an approach for developing a peak hour forecasting model, selecting the optimal model by applying a series of machine learning algorithms. To evaluate the model, they work with data from 57 regions of Russia corresponding to the period from January 2016 to January 2020. The algorithms considered were random forest classifier (RFC), decision tree classifier (DTC), nearest neighbors classifier (K-NN), and extra-tree classifier (ETC). The evaluation of the model performance is made based on the accuracy of the actual peak hour with respect to the predicted peak hour, considering one-, two-, and three-hour intervals of the most probable peak hours. The highest accuracy was obtained using the extra-tree classifier.

In their research, [8] presents a methodology to predict coincident peak demand events, including the day and time of event occurrence. The approach is based on performing Monte Carlo simulations to generate scenarios and estimators for this type of event. Additionally, [4] presents an open-source tool for electric energy forecasting through which a variety of methods for forecasting maximum demand, as well as its time of occurrence, are implemented. One of the case studies presented consisted of forecasting the hours of the day or days in a month or year, in which the maximum demand occurs. They worked with electricity consumption data, with a 5-minute resolution, from the New England region in the United States in 2020. It turned out that the model obtained with an artificial neural network of the LSTM type had better performance, with values for the metrics precision, recall, and accuracy, of 0.84, 0.84, and 0.83, respectively. Similarly, [9] proposes a model based on deep learning, to predict the k hours of the day with the highest and lowest demands. They evaluate their model using data from two years of electrical demand from an electrical microgrid that supplies 156 buildings. Its model based on an artificial neural network of the LSTM type was compared with other models: Linear Regression, Arima, and artificial neural network. They obtained that their four-layer LSTM network model had the best performance for both the k hours of high demand and for the k hours of low demand. Finally, [10] present a methodology to forecast the magnitude of maximum demand, as well as its time of occurrence, using two machine learning algorithms: Multiple Linear Regression and Gradient Boosting Machine. Among the explanatory variables, they consider the time, day of the week, month, holidays, and temperature. The results indicate that the regression performs better during seasons with low time variability, while the ensemble methods show greater accuracy in general.

The rest of the article is distributed as follows. Section 2 presents the theoretical background. Then, in section 3 the materials and method used in the research. Next, in section 4 the results obtained are analyzed and discussed. After that, there are the conclusions derived from the research carried out. Finally, bibliographic references are presented.

II. THEORETICAL REFERENCE

II.1 CLASS BALANCING TECHNIQUES

A data set is said to be imbalanced if the class we are interested in falls in the minority class and appears sparsely compared to the majority class, the minority class is also known as positive class, while the majority class is also known as negative class [11]. On the other hand, class imbalance is described as a large discrepancy between two classes of the same target variable, where one class is represented by many instances, while the other is only represented by a small number of instances [12].

Two strategies can be considered to counteract class imbalance: the data sampling approach, and the cost-sensitive learning approach. In this research we worked with the first approach, which could be approached using a sub-sampling method or an over-sampling method. The sampling methodology that selectively strips the majority class, while ensuring that the data set retains the meaningful information associated with this majority class, is known as a sub-sampling approach. While the sampling approach in which instances of the minority class are frequently replicated until a balanced class distribution is reached is known as an oversampling approach. In this work, over-sampling methods are used, among which the SMOTE method (Synthetic Minority Over-sampling Technique), the SMOTE-NC method (Synthetic Minority Over-sampling Technique for Nominal and Continuous), among others, stand out.

II.2 CLASSIFICATION MODELS PERFORMANCE METRICS

The following metrics are considered to evaluate the performance of the classification models: Accuracy, Precision, Recall, and F1. According to what was mentioned by [13], Accuracy “is defined as the sum of all correct predictions divided by the sum of all predictions”, it is obtained using Equation (1). Precision “is related to the number of correct positive predictions,” and is generated through Equation (2). The Recall metric “is related to the number of positive events correctly predicted” is achieved using Equation (3). The F1 metric “is known as the harmonic mean between precision and recall”, is computed with Equation (4).

