

Research Article

The Effectiveness of Feature Selection Method in Solar Power Prediction

Md Rahat Hossain, Amanullah Maung Than Oo, and A. B. M. Shawkat Ali

Power Engineering Research Group (PERG), Central Queensland University, Rockhampton, QLD 4701, Australia

Correspondence should be addressed to Md Rahat Hossain; m.hossain@cqu.edu.au

Received 13 December 2012; Accepted 11 July 2013

Academic Editor: Czarena Crofcheck

Copyright © 2013 Md Rahat Hossain et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

This paper empirically shows that the effect of applying selected feature subsets on machine learning techniques significantly improves the accuracy for solar power prediction. Experiments are performed using five well-known wrapper feature selection methods to obtain the solar power prediction accuracy of machine learning techniques with selected feature subsets. For all the experiments, the machine learning techniques, namely, least median square (LMS), multilayer perceptron (MLP), and support vector machine (SVM), are used. Afterwards, these results are compared with the solar power prediction accuracy of those same machine learning techniques (i.e., LMS, MLP, and SVM) but without applying feature selection methods (WAFS). Experiments are carried out using reliable and real life historical meteorological data. The comparison between the results clearly shows that LMS, MLP, and SVM provide better prediction accuracy (i.e., reduced MAE and MASE) with selected feature subsets than without selected feature subsets. Experimental results of this paper facilitate to make a concrete verdict that providing more attention and effort towards the feature subset selection aspect (e.g., selected feature subsets on prediction accuracy which is investigated in this paper) can significantly contribute to improve the accuracy of solar power prediction.

1. Introduction

Feature selection can be considered one of the main pre-processing steps of machine learning [1]. Feature selection is different from feature extraction (or feature transformation), which creates new features by combining the original features [2]. The advantages of feature selection are manifold. First, feature selection significantly saves the operating time of a learning procedure by eliminating irrelevant and redundant features. Second, without the intervention of irrelevant, redundant, and noisy features, learning algorithms can centrally point on most essential features of data and build simpler but more precise data models. Third, feature selection can help build a simpler and more common model and get a better insight into the fundamental perception of the task [3–5]. The feature selection aspect is fairly significant for the reason that with the same training data, it may happen that an individual regression algorithm can perform better with different feature subsets. The success of machine learning on a particular task is affected by

many factors. Among those factors first and foremost is the representation and quality of the instance data [6]. The training stage becomes critical with the existence of noisy, irrelevant, and redundant data. Sometimes, the real life data contain too much information; among those, very little is useful for desired purpose. Therefore, it is not important to include every piece of information from the raw data source for modelling.

Usually, features are differentiated [7] as

- (1) relevant: this class of features has strong impact on the output;
- (2) irrelevant: opposite to relevant features, irrelevant features do not have any bias on the output;
- (3) redundant: a redundancy occurs when a feature captures the functionality of other;

All the algorithms to perform feature selection consist of two common aspects. One is the search method which is actually a selection algorithm to generate designed feature subsets

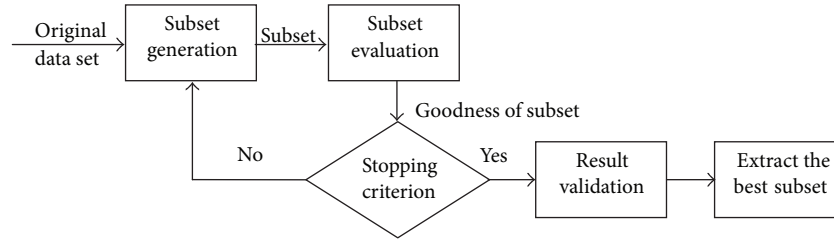


FIGURE 1: Key sequences of feature selection.

and attempts to reach the most advantageous ones. Another aspect is called evaluator which is basically an evaluation algorithm to make a decision about the goodness of the planned feature subset and finally returns the assessment about righteousness of the search method [8]. On the other hand, lacking an appropriate stopping condition, the feature selection procedure could run exhaustively or everlastingly all the way throughout the raw dataset. It may be discontinued whenever any attribute is inserted or deleted but ultimately not producing a better subset or whenever a subset is produced which provides the maximum benefits according to some assessing functions. A feature selector may stop manipulating features when the merit of a current feature subset stops improving or conversely does not degrade. A usual feature selection procedure revealed in Figure 1 consists of four fundamental steps: (A) subset generation; (B) subset evaluation; (C) stopping criterion; and (D) result validation [9]. The procedure begins with subset generation that utilizes a definite search strategy to generate candidate feature subsets. Then, each candidate subset is evaluated according to a definite evaluation condition and compared with the previous best one. If it is better, it replaces the previous best. The process of subset generation and evaluation is repetitive until a given stopping condition is satisfied. Finally, the selected best feature subset is validated by some test data. Figure 1 graphically demonstrates the previous mentioned steps and procedures of feature selection process.

