Araştırma Makalesi / Research Article

Developing Prediction System for Solar Power Plant Using Machine Learning Algorithms

Yusuf UZUN*

Necmettin Erbakan University, Faculty of Seydisehir A.C. Engineering, Department of Computer Engineering, Konya, Turkey
(ORCID: 0000-0002-7061-8784)

Abstract
The use of renewable energy sources in the production of electricity has become inevitable in order to reduce the greenhouse gases left in the atmosphere that cause the Earth to warm up. Although countries on a national basis have implemented a number of policies to support electricity generated from renewable energy sources, investments to produce electricity without a license on a local basis are not desirable. According to the climatic conditions of the power plant of 1 MW installed founded in Konya and power plant production data are monitored. Machine learning is a sub-branch of artificial intelligence that deals with the design and development of algorithms that allow computers to develop their behavior based on experimental data. In this study, Naive Bayes, Decision Tree, CN2 Rule Induction, Random Forest, Support Vector Machine, k-Nearest Neighbor, Artificial Neural Network, Logistic Regression and AdaBoost machine learning algorithms are used for prediction and classification. Generally, energy investors are curious about the return on their investment. It is very important for energy providers to predict how much electricity will be generated from existing solar power plants and accordingly determine the measures they will take to meet the electricity demand in the future. ROC analyzes were performed for machine learning models and performance evaluation was performed. In this study, the best performance estimation value obtained from the solar power plant depending on the weather conditions was obtained with 92.24% accuracy.

Keywords: Renewable sources, machine learning, ROC analysis.

Makine Öğrenmesi Algoritmalarını Kullanarak Güneş Enerjisi Santrali İçin Tahmin Sistemi Geliştirme

Öz
Dünya'nın ısınmasına neden olan atmosfere bırakılan sera gazlarını azaltmak için elektrik üretiminde yenilenebilir enerji kaynaklarının kullanılması kaçınılmaz hale gelmiştir. Ulusal bazda olan ülkeler, yenilenebilir enerji kaynaklarından üretilen elektriği desteklemek için bir dizi politika uygulamış olsalar da, yerel olarak lisanssız elektrik üretmek için yatırım yapılması istenmemektedir. Konya'da kurulan 1 MW'lık santralin iklim koşullarına göre santral üretim verileri izlenmektedir. Makine öğrenmesi, bilgisayarların deneysel verilere dayalı davranışlarını geliştirmelerine izin veren algoritmaların tasarımı ve geliştirilmesi ile ilgilenen yapay zekânın alt koludur. Bu çalışmada Naive Bayes, Karar Ağacı, CN2 Kural İndüksiyonu, Rastgele Orman, Destek Vektör Makinesi, k-En Yakın Komşu, Yapay Sinir Ağı, Lojistik Regresyon ve AdaBoost makine öğrenme algoritmaları, tahmin ve sınıflandırma işlemleri için kullanılmıştır. Genel olarak, enerji yatırımcıları yatırımlarının geri dönüşünü merak etmektedirler. Enerji tedarikçileri için mevcut güneş enerji santrallerinden ne kadar elektrik üretileceğini tahmin etmek ve buna bağlı olarak gelecekte elektrik talebini karşılamak için alabilecekleri önlemleri belirlemek çok önemlidir. Makine öğrenmesi modelleri için ROC analizleri ve performans değerlendirmesi yapıldı. Bu çalışmada, hava şartlarına bağlı olarak güneş enerjisi santralinden elde edilen en iyi performans tahmin değeri % 92,24 doğrulukla elde edilmiştir.

Anahtar kelimeler: Yenilenebilir kaynaklar, makine öğrenmesi, ROC analizi.

