Electric Power Systems Research 199 (2021) 107436

Contents lists available at ScienceDirect

Electric Power Systems Research

journal homepage: www.elsevier.com/locate/epsr

Energy management of hybrid energy system sources based on machine learning classification algorithms

Hmeda Musbah *, Hamed H. Aly, Timothy A. Little

Department of Electrical and Computer Engineering, Dalhousie University, Halifax, Canada

ARTICLE INFO

Keywords:
Scheduling and managing
Hybrid energy systems
Confusion matrix
Machine learning

ABSTRACT

Hybrid energy systems (HES) that contain renewable energy sources, such as wind and solar energy help to minimize CO_2 emissions. Therefore, studying these systems to improve their performance has become one of the critical needs these days due to the environmental crisis. Within HES, energy management (EM) of HES is an essential topic that has been covered in detail by numerous studies, as errors in EM can lead to HES blackouts. Recent research has experimented with energy management strategy (EMS) to achieve optimal EM. This work aims to generate a robust forecasting model for one hour ahead of EM. The present research work has two main objectives. The first objective is to determine which energy source should supply the demand side, using different machine-learning algorithms such as Random Forest (RF), Decision Tree (DT), Gaussian Naive Bayes (Gaussian NB) and K-Nearest Neighbors (KNN). The second objective is to compare the results of these algorithms to choose the algorithm with the best performance and to rank them based on performance as well as accuracy. The work is validated using different algorithms. The results show that DT algorithm has achieved the best performance compared to the RF and Gaussian NB algorithms. KNN algorithm gives the lowest accuracy especially over class 3. The results proof that RF, DT, and Gaussian NB algorithms are reliable.

1. Introduction

Forecasting plays a vital role in the electrical engineering industry, as it gives designers a clear idea about energy system configuration over short-term periods. Such forecasts help in planning and managing the generated renewable energy as an alternative, to reduce the overall cost and CO_2 emissions from conventional energy resources. If the forecasting is not accurate, the energy system can break down. Hybrid energy systems (HES) consist of renewable energy sources such as wind and solar that are fluctuating, intermittent and nonlinear and need accurate forecasting models. The performance of HES depends on several factors that should be taken into account such as the demand for power, power production, weather conditions and power management. Proper demand forecast allows designers to determine power production capacity. Also, knowing the weather conditions enable designers to predict optimum solar and wind energy availability. These factors can either enhance or reduce the performance of the HES. To date, several studies have been conducted, aimed at forecasting and energy management, and several methods have been applied to achieve better results [1,2]. However, energy management of hybrid systems has attracted many researchers, as it plays a vital role in transmitting the energy through the hybrid energy system. The energy management strategy and optimization usually work side by side to guarantee load electrification and to minimize the cost of energy production. A successful energy management strategy gives the hybrid energy system stability and protects its components from damage due to overloading [3].

As HES include solar or wind energy, the energy management strategy becomes mandatory because solar and wind are intermittent and insufficient. Fig. 1 represents the schematics of the methodology of energy management in HES that contains solar and wind. In the first step, a historical dataset of the demand side and weather conditions are used to forecast future demand side and renewable energy source production, respectively. The second step consists of obtaining the energy management through genetic algorithm (GA), differential evolution (DE), neural network, fuzzy logic, and neuro-fuzzy techniques. In the proposed methodology, information obtained from the energy management technique is used to generate dataset. Consequently, a historical dataset of energy management obtained from the energy management techniques is used to determine the sources of the hybrid energy requirement of the load. To the best of our knowledge, no

* Corresponding author.
E-mail addresses: Hm392855@dal.ca (H. Musbah), hamed.aly@dal.ca (H.H. Aly), Timothy.Little@Dal.Ca (T.A. Little).

https://doi.org/10.1016/j.epsr.2021.107436
Received 21 February 2021; Received in revised form 30 May 2021; Accepted 11 June 2021
Available online 24 June 2021

previous study has investigated this topic. The present study has two main objectives: 1. to forecast the scheduling of the energy sources using machine learning algorithms such, as Random Forest (RF), Decision Tree (DT), Gaussian Naive Bayes (Gaussian NB), and K-Nearest Neighbors (KNN); and 2. to compare the results of the above-named algorithms. The novelty of the work is:

1 To develop different intelligent approaches for the preprocessed data to predict different energy resources.
2 To propose a new methodology to optimize the management between different energy resources to minimize the fossil fuel emission and the overall cost as well as to increase the penetration of the renewable energy resources.
3 To modify the model's parameters based on the optimal size of different forecasting approaches to achieve the optimal accuracy for the models which will lead to the system convergence and increase the accuracy of the network and reduce the training duration.
4 To propose a model for scheduling prediction which is the main factor for energy management.
5 To validate the proposed work by using different intelligent approaches.

2. Literature review

The demand side problem is one of the factors that should be taken into consideration during the building of the HES. This problem can be solved by applying different kinds of methods and techniques, such as regression analysis, time series analysis, artificial neural networks, genetic algorithms, support vector machine, fuzzy logic, and adaptive network-based fuzzy inference. Recently, hybrid methods and intelligent approaches have been attractive to researchers in solving the demand side problem [4].

In [5], hourly short-term electric load data were predicted by applying the Seasonal Autoregressive Integrated Moving Average (SARIMA) model. The researchers used the Fast Fourier Transformation algorithm (FFT) to detect the existence of seasonality in the time series of electrical load data. In [6], a fuzzy logic and adaptive neuro-fuzzy inference system (ANFIS) were used to forecast hourly short-term load in Turkey, with Artificial Intelligence (AI) being described as a powerful technique for determining the demand side problem. In addition, the researchers stated that using Artificial Neural Networks (ANNs) with Particle Swarm Optimization (PSO), Back Propagation Algorithm (BPA) or Fuzzy Logic (FL) as hybrid methods would increase the accuracy of solving the demand side problem [7]. Several different ANN architecture performances in forecasting the demand side were evaluated in [8]. The authors affirmed that intelligent forecasting methods are superior to conventional methods with regard to accuracy. Multivariate adaptive regression splines (MARS), artificial neural network (ANN) and linear regression (LR) methods were used to determine short, mid- and long-term load forecasting [9]. Several factors that affect HES indirectly have been summarized in detail in [10].