Based on some evaluation functions and calculations, feature selection methods find out the best feature from different candidate subsets. Usually, feature selection methods are classified into two general groups (i.e., filter and wrapper) [10]. Inductive algorithms are used by wrapper methods as the evaluation function whereas filter methods are independent of the inductive algorithm. Wrapper methods work along wrapping the feature selection in conjunction with the induction algorithm to be used, and to accomplish this wrapper methods use cross validation. With the same training data, it may happen that individual regression algorithm can perform better with different feature subsets [11]. Widely used wrapper selection methods are briefly discussed in the following section. Section 3 deals with real life data collection and analysis of the dataset. Sections 4 and 5 show the experimental results using machine learning techniques and selected feature subsets. Comparison of the results and graphical presentation of those results are presented in those two sections. Section 6 demonstrates the performance of those obtained prediction results by paired t -tests. The results

from the experiments demonstrate that LMS, MLP, and SVM supplied with selected feature subsets provide better prediction accuracy (i.e., reduced MAE and MASE) than when they are without the selected feature subsets. Concluding remarks are provided in final section of this paper.

2. Wrapper Methods of Feature Selection

In the field of machine learning, the feature subset selection that is also named as attribute subset selection, variable selection, or variable subset selection is an important method that helps to select an appropriate subset of significant features for model development. In machine learning based advanced and sophisticated applications (e.g., solar power prediction), feature selection methods have become an obvious need. Generally, feature selection methods are categorised into three different classes [3, 12]: *filter selection methods* select feature subsets as a preprocessing act, autonomously of the selected predictor; *wrapper selection methods* exploit the predictive power of the machine learning technique to select suitable feature subsets; and to conclude, *embedded selection methods* usually select feature subsets in the course of training. Wrapper methods were used for the experiments in this paper. Therefore, in the subsequent section, widely used and accepted wrapper methods are placed.

The wrapper methods use the performance (e.g., regression, classification, or prediction accuracy) of an induction algorithm for feature subset evaluation. Figure 2 shows the ideas behind wrapper approaches [3]. For each generated feature subset S , wrappers evaluate its goodness by applying the induction algorithm to the dataset using features in subset S . Wrappers can find feature subsets with high accuracy because the features match well with the learning algorithms.

The easiest method among all the wrapper selection algorithms is the forward selection (FS). This method starts the procedure without having any feature in the feature subset and follows a greedy approach so that it can sequentially add features until no possible single feature addition results in a higher valuation of the induction function. Backward elimination (BE) begins with the complete feature set and gradually removes features as long as the valuation does not degrade. Description about forward selection (FS) and backward selection (BS) can be found in [13] where the authors proved that wrapper selection methods are better than methods having no selection.

Starting with an empty set of features, the best first search (BFS) produces every possible individual feature extension

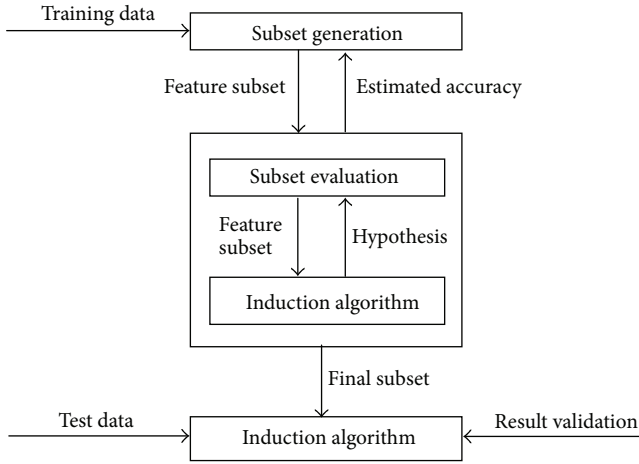


FIGURE 2: The wrapper approach for feature selection [3].

[3]. BFS exploits the greedy hill climbing approach in conjunction with backtracking to search the space of feature subsets. BFS has all the flexibility to start with an empty subset and search in forward direction. Alternatively, it can start having full set of attributes and search in backward direction or it can start randomly from any point and move towards any direction. Extension of the BFS is the linear forward selection (LFS). A limited number of attributes k are taken into consideration by LFS. This method either selects the top k attributes by initial ordering or it can carry put a ranking [14, 15].

Subset size forward selection (SSFS) is the extension of LFS. SSFS carries out an internal cross validation. An LFS is executed on every fold to find out the best possible subset size [15, 16]. Through the individual evaluations, attributes are ranked by the ranker search. It uses this search in combination with attribute evaluators [16].

