

Analyzing the Impact of Feature Selection on Crop Yield Prediction

Jayshree Hajgude, Tanuja Sarode

Department of Computer Engineering, Thadomal Sahani Engineering College, Mumbai University, Mumbai, India

Email: Jayashree.hajgude@ves.ac.in, Tanuja.sarode@thadomal.org

How to cite this paper: Hajgude, J. and Sarode, T. (2024) Analyzing the Impact of Feature Selection on Crop Yield Prediction. *Journal of Computer and Communications*, 12, 278-291.

<https://doi.org/10.4236/jcc.2024.128017>

Received: June 26, 2024

Accepted: August 27, 2024

Published: August 30, 2024

Copyright © 2024 by author(s) and Scientific Research Publishing Inc. This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

Abstract

In agriculture sector, machine learning has been widely used by researchers for crop yield prediction. However, it is quite difficult to identify the most critical features from a dataset. Feature selection techniques allow us to remove the extraneous and noisy features from the original feature set. The feature selection techniques help the model to focus only on the important features of the data, thus reducing execution time and improving efficiency of the model. The aim of this study is to determine relevant subset features for achieving high predictive performance by using different feature selection techniques like Filter methods, Wrapper methods and embedded methods. In this work, different feature selection techniques like Rank-based feature selection technique, weighted feature selection technique and Hybrid Feature Selection Technique have been applied to the agricultural data. The optimal feature set returned by different feature selection techniques is used for yield prediction using Linear regression, Random Forest, and Decision Tree Regressor. The accuracy of prediction obtained using the above three methods has been analyzed by using different evaluation parameters. This study helps in increasing predictive accuracy with the minimum number of features.

Keywords

Crop Yield, Feature Selection, Hybrid, Embedded, Analysis

1. Introduction

Recently, data has been rapidly increasing in both the number of samples and features. To determine data trends, it is necessary to study how to utilize these large-scale data. The huge number of high dimensional data has a big challenge due to the presence of noisy, redundant, and irrelevant data as it increases the computational complexity of the model and overcomes overfitting problem. So,

feature selection is necessary as it allows us to choose relevant features from the original data by removing noisy, irrelevant, and redundant features.

In machine learning problems, high dimensional data, especially in terms of many features, is increasing these days [1]. In data analysis applications, the dataset may have many features which determine the applicability and usability of the data [2]. The reduction of the large number of features to smaller ones with preserving the relevant information while discarding the redundant features is referred to as feature selection [3]. Nowadays agriculture data is becoming high dimensional data, especially in terms of features. So, we must determine relevant features from this high-dimensional data. We can choose correlated and non-redundant features, so feature selection plays an important role.

Feature selection process involves the following steps.

- Understand the Problem: Gain insights into the problem domain, including the factors that influence crop yield such as weather conditions, soil characteristics, crop type, farming practices, etc.
- Data Collection: Gather a comprehensive dataset containing various features related to crop yield, including historical yield data, weather data, soil data, crop management practices, etc.
- Exploratory Data Analysis: Perform Exploratory Data Analysis to understand the relationships between different features and the target variable. This involves visualizing data, computing summary statistics, and identifying correlations.

Agriculture plays an impactful role in Indian economy and crop yield has a major contribution. This section includes details about different methods for crop yield prediction. Feature selection techniques are classified into three major types called filter, wrapper, and Embedded approaches [4] [5]. The wrapper method is used for selecting features, but the approach is computationally expensive and does not provide optimal number features. Filter approach is independent of any learning algorithm [6]. The embedded approach uses the advantage of both supervised and unsupervised approaches. By combining feature selection techniques, we can improve the learning performance [7]. Forward selection, wrapper technique is used for selecting important features [8], correlation-based filter feature selection algorithm is used in [9], random forest variable Importance embedded technique is used in [10]. These algorithms do not result in significant features. To overcome this drawback a hybrid approach is used which is combination of Correlation and Sequential Forward Feature Selection [11]. Maya Gopal and Bhargavi [12] analyzed the performance of machine learning algorithms using various feature selection techniques for crop yield prediction. Maya Gopal and Bhargavi [13] proposed sequential forward selection technique to find the “optimal” feature subset by iteratively selecting features based on the performance of the classifiers. Zahra Karimi *et al.* [14], implemented hybrid filter feature selection scheme for intrusion detection. The experimental results show that the proposed technique offers higher accuracy than other methods. Surabhi Chouhan *et al.* [15], proposed a method of applying the