*Sorumlu yazar: yuzun@erbakan.edu.tr

Geliş Tarihi: 18.06.2019, Kabul Tarihi: 05.12.2019

1. Introduction

Since electric energy is easily transformed into energy types, it is one of the most widely used energy types today. As a result of the use of coal and petroleum-based energy sources used in electricity generation, greenhouse gases released to the atmosphere cause global warming. In global warming, floods in some parts of the country due to climate change are seen as global problems such as drought in some regions and desertification in the world. The constant increase of the human population and the widespread use of electric energy cause these problems to grow even more. Energy supply companies are trying to meet this increasing energy demand. For the energy supply companies, the change in the production amounts due to the climate conditions in the power plants that generate electricity from the solar, wind and wave energy is an important problem. Accurate estimation of the amount of energy produced by these plants before production is very important in terms of network reliability. Inaccurate estimation of the amount of energy produced by renewable energy sources, voltage drop in the network, frequency change etc. causes undesirable situations. Machine learning is a discipline related to learning and extracting rules from data. It is used to enable computers to learn information by using a set of experiences. These methods used in machine learning include concept learning from data. Researchers grouped machine learning techniques into three categories. These categories are active learning through interaction method, learning by using previous knowledge and learning step by step [1]. Machine learning is the operation of a number of numerical methods in order to improve system performance by gaining knowledge from the experiences obtained [2]. Yumurtaci suggested an ANN controller approach to provide energy management and performance of a hybrid system including hydrogen, wind and solar energy technologies [3]. Kulaksiz et al., presented a Genetic Algorithm (GA) method to improve the maximum power point tracing capacity of a photovoltaic system [4]. Sharma et al., proposed an optimal power point tracking and control method for a hybrid renewable energy system under a independent environment [5]. Raju et al., suggested the application of the improved distributed energy management system and request management of a solar microgrid using a multiagent system coordination method [6].

Researchers working in the field of machine learning are working on a number of methods to obtain accurate and understandable rules and to achieve high accuracy estimates. In this study, a prediction system will be developed by obtaining correct, understandable and valid rules using Naive Bayes, Decision Tree, CN2 Rule Induction, Random Forest, Support Vector Machine, k-Nearest Neighbor, Artificial Neural Network, Logistic Regression and AdaBoost machine learning algorithms. ROC analyzes were performed for machine learning models and performance evaluation was performed. Estimation and rule extraction will be carried out on the dataset obtained from the solar power plant.

2. Material and Methods

2.1. Machine learning algorithms

In this study, Naive Bayes, Decision Tree, CN2 Rule Induction, Random Forest, Support Vector Machine (SVM), k-Nearest Neighbor (kNN), Artificial Neural Network (ANN), Logistic Regression and AdaBoost machine learning algorithms were used to develop the prediction system.

Naive Bayes is a simple and fast statistical machine learning algorithm that uses a simplified version of the Bayes rule to calculate the probability of categorical attribute values given as input in a dataset. The previous probability states and conditional attributes in the categories are estimated from the frequency values calculated from the training data [7]. In the Decision Tree, which is commonly used in classification processes, the aim is to create a decision model that predicts the value of the class attribute based on the input attributes in the dataset. A decision tree model is similar to a tree. Here, each branch specifies the result value of the comparison operation on the previous node, and after each calculation of the attribute values, each sheet node indicates a class label. The paths from root to leaf represent classification rules [8]. The CN2 Rule Induction is a machine learning algorithm of classification and estimation developed for efficient and efficient extraction of simple, understandable rules in the form of "if - then - class" [9]. The Random Forest algorithm proposed by Ho is a learning method used in a number of tasks such as classification, regression and prediction. Random Forest was created by combining decision tree models. The structure of each tree in the model was developed from a boot sample taken from the training data [10]. The kNN algorithm is used for estimation problems such as classification and regression. The k in the sample area is a method that searches for the nearest educational samples and uses their averages in prediction procedures. The nearest neighbors to K is a method used to find the closest point k to a given data point at a given metric value [11]. SVM is a popular machine learning algorithm that performs classification tasks by creating hyperplanes in a multidimensional area that separates the status of different class labels. SVM is used for both regression and classification operations with large-scale continuous or categorical values [12]. ANN is a very popular method which has been developed by inspiring from the neural networks in the brain. The backpropagation algorithm for the calculation of the weights to be used in the ANN was used. Basically, ANN consists of a single input layer, at least one hidden layer, a single output layer, and weighted interconnection that connect the nodes in those layers. The number of nodes used in the input and output layers depends on the type of problem [13]. AdaBoost, the abbreviation of Adaptive Boost algorithm, is a machine learning algorithm created by combining several weak classifiers. This is a meta algorithm and can be used with many other learning algorithms to improve their performance. AdaBoost is sensitive to noisy data and outliers. Otherwise, the problem of over-learning is less sensitive than most learning algorithms [14].