GA performs a search using the simple genetic algorithm described in Goldberg's study [17]. Genetic algorithms are random search techniques based on the principles of natural selection [17]. They utilize a population of competing solutions evolved to an optimal solution. For feature selection, a solution is a fixed length binary sequence representing a feature subset. The value of each position—typically 1 or 0—in the sequence represents the presence or absence of a particular feature, respectively. The algorithm proceeds in an iterative manner where each successive generation is produced by applying genetic operators to the members of the current generation. Nonetheless, GAs naturally involves a huge quantity of evaluations or iterations to achieve optimal solution. Other than all these conventional methods, we have experimentally verified an unconventional approach. In this method, we calculated the correlation coefficient for each (except the target attribute) of the competing attributes with respect to the target attribute of the used dataset. For this purpose, we used Pearson's correlation coefficient formula which is described in the next section. After the attribute wise calculation, we selected those attributes whose correlation coefficient values are positive only as feature subset. The attributes having negative correlation coefficient are ignored

for this case. We named this method positive correlation coefficient selection (PCCS) [18].

3. Real Life Data Collection and Analysis

One of the key conditions to successfully perform the experiments of this paper is to collect recent, reliable, accurate, and long-term historical weather data of the particular location. However, finding accurate multiyear data near the experiment site has always proved to be challenging because these data are not readily available due to the cost and difficulty in measurement [19]. There are only few sites in Australia providing research-quality data of solar radiation; so, these data are generally not available. Rockhampton, a subtropical town in North Australia, was chosen for the experiments of this research. The selected station is "Rockhampton Aero," having latitude of -23.38 and longitude of 150.48 . According to the Renewable Energy Certificates (RECs) zones within Australia, Rockhampton is identified within the most important zone [20]. The recent data were collected from Commonwealth Scientific and Industrial Research Organization (CSIRO), Australia.

Data were also collected from the Australian Bureau of Meteorology (BOM), the National Aeronautics and Space Administration (NASA), and the National Oceanic and Atmospheric Administration (NOAA). All of this mission-critical information on solar radiation was being estimated from cloud cover and humidity at airports. Free data are available from National Renewable Energy Laboratory (NREL) and NASA. These are excellent for multiyear averages but perform poorly for hourly and daily measurements. After analyzing the raw data collected from different sources, the data provided by CSIRO were finally selected. The data used in this paper are based on hourly global solar irradiance ground measurements which are a significant aspect of the dataset. These data were gathered for a period of five years from 2006 to 2010.

The attributes in the used dataset are average air temperature, average wind speed, current wind direction, average relative humidity, total rainfall, wind speed, wind direction, maximum peak wind gust, current evaporation, average absolute barometer, and average solar radiation. Table 1 represents the statistical properties of the raw data.

4. Applying Feature Selection Techniques on the Dataset

All the research works related to solar radiation prediction select the input features or attributes randomly. Unlike the conventional way, this research experimented with the maximum number of features and found out the best possible combination of features for the individual learning models of the hybrid model. To perform the experiments for selecting significant feature subsets for individual machine learning technique, the traditional BFS, LFS, SSFS, ranker search, GS, and our very own PCCS selection methods are used. To

TABLE 1: Statistical description of the raw data set.

	Min.	Max.	Mean	Std. dev
Avg. air temp. (DegC)	-5.8	40.1	20.47	6.99
Avg. wind speed (Km/h)	0	27.1	6.99	4.78
Current wind dir. (Deg)	0	359	158.91	103.66
Avg. relative humidity (%)	0	100	55.11	24.26
Total rainfall (mm)	0	30.4	0.07	0.69
Wind speed (Km/h)	0	24.83	5.77	4.38
Wind direction (Deg)	0	360	169.91	109.84
Max. peak wind gust (Km/h)	0	106	20.45	11.33
Current evaporation (mm)	-1.36	1.36	0.31	0.28
Avg. abs. barometer (hPa)	921	1020	966.59	12.09
Solar radiation (W/m ²)	1	1660	300.75	325.17

carry out experiments, three algorithms for machine learning technique, namely, least median square [21], multilayer perceptrons [22], and support vector machine [23], are used.

Evaluating the degree of fitness, that is, how well a regression model fits to a dataset, is usually obtained by correlation coefficient. Assuming the actual values as a_1, a_2, \dots, a_n and the predicted values as p_1, p_2, \dots, p_n , the correlation coefficient is known by the equation:

$$R = \frac{S_{PA}}{\sqrt{S_P S_A}}, \quad (1)$$

where,

$$S_{PA} = \frac{\sum_i (p_i - \bar{p})(a_i - \bar{a})}{n-1}, \quad (2)$$

$$S_P = \frac{\sum_i (p_i - \bar{p})^2}{n-1}, \quad S_A = \frac{\sum_i (a_i - \bar{a})^2}{n-1}.$$