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN} \tag{1}$$

$$Precision = \frac{TP}{TP+FP} \tag{2}$$

$$Recall = \frac{TP}{TP+FN} \tag{3}$$

$$F1 = \frac{2\cdot(precisión\cdot recall)}{precisión+recall} \tag{4}$$

Where:

TP: True positives
TN: True negatives
FP: False positives
FN: False negatives

II.3 MULTI-CRITERIA DECISION MAKING

Multicriteria Decision Making (MCDM) is related to the treatment of decision problems in which more than one decision criterion is present to be considered for choosing the best option, within a group of alternatives [14]. MCDM is divided into two: multi-objective decision making (MODM), and multi-attribute decision making (MADM). Typically, MODM has an explicit goal and a continuous decision space, while MADM has an implicit goal and a discrete decision space [15].

The multi-attribute decision problem can be represented by its decision matrix of M rows and N columns. The element a_{ij} of this matrix shows the performance of the alternative A_i when evaluated by the decision criterion C_j, (i = 1, 2, 3, ..., M, and j = 1, 2, 3, ..., N). Each of the criteria has a relative importance weight wj, which is generally determined by the "decision maker". This is how, given a set of alternatives and decision criteria, we seek to establish the optimal alternative with the highest degree of "desirability" with respect to the decision criteria [16].

Ishizaka & Nemery [17] mention the taxonomy of decision problems proposed by Roy in 1981, that is: selection problems, classification problems, hierarchy problems, and description problems. In this research, the selection problem is addressed, in which the goal is to select the best alternative, according to a set of predetermined criteria. Similarly, we work with the Simple Additive Weighting (SAW) multi-criteria technique, also known as the weighted sum method, which is one of the simplest and most used decision-making methods [18]. It consists of four steps: preparing the decision matrix (M×N), preparing the normalized decision matrix, calculating the magnitude that represents the performance of each alternative, and ranking the alternatives according to their performance to select the best of them.

To generate the matrix of the second step, each of its components must be normalized using Equation (5) if it is a cost criterion, and Equation (6) if it is a benefit criterion. Likewise, to obtain the magnitude S_i of alternative i, mentioned in the third step, Equation (7) must be used.

$$\widetilde{a_{ij}} = \frac{\min a_{ij}}{a_{ij}} \tag{5}$$

$$\widetilde{a_{ij}} = \frac{a_{ij}}{\max a_{ij}} \tag{6}$$

$$S_i = \sum_{j=1}^{N} w_j \cdot \widetilde{a_{ij}} \tag{7}$$

III. MATERIALS AND METHODS

To analyze the data, the stages that make up a typical data science project were followed. These stages are establishment of the project objective(s), search for the data to be used, data processing, exploratory data analysis, data modeling, and decision making [19]. Figure 1 presents a diagram of the stages followed in the methodology.

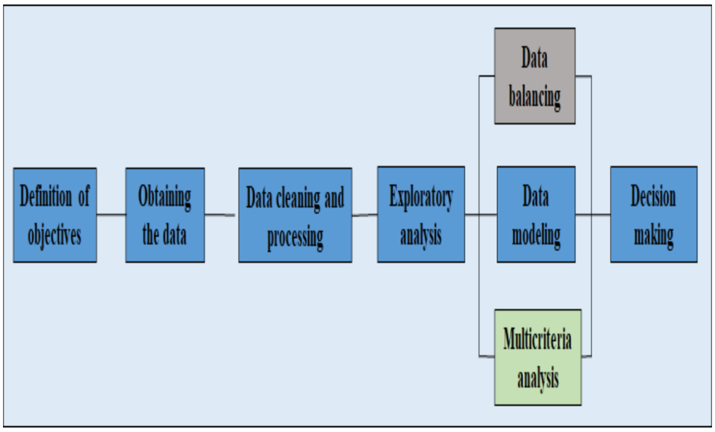

Figure 1: Research methodology.
Source: Authors, (2024).

