CAMES, **28**(2): 105–118, 2021, doi: 10.24423/cames.324

A Statistical Comparison of Feature Selection Techniques for Solar Energy Forecasting Based on Geographical Data

Saloua EL MOTAKI[1)*], Abdelhak EL FENGOUR[2), 3)]

[1)] *Department of Computer Science*
Faculty of Science Dhar El Mahraz
University Sidi Mohamed Ben Abdellah
Fez, Morocco
*Corresponding Author e-mail: saloua.elmotaki@usmba.ac.ma

[2)] *University Ibn Tofail*
Kenitra, Morocco

[3)] *Faculty of Environmental Sciences and Biochemistry*
University of Castilla-La Mancha
Toledo, Spain
e-mail: elfengourabdo@gmail.com

In recent years, solar energy forecasting has been increasingly embraced as a sustainable low-energy solution to environmental awareness. It is a subject of interest to the scientific community, and machine learning techniques have proven to be a powerful means to construct an automatic learning model for an accurate prediction. Along with the various machine learning and data mining utilities applied to solar energy prediction, the process of feature selection is becoming an ultimate requirement for improving model building efficiency. In this paper, we consider the feature selection (FS) approach potential. We provide a detailed taxonomy of various feature selection techniques and examine their usability and ability to deal with a solar energy forecasting problem, given meteorological and geographical data. We focus on filter-based, wrapper-based, and embedded-based feature selection methods. We use the reduced number of selected features, stability, and regression accuracy and compare feature selection techniques. Moreover, the experimental results demonstrate how the feature selection methods studied can considerably improve the prediction process and how the selected features vary by method, depending on the given data constraints.

Keywords: feature selection, filter method, wrapper method, embedded method, solar energy forecasting, regression performance, smart environment.

1. Introduction

Photovoltaic systems have proven to be increasingly prominent sources of energy in power grids. Nevertheless, utilities are expected to ensure electrical supply under certain constraints for the forthcoming periods, which is a challenge due to the varying aspects of the weather. Accordingly, it is important to accurately predict solar energy patterns with respect to alternative renewable energy sources. On the other hand, the energy produced by concentrating solar power and photovoltaic systems is among the most difficult variables to predict [1]. Topography, clouds, and aerosols all have an impact on solar energy forecasts. Thus, statistical and machine learning techniques that rely on historical solar production data have become of paramount interest.

In recent years, the data obtainable from real-world applications has increased significantly. Such applications can be distinguished by their quantifiable characteristics called features (attributes), which are, in their turn, exploited for knowledge extraction by machine learning techniques. Specifically, for regression tasks, a wide range of features can cause several concerns, including the loss of accuracy, higher computational load, and bias occurrence. Feature selection (FS) consists of finding a small set of characteristics that describe a given process more effectively [2].

FS involves the selection of a smaller subset from a large set of features that are relevant to a given problem. FS is a crucial phase of knowledge discovery and processing that makes data mining and machine learning tools more insightful and accurate. However, various features that are not relevant, misleading, or overlapping may arise in high-dimensional data. In addition, a number of learning algorithms typically underperform over a broad selection of features. Practitioners and scientists have accordingly chosen the FS process to preprocess the data prior to the application of mining or learning techniques, which accelerates the training and testing mechanism of the model by considerably lowering the computation time; it also improves the interpretability of the results and the prediction reliability, as well as reduces the occurrence of over-fitting. The FS techniques are generally divided into three broad categories outlined below:

- Filter-based FS: used to select relevant features, regardless of the applied data mining algorithm, through statistical measures. Such approaches include univariate models that examine each feature independently and attach a corresponding score to it; which is then used to develop a ranking-based selection of the features [3, 4]. This category also includes multivariate models; in this case, several possible subsets of features are investigated and scored for potential discrimination [5].
- Wrapper FS: the FS method in this category depends on the output of a predesigned data mining or machine learning algorithm used to determine

the goodness of a certain subset of features. The predictor is considered as a black box and the performance of the predictor is taken as an objective function to evaluate the subset of variables, thus to select features based on their performance [6].

- Embedded FS: this category is a combination of filtering and wrapping approaches. The basic concept is to integrate FS into the training process. That is, the methods are based on learning mechanisms that have built-in methods for selecting features independent test and performance evaluation functions [5, 7].

