JAMBURA JOURNAL OF MATHEMATICS
Jambura J. Math. Vol. 5, No. 2, pp. 351–362, August 2023
Journal Homepage: http://ejurnal.ung.ac.id/index.php/jjom
DOI: https://doi.org/10.34312/jjom.v5i2.20397

Unsupervised Feature Selection Based on Self-configuration Approaches using Multidimensional Scaling

Ridho Ananda[1,*], Atika Ratna Dewi[2], Maifuza Binti Mohd Amin[3], Miftahul Huda[4], Gushelmi[5]

[1]*Program Studi Teknik Industri, Institut Teknologi Telkom Purwokerto, Purwokerto, Indonesia*
[2]*Program Studi Data Sains, Institut Teknologi Telkom Purwokerto, Purwokerto, Indonesia*
[3]*Politeknik Port Dickson, Negeri Sembilan, Malaysia*
[4]*Program Studi Statistika, Universitas Bina Bangsa, Banten, Indonesia*
[5]*Universitas Putra Indonesia YPTK, Padang, Indonesia*

[*]*Corresponding author. Email: ridho@ittelkom-pwt.ac.id*

ABSTRACT

Some researchers often collect features so the principal information does not lose. However, many features sometimes cause problems. The truth of analysis results will decrease because of the irrelevant or repetitive features. To overcome it, one of the solutions is feature selection. They are divided into two, namely supervised and unsupervised learning. In supervised, the feature selection can only be carried out on data containing labels. Meanwhile, in unsupervised, there are three approaches correlation, configuration, and variance. This study proposes an unsupervised feature selection by combining correlation and configuration using multidimensional scaling (MDS). The proposed algorithm is MDS-Clustering, which uses hierarchical and non-hierarchical clustering. The result of MDS-clustering is compared with the existing feature selection. There are three schemes in the comparison process, namely, 75%, 50%, and 25% feature selected. The dataset used in this study is the UCI dataset. The validities used are the goodness-of-fit of the proximity matrix (GoFP) and the accuracy of the classification algorithm. The comparison results show that the feature selection proposed is certainly worth recommending as a new approach in the feature selection process. Besides, on certain data, the algorithm can outperform the existing feature selection.

Keywords:
Feature Selection; Multidimensional Scaling; Clustering

Citation Style:

R. Ananda, *et al.*, "Unsupervised Feature Selection Based on Self-configuration Approaches using Multidimensional Scaling", *Jambura J. Math.*, vol. 5, No. 2, pp. 351–362, 2023, doi: https://doi.org/10.34312/jjom.v5i2.20397

1. Introduction

In quantitative research, the conclusion of the phenomenon observed emerges from the results analysis of information that affects the phenomenon. Information and phenomena are said features. In research, some researchers often collect many features so that the principal information do not lose. However, a large number of features sometimes gives a new problem. The truth of the analysis results will decrease because there are irrelevant

e-ISSN: 2656-1344
Received: 2023-06-12 | Revised: 2023-07-07 | Accepted: 2023-07-11| Published: 2023-08-01

or repetitive features [1]. One of the solutions that can be used to overcome this problem is the application of the dimensionality reduction method.

Dimensionality reduction is the pre-processing step to take irrelevant or repetitive features away so that it can improve the learning feature's accuracy and reduce the training time. In dimensionality reduction techniques, feature selection and feature extraction were the methods that had been proposed and used in general [2]. Feature selection methods select significant features and remove the other feature from the dataset, whereas feature extraction transforms the existing features into new features with the lower dimension. If the aim is to find out significant features, then it should be used in feature selection.

The algorithms for feature selection can be classified into supervised and unsupervised algorithms. Supervised algorithms integrate the labels of the objects into the selection process. Meanwhile, in unsupervised algorithms, the process is done based on the features without including labels [3, 4]. If the label is basically unknown on the dataset object, then only unsupervised algorithms can be used in the selection process. There are four approaches to the supervised algorithm, namely wrapper [5], tree [6], statistical test [7], and filter to get the most relevant feature [8]. While unsupervised is based on correlation [9, 10], suitability of object configuration [11], and variance [12]. The unsupervised algorithm based on correlation using principal component analysis (PCA) has high time complexity similarly, the approach to the suitability of the configuration using Procrustes analysis. Still based on correlation with hierarchical clustering method has also been proposed. This method yields lower time complexity but is not as effective as PCA results. At the same time, the variance approach has been proposed as a variance threshold algorithm. The algorithm eliminates features below a certain threshold value of variance without involving correlation between features. In this algorithm, the determination of the threshold should be the researcher's attention.

Based on the description above, this study proposes a new approach for doing a feature selection process based on configuration. The configuration of features is obtained from the correlation matrix by using the multidimensional scaling (MDS) algorithm. The results of the configuration features are then grouped using the clustering algorithm. The advantage of this algorithm is that the clustering process can be generalized hierarchically or non-hierarchically. In addition, variance and the correlation between features are also involved in the selection process. With these characteristics, it is hoped that the feature selection process will be more effective and efficient. The data used in this study is the UCI datasets. The results of the feature selection are then compared with several supervised and unsupervised algorithms that already exist. The performance indicators used are the size of the goodness-of-fit of the object configuration, time complexity, and accuracy of the classification results.