Particle Swarm Optimization and Support Vector Machine (PSO-SVM) to select features from a dataset. Akhiat, Y [16] proposed an ensemble selection framework to improve the proposed feature selection algorithm to reduce the overfitting problem using multi-bagging and dropout technique. According to G. Mariammal *et al.* [17] a modified recursive feature elimination approach was presented for selecting acceptable features for crop prediction using bagging methodologies. A. Suruliandi *et al.* [18], carried out comparative examination of multiple wrapper feature selection methodologies to recommend the best suited crop. It is found that fusion of Recursive Feature Elimination approach and Adaptive Bagging classifier performed well as compared to remaining methods. Alberto Gonzalez-Sanchez *et al.* [19], evaluated different model correctness and used artificial neural networks to demonstrate result. Hsu HH, Hsieh CW, Da LuM [20] used hybrid feature selection technique which is combination of filters and wrappers. Dewi C [21], used random forest and support vector machine techniques to improve the model accuracy using feature selection. Nicholas Pudjihartono [22] reviewed different feature selection techniques for disease risk prediction. M.S. Suresh Sumi [23], used combination of Filter and Wrapper feature selection techniques to improve classification Accuracy. S. P. Raja [24] applied different feature selection techniques and used ensemble model to improve the performance. Thomas. Rincy [25], Kajal P Parmar [26], Younes Bouchlaghem [27] performed a detail review of different feature selection techniques that can be used for effective accuracy of the machine learning model. Mrs. K.A. Poornima [28], combined Correlation and Sequential Forward Feature Selection techniques to improve accuracy in crop prediction.

The Efficient method for prediction of crop yield is an important process of agriculture and machine Learning is playing a major role in this area in recent years. Feature selection and prediction are critical machine learning techniques, and it improves prediction accuracy and helps to avoid processing of unwanted feature data and minimizing execution time. For selecting relevant features in literature mainly filter, wrapper, and embedded methods are used frequently. To improve the accuracy of prediction efficient feature selection techniques should be introduced. As a part of research this study implemented hybrid feature selection technique namely Rank-based W2FE feature selection techniques which is a fusion of the efficient techniques from Filter, Wrapper and Embedded method. The Wrapper 2 Filter Embedded (W2FE) method is a combination of two wrapper methods, one Filter and one embedded method. The result of W2FE feature selection model has been given to rank-based and weighted rank-based algorithm to determine the most suitable features for crop yield prediction. Weighted rank-based W2FE feature selection algorithm gives better results as compared to the Rank-based W2FE feature selection techniques. The main contributions of this research are summarized as follows. In this research work various features affect crop yield. has been studied. Different feature selection techniques have been studied and implemented, to determine relevant features for predicting crop yield. In this study Linear regression, Random Forest,

and Decision Tree regressor methods are used for assessing the results given by different feature selection techniques. The most suitable methods are chosen after evaluating performance using different machine learning models. hybrid feature selection techniques, namely rank-based W2FE feature selection algorithm and Weighted rank-based W2FE feature selection algorithm have been proposed and implemented by integrating efficient feature selection techniques to select optimal feature sets and hence improves efficiency of crop yield prediction.

2. Method

In this work, different feature selection techniques have been implemented and performance of different feature selection techniques has been analyzed for prediction of the crop yield. Hybrid feature selection techniques have been implemented by combining suitable methods from different feature selection techniques to select a minimal set of features hence improved crop yield prediction. This study considered metrological data of Marathwada region from India country. Weather data is collected from the Meteorological Department Pune, crop production data has been collected from Government Website.

2.1. Study Area

This study area has been used as Marathwada region from Maharashtra state in India. The Marathwada have different principal foods like Soybean, Wheat, Jawar etc. This study considered 8 districts from the Marathwada region.

2.2. Dataset Details

This study has used crop production Dataset used from publicly available records of the Indian Government for the years 2001 to 2019 and the weather features details have been sourced from Indian Meteorological Department in Pune. Features considered for this study are rainfall details, temperature details, precipitation, evaporation, and area.

2.3. Methodology

1) Understand the Problem: Gain insights into the problem domain, including the factors that influence crop yield such as weather conditions, soil characteristics, crop type, farming practices, etc.