Knowing in advance the amount of energy produced from renewable and traditional energy sources in a HES is recognized as a fundamental process in HES design. Forecasting of the power production has received considerable attention in recent years, with many researchers applying a variety of approaches and techniques to achieve high forecasting accuracy. In [11], a multi-layer perceptron (MLP) model was employed to forecast wind power production 24 h in advance. Meanwhile, the authors in [12] used Recurrent Neural Network (RNN) to forecast the solar power production from a photovoltaic power plant, and the researchers in [13] employed three different methods to forecast photovoltaic power production, namely: Auto Regressive Integrated Moving Average (ARIMA), Radial Basis Function Neural Network (RBFNN), and Least Squares Support Vector Machine (LS-SVM). In [14], the solar power production was forecast using hybrid Wavelet-PSO-SVM forecasting model based on SCADA.

Weather conditions, wind speed, solar irradiation and temperature all have a direct effect on the amount of power produced in a HES [15]. Wind speed was predicted in [16] using a novel hybrid forecasting system containing three modules (a data preprocessing module, an optimization module, and a forecasting module). The authors in [17] applied an autoregressive moving average with echo state network compensation to improve the accuracy of short-term wind speed forecasting. A hybrid deep learning model that contains a gated recurrent unit (GRU) neural network and an attention mechanism was used to forecast the solar irradiance changes in four different seasons in [18], while the authors in [19] demonstrated that forecasting solar irradiance is vital to renewable energy generation.

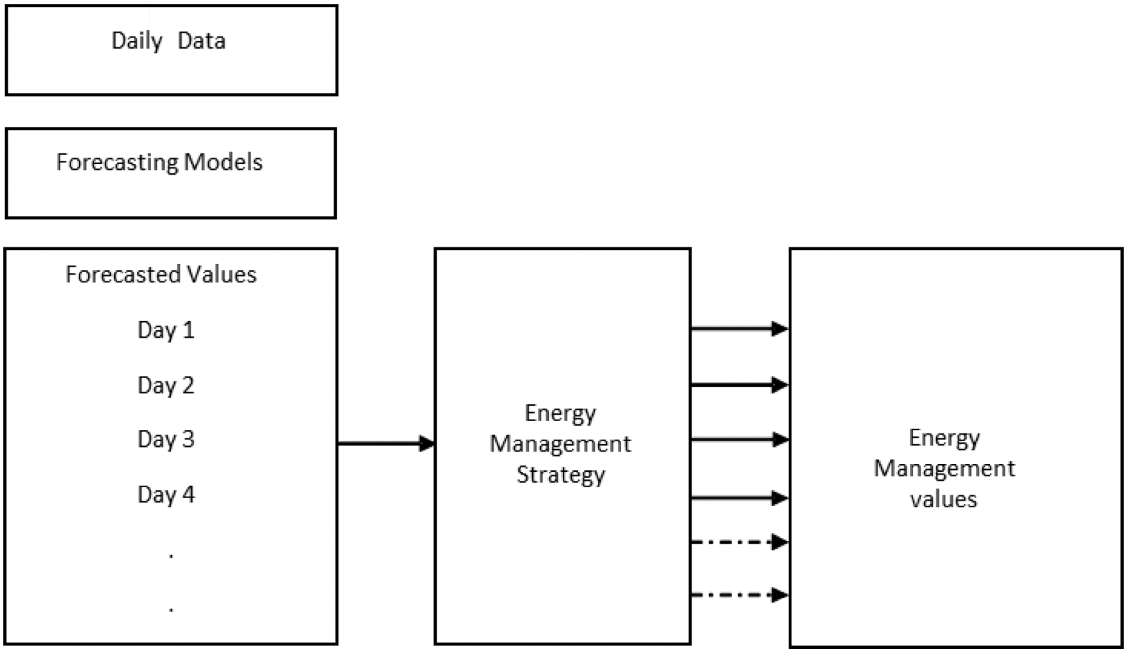


Fig. 1. Common methodology.

An intelligent hybrid clustered for wind speed forecasting model was proposed based on different combinations of ANN, WNN and least square methods. The model is based on two different steps of forecasting using back-to-back results to increase the number of inputs of the second stage to improve the system accuracy. The model uses preprocessed data based on clustering techniques and at the end, the forecasted data are aggregated [1]. A hybrid model consisting of Neuro Wavelet (WNN), Time Series and Recurrent Kalman filter for wind speed forecasting was proposed based on two different stages. The model depended on using the error from the first stage as an input for the second stage to improve the system accuracy and reduce the training time [3]. Hybrid Models of ANN, WNN, and Kalman filter based on clustering techniques for smart grid integration were proposed for short-term load forecasting. The model convergence very fast but it is more complicated and if there is any error in any stage, the error will be accumulated [2]. An adaptive method based on multi-model partitioning algorithm (MMPA) was developed to forecast a short-term electricity load using a historical data. Different real cases derived from measurement loads taken at Hellenic Public Power Cooperative Company have been studied. The obtained results showed that the proposed method is able to determine the component of electricity load time series [20]. Several ANN models have been built based on different combination, learning algorithms and transfer functions. Real data were divided into three stages: training, validation and testing stages. The model's outputs were compared to each other to identify the most reliable model. The selected model was used to forecast energy consumption years ahead [21]. In order to obtain short-term load forecasting, a technique based on ANN methods and wavelet denoising algorithm were applied to real data collected from the Bulgarian power system grid. The obtained results show that the proposed method is successful in reducing the standard deviation between actual and forecasted data [22]. HOMER software was used to analyze the technical and economic viability of hybrid energy systems in the Masirah Island power system in Oman. They evaluated different scenarios using package DIgSILENT. The authors stated that the hybrid energy system containing diesel, photovoltaic and wind turbine is a good choice as it reduces the operation cost [23].

A comprehensive study has been done to predict the hourly energy from a solar thermal collector system. The authors used random forest (RF), extra trees (ET), decision trees, and support vector regression (SVR). These models were evaluated based on ability (stability), accuracy and computational cost. The obtained results showed that RF and ET performances are equal, and they are more accurate than DT [24].

The daily total energy generation of an installed photovoltaic system was predicted using the Naïve Bayes classifier. The classifier applied to a one-year historical dataset such as daily average temperature, daily total sunshine duration, daily total global solar radiation and daily total photovoltaic energy generation parameters. The results proved that the Naïve Bayes classifier is effective in predicting the total energy generation, where its accuracy is 82.1917% [25].

