Electrical Power and Energy Systems 60 (2014) 126–140

Contents lists available at ScienceDirect

Electrical Power and Energy Systems

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

Prediction of full load electrical power output of a base load operated combined cycle power plant using machine learning methods

Pınar Tüfekci *

Department of Computer Engineering, Faculty of Çorlu Engineering, Namık Kemal University, TR-59860 Çorlu, Tekirdağ, Turkey

ARTICLE INFO

Article history:
Received 7 April 2013
Received in revised form 15 February 2014
Accepted 25 February 2014
Available online 24 March 2014

Keywords:
Prediction of electrical power output
Combined cycle power plants
Machine learning methods

ABSTRACT

Predicting full load electrical power output of a base load power plant is important in order to maximize the profit from the available megawatt hours. This paper examines and compares some machine learning regression methods to develop a predictive model, which can predict hourly full load electrical power output of a combined cycle power plant. The base load operation of a power plant is influenced by four main parameters, which are used as input variables in the dataset, such as ambient temperature, atmospheric pressure, relative humidity, and exhaust steam pressure. These parameters affect electrical power output, which is considered as the target variable. The dataset, which consists of these input and target variables, was collected over a six-year period. First, based on these variables the best subset of the dataset is explored among all feature subsets in the experiments. Then, the most successful machine learning regression method is sought for predicting full load electrical power output. Thus, the best performance of the best subset, which contains a complete set of input variables, has been observed using the most successful method, which is Bagging algorithm with REPTree, with a mean absolute error of 2.818 and a Root Mean-Squared Error of 3.787.

1. Introduction

In order for accurate system analysis with thermodynamical approaches, a high number of assumptions is necessary such that these assumptions account for the unpredictability in the solution. Without these assumptions, a thermodynamical analysis of a real application compels thousands of nonlinear equations, whose solution is either almost impossible or takes too much computational time and effort. To eliminate this barrier, the machine learning approaches are used mostly as alternative instead of thermodynamical approaches, in particular, to analyze the systems for arbitrary input and output patterns [1].

Predicting a real value, which is known as regression, is the most common problem researched in machine learning. For this reason, machine learning algorithms are used to control response of a system for predicting a numeric or real-valued target feature. Many real-life problems can be solved as regression problems, and evaluated using machine learning approaches to develop predictive models [2].

This paper deals with several machine learning regression methods for a prediction analysis of a thermodynamic system, which is a combined cycle power plant (CCPP) with two gas turbines, one steam turbine, and two heat recovery systems. Predicting electrical power output of a power plant has been considered a critical real-life problem to construct a model using machine learning techniques. To predict full load electrical power output of a base load power plant correctly is important for the efficiency and economic operation of a power plant. It is useful in order to maximize the income from the available megawatt hours (MW h). The reliability, and sustainability of a gas turbine depend highly on prediction of its power generation, particularly when it is subject to constraints of high profitability and contractual liabilities.

Gas turbine power output primarily depends on the ambient parameters which are ambient temperature, atmospheric pressure, and relative humidity. Steam turbine power output has a direct relationship with vacuum at exhaust. In the literature, the effects of ambient conditions are studied with intelligent system tools such as Artificial Neural Networks (ANNs) for prediction of electrical power (P_E) [1,3,4]. In [1], the effects of ambient-pressure and temperature, relative humidity, wind-velocity and direction on the plant power are investigated using the ANN model, which is based on the measured data from the plant. In [4], the ANN model is used to predict the operational and performance parameters of a gas turbine for varying local ambient conditions.

* Tel.: +90 282 250 23 00; fax: +90 282 652 93 72.
E-mail address: ptufekci@nku.edu.tr

http://dx.doi.org/10.1016/j.ijepes.2014.02.027
0142-0615/

The intelligent systems are also used for modeling a stationary gas turbine. For instance, ANNs identification techniques are developed in [5] and the results show that ANN system identification are perfectly applicable to estimate gas turbine behaviors in wide range of operating points from full speed no load to full load conditions. In [6], Multi Layer Perceptron (MLP) and Radial Basis Function (RBF) Networks are used for identification of stationary gas turbine in startup stage. In [7], dynamic linear models and Feed Forward Neural Networks are compared for gas turbine identification and the neural network is found as a predictor model to identify with better performances than the dynamic linear models. In [8,9], the ANNs models are also used for performance analysis, anomaly detection, fault detection and isolation of gas turbine engines.

Furthermore, in the literature, several studies, e.g., [10–16] have been undertaken to predict electricity energy consumption using machine learning tools, only a few studies such as [1], which is related to prediction of the total electric power of a cogeneration power plant with three gas turbines, one steam turbine and a district a heating system, are found to be as a similar study of this paper.

Pertaining to power plants, it is needed to develop a reliable predictive model for the following day's net energy yield (full load electrical power output) per hour by using real-valued target feature. For this task, there are two main purposes of this study. First one is to determine the best subset of the dataset, which gives the highest predictive accuracy with a combination of the input parameters defined for gas and steam turbines such as ambient temperature, vacuum, atmospheric pressure, and relative humidity. For this purpose, the effects of different combinations of the parameters were investigated and analyzed on predicting full load electrical power output by using 15 machine learning regression methods in WEKA [17] toolbox. Afterwards, the results of the predictive accuracies for the different combinations of the parameters were compared and evaluated to find out the best subset of the dataset. This paper compared the predictive accuracies of the regression methods as the second purpose, which was found out the most successful regression method in the prediction of full load electrical power output of a base load operated CCPP with the highest prediction accuracy.

The remainder of this paper is organized as follows. In Section 2 materials and methods are elaborated, whereas the experimental work is given in Section 3. In Section 4 we provide a discussion of the study and then we conclude in Section 5.

2. Materials and methods

2.1. System description

A combined cycle power plant is composed of gas turbines (GT), steam turbines (ST) and heat recovery steam generators (HRSG). In a CCPP, the electricity is generated by gas and steam turbines, which are combined in one cycle, and is transferred from one turbine to another [18]. A gas turbine in a combined cycle system does not only generate the electrical power but also generates fairly hot exhaust. Routing these gases through a water-cooled heat exchanger produces steam, which can be turned into electric power with a coupled steam turbine and generator. Hence, a gas turbine generator generates electricity and waste heat of the exhaust gases is used to produce steam to generate additional electricity via a steam turbine. This type of power plant is being installed in increasing numbers around the world where there is access to substantial quantities of natural gas [19].

The CCPP,[1] which supplied the dataset for this study, is designed with a nominal generating capacity of 480 MW, made up of 2 × 160 MW ABB 13E2 Gas Turbines, 2 × dual pressure Heat Recovery Steam Generators (HRSG) and 1 × 160 MW ABB Steam Turbine as illustrated in Fig. 1.

Gas turbine load is sensitive to the ambient conditions; mainly ambient temperature (AT), atmospheric pressure (AP), and relative humidity (RH). However, steam turbine load is sensitive to the exhaust steam pressure (or vacuum, V). These parameters of both gas and steam turbines, which are related with ambient conditions and exhaust steam pressure, are used as input variables in the dataset of this study. The electrical power generating by both gas and steam turbines is used as a target variable in the dataset. All the input variables and target variable, which are defined as below, correspond to average hourly data received from the measurement points by the sensors denoted in Fig. 1.

(1) *Ambient Temperature (AT):* This input variable is measured in whole degrees in Celsius as it varies between 1.81 °C and 37.11 °C.
(2) *Atmospheric Pressure (AP):* This input variable is measured in units of minibars with the range of 992.89–1033.30 mbar.
(3) *Relative Humidity (RH):* This variable is measured as a percentage from 25.56% to 100.16%.
(4) *Vacuum (Exhaust Steam Pressure, V):* This variable is measured in cm Hg with the range of 25.36–81.56 cm Hg.
(5) *Full Load Electrical Power Output (P_E): P_E* is used as a target variable in the dataset. It is measured in mega watt with the range of 420.26–495.76 MW.

2.2. Feature subset selection

Data pre-processing is a significant process that contains the processes of cleaning, integration, transformation, and reduction of data for using quality data in machine learning (ML) algorithms. Data sets may vary in dimension from two to thousands of features, and many of these features may be irrelevant or redundant. Feature subset selection decreases the data set dimension by removing irrelevant and redundant features from an original feature set. The objective of feature subset selection is to procure a minimum set of original features. Using the decreased set of original features enables ML algorithms to operate faster and more effectively. Therefore, it helps to predict more correctly by increasing learning accuracy of ML algorithms and improving result comprehensibility [20].

The feature selection process begins by inputting an original feature set, which includes n number of features or input variables. At the first stage of feature selection, which is called subset generation, a search strategy is used for producing possible feature subsets of the original feature set for evaluation. Abstractly, the current best subset of the original feature set can be performed by evaluating all the possible feature subsets, which are all the contending 2^n possible subsets. This search is known as exhaustive search, which is too costly and impracticable if the original feature set consists of enormous features [21]. There are also some several search procedures to find the optimal subset of the original feature set, which are more realistic, easier and more practical. However, in this study, the exhaustive search is used as search procedure. Therefore, every feature combination is tried and marked with a score by using ML regression methods, which equals a value of the prediction accuracy. Then the results of each ML regression method are compared to find the feature subset with the best prediction accuracy, which is called as the best subset.

2.3. Machine learning regression methods

A machine learning (ML) algorithm estimates an unknown dependency between the inputs, which are independent variables,

[1] The name of donor power plant is kept confidential.