The objective is to determine the time of occurrence of the maximum demand in a geographic area. The data is obtained from external and/or internal sources, in this case they correspond to the measurements of the maximum and minimum electrical demand, and their times of occurrence, for the area under study. Once the data is obtained, it usually must be cleaned and processed, applying the techniques described in [20]. For the exploratory data analysis, graphical and non-graphical descriptive statistics techniques are used, and from this analysis significant information could be obtained from the data. Modeling is then done using machine learning algorithms. Specifically, classification algorithms are used to forecast the time at which peak demand will occur in the coming days. With the results obtained in the two previous stages, we proceed to the decision-making stage. The stages mentioned so far, corresponding to the blue blocks in Figure 1, are typical of a data science project. In this research, a balancing stage is incorporated, since the data used is unbalanced and this characteristic affects the result returned by the models. Likewise, the multi-criteria analysis stage is incorporated to select the optimal classification model that is used to predict the time of occurrence of the maximum demand.

All the stages just presented are developed using the Python programming language and its respective libraries.

III.1 DATA CLEANING AND PROCESSING

The original data consists of the historical time series of the maximum hourly demand for the period 2021-2024 of a South American country. This series is processed to generate a time series with daily resolution, which contains the maximum demand of the day, the time at which this demand occurs, the minimum demand for the day, and the time at which this minimum demand occurs.

Additionally, the series includes the year, month, week, and day of the week. This daily time series includes information on holidays and working days, as well as a column for the maximum ambient temperature, and another for the average ambient temperature.

Considering the column of the months of the year, a column is created that indicates whether the corresponding day belongs to the historical rainy period or the historical drought period, since during the drought period the ambient temperature increases, and consequently the electricity demand grows, somehow affecting the time of occurrence of this demand.

Finally, from the peak demand hour column, a column is created that indicates whether for the respective day, the peak demand hour belongs to the afternoon hours (around 2 pm) or belongs to the hours of the night (around 8 pm).

IV. RESULTS AND DISCUSSIONS

This section analyzes the results obtained in the exploratory data analysis and data modeling stages.

IV.1 EXPLORATORY DATA ANALYSIS

First, it is important to determine the distribution of hours of maximum demand throughout the 2021-2024 study period, which is presented in Figure 2.

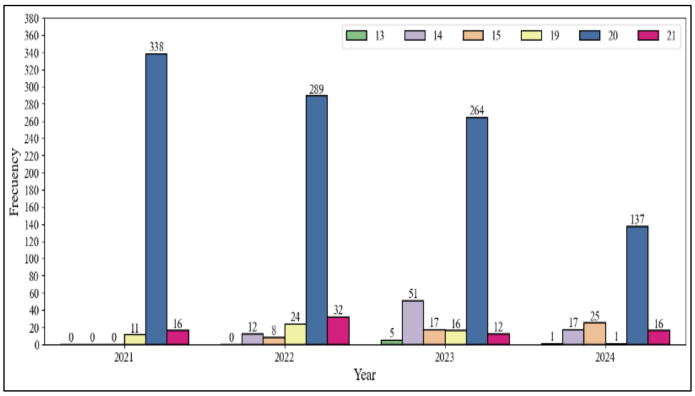

Figure 2: Distribution of hours of maximum demand per year.
Source: Authors, (2024).

From Figure 2, during 2021, the hours of maximum demand were all at night (hours 19, 20, and 21 of the day), which makes sense since in that year there was a greater presence of teleworking due to the pandemic. Likewise, for the entire period the time of maximum demand was mostly 8 pm, and that starting in 2022 the afternoon hours began to have a certain presence in the data on the time of maximum demand.

On the other hand, Figure 3 shows the distribution of the hours of maximum demand for each of the months of the year. It is observed that during the months of the last third of the year, 7 pm as the time for maximum demand has the greatest incidence, while until the month of August its presence is almost non-existent. Additionally, it is noted that the hours of maximum daytime demand intensify from the month of June until the month of October.