This paper is arranged as follows. Section II describes a brief of principal coordinate analysis and the research methods used. Section III describes the results and discussion. In the last section, the conclusion and suggestions are described.

2. Methods

2.1. A Brief of Multidimensional Scaling

Multidimensional scaling (MDS) is an analysis that can be used to represent proximities among objects or find their configuration in a low-dimensional space [13]. Based on the

characteristic of proximities used, MDS is divided into metric and ordinal scaling. Metric scaling was proposed by Young and Householder, who showed how with the proximities matrix in Euclidean space [14]. The configuration of points can be obtained where it is still maintaining the distance. This problem was later popularized by Gower under the name of principal coordinate analysis (PCoA) [15]. The ordinal scaling procedure was first introduced by Kruskal [16]. For example, we will determine the configuration of n objects from the proximities matrix of these objects in k dimensions by using MDS. The procedure of MDS is shown in the following steps:

1. Suppose that $\mathbf{D} = \left(\delta_{ij}\right)$ is a proximities matrix where δ_{ij} is the value of the proximities between i-th object and j-th object for every i, j.
2. Construct $\mathbf{A} = \left(a_{ij}\right)$ where $a_{ij} = -\frac{1}{2}\delta_{ij}^2$ for every i, j.
3. Compute $\mathbf{B} = \left(b_{ij}\right)$ where

$$\begin{aligned} b_{ij} =& a_{ij} - \overline{a}_{i.} - \overline{a}_{.j} + \overline{a}_{..} \\ \overline{a}_{i.} =& \sum_{j=1}^{n} \frac{a_{ij}}{n};\ \overline{a}_{.j} = \sum_{i=1}^{n} \frac{a_{ij}}{n};\ \text{and } \overline{a}_{..} = \sum_{i=1}^{n}\sum_{j=1}^{n} \frac{a_{ij}}{n}. \end{aligned}$$

4. Use the spectral decomposition process to get factorization of $\mathbf{B}$ as shown in Equation (1).

$$\mathbf{B} = \mathbf{V}\mathbf{\Lambda}\mathbf{V}' = \mathbf{V}\mathbf{\Lambda}^{1/2}\mathbf{\Lambda}^{1/2}\mathbf{V}' \tag{1}$$

   $\mathbf{V}$ and $\mathbf{\Lambda}$ are the matrix of eigenvector of $\mathbf{B}$ and the diagonal matrix of eigenvalue of $\mathbf{B}$ respectively. If the rank of $\mathbf{B}$ is q then $\mathbf{\Lambda} = \text{diag}(\lambda_1, \lambda_2, \ldots, \lambda_q)$ where $i < j \Rightarrow \lambda_i \geq \lambda_j > 0$, $\forall i, j$ and $\mathbf{V} = \begin{pmatrix} \mathbf{v}_1 & \mathbf{v}_2 & \ldots & \mathbf{v}_q \end{pmatrix}$.
5. Let ${}_n\mathbf{Z}_q = \mathbf{V}\mathbf{\Lambda}^{1/2}$ so we get $\mathbf{Z}$ matrix where the proximity of two arbitrary objects in $\mathbf{Z}$, for instance objects i and j, similar to δ_{ij} for every i, j.
6. It show that ${}_n\mathbf{Z}_q$ is the configuration object of $\mathbf{D}$ in q dimension.

If we prefer a small dimension k, $k \ll q$, we can use the first k eigenvalues and corresponding eigenvectors where the percentage of k eigenvalues is expected to meet Equation (2) [17] as follows:

$$\frac{\sum_{i=1}^{k} \lambda_i}{\sum_{i=1}^{q} \lambda_j} \times 100\% \geq 75\%. \tag{2}$$

2.2. MDS-Clustering Algorithm

The basic idea of the proposed algorithm is initiated from how to perform the feature selection process using hierarchical and non-hierarchical clustering algorithms in general. Based on the objective, configuration information is certainly needed from features in the dataset. By assuming the correlation value is a measure of similarity in [0,1], the direction of the correlation between features is ignored, so a proximity matrix between components can be obtained. The matrix can be used to obtain configurations between features by using the MDS algorithm. Furthermore, each feature can be clustered based on the proximity of its configuration.

The clustering algorithm is one of the data mining algorithms where objects are clustered by maximizing similarity measures among objects in the same cluster and

minimizing similarity among objects in different clusters [18]. The clustering procedure is generally divided into hierarchical and non-hierarchical clustering algorithms. The hierarchical clustering algorithms are single linkage, complete linkage, average linkage, and ward linkage. Meanwhile, in non-hierarchical clustering, the k-means algorithm is the most popular and widely applied in various fields like education [19–21], the problems in incomplete data [22], and the problems in shape data [23].