2.2. Solar power plant

The use of renewable energy sources is increasing day by day and the infrastructure of the production system is shifting from the central network to many small capacity production centers called as distributed network. As of June 2, 2015, 3152.7 GWh energy was generated from the installation of the solar power plant in the KONYA Organized Industry. For example, on June 16, 2015, 5137 KWh was produced and on June 16, 2016, approximately 5943 KWh was produced in daily energy production. The cost of the installed solar power plant is approximately 1.1 million dollars and the income obtained in 16 months is 409859 dollars. According to our calculations, the cost of the plant can be met in approximately 4 years. The ground floor of the power plant is approximately 19000 quadrat meters and the floor costs about 800000 dollars.

A representative view of the solar power plant in Konya is shown in Figure 1 (a). 1 MW grid connected solar power plant has a surface area of 20000 m2. SPP location has 37 degrees north latitude and 32 degrees east longitude. Sunshine duration is 2640 hour/year and annual solar radiation is 1311 kWh/m2 year. Solar power plant has 4200 solar panels having 240 W power rating and 55 solar inverters having 20 kW power output. Solar panels with stainless steel structure have 20 degrees angle on the South-East direction. The solar power plant field has about 0.5 km distance from the center of the power distribution substation. The single line diagram of the solar power plant is shown in Figure 1 (b).

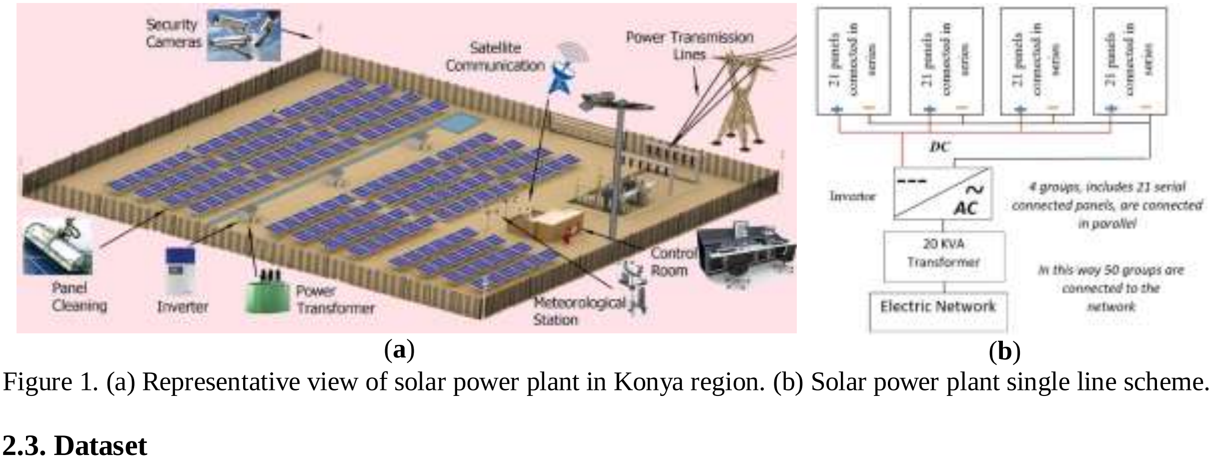

Security Cameras
Satellite Communication
Power Transmission Lines
Panel Cleaning
Inverter
Power Transformer
Control Room
Meteorological Station
(a)
21 panels connected in series
21 panels connected in series
21 panels connected in series
21 panels connected in series
DC
Invertor
AC
4 groups, includes 21 serial connected panels, are connected in parallel
20 KVA Transformer
In this way 50 groups are connected to the network
Electric Network
(b)

Figure 1. (a) Representative view of solar power plant in Konya region. (b) Solar power plant single line scheme.

2.3. Dataset

The Solar Power Plant dataset obtained from the 1 MW solar power plant in Konya was used for the development of the estimation system and the rule extraction by using machine learning algorithms. The SPP dataset with two class values consists of 776 samples and contains 7 attributes. Before starting the experimental study, missing and inconsistent data were extracted from the SPP dataset. The attribute names, and data type (continuous/categorical) of attributes are shown in Table 1.

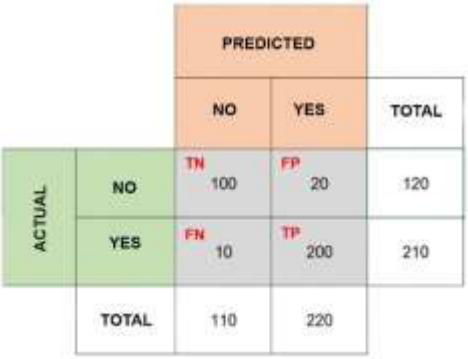
Table 1. SPP dataset.

