Article

Evaluation of Feature Selection Methods on Psychosocial Education Data Using Additive Ratio Assessment

Fitriani Muttakin ¹, Jui-Tang Wang ^{2,*}, Mulyanto Mulyanto ² and Jenq-Shiou Leu ²

¹ Information Systems Department, Universitas Islam Negeri Sultan Syarif Kasim, Pekanbaru 28293, Indonesia; fitrianimuttakin@uin-suska.ac.id

² Electronic and Computer Engineering, National Taiwan University of Science and Technology, Taipei City 106335, Taiwan; d10602813@mail.ntust.edu.tw (M.M.); jsleu@mail.ntust.edu.tw (J.-S.L.)

* Correspondence: rtwang@mail.ntust.edu.tw

Abstract: Artificial intelligence, particularly machine learning, is the fastest-growing research trend in educational fields. Machine learning shows an impressive performance in many prediction models, including psychosocial education. The capability of machine learning to discover hidden patterns in large datasets encourages researchers to invent data with high-dimensional features. In contrast, not all features are needed by machine learning, and in many cases, high-dimensional features decrease the performance of machine learning. The feature selection method is one of the appropriate approaches to reducing the features to ensure machine learning works efficiently. Various selection methods have been proposed, but research to determine the essential subset feature in psychosocial education has not been established thus far. This research investigated and proposed methods to determine the best feature selection method in the domain of psychosocial education. We used a multi-criteria decision system (MCDM) approach with Additive Ratio Assessment (ARAS) to rank seven feature selection methods. The proposed model evaluated the best feature selection method using nine criteria from the performance metrics provided by machine learning. The experimental results showed that the ARAS is promising for evaluating and recommending the best feature selection method for psychosocial education data using the teacher's psychosocial risk levels dataset.

Keywords: evaluation feature selection; evaluation model; decision model; psychosocial education

Citation: Muttakin, F.; Wang, J.-T.; Mulyanto, M.; Leu, J.-S. Evaluation of Feature Selection Methods on Psychosocial Education Data Using Additive Ratio Assessment. *Electronics* **2022**, *11*, 114. <https://doi.org/10.3390/electronics11010114>

Academic Editors: Agnieszka Konys and Agnieszka Nowak-Brzezińska

Received: 14 November 2021

Accepted: 28 December 2021

Published: 30 December 2021

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Copyright: © 2021 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

Psychosocial education is multidisciplinary and covers a vast field of study. Therefore, it is not surprising that research in psychosocial education encompasses an abundance of environments and features that are logically expected to be linked to the problem-solving of educational quality improvements. Research from various perspectives, such as personal environment [1], family [2], nutrition [3], and physical activities [4], has been conducted to get an overview of the various psychosocial relationships in education. Accordingly, research linked to psychosocial education is categorized as one of the most active in education. Indeed, the search using the keyword “psychosocial education” in Google Scholar shows 212,000 results of research published between 2017 and 2021.

On the other hand, the success of artificial intelligence and big data influences decision-making perspectives, particularly those based on predictive problems. Big data can effectively handle more large-scale amounts, more complex varieties, and higher data dimensions [5]. Meanwhile, artificial intelligence, especially machine learning, significantly improves the quality of decision models [6,7]. These two factors encourage researchers to collect more data with massive features.

Theoretically, the more data that are collected, the more information that is obtained. The more information obtained, the better the prediction will be generated to be. However, the increase in the number of variables and the volume of data impacts data sparsity,

especially if the data quality is poor. The increase in sparsity makes it much more difficult to find data representative of the population. Furthermore, it makes machine learning challenging to generalize to the domain problem. The vague generalization will cause machine learning to lose its ability to adapt to new problems [8,9].

Instead of thrusting all features into machine learning, performing input feature optimization is often more efficient and effective. Feature selection can eliminate all features that are irrelevant to the prediction target. There have been various methods of selecting a feature that has been proposed and proven to impact machine learning performance. With many feature selection methodologies and different approaches in each method, it is relatively easy to raise a question about which method can give the optimum and effective results in machine learning, especially regarding the psychosocial education problem.