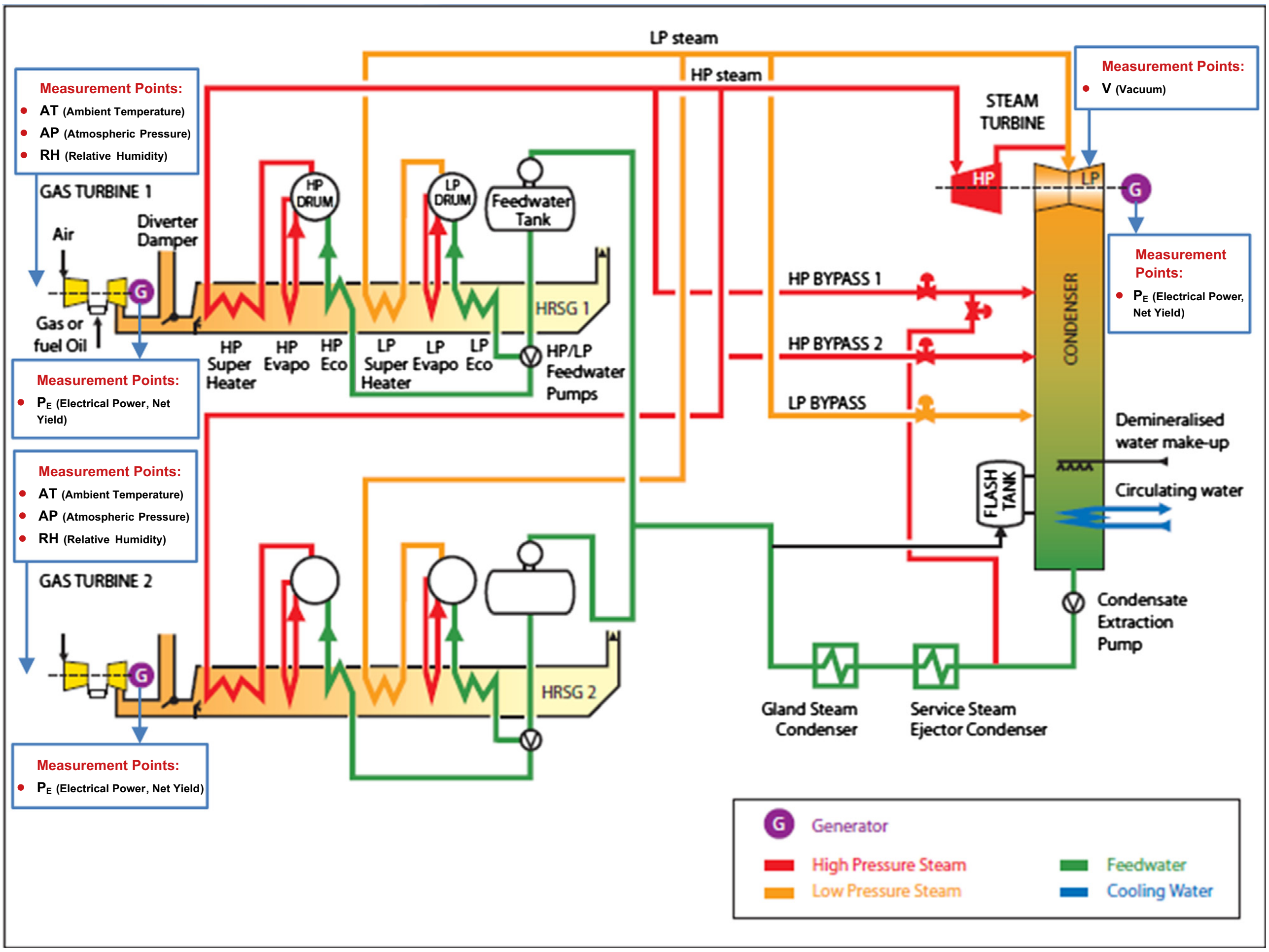

Fig. 1. The combined cycle power plant layout.

and output, which is a dependent variable, from a dataset. In this study, regression methods are generated as learning algorithms to predict full load electrical power output of a combined gas and steam turbines and the dataset is considered as a pair (X_i, Y_i), that is known as an instance. A machine learning regression method, which builds a mapping function $\widehat{Y} = (X_i, Y_i)$ by using these pairs, behaves as shown in Fig. 2. The purpose of a machine learning regression method is to select the best function, which minimizes the error between the actual output (Y) of a system and predicted output $(\widehat{Y})$ based on instances of the dataset, which are called training dataset [22].

Table 1 shows a list of the ML regression methods, which are used in this study. Most of these regression methods have been widely used for modeling many real-life regression problems. These methods are divided into five categories such as Functions, Lazy-learning Algorithms, Meta-learning Algorithms, Rule-based Algorithm, and Tree-based Learning Algorithms, stated by the WEKA statistical analysis package. Functions contain algorithms,

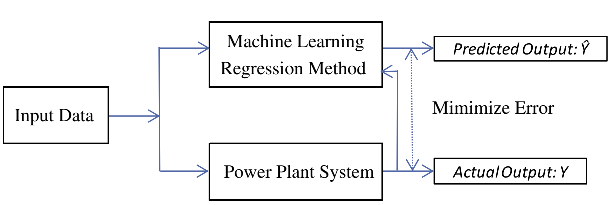

Fig. 2. A machine learning regression method using real system data to predict output.

which are based on the mathematical models. Lazy-learning algorithms delay dealing with training data until a query is answered. They store the training data in memory and find relevant data in the database to answer a particular query. Meta-learning algorithms integrate different kinds of learning algorithms to enhance the achievement of the used current learning algorithms. Rule-based algorithm uses decision rule for regression model. Tree-based learning algorithms are used for making predictions via a tree structure. Leaves of the tree structures illustrate classifications and branches of the tree structures denote conjunctions of features.

Here we present a brief summary of the methods in Table 1.

(1) *Simple Linear Regression (SLR):* The SLR generates a regression model, which has lowest squared error. This model fits straight models between each input attribute (a_0 and a_1) and output (x) as in Eq. (1), in which the values of w and w_0, which are the weight of a_0 and a_1, are estimated by the method of least squares. In Eq. (1), a_0 is assumed as the constant 1.

$$x = w_0 + wa_1 \tag{1}$$

The predictive model is chosen by minimizing the squared error, which is the difference between the observed values and the predicted values as illustrated in Fig. 2 [23].

(2) *Linear Regression (LR):* The LR presents a mathematical model of the relationship between a dependent variable

Table 1
Regression methods used in this study.

Categories	Method name	Abbreviation
Functions	Simple Linear Regression	SLR
	Linear Regression	LR
	Least Median Square	LMS
	Multilayer Perceptron	MLP
	Radial Basis Function Neural Network	RBF
	Pace Regression	PR
	Support Vector Poly Kernel Regression	SMOReg
Lazy-learning algorithms	IBk Linear NN Search	IBk
	KStar	K*
	Locally Weighted Learning	LWL
Meta-learning algorithms	Additive Regression	AR
	Bagging REP Tree	BREP
Rule-based algorithm	Model Trees Rules	M5R
Tree-based learning algorithms	Model Trees Regression	M5P
	REP Trees	REP

and one or more independent variables. The data is observed using a linear model by this algorithm. LR method deals with weighted instances to create a prediction model. The least squares regression is performed to specify linear relations in the training data. If there is some linear dependency among the data, LR may create a best predictive model, which is a linear regression equation to predict the output value (x) for a set of input attributes $a_1, a_2, \ldots, a_k$. In Eq. (2), $w_0, w_1, \ldots, w_k$, are the weights respectively of each input attribute, where w_1 is the weight of a_1 and a_0 is always considered as the constant 1.

$$x = w_0 + w_1 a_1 + \cdots + w_k a_k \tag{2}$$

The weights must be selected to minimize the difference between the actual output value and predicted output value. In Eq. (3), the predicted output value for the first instance of a training dataset is calculated as

$$w_0 + w_1 a_1^{(1)} + \cdots + w_k a_k^{(1)} = \sum_{j=0}^{k} w_j a_j^{(1)} \tag{3}$$

After the predicted outputs for all instances are calculated, the weights are updated to minimize sum of squared differences between the actual and predicted outcome as in Eq. (4) [24,25].

$$\sum_{i=1}^{n} \left(x^{(i)} - \sum_{j=0}^{k} w_j a_j^{(i)} \right) \tag{4}$$

(3) *Least Median Square (LMS):* The LMS method is a linear regression method reduces the median squared error. In LMS algorithm, the weights are reassigned to minimize the median of the squares of the difference between the actual output and the predicted outcome using the regression equation as in Eq. (5) [26].

$$\underset{i}{median} \left(x^{(i)} - \sum_{j=0}^{k} w_j a_j^{(i)} \right) \tag{5}$$

(4) *Multi Layer Perceptron (MLP):* The MLP is a feed forward artificial neural network model that consists of neurons with massively weighted interconnections, where signals always travel in the direction of the output layer. These neurons are mapped as sets of input data onto a set of appropriate outputs with hidden layers as illustrated in Fig. 3. The input signals are sent by the input layer to the hidden layer without performing any operations. Then the hidden and output layers multiply the input signals by a set of weights, and either linearly or nonlinearly transforms results into output values. These weights are optimized to obtain reasonable prediction's accuracy. A typical MLP with one hidden layer can be mathematically defined in Eqs. (6)–(10) as below [27]:

$$u_j = \sum_{i=1}^{N_{inp}} X_i a_{ij} + a_{oj} \tag{6}$$

Eq. (6) expresses summing products of the inputs (X_i) and weight vectors (a_{ij}) and a hidden layer's bias term (a_{0j}). In Eq. (7), the outputs of hidden layer (Z_j) are obtained as transforming this sum, which is defined in Eq. (6), using transfer function (activation function) g.

$$Z_j = g(u_j) \tag{7}$$

The most widely used transfer function is sigmoid function, which is defined in Eq. (8) for input x. The hidden and output layers are based on this sigmoid function.

$$g(x) = sigmoid(x) = \frac{1}{(1 + e^{-x})} \tag{8}$$

Eq. (9) defines summing products of hidden layer's outputs (Z_j) and weight vectors (b_{jk}) and output layer's bias term (b_{0k}).

$$v_k = \sum_{j=1}^{N_{hid}} Z_j b_{jk} + b_{0k} \tag{9}$$

In Eq. (10), the outputs of the output layer (Y_k) are obtained by transforming this sum, which is obtained in Eq. (9), using sigmoid function g, which is defined in Eq. (8) [22].

$$Y_k = g(v_k) \tag{10}$$

(5) *Radial Basis Function Neural Network (RBF):* The RBF Neural Network is emerged as a variant of neural network. It uses a normalized distance between the input points and the hidden nodes to define the activation of each node [27]. RBF Network maps inputs from the input layer to each of the hidden units, which use radial functions for activation. A Gaussian function is useful in finding the activation at a hidden unit. The Gaussian function is defined as [28]:

$$h(x) = \exp\left(\frac{-(x-c)^2}{r^2} \right) \tag{11}$$

where c is the center of bell-shaped Gaussian and r is the width.