The advantage of obtaining the feature configuration, the performance improvement algorithm from k-means can be applied in the feature selection process, for instance, the algorithms for determining the optimal k cluster. The procedure of the MDSclustering algorithm is described in the following steps:

1. Suppose that $\mathbf{X} = (x_{ij})$ n-by-p is a matrix dataset whose n objects and p features.
2. Calculate the correlation matrix $\mathbf{R}= (|r_{ij}|)$ for the features in the dataset where the direction of correlation is ignored. So, $|r_{ij}| \in [0,1]$ where $r_{ij} \approx 0$ means there is almost no correlation between i-th and j-th features. Conversely, if $r_{ij} \approx 1$ then there is strong correlation between them.
3. Transform the matrix $\mathbf{R}$ into a proximity matrix $\mathbf{D}$ using Equation (3).

$$\mathbf{D} = \mathbf{J} - \mathbf{R}. \tag{3}$$

   $\mathbf{J}$ is a unit matrix with size $p \times p$. $\mathbf{D} = (d_{ij})$ is the proximity matrix of features where $d_{ij} \in [0,1]$. $d_{ij} \approx 0$ means the proximity between the two features is strong which means the correlation between them is also it. Meanwhile, if $d_{ij} \approx 1$ then it is the opposite condition from $d_{ij} \approx 0$.
4. Using the MDS algorithm to obtain the configuration of features, $\mathbf{Z}$ matrix, from proximity matrix $\mathbf{D}$.
5. By using the $\mathbf{Z}$ matrix, the clustering algorithm is processed. If we would like to select k features, then k clusters must be gained in the clustering process. Furthermore, one feature with the most significant variance is selected for each cluster, and the others are deleted.

This study used hierarchical and nonhierarchical clustering in the algorithm proposed. The clustering algorithms used are single linkage, complete linkage, average linkage, ward linkage, and k-means. Then, the results will be compared with the unsupervised feature selection on labeled and unlabeled data and the supervised feature selection on labeled data.

2.3. Unsupervised Feature Selection

The unsupervised method has three approaches in the feature selection process, namely the correlation approach, object configuration, and variance. Unsupervised feature selection algorithms with a correlation approach are B2, B4, and Clustering [9, 10]. Feature selection with a configuration approach utilizes Procrustes analysis [11]. While the selection of features with the variance approach uses the variance threshold [12]. In this study, because the feature selection process is based on the number of selected features, the algorithm that utilizes the variance threshold is not included in the comparison process.

2.4. Supervised Feature Selection

In the supervised feature selection algorithms, four approaches can be used, namely wrapped, intrinsic, filter based on statistical tests, and filter based on the measure of the feature importance. If a wrapped approach is chosen, one of the algorithms that can be used is recursive feature elimination (RFE) [24]. Meanwhile, if the intrinsic approach is selected, the decision tree algorithm is one of the algorithms in this approach [25]. It is different if we choose a filter approach based on statistical tests, we can utilize the Anova algorithm [26]. The last approach is a filter based on the measure of the feature importance, which can employ the information gain algorithm [27]. Those supervised algorithms will be utilized to be compared with the algorithm proposed.

2.5. Feature Selection Validity

Feature selection validity is the technique to justify whether a certain algorithm has higher performance than others in the feature selection process. Siswadi et al. suggest an efficiency score that is measured according to the goodness-of-fit of Procrustes (GoFP) by the matrices [11]. Suppose that $\mathbf{X}$ is the initial data matrix and $\mathbf{X}_q$ is the data by keeping q variables of $\mathbf{X}$. We define $\mathbf{D}$ and $\mathbf{D}_q$ that are proximity matrices from $\mathbf{X}$ and $\mathbf{X}_q$, respectively. Then, the GoFP is calculated according to Equation (4) [28].

$$\text{GoFP}\left(\mathbf{D}_q, \mathbf{D}\right) = \left(\sum_{i=1}^{r} \sigma_{ii}\right)^2. \tag{4}$$

r and σ_{ii} are rank and singular values respectively from the GoFP processed. The measure is at the interval [0,1]. The closer to 1, it shows that the higher the compatibility of $\mathbf{D}_q$ to $\mathbf{D}$. It means that the information of proximity matrix from $\mathbf{X}$ is not lost significantly because of the feature selection. On labeled data, the GoFP size can still be used regardless of the label on the data.

On the other hand, the validity of feature selection can also be gained through the accuracy result of the classification algorithms in the labeled data [3]. Still referring to the previous information, suppose that $\mathbf{X}_q$ is data after being taken q variables of $\mathbf{X}$ and corresponding to k that is a vector of the data label. The feature selection process has good quality if the results of the classification algorithm using $\mathbf{X}_q$ have high accuracy. This study uses the K-Nearest Neighbor (KNN) algorithm for the classification process. The KNN algorithm is popular enough in Machine Learning. And also, its classification process is based on the proximities of the object using their configuration [29].