Attributes	Data type
Datatime	Continuous
Energy Generation	Continuous
Tilted Irradiation	Continuous
Average Time to Sun	Continuous
Average Humidity	Continuous
Average Wind Speed	Continuous
Performance (Class)	Categorical

The continuous values of the performance class attribute of the SPP dataset were converted to 2-class categorical values with 0 and 1 format.

3. Results and Discussion

In this study, classification, estimation and rule extraction were performed by using machine learning algorithms using SPP data set. In order to determine the success of the results, ROC analysis method was applied to test procedures. The confusion matrix is frequently used to compare the estimates of the target (class) attribute and the actual values in order to assess the performance of the classification models used in machine learning. The sample confusion matrix of a data set with "yes", "no" class values were shown in Figure 2.



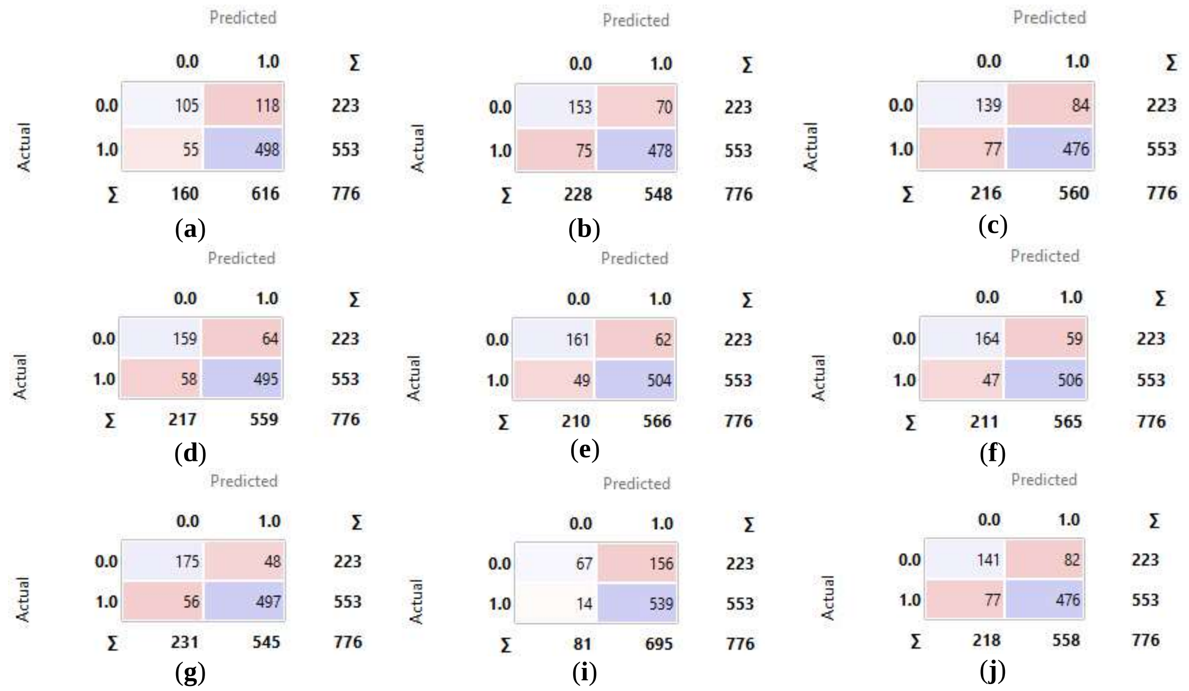
		PREDICTED		
		NO	YES	TOTAL
ACTUAL	NO	TN 100	FP 20	120
	YES	FN 10	TP 200	210
	TOTAL	110	220	

Figure 2. Confusion matrix structure.

Each row of the matrix represents the instances in an actual class while each column represents the instances in a predicted class. Where the classification estimates will have one of the following four evaluations [15].

- True Positive (TP): The algorithm output "yes", in the actual case "yes".
- False Positive (FP): The algorithm outputs "yes", but the actual status is "no".
- False Negative (FN): The algorithm outputs "no", but the actual status is "yes".
- True Negative (TN): The algorithm output "no", in the actual case "no".