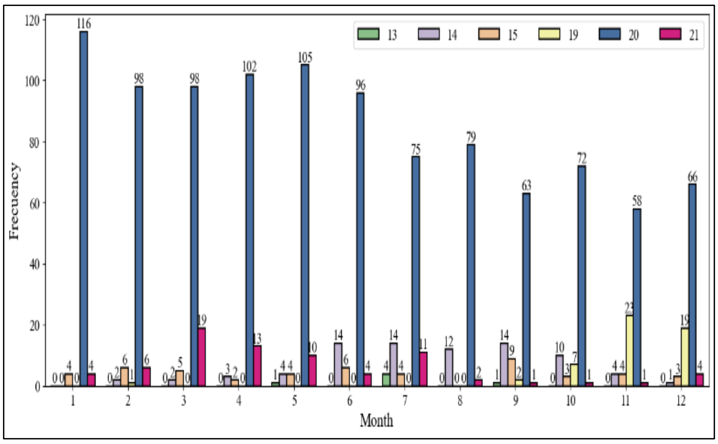

Figure 3: Distribution of hours of maximum demand per year.
Source: Authors, (2024).

Likewise, it is of interest to determine the distribution of the hours of maximum demand by weekday, which is presented in Figure 4. In this case, Monday is represented by the number “1”, and Sunday is represented with the number “7”.

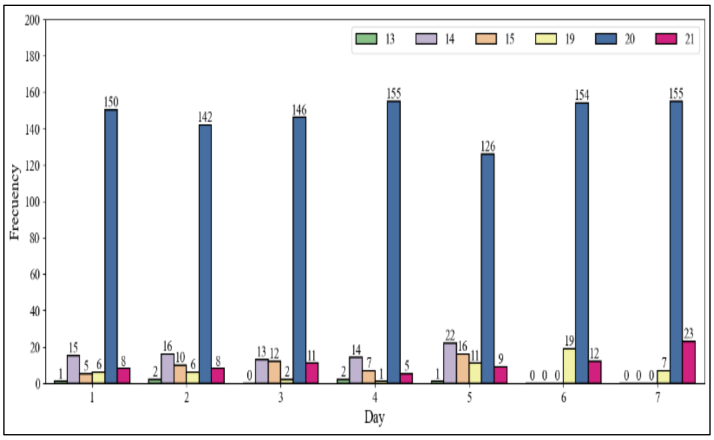

Figure 4: Distribution of hours of maximum demand by day.
Source: Authors, (2024).

It can be noted that, on Saturdays and Sundays, the hours of maximum demand correspond only to nighttime hours. On the other hand, it is observed that Fridays (day 5) are the days on which there are more hours of maximum demand during the afternoon, and they are also the days on which fewer hours of maximum demand coincide with 8 pm.

Additionally, Figure 5 presents the distribution of hours of maximum demand according to the type of day, to compare the behavior of weekdays with holidays and weekends. During weekdays 8 pm predominates, and 15.4% of the hours of maximum demand occur during the afternoons. The behavior of holidays is like that of Saturdays and Sundays, with 100% of the hours of maximum demand during the night and with a proportion of around 80% of the hours coinciding with 8 pm.

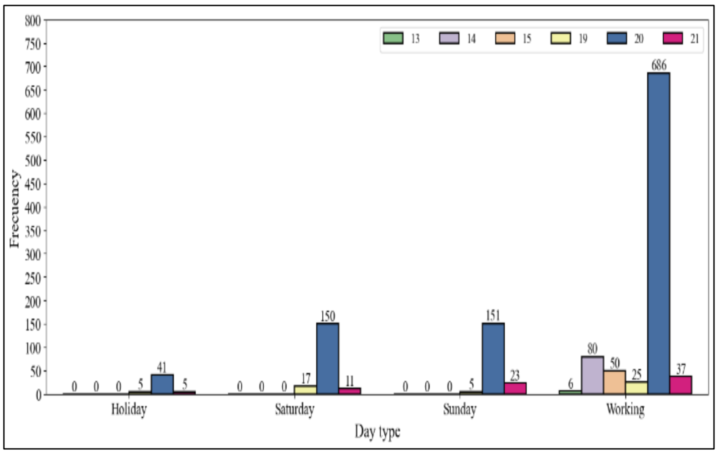

Figure 5: Distribution of hours of maximum demand by type of day.
Source: Authors, (2024).