Hence, this paper proposed a methodology to evaluate the best feature selection method in the domain of psychosocial education. The evaluation was performed using a decision model approach that utilized multi-criteria decision making (MCDM). Furthermore, additive ratio assessment (ARAS) was adopted to evaluate and rank the best feature selection method. The evaluation and ranking used the metrics from the machine learning classification performance on the teacher's psychosocial risk level dataset.

2. Related Work

Feature selection is one of the critical stages in machine learning modeling, and the relevant feature has implications for better stability, robustness, and generalization of machine learning [10]. The feature selection method can be divided into three approaches [11–13]: filtering, wrapper, and embedded method.

Moorthy and Gandhi [14] previously conducted research using the filtering method. They optimized medical data using feature selection techniques for classification problems. They combined analysis of variance (ANOVA) and whale optimization (WO) to give a better result for SVM and k-NN classifiers than the one without ANOVA-WO. Ding and Li [15] also conducted a similar study identifying mitochondrial proteins in malaria by combining ANOVA and incremental feature selection (IFS) methods to find the most optimal feature. The proposed model achieved 97.1% accuracy compared to 92.0% on the comparison model. Next, Utama [16] performed feature selection using the mutual information (MI) model to predict the airline's tweet sentiment analysis. The feature selection made contributions to the classifier improvement.

Similarly, the wrapper method also gives promising results. Richhariya et al. [17] proposed a Universum support vector machine based on the recursive feature elimination (USVM-RFE) method to diagnose Alzheimer's. Feature selection was performed on the MRI data of brain tissue, and the classification using USVM-RFE showed better results than the one using SVR-RFE. The implementation of RFE was also done in the study [18], where RFE-SVM was used to determine the best feature among the various heart rate variability (HRV) data. The study showed that RFE-SVM could identify the HRV feature and detect the stress level better.

The approach using the embedded method has been widely used. Liu et al. [19] implemented feature selection using the embedded method. The implementation was performed during a cyberattack on the Internet of Things (IoT) data. The accuracy of the proposed method was relatively comparable to that of the comparison model. However, it was better in training speed, 1000 times faster than the overall features model. The implementation of the embedded method as the feature selection was also conducted by Loscalzo et al. [20]. Feature selection was used to remove unneeded input in robotic sensors. The paper showed that the embedding methodology significantly reduces unimportant sensors. Lastly, Liu et al. [21] compared embedding methodology to the others, such as Chi-Square, F-Statistic, and Gini Index. The experiment showed that the weighted Gini Index (WGI) method was better than the other methodologies on the data with limited features.

Given the importance of choosing a suitable feature selection method for the data characteristics of domain problems, selecting the best feature selection method is quite

challenging. There are various techniques for selecting feature selection methods, one of which is the decision system model approach. Kou et al. [22] conducted a study to select the best subset feature for a text classification case. The study compared several models from the MCDM, such as TOPSIS, GRA, WSM, VIKOR, and PROMOTHEE. The results showed that PROMOTHEE was better for evaluation models in the text-based classification case than other models. Hasemi et al. [23] proposed the EFS-MCDM method to determine the best feature on the computer network dataset. The features ranking in the EFS-MCDM delivered more optimal and efficient results in the measurement of accuracy, f-score, and run-time algorithm compared to other methods. Similarly, Singh [24] implemented TOPSIS to select features in the network traffic dataset. The research concluded that the classification model with the TOPSIS-based feature subset had the same accuracy yet much lower computation time.

Despite all the studies conducted on selecting existing feature selection methods so far, to the best of the authors' knowledge, there has been no study comparing and evaluating the best feature selection method to be implemented in psychosocial education. Previous studies on psychosocial education only implemented machine learning without extensive analysis of the used features.