In this paper, we will investigate these methods. The three previously mentioned categories differ in the accuracy, computational performance, and susceptibility to over-fitting. Filter-based methods are used as pre-processing to sort features, whereby the highest-rated features are picked and fitted to a predictor. Thus, they are able to efficiently adapt to a high volume of data with minimal computational cost and reduced over-fitting risk. Nevertheless, their accuracy is relatively restricted as compared to the wrapping approaches [6]. For the latter, the predictor's performance is the criterion for FS, that is, the predictor is wrapped into a search algorithm to provide a subset with the best performance of the predictor, which allows achieving a high accuracy although they are prone to be more computationally expensive. The introduction of embedded FS methods aims to provide an appropriate compromise between the accuracy of the wrapper approach and the usability of filter-based selection [8].

We focus on FS methods that involve a supervised learning concept. We conduct a comparative investigation of six different FS methods applied to solar energy forecasting: correlation criteria (CC) and mutual information (MI) as filter-based methods [8], stepwise algorithm [9] and sequential feature selection (SFS) algorithm [10] as wrapper FS, and random forrest (RF) and least absolute shrinkage and selection operator (lasso) as embedded FS approaches. Our statistical analysis includes an evaluation of the accuracy, stability and performance of each FS technique based on various statistical indices.

We organize the rest of this paper as follows: a review of some recent work on FS methods for predicting solar energy is given in Sec. 2. In Sec. 3, the FS methods studied in this work are described. Section 4 discusses the results of the application of FS methods via a real-world case. Section 5 closes this work with a brief discussion.

2. Related work

Feature selection has been a highly active field of development and research since the 1970s. It has been a key contributor to several fields including data mining, machine learning, bioinformatics, environmental disaster management, etc. [2, 11–19]. The prediction of solar energy has become one of the most recent and most challenging areas of application for FS techniques. A variety of alternative methods has been developed for this purpose. In [19], a model based on FS algorithms including linear correlation (LC), reliefF, and logical information analysis (LIA) has been introduced. The objective was to enhance the forecasting process of solar energy production at multiple grid stations through the selection of the most relevant meteorological features. Similarly, the authors in [20] have discussed the LC and reliefF methods, and they employed these FS methods for selecting significant features to improve numerical weather prediction performance. In addition, they have implemented a novel FS method based on the Local Information Analysis (LIA).

Likewise, D. O'Leary and J. Kubby [21] have used the correlation-based FS to further improve the artificial neural networks (ANN) forecasting accuracy. They demonstrated the high priority of eliminating noisy and confusing meteorological features in improving the performance of solar energy prediction. Similarly, a neural network coupled with an improved version of the shark smell optimization algorithm was adopted in [22] to build a hybrid forecasting system of solar energy. Namely, a metaheuristic approach has been conceived to adjust the neural network's parameters. In addition, a two-phase FS algorithm using the MI and interaction profit theoretical criteria (IPTC) was used to discard irrelevant input features.

The authors in [23] have proposed a systematic FS scheme to develop a predictive model for building energy forecasting. They have developed a model for coupling statistical data analysis, physical building analysis, and engineering experiments. The system involves a preprocessing phase that applies domain knowledge along with an FS statistical method to reject duplicate and irrelevant features, specifically a filtering method. Additionally, the wrapping approach for determining the optimal set of features was performed.

Similarly, authors in [24] have proposed a novel FS approach that evaluates each feature importance via a bootstrapping of support vector machine (SVM) classifiers. For the best solar radiation forecast, the suggested technique selects the most applicable features. In the same vein as utilizing FS techniques to improve solar forecasting, authors in [25] have shown that FS methods can greatly improve the performance of machine learning models trained on real-world historical meteorological data.

In the same vein, we propose a study that aims to compare six approaches for predicting solar energy, which use distinct selection procedures that have never been used before. The following are the main contributions of this work:

- We present a comprehensive investigation of multiple FS techniques for solar energy prediction based on geographical data, including CC, M, SFS, stepwise, lasso, and RF. To our knowledge, no earlier research has looked into these algorithms for determining photovoltaic predictors. Linear regression (LR) and support vector regression (SVR) prediction methods are used to assess the outcomes.