These evaluations were used for calculate the sensitivity, specificity, selectivity and accuracy values. Confusion matrix diagrams which are used in ROC analysis of machine learning algorithms were calculated and were shown in Figure 3.



Actual	Predicted 0.0	Predicted 1.0	Σ
0.0	105	118	223
1.0	55	498	553
Σ	160	616	776

(a)

Actual	Predicted 0.0	Predicted 1.0	Σ
0.0	153	70	223
1.0	75	478	553
Σ	228	548	776

(b)

Actual	Predicted 0.0	Predicted 1.0	Σ
0.0	139	84	223
1.0	77	476	553
Σ	216	560	776

(c)

Actual	Predicted 0.0	Predicted 1.0	Σ
0.0	159	64	223
1.0	58	495	553
Σ	217	559	776

(d)

Actual	Predicted 0.0	Predicted 1.0	Σ
0.0	161	62	223
1.0	49	504	553
Σ	210	566	776

(e)

Actual	Predicted 0.0	Predicted 1.0	Σ
0.0	164	59	223
1.0	47	506	553
Σ	211	565	776

(f)

Actual	Predicted 0.0	Predicted 1.0	Σ
0.0	175	48	223
1.0	56	497	553
Σ	231	545	776

(g)

Actual	Predicted 0.0	Predicted 1.0	Σ
0.0	67	156	223
1.0	14	539	553
Σ	81	695	776

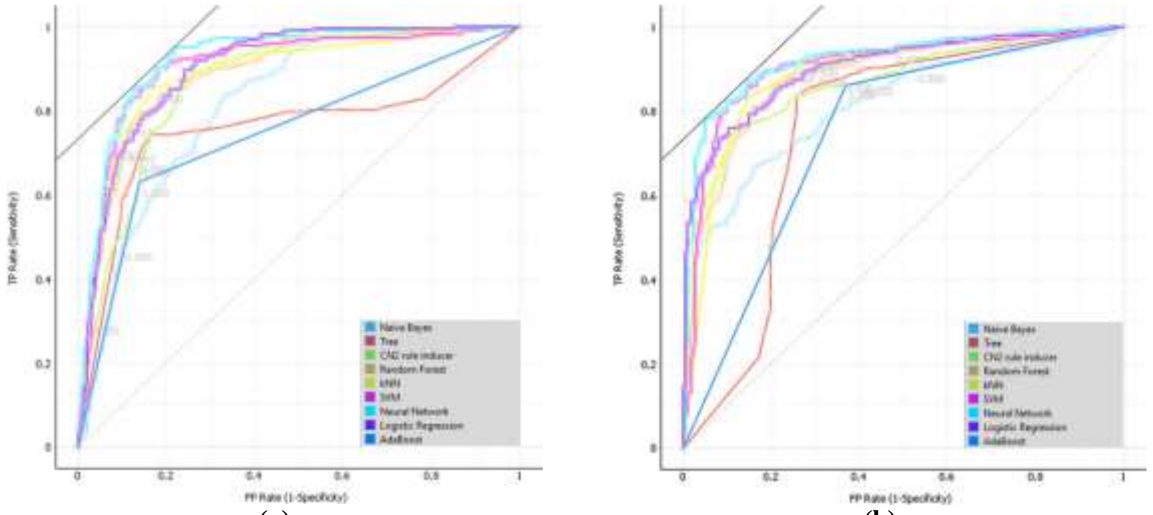
(i)

Actual	Predicted 0.0	Predicted 1.0	Σ
0.0	141	82	223
1.0	77	476	553
Σ	218	558	776

(j)

Figure 3. (a) Confusion matrix for Naive Bayes. (b) Confusion matrix for Decision Tree. (c) Confusion matrix for CN2 Rule Inducer. (d) Confusion matrix for Random Forest. (e) Confusion matrix for kNN. (f) Confusion matrix for SVM. (g) Confusion matrix for ANN. (i) Confusion matrix for Logistic Regression. (j) Confusion matrix for AdaBoost.

The performance evaluation of the used machine learning algorithms has been carried out by performing ROC (Receiver Operating Characteristics) analysis. The ROC curve results from the ratio of the sensitivity value to the precision value when the threshold value used to differentiate in binary classification systems differs. In simple terms, it can be expressed as the ratio of right positive to false positives. In this analysis, it is aimed to use a variable (continuous variable) that takes continuous values in a certain definition range as a diagnostic test. The ROC curves of the machine learning algorithms revealed according to the state of the class values were shown in Figure 4.