IV.2 DATA MODELING

Next, the data modeling is presented to estimate the time of occurrence of the maximum electrical demand, for which supervised machine learning classification algorithms are used: K nearest neighbors, random forests, support vector machine, and logistic regression. In [21] the same classification algorithms are used to estimate the performance ratio of a solar PV plant, which also corresponds to a binary classification problem.

In this research, use is made of the daily maximum electricity demand data for the period 2023-2024, until 07/14/2024, to generate and train the models, and the remaining seventeen days until 07/31/2024 are used. to recalculate the performance evaluation metrics of said models. Likewise, data balancing techniques are used in combination with each of the models, since as observed in the exploratory analysis of the data, they are unbalanced around 8 pm, the time of maximum maximum demand. Specifically, the SMOTE-NC method was used, which is convenient when there is a data set with both numerical variables and categorical variables.

For each of the models, the data set was divided into two parts: training and testing, with a proportion of 80% for training and 20% for testing. Two possible times are considered: 2 pm and 8 pm, both with a variation of more or less one hour.

IV.2.1 RANDOM FOREST ALGORITHM

This is an algorithm that uses the bagging technique, which means aggregation by resampling, involving the manipulation of the training data set with resampling. From the training data set, multiple data sets are generated for training, and classification models are generated from each of them [22]. By having the classification models, predictions are made with each of them and the results are combined by taking the majority vote, that is, the result that is most repeated in the case of classification problems [23].

For the parameterization of the model, the number of estimators was set at 150, and to measure the quality of the forest the "gini" metric was used. To evaluate the model, the metrics were used: Accuracy, Precision, F1, and Recall. Table 1 shows the results obtained from the evaluation of the model in the testing phase.

Subsequently, the model is used to estimate the hour of maximum demand for the following seventeen days, from July 15 to 31, considering two options: 2 pm and 8 pm, both classes with a tolerance of 1 hour, that is, a range from 1 pm to 3 pm (13_14_15), and the other range from 7 pm to 9 pm (19_20_21). The results are compared with the actual values, and the new values of the evaluation metrics are presented in Table 2, while the estimated hours are presented in Table 3.

IV.2.2 K NEAREST NEIGHBORS ALGORITHM

This is an algorithm that is used both to generate regression and classification models, and falls within the category of supervised machine learning, and is also non-parametric [24]. However, the basic principle of this algorithm is to consider elements that are similar to each other (closest neighbors), so it is required to previously know the number of neighbors K to take into account. There are several methods to obtain the value of K, for example, [25] propose using cross-validation techniques. In our case, the optimal value of K was obtained by plotting the number of nearest neighbors versus accuracy, and the value of K that maximized this metric was selected. After evaluating the model in the testing phase, the values of the metrics presented in Table 1 were obtained.

Next, the model is used to estimate the hour of maximum demand for the period from July 15 to 31, considering two options: 2 pm and 8 pm, both classes with a tolerance of 1 hour, that is, a range from 1 pm to 3 pm, and the other range from 7 pm to 9 pm. The results are presented in Tables 2 and 3.

IV.2.3 SUPPORT VECTOR MACHINE ALGORITHM

Like the previous models, support vector machines can be used for both classification problems and regression problems, using the same operating principle. This algorithm uses the concept of kernel to convert the given data into a higher dimension, to achieve the so-called hyperplanes. The points located on each side of the hyperplane and that are closest to it are known as support vectors. There are four main types of kernels, namely linear, polynomial, sigmoid and radial basis function [26].

For our case study, the support vector classifier is used, and its default parameters are considered, which includes a radial basis function kernel. The trained model is used to make the forecast from the test data, and the results obtained are evaluated through the corresponding metrics. The metric values are presented in Table 1.

Next, the model is used to estimate the hour of maximum demand for the period from July 15 to 31, considering two options: 2 pm and 8 pm, both classes with a tolerance of 1 hour, that is, a range from 1 pm to 3 pm, and the other range from 7 pm to 9 pm. The results are presented in Tables 2 and 3.