- Based on realistic settings, a comparative analysis is carried out. The used data collection comprises several criteria, including stability and performance with meteorological and geographic data (see Sec. 4 for more information), enabling more comprehensive evaluation.

3. Feature selection algorithms

3.1. Filter-based FS

- **Correlation criteria (CC)**: the Pearson correlation coefficient [8] is one of the most basic criterion. It is defined as follows:

$$R(i) = \frac{cov(x_i, y)}{(var(x_i) \times var(y))^{1/2}}, \tag{1}$$

where $y = (y_1, y_2, ..., y_n)$ denotes a $1 \times n$-vector that appoints the predicted feature (output labels), and $X = (1, x_1, x_2, ..., x_n)$ represents a $n \times (p+1)$-matrix of predictors. Moreover, x_i denotes the i-th variable, and $cov(.)$ and $var(.)$ are the covariance and variance functions, respectively.

- **Mutual information (MI)**: it measures the dependence between random variables. It differs from the correlation coefficient that measures only the linear dependencies between these variables [26]. The MI has zero value if and only if the variables are strictly independent and increases with the dependence. A rapid and accurate method for estimating MI has been developed in [27]. This method is based on an estimation of the entropy. Indeed, the MI and the entropy are linked by the following formula:

$$\mathrm{MI}(x_i, y) = H(x_i) + H(y) - H(x_i, y) \tag{2}$$

with $H(.)$ being the entropy of one or more variables. This entropy is computed by the nearest neighbor method [27]. The resulting MI between X and y means that if the variables are independent, the MI would be null and greater than zero if they are dependent. This implies that one variable can yield insight into the other, which proves dependence.

3.2. Wrapper FS methods

- **Sequential feature selection (SFS)**: the algorithm begins with an initially empty set. In the first step, the feature that provides the highest value for the objective function is added. Then, the remaining characteristics are individually included in the subset in the second phase, and the new subset is assessed. If an individual feature keeps the highest classification accuracy, it is constantly involved in the subset. The approach is continued until all of the important features are included [10].

- **Stepwise algorithm**: for linear regression, logistic regression, and other classical regression models, stepwise regression is a wrapper approach that identifies the optimal predictive elements to use in a model from a larger collection of potentially predictive features [9]. The implementation of the stepwise regression involves two possible pathways. The first way (called decreasing FS) is to use a model that incorporates all the features that are likely to impact the target variable and then gradually withdraw the weakest features from the initial model based on a goodness-of-fit measure that adjusts the number of features in the model. The procedure goes on, and further attributes are eliminated in progressively decreasing levels until the improvement metric cannot be adjusted. The second fundamental technique (referred to as ascending FS) begins with a model that contains only one fixed variable, and then it enlarges this model to incorporate the features of a range of attributes that make the fit measure more meaningful. The method is reproduced by adding more characteristics with a series of gradual upward increments and finishes when the modified fitting measures cannot be improved.

  We use a hybrid of ascending and descending FS approaches for this study. First, we define a selection of features that might be relevant using a ascending method. The descending FS approach is then used to identify features that require elimination from the formerly constructed subset. We utilize the adjusted R-squared to monitor this process.

3.3. Embedded FS methods

- **A least absolute shrinkage and selection operator (lasso)**: it was introduced by R. Tibshirani [28], lasso is a retraction and an FS approach for linear regression patterns. Lasso's aim is to build a collection of predictors in a manner that reduces the loss function on a quantitative predictor variable by setting constraints to the model that pushes some coefficients down to zero. As a result, variables with non-zero coefficients are more strongly attached to the outcome variable, while variables with zero coefficients are eliminated. Lasso is a multiple regression function that is estimated mathematically. A penalization function K can be used to define it as follows:

$$K(\beta) = \arg\min \left\{ \frac{1}{2} \sum_{i=1}^{n} \left(y_i - \beta_0 - \sum_{j=1}^{p} x_{ij}\beta_j \right)^2 + \lambda \sum_{j=1}^{p} |\beta_j| \right\}, \quad (3)$$

where $\beta = (\beta_0, \beta_1, ..., \beta_p)$ is a p+1-column vector that represents the related affects per each predictor in X, and λ stands for the penalty parameter that regulates the compromise between model complexity and accuracy loss. By precisely setting some of the coefficients to zero, lasso reduces the variability of the estimations, allowing for the development of easily interpretable models.