2) Data Collection: Gather a comprehensive dataset containing various features related to crop yield, including historical yield data, weather data, soil data, crop management practices, etc.

3) Exploratory Data Analysis: Perform Exploratory Data Analysis to understand the relationships between different features and the target variable (crop yield). This involves visualizing data, computing summary statistics, and identifying correlations.

4) Feature Selection Techniques: Apply different feature selection techniques

and analyze the performance.

2.4. Proposed Framework

In the proposed work, a hybrid method-W2FE Rank-based/Weighted Rank-based Feature selection technique for feature selection has been proposed which consists of fusion of different feature selection techniques. The architecture of the proposed model is represented in **Figure 1**. Different feature selection techniques have been implemented and performance has been evaluated.

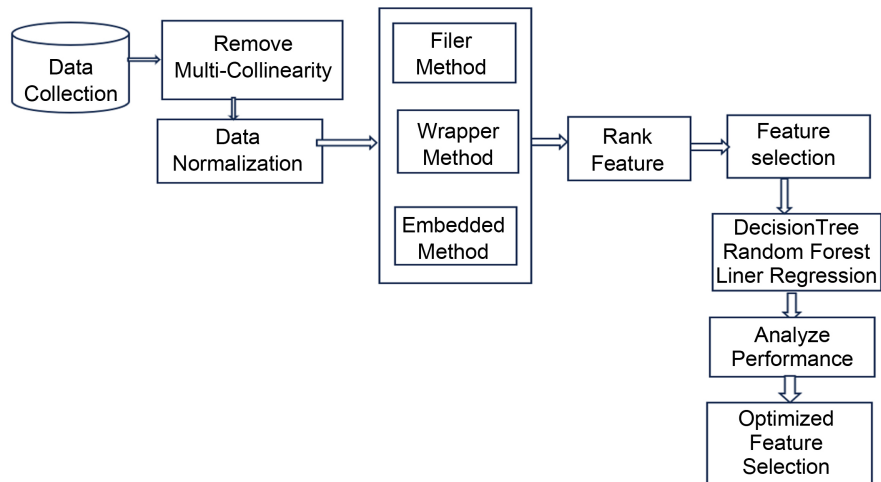


Figure 1. The architecture of the proposed feature selection model.

Architecture of proposed hybrid feature selection technique model consists of following major components:

- 1) **Data Preprocessing component:** In data pre-processing multicollinearity among the data has been removed and data has been standardized by using Min-Max normalization technique.
- 2). **Feature selection component:** Preprocessed data has been given as a input to the feature selection module. In this module, the most suitable methods from different types of feature selection technique is being used, and the best features are identified for crop yield prediction. Rank-based feature elimination technique is used to rank features while determining relevant features for crop yield prediction.

2.5. Implementation Details

As a part of the research, this study implemented hybrid model: W2FE Rank-based/Weighted Rank-based Feature selection algorithm which is a fusion of the most suitable techniques from Filter, Wrapper and Embedded method. The result of hybrid model has been given to rank-based and weighted rank-based algorithm to determine the most suitable features for crop yield prediction rank-based feature selection algorithm, weighted rank-based feature.

Algorithmic Details:

1) W2FE Rank-based /Weighted Rank-based Feature Selection Algorithm:

Input: feature set $F = \{f_1 \succ f_2 \succ \dots \succ f_m\}$ //m is the number of features

Output: optimal feature set R

- a) Create a Feature Score Table (FST) of 8 rows and 7 columns, having the evaluation scores provided by the different methods for feature selection of each feature.
- b) F1 = Feature set after applying Filter based technique.
- c) F2 = Feature set after applying Wrapper based technique.
- d) F3 = Feature set after applying Embedded technique.
- e) $F = F1 \cup F2 \cup F3$.
- f) Apply Rank-based and Weighted Rank-based algorithm.
- g) Calculate the total score by taking sum of all scores for every feature.
- h) If score $> 0.5 * \text{number of features}$, then.
- i) Add feature in the selected features.

2) Feature Selection Techniques:

a) Filter Methods: Uses statistical measures like correlation, chi-square, or mutual information to rank features based on their relevance to the target variable.

b) Wrapper Methods: Employ algorithms like Recursive Feature Elimination (RFE) or Forward/Backward Selection, which select subsets of features based on the performance of a specific machine learning algorithm.

c) Embedded Methods: Some machine learning algorithms inherently perform feature selection during training. Examples include LASSO (L1 regularization) and tree-based methods like Random Forests.