Research Contribution

Based on the knowledge gaps derived from the previous studies, this paper would advance the body of knowledge about the feature selection method in two primary contributions:

1. This paper provides a systematic model for determining the best feature selection method using an adapted additive ratio assessment model [24]. Specifically, the selection of the feature selection method is implemented in the psychosocial education dataset.
2. This paper offers a comprehensive study and evaluation by comparing the performance of machine learning in every feature selection method. ARAS used the performance metrics from machine learning as criteria in determining the best feature selection method.

3. Methodology

3.1. Theoretical Overview

3.1.1. Artificial Intelligence Research on Psychosocial Education

Nowadays, the research in the education field focuses not only on the academic aspects, such as academic achievement, graduation level, academic grading, and teaching methods, but also on non-academic aspects, such as community relationships [25] and psychosocial. As such, the non-academic aspect also influences the quality of education [26–28].

On the other hand, the flourishing of research in artificial intelligence has made an impressive contribution to the psychosocial education field. Numerous artificial intelligence-based studies have successfully revealed psychosocial phenomena that influence education development. In a study conducted by Navarro [29], artificial intelligence was successfully used to predict the link between the condition of the environment and educators' stress levels. In addition, the research successfully interviewed and provided 4890 data points with 118 features used in predicting the level of stress on the educators. An extensive amount of data and high-dimensional features in the study indicated that psychosocial research is essential and exciting to be carried out.

3.1.2. Feature Selection Methods

Real-world problems are often represented by extensive data collection and high-dimensional features. Occasionally, existing features may not directly relate to the target problems that need to be solved [30,31]. Under such circumstances, the selection of features becomes critical. Selecting the right features makes it possible to improve model performance and efficiency in the computation process [32,33].

Three approaches are available to select features. The first approach, the filtering method, performs a selected subset of features based on the characteristics of the feature itself. The best feature is obtained from the statistical analysis of each feature with other features or target data. Next, the wrapper method uses machine learning to select the best data subsets for analysis. The wrapper method uses machine learning to reconstruct the feature subset and tests it using statistical modeling. The third approach, the embedded method, uses the same principle as the wrapper method, however, in evaluating the feature subset by analyzing the performance of machine learning.

Next, this section will briefly describe eight feature selection methods evaluated in this paper. There are three filtering methods: analyzing variance, mutual information, and chi-square; the exhaustive search feature is a wrapped method; embedding random forest, Lasso, and recursive feature elimination are embedded methods. Those methods would be compared to the models of machine learning using the baseline feature.

3.1.3. ANOVA

ANOVA is a statistical analysis used to calculate the distance of difference (variance) between two clusters [34]. ANOVA uses the *f-ratio* to calculate the magnitude of every feature and target class. Magnitude values above the *f-ratio* will be retained, and others will be discarded. In an ANOVA with class k , the variance among classes is defined as follows [35]:

$$\sigma_{v-all}^2 = \frac{\sum (\bar{x}_i - \bar{x})^2 n_i}{(k-1)} \quad (1)$$

where n_i is the value discovered from the calculation on the i -th class, $\bar{x}_i$ is the mean of the i -th class, and $\bar{x}$ is the mean of all classes; the class variance is defined as follows:

$$\sigma_{v-class}^2 = \frac{(\sum \sum (\bar{x}_{ij} - \bar{x})^2) - (\sum (\bar{x}_i - \bar{x})^2 n_i)}{(R-k)} \quad (2)$$

Then, the *f-ratio* is calculated based on the degree of the two variances:

$$f\text{-ratio} = \frac{\sigma_{v-all}^2}{\sigma_{v-class}^2} \quad (3)$$

3.1.4. Chi-Square

Chi-square is a statistical method that is widely used for calculating the correlation between two variables [36–38]. The implementation of Chi-square as the method to select subset features in machine learning can be done by calculating the dependency level of each feature towards the target data [39,40]. If n is the observed frequency and μ is the expected frequency, then the Chi-square (X^2) for a feature with a number of f and class C is defined as follows:

$$X^2 = \sum_{i=1}^f \sum_{j=1}^C \frac{(n_{ij} - \mu_{ij})^2}{\mu_{ij}} \quad (4)$$