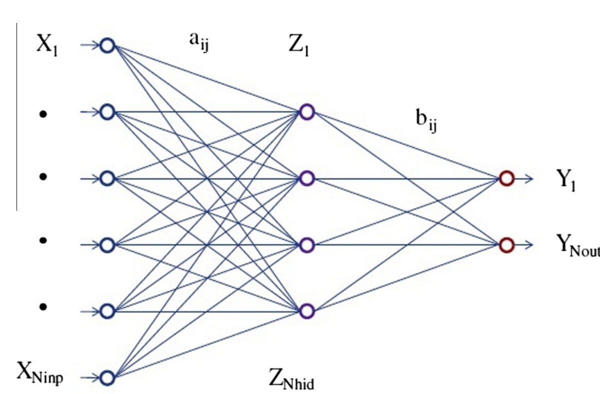

Fig. 3. Multi layer perceptron neural networks.

(6) *Pace Regression (PR):* The PR method creates a predictive model by evaluating the effects of each variable. It uses a cluster analysis to enhance the statistical basis for estimating their contribution to the overall regression [23].
(7) *Support Vector Poly Kernel Regression (SMOReg):* Support vector machines (SVM) are kernel based learning algorithm for solving classification and regression problems. Support vector regression (SVR) maps the input data x into a higher dimensional feature space by nonlinear mapping to solve a linear regression problem in this feature space. This transformation can be done using a kernel function. The most common existing kernel functions are linear kernel, polynomial kernel, Gaussian (RBF) kernel, and sigmoid (MLP) kernel [28]. In this study, Support Vector Poly Kernel Regression is used as an implementation of the sequential minimal optimization algorithm for training a support vector regression model.
(8) *IBk Linear NN Search (IBk):* The IBk instance-based learning that works as a k-nearest-neighbor classifier, which is the most commonly used instance-based or lazy method for both classification and regression problems. In this paper, it is used for a regression problem. The algorithm normalizes attributes by default and can do distance weighting. A variety of different search algorithms are used to speed up the task of finding the nearest neighbors [20].

The KNN algorithm first measures distances between each instance in the training dataset and the test instance according to a distance metric, which is often used Euclidian distance, then the nearest instances to the test instance are determined as the target value. For this purpose, the KNN algorithm gives higher values to the weights of the nearer neighbors. Thus, Eq. (12) is used to predict the target value of the test instance (q_i) [2].

$$\mathrm{Sim}(q_i, s_j) = 1 - \frac{\sqrt{\sum_{f-1}^{N} \delta(q,s,f)^2}}{\sqrt{N}} \tag{12}$$

(9) *KStar (K^*):* KStar method is also an instance-based classifier used for regression. It generates a predictive model by using some similarity function based on an entropy-based distance function [29].
(10) *Locally Weighted Learning (LWL):* The LWL uses an instance-based algorithm, which assigns instance weights. This algorithm can perform classification or regression [20].
(11) *Additive Regression (AR):* The AR is a meta-learning algorithm, which produces predictions by combining contributions from an ensemble (collection) of different models. Each iteration fits a model to the residuals left by the classifier on the previous iteration. Prediction is accomplished by adding the predictions of each classifier. Reducing the shrinkage parameter helps prevent over fitting and has a smoothing effect, although it increases the learning time [30].
(12) *Bagging REP Tree (BREP):* Bagging or bootstrap aggregating is general technique for improving prediction rules by creating various sets of the training sets. Bagging algorithm is applied to tree-based methods such as REP Trees to reduce the variance associated with prediction, and therefore, increase the accuracy of the resulting predictions. Bagging can be formalized as follows

$$\hat{y}_{BAG} = \frac{1}{B}\sum_{b=1}^{B} \emptyset(x; T_b) \tag{13}$$

where B is the number of bootstrap samples of training set T and x is the input. $\hat{y}_{BAG}$ is the average of the different estimated trees. A bootstrap sample is randomly drawn from the training set, but with replacement. The purpose is to create numerous similar training sets using, sampling and train a new function for each of these sets. The functions learned from these sets are then used collectively to predict the test set [31,32]. In this study, we use Bag Size Percentage of 100, with 10 iterations and REP Tree as a predictor. These are the default settings provided by the WEKA tool.
(13) *Model Trees Rules (M5R):* The M5R is a rule-based algorithm, based on M5 algorithm, which builds a tree to predict numeric values for a given instances. For a given instance, the tree is traversed from top to bottom until a leaf node is reached and the best leaf into a rule is made using a decision list [23].
(14) *Model Trees Regression (M5P):* The M5P is a regression-based decision tree algorithm, which builds a model tree using the M5 algorithm. For a given instance, the tree is traversed from top to bottom until a leaf node is reached. At each node in the tree, a decision is made to follow a particular branch based on a test condition on the attribute associated with that node. As the leaf nodes contain a linear regression model to obtain the predicted output, the tree is called a model tree. The M5P algorithm builds a model tree using to divide and conquer method [33].
(15) *Reduced Error Pruning (REP) Trees:* The REP Trees algorithm creates a regression tree using the node statistics such as information gain or variance reduction measured in the top-down phase, and prunes it using reduced-error pruning [34].

3. Comparative analysis

This section discusses the comparative analysis that was conducted to evaluate and compare the some machine learning regression methods available for predicting hourly full load electrical power output of a CCPP, which is combined with gas and steam turbines. First, the best subset of the dataset is explored among all feature subsets in the experiments. Then, the most successful machine learning regression method is sought for predicting full load electrical power output. The flow diagram of the prediction process is illustrated in Fig. 4.

3.1. Dataset description

The dataset used in this study, which consists of four input variables and a target variable, was collected over a six-year period (2006–2011). It is composed of 9568 data points collected when the CCPP is set to work with a full load over 674 different days. The input variables correspond to average hourly data received from the measurement points by the sensors denoted in Fig. 1.

The input variables affect full load electrical power output (P_E), which is considered as the target variable in the dataset and corresponds to average hourly full load electrical power output data received from the control system when the power plant was at base load.

At the beginning of the pre-processing stage shown in Fig. 4, the dataset was consisting of 674 datasheets (formatted in .xls) of 674 different days with some noisy and incompatible data. The data is cleaned by filtering the incompatible data, which is out of range meaning that if the power plant is operating below 420.26 MW, and the noisy data, which occurs when the signal is interfered from electrical disturbance.

Then the datasets are merged to eliminate any duplication data and to be integrated dataset. After that, the dataset, which includes 674 files formatted as .xls, is integrated as a file formatted in .xls. Then it has been randomly shuffled five times and transformed to a file formatted as .arff that is necessary for

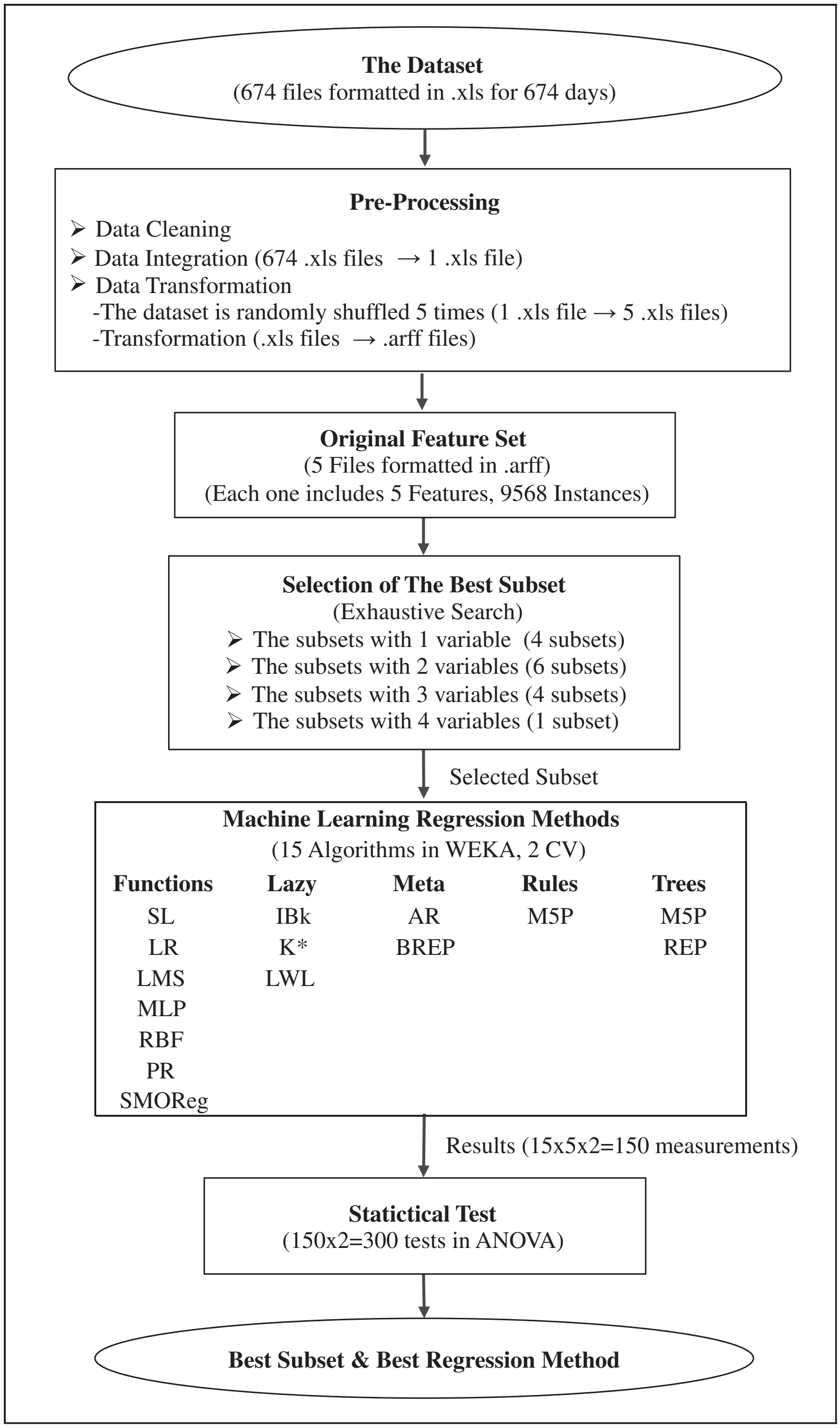

Fig. 4. The flow diagram of the prediction process.

processing in WEKA tool. At the end of pre-processing, the original feature set is composed of five files formatted as .arff and each file consists of 9568 data points with five totally parameters as the integrated data set.