2.6. Research Flow

The research flow used in this study consists of four main steps: preprocessing the data, processing the feature selection algorithms, calculating the feature selection validity, and comparing the validity results. The UCI datasets are used in this study. In the preprocessing stage, features containing missing values are deleted. Furthermore, especially for UCI data, a standardization process must be carried out because of the difference of units among attributes. Then the feature selection process is carried out using the proposed algorithm on the dataset as well as the unsupervised algorithms, which will be used as a comparison. Meanwhile, the supervised algorithm that has been determined will carry out the selection process on the labeled data. The results of the

feature selection are then determined by the value of GOFP and the accuracy of the classification results. The validity results are compared to know the performance of the proposed algorithm. In simple terms, the research is shown in Figure 1.

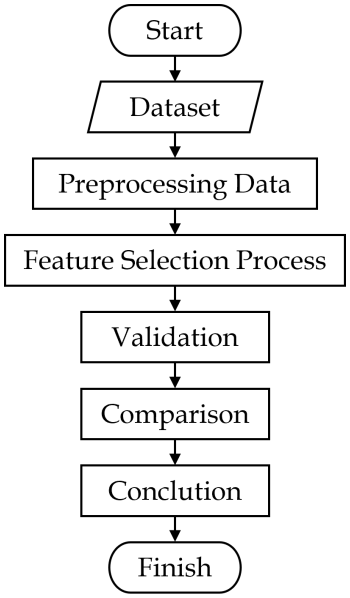

Figure 1. The used research flowchart

3. Results and Discussions

3.1. The Data Used and Preprocessing Data

The data used in this study are secondary data obtained from the UCI dataset, namely climate, E.coli, cortex nuclear, Parkinsons, wine red, wine white, and ionosphere. Those datasets have numerical attributes type. The description of the data is shown in Table 1. The table gives information about the number of objects, features, and classes in each data. As described in the research flow, before the feature selection algorithms are processed on the dataset, a preprocessing process is carried out for the data, namely standardization using the Z-score.

Table 1. The data used

Symbol	Dataset	n objects	n Features	n class
D1	Climate	540	19	2
D2	Ecoli	336	8	8
D3	Cortex Nuclear	552	78	8
D4	Parkinsons	195	17	2
D5	Wine Red	1599	12	6
D6	Wine White	4898	12	7
D7	Ionosphere	351	34	2

To get a good conclusion, in the feature selection process, several schemes are carried out. The first scheme is that 75% of the total number of features are selected, and the remains are deleted. The second scheme is that 50% of the features are selected. The last scheme is as many as 25% of the selected features. Table 2 shows the number of variables selected in the schema for each dataset.

Table 2. The number of features selected in each scheme

Dataset	*n* feature	75% selected	50% selected	25% selected
D1	19	15	9	5
D2	8	6	4	2
D3	78	59	39	20
D4	17	13	8	5
D5	12	9	6	3
D6	12	9	6	3
D7	34	26	17	9

3.2. Comparison of Feature Selection Algorithms

After the feature selection process is carried out on each scheme in the dataset used, the validity of feature selection algorithms is determined. Table 3 shows the GoFP of 75% of the data selected for each algorithm used in the dataset. The algorithm shown in Table 3 consists of the proposed feature selection algorithms and the existing variable selection algorithms. MDS-Kmeans is the proposed algorithm based on the configuration of the features, which in the selection process uses a nonhierarchical clustering approach, namely k-means. Meanwhile, MDS-WL, MDS-AL, MDS-SL, and MDS-CL are feature selection algorithms, but the selection uses hierarchical clustering, namely ward linkage, average linkage, single linkage, and centroid linkage, respectively. All of the proposed algorithms were compared with the unsupervised and supervised variable selection algorithms using clustering, B2, B4, Procrustes, ANOVA, RFE, Decision tree (DT), and information gain (IG) based on the validity used.

Table 3. The GoFP at 75% features selected

Algorithm	D1	D2	D3	D4	D5	D6	D7
MDS-kmeans	0.700	0.967	0.861	0.947	0.947	0.989	0.857
MDS-WL	0.703	0.965	0.923	0.947	0.992	0.908	0.859
MDS-AL	0.703	0.965	0.923	0.947	0.992	0.908	0.851
MDS-SL	0.686	0.965	0.923	0.947	0.992	0.990	0.842
MDS-CL	0.703	0.965	0.923	0.947	0.992	0.908	0.859
Clustering	0.710	0.953	0.880	0.947	0.947	0.990	0.848
B2	0.706	0.967	0.923	0.947	0.947	0.989	0.875
B4	0.711	0.879	0.923	0.947	0.947	0.989	0.846
Procrustes	0.705	0.967	0.923	0.947	0.947	0.989	0.877
Anova	0.724	0.803	0.932	0.405	0.997	0.990	0.865
RFE	0.718	0.762	0.932	0.405	0.481	0.999	0.886
DT	0.236	0.742	0.813	0.949	0.949	0.465	0.473
IG	0.710	0.668	0.932	0.966	0.996	0.990	0.890
Average	0.670	0.890	0.908	0.865	0.933	0.931	0.833
Stdev	0.131	0.108	0.036	0.204	0.138	0.144	0.109