To find out the correlation coefficient of the model, the full training set is partitioned into ten mutually exclusive and same-sized subsets. The performance of the subset depends on the accuracy of predicting test values. For every individual algorithm, this cross validation method was run over ten times, and finally, the average value for 10 cross validations was calculated. In k -cv, a dataset S_n is uniformly partitioned into k folds of similar size $P = \{P_1, \dots, P_k\}$. For the sake of clarity and without loss of generality; it is supposed that n is multiple of k . Let $T_i = S_n/P_i$ be the complement dataset of P_i . Then, the algorithm $A(\cdot)$ induces a classifier from T_i , $\psi_i = A(T_i)$ and estimates its prediction error with P_i . The k -cv prediction error estimator of $\psi = A(S_n)$ is defined as follows [24]:

$$\xi_k(S_n, P) = \frac{1}{n} \sum_{i=1}^k \sum_{(x,c) \in P_i} 1(c, \psi_i(x)), \quad (3)$$

where $1(i, j) = 1$ if and only if $i \neq j$ and equal to zero otherwise. So, the k -cv error estimator is the average of the errors made by the classifiers ψ_i in their respective divisions P_i .

According to Zheng and Kusiak in [24], the mean absolute error (MAE) and mean absolute percent error (MAPE) are used to measure the prediction performance; we have also used these evaluation metrics for our experiments. The definitions are expressed as

$$\text{MAE} = \frac{\sum_{i=1}^n |p_i - a_i|}{n}, \quad (4)$$

$$\text{MAPE} = \frac{(\sum |PE|)}{n},$$

where $PE = (E/a) * 100$, $E = (a - p)$, a = actual values, p = predicted values, and n = number of occurrences.

Error of the experimental results was also analyzed according to mean absolute scaled error (MASE) [25]. MASE is scale free, less sensitive to outlier; its interpretation is very easy in comparison to other methods and less variable to small samples. MASE is suitable for uncertain demand series as it never produces infinite or undefined results. It indicates that the prediction with the smallest MASE would be counted the most accurate among all other alternatives [25]. Equation (5) states the formula to calculate MASE as

$$\text{MASE} = \frac{\text{MAE}}{(1/n - 1) \sum_{i=2}^n |a_i - a_{i-1}|}, \quad (5)$$

where,

$$\text{MAE} = \frac{1}{n} \sum_{i=1}^n |p_i - a_i|. \quad (6)$$

4.1. Prediction of the Machine Learning Techniques Using the Selected Feature Subsets. Various feature subsets were generated or selected using different wrapper feature selection methods. Afterwards, six-hours-ahead solar radiation prediction by the selected machine learning techniques, namely, LMS, MLP, and SVM, was performed. For this instance, the selected feature subsets were supplied to the individual machine learning techniques. The intention of this experiment was to observe whether this initiative produces any improvement in the error reduction of those selected machine learning techniques or not. For these experiments, any tuning of the particular algorithms to a definite dataset was avoided. For all the experiments, default values of learning parameters were used. In general, in the following tables, one can see the CC, MAE, MAPE, and MASE of six-hours-ahead prediction for each machine learning technique supplied with different feature subsets. For all the experiments, “W” is used to indicate that a particular machine learning technique supplied with the selected feature subsets statistically outplays the one without applying feature selection (WAFS) methods.

Tables 2 and 3 represent the obtained CC and MAE for applying LMS, MLP, and SVM machine learning technique for six hours in advance prediction on the used dataset before and after feature selection process.

In Tables 4 and 5, the MAPE and MASE are shown before and after feature selection processes are applied to LMS, MLP, and SVM machine learning technique for the same purpose.

TABLE 2: Achieved CC after applying various wrapper selection methods on LMS, MLP, and SVM.

	WAFS	BFS	LFS	SSFS	Ranker		GS	PCCS	
LMS	0.95	0.96	0.96	0.96	0.96		0.96	0.97	W
MLP	0.98	0.97	0.97	0.98	0.99	W	0.97	0.98	
SVM	0.96	0.96	0.96	0.96	0.97	W	0.96	0.96	

TABLE 3: Achieved MAE after applying various wrapper selection methods on LMS, MLP, and SVM.

	WAFS	BFS	LFS	SSFS	Ranker		GS	PCCS	
LMS	77.19	76.81	74.49	74.12	74.93		87.53	73.37	W
MLP	91.02	168.34	222.73	119.11	84.31	W	288.83	110.57	
SVM	126.88	123.46	129.51	123.42	122.12	W	125.59	124.52	

TABLE 4: Achieved MAPE after applying various wrapper selection methods on LMS, MLP, and SVM.

	WAFS	BFS	LFS	SSFS	Ranker		GS	PCCS	
LMS	17.65	19.53	17.08	17.04	16.93		17.49	16.82	W
MLP	20.17	50.53	41.83	23.46	17.87	W	32.83	21.50	
SVM	21.72	21.53	22.35	21.35	20.88	W	21.35	21.65	

TABLE 5: Achieved MASE after applying various wrapper selection methods on LMS, MLP, and SVM.