The authors in [26] claimed that many machine learning algorithms such, as linear regression (LR), K nearest neighbor regression (KNN), support-vector machine regression (SVMR), and decision-tree regression (DTR) are used in renewable-energy predictions. They stated that the most used machine-learning algorithms were for solar energy and wind-energy predictions. The following section sums up the different algorithms encountered in machine learning.

3. Machine learning

Machine learning is an application of artificial intelligence (AI). It is widely used in every sphere of human life because its ability in solving real life problem. Fig. 2 shows the two main steps in achieving machine-learning (ML) algorithms. The two stages of the dataset are divided into two unequal groups—training and testing datasets—designating training and testing stages. In the training stage, the dataset is used as input to the selected algorithm to train it. The trained selected algorithm is then fed by testing the dataset to evaluate the selected algorithm performance in the testing stage. Machine learning problems can be divided into three types which are supervised, unsupervised and reinforcement problem. There's no one specific algorithm can solve machine learning problems because of the simplicity and complexity of their classification, which sometimes requires a unique algorithm [27]. The main reason that makes us select the Random Forest (RF), Gaussian Naive Bayes (Gaussian NB), Decision Tree (DT) algorithm and K-Nearest Neighbor (KNN) is problem type where these algorithms are able to solve the classification problem. The second reason is the number of data points and features where the algorithms can be used to handle different sized dataset. Also, these algorithms do not require normalization of data. Moreover, the algorithms are simple and easy to be implemented.

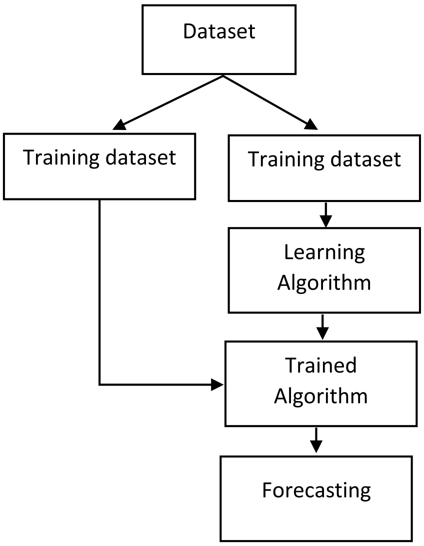


Fig. 2. Two main steps in achieving ML algorithms.

- Random Forest (RF)

Breiman [28] introduced RF in 2001. Random Forest is a supervised machine-learning algorithm that has been widely used due to its robust performance [29]. However, because the imbalances inherent in most practical classification problems, other common algorithms cannot accurately deal with the problems. RF has the ability to

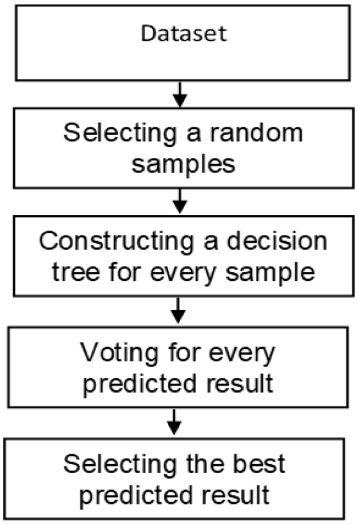

Dataset
Selecting a random samples
Constructing a decision tree for every sample
Voting for every predicted result
Selecting the best predicted result

Fig. 3. Main steps involved in the Random Forest (RF) algorithm.

overcome these problems by applying a cost-sensitive learning and sampling technique [30]. Fig. 3 presents the main steps of RF.

- Gaussian Naive Bayes (Gaussian NB)

Gaussian Naive Bayes is a simple probabilistic algorithm. It is one of the most well-known of the Naive Bayes (NB) algorithms that uses the Bayes' theorem. The approach is designed to handle continuous attributes, which is associated with each class that is distributed according to Gaussian distribution [31]. The major advantages of the NB family are that it can be applied to practical classification problems, it requires less training data, and can be trained very effectively in supervised learning. A major drawback of the NB family is that the attributes are assumed to be independent, which is almost impossible [32].

- Decision Tree (DT) algorithm

Decision Tree algorithms were first introduced by Ho in 1995. Later, a multiple DT algorithm was used to form the Random Forest algorithm [33]. The DT approach can be employed to solve classification or regression problems. Unlike many other algorithms, DT has the ability to handle a wide range of attributes and does not require scale normalization before model building and application. Furthermore, regarding data preparation, the DT algorithm is not affected by missing data [34].

- K-Nearest Neighbor (KNN)

K Nearest Neighbor (KNN) is one of the most frequently used algorithms in the machine-learning field due to its ease of use and versatility [35]. However, because KNN uses all the training data, it requires time to read it as well as memory to store it. The authors in [36] provide a good summary of the advantages and disadvantages of KNN techniques.

The letter "K" indicates the number of nearest neighbors, while the term "nearest neighbor" indicates that the algorithm is searching for the closest point needed for classification and labeling of the nearest point assigned to it. The nearest neighbor distance between two points can thus be calculated using a Euclidean distance function, as shown in Eq. (1):

$$D = \sqrt{\sum_{i=1}^{n} \left(b_i - a_j\right)^2} \quad (1)$$

where D is the distance between the two points a and b.

Algorithms such as KNN and support vectors that use distance measures between input variables could face some issues, one of which is the differences in input variable scales. This issue could lead to difficulty during model creating and their performance could be poor during learning. Therefore, standardizing or normalizing data on the same scale is highly recommended [37]. Eqs. (2) and (3) can be used to standardize or normalize the data, respectively.

$$\text{normalization value} = \frac{(\boldsymbol{x} - \boldsymbol{\mu})}{\boldsymbol{\sigma}} \quad (2)$$

where μ is the mean of the feature values and σ is the standard deviation of the feature values.

$$\text{standardization value} = \frac{= (\boldsymbol{x} - x_{min})}{(\boldsymbol{x}_{max} - \boldsymbol{x}_{min}} \quad (3)$$

where Xmax and Xmin are the maximum and the minimum values of the feature, respectively.