The results shown in Table 3 provide information that the GoFP of the proximity matrix from the feature selection results is generally not much different. It is shown by the relatively small standard deviation value, namely, 0.131 on average. For the proposed algorithms, the majority of the GoFP obtained shows that those algorithms get good enough in the value of GoFP. In D2 data, namely E.coli data, MDS-Kmeans gives the highest GoFP results compared to other algorithms at 75% of the selected data. The black watermark indicates the maximum validity.

Furthermore, Table 4 shows the accuracy results of the KNN classification algorithm in the dataset with 75% of features selected. The table provides information that in all datasets, the performance of the proposed algorithm in the classification process gives quite good results. In dataset D2 until D7, some proposed algorithms give maximum accuracy in the classification process.

Table 4. The accuracy of classification at 75% of features selected

Algorithm	D1	D2	D3	D4	D5	D6	D7
MDS-kmeans	0.922	0.996	0.845	0.903	0.699	0.691	0.866
MDS-WL	0.919	0.995	0.896	0.903	0.707	0.694	0.852
MDS-AL	0.919	0.995	0.896	0.903	0.707	0.694	0.872
MDS-SL	0.926	0.995	0.896	0.903	0.707	0.685	0.869
MDS-CL	0.919	0.995	0.896	0.903	0.707	0.694	0.852
Clustering	0.928	0.993	0.890	0.903	0.700	0.685	0.858
B2	0.920	0.996	0.896	0.903	0.699	0.691	0.858
B4	0.915	0.993	0.896	0.903	0.699	0.691	0.858
Procrustes	0.935	0.996	0.896	0.903	0.699	0.691	0.846
Anova	0.944	0.946	0.824	0.744	0.481	0.438	0.859
RFE	0.963	0.901	0.824	0.744	0.441	0.450	0.817
DT	0.944	0.973	0.721	0.872	0.541	0.454	0.845
IG	0.889	0.730	0.824	0.821	0.513	0.439	0.761
Average	0.926	0.962	0.861	0.869	0.638	0.615	0.847
Stdev	0.018	0.075	0.053	0.061	0.103	0.118	0.029

In the 50% feature selected scheme, the results of the suitability measure of the proximity matrix using GoFP are shown in Table 5. In the table, it can be seen in the D1 or climate dataset that the proposed algorithms and the existing algorithms provide a relatively low GoFP value with an average below 0.5, namely, 0.471. Although low, the maximum GoFP value in D1 data is obtained by MSD-SL. Meanwhile, in datasets D2 to D7, the proposed algorithm gives a relatively good score, with the majority above the average. In D2 and D6 data, the MDS-Kmeans algorithm obtains the maximum value on the goodness-of-fit of the proximity matrix.

Table 5. The GoFP at 50% features selected

Algorithm	D1	D2	D3	D4	D5	D6	D7
MDS-kmeans	0.496	0.970	0.691	0.947	0.947	0.999	0.798
MDS-WL	0.473	0.967	0.772	0.947	0.992	0.908	0.794
MDS-AL	0.504	0.945	0.772	0.947	0.992	0.908	0.763
MDS-SL	0.510	0.942	0.772	0.947	0.992	0.908	0.720
MDS-CL	0.507	0.966	0.772	0.947	0.992	0.908	0.792
Clustering	0.506	0.942	0.697	0.947	0.993	0.915	0.764
B2	0.471	0.970	0.943	0.947	0.947	0.989	0.799
B4	0.491	0.876	0.939	0.947	0.947	0.989	0.776
Procrustes	0.474	0.970	0.881	0.947	0.947	0.989	0.843
Anova	0.493	0.762	0.856	0.405	0.949	0.990	0.734
RFE	0.502	0.722	0.571	0.008	0.005	0.009	0.792
DT	0.236	0.730	0.813	0.949	0.949	0.480	0.568
IG	0.462	0.609	0.365	0.024	0.482	0.468	0.761
Average	0.471	0.875	0.757	0.762	0.856	0.805	0.762
Stdev	0.072	0.124	0.156	0.363	0.290	0.300	0.066

Table 6 shows the results of the accuracy of the classification process at 50% of features selected using the feature selection algorithms. Based on the table, the classification process shows that the proposed algorithm provides maximum accuracy in the D2, D4, and D5 datasets, namely the Ecoli, Parkinsons, and wine red, respectively. Meanwhile, the accuracy results of the proposed algorithm are generally above average.