- **Random forest (RF)**: it is based on the principle of model aggregation, is a popular and accurate method for high and poorly structured classification and regression problems [29]. The main principle underlying the RF framework is to build several impartial decision trees using stratified random training data with replacement, in which each tree determines a class and the forest decides the value of the highest among all trees.

  In essence, we assume that the prediction of the tree t_k is given by $\widetilde{y}_k$ for an input X and a number of trees equal to K:

  $$\widetilde{y} = \text{majority Vote} \left\{\widetilde{y}_k\right\}_1^K . \quad (4)$$

  L. Breiman [29] presented RF as the main approach to measure the relevance of the features involved in the prediction by scoring the out-of-bag (OOB) value. It aims to compute the difference between the baseline mean error and the random mean error of the OOB pattern. The approach replenishes the stochastic usage of the RF model to estimate this allowance and compute the mean error for all trees of the given feature in OOB. The purpose of this permutation is to override the actual correlation between the feature given and the y-values, and then to consider the implication for the RF model of such an override. The most relevant features are those with a large decrease in mean error.

4. Real-world application

4.1. Dataset

We use a dataset from the Open Power System Data (OPSD) framework for this research, which provides free and public data for power system analysis [30]. This database encompasses the data of conventional power plants for various European countries, including Germany. It consists of the technical characteristics of each particular power station, e.g., geographical information, principal energy source, etc. We process and extend the selected dataset to include weather data available on the same platform. Then, the dataset is processed and coupled with a weather dataset provided by the same platform. The reader is directed to [31] for further details. Typically, the actual solar energy production (given in MW) is assumed to be the predicted variable, and a summary of certain geographical and meteorological features considered in this investigation is provided in Table 1.

Table 1. The information used to estimate solar energy output, including ground, wind, temperature, and air characteristics.

Feature	Description
Lat	The latitude of each geographical chunk examined [m]
Lon	The longitude of each geographical chunk examined [m]
h1	Height above ground level [2 m above height of displacement]
h2	Height above ground level [10 m above height of displacement]
v1	Velocity at height h1 [m/s]
v2	Velocity at height h2 [m/s]
v_50	Velocity at height 50 m above ground level [m/s]
z0	Roughness length [m]
SWTDN	Full horizontal radiation at the top of the atmosphere [W/m^2]
SWGDN	Full ground horizontal radiation [W/m^2]
T	Temperature at h1 level [K]
Rho	Surface air density [kg/m^3]
p	Surface air pressure [Pa]

4.2. Feature selection results and discussion

Having performed the considered FS methods on the above-mentioned dataset, we list the most relevant features to predict solar energy. We employ Python's Scikit-learn package, which provides a variety of machine learning packages and built-in functions, including MI, CC, SFS, RF, and lasso models. We, however, implement the stepwise technique since the Python package statsmodels does not support built-in functions for ascending and descending stepwise.

Figure 1 exhibits features selected by the methods addressed in this work. We observe that by employing the classifiers, the SWTDN, SWGDN, and T are selected as the most relevant features by most of the algorithms used with a high score. On the other hand, the cumulated hours characteristic (Figs 1a and 1b) is denoted as a significant feature only by embedded approaches (lasso and RF), which signifies that the selection results on the strategy followed. We remark, moreover, that the sub-set of features that are selected by all FS algorithms does not have the same importance. For example, the feature latitude (lat) is depicted