4. Evaluating metrics

The essential step in building a machine-learning model is evaluating its performance. Various metrics can be used for this purpose as well as for comparative purposes. Accuracy, which is a common evaluation metric for classification problems, is calculated as shown in Eq. (4):

$$\text{Accuracy} \ \frac{TP + TN}{T_O} \quad (4)$$

Where: TP is true positive, TN is true negative, and T_O is total number of predictions. The TP is when the predicted value is yes and the actual output is also yes, while the TN is when the predicted value is no and the actual output is no as well.

The overall classification accuracy is often not an appropriate metric for evaluating the model performance in the case of a dataset with imbalanced data [38]. In addition, sometimes the algorithm understands only one or two classes, which means that the algorithm could be biased towards one class over the others. A few of the more powerful metrics that can provide a clear idea about model performance when dealing with imbalanced data are given below:

- Precision: The precision metric can be calculated as shown in Eq. (5) by the number of true positives (TPs) divided by the number of TPs and False Positives (FPs).

$$\text{Precision} = \frac{TP}{TP + FP} \quad (5)$$

- Recall: Recall is another important metric, which is defined as the number of TPs divided by the number of TPs and the number of FNs, as expressed in Eq. (6):

$$\text{Recall} = \frac{TP}{TP + FN} \quad (6)$$

- F1 Score: The F1 Score metric shows the robustness and precision of the model and seeks to find the balance between precision and recall. Mathematically, it can be expressed as:

$$\text{F1} - \text{score} = \frac{2}{\frac{1}{Precision} + \frac{1}{Recall}} \quad (7)$$

Many assessment methods were used in the literature. Mean absolute percentage error (MAPE) and the normalized root mean square error (nRMSE) are used in this work to asset the model's accuracy as shown in Eqs. (1) and (2) [39,40].

$$MAPE = \frac{1}{N}\sum_{1}^{N}\left|\frac{x_i - y_i}{y_i}\right| x\ 100 \quad (8)$$

where, x_i is the predicted value, y_i is the actual value, and N is the number of observations.

$$nRMSE = \frac{\sqrt{\frac{1}{N}\sum_{1}^{N}(x_i - y_i)^2}}{\bar{y}} \quad (9)$$

where $\bar{y}$ is mean of the actual data.

The common methodology of energy management in HES that contains solar, and wind is introduced. In the first step, a historical dataset of the demand side and weather conditions are used to forecast future demand side and renewable energy source production, respectively. The second step is to obtain the energy management through genetic algorithm (GA), differential evolution (DE), neural network, fuzzy logic, and neuro-fuzzy techniques. In the proposed methodology, the energy management that is obtained from the energy management techniques is collected form the dataset. In the proposed methodology, a historical dataset of energy management obtained from the energy management techniques is used to determine the sources of the hybrid energy to be connected to the load.

5. Data analysis and hybrid energy system description

The scheduling dataset includes 336 instances, 5 attributes, 4 inputs, and 1 output variable. The inputs are hourly demand side, temperature and availability of solar and wind, while the output is the scheduling of the hybrid energy sources. The inputs are numeric and have values across various ranges. The last attribute is the output variable, which is called "class".

The class type is nominal and has 6 values. These values are encoded as presented in Table 1. Fig. 4a–d shows the graphical distribution of each attribute, every color represents the class and number of attributes. As can be seen, there is a different overlap distribution for the class values on each of the attributes. The temperature and demand side attributes have a Gaussian-like distribution and a nearly Gaussian distribution with a skew, respectively. If there were significantly more data, Gaussian would be clear in both attributes. The wind and sun attribute values are 0 and 1, where 0 means there is no wind or solar and 1 means there is wind or solar. The classes are imbalanced, which indicates there is an unequal number of instances in each class.

6. The focus of the present study

A HES consisting of two renewable energy sources (solar and wind energy) and two traditional energy sources (gasoline and diesel generators) is studied in this work. The power output of the solar and wind turbine is 20 kW and 25 kW, respectively, while the output for the gasoline and diesel generators is 50 kW and 55 kW, respectively.

The system is used to supply a remote community. The maximum demand side in the remote community is 100 kW and the minimum demand side is 23.6 kW. The sources of the HES have been scheduling more than 336 h over two weeks.

The machine-learning algorithms are employed to predict the scheduling of the HES sources. Specifically, the scheduling dataset is divided into two groups: 70% of the dataset for training the machine-learning algorithms, and the rest of dataset for testing the algorithms.

7. Results and discussion

Predicting the source that should meet the demand is one of the important factors that should be taken into consideration when designing a hybrid energy system. As we stated in the literature review, those studies paid attention to power consumption, power generation and weather prediction, but not scheduling prediction. The results of schedule prediction for the hybrid energy system were successfully obtained in the present study. Table (2) shows the overall accuracy of the algorithms applied to the scheduling dataset. Clearly, RF, Gaussian NB and DT algorithms resulted in reliable percentages. The accuracy can be used for evaluating binary and multiclass classifiers. However, due to the fact that the data is imbalanced, the overall accuracy cannot be a reliable metric to evaluate the algorithms, because the overall accuracy treats all the classes equal and does not give an attention to minority classes, which is called accuracy paradox. Therefore, this problem has been solved by using precision, recall and F1-score metrics. These metrics show how the algorithms deals with individual classes. Tables 3–5 display the precision, recall and F1-score metrics for evaluating the RF, Gaussian NB, KNN, KNN Standard Scaler and DT algorithms. These tables were converted to figures for concept clarification.

Fig. 5 presents the use of precision metric for evaluating the algorithms over the classes. As can be seen, the DT algorithm shows overall excellent performance, followed by the RF and Gaussian NB algorithms. The KNN algorithm shows the worse performance, especially in class 3. Fig. 6 shows the use of recall metric for assessing the algorithms over classes 1, 2, 3, 4, 5 and 6. As in Fig. 6, it is clear that the DT algorithm has the highest performance of the classes, while the RF and Gaussian NB algorithms and the KNN show the lowest.

As shown in Fig. 7, utilizing the F1-score metric gives nearly the same results as the precision and recall metrics regarding performance. Even though class 3 recurs only 13 times in the dataset, the RF, Gaussian NB and DT present a robust performance in every performance stages. The obtained results from precision, recall and F1-score metrics show the real algorithms performance and how the overall accuracy gives misleading results over imbalanced data. In general, it is noticeable that the RF, Gaussian NB and KNN algorithms are biased to specific classes, whereas the DT algorithm understands all the classes.