3) Rank-Based Feature Selection Algorithm:

Input: feature set $F = \{f_1 \succ f_2 \succ \dots \succ f_m\}$ //m is the number of features

Output: ranked features R

- Create a Feature Score Table (FST) of 8 rows and 12 columns, having the evaluation scores provided by the different methods for feature selection of each feature.
- If feature has been selected, then provide score 1 otherwise 0
- Calculate Total Score by taking sum of all scores for every feature.
- If score $> 0.5 * \text{number of features}$, then
Add feature in the selected features.
- Return features.

4) Weighted Rank-Based Feature Selection Algorithm:

Input: feature set $F = \{f_1 \succ f_2 \succ \dots \succ f_m\}$ //m is the number of features

Output: ranked features R

- Create a Feature Score Table (FST) of 8 rows and 7 columns, having the evaluation scores provided by the different methods for feature selection of each feature.
- Create rank table from FST.
- For every selected feature assign weight depending on rank of feature.

Calculate the total score by taking sum of all scores for every feature.
 If score $>0.5 \times$ number of features, then.
 Add feature in the selected features.

- Return features.

3. Results and Discussion

This research work analyses different types of feature selection techniques like Filter feature selection technique, Wrapper feature selection technique and Embedded feature selection techniques, Rank-based feature selection technique, Weighted feature selection technique and Hybrid feature section techniques: W2FE Rank-Based/Weighted Rank-Based Feature selection algorithm.

1) Analysis of Basic Feature Selection Techniques: The feature subsets returned by different basic feature selection techniques are shown in **Table 1**.

Table 1 depicts the result of different rank-based Feature selection Algorithm implemented as a part of said research.

Table 1. Feature selection algorithm result.

Feature Selection Method/Feature	MMAX	HMAX	MMIN	LMIN	TRMF	HVRYF	MWS	AREA	RD	MEVP	MWS	PRCP	HMD
Correlation	N	N	N	N	Y	Y	N	Y	Y	N	N	N	N
Chi-Square	N	N	Y	N	Y	N	N	N	N	Y	N	Y	N
Forward Elimination Technique	N	N	N	Y	N	Y	Y	N	N	Y	Y	N	N
Backward Elimination Selection	N	N	N	Y	Y	N	Y	N	N	Y	Y	N	N
SFS Linear Regression	N	Y	Y	N	N	Y	N	N	N	Y	N	N	N
SFS KNN	Y	Y	N	Y	N	N	N	N	Y	N	N	N	N
RFE	N	N	Y	Y	N	N	N	N	N	N	N	Y	Y
Random Forest Importance	Y	N	N	N	Y	Y	N	Y	N	N	N	N	N
Embedded LGB	Y	Y	N	Y	N	N	N	N	N	N	Y	N	N

2) Rank-based Feature Selection Technique: After applying different feature selection techniques, the Rank-based feature selection technique has been applied on the features returned by different feature selection techniques. Rank-based feature selection technique returns features whose score is minimum 0.5.

Table 2 shows the score of every feature after applying different feature selection techniques as a part of said research.

Table 2. Rank-based feature selection algorithm.

Feature Selection Method/Feature	MMAX	HMAX	MMIN	LMIN	TRMF	HVRYF	MWS	AREA	RD	MEVP	MWS	PRCP	HMD
Correlation	N	N	N	N	Y	Y	N	Y	Y	N	N	N	N
Chi-Square	N	N	Y	N	Y	N	N	N	N	Y	N	Y	N
Forward Elimination Technique	N	N	N	Y	N	Y	Y	N	N	Y	Y	N	N
Backward Elimination Selection	N	N	N	Y	Y	N	Y	N	N	Y	Y	N	N
SFS Linear Regression	N	Y	Y	N	N	Y	N	N	N		N	N	N
SFS KNN	Y	Y	N	Y	N	N	N	N	Y	N	N	N	N
RFE	N	N	Y	Y	N	N	N	N	N	N	N	Y	Y
Random Forest Importance	Y	N	N	N	Y	Y	N	Y	N	N	N	N	N
Embedded LGB	Y	Y	N	Y	N	N	N	N	N	N	Y	N	N
Total Score	4	3	2	5	4	4	2	2	2	3	3	2	1

After applying Rank-based technique on the above result it has returned [MMAX, HMAX, TRMF, HVRF] features.