3.1.5. Mutual Information (MI)

MI is used to calculate the distance of random vectors between clusters [41,42]. Mutual information looks for the similarity value between the distribution of probability $P(X, Y)$ and product of entropy $P(X)P(Y)$ [43]. Mutual information between two random vectors X and Y is defined as follows:

$$MI(X, Y) = \sum_{x \in X} \sum_{y \in Y} P(X = x, Y = y) \ln \frac{P(X = x, Y = y)}{P(X = x)P(Y = y)} \quad (5)$$

In feature selection problems, the implementation of mutual information was used to calculate the information value of how significant the contribution of a feature is towards

the prediction of the target class [44–46]. Mutual information for feature set S and m feature, which have a large dependence on the target class C , is defined as follows:

$$MI(S_m, C_i) = \frac{\log(P(S_m, C_i))}{P(S_m) * P(C_i)} \quad (6)$$

3.1.6. Exhaustive Search Feature (EFS)

In the EFS method, the algorithm performance is obtained by evaluating the existing features in all possible combinations. The feature subset with the highest performance will be selected [47,48]. EFS works by finding the value of validity (P, S) , assessing the entire subset of candidate feature S for a whole solution to a problem P . The result is obtained from Output (P, S) , in which the entire values of S are suitable for the problem P . The EFS method is a greedy algorithm, as it uses a brute force approach to find the best possible feature subset. Due to its exhaustive nature, EFS usually requires large amounts of resources.

3.1.7. Embedding Random Forest (ERF)

ERF is an ensemble method to reconstruct the average output of an individual tree [49]. The recursive approach is needed to find the best value from the feature subset during the elimination process, especially for highly correlated features [50]. Evaluating the high correlation can be done using the mean decreasing impurity approach. The Gini Index is one of the most popular measures of mean decreasing impurities, and it is defined as follows:

$$Gini = 1 - \sum_{i=1}^n (P_i)^2 \quad (7)$$

3.1.8. Lasso

The least absolute shrinkage and selection operator (Lasso) is one of the shrinkage techniques. Lasso selects the variables by minimizing the number of squared errors using penalty regularization [49,51]. Shrinkage regression is carried out towards zero along with the increase in the value of the lambda (λ) parameter used to control the number of shrinkages [52]. Lasso is defined as follows:

$$L_{lasso} = (\hat{\beta}) = \sum_{i=1}^n (y_i - x_i' \hat{\beta})^2 + \lambda \sum_{j=1}^m |\hat{\beta}_j| \quad (8)$$

3.1.9. Recursive Feature Elimination (RFE)

RFE is a feature selection method that works iteratively to rank features' importance [50]. In minimizing computational resources, some approaches eliminate instead of one by one but based on a subset of features [53]. An analysis and elimination process is performed in each iteration on the feature subsets with low relevance values. Two components of RFE are the number of features and the algorithm used to analyze the performance of the feature subsets. Generally, the iteration procedure of the RFE is performed as follows [54]: (1). Train each feature subset with a classifier, (2). Regarding the ranking of the feature subsets, calculate each feature subset's ranking, (3). Removing the feature subset that has low significance.

3.1.10. ARAS: Decision System Approach for the Feature Evaluation Method

Additive ratio assessment (ARAS) is one of the MCDM modeling techniques. ARAS is a method that relies on the intuitive principle that the best solution must have the largest ratio. Ranking using the ARAS method is performed by comparing the value of each criterion on each alternative by looking at its weight to obtain the ideal alternative [55,56].

The ARAS method utilizes a function value that determines the complexity of feasible alternatives. The ARAS method was directly proportional to the values and weights of the

main criteria considered to determine the best alternative. ARAS is based on the argument that complex problems can be understood simply by using relative comparisons. In ARAS, the ratio of the sum of normalized and weighted criteria values describes the possible alternatives to obtaining the optimal alternative rank. The ARAS method compares the utility functions of alternatives with optimal utility function values [57].