Standardizing the dataset is an amendment to the performance of the KNN algorithm. Figs. 8–10 manifest the performance of the KNN algorithm after standardizing the dataset. It is noticeable that there is a big change in KNN performance where the KNN algorithm was capable of understanding all the classes. Also, it is noticeable that the algorithm could increase the performance over class 1 using oversampling or undersampling technique or increasing the number of features. Undoubtedly, as a result of the amendment, the KNN algorithm competed with the RF and Gaussian NB in terms of some of the classes.

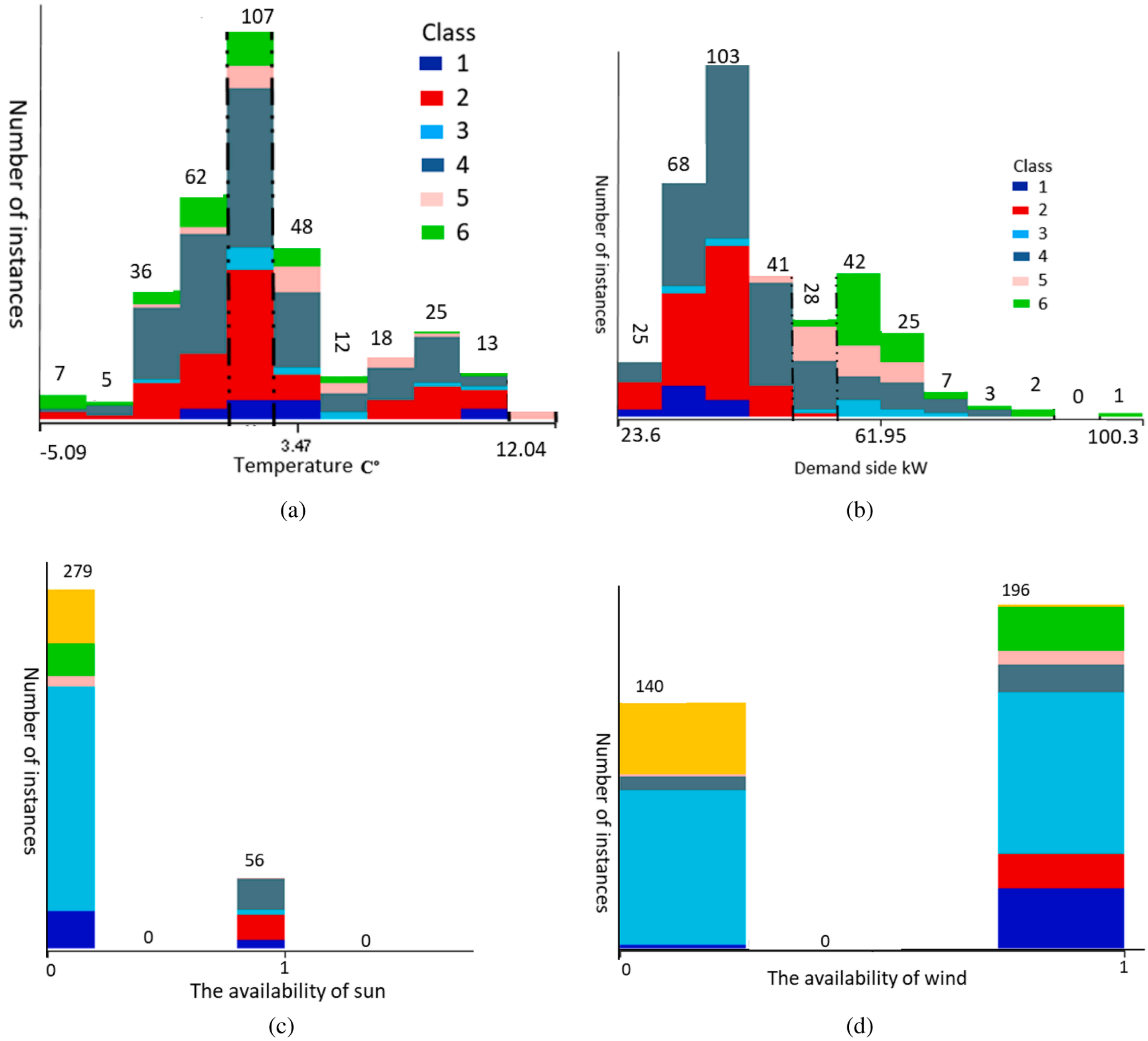
For summarizing the performance of the algorithms, the confusion matrix was generated as shown in Fig. 11. The columns represent the actual results while the rows represent the predicted results; the correct predictions are highlighted in red color. Clearly, 7 values were correctly classified as class 3. Reading down the class 6 column, one value that should be class 6 were classified as 3. Also, 23 values were correctly classified as class 4. Reading down the class 12 column, 3 values that should be class 12 were classified as 4. The Rf algorithm correctly classified the other values, namely classes 6, 8 and 10.

8. Conclusion

This work has proposed a novel technique to forecast the next hour for energy management of the hybrid energy systems. Machine-learning algorithms, such as Random Forest (RF), Decision Tree (DT), Gaussian Naive Bayes (Gaussian NB) and K-Nearest Neighbors (KNN) were applied to Energy Management datasets in this study to forecast which sources should connect to supply the demand side. The work was validated by making comparison with four different techniques to choose the best one. The results from the overall accuracy metric indicate that while the algorithms are reliable in forecasting the Energy Management, this is somewhat misleading, as the algorithms demonstrate biases to specific classes. Thus, the classification report has been used instead of the overall accuracy metric. In utilizing classification, it is found that the DT algorithm achieved excellent performance compared to the RF and Gaussian NB algorithms, while the KNN algorithm presented a weak performance, especially over class 3. Furthermore, the RF, DT, and Gaussian NB algorithms were found to be reliable. Finally, after standardizing the Energy Management dataset, the KNN algorithm is able to compete with the RF and Gaussian NB algorithms in some of the classes. This work proposes different models, and the model parameters are

Table 1
Encoded values for different classes.

Class	Class Encoding
Solar and Wind	1
Gasoline Generator	2
Solar and Gasoline Generator	3
Wind and Gasoline Generator	4
Solar, Wind and Gasoline Generator	5
Gasoline Generator and Diesel Generator	6



(a)

(b)

(c)

(d)

Fig. 4. (a) the graphical distribution of temperature. (b) the graphical distribution of demand side. (c) the graphical distribution of the availability of sun. (d) the graphical distribution of the availability of wind.