(a) (b)

Figure 4. (a) ROC curves of machine learning algorithms according to the state of class value 0. (b) ROC curves of machine learning algorithms according to the state of class value 1.

The ROC curve was drawn by placing the coordinate system on sensitivity to the ordinate and false positivity (1-selectivity) on the abscissa. The size of the area under the ROC curve determines the degree of accuracy of the classification. This size is between 0.50 and 1.00 and the closer to 1.00, the better the classification is made. A score of 1.00 indicates perfect classification. Sensitivity and specificity concepts were used to determine the accuracy of the estimates. Equations 1, 2, 3 and 4 were used to calculate the sensitivity, specificity, selectivity and accuracy values.

$$Sensitivity = \frac{TP}{TP + FN} \tag{1}$$

$$Specificity = \frac{TN}{TN + FP} \tag{2}$$

$$Selectivity = \frac{TP}{TP + FP} \tag{3}$$

$$Accuracy = \frac{Sensitivity + Specificity}{2} \tag{4}$$

After the ROC curves were drawn, the areas (Area Under Curve - AUC) under the curves were calculated. The higher the AUC value, the better is the predictive test.

Cross-validation test was used to determine the accuracy of the methods in the evaluation of machine learning algorithms. The criterion value was selected as 10 in the cross-validation of the data set. Here, the data set was divided into 10 randomly equal parts by the learning algorithm and each piece was used as a test and the remaining 9 parts were used for training. The mean value of the obtained results was calculated and the total classification result was obtained. Performance values of Machine Learning algorithms are shown in Table 2.

Table 2. Performance values for machine learning algorithms.

Methods	AUC	CA	F1	Precision	Recall
CN2 Rule Inducer	0.865	0.793	0.792	0.791	0.793
Naive Bayes	0.829	0.777	0.765	0.765	0.777
Random Forest	0.880	0.843	0.842	0.842	0.843
SVM	0.907	0.863	0.862	0.862	0.863
Decision Tree	0.752	0.813	0.814	0.814	0.813
kNN	0.889	0.857	0.856	0.855	0.857
Neural Network	**0.924**	**0.866**	**0.867**	**0.868**	**0.866**
Logistic Regression	0.906	0.781	0.742	0.790	0.781
AdaBoost	0.747	0.795	0.794	0.794	0.795

Where AUC is area under ROC curve, CA (Classification Accuracy) is accuracy classification score, F1 is a weighted harmonic mean of precision and recall, precision is the proportion of true positives among instances classified as positive and recall is the proportion of true positives among all positive instances in the data. Equation 5 was used to calculate F1.

$$F1 = 2\,x\,\frac{Selectivity\ x\ Sensitivity}{Selectivity +\ Sensitivity} \tag{5}$$

When the areas under the ROC curve (AUC) were examined in Table 1, the neural network machine learning algorithm had the best test with a value of 0.924. All other algorithms were found to have AUC values above 0.5. When the classification accuracy (CA) was examined, the neural network machine learning algorithm performed the best classification with a value of 0.866. The SVM and Logistic Regression algorithms provided a classification success which was very close to the neural network method with a value of 0.863. When the other classification qualifications were examined, the best classification test showed the neural network algorithm.

A total of 156 rules were obtained with the CN2 rule inducer algorithm. Some of the rules in the "if-then" structure obtained by the CN2 rule inducer algorithm from the SPP dataset were shown in Table 3. The rules obtained from the dataset were used to develop the estimation system.

Table 3. The example rules obtained by the CN2 rule inducer algorithm.

Rules
If Energy Generation >=0.732 and Datatime>=0.433 then Performance =1
If Tilted Irradiation>=0.466 and Energy Generation <=0.362 and Datatime>=0.366 then Performance =0
If Tilted Irradiation>=0.624 and Datatime>=0.833 and Energy Generation >=0.545 then Performance =0
If Tilted Irradiation<=0.555 and Energy Generation >=0.403 and Average Time to Sun>=0.286 then Performance =1
If Average Wind Speed>=0.239 and Energy Generation >=0.601 then Performance =0

In another method used in rule extraction, the decision tree is the algorithm. The structure of the decision tree with 5 depth levels used in this study was shown in Figure 5.