Table 6. The accuracy of classification at 50% of features selected

Algorithm	D1	D2	D3	D4	D5	D6	D7
MDS-kmeans	0.931	0.996	0.857	0.903	0.700	0.680	0.869
MDS-WL	0.915	0.993	0.887	0.903	0.706	0.684	0.877
MDS-AL	0.924	0.993	0.887	0.903	0.706	0.684	0.889
MDS-SL	0.930	0.993	0.887	0.903	0.706	0.68	0.877
MDS-CL	0.920	0.993	0.887	0.903	0.706	0.684	0.869
Clustering	0.919	0.993	0.824	0.903	0.692	0.693	0.866
B2	0.919	0.996	0.875	0.903	0.701	0.692	0.886
B4	0.915	0.993	0.89	0.903	0.701	0.692	0.875
Procrustes	0.913	0.996	0.842	0.903	0.701	0.692	0.877
Anova	0.954	0.964	0.735	0.744	0.466	0.439	0.915
RFE	0.954	0.901	0.735	0.692	0.544	0.447	0.859
DT	0.944	0.964	0.721	0.872	0.541	0.43	0.845
IG	0.880	0.694	0.662	0.667	0.444	0.407	0.817
Average	0.924	0.959	0.822	0.854	0.639	0.608	0.871
Stdev	0.020	0.084	0.08	0.089	0.101	0.123	0.023

Table 7. TThe GoFP at 25% features selected

Algorithm	D1	D2	D3	D4	D5	D6	D7
MDS-kmeans	0.268	0.968	0.495	0.767	0.995	0.915	0.641
MDS-WL	0.277	0.932	0.495	0.965	0.993	0.915	0.623
MDS-AL	0.309	0.947	0.495	0.965	0.994	0.915	0.589
MDS-SL	0.309	0.800	0.495	0.965	0.994	0.915	0.561
MDS-CL	0.279	0.948	0.743	0.965	0.993	0.915	0.609
Clustering	0.264	0.932	0.495	0.947	0.995	0.914	0.581
B2	0.270	0.98	0.759	0.947	0.949	0.990	0.618
B4	0.270	0.871	0.743	0.947	0.949	0.990	0.626
Procrustes	0.279	0.977	0.759	0.947	0.949	0.990	0.667
Anova	0.290	0.734	0.539	0.405	0.949	0.915	0.521
RFE	0.281	0.526	0.423	0.006	0.005	0.014	0.557
DT	0.236	0.744	0.813	0.949	0.949	0.465	0.436
IG	0.262	0.564	0.052	0.011	0.042	0.009	0.543
Average	0.276	0.84	0.562	0.753	0.827	0.759	0.582
Stdev	0.019	0.156	0.206	0.365	0.357	0.358	0.061

In 25% of selected features, as shown in Table 7, the quality of the goodness-of-fit of the proximity matrix using GoFP for all feature selection algorithms got almost the same value in datasets D1 and D7, namely climate and ionosphere, respectively. It is also known that the GoFP value decreased significantly in the D1 dataset. In the data, all of the feature selection algorithms used get low GoFP values. However, in other data, the proposed variable selection algorithm generally obtains a relatively high GoFP value. And then, in the D4 or Parkinson, the maximum GoFP is obtained by MDS-WL, MDS-

AL, MDS-SL, and MDS-CL. Meanwhile, in D5 or wine red dataset, the maximum GoFP is received by MDS-Kmeans.

Table 8. The accuracy of classification at 25% of features selected

Algorithm	D1	D2	D3	D4	D5	D6	D7
MDS-kmeans	0.915	0.991	0.807	0.892	0.687	0.682	0.909
MDS-WL	0.915	0.991	0.708	0.892	0.687	0.680	0.886
MDS-AL	0.926	0.986	0.708	0.892	0.706	0.678	0.903
MDS-SL	0.926	0.995	0.708	0.892	0.706	0.678	0.906
MDS-CL	0.917	0.993	0.848	0.892	0.687	0.680	0.903
Clustering	0.924	0.989	0.708	0.903	0.669	0.695	0.860
B2	0.915	0.995	0.815	0.903	0.670	0.690	0.889
B4	0.915	0.989	0.848	0.903	0.670	0.690	0.883
Procrustes	0.911	0.995	0.815	0.903	0.670	0.690	0.892
Anova	0.963	0.964	0.632	0.744	0.541	0.455	0.873
RFE	0.963	0.946	0.588	0.692	0.556	0.507	0.873
DT	0.944	0.973	0.721	0.872	0.541	0.454	0.873
IG	0.935	0.568	0.544	0.590	0.478	0.437	0.789
Average	0.276	0.84	0.562	0.753	0.827	0.759	0.582
Stdev	0.019	0.156	0.206	0.365	0.357	0.358	0.061

Table 8 shows the results of the accuracy of the classification process on the data that has been selected by the variables. The results of the table show that the performance of each variable selection algorithm is relatively the same. However, when examined in more detail, the proposed variable selection algorithm gets the maximum classification accuracy value on some data, namely D2, D3, D5, and D7. The best accuracy in D2, D3, D5 and D7 are obtained by MDS-SL, MDS-CL, MDS-AL, and MDS-kmeans respectively.