Table 2
Overall accuracy of algorithms.

Algorithm	Accuracy of training data	Accuracy of testing data
RF	99.5	95.05
Gaussian NB	95	95
DT	100	95
KNN	32.34	38.61
KNN Standard Scaler	97	95

Tables 3
Precision metrics of algorithms.

Class	RF	Gaussian NB	DT	KNN	KNN- standardscaler
1	100%	100%	100%	10%	100%
2	100%	84%	100%	57%	93%
3	80%	100%	100%	0%	67%
4	98%	98%	98%	83%	100%
5	100%	73%	100%	29%	100%
6	67%	100%	100%	44%	83%

Tables 4
Call metric of algorithms.

Class	RF	Gaussian NB	DT	KNN	KNN- standardscaler
1	88%	62%	100%	75%	75%
2	88%	100%	100%	46%	100%
3	100%	100%	100%	0%	100%
4	100%	100%	100%	33%	96%
5	100%	100%	100%	40%	100%
6	86%	50%	92%	33%	83%

Tables 5
F1-score metric of algorithms.

Class	RF	Gaussian NB	DT	KNN	KNN- standardscaler
1	93%	77%	100%	18%	86%
2	94%	92%	100%	51%	97%
3	89%	100%	100%	0%	80%
4	99%	99%	99%	48%	98%
5	100%	84%	100%	33%	100%
6	75%	67%	96%	38%	83%

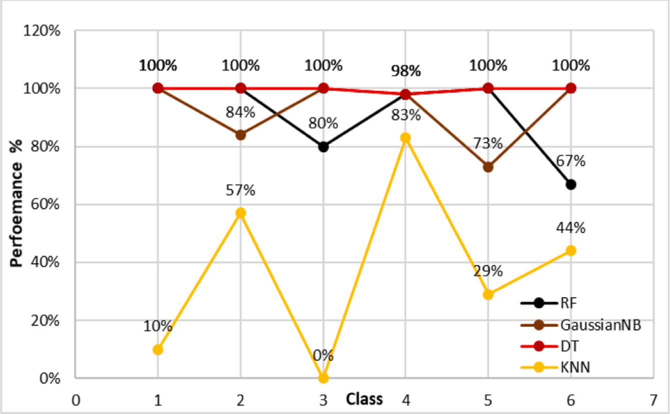


Fig. 5. Precision metric.

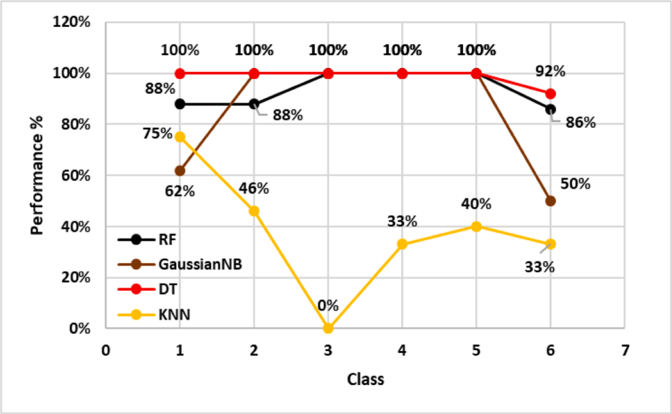


Fig. 6. Recall metric.

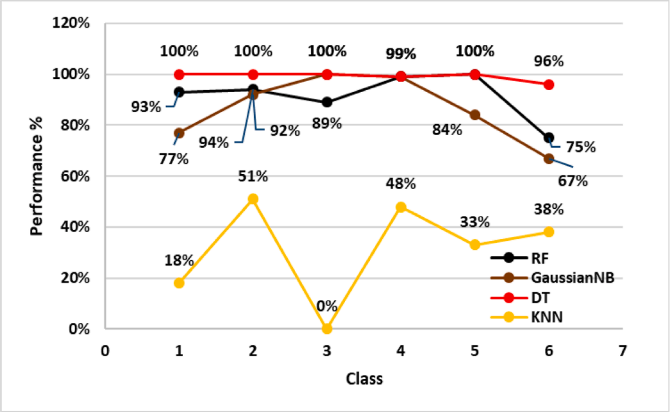


Fig. 7. F1-score metric.

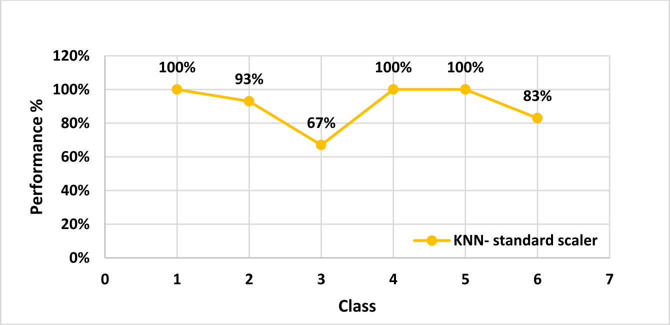


Fig. 8. Precision metric- KNN algorithm performance post-dataset standardizing.

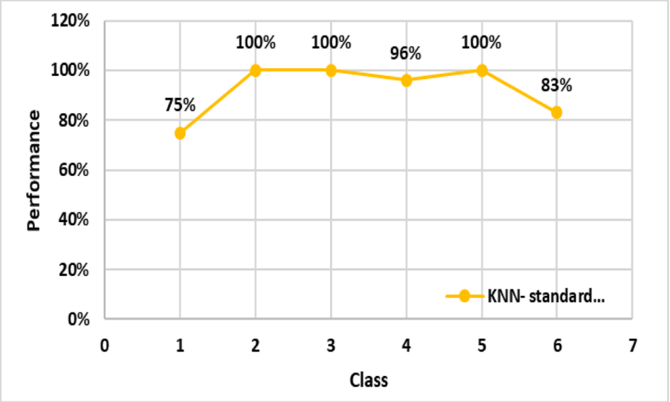


Fig. 9. Recall metric -KNN algorithm performance post-dataset standardizing.