Table 2 denotes simple statistics of the dataset.

Table 3 is the covariance matrix which indicates that the parameters are not independent.

Table 4 illustrates the parameters' cross-correlation. The scatter plot of the data used is given in Fig. 5. When Table 4 and Fig. 5 are examined together, the highest correlation among input features is observed between AT and V (0.84). Moreover, the highest correlations with target variable (P_E) are also observed with AT (−0.95) and V (−0.87).

According to Fig. 5, atmospheric pressure (AP) and relative humidity (RH) do not have a strong correlation with the target variable (P_E) sufficient for an individual prediction. When other factors remain constant, it has been shown that P_E increases with increasing AP and RH individually. Here the effects of each input variable individually on target variable are presented as below:

(1) *Effect of Ambient Temperature (AT):* The effect of AT on the performance is the most widely studied subject about gas turbines [35–37]. This can be appreciated since AT is the most influential factor showing a correlation around −0.95 with P_E as illustrated in Fig. 5.

Fig. 6 shows the scatter diagram of AT vs. P_E and the linear regression model fit to data. This illustration is done using the whole dataset for the purpose of preliminary investigation. The resulting predictive model

$$y = -2.1713x + 497.03 \tag{14}$$

can be interpreted as unit (°C) increase in AT accounts for a reduction of 2.1713 MW in P_E. The performance reduction due to increase in temperature is known to stem from the decrease in the density of inlet air.

(2) *Effect of Ambient Pressure (AP):* Among the ambient variables, the second most influential one is AP [35]. However, it does not have a strong correlation with the target variable sufficient for an individual prediction, as it can be observed from the scatter diagram shown in Fig. 7, which indicates the predictive model for whole dataset as below:

$$y = 1.4335x - 998.78 \tag{15}$$

Similar to affect the mechanism of AT, AP is responsible for the density of inlet air. The slope of the linear regression function tells us that unit increase in AP corresponds to 1.4335 MW increase in P_E. However, this is a naïve estimation taking into account the other factors. When other factors remain constant, it has been shown that P_E increases with increasing AP [35].

(3) *Effect of Relative Humidity (RH):* When other variables are kept constant, the performance is increased with RH [35,38]. However, as it can be seen from Fig. 8, which gives the predictive model as in Eq. (16), it is not self-sufficient for prediction.

$$y = 0.4556x + 420.96 \tag{16}$$

Higher relative humidity increases exhaust-gas temperature of gas turbine which leads to an increase in the power generated by the steam turbine [35].

(4) *Effect of Vacuum (V):* As seen in Fig. 1, the plant also employs a steam turbine which leads to considerable increase in total electrical efficiency. V variable (exhaust vacuum in cm Hg) is collected from steam turbine and is found influential in its performance as in Fig. 5. When all other variables are kept constant exhaust-vacuum is known to have a negative influence on condensing-type turbine efficiency [39]. As depicted in Fig. 9, P_E is inversely proportional to V as in Eq. (17).

$$y = -1.1681x + 517.8 \tag{17}$$

It can also be observed that the slope of linear regression function is greater than those of RH and AP but less than AT.

Table 2
Basic statistics of dataset.

	AT	V	AP	RH	P_E
Min	1.81	25.36	992.89	25.56	420.26
Max	37.11	81.56	1033.30	100.16	495.76
Mean	19.65	54.31	1013.26	73.31	454.37
Variance	55.54	161.49	35.2	213.17	291.28

Table 3
Covariance matrix.

	AT	V	AP	RH	P_E
AT	55.54	79.94	−22.46	−59.03	−120.59
V		161.49	−31.21	−57.92	−188.64
AP			35.27	8.63	52.55
RH				213.17	97.13
P_E					291.28

Table 4
Correlation matrix.

	V	AP	RH	P_E
AT	0.84	−0.51	−0.54	−0.95
V		−0.41	−0.31	−0.87
AP			0.10	0.52
RH				0.39

3.2. Prediction accuracy

The prediction accuracy of each machine learning regression method is used to evaluate the overall match between actual and predicted values. In this paper, the prediction accuracy is evaluated by using the following performance criteria such as Mean Absolute Error (MAE) and Root Mean Squared Error (RMSE) for continuous variables.

(1) *Mean Absolute Error (MAE):* Mean absolute error is the average of the difference between predicted and actual value in all test cases, without considering their direction [40].

$$\mathrm{MAE} = \frac{|a_1 - c_1| + |a_2 - c_2| + \cdots + |a_n - c_n|}{n} \tag{18}$$

(2) *Root Mean-Squared Error (RMSE)*: Root Mean Square Error, RMSE is a frequently used measure of differences between values predicted by a model or estimator and the values actually observed from the process being modeled or estimated [40].

$$\mathrm{RMSE} = \sqrt{\frac{(a_1 - c_1)^2 + (a_2 - c_2)^2 + \cdots + (a_n - c_n)^2}{n}} \tag{19}$$

In the above Eqs. (18) and (19), a is the actual value of the output, c is the predicted value of the output. In all above error measurements, a lower value means a more precise model, with a value of 0 depicting a statistically perfect model [41].

3.3. Producing feature subsets

Selection of the best subset is related to choosing a subset, among the all feature subsets produced. The best subset is a feature set, which indicates the best performance in prediction accuracy. Theoretically, the best subset can be found by evaluating all the possible subsets, which is known as exhaustive search. An exhaustive search of the feature space needs to search all of 2^n possible subsets of n features. In this study, the exhaustive search is applied to the original dataset, which consists of four parameters as input variables and a target parameter as a response. The aim is to choose a minimal model with the best subset, which correctly predicts the response [42]. For this aim, after collecting preliminary statistical data, we applied an exhaustive search to the original dataset to find the best subset by evaluating all the competing candidate subsets ($2^4 - 1 = 15$) in the experiments. Moreover, we have divided the experiments to four categories, in which the subsets with one, two, three, and four parameters are applied to the regression methods. In these experiments, we have determined

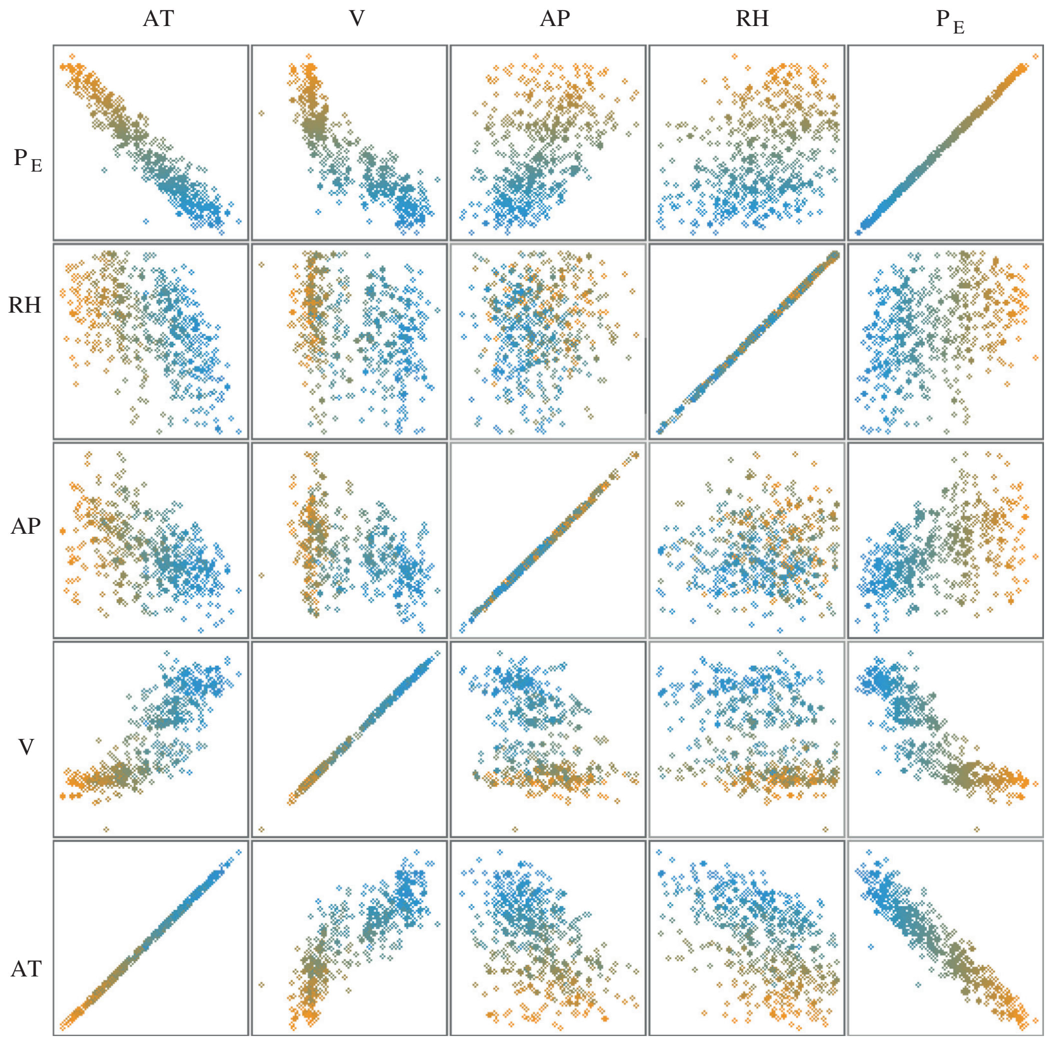

Fig. 5. Pairwise scatter plot of parameters.