IV.2.4 LOGISTIC REGRESSION ALGORITHM

Through the application of this algorithm, models for binary classification can be generated. According to [27], this technique "Is one of the most used linear statistical models for discriminant analysis." After performing a linear regression, the algorithm converts the output of this regression through a logistic function (hence its name), which is commonly the sigmoid function. This last function assigns a conditional probability for each of the classes. When applying this algorithm, all default parameters were taken, which included the Limited-memory Broyden, Fletcher, Goldfarb, and Shanno (lbgfs) optimization method. As was done with the other algorithms, the trained model is used to make predictions using the test set, and it was evaluated considering the corresponding metrics, whose values are presented in Table 1.

As with the previous models, Tables 2 and 3 present the results obtained after using the model to estimate the time of maximum demand for the period from July 15 to July 31.

Table 1: Metric results in the testing phase.

Metric	K-NN	RF	SVC	LR
Accuracy	0.765	0.844	0.777	0.737
Precision	0.793	0.848	0.807	0.735
F1	0.761	0.843	0.769	0.726
Recall	0.770	0.843	0.772	0.723

Source: Authors, (2024).

From Table 1 it can be noted that the random forest algorithm was the one that had the best performance according to the evaluation metrics considered. From Table 2 it can be noted that the K-NN and SVC algorithms were the best evaluated according to the metrics used.

Table 2: Metric results with new data.

Metric	K-NN	RF	SVC	RL
Accuracy	0.882	0.823	0.882	0.824
Precision	0.882	0.875	0.882	0.826
F1	0.882	0.813	0.882	0.824
Recall	0.882	0.813	0.882	0.826

Source: Authors, (2024).

In the same order of ideas, Table 3 presents the estimates of the time of maximum demand. It can be noted that using the K-NN model, the actual time of maximum demand was within the estimated range on 15 of the 17 days analyzed, as was the SVC model. In the case of the RF and RL models, the success was on 14 of the 17 days.

Table 3: Maximum demand hour estimation results.

Date	Actual time	K-NN	RF	SVC	RL
7/15/2024	14	13_14_15	19_20_21	13_14_15	13_14_15
7/16/2024	20	19_20_21	19_20_21	19_20_21	13_14_15
7/17/2024	15	13_14_15	19_20_21	13_14_15	13_14_15
7/18/2024	15	19_20_21	19_20_21	19_20_21	19_20_21
7/19/2024	20	13_14_15	19_20_21	13_14_15	13_14_15
7/20/2024	20	19_20_21	19_20_21	19_20_21	19_20_21
7/21/2024	20	19_20_21	19_20_21	19_20_21	19_20_21
7/22/2024	14	13_14_15	13_14_15	13_14_15	13_14_15
7/23/2024	15	13_14_15	13_14_15	13_14_15	13_14_15
7/24/2024	20	19_20_21	19_20_21	19_20_21	19_20_21
7/25/2024	14	13_14_15	13_14_15	13_14_15	13_14_15
7/26/2024	14	13_14_15	13_14_15	13_14_15	13_14_15
7/27/2024	20	19_20_21	19_20_21	19_20_21	19_20_21
7/28/2024	20	19_20_21	19_20_21	19_20_21	19_20_21
7/29/2024	13	13_14_15	13_14_15	13_14_15	13_14_15
7/30/2024	20	19_20_21	19_20_21	19_20_21	19_20_21
7/31/2024	20	19_20_21	19_20_21	19_20_21	19_20_21

Source: Authors, (2024).

IV2.5 MULTI-CRITERIA MODEL SELECTION

The SAW method is used to select the best model. The decision criteria to consider are Accuracy and F1, obtained in the testing phase, Accuracy and F1 obtained with new data, and the number of days that the respective model guessed correctly the time of maximum demand. The alternatives of the decision problem are the four models considered: K-NN, RF, SVC, and RL. The normalized decision matrix is shown in Table 4.

Table 4: Normalized decision matrix.

Decision alternatives	Decision criteria				
	Accuracy test	F1 test	Accuracy new	F1 new	Days of success
K-NN	0.906	0.903	1.000	1.000	1.000
RF	1.000	1.000	0.933	0.922	0.933
SVC	0.921	0.912	1.000	1.000	1.000
RL	0.873	0.861	0.934	0.934	0.933

Source: Authors, (2024).