3.3. Discussion

Based on the goodness-of-fit of the proximity matrix, the proposed algorithm has a GoFP value that is relatively not much different from other algorithms. Even on some datasets, the proposed algorithm obtains the maximum GoFP. This is shown in the 75% feature selected scheme, where MDS-Kmeans get the highest score in the E. coli data. This also happens in the 50% selected variable scheme, where the MDS-Kmeans on the data also get the maximum value. While in the 25% feature selected scheme, there is an algorithm that provides a low value in the goodness-of-fit of the proximity matrix using GoFP. However, in the majority of datasets, the proposed algorithm scores fairly well.

While the accuracy results from the classification process, the proposed algorithm, in general, gets a pretty good score. In some schemes, it is also shown that the accuracy of one of these algorithms gets the maximum value compared to the existing feature selection algorithm. It shows that the proposed algorithm is superior to other algorithms on certain data.

4. Conclusion

This paper discussed the feature selection algorithms proposed based on self-configuration approaches using multidimensional scaling. The results of the configuration are then carried out by a feature selection process using hierarchical or nonhierarchical clustering procedures. The feature selection algorithms proposed are MDS-Kmeans, MDS-WL, MDS-AL, MDS-SL, and MDS-CL. Based on the comparison

among the algorithms proposed and the existing algorithms, it concluded that the proposed variable selection algorithm has a performance that is not much different from the previous algorithm. In certain datasets, the proposed algorithm is able to maintain the suitability of the proximity matrix of the data and has a high accuracy of classification results. Based on the results, the feature selection algorithms proposed certainly deserve to be recommended as a new approach in the feature selection process.

Acknowledgements

We thank to support of Ristekdikti Through a 2022 external grant for the financial support of the research.

References

[1] V. C. P and A. A. Chikkamannur, "Feature selection: An empirical study," *International Journal of Engineering Trends and Technology*, vol. 69, no. 2, pp. 165–170, 2021, doi: 10.14445/22315381/IJETT-V69I2P223.

[2] S. Velliangiri, S. Alagumuthukrishnan, and S. I. T. joseph, "A review of dimensionality reduction techniques for efficient computation," *Procedia Computer Science*, vol. 165, pp. 104–111, 2019, doi: 10.1016/j.procs.2020.01.079.

[3] S. Solorio-Fernández, J. A. Carrasco-Ochoa, and J. F. Martínez-Trinidad, "A review of unsupervised feature selection methods," *Artificial Intelligence Review*, vol. 53, no. 2, pp. 907–948, 2020, doi: 10.1007/s10462-019-09682-y.

[4] M. Kuhn and K. Johnson, *Applied Predictive Modeling*. Springer New York, 2013, doi: 10.1007/978-1-4614-6849-3.

[5] Y. Xue, L. Zhang, B. Wang, Z. Zhang, and F. Li, "Nonlinear feature selection using gaussian kernel svm-rfe for fault diagnosis," *Applied Intelligence*, vol. 48, no. 10, pp. 3306–3331, 2018, doi: 10.1007/s10489-018-1140-3.

[6] H. Rao, X. Shi, A. K. Rodrigue, J. Feng, Y. Xia, M. Elhoseny, X. Yuan, and L. Gu, "Feature selection based on artificial bee colony and gradient boosting decision tree," *Applied Soft Computing*, vol. 74, no. 10, pp. 634–642, 2019, doi: 10.1016/j.asoc.2018.10.036.

[7] Antonelli, Claggett, Henglin, Kim, Ovsak, Kim, Deng, Rao, Tyagi, Watrous, Lagerborg, Hushcha, Demler, Mora, Niiranen, Pereira, Jain, and Cheng, "Statistical workflow for feature selection in human metabolomics data," *Metabolites*, vol. 9, no. 7, p. 143, 2019, doi: 10.3390/metabo9070143.

[8] S. Hooker, D. Erhan, P.-J. Kindermans, and B. Kim, *A Benchmark for Interpretability Methods in Deep Neural Networks*. Red Hook, NY, USA: Curran Associates Inc., 2019.

[9] I. T. Jolliffe, "Discarding variables in a principal component analysis. i: Artificial data," *Applied Statistics*, vol. 21, no. 2, pp. 160–173, 1972, doi: 10.2307/2346488.

[10] ——, "Discarding variables in a principal component analysis. ii: Real data," *Applied Statistics*, vol. 22, no. 1, pp. 21–31, 1973, doi: 10.2307/2346300.

[11] Siswadi, A. Muslim, and T. Bakhtiar, "Variable selection using principal component and procrustes analysis and its application in educational data," *Journal of Asian Scientific Research*, vol. 2, no. 12, pp. 856–865, 2012, [online]. available: https://archive.aessweb.com/index.php/5003/article/view/3435.