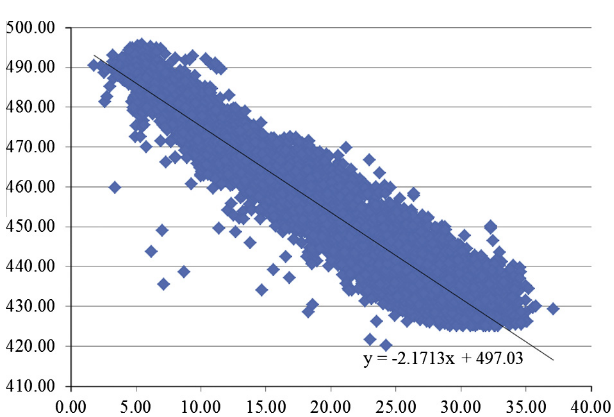

Fig. 6. Scatter diagram of AT (°C, x axis) vs. P_E (MW, y axis).

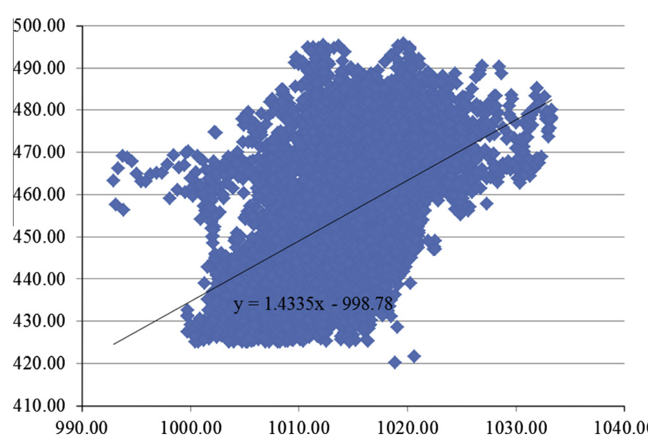

Fig. 7. Scatter diagram of AP (mb, x axis) vs. P_E (MW, y axis).

the best subset of each experiment by evaluating and comparing the results of all regression methods for the candidate subsets, which are shown in Table 5.

3.4. Cross-validation

For subsequent experiments of this study, in order to compare performance of different learning algorithms, 5×2 cross-validation (CV) [43] was applied. In this scheme, the dataset is randomly shuffled five times and each of them used in 2-fold cross-validation.

CV is a validation scheme where the dataset is partitioned into equal sized subsets. In the general case of K-fold CV, at each machine learning experiment one subset is used for validation (i.e. to test the predictive model) and the rest is for training. 2-fold

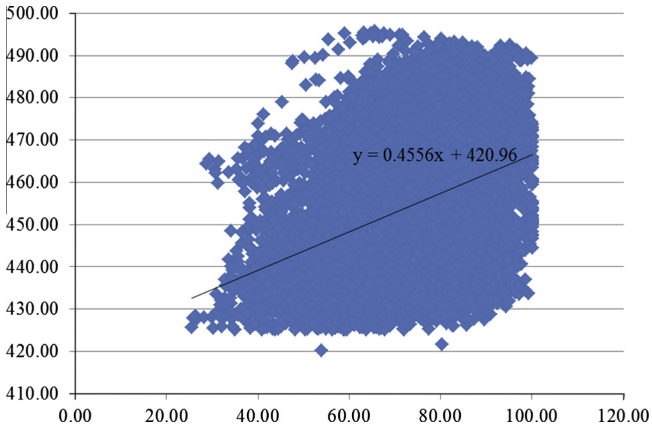

Fig. 8. Scatter diagram of RH (*x* axis) vs. P_E (MW, *y* axis).

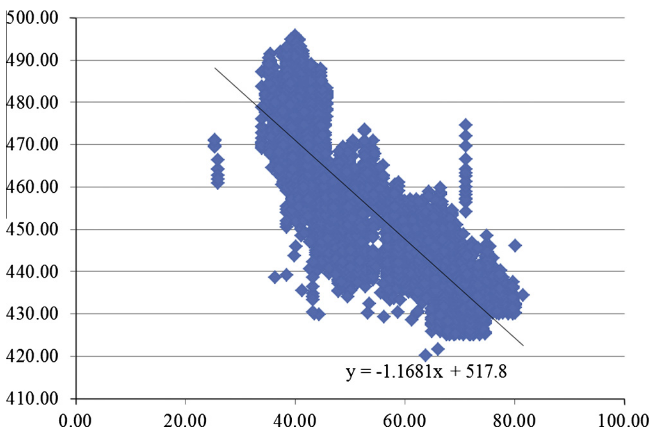

Fig. 9. Scatter diagram of V (cm Hg, *x* axis) vs. P_E (MW, *y* axis).

Table 5
All possible subsets of the dataset.

Subsets	AT	V	AP	RH
With 1 parameter				
AT	1	0	0	0
V	0	1	0	0
AP	0	0	1	0
RH	0	0	0	1
With 2 parameters				
AT–V	1	1	0	0
AT–AP	1	0	1	0
AT–RH	1	0	0	1
V–AP	0	1	1	0
V–RH	0	1	0	1
AP–RH	0	0	1	1
With 3 and 4 parameters				
AT–V–AP	1	1	1	0
AT–V–RH	1	1	0	1
AT–AP–RH	1	0	1	1
V–AP–RH	0	1	1	1
AT–V–AP–RH	1	1	1	1

CV is a special case where at first experiment one half is trained to validate on the second half and then the roles of the initial subsets are swapped [44]. For the generalization of results, each cell in the following tables (Tables 6–10) corresponds to an average of 10 (5 × 2 CV) [30] runs in WEKA toolbox.

3.5. Analysis of variance test

The resulting validation set performances of size 10 for each learning algorithm are used for a statistical significance test. Thus, one-way Analysis of Variance (ANOVA) has been used for the statistical tests, which are applied to average results obtained from each 2-fold CV. ANOVA is a parametric test in which the mean across different groups is compared, and one-way ANOVA is chosen here to analyze the significance of difference with respect to a single-factor [45]. It is also used to compare results of machine learning experiments [44]. The idea is to decompose the total variability into within group and between group variability, which provides estimators to dataset variance. Then the ratio of mean between and mean within group variance provide the test statistic. One-way ANOVA tests the null hypothesis that the means of all treatments are equal, and the alternative hypothesis is that at least one pair is significantly different. When the null hypothesis is rejected post hoc tests are issued to evaluate pairwise differences [44,45]. In our machine learning experiments, we test one-way ANOVA for comparing algorithms and for comparing feature subsets independently. Therefore, 300 measurements were obtained over 15 × 5 × 2 × 2, which correspond to the number of methods, 2-fold average and features, respectively.

3.6. Selection of the best subset

The controlled experimental results, which belong to following four experiments, have shown by using which subset the response

Table 6
Regression errors with RMSE performances on the subsets with one parameter.

Categories	Regression methods	AT	V	AP	RH
Functions	SLR	5.426	8.423	14.597	15.720
	LR	5.426	8.423	14.597	15.720
	LMS	5.433	8.424	14.847	15.733
	MLP	6.483	9.454	15.601	18.031
	RBF	7.501	10.138	14.586	15.869
	PR	5.426	8.423	14.597	15.720
	SMOReg	5.433	8.423	14.950	15.826
Lazy-learning algorithms	IBk	6.377	5.266	17.448	20.585
	K*	5.381	7.828	14.377	15.729
	LWL	8.005	8.954	14.898	15.947
Meta-learning algorithms	AR	5.933	7.746	14.342	15.760
	BREP	5.208	5.372	14.559	16.020
Rule-based algorithm	M5R	5.085	6.370	14.300	15.720
Tree-based learning algorithms	M5P	5.086	6.476	14.295	15.720
	REP	5.229	5.520	14.467	15.829
Mean		5.829	7.683	14.831	16.262

Table 7
Regression errors with RMSE performances on the subsets with two parameters.

Categories	Regression methods	AT–V	AT–AP	AT–RH	V–AP	V–RH	AP–RH
Functions	SLR	5.425	5.426	5.426	8.423	8.423	14.597
	LR	4.956	5.376	4.800	7.884	8.151	13.395
	LMS	4.968	5.384	4.808	7.923	8.152	13.612
	MLP	6.091	6.399	5.687	8.579	9.498	14.310
	RBF	8.757	7.719	10.665	9.962	9.903	15.558
	PR	4.956	5.376	4.800	7.884	8.151	13.395
	SMOReg	4.968	5.392	4.811	7.900	8.162	13.652
Lazy-learning algorithms	IBk	5.560	6.619	6.239	8.039	9.407	17.562
	K*	4.634	5.157	4.903	6.926	7.481	12.996
	LWL	7.915	8.211	8.254	8.918	9.080	14.399
Meta-learning algorithms	AR	5.545	5.885	5.899	7.514	7.649	13.387
	BREP	4.026	4.900	4.607	5.300	5.796	12.948
Rule-based algorithm	M5R	4.419	4.904	4.554	6.747	7.109	13.074
Tree-based learning algorithms	M5P	4.359	4.897	4.558	6.475	6.905	13.041
	REP	4.339	5.059	4.772	5.888	6.476	13.248
Mean		5.395	5.780	5.652	7.624	8.023	13.945

Table 8
Regression errors with RMSE performances on the subsets with three parameters.

Categories	Regression methods	AT–V–AP	AT–V–RH	AT–AP–RH	V–AP–RH
Functions	SLR	5.426	5.426	5.426	8.423
	LR	4.891	4.570	4.800	7.562
	LMS	4.905	4.580	4.809	7.588
	MLP	5.794	5.341	5.544	8.492
	RBF	8.457	8.695	9.166	9.571
	PR	4.891	4.570	4.799	7.562
	SMOReg	4.905	4.585	4.812	7.569
Lazy-learning algorithms	IBk	4.921	5.282	5.737	7.775
	K*	4.201	4.331	4.717	6.245
	LWL	8.062	8.211	8.322	8.980
Meta-learning algorithms	AR	5.552	5.549	5.907	7.415
	BREP	3.855	3.922	4.438	5.324
Rule-based algorithm	M5R	4.268	4.217	4.505	6.552
Tree-based learning algorithms	M5P	4.205	4.178	4.504	6.354
	REP	4.232	4.291	4.748	6.072
Mean		5.238	5.183	5.482	7.432

can be predicted with the highest prediction accuracy by the regression methods. The RMSE measure is used in the tables to represent the performance of the regression methods to determine the best subset.