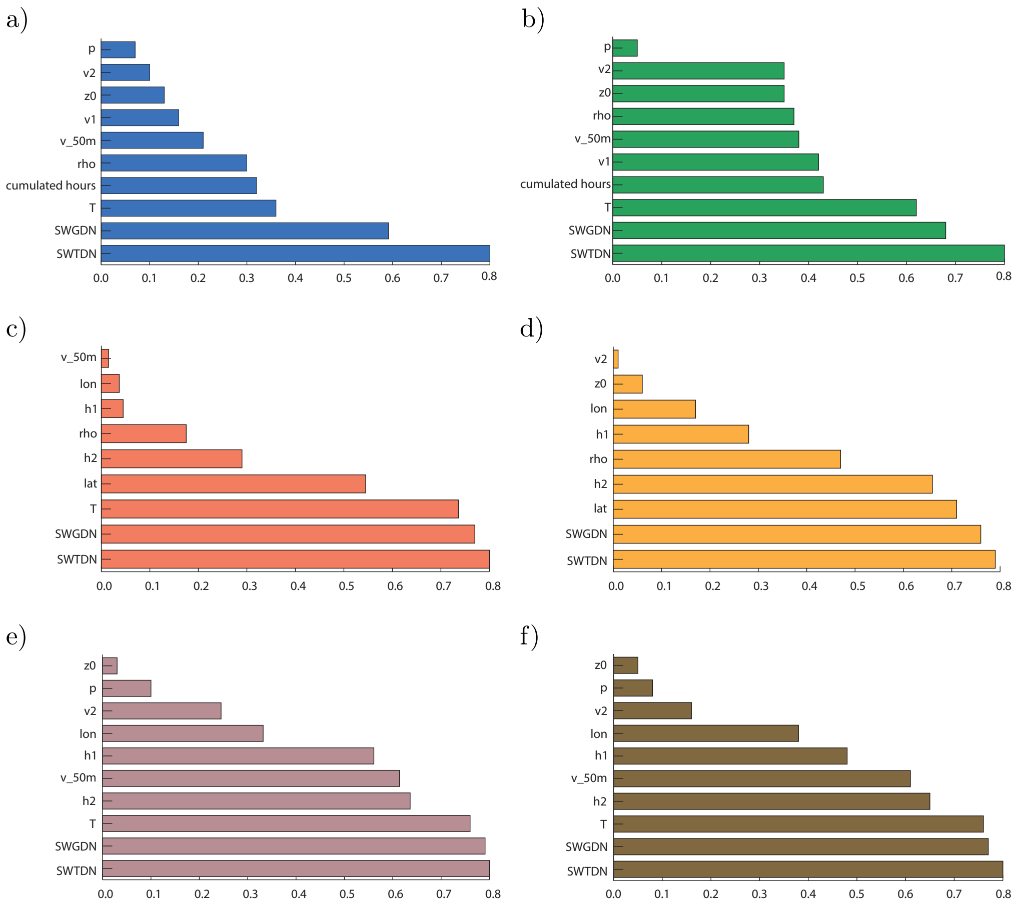

FIG. 1. Top most important features selected by:
a) lasso, b) RF, c) stepwise, d) SFS, e) CC, and f) MI selection.

by the SFS selection with a score of about 0.73 (Fig. 1d); the same feature is selected by stepwise method with 0.54 (Fig. 1c). We note that latitude is denoted as a relevant feature only by the two wrapper FS techniques. Similarly, the feature $v_{5}0m$ is selected by different FS techniques with various scores ranging from 0.15 with stepwise method to 0.58 with the CC selection (Fig. 1e).

4.2.1. Stability analysis. Various FS algorithms may be introduced to a specific application, and the optimal one that satisfies the appropriate criterion can be chosen. An underestimated challenge is the stability of the function selection algorithms. When new training samples are introduced or any training samples are omitted, the stability of the FS algorithm can be interpreted as the algorithm's potential to generate a consistent feature subset [31]. Many measures exist in the literature to calculate stability. In this study, we use the Jaccard index (Appendix).

Figure 2 exhibits the Jaccard index values as a function of the number of selected features. For this test, we have used all the features included in the dataset. From the analysis of the figure, it follows that filter-based methods generate the most stable sets of selected features given their robust approach in terms of overfitting. They typically involve either univariate or multivariate statistical models and are not influenced by any learning techniques. However, wrapper-based and embedded-based FS do not always perform well with classifiers as they both use their own learning process for selecting features.