After calculating the magnitudes of each of the alternatives, and ranking them from highest to lowest, the results presented in Table 5 are obtained, from which it is deduced that the optimal classification model is the one derived from the support vector machine algorithm (SVC), followed by the K-nearest neighbors (K-NN) model.

Table 5: Model hierarchy.

Model	S_i
SVC	0.982
K-NN	0.979
RF	0.945
RL	0.919

Source: Authors, (2024).

V. CONCLUSIONS

A methodology is presented that allows estimating the time of occurrence of the maximum electrical demand, using the optimal classification model, which was selected using multicriteria decision-making analysis. Within the methodology, a stage is included to balance the classes. After applying the weighted sum multicriteria decision-making method, it was obtained that the model derived from the support vector machine algorithm is optimal for estimating the time of occurrence of the maximum electrical demand. When evaluating the models with new data, this model correctly predicted the time of occurrence on 15 of the 17 days considered. According to historical data, the time of occurrence is in the afternoon around 2 pm, or at night around 8 pm. There is an imbalance in these hours of occurrence, with a clear tilt towards 8 pm, with a ratio of 75/25 with respect to all other hours for the period 2021-2024, and a ratio of 88/12 in the hours of occurrence. occurrence of the night with respect to the hours of the day, for the same study period.

VI. AUTHOR'S CONTRIBUTION

Conceptualization: César Yajure.
Methodology: César Yajure.
Investigation: César Yajure.
Discussion of results: César Yajure.
Writing – Original Draft: César Yajure and Valesca Fuenzalida.
Writing – Review and Editing: César Yajure and Valesca Fuenzalida.
Resources: César Yajure.
Supervision: César Yajure.
Approval of the final text: César Yajure.

VIII. REFERENCES

[1] Y. Xie, «Peak Load Hour Prediction,» 01 01 2022. [En línea]. Available: https://exergyenergy.com/wp-content/uploads/2022/01/Peak-load-Prediction-White-Paper.pdf.

[2] C. Candia, «ENERLINK,» 10 Julio 2024. [En línea]. Available: https://blog.enerlink.com/las-claves-del-cobro-de-demanda-en-horas-punta. [Último acceso: 10 Julio 2024].

[3] J. Torriti, «Understanding the timing of energy demand through time use data: Time of the day dependence of social practices,» Energy Research and Social Science, pp. 37-47. https://doi.org/10.1016/j.erss.2016.12.004, 2017.

[4] P. Bovornkeeratiroj, J. Wamburu, D. Irwin y P. Shenoy, «PeakTK: An Open Source Toolkit for Peak Forecasting in Energy Systems,» de Computing and Sustainable Societies, Seattle, WA, USA, 2022.

[5] T. Fu, H. Zhou, X. Ma, Z. J. Hou y D. Wu, «Predicting peak day and peak hour of electricity demand with ensemble machine learning,» Frontiers in Energy Research, pp. 1-11. https://doi.org/10.3389/fenrg.2022.944804, 2022.

[6] J. Liu y L. E. Brown, «Prediction of Hour of Coincident Daily Peak Load,» de IEEE Power & Energy Society Innovative Smart Grid Technologies Conference, Washington DC, USA , 2019.

[7] V. Voronin, F. Nepsha y M. Krasilnikov, «Short term forecasting peak load hours of regional power systems using machine learning methods,» CIGRE Science & Engineering (CSE), pp. 1-18, 2023.

[8] R. Carmona, X. Yang y C. Zeng, «Coincident Peak Prediction for Capacity and Transmission Charge Reduction,» arXiv:2407.04081, p. https://doi.org/10.48550/arXiv.2407.04081, 2024.

[9] A. Soman, A. Trivedi, D. Irwin, B. Kosanovic, B. McDaniel y P. Shenoy, «Peak Forecasting for Battery-based Energy Optimizations in Campus Microgrids,» de e-Energy '20: Proceedings of the Eleventh ACM International Conference on Future Energy Systems, Virtual, 2020.