In the first experiment, each input variable of the dataset has applied individually to the regression methods. Thus, the performance of the methods for predicting P_E for the subsets, which consist of only one parameter, is presented in Table 6. According to the mean performance of each subset with one parameter for whole regression methods in this table, the best subset of this experiment is found as the subset with AT parameter with the highest prediction accuracy, which is the lowest error in terms of a RMSE of 5.829.

In addition to that, a one-way ANOVA was used to test the statistical significance of the difference in predictive performances among machine learning algorithms using only one feature. The test result indicated no significant difference, $F(14,285) = .48$, $p = .940$. A second one-way ANOVA was used to test for performance differences among machine features over all algorithms. The test result indicated significant difference between features, $F(3,296) = 1461.38$, $p < .05$. A multiple comparison test based on Tukey-HSD indicated that all groups differ significantly from each other and the performance ordering is AT < V < AP < RH, which implied AT yields the minimal error which validates the ANOVA result (see Fig. 10) as also the mean performance results shown in Table 6.

In the second experiment, the performance of the methods for predicting P_E for the subsets, which consist of only two parameters, is compared in Table 7. This table indicates that the highest prediction accuracy is found as a mean RMSE value of 5.395 for the subset with AT, and V parameters, which is the best subset among the subsets with two parameters.

On the other hand, statistical tests indicated that there is a significant difference between performances of algorithms when tested under one-way ANOVA, $F(14,435) = 4.17$, $p < 0.05$. However, the post hoc tests had shown that the performance of the best algorithm with two-parameter models was found to be significantly better than only LWL and RBF. A one-way ANOVA was used to test performances of subsets with two parameters. Results indicated a significant difference among groups, $F(5,444) = 485.76$, $p < 0.05$. Multiple comparison tests indicated no significant difference among subsets containing AT (i.e. AT–V, AT–AP, and AT–RH). Moreover, all subsets, including AT were found to be significantly better than the subsets without AT. The subset with AP, and RH parameters was found to be significantly the worst as can be observed in Table 7.

In the third experiment, the subsets with three parameters are applied to all regression methods. Table 8 illustrates that the BREP

Table 9
Regression errors with RMSE performances on the subsets with four parameters and the best RMSE performances of the each previous experiments applied to subsets with one. two. and three parameters.

Categories	Regression methods	AT	AT–V	AT–V–RH	AT–V–AP–RH	Mean
Functions	SLR	5.426	5.425	5.426	5.426	5.426
	LR	5.426	4.956	4.570	4.561	4.878
	LMS	5.433	4.968	4.580	4.572	4.888
	MLP	6.483	6.091	5.341	5.399	5.829
	RBF	7.501	8.757	8.695	8.487	8.360
	PR	5.426	4.956	4.570	4.561	4.878
	SMOReg	5.433	4.968	4.585	4.563	4.887
Lazy-learning algorithms	IBk	6.377	5.560	5.282	4.656	5.469
	K*	5.381	4.634	4.331	3.861	4.552
	LWL	8.005	7.915	8.211	8.221	8.088
Meta-learning algorithms	AR	5.933	5.545	5.549	5.556	5.646
	BREP	5.208	4.026	3.922	3.779	4.234
Rule-based algorithm	M5R	5.085	4.419	4.217	4.128	4.462
Tree-based learning algorithms	M5P	5.086	4.359	4.178	4.087	4.428
	REP	5.229	4.339	4.291	4.211	4.518
Mean		5.829	5.395	5.183	5.071	5.370

Table 10
Results of the best regression methods of each category for the best subsets of the experiments.

Categories	Regression methods	AT		AT–V		AT–V–RH		AT–V–AP–RH		Mean	
		MAE	RMSE	MAE	RMSE	MAE	RMSE	MAE	RMSE	MAE	RMSE
Functions	LMS	4.285	5.433	3.912	4.968	3.619	4.580	3.621	4.572	3.859	4.888
	SMOReg	4.284	5.433	3.913	4.968	3.620	4.585	3.620	4.563	3.859	4.887
Lazy-learning algorithms	K*	4.260	5.381	3.628	4.634	3.358	4.331	2.882	3.861	3.532	4.552
Meta-learning algorithms	BREP	4.074	5.208	3.035	4.026	2.952	3.934	2.818	3.787	3.220	4.239
Rule-based algorithm	M5R	3.985	5.085	3.418	4.419	3.265	4.217	3.172	4.128	3.460	4.462
Tree-based learning algorithms	M5P	3.982	5.086	3.362	4.359	3.229	4.178	3.140	4.087	3.428	4.428
	REP	4.089	5.229	3.258	4.339	3.214	4.291	3.133	4.211	3.424	4.518

method gives the highest prediction accuracy for each subset with three parameters in RMSE measure. According to the mean results of Table 8, the best subset with three parameters of this experiment is obtained as the subset comprising AT, V, and RH parameters with RMSE of 5.183 as a mean value, which is the average of the results of all regression methods for the selected subset. Similarly, a one-way ANOVA indicated significant difference between groups, $F(14,285) = 31.13$, $p < 0.05$. Post-hoc tests indicated that BREP is significantly better than 6 out of 14 methods, namely AR, IBk, LWL, MLP, RBF, and SLR. MLP was found to be significantly better than the least 2 and significantly worse than the top 5 ordered by means. Post-hoc tests after issuing ANOVA revealed significant difference between all subsets containing AT and V-AP-RH. However, no significant difference was observed among the performances of the subsets with three parameters containing AT (see Fig. 11).

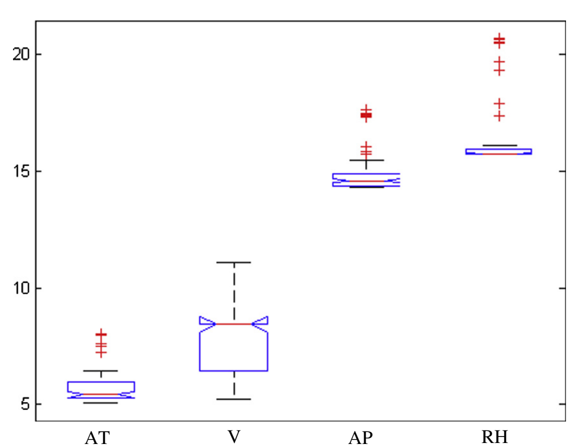

Fig. 10. The ANOVA plot of RMSE performances of models using a single parameter. Models enumerated correspond to AT, V, AP, and RH, respectively.

In the last experiment, there is only one subset with four parameters to apply to the regression methods. Table 9 illustrates the best results of the other three experiments and the results of the subset with four parameters. According to the mean RMSE values of Table 9, the best subset is found to be the subset with AT, V, AP, and RH parameters, which can be used for predicting P_E with the highest prediction accuracy, with the mean RMSE value of 5.071.

3.7. Selection of the best regression method

Table 9 is a brief table, which indicates the best subset and the best regression method with the highest prediction accuracy of P_E. According to this table, the subset with four parameters has been found as the best subset. Moreover, the comparison of the mean performances indicates that BREP method, which is a meta classifier used as a learner combination, is the best regression method, which is the most successful regression method with the highest mean prediction accuracy with a mean RMSE value of 4.234 for all the best subsets of previous experiments in this table. Afterward, BREP is followed by respectively M5P, M5R, REP, K*, LR, PR, SMOReg, LMS, SLR, IBk, AR, MLP, LWL, and RBF regression methods on all evaluation methods. The RBF method is found to be the poorest performing predictive model.

In addition to that, the M5R method is the best method with a RMSE of 5.085 for the subset AT, which is the best subset of the

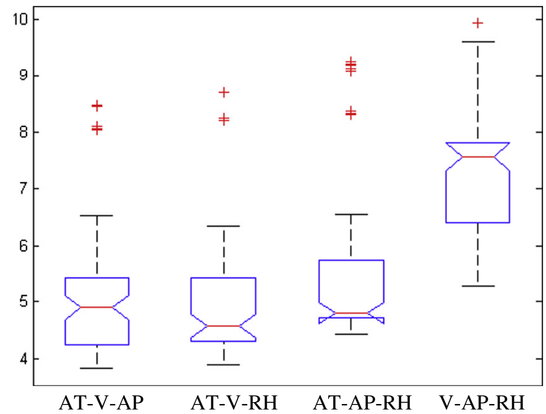

Fig. 11. The ANOVA plot of models (over all algorithms) with 3-parameter subsets. As can be seen, the three subsets containing AT do not significantly differ among themselves however all are significantly better than the one without AT (i.e. subset with V–AP–RH).

subsets with one parameter. However, the BREP method is the best method with a RMSE of 4.026 for the subset AT, and V, which is the best subset of the subsets with two parameters, and with a RMSE of 3.922 for the subset AT, V, and RH, which is the best subset of the subsets with three parameters, and with a RMSE of 3.779 for the subset AT, V, AP, and RH, that is the best subset of allover the produced subsets.

According to the performances in Table 9, the scatter plots of the most successful algorithms for each best subset with one, two, three and four parameters are presented for the actual and observed P_E in Fig. 12. This figure illustrates that BREP performance for the best subset with four parameters fits best to the ideal line (i.e., the diagonal line), followed by BREP for the subsets with three, and two parameters, and M5R for the subset with one parameter. Most of the predicted values of the M5R method are above the ideal prediction line. This implies that the M5R method over estimates the overall predictions.

The best performed algorithms in each category for the best subsets of the subsets with one, two, three, and four parameters are denoted in Table 10. As can be seen this table, the SMOReg method has the best performances of the functions for the subset AT, and the subset AT, V, AP, and RH; the LMS method has the best performance of functions for the subset AT, and V, and the subset AT, V, and RH. The K* is the best method of the lazy-learning algorithms for the all subsets with one, two, three, and four parameters. The BREP is the most successful algorithm of the meta-learning algorithms used in this study for all the subsets with one, two, three, and four parameters. The M5R is the only rule-based algorithm used for this study. Though the M5P is the best method of tree-based algorithms for the subset AT, the BREP is also the best method of tree-based learning algorithm for the subsets AT and V; AT, V, and RH; AT, V, AP, and RH.