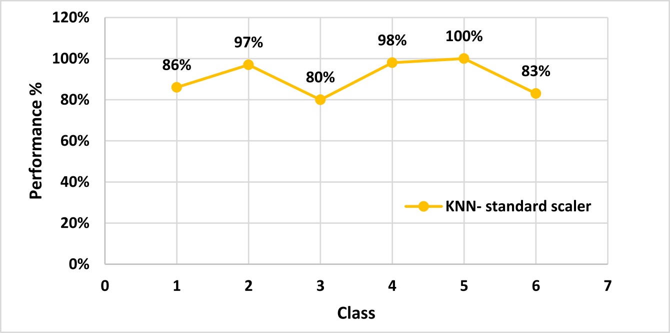


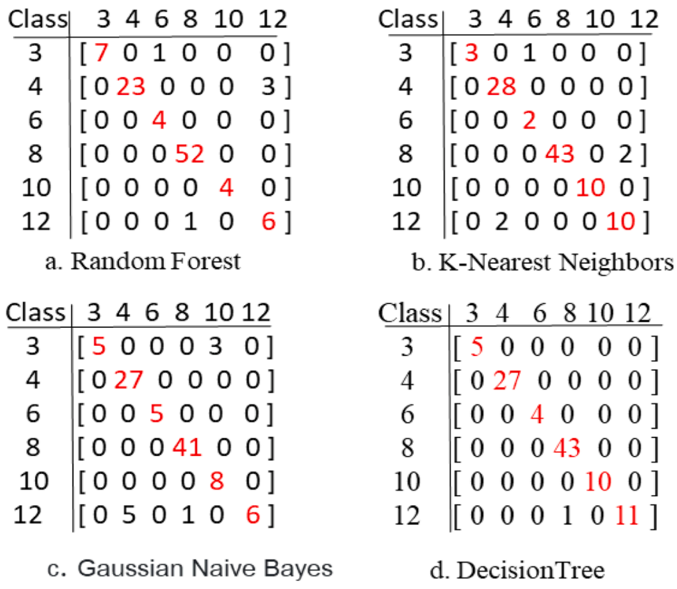
Fig. 10. F1-score metric- KNN algorithm performance post-dataset standardizing.

optimized based on the optimal size of different forecasting approaches to achieve the optimal accuracy for the models which will lead to the system convergence and increase the accuracy of the network as well as reduce training duration. The results show the effectiveness of the proposed models for scheduling prediction. The proposed work for the EM enables the use of renewable energy in highly efficient and accurate ways. Based on the forecasting system, the designer can take action in advance to turn on or shut down some conventional energy resources to reduce CO_2 emission without overloading. The proposed forecasting algorithms are very simple and easy to be used compared to more complicated hybrid algorithms and this helps in reducing the error compared to hybrid models that need more stages and this increases the overall error for the overall systems for these hybrid models. In this work only one algorithm is used, and this reduces the training time and the error at the same time. All proposed models are highly accurate but the best one is DT algorithm.

Author statement

1. Hmeda Musbah (PhD student)
2. Hamed H. Aly (co- supervisor)
3. Timothy A. Little (supervisor)

This work is part of my thesis PhD.



Class	3	4	6	8	10	12
3	7	0	1	0	0	0
4	0	23	0	0	0	3
6	0	0	4	0	0	0
8	0	0	0	52	0	0
10	0	0	0	0	4	0
12	0	0	0	1	0	6

a. Random Forest

Class	3	4	6	8	10	12
3	3	0	1	0	0	0
4	0	28	0	0	0	0
6	0	0	2	0	0	0
8	0	0	0	43	0	2
10	0	0	0	0	10	0
12	0	2	0	0	0	10

b. K-Nearest Neighbors

Class	3	4	6	8	10	12
3	5	0	0	0	3	0
4	0	27	0	0	0	0
6	0	0	5	0	0	0
8	0	0	0	41	0	0
10	0	0	0	0	8	0
12	0	5	0	1	0	6

c. Gaussian Naive Bayes

Class	3	4	6	8	10	12
3	5	0	0	0	0	0
4	0	27	0	0	0	0
6	0	0	4	0	0	0
8	0	0	0	43	0	0
10	0	0	0	0	10	0
12	0	0	0	1	0	11

d. DecisionTree

Fig. 11. Confusion matrix of the algorithms.

Declaration of Competing Interest

None.

References

[1] H.H. Aly, A novel deep learning intelligent clustered hybrid models for wind speed and power forecasting, Energy 213 (2020), 118773.

[2] H.H. Aly, A proposed intelligent short-term load forecasting hybrid models of ANN, WNN, and KF based on clustering techniques for smart grid, Elsevier Int. J. Electr. Energy Res. 182 (2020), 106191, https://doi.org/10.1016/j.epsr.2019.106191.

[3] H.H. Aly, An intelligent hybrid model of neuro wavelet, time series and recurrent Kalman filter for wind speed forecasting, Elsevier J. Sustain. Energy Technol. Assess. 41 (2020), 100802, https://doi.org/10.1016/j.seta.2020.100802.

[4] G. Aburiyana, M. El-Hawary, An overview of forecasting techniques for load, wind and solar powers, IEEE Electr. Power Energy Conf. (2017) 1–7.

[5] H. Musbah, M. El-Hawary, SARIMA model forecasting of short-term electrical load data augmented by fast Fourier transform seasonality detection, in: Proceedings of the IEEE Canadian Conference of Electrical and Computer Engineering (CCECE), 2019, pp. 1–4.

[6] H.H. Çevik, M. Çunkaş, Short-term load forecasting using fuzzy logic and ANFIS, Neural Comput. Appl. 26 (6) (2015) 1355–1367.

[7] A. Baliyan, K. Gaurav, S.K. Mishra, A review of short-term load forecasting using artificial neural network models, Proc. Comput. Sci. 48 (2015) 121–125.

[8] S.T. Mehmood, Performance Evaluation of New and Advanced Neural Networks for Short Term Load Forecasting: Case Studies for Maritimes and Ontario, 2014.

[9] G. Nalcaci, A. Özmen, G.W. Weber, Long-term load forecasting: models based on MARS, ANN and LR methods, Cent. Eur. J. Oper. Res. 27 (4) (2019) 1033–1049.

[10] V. Gupta, S. Pal, An overview of different types of load forecasting methods and the factors affecting the load forecasting, Int. J. Res. Appl. Sci. Eng. Technol. (IJRASET) 5 (IV) (2017) 729–733.

[11] J. Wasilewski, D. Baczynski, Short-term electric energy production forecasting at wind power plants in Pareto-optimality context, Renew. Sustain. Energy Rev. 69 (2017) 177–187.