3) Weighted Rank-based feature selection technique: For improving performance Weighted Rank-based feature selection technique has been implemented. In weighted Rank-based selection technique weights are assigned to every feature depending on the importance given by different feature selection technique. **Table 3** shows the score of every feature after applying Weighted Rank-based feature selection technique returns features whose weighted score is minimum 0.5.

The optimal features selected by rank-based technique are “MMAX”, “HMAX”, “LMIN”, “HVRF” features.

4) Performance Evaluation of Feature Selection Techniques using Machine Learning Models:

The performance of these techniques has been evaluated using different machine learning models like Linear Regressor, Decision Tree and Random Forest. **Table 4** shows the performance of different feature selection techniques.

After applying different feature selection techniques, it is found that Embedded LGB (E), SFS LR (W), Chi-Square (F), FST/BET (W) performed better as compared to other feature selection techniques.

5) Performance Evaluation of W2FE Rank-Based and Weighted Rank-Based Feature Selection Techniques using Machine Learning Models:

Table 3. Weighted rank-based feature selection algorithm.

Feature Selection Method/Feature	MMAX	HMAX	MMIN	LMIN	TRMF	HVRYF	MWSAREA	RD	MEVP	MWS	PRCP	HMD
Correlation	N	N	N	N	Y(3)	Y(4)	N	Y(2)	Y(2)	N	N	N
Chi -Square	N	N	Y	N	Y	N	N	N	N	Y	N	Y(1)
Forward Elimination Technique	N	N	N	Y(2)	N	Y(4)	N	N	N	Y(1)	Y(3)	N
Backward Elimination Selection	N	N	N	Y(2)	N	Y(4)	N	N	N	Y(1)	Y(3)	N
SFS Linear Regression	N	Y(4)	Y(3)	N	N	Y(2)	N	N	N	Y(1)	N	N
SFS KNN	Y(4)	Y(3)	N	Y(2)	N	N	N	N	Y(1)	N	N	N
RFE	N	N	Y(4)	Y(3)	N	N	N	N	N	N	N	Y(2)
Random Forest Importance	Y(4)	N	N	N	Y(3)	Y(2)	N	Y(1)	N	N	N	N
Embedded LGB	Y(4)	Y(3)	N	Y(2)	N	N	N	N	N	N	Y(1)	N
Total Score	12	10	7	9	6	12	2	2	2	3	7	3

Table 4. Performance evaluation of different feature selection techniques.

Feature Selection Technique	Features Selected	ML Algorithm	MAE	MSE	RMSE
FST/BET (W)	HVRYF, "MWS", "LMIN", "MEVP"	LR	2.38	5.87	0.21
		DT	0.21	0.05	0.21
		RF	0.23	0.06	0.25
Co-relation (F)	"HVRYF", "TMRF", "Area", "RD"	LR	0.95	1.17	1.08
		DT	0.45	0.37	0.61
		RF	0.27	0.11	0.33
Chi-Square (F)	"MMIN", "TMRF", "MEVP", "PRCP"	LR	0.21	0.06	0.24
		DT	0.33	0.16	0.41
		RF	0.22	0.08	0.28
SFS LR (W)	"HMAX", "MMIN", "HVRYF", "MEVP"	LR	0.95	0.93	0.97
		DT	0.13	0.02	0.16
		RF	0.14	0.03	0.18
SFS KNN (W)	"MMAX", "HMAX", "LMIN", "RD"	LR	0.43	0.24	0.49
		DT	0.29	0.12	0.34
		RF	0.3	0.12	0.32
RFE (W)	"MMIN", "LMIN", "PRCP", "HMD"	LR	0.61	0.44	0.67

Continued

			DT	0.45	0.23	0.48
			RF	0.46	0.22	0.47
			LR	0.62	0.41	0.64
RFI (E)	"MMAX", "TMRF", "HVYRF", "AREA"		DT	0.43	0.39	0.62
			RF	0.31	0.12	0.34
			LR	0.18	0.05	0.23
Embedded LGB (E)	"MMAX", "HMAX", "MMIN", "LMIN"		DT	0.31	0.23	0.48
			RF	0.27	0.1	0.32
			LR	0.3	0.15	0.38
Rank-based Technique	MMAX, HMAX, TRMF, HVRF		DT	0.66	0.5	0.71
			RF	0.35	0.13	0.36
			LR	0.45	0.25	0.5
Weighted Rank Based Technique	"MMAX", "HMAX", "LMIN", "HVRF"		DT	0.45	0.25	0.53
			RF	0.42	0.25	0.5

Table 5. (a). W2FE rank-based feature selection techniques. (b). W2FE weighted rank-based feature selection techniques.