Lastly, we carried out more learner combination experiments with the best performing N ($N = 2\ldots6$) predictors, namely the best performing algorithms of each category for the best subsets with one, two, three, and four parameters shown in Table 10. While the advantages of learner combination are manifold (avoiding local minima of learners, providing different views of the data, reducing estimation variance, etc.), the main criticism is the increased model complexity. As one of the core values of science is simplicity, usually referred to as Ockham's razor, we wish to seek a learner combiner method as simple and accurate as possible. We see that the learner combination, namely voting the best N learners, did not improve over the performance of Bagging REPTree except with AT feature alone. The best overall performance obtained with the full set of features (also using Bagging REPTree) did not improve by any combination of best N predictors as denoted in Table 11. In case the results have improved slightly due to learner combination, we

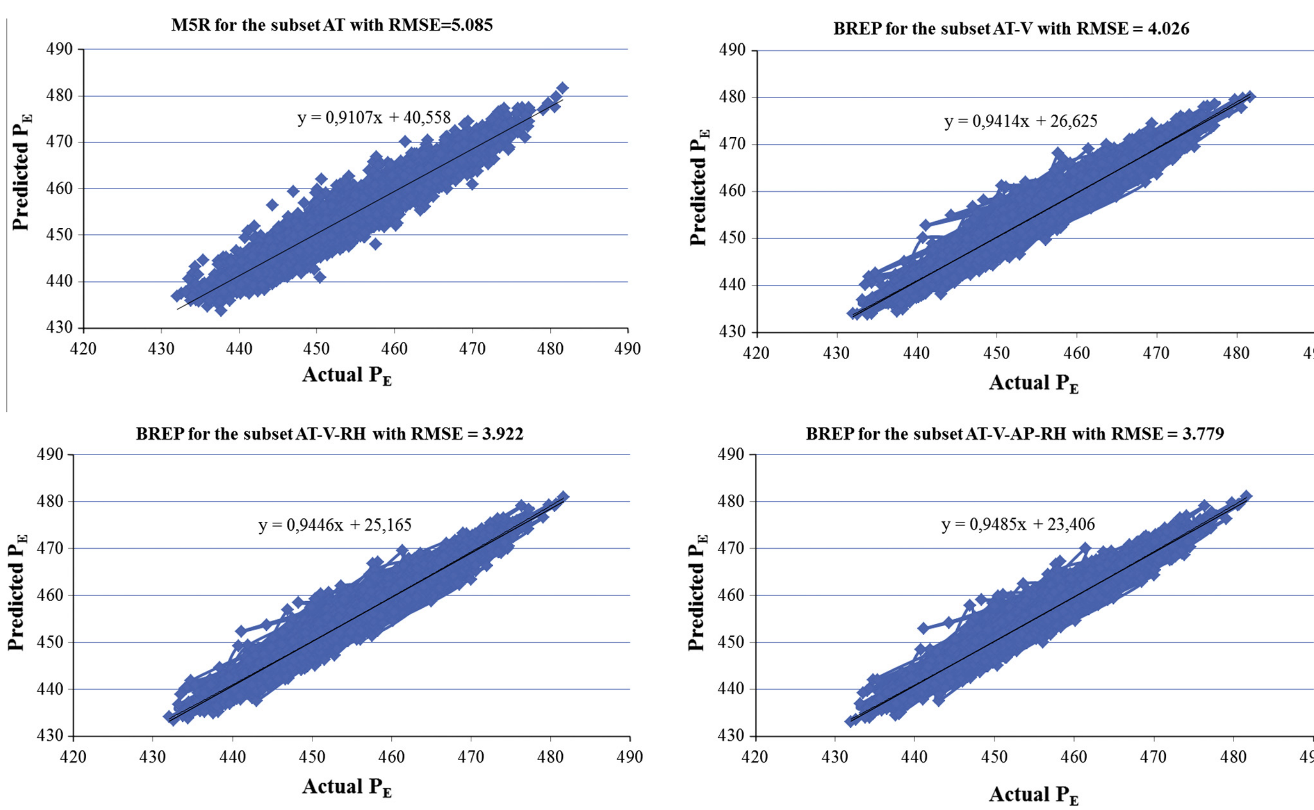

Fig. 12. Scatter plots of the actual and predicted P_E for the subsets with one, two, three, and four parameters.

Table 11
Comparing the results of the best single regression method with the results of voting the methods.

Categories	Regression methods	AT		AT–V		AT–V–RH		AT–V–AP–RH		Mean	
		MAE	RMSE	MAE	RMSE	MAE	RMSE	MAE	RMSE	MAE	RMSE
The best single method	BREP	4.074	5.208	3.035	4.026	2.952	3.934	2.818	3.787	3.220	4.239
Voting first 2 methods	BREP, M5P	4.000	5.111	3.147	4.121	3.027	3.971	2.911	3.848	3.271	4.263
Voting first 3 methods	BREP, M5P, M5R	3.986	5.091	3.213	4.185	3.085	4.023	2.972	3.904	3.314	4,301
Voting first 4 methods	BREP, M5P, M5R, REP	3.989	5.095	3.217	4.190	3.036	3.980	2.948	3.886	3.298	4.288
Voting first 5 methods	BREP, M5P, M5R, REP, K*	4.003	5.099	3.202	4.163	3.052	3.985	2.855	3.783	3.278	4,258
Voting first 6 methods	BREP, M5P, M5R, REP, K*, SMOReg	4.001	5.095	3.248	4.212	3.091	4.015	2.912	3.830	3.313	4,288
Voting first 7 methods	BREP, M5P, M5R, REP, K*, SMOReg, LMS	4.012	5.108	3.303	4.270	3.137	4.046	2.977	3.887	3.358	4.328

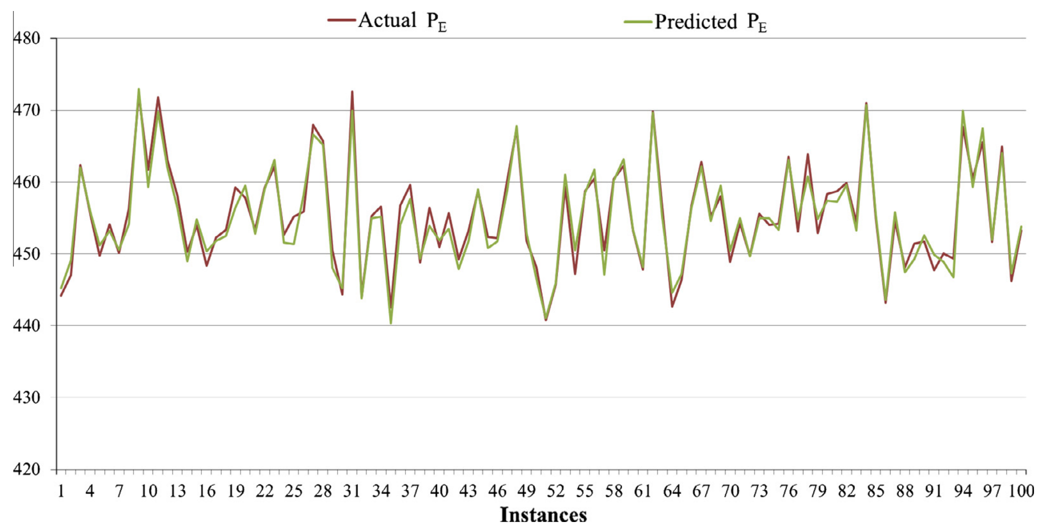

Fig. 13. Scatter plot of the actual P_E and predicted P_E with the BREP method for the best subset with AT, V, AP, and RH parameters.

would still recommend using a simpler model due to aforementioned rule of simplicity.

The scatter plot of the best predictive model of this study, which predicts full load electrical power output of a CCPP, is denoted in Fig. 13. This plot is the scatter plot of the actual P_E and predicted P_E with the best regression method, which is found as BREP method, for the best subset, which is found as the subset with four parameters such as AT, V, AP, and RH, with MAE of 2.818 and with RMSE of 3.787.

4. Discussion

The ultimate aim of machine learning study is to provide a generalizable algorithm to predict future, unseen data. From machine learning perspective, the answer to generalization capability is two folds: the generalization ability of the algorithm and the generalization ability of the trained model. First, we need to distinguish the model from the method or the algorithm: the methods consist of the algorithms as well as the abstractions learned by these algorithms from data. The methods optimize an objective function and learn the abstractions, which collectively and compactly form the model. Examples to methods are feed-forward neural networks that utilize error back-propagation with gradient descent to update the network parameters, namely weights. The model here is a set of weights connecting the input, hidden and output layers. In the case of decision trees, a hierarchical ordering of features with respective thresholds is obtained while the algorithm tries to reduce the error (in regression) or increase the information gain (in classification). Therefore the algorithm validity and the learned model's validity are different.

4.1. Validity of learned model

The generalization ability of the learned model can be estimated at lab environment by the performance on unseen test data. In our simulations, we carry out experiments in 5×2 cross validation scheme. In this scheme the data, which are collected over a 6 year-period, are split into a training set and a testing set. The training data is used by the algorithm to train a machine learning model (i.e. to learn model parameters) and the independent testing set is used to evaluate the model's performance. Also the model's hyper-parameters can be fine tuned using the validation set performance. However, in order to avoid over-fitting to data, which worsen the generalization power of the learner, we do not carry out fine tuning using validation set. Moreover, we carry out statistical tests to show the relative performance of learned models as well as the relevant algorithm on unseen test sets.

To assert the validity/generalization power of the model on new data, we first need to make sure that the data is drawn from the same underlying distribution. This means that the data should be generated from the same or similar physical data generation process. The proposed methods are applied in the power plant (whose identity is kept confidential) to predict next day's hourly power output with a high accuracy (less than 2% mean relative error). This is an "on-site, in-the-wild" confirmation of the proposed machine learning methods. If the model is to be used in a different power plant, first the collected data should match in terms of their statistics. In case of a totally different ambient condition set (such as a co-generation power plant in Sweden, where the ambient conditions differ dramatically with respect to Turkey), the model should be trained from the data collected from this plant. The validity/suitability of the algorithms will be discussed in the next subsection.

4.2. Validity of the training algorithm

As mentioned before, the data are collected over a long period, therefore are highly representative of the population. Similarly, the numbers of samples as opposed to the number of features are sufficient to learn a regression model, which can be subjected to statistical significance analysis. Without fine tuning the method's hyper-parameters (such as the number of hidden nodes in a neural network), the statistical tests provide a confident estimate of relative performance. We use 5×2 CV to obtain 10 simulations to measure the statistical significance. Based on the findings, we can argue that in a similar study even with different ambient conditions, the four features collectively are most likely to provide the best results. Similarly, the best performing algorithms are encouraged to be tested first.