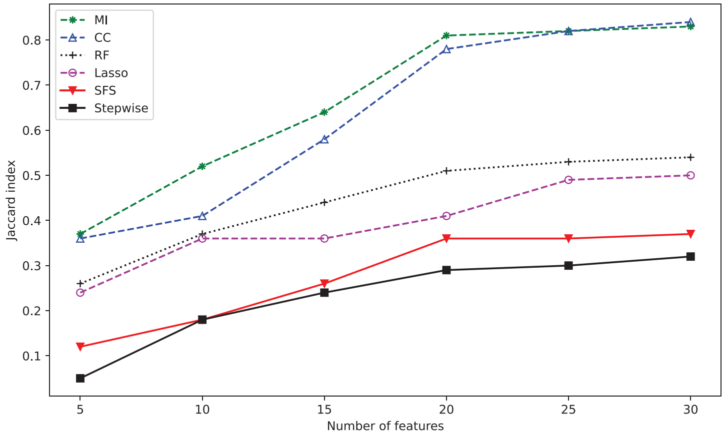

FIG. 2. Typical FS stability measured by Jaccard's index for the six FS techniques.

4.2.2. Performance analysis Having ensured that our set of FS methods is stable to data variations, we check if the ultimate selection of features is relevant. The universal metric for evaluating a learning system is performance, which is typically measured in the case of classification or regression as prediction accuracy. To analyze the accuracy of the selected features, Fig. 3 gives the results of applying a LR and SVR with the selected subsets of features. We can observe that the performance of regression methods is low due to the calculation of MI and CC. In fact, by estimating the pair distribution function (PDF) of the features and the distribution of the output class, the computation of MI and CC might not be accurate and readily be driven by the incremental densities. Typically, the plot of the filter-based methods demonstrates that the ranking techniques are proven to be trivial. On the other hand, we can clearly see that the maximum performance is achieved by the stepwise algorithm for both regression approaches, as it provides an accuracy of more than 96% with LR. According to the results obtained, the wrapper methods provide the best performance among the different techniques studied.

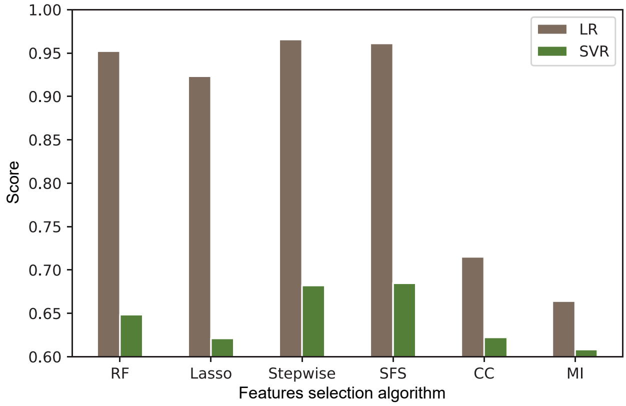

FIG. 3. Bar chart exhibiting the R-squared score values obtained by LR and SVR models applied with the most relevant features selected.

In general, the LR performance achieved is acceptable using all of the selection strategies explored. Nevertheless, the regression using SVR lacks accuracy as it does not reach 70% for any of the FS methods, especially with the features selected using filter-based approaches.

5. Conclusion

Feature selection becomes an essential part of regression and classification problems when dealing with high volumes of datasets containing redundant, noisy, and misleading data. In this study, we used six different techniques to select the features that are most relevant for predicting solar energy generation as a function of different geographical and meteorological parameters. FS methods prove that having too much information does not always lead to effective machine learning applications. Using the basic classification performance ratings, we select the most appropriate FS algorithm. Accordingly, we evaluate the FS algorithms, in this work, following the considerations: reduced number of features, stability, and regression performance.

From the results obtained, we conclude that filter-based approaches (CC and MI) show significant stability compared to both embedded-based methods (RF and lasso) and wrapper FS methods (stepwise and SFS); these latter ones are the most likely to be unstable. However, the performance of the wrapper methods – stepwise as an example – allows selecting a set of features that can improve the accuracy of the regression approaches used for the evaluation.

Overall, the application of FS is always likely to provide the advantages of improved data insight, enhanced learning model, increased standardization, as well as the elimination of irrelevant features. We have used methods from different categories to select the most important features to predict solar energy generation in a smart grid successfully. However, an additional step was required to improve the prediction accuracy with the SVR. It involved excluding correlated features despite their selection as relevant. A possible explanation lies in the type of data handled. For example, the characteristics involving the geographical pattern may be dependent, making it convenient to retain just the highly correlated feature with the forecasted variable and discard the others. Note that we apply the regression algorithms with hyper-parameters by default.