(a)									
Feature Selection Technique	MMAX	HMAX	MMIN	LMIN	HVYRF	TMRF	MEVP	MWS	PRCP
Embedded LGB (E)	Y	Y	Y	Y	-	-	-	-	-
SFS LR (W)	-	Y	Y	-	Y	-	Y	-	-
Chi-Square (F)	-	-	Y	-	-	Y	Y	-	Y
FST/BET (W)	-	-	-	Y	Y	-	Y	Y	-
Total Rank	1	2	3	2	2	1	3	1	1
(b)									
Feature Selection Technique	MMAX	HMAX	MMIN	LMIN	HVYRF	TMRF	MEVP	MWS	PRCP
Embedded LGB (E)	Y(4)	Y(3)	Y(2)	Y(1)	-	-	-	-	-
SFS LR (W)	-	Y(4)	Y(3)	-	Y(2)	-	Y(1)	-	-
Chi-Square (F)	-	-	Y(4)	-	-	Y(3)	Y(2)	-	Y(1)
FST/BET (W)	-	-	-	Y(2)	Y(4)	-	Y(1)	Y(3)	-
Total Weighted Rank	4	7	9	3	6	3	4	3	1

Features selected by Embedded LGB (E), SFS LR (W), Chi-Square (F), FST/BET (W) methods have been integrated and Rank-based and Weighted Rank-based techniques have been applied on the given data. **Table 5(a)** and **Table 5(b)** show the result of W2FE rank-based feature selection technique and W2FE Weighted Rank-based Feature Selection Technique.

The Rank-based Selective Feature Selection Techniques returns HMAX, MMIN, LMIN, HVRF, MEVP as optimal feature set.

The Weighted Rank-based Selective Feature Selection Techniques returns MMIN, HMAX, HVYRF as optimal feature set.

Table 6. Performance evaluation of W2FE rank-based and weighted rank-based feature selection techniques.

Feature Selection Technique	Features Selected	ML Techniques	MAE	MSE	RMSE
W2FE Rank-Based Feature Selection Technique	HMAX,	LR	1.15	1.5	1.22
	MMIN,				
	LMIN	DT	0.23	0.12	0.34
	HVRF, MEVP	RF	0.31	0.12	0.35
W2FE Weighted Rank-Based Feature Selection Technique		LR	0.3	0.15	0.38
	HMAX,				
	MMIN,	DT	0.22	0.08	0.29
	HVRF	RF	0.18	0.05	0.22

Table 6 shows the performance evaluation of W2FE Rank-based and Weighted Rank-based Feature Selection Techniques.

After evaluation these techniques it is found that W2FE Weighted Rank-based Feature technique performed well as compared to the W2FE Rank-based Selection Technique.

4. Conclusion

In this paper, different types of feature selection techniques have been implemented and performance of those techniques have been analyzed by using different machine learning models. This research work used a two-stage feature selection approach which is integration of efficient feature selection techniques which has been found after analyzing performance of basic feature selection techniques and rank-based techniques. In W2FE Rank-based Feature Selection Technique and Weighted Rank-based Feature Selection Techniques, a filter, wrapper and embedded methods are used to determine features which are mostly selected by different feature selection techniques. These feature candidate sets are further pruned by using rank-based techniques. It is found that by exploiting strengths of the different methods and creating fusion of different methods allow better accuracy and stability than relying on any single feature selection method. The paper analyzed performance of two feature elimination method namely W2FE Weighted Rank-based Selective Feature Selection Techniques and W2FE Rank-based Feature Selection Techniques and found that

Weighted Rank-based Selective Feature Selection Techniques gives better results as compared to Rank-based Selective Feature Selection Techniques.

Conflicts of Interest

The authors declare no conflicts of interest regarding the publication of this paper.