	WAFS	BFS	LFS	SSFS	Ranker		GS	PCCS	
LMS	0.63	0.71	0.61	0.60	0.61		0.62	0.59	W
MLP	0.74	2.35	1.81	0.97	0.68	W	1.37	0.90	
SVM	1.03	1.02	1.05	1.00	0.99	W	1.00	1.01	

The results from the experimental results show that the PCCS is somewhat a superior feature selection method for LMS algorithm considering all the instances. It is noticeable that all the feature selection methods contributed to improve the CC of LMS algorithm. However, in the case of MAE, all the selection algorithms except the GS improve the results for LMS. In both the case of MAPE and MASE, BFS is the only selection method which does not improve the results for LMS. It is found from those results that the ranker search is to some extents superior feature selection method for MLP algorithm. It is noticeable that all the feature selection methods present a nearly close CC for MLP algorithm but in the case of MAE, MAPE, and MASE, ranker search is the only selection method which improves the results. Finally, the obtained results illustrate that again the ranker search is to some extent a superior feature selection method for SVM. It is also noticeable that all the feature selection methods present either nearly close or equal CC for SVM. However, in the case of MAE, MAPE, and MASE, LFS is the only one which is unable to improve the results for SVM.

5. Prediction Results: Before versus after Applying the Feature Selection Techniques

In Table 6, the prediction errors (MAE and MASE) of the individual machine learning techniques are compared on the basis of before supplying selected feature subsets and after supplying selected feature subsets on them. The comparative results show that errors are reduced for all the instances after supplying selected feature subsets. The terms MAE_BEFORE

and MASE_BEFORE represent the results for having no selected feature subsets for MAE and MASE, respectively, whereas the terms MAE_AFTER and MASE_AFTER represent the results having selected feature subsets for MAE and MASE, respectively.

Figure 3(a) graphically demonstrates the prediction accuracy comparison of LMS, MLP, and SVM before and after applying selected feature subsets in terms of MAE. Figure 3(b) graphically demonstrates the prediction accuracy comparison of LMS, MLP, and SVM before and after applying selected feature subsets in terms of MASE.

Figure 4(a) graphically illustrates the significance of applying and without applying (WAFS) various feature selection methods on LMS in terms of CC. Figure 4(b) graphically illustrates the significance of applying and without applying (WAFS) feature selection methods on MLP in terms of MAE. In both the cases, results are improved after applying selected feature subsets.

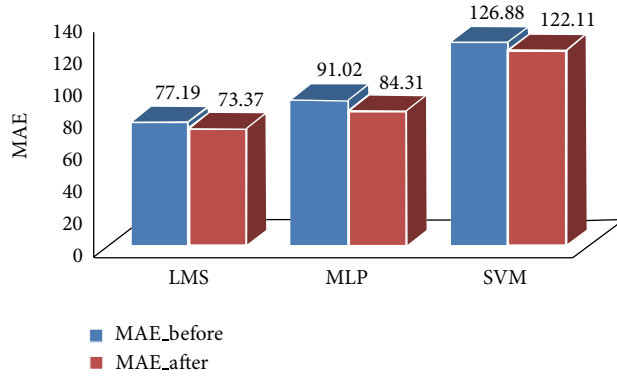
Figure 5 graphically illustrates the significance of applying and without applying (WAFS) feature selection methods on SVM in terms of MAPE. The graphical representation clearly shows that the result is improved in terms of MAPE after applying selected feature subsets.

6. Performance Analysis through Statistical Tests

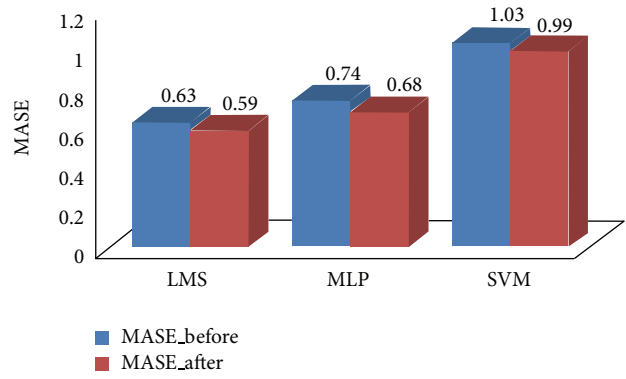
A statistical test provides a mechanism for making quantitative decisions about a process or processes. The intention is to determine whether there is enough evidence to "reject"

TABLE 6: Error measurements of the top most three decisive regression algorithms' prediction accuracy with feature selection.