Appendix

Jaccard index

The Jaccard index is a metric used to measure the average similarity of all selected pairwise feature subsets (W) [32]. It is given by the following:

$$\mathrm{JS} = \frac{2}{W \times (|W-1|)} \sum_{i=1}^{|W|-1} \sum_{i+1}^{|W|} J(S_i, S_j), \tag{5}$$

where $S = (S_1, S_2, ..., S_n)$ is the set of selected features and $J(S_i, S_j) = \frac{|S_i \cap S_j|}{|S_i \cup S_j|}$. The stability index (JS) provides an output within a range of 0 indicating that the FS algorithm is unstable and a value close to 1 signifies that the algorithm is stable [33].

References

1. M. Diagne, M. David, Ph. Lauret, J. Boland, N. Schmut, Review of solar irradiance forecasting methods and a proposition for small-scale insular grids, *Renewable and Sustainable Energy Reviews*, **27**: 65–76, 2013, doi: 10.1016/j.rser.2013.06.042.

2. H. Liu, L. Yu, Toward integrating feature selection algorithms or classification and clustering, *IEEE Trans. on Knowledge and Data Engineering*, **17**(4): 491–502, 2005, doi: 10.1109/TKDE.2005.66.

3. M.A. Hall, Correlation-based feature selection for discrete and numeric class machine learning, [in:] *Proceedings of the Seventeenth International Conference on Machine Learning*, ICML '00, pp. 359–366, Morgan Kaufmann Publishers Inc., 2000.

4. M. Dash *et al.*, Feature selection for clustering – a filter solution, [in:] *Proceedings of the 2002 IEEE International Conference on Data Mining*, ICDM '02, pp. 115–122, Washington, DC, USA, IEEE Computer Society, 2002.

5. Y. Saeys, I. Inza, P. Larrañaga, A review of feature selection techniques in bioinformatics, *Bioinformatics*, **23**(19): 2507–2517, 2007 doi: 10.1093/bioinformatics/btm344.

6. R. Kohavi, G.H. John, Wrappers for feature subset selection, *Artificial Intelligence*, **97**(1–2): 273–324, 1997, doi: https://doi.org/10.1016/S0004-3702(97)00043-X.

7. L. Rangarajan, Veerabhadrappa. Bi-level dimensionality reduction methods using feature selection and feature extraction, *International Journal of Computer Applications*, **4**(2): 33–38. 2010.

8. I. Guyon, A. Elisseeff, An introduction to variable and feature selection, *Journal of Machine Learning Research*, **3**: 1157–1182, 2003.

9. R. Mundry, C.L. Nunn, Stepwise model fitting and statistical inference: turning noise into signal pollution, *The American Naturalist*, **173**(1): 119–123, 2009, doi: 10.1086/593303.

10. J Reunanen, Overfitting in making comparisons between variable selection methods, *Journal of Machine Learning Research*, **3**:1371–1382, 2003.

11. J. Cai, J. Luo, S. Wang, S. Yang, Feature selection in machine learning: A new perspective, *Neurocomputing*, **300**: 70–79, 2018, doi: 10.1016/j.neucom.2017.11.077.

12. J. Brownlee, *Data Preparation for Machine Learning: Data Cleaning, Feature Selection, and Data Transforms in Python*, Machine Learning Mastery, 2020.

13. J. Li *et al.*, Feature selection: A data perspective, *ACM Computing Surveys*, **50**(6): 1–45, 2017, doi: 10.1145/3136625.

14. G. Georgiev, I. Valova, N. Gueorguieva, Feature selection for multiclass problems based on information weights, *Procedia Computer Science*, **6**: 189–194, 2011, doi: 10.1016/j.procs.2011.08.036.

15. L. Wang, Y. Wang, Q. Chang, Feature selection methods for big data bioinformatics: A survey from the search perspective, *Methods*, **111**: 21–31, 2016, doi: 10.1016/j.ymeth.2016.08.014.

16. P. Drotár, J. Gazda, Z. Smékal, An experimental comparison of feature selection methods on two-class biomedical datasets, *Computers in Biology and Medicine*, **66**: 1–10, 2015, doi: 10.1016/j.compbiomed.2015.08.010.