References

- [1] Wang, X., Chen, R., Yan, F., Zeng, Z. and Hong, C. (2019) Fast Adaptive K-Means Subspace Clustering for High-Dimensional Data. *IEEE Access*, **7**, 42639-42651. <https://doi.org/10.1109/access.2019.2907043>
- [2] Jaiswal, J.K. and Samikannu, R. (2017) Application of Random Forest Algorithm on Feature Subset Selection and Classification and Regression. 2017 *World Congress on Computing and Communication Technologies*, Tiruchirappalli, 2-4 February 2017, 65-68. <https://doi.org/10.1109/wccct.2016.25>
- [3] Blum, A.L. and Langley, P. (1997) Selection of Relevant Features and Examples in Machine Learning. *Artificial Intelligence*, **97**, 245-271. [https://doi.org/10.1016/s0004-3702\(97\)00063-5](https://doi.org/10.1016/s0004-3702(97)00063-5)
- [4] Chen, C., Tsai, Y., Chang, F. and Lin, W. (2020) Ensemble Feature Selection in Medical Datasets: Combining Filter, Wrapper, and Embedded Feature Selection Results. *Expert Systems*, **37**, e12553. <https://doi.org/10.1111/exsy.12553>
- [5] Hancer, E., Xue, B. and Zhang, M. (2020) A Survey on Feature Selection Approaches for Clustering. *Artificial Intelligence Review*, **53**, 4519-4545. <https://doi.org/10.1007/s10462-019-09800-w>
- [6] Cekik, R. and Uysal, A.K. (2020) A Novel Filter Feature Selection Method Using Rough Set for Short Text Data. *Expert Systems with Applications*, **160**, Article 113691. <https://doi.org/10.1016/j.eswa.2020.113691>
- [7] Karasu, S., Altan, A., Bekiros, S. and Ahmad, W. (2020) A New Forecasting Model with Wrapper-Based Feature Selection Approach Using Multi-Objective Optimization Technique for Chaotic Crude Oil Time Series. *Energy*, **212**, Article 118750. <https://doi.org/10.1016/j.energy.2020.118750>
- [8] Ren, K., Fang, W., Qu, J., Zhang, X. and Shi, X. (2020) Comparison of Eight Filter-Based Feature Selection Methods for Monthly Streamflow Forecasting—Three Case Studies on CAMELS Data Sets. *Journal of Hydrology*, **586**, Article 124897. <https://doi.org/10.1016/j.jhydrol.2020.124897>
- [9] Kurniawan, Y.I., Cahyono, T., Nofiyati, Maryanto, E., Fadli, A. and Indraswari, N.R. (2020) Preprocessing Using Correlation Based Features Selection on Naive Bayes Classification. *IOP Conference Series: Materials Science and Engineering*, **982**, Article 012012. <https://doi.org/10.1088/1757-899x/982/1/012012>
- [10] Gholamnezhad, P., Broumandnia, A. and Seydi, V. (2020) An Inverse Model-Based Multi-Objective Estimation of Distribution Algorithm Using Random-Forest Variable Importance Methods. *Computational Intelligence*, **38**, 1018-1056. <https://doi.org/10.1111/coin.12315>
- [11] Folli, G.S., Nascimento, M.H.C., de Paulo, E.H., da Cunha, P.H.P., Romão, W. and Filgueiras, P.R. (2020) Variable Selection in Support Vector Regression Using Angular Search Algorithm and Variance Inflation Factor. *Journal of Chemometrics*, **34**, e3282. <https://doi.org/10.1002/cem.3282>