	MAE_before	MAE_after	MASE_before	MASE_after	Rank
LMS	77.19	73.37	0.63	0.59	1
MLP	91.02	84.31	0.74	0.68	2
SVM	126.88	122.11	1.03	0.99	3

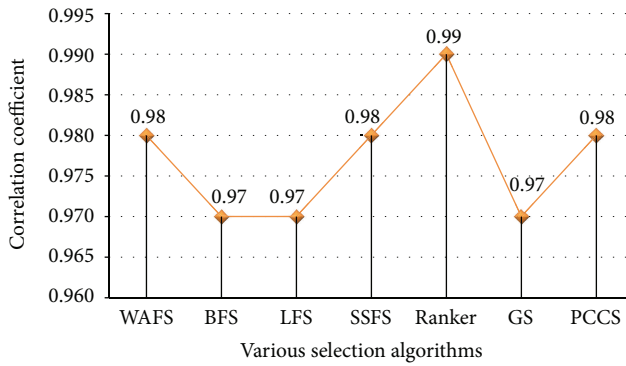


(a)

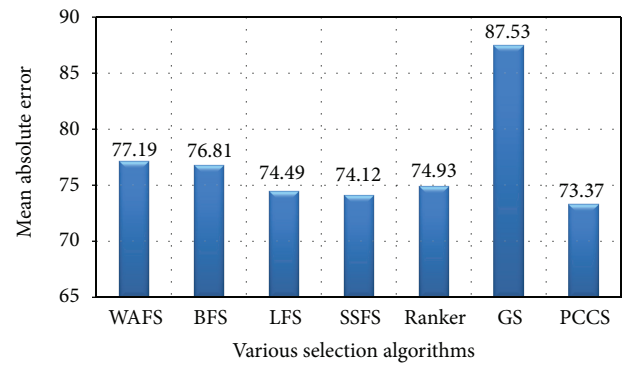


(b)

FIGURE 3: MAE (a) and MASE (b) comparison of LMS, MLP, and SVM before and after feature selection process.



(a)



(b)

FIGURE 4: Effects of applying and without applying various feature selection algorithms on MLP (a) and LMS (b).

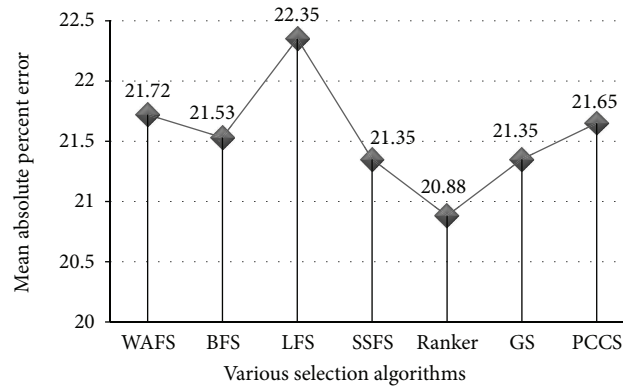


FIGURE 5: Effects of applying and without applying various feature selection algorithms on SVM.

TABLE 7: Paired samples statistics.

	Mean	<i>N</i>	Std. deviation	Std. error mean
Pair 1				
Actual data	639.17	6	177.41	72.43
LMS prediction	712.54	6	107.25	43.78
Pair 2				
Actual data	639.17	6	177.41	72.43
MLP prediction	584.77	6	263.11	107.41
Pair 3				
Actual data	639.17	6	177.41	72.43
SVM prediction with RBF kernel	504.13	6	209.35	85.47
Pair 4				
Actual data	639.17	6	177.41	72.43
SVM prediction with poly kernel	517.05	6	183.19	74.79

TABLE 8: Paired samples correlations.

	<i>N</i>	Correlation	Sig.
Pair 1			
Actual data & LMS prediction	6	0.967	0.002
Pair 2			
Actual data & MLP prediction	6	0.985	0.000
Pair 3			
Actual data & SVM prediction with RBF kernel	6	0.961	0.002
Pair 4			
Actual data & SVM prediction with poly kernel	6	0.962	0.002

an inference or hypothesis about the process. For a single group, the means of two variables are usually compared by the paired-samples t -test. It considers the differences between the values of two variables in a case-by-case basis and examines whether the average varies from 0 or not. It carries out various types of outputs such as statistical description for the test variables, the correlation among the test variables, meaningful statistics for the paired differences, the t -test itself, and a 95% confidence interval. The paired-samples t -tests are carried out in order to justify whether any significant difference exists between the actual and predicted results achieved by the selected three machine learning techniques for ensemble method. The t -test is executed with the SPSS package—PASW Statistics 20 [26]. In this paper, the paired t -test is employed to verify the performance of the machine learning techniques used in the experiments. Here, the null hypothesis and the alternative hypothesis are termed as H_0 and H_a , respectively, where H_0 means there is no significant difference between the actual and predicted mean values, H_a means there is significant difference between the actual and predicted mean values.