[12] Y. Siti Ambarwati and S. Uyun, "Feature selection on magelang duck egg candling image using variance threshold method," in *2020 3rd International Seminar on Research of Information Technology and Intelligent Systems (ISRITI)*, 2020, pp. 694–699, doi: 10.1109/ISRITI51436.2020.9315486.

[13] P. Mair, I. Borg, and T. Rusch, "Goodness-of-fit assessment in multidimensional scaling and unfolding," *Multivariate Behavioral Research*, vol. 51, pp. 772–789, 2016, doi: 10.1080/00273171.2016.1235966.

[14] G. Young and A. S. Householder, "Discussion of a set of points in terms of their mutual distances," *Psychometrika*, vol. 3, no. 1, pp. 19–22, 1938, doi: 10.1007/BF02287916.

[15] J. C. Gower, "Some distance properties of latent root and vector methods used in multivariate analysis," *Biometrika*, vol. 53, no. 3/4, p. 325, 1966, doi: 10.2307/2333639.
[16] J. B. Kruskal, "Multidimensional scaling by optimizing goodness of fit to a nonmetric hypothesis," *Psychometrika*, vol. 29, no. 1, pp. 1–27, 1964, doi: 10.1007/BF02289565.
[17] I. T. Jolliffe, *Principal Component Analysis*. Springer-Verlag, 2002, doi: 10.1007/b98835.
[18] M. Pavithra and R.M.S.Parvathi, "A survey on clustering high dimensional data techniques," *International Journal of Applied Engineering Research*, vol. 12, no. 11, pp. 2893–2899, 2017.
[19] R. Ananda, "Silhouette density canopy k-means for mapping the quality of education based on the results of the 2019 national exam in banyumas regency," *Khazanah Informatika : Jurnal Ilmu Komputer dan Informatika*, vol. 5, no. 2, pp. 158–168, 2019, doi: 10.23917/khif.v5i2.8375.
[20] R. Ananda, M. Z. Naf'an, A. B. Arifa, and A. Burhanuddin, "Recommendation system for specialization selection using k-means density canopy," *Jurnal RESTI (Rekayasa Sistem dan Teknologi Informasi)*, vol. 4, no. 1, pp. 172–179, 2020, doi: 10.29207/resti.v4i1.1531.
[21] R. Adhitama, A. Burhanuddin, and R. Ananda, "Penentuan jumlah cluster ideal smk di jawa tengah dengan metode x-means clustering," *JIKO (Jurnal Informatika dan Komputer)*, vol. 3, no. 1, pp. 1–5, 2020, doi: 10.33387/jiko.v3i1.1635.
[22] R. Ananda and A. Z. Yamani, "Determination of initial k-means centroid in the process of clustering data evaluation of teaching lecturers," *Jurnal RESTI (Rekayasa Sistem dan Teknologi Informasi)*, vol. 4, no. 3, pp. 544–550, 2020, doi: 10.29207/resti.v4i3.1896.
[23] R. Ananda and A. Prasetiadi, "Hierarchical and k-means clustering in the line drawing data shape using procrustes analysis," *JOIV : International Journal on Informatics Visualization*, vol. 5, no. 3, p. 306, 2021, doi: 10.30630/joiv.5.3.532.
[24] A. Bahl, B. Hellack, M. Balas, A. Dinischiotu, M. Wiemann, J. Brinkmann, A. Luch, B. Y. Renard, and A. Haase, "Recursive feature elimination in random forest classification supports nanomaterial grouping," *NanoImpact*, vol. 15, p. 100179, 2019, doi: 10.1016/j.impact.2019.100179.
[25] S. Bashir, Z. S. Khan, F. Hassan Khan, A. Anjum, and K. Bashir, "Improving heart disease prediction using feature selection approaches," in *2019 16th International Bhurban Conference on Applied Sciences and Technology (IBCAST)*, 2019, pp. 619–623, doi: 10.1109/IBCAST.2019.8667106.
[26] Nurrahma and R. Yusuf, "Comparing different supervised machine learning accuracy on analyzing covid-19 data using anova test," in *2020 6th International Conference on Interactive Digital Media (ICIDM)*, 2020, pp. 1–6, doi: 10.1109/ICIDM51048.2020.9339676.
[27] E. O. Omuya, G. O. Okeyo, and M. W. Kimwele, "Feature selection for classification using principal component analysis and information gain," *Expert Systems with Applications*, vol. 174, p. 114765, 2021, doi: 10.1016/j.eswa.2021.114765.
[28] T. Bakhtiar and S. Siswadi, "On the symmetrical property of procrustes measure of distance," *International Journal of Pure and Apllied Mathematics*, vol. 99, no. 3, pp. 315–324, 2015, doi: 10.12732/ijpam.v99i3.7.
[29] R. Ananda and A. Prasetiadi, "Classification based on configuration objects by using procrustes analysis," *Jurnal Infotel*, vol. 13, no. 2, pp. 76–83, 5 2021, doi: 10.20895/infotel.v13i2.637.