- [12] Maya Gopal, P.S. and Bhargavi, R. (2019) Performance Evaluation of Best Feature Subsets for Crop Yield Prediction Using Machine Learning Algorithms. *Applied Artificial Intelligence*, **33**, 621-642. <https://doi.org/10.1080/08839514.2019.1592343>
- [13] Maya Gopal, P.S. and Bhargavi, R. (2019) Selection of Important Features for Optimizing Crop Yield Prediction. *International Journal of Agricultural and Environmental Information Systems*, **10**, 54-71. <https://doi.org/10.4018/ijaeis.2019070104>
- [14] Karimi, Z., Mansour Riahi Kashani, M. and Harounabadi, A. (2013) Feature Ranking in Intrusion Detection Dataset Using Combination of Filtering Methods. *International Journal of Computer Applications*, **78**, 21-27. <https://doi.org/10.5120/13478-1164>
- [15] Chouhan, S., Singh, D. and Singh, A. (2016) An Improved Feature Selection and Classification Using Decision Tree for Crop Datasets. *International Journal of Computer Applications*, **142**, 5-8. <https://doi.org/10.5120/ijca2016909966>
- [16] Akhiat, Y., Chahhou, M. and Zinedine, A. (2019) Ensemble Feature Selection Algorithm. *International Journal of Intelligent Systems and Applications*, **11**, 24-31. <https://doi.org/10.5815/ijisa.2019.01.03>
- [17] Mariammal, G., Suruliandi, A., Raja, S.P. and Poongothai, E. (2021) Prediction of Land Suitability for Crop Cultivation Based on Soil and Environmental Characteristics Using Modified Recursive Feature Elimination Technique with Various Classifiers. *IEEE Transactions on Computational Social Systems*, **8**, 1132-1142. <https://doi.org/10.1109/tcss.2021.3074534>
- [18] Suruliandi, A., Mariammal, G. and Raja, S.P. (2021) Crop Prediction Based on Soil and Environmental Characteristics Using Feature Selection Techniques. *Mathematical and Computer Modelling of Dynamical Systems*, **27**, 117-140. <https://doi.org/10.1080/13873954.2021.1882505>
- [19] Gonzalez-Sanchez, A., Frausto-Solis, J. and Ojeda-Bustamante, W. (2014) Attribute Selection Impact on Linear and Nonlinear Regression Models for Crop Yield Prediction. *The Scientific World Journal*, **2014**, 1-10. <https://doi.org/10.1155/2014/509429>
- [20] Hsu, H., Hsieh, C. and Lu, M. (2011) Hybrid Feature Selection by Combining Filters and Wrappers. *Expert Systems with Applications*, **38**, 8144-8150. <https://doi.org/10.1016/j.eswa.2010.12.156>
- [21] Dewi, C. and Chen, R.C. (2019) Random Forest and Support Vector Machine on Features Selection for Regression Analysis. *International Journal of Innovative Computing, Information & Control*, **15**, 2027-2037.
- [22] Pudjihartono, N., Fadason, T., Kempa-Liehr, A.W. and O'Sullivan, J.M. (2022) A Review of Feature Selection Methods for Machine Learning-Based Disease Risk Prediction. *Frontiers in Bioinformatics*, **2**, Article 927312. <https://doi.org/10.3389/fbinf.2022.927312>
- [23] Suresh Sumi, M.S. and Narayanan, A. (2019) Improving Classification Accuracy Using Combined Filter+Wrapper Feature Selection Technique. 2019 *IEEE International Conference on Electrical, Computer and Communication Technologies*, Coimbatore, 20-22 February 2019, 1-6.
- [24] Raja, S.P., Sawicka, B., Stamenkovic, Z. and Mariammal, G. (2022) Crop Prediction Based on Characteristics of the Agricultural Environment Using Various Feature Selection Techniques and Classifiers. *IEEE Access*, **10**, 23625-23641. <https://doi.org/10.1109/access.2022.3154350>
- [25] Thomas, R.N. and Gupta, R. (2020) Feature Selection Techniques and Its Importance in Machine Learning: A Survey. 2020 *IEEE International Students' Con-*

- ference on Electrical, Electronics and Computer Science*, Bhopal, 22-23 February 2020, 1-6. <https://doi.org/10.1109/sceecs48394.2020.189>
- [26] Parmar, K.P. and Bhatt, T. (2022) Crop Yield Prediction Based on Feature Selection and Machine Learners: A Review. *Proceedings of the Second International Conference on Artificial Intelligence and Smart Energy*, Coimbatore, 23-25 February 2022, 354-358. <https://doi.org/10.1109/ICAIS53314.2022.9742891>
- [27] Bouchlaghem, Y., Akhiat, Y. and Amjad, S. (2022) Feature Selection: A Review and Comparative Study. *E3S Web of Conferences*, **351**, Article No. 01046. <https://doi.org/10.1051/e3sconf/202235101046>
- [28] Poornima, K.A. and Dheepa, G. (2022) An Efficient Feature Selection and Classification for the Crop Field Identification: A Hybridized Wrapper Based Approach. *Turkish Journal of Computer and Mathematics Education*, **13**, 241-254.