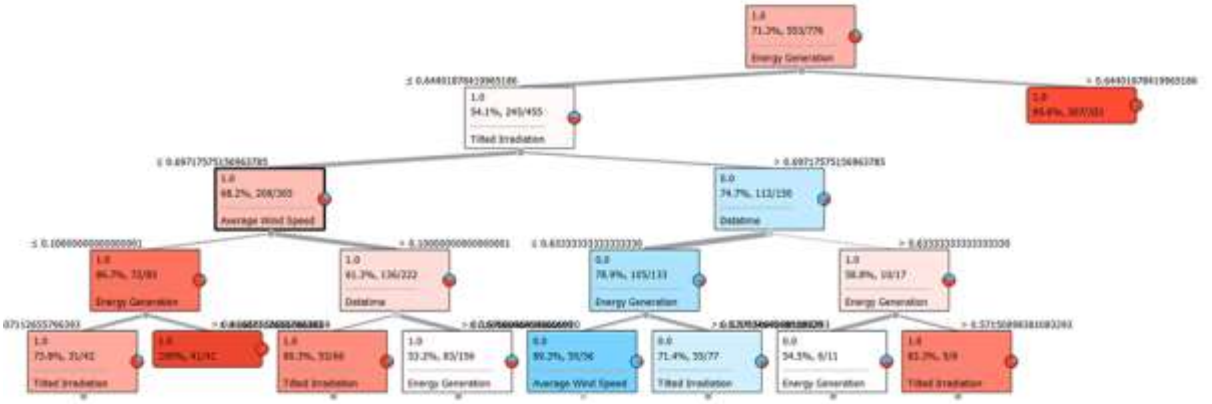


Figure 5. Structure of decision tree with 5 depth levels.

In the decision tree method, the rules were obtained by moving from the root to the leaves. In this method, the rules of the "if-then" structure were extracted from the SPP dataset. The rules obtained from the dataset were used to develop the estimation system.

Rapid and unpredictable changes can be observed due to instant weather events in electricity generation from renewable energy sources. During the increase in demand, fossil fuel-powered power plants are put into operation in order to keep the mains voltage and frequency at the required values and to meet the energy demand. Due to the nature of such resources, the differences between the estimated production values and real-time production amounts vary and are uncertain. As a result of these volatility and uncertainties, the need for network flexibility regarding network planning and management appears. In the event that large-scale renewable energy sources are disconnected or connected to the network, it is difficult to respond to changes in demand in minutes in peak time. In order to ensure that this rapid descent and output is provided by the system, the generation plants outside of the renewable energy sources must be available in such a way as to be able to capture these accelerations. In this study, we investigated how estimation and rule learning process can be improved by machine learning algorithms using solar power plant data. In the rule learning study, a more effective and efficient working environment was provided by using machine learning methods. Researchers working on solar power plants can achieve the necessary estimates and inferences easily and quickly with this application.

4. Conclusion

The impact of renewable energy sources, which are headed by solar and wind energy, on the electricity grid is increasing day by day. Network flexibility is an important parameter in real-time electrical network operations. In systems with low flexibility, integration of new wind and solar plants into the grid reduces system durability, leads to high costs and makes it difficult to incorporate sustainable energy into the system. Having more reserve power plants increase the cost of unit energy. Failure to have as many spare power plants as necessary may result in loss of energy balance, undesirable changes in network frequency, and often service disruptions. This study is the first step of obtaining the data required for predetermining network requirements by predicting the energy that can be produced depending on the weather conditions of any solar power plant. In this study, SPP datasets were evaluated with different machine learning methods and ROC analysis was performed to verify the accuracy of the tests. With the proposed methods, it is also aimed to create a substructure that allows estimation of auxiliary power plants and forces to be kept backup by estimating the production values of the existing other solar plants. This study investigated whether it is beneficial to use defined methods to perform performance estimation by using machine learning algorithms in some data obtained from solar power plant. Our study shows that it can benefit the researchers by providing an interactive, interesting and unique study experience. This study provides experimental support for estimation and optimization of solar power plant environments for existing literature. In the future, we plan to make more efficient and realistic studies by adapting the metaheuristic methods to some machine learning algorithms. We plan to increase the motivation of the researchers by preparing more realistic and interactive models in future studies. Through this system, companies producing energy can realize their commitments against companies that demand by estimating the value of electricity they will produce in a more planned and responsible way. Because production depends on natural conditions, prediction is very important. Investors may be able to anticipate the potential energy to be generated based on weather forecasts, so that they can be informed in advance about the procedures required to manage their financial resources. In this study, energy provider companies will be able to make weekly, monthly and annual energy planning by using the predicted data from their own renewable energy power plants depending on the weather conditions. Again, with the help of a software that they can develop, they will be able to create this work in a more profitable way. By making the predictions correctly, the network reliability will be stabilized, thus preventing unnecessary operation of the power plants, resulting in a significant increase in the rate of profitability. The aim of this study is to improve and develop with different methods in the future.