17. S. Khalid, T. Khalil, S. Nasreen, A survey of feature selection and feature extraction techniques in machine learning, [in:] *2014 Science and Information Conference*, pp. 372–378, Aug. 2014, doi: 10.1109/SAI.2014.6918213.

18. W. Awada, T.M. Khoshgoftaar, D. Dittman, R. Wald, A. Napolitano, A review of the stability of feature selection techniques for bioinformatics data, [in:] *2012 IEEE 13th International Conference on Information Reuse & Integration (IRI)*, pp. 356–363, 2012, doi: 10.1109/IRI.2012.6303031.

19. R. Martin, R. Aler, J.M. Valls, I.M. Galvan, Machine learning techniques for daily solar energy prediction and interpolation using numerical weather models, *Concurrency and Computation: Practice and Experience*, **28**(4): 1261–1274, 2016, doi: 10.1002/cpe.3631.

20. R. Aler, R. Martín, J.M. Valls, I.M. Galván, A study of machine learning techniques for daily solar energy forecasting using numerical weather models, [in:] D. Camacho, L. Braubach, S. Venticinque, C. Badica [Eds], *Intelligent Distributed Computing VIII*, *Studies in Computational Intelligence*, Vol. 570, pp. 269–278, Springer International Publishing, 2015, doi: 10.1007/978-3-319-10422-5_29.

21. D. O'Leary, J. Kubby, Feature selection and ANN solar power prediction, *Journal of Renewable Energy*, **2017**: 1–7, 2017, doi: 10.1155/2017/2437387.

22. O. Abedinia, N. Amjady, N. Ghadimi, Solar energy forecasting based on hybrid neural network and improved metaheuristic algorithm, *Computational Intelligence*, **34**(1): 241–260, 2018, doi: 10.1111/coin.12145.

23. L. Zhang, J. Wen, A systematic feature selection procedure for short-term data-driven building energy forecasting model development, *Energy and Buildings*, **183**: 428–442, 2019, doi: 10.1016/j.enbuild.2018.11.010.

24. O. Garcia-Hinde *et al.*, Feature selection in solar radiation prediction using bootstrapped SVRs, [in:] *2016 IEEE Congress on Evolutionary Computation (CEC)*, pp. 3638–3645, 2016, doi: 10.1109/CEC.2016.7744250.

25. M.R. Hossain, A.M.T. Oo, A.B.M.S. Ali, The effectiveness of feature selection method in solar power prediction, *Journal of Renewable Energy*, **2013**, Article ID: 952613, 9 pages, 2013, doi: 10.1155/2013/952613.

26. C. Lazar *et al.*, A survey on filter techniques for feature selection in gene expression microarray analysis, *IEEE/ACM Transactions on Computational Biology and Bioinformatics*, **9**(4): 1106–1119, 2012, doi: 10.1109/TCBB.2012.33.

27. A. Kraskov, H. Stögbauer, P. Grassberger, Estimating mutual information, *Physical Review E*, **69**: 066138, 2004, doi: 10.1103/PhysRevE.69.066138.

28. R. Tibshirani, Regression shrinkage and selection via the lasso, *Journal of the Royal Statistical Society: Series B (Methodological)*, **58**(1): 267–288, 1996, doi: 10.1111/j.2517-6161.1996.tb02080.x.

29. L. Breiman, Random Forests, *Machine Learning*, **45**(1): 5–32, 2001, doi: 10.1023/A:1010933404324.

30. *Open Power System Data – A platform for open data of the European power system*, https://data.open-power-system-data.org/conventional_power_plants/2018-12-20 (accessed: 2019-09-14).

31. A.-C. Haury, P. Gestraud, J.-P. Vert, The influence of feature selection methods on accuracy, stability and interpretability of molecular signatures, *PloS ONE*, **6**(12): e28210, 2011, doi: 10.1371/journal.pone.0028210.

32. P. Mohana Chelvan, K. Perumal, A survey on feature selection stability measures, *International Journal of Computer and Information Technology*, **5**(1): 98–103, 2016.

33. U.M. Khaire, R. Dhanalakshmi, Stability of feature selection algorithm: A review, *Journal of King Saud University – Computer and Information Sciences*, 2019, doi: 10.1016/j.jksuci.2019.06.012.

Received February 6, 2021; revised version May 29, 2021.