In Tables 7, 8, and 9, the paired samples statistics, the paired sample correlations, and paired samples tests for the actual and predicted values of LMS, MLP, and SVM are represented, respectively. Observing Table 9, we find that for pair 1, $t(5) = -2.28$, $P > 0.0001$; for pair 2, $t(5) = 1.43$, $P > 0.0001$; for pair 3, $t(5) = 5.28$, $P > 0.0001$; and for pair 4, $t(5) = 5.99$, $P > 0.0001$. Due to the means of the actual and

predicted values of each pair and the direction of the t values, it can be concluded that there was no statistically significant difference between actual and predicted values for all the cases. Therefore, the test failed to reject the null hypothesis H_0 .

7. Conclusions

Feature selection is a fundamental issue in both the regression and classification problems especially for the dataset having a very high volume of data. Applying feature selection methods on machine learning techniques may significantly contribute to increase performance in terms of accuracy. In this paper, various methods of feature selection methods have been briefly described. In particular, the wrapper feature selection methods are found better, which is also justified by the results obtained from the experiments performed in this paper. From the experiments performed in this paper, it is found that for LMS, the MAE before and after applying selected feature subsets is 77.19 and 73.37, respectively, and MASE 0.63 and 0.59, respectively. In the case of MLP, the MAE before and after applying selected feature subsets is 91.02 and 84.31, respectively, and MASE 0.74 and 0.68, respectively. For SVM, the MAE before and after applying selected feature subsets is 126.88 and 122.11, respectively, and MASE 1.03 and 0.99, respectively. The comparison between the results clearly shows that LMS, MLP, and SVM provide better prediction accuracy (i.e., reduced MAE and MASE) with selected feature

TABLE 9: Paired samples test.

		Paired differences				<i>t</i>	df	Sig. (2-tailed)	
	Mean	Std. deviation	Std. error mean	95% confidence interval of the difference					
				Lower	Upper				
Pair 1									
Actual data - LMS prediction	−73.37	78.62	32.09	−155.88	9.14	−2.28	5	0.071	
Pair 2									
Actual data - MLP prediction	54.39	93.47	38.16	−43.69	152.48	1.43	5	0.213	
Pair 3									
Actual data - SVM prediction with RBF kernel	135.04	62.63	25.57	69.31	200.76	5.28	5	0.003	
Pair 4									
Actual data - SVM prediction with poly kernel	122.11	49.95	20.39	69.69	174.53	5.99	5	0.002	

subsets than without selected feature subsets. Experimental results of this paper facilitate to make a concrete verdict that providing more attention and effort towards the feature subset selection aspect (e.g., selected feature subsets on prediction accuracy which is investigated in this paper) can significantly contribute to improve the accuracy of solar power prediction.

It is mentionable that for these experiments, the machine learning techniques were applied with the default learning parameter settings. In the near future, the new experiments will be performed with the intention to achieve better prediction accuracy of the selected machine learning techniques by applying both the optimized or tuned learning parameter settings and selected feature subsets on them.

Nomenclature

a_i :	Actual values
H_a :	Alternative hypothesis
C :	Class attribute; $C \in \{1, \dots, r\}$
$T_i = S_n/P_i$:	Complement dataset of P_i
cv:	Cross validation
P_i :	Divisions
k -cv:	Error estimator
$A(\cdot)$:	Induction algorithm
M :	Mean
H_0 :	Null hypothesis
k :	Number of folds
N :	Number of instances
P :	Set of n points
S :	Standard deviation
S_n :	Training set
P :	Two-tailed significance value
r_i :	Unknown errors or residuals
x :	Unlabeled instance.

Greek Symbols

Ψ_i :	Classifiers
ξ_k :	Randomized error estimator
α :	Significance level = 0.05.

Abbreviations

BOM:	Bureau of meteorology
BE:	Backward elimination
BS:	Backward selection
BFS:	Best first search
CSIRO:	Commonwealth Scientific and Industrial Research Organization
CC:	Correlation coefficient
Deg:	Degree
DegC:	Degree Celsius
FCBF:	Fast correlation-based feature selection
FS:	Forward selection
GA:	Genetic algorithm
GS:	Genetic search
hPa:	Hectopascals
Km/h:	Kilometer per hour
LFS:	Linear forward selection
LMS:	Least median square
MAE:	Mean absolute error
MASE:	Mean absolute scaled error
MAPE:	Mean absolute percent error
MLP:	Multilayer perceptron
PASW:	Predictive analytic software
RECs:	Renewable energy certificates
mm:	Millimeter
NASA:	National Aeronautics & Space Admin
NOAA:	National Oceanic & Atmospheric Admin
NREL:	National Renewable Energy Laboratory

PCCS: Positive correlation coefficient selection
 SPSS: Statistical Package for Social Sciences
 SSFS: Subset size forward selection
 SVM: Support vector machine
 WAFS: Without applying feature selection
 W/m²: Watt per meter square.