5. Conclusion

This study presented an alternative solution model for a prediction of the electrical power output of a base load operated CCPP, when it was full load. Instead of thermodynamical approaches, which involve some assumptions with intractably many nonlinear equations of a real application of a system, machine learning approaches were preferred to use for accurate prediction. The analysis of a system by using thermodynamical approaches takes too much computational time and effort, and sometimes the result of this analysis might be unsatisfactory and unreliable due to many assumptions taken into account and nonlinear equations. In order to overcome this obstacle, the analysis of several machine learning regression methods for predicting output of a thermodynamic system, which is a CCPP with two gas turbines, one steam turbine and two heating systems, was presented as an alternative analysis.

There were two main purposes of this study. The first was to discover the best subset of our dataset among all other subset configurations in the experiments. For this purpose, we investigated which parameter or combinations of parameters were the most influential on the prediction of the target parameter. Secondly, we aimed to find out which machine learning regression method was the most successful in prediction of full load electrical power output.

In order to find out the most influent individual variables or combination of the variables, all possible subsets of the dataset, which include 15 different combinations of four variables such as AT, V, AP and RH, were applied to 15 different machine learning regression methods. As a result of the experiments, the subset, which consists of a complete set of parameters, was found to be the best subset of the dataset among all possible subsets yielding MAE of 2.818 and RMSE of 3.787 in prediction of electrical power output. Besides, the best accuracy was obtained by applying the subset with four parameters using Bagging method with REPTree predictor. Similarly, according to the average results of the comparative experiments, the most successful method, which might predict the full load electrical power output of a base load operated CCPP with the highest prediction accuracy, was found as Bagging REP Tree method, resulting in MAE of 3.220 and in RMSE of 4.239.

The CCPP, where the dataset is supplied for this study, has started to use this developed predictive model for next day's hourly energy output. As input the CCPP uses the next day's temperature forecast given by the state's meteorology institute. In future works, we plan to perfect the input to this predictive model by first predicting the next day's ambient variables more precisely, and also investigate prediction of electrical power output for different types of power plants.

Acknowledgments

Finally, I would like to thank Dr. Erdinç Uzun and Heysem Kaya, who is pursuing his PhD degree in Boğaziçi University, Department of Computer Engineering in Istanbul, for their sincere supports, and helps.

References

[1] Kesgin U, Heperkan H. Simulation of thermodynamic systems using soft computing techniques. Int J Energy Res 2005;29:581–611.
[2] Güvenir HA. Regression on feature projections. Knowl-Based Syst 2000;13:207–14.
[3] Kaya H, Tüfekci P, Gürgen FS. Local and global learning methods for predicting power of a combined gas & steam turbine. In: International conference on emerging trends in computer and electronics engineering (ICETCEE'2012), Dubai, March 24–25, 2012.
[4] Fast M, Assadi M, Deb S. Development and multi-utility of an ANN model for an industrial gas turbine. Appl Energy 2009;86(1):9–17.
[5] Rahnama M, Ghorbani H, Montazeri A. Nonlinear identification of a gas turbine system in transient operation mode using neural network. In: 4th Conference on thermal power plants (CTPP), IEEE Xplore; 2012.
[6] Refan MH, Taghavi SH, Afshar A. Identification of heavy duty gas turbine startup mode by neural networks. In: 4th Conference on thermal power plants (CTPP), IEEE Xplore; 2012.
[7] Yari M, Shoorehdeli MA. V94.2 gas turbine identification using neural network. In: First RSI/ISM international conference on robotics and mechatronics (ICRoM), IEEE Xplore, 2013.
[8] Kumar A, Srivastava A, Banerjee A, Goel A. Performance based anomaly detection analysis of a gas turbine engine by artificial neural network approach. In: Procee. annual conference of the prognostics and health management society; 2012.
[9] Tayarani-Bathaie SS, Sadough Vanini ZN, Khorasan K. Dynamic neural network-based fault diagnosis of gas turbine engines. Neurocomputing 2014;125(11):153–65.
[10] Tso GKF, Yau KKW. Predicting electricity energy consumption: a comparison of regression analysis, decision tree and neural networks. Energy 2007;32(9):1761–8.
[11] Azadeh A, Saberi M, Seraj O. An integrated fuzzy regression algorithm for energy consumption estimation with non-stationary data: a case study of Iran. Energy 2010;35(6):2351–66.
[12] Ekonomou L. Greek long-term energy consumption prediction using artificial neural networks. Energy 2010;35(2):512–7.
[13] Che J, Wang J, Wang G. An adaptive fuzzy combination model based on self-organizing map and support vector regression for electric load forecasting. Energy 2012;37(1):657–64.
[14] Kavaklioglu K. Modeling and prediction of Turkey's electricity consumption using support vector regression. Appl Energy 2011;88:368–75.
[15] Leung PCM, Lee EWM. Estimation of electrical power consumption in subway station design by intelligent approach. Appl Energy 2013;101:634–43.
[16] Alrashidi MR, El-Naggar KM. Long term electric load forecasting based on particle swarm optimization. Appl Energy 2010;87:320–6.
[17] Machine Learning Group at University of Waikato. <www.cs.waikato.ac.nz/ml/weka/> [accessed: 15.10.12].
[18] Niu LX. Multivariable generalized predictive scheme for gas turbine control in combined cycle power plant. In: IEEE conference on cybernetics and intelligent systems; 2009.
[19] Ramireddy V. An overview of combined cycle power plant. <http://electrical-engineering-portal.com/an-overview-of-combined-cycle-power-plant> [accessed: 02.03.13].
[20] Han J, Kamber M. Data mining: concepts and techniques. San Francisco: Morgan Kauffmann Publishers; 2001.
[21] Dash M, Liu H. Consistency-based search in feature selection. Artif Intell 2003;151:155–76.
[22] Betrie GD, Tesfamariam S, Morin KA, Sadiq R. Predicting copper concentrations in acid mine drainage: a comparative analysis of five machine learning techniques. Environ Monit Assess 2012. http://dx.doi.org/10.1007/s10661-012-2859-7.
[23] Ekinci S, Celebi UB, Bal M, Amasyalı MF, Boyacı UK. Predictions of oil/chemical tanker main design parameters using computational intelligence techniques. Appl Soft Comput 2011;11:2356–66.
[24] Witten IH, Frank E. Data mining: practical machine learning tools and techniques with java implementations. San Francisco: Morgan Kaufmann; 2005.
[25] Crouch C, Crouch D, Maclin R, Polumetla A. Automatic detection of RWIS sensor malfunctions (Phase I). Final report, published by Intelligent Transportation Systems Institute Center for Transportation Studies University of Minnesota, CTS 09-10, March 2009.

[26] Rousseeuw PJ. Least median of squares regression. J Am Stat Assoc 1984;79(388) [Theory and Methods Section].
[27] Haykin S. Neural networks, a comprehensive foundation. 2nd ed. New Jersey: Prentice Hall; 1999.
[28] Elish MO. A comparative study of fault density prediction in aspect-oriented systems using MLP, RBF, KNN, RT, DENFIS and SVR models. Artif Intell Rev 2012. http://dx.doi.org/10.1007/s10462-012-9348-9.
[29] Cleary JG, Trigg LE. K*: an instance-based learner using an entropic distance measure. In: 12th International conference on machine learning; 1995. p. 108–14.
[30] Friedman JH. Stochastic gradient boosting (tech. rep.). Palo Alto (CA): Stanford University, Statistics Department; 1999.
[31] Breiman L. Bagging predictors. Mach Learn 1996;24(2):123–40.
[32] D'Haen J, Van den Poel D. Temporary staffing services: a data mining perspective. In: 2012 IEEE 12th international conference on data mining workshops; 2012.
[33] Wang Y, Witten I. Inducing model trees for continuous classes. In: Ninth European conference on machine learning, Prague, Czech Republic; 1997. p. 128–37.
[34] Portnoy S, Koenker R. The Gaussian hare and the Laplacian tortoise: computability of squared-error versus absolute-error estimators. Stat Sci 1997;12(4):279–300.
[35] Arrieta FRP, Lora EES. Influence of ambient temperature on combined-cycle power-plant performance. Appl Energy 2005;80(3):261–72.
[36] De Sa A, Al Zubaidy S. Gas turbine performance at varying ambient temperature. Appl Therm Eng 2011;31(14–15):2735–9.
[37] Erdem HH, Sevilgen SH. Case study: effect of ambient temperature on the electricity production and fuel consumption of a simple cycle gas turbine in Turkey. Appl Therm Eng 2006;26(2–3):320–6.
[38] Lee JJ, Kang DW, Kim TS. Development of a gas turbine performance analysis program and its application. Energy 2011;36(8):5274–85.
[39] Improve steam turbine efficiency. <http://www.iffco.nic.in/applications/brihaspat.nsf/0/fddd5567e90ccfbde52569160021d1c8/$FILE/turbine.pdf> [accessed: January 2012].
[40] Challagulla VUB, Bastani FB, Yen IL, Paul RA. Empirical assessment of machine learning based software defect prediction techniques. In: 10th IEEE international workshop on object-oriented real-time dependable systems, Sedona, 2–4 February, 2005. p. 263–70. http://dx.doi.org/10.1109/WORDS.2005.32.
[41] Liu H, Gopalkrishnan V, Quynh KTN, Ng W. Regression models for estimating product life cycle cost. J Intell Manuf 2009;20:401–8. http://dx.doi.org/10.1007/s10845-008-0114-4.
[42] Wallet BC, Marchette DJM, Solka JL, Wegman EJ. A genetic algorithm for best subset selection in linear regression. Comput Sci Stat 1997;28:545–50.
[43] Dietterich TG. Approximate statistical tests for comparing supervised classification learning algorithms. Neural Comput 1998;10:1895–923.
[44] Alpaydın E. Introduction to machine learning. 2nd ed. MIT Press; 2010.
[45] Montgomery DC. Design and analysis of experiments. 6th ed. John Wiley & Sons; 2007.