Like the classical MCDM approach, ARAS focuses on the ranking of criteria. Ranking with ARAS is done in several stages [55]. The first stage is forming a decision-making matrix. The matrix consists of $0 - m$ alternatives (rows) and $1 - n$ criteria (columns). If i represents the number of alternatives, j is the number of criteria. The decision-making matrix is denoted as follows:

$$X = \begin{bmatrix} x_{01} & \dots & x_{0j} & \dots & x_{0n} \\ \vdots & \ddots & \vdots & \ddots & \vdots \\ x_{i1} & \dots & x_{ij} & \dots & x_{in} \\ \vdots & \ddots & \vdots & \ddots & \vdots \\ x_{m1} & \dots & x_{mj} & \dots & x_{mn} \end{bmatrix}; i = \overline{0, m}; j = \overline{1, n}; \quad (9)$$

The x_{0j} optimal value of the criterion is the best value that can be used to represent the performance on each j criterion. In this paper, x_{0j} optimal criterion is defined as follows:

$$\begin{aligned} x_{0j} &= \max_i x_{ij}, \text{ if } \max_i x_{ij} \text{ is benefit;} \\ x_{0j} &= \min_i x_{ij}^*, \text{ if } \min_i x_{ij}^* \text{ is cost;} \end{aligned} \quad (10)$$

The next stage is normalizing all the criteria defined from $\bar{x}_{ij}$ of the matrix $\bar{X}$. The normalized decision-making matrix $\bar{X}$ is defined as follows:

$$\bar{X} = \begin{bmatrix} \bar{x}_{01} & \dots & \bar{x}_{0j} & \dots & \bar{x}_{0n} \\ \vdots & \ddots & \vdots & \ddots & \vdots \\ \bar{x}_{i1} & \dots & \bar{x}_{ij} & \dots & \bar{x}_{in} \\ \vdots & \ddots & \vdots & \ddots & \vdots \\ \bar{x}_{m1} & \dots & \bar{x}_{mj} & \dots & \bar{x}_{mn} \end{bmatrix}; i = \overline{0, m}; j = \overline{1, n}; \quad (11)$$

Normalization of benefits criteria can be done using the following formula:

$$\bar{x}_{ij} = \frac{x_{ij}}{\sum_{i=0}^m x_{ij}} \quad (12)$$

Meanwhile, normalization of cost criteria can be done using the normalized two-stage procedure following the notation:

$$\bar{x}_{ij} = \frac{1}{x_{ij}^*}; \bar{x}_{ij} = \frac{x_{ij}}{\sum_{i=0}^m x_{ij}} \quad (13)$$

The next step is defining the normalized-weighted matrix, starting with determining the value of w_j . The sum of weights of all the criteria is 1, and the weight w_j is limited as follows:

$$\sum_{j=1}^n w_j = 1 \quad (14)$$

After that, the normalized-weighted matrix is calculated using the following formula:

$$\hat{X} = \begin{bmatrix} \hat{x}_{01} & \dots & \hat{x}_{0j} & \dots & \hat{x}_{0n} \\ \vdots & \ddots & \vdots & \ddots & \vdots \\ \hat{x}_{i1} & \dots & \hat{x}_{ij} & \dots & \hat{x}_{in} \\ \vdots & \ddots & \vdots & \ddots & \vdots \\ \hat{x}_{m1} & \dots & \hat{x}_{mj} & \dots & \hat{x}_{mn} \end{bmatrix}; i = \overline{0, m}; j = \overline{1, n} \quad (15)$$