In this study, Orange software was used to realize machine learning algorithms. The algorithms were run on a computer with i7 processor and 16 GB memory configuration. When the success rates of the test results given in Table 2 are examined, it can be said that the most successful machine learning methods are Neural Network, SVM and Logistic Regression algorithms. To obtain the ROC analysis graph, the confusion matrix values specified in the figure were calculated. When the area under the ROC curve (AUC) is calculated in Figure 4, Neural Network, SVM and Logistic Regression algorithms are more successful. The higher the AUC value, the better is the prediction test. However, it is not correct to make decisions based on the results of only training and testing procedures on important topics such as forecasting. The results to be drawn from the ROC curves and the sensitivity value should be high. As a result, ROC analysis; where the diagnostic process will take a long time, high cost, special method-equipment and qualified manpower will be required; it is an analysis method that will make a significant contribution to the decision-making process by determining the appropriate cutting points for the easily available markers in a short time.

Author’s Contributions

All contribution belongs to myself in the article.

Statement of Conflicts of Interest

No potential conflict of interest was reported by the author.

Statement of Research and Publication Ethics

The author declares that this study complies with Research and Publication Ethics.

References

[1] Mitchell T.M. 1997. Machine Learning. McGraw-Hill Science.

[2] Langley P., Simon P. 1995. Applications of machine learning rule induction. Communications of the ACM, 38 (11): 11-46.

[3] Yumurtaci R. 2013. Role of energy management in hybrid renewable energy systems: case study-based analysis considering varying seasonal conditions. Turk J Elec Eng & Comp Sci, 21 (4): 1077-1091.

[4] Kulaksiz A.A., Akkaya R. 2012. Training data optimization for ANNs using genetic algorithms to enhance MPPT efficiency of a stand-alone PV system. Turk J Elec Eng & Comp Sci., 20: 1-14.

[5] Sharma R., Suhag S. 2017. Novel control strategy for hybrid renewable energy-based standalone system. Turk J Elec Eng & Comp Sci., 25 (3): 2261-2277.

[6] Raju L., Sakaya M., Mahadevan S. 2017. Implementation of energy management and demand side management of a solar microgrid using a hybrid platform. Turk J Elec Eng & Comp Sci., 25 (3): 2219-2231.

[7] John G., Cleary E., Leonard E. 1995. K*: An instance-based learner using an entropic distance measure. In: 12th International Conference on Machine Learning, pp. 108-114.

[8] Quinlan J.R. 1986. Induction of decision trees. Machine Learning, 1: 81-106.

[9] Clark P., Niblett T. 1989. The CN2 induction algorithm. Machine Learning, 3 (4): 261-283.

[10] Ho T.K. 1995. Random decision forests. Proceedings of the 3rd International Conference on Document Analysis and Recognition, Montreal, QC. pp. 278-282.

[11] Garcia S., Derrac J., Cano J., Herrera F. 2012. Prototype selection for nearest neighbor classification: Taxonomy and empirical study. IEEE Transactions on Pattern Analysis and Machine Intelligence, 34 (3): 417-435.

[12] Steinwart I., Christmann A. 2008. Support Vector Machines. Springer-Verlag, New York.

[13] Uzun Y., Arıkan H., Tezel G. 2016. Rule extraction from training artificial neural network using variable neighbourhood search for wisconsin breast cancer. Journal of Multidisciplinary Engineering Science and Technology (JMEST), 3 (8): 5452-5458.

[14] Freund Y., Schapire R.E. 1997. A decision-theoretic generalization of on-line learning and an application to boosting. Journal of Computer and System Sciences, 55: 119-139.

[15] Metz C.E. 1978. Basic principles of ROC analysis. Sem Nuc Med., 283-298.