References

- [1] L. Yu and H. Liu, "Feature selection for high-dimensional data: a fast correlation based filter solution," in *Proceedings of the 20th International Conference on Machine Learning*, pp. 856–863, August 2003.
- [2] F. Tan, *Improving feature selection techniques for machine learning [Ph.D. thesis]*, 2007, Computer Science Dissertations: Paper 27, http://digitalarchive.gsu.edu/cs_diss/27.
- [3] R. Kohavi and G. H. John, "Wrappers for feature subset selection," *Artificial Intelligence*, vol. 97, no. 1-2, pp. 273–324, 1997.
- [4] D. Koller and M. Sahami, "Toward optimal feature selection," in *Proceedings of the 13th International Conference on Machine Learning*, pp. 284–292, 1996.
- [5] M. Dash and H. Liu, "Feature selection for classification," *Intelligent Data Analysis*, vol. 1, no. 3, pp. 131–156, 1997.
- [6] T. Mitchell, *Machine Learning*, McGraw Hill, 1997.
- [7] D. A. Bell and H. Wang, "A formalism for relevance and its application in feature subset selection," *Machine Learning*, vol. 41, no. 2, pp. 175–195, 2000.
- [8] M. Karagiannopoulos, D. Anyfantis, S. B. Kotsiantis, and P. E. Pintelas, "Feature selection for regression problems," in *Proceedings of the 8th Hellenic European Research on Computer Mathematics & Its Applications (HERCMA '07)*, pp. 20–22, 2007.
- [9] H. Liu and L. Yu, "Toward integrating feature selection algorithms for classification and clustering," *IEEE Transactions on Knowledge and Data Engineering*, vol. 17, no. 4, pp. 491–502, 2005.
- [10] P. Langley, "Selection of relevant features in machine learning," in *Proceedings of the AAAI Fall Symposium on Relevance*, pp. 1–5, 1994.
- [11] A. L. Blum and P. Langley, "Selection of relevant features and examples in machine learning," *Artificial Intelligence*, vol. 97, no. 1-2, pp. 245–271, 1997.
- [12] I. Guyon and A. Elisseeff, "An introduction to variable and feature selection," *Journal of Machine Learning Research*, vol. 3, pp. 1157–1182, 2003.
- [13] R. Caruana and D. Freitag, "Greedy attribute selection," in *Proceedings of the 11th International Conference on Machine Learning*, San Francisco, Calif, USA, 1994.
- [14] M. Gütlein, E. Frank, M. Hall, and A. Karwath, "Large-scale attribute selection using wrappers," in *Proceedings of the IEEE Symposium on Computational Intelligence and Data Mining (CIDM '09)*, pp. 332–339, April 2009.
- [15] M. Gutlein, *Large scale attribute selection using wrappers [Diploma of Computer Science thesis]*, Albert-Ludwigs-Universität-Freiburg, 2006.
- [16] M. Hall, E. Frank, G. Holmes, B. Pfahringer, P. Reutemann, and H. Witten, "The WEKA data mining software: an update," *ACM SIGKDD Explorations Newsletter*, vol. 11, no. 1, pp. 10–18, 2009.
- [17] E. Goldberg, *Genetic Algorithms in Search, Optimization and Machine Learning*, Addison-Wesley, 1989.
- [18] M. R. Hossain, A. M. T. Oo, and A. B. M. S. Ali, "The combined effect of applying feature selection and parameter optimization on machine learning techniques for solar Power prediction," *American Journal of Energy Research*, vol. 1, no. 1, pp. 7–16, 2013.
- [19] S. Coppolino, "A new correlation between clearness index and relative sunshine," *Renewable Energy*, vol. 4, no. 4, pp. 417–423, 1994.
- [20] <http://www.solarguys.com.au/>.
- [21] P. J. Rousseeuw, "Least median of squares regression," *Journal of the American Statistical Association*, vol. 79, no. 388, pp. 871–880, 1984.
- [22] S. Haykin, *Neural Networks: A Comprehensive Foundation*, Prentice Hall, 1999.
- [23] S. K. Shevade, S. S. Keerthi, C. Bhattacharyya, and K. R. K. Murthy, "Improvements to the SMO algorithm for SVM regression," *IEEE Transactions on Neural Networks*, vol. 11, no. 5, pp. 1188–1193, 2000.
- [24] H. Zheng and A. Kusiak, "Prediction of wind farm power ramp rates: a data-mining approach," *Journal of Solar Energy Engineering*, vol. 131, no. 3, Article ID 031011, 8 pages, 2009.
- [25] R. J. Hyndman and A. B. Koehler, "Another look at measures of forecast accuracy," *International Journal of Forecasting*, vol. 22, no. 4, pp. 679–688, 2006.
- [26] *IBM SPSS Statistics for Windows-Version 20.0*, IBM Corporation, New York, NY, USA, 2011.