[10] D. L. Donaldson, J. Browell y C. Gilbert, «Predicting the magnitude and timing of peak electricity demand: A competition case study,» IET Smart Grid, pp. 1-12. https://doi.org/10.1049/stg2.12152, 2023.

[11] A. Babajide Ebenezer1, O. Boyinbode y M. I. Oladunjoye, «A Comprehensive Analysis of Handling Imbalanced Dataset,» International Journal of Advanced Trends in Computer Science and Engineering, pp. 454-463. https://doi.org/10.30534/ijatcse/2021/031022021, 2021.

[12] T. Wongvorachan, S. He y O. Bulut, «A Comparison of Undersampling, Oversampling, and SMOTE Methods for Dealing with Imbalanced Classification in Educational Data Mining,» Information MDPI, pp. 4-15. https://doi.org/10.3390/info14010054, 2023.

[13] W. M. Lee, Python Machine Learning, Indianapolis: John Wiley & Sons, Inc., 2019.

[14] C. A. Yajure-Ramírez, «Enfoque multicriterio para la selección óptima de variables explicativas para modelos de pronóstico de la energía eléctrica de plantas solares fotovoltaicas,» Revista Tecnológica - Espol, pp. 83-98. https://doi.org/10.37815/rte.v35n3.1045, 2023.

[15] F. Eltarabishi, O. H. Omar, I. Alsyouf y M. Bettayeb, «Multi-Criteria Decision Making Methods And Their Applications– A Literature Review,» de Proceedings of the International Conference on Industrial Engineering and Operations Management, Dubai, UAE, 2020.

[16] E. Triantaphyllou, B. Shu, S. Nieto Sanchez y T. Ray, «Multi-Criteria Decision Making: An Operations Research Approach,» Encyclopedia of Electrical and Electronics Engineering, pp. 175-186., 1998.

[17] A. Ishizaka y P. Nemery, Multi-Criteria Decision Analysis - Methods and Software, West Sussex, United Kingdom: John Wiley & Sons, Ltd, 2013.

[18] H. Taherdoost, «Analysis of Simple Additive Weighting Method (SAW) as a Multi-Attribute Decision-Making Technique: A Step-by-Step Guide,» Journal of Management Science & Engineering Research, pp. 21-24. DOI: https://doi.org/10.30564/jmser.v6i1.5400, 2023.

[19] D. Cielen, A. D. B. Meysman y M. Ali, Introducing Data Science, Shelter Island, NY: Manning Publications Co., 2016.

[20] S. K. Mukhiya and U. Ahmed, Hands-On Exploratory Data Analysis with Python, Birmingham, UK: Packt Publishing Ltd., 2020.

[21] C. A. Yajure Ramírez, «Selección del modelo óptimo de predicción de la relación de desempeño de una planta solar fotovoltaica. Un enfoque multicriterio basado en algoritmos de aprendizaje automático,» Ciencia, Ingenierías y Aplicaciones INTEC, pp. 7-29. DOI: https://doi.org/10.22206/cyap.2023.v6i2.2935, 2023.

[22] J. Rogel-Salazar, Data Science and Analytics with Python, Boca Raton, FL: CRC Press Taylor & Francis Group, 2017.

[23] E. Russano y E. Ferreira Avelino, Fundamentals of Machine Learning Using Python, Oakville, Canadá: Arcler Press, 2020.

[24] M. E. Fenner, Machine Learning with Python for Everyone, Boston: Pearson Education, Inc., 2020.

[25] M. Maleki, N. Manshouri y T. Kayıkçıoğlu, «A Novel Simple Method to Select Optimal k in k-Nearest Neighbor Classifier,» International Journal of Computer Science and Information Security, pp. 464-469, 2017.

[26] R. Muthukrishnan y M. Jamila. S, «Predictive Modeling Using Support Vector Regression,» International Journal of Scientific & Technology Research, pp. 4863-4865., 2020.

[27] K. Kirasich, T. Smith y B. Sadler, «Random Forest vs Logistic Regression: Binary Classification for Heterogeneous Datasets,» SMU Data Science Review, p. https://scholar.smu.edu/datasciencereview/vol1/iss3/9, 2018.