[12] S. Misak, T. Burianek, J. Stuchly, Solar power production forecasting based on recurrent neural network, in: Proceedings of the Afro-European Conference for Industrial Advancement, Advances in Intelligent Systems and Computing, 2016, pp. 195–204.

[13] Y. Zhang, M. Beaudin, H. Zareipour, D. Wood, Forecasting solar photovoltaic power production at the aggregated system level, in: Proceedings of the IEEE North American Power Symposium, 2014, pp. 1–6.

[14] A.T. Eseye, J. Zhang, D. Zheng, Short-term photovoltaic solar power forecasting using a hybrid Wavelet-PSO-SVM model based on SCADA and Meteorological information, Renew. Energy 118 (2018) 357–367.

[15] S. Ramaswamy, P.K. Sadhu, Forecasting PV power from solar irradiance and temperature using neural networks, in: Proceedings of the IEEE, International Conference on Infocom Technologies and Unmanned Systems (Trends and Future Directions), 2017, pp. 244–248. December).

[16] P. Jiang, H. Yang, J. Heng, A hybrid forecasting system based on fuzzy time series and multi-objective optimization for wind speed forecasting, Appl. Energy 235 (2019) 786–801.

[17] Z. Tian, G. Wang, Y. Ren, Short-term wind speed forecasting based on autoregressive moving average with echo state network compensation, Wind Eng. 44 (2) (2020) 152–167.

[18] K. Yan, H. Shen, L. Wang, H. Zhou, M. Xu, Y. Mo, Short-term solar irradiance forecasting based on a hybrid deep learning methodology, Information 11 (1) (2020) 32.

[19] A. Alzahrani, P. Shamsi, C. Dagli, M. Ferdowsi, Solar irradiance forecasting using deep neural networks, Proc. Comput. Sci. 114 (2017) 304–313.

[20] S.S. Pappas, L. Ekonomou, V.C. Moussas, P. Karampelas, S.K. Katsikas, Adaptive load forecasting of the Hellenic electric grid, J. Zhejiang Univ. Sci. A 9 (12) (2008) 1724–1730.

[21] P. Karampelas, V. Vita, C. Pavlatos, V. Mladenov, L. Ekonomou, Design of artificial neural network models for the prediction of the Hellenic energy consumption, in: Proceedings of the 10th Symposium on Neural Network Applications in Electrical Engineering, Belgrade, Serbia, 2010, pp. 41–44, 23-25.

[22] Hamed Aly, Mohamed El-Hawary, A Proposed ANN and FLSM Hybrid Model for Tidal Current Magnitude and Direction Forecasting, IEEE Journal of Oceanic Engineering 39 (1) (2014) 26–31, https://doi.org/10.1109/JOE.2013.2241934.

[23] H.M. Al Ghaithi, G.P. Fotis, V. Vita, Techno-economic assessment of hybrid energy off-grid system - a case study for Masirah island in Oman, Int. J. Power Energy Res. 1 (2017) 103–116.

[24] M.W. Ahmad, J. Reynolds, Y. Rezgui, Predictive modelling for solar thermal energy systems: a comparison of support vector regression, random forest, extra trees and regression trees, J. Clean. Prod. (2018) 810–821.

[25] R. Bayindir, M. Yesilbudak, M. Colak, N. Genc, A novel application of Naive Bayes classifier in photovoltaic energy prediction, in: Proceedings of the 16th IEEE International Conference on Machine Learning and Applications (ICMLA), 2017, pp. 523–527.

[26] J.P. Lai, Y.M. Chang, C.H. Chen, P.F. Pai, A survey of machine learning models in renewable energy predictions, Appl. Sci. 10 (17) (2020) 5975.

[27] P. Gupta, Introduction to Machine Learning in the Cloud with Python: Concepts and Practices, Springer Nature, 2021, https://doi.org/10.1007/978-3-030-71270-9.

[28] L. Breiman, Random forests, Mach. Learn. 45 (1) (2001) 5–32.

[29] M. Fernández-Delgado, E. Cernadas, S. Barro, D. Amorim, Do we need hundreds of classifiers to solve real world classification problems? J. Mach. Learn. Res. 15 (1) (2014) 3133–3181.

[30] C. Chen, A. Liaw, L. Breiman, Using Random Forest to Learn Imbalanced Data, 110, University of California, Berkeley, 2004, p. 24.

[31] B.M. Gayathri, C.P. Sumathi, An automated technique using Gaussian naïve Bayes classifier to classify breast cancer, Int. J. Comput. Appl. 148 (6) (2016) 16–21.

[32] H. Kamel, D. Abdulah, J.M. Al-Tuwaijari, Cancer classification using Gaussian Naive Bayes algorithm, in: Proceedings of the IEEE, 2019 International Engineering Conference, 2019, pp. 165–170.

[33] C. Reinders, H. Ackermann, M.Y. Yang, B. Rosenhahn, Learning convolutional neural networks for object detection with very little training data, In. Multimodal Scene Understanding, Academic Press, 2019, pp. 65–100.

[34] V. Kotu, B. Deshpande, Predictive Analytics and Data Mining: Concepts and Practice with RapidMiner, Morgan Kaufmann Publisher, 2015.

[35] L.Y. Hu, M.W. Huang, S.W. Ke, C.F. Tsai, The distance function effect on k-nearest neighbor classification for medical datasets, SpringerPlus 5 (1) (2016) 1–9.

[36] N. Bhatia, Vandana, Survey of nearest neighbor techniques, Int. J. Comput. Sci. Inf. Secur. 8 (2) (2010) 302–305.

[37] E. Frank, A. Mark, H. Ian, The WEKA Workbench. Online Appendix for Data Mining: Practical Machine Learning Tools and Techniques, (2016).

[38] J. Akosa, Predictive accuracy: a misleading performance measure for highly imbalanced data, in: Proceedings of the SAS Global Forum, 2017, pp. 2–5 (April).

[39] H.H. Aly, Intelligent optimized deep learning hybrid models of neuro wavelet, Fourier series and recurrent Kalman filter for tidal currents constitutions forecasting, Ocean Eng. 218 (2020), 108254.

[40] H.H. Aly, A novel approach for tidal currents harmonic constitutions forecasting hybrid models based on clustering techniques for smart grid, Renew. Energy 147 (Part 1) (2019) 1554–1564.