Normalization using weight w_j for all criteria can be calculated using the following formula:

$$\hat{x}_{ij} = \bar{x}_{ij} * w_j; i = \overline{0, m}, \quad (16)$$

where w_j is the weight of criterion j , and $\hat{x}_{ij}$ is the normalized ranking of criterion j . The next step is to calculate the values of the optimality function using the following formula:

$$S_i = \sum_{j=1}^n \hat{x}_{ij}; i = \overline{0, m}, \quad (17)$$

The final step in the ARAS model is to determine the ranking of the alternatives. If S_i and S_0 are optimality criterion values, then the ranking K for alternatives i follows the definition:

$$K_i = \frac{S_i}{S_0}; i = \overline{0, m}, \quad (18)$$

3.2. Experimental Design

In this section, the stages of the proposed methodology will be discussed. Three steps comprise the proposed method: preprocessing, machine learning, and the decision system. The first step is preprocessing the dataset, and the preprocessing stages aim to improve the quality of the data. Furthermore, preprocessing is performed to make the dataset more visible and is considered to improve the machine learning algorithm [58,59]. Data preprocessing concerns cleaning data, transforming categorical data to numerical form, and normalizing data.

After the preprocessing, the next step is the machine learning phase. This phase involves feature selection methods, classification, and performance evaluation. The feature selection method determines the best subset of features from the dataset. In the classification stage, a decision tree classifier is employed to generate the performance of models such as accuracy, precision, recall, f1-score, weighted precision, weighted recall, weighted f1-score, train time, and inference time using the selected feature from the previous stage.

The next stage is the decision system phase. In this step, the performance metrics are compared to determine the rank of the feature selection methods. ARAS uses the performance matrices as the ranking criteria, and this method is essential for formulating the best feature selection methods. The final result, ARAS, presented the feature selection method ranking. The stages of the proposed methodology are depicted in Figure 1.

Figure 1. The proposed method of decision model to evaluate feature selection methods.

The proposed model evaluates the feature selection method using two metrics, i.e., model performance and computation performance. Model performance is a measurement of machine learning performance using selected features, and in contrast, computational performance refers to computational capabilities during the training and inference process. Experiments and evaluations are carried out on seven methods and one baseline model, which is a model that uses all features. The schematic detail of the criteria selection of the feature selection method is portrayed in Figure 2.

Figure 2. Schematic diagram of decision model to evaluate feature selection methods.

3.3. Dataset Description

The psychosocial education dataset used here refers to the research [29,60] to test the proposed method. It is a public dataset obtained from a psychosocial assessment to identify Colombia's teachers' stress levels. The dataset consists of 4890 instances and 118 features divided into six domains. The complete specification of the dataset can be seen in Table 1.

Table 1. Detailed specification of dataset.

No	Detail	Specification
1	Number of Features	118
2	Number of Classes	4
3	Number of Instances	4890
4	Classes Name	Low, Medium, High, Very High
5	Features Domain	Sociodemographic (S), Demands of the Job (D), Control over Work (C), Leadership and Social Relations at Work (L), Rewards (R)

3.4. Dataset Preprocessing

In a machine learning problem, the dataset is present to demonstrate the effectiveness of the proposed method. Therefore, a high-quality dataset is required to evaluate the proposed model against the existing model. Data preprocessing is a well-known technique to improve dataset quality.

The teacher's psychosocial risk level dataset is valuable and pristine, and it provides the basis for delivering research on the degree of psychosocial distress among teachers in Columbia. Several studies have been conducted using the same dataset [29,60]. Primarily, the dataset was preprocessed appropriately. It will still be necessary for us to perform several preprocessing steps to prepare a suitable dataset for the proposed methods.

The first step involves performing common preprocessing steps, such as clearing improper data and handling missing values. Then, we divided the data into two subsets by following the Pareto distribution rule [61]. In this case, 80% of the data was used for

training, and 20% was used for testing. A randomly selected distribution is made to ensure fair data distribution.

The next step is to apply standardization to rescale the distribution of each dataset subset. By performing a standardization transformation of the dataset, each feature dataset will have a mean value of 0 with a standard deviation value of 1. Hopefully, a preprocessed dataset will lead machine learning to the optimal model.

3.5. Evaluation of Performance Metrics for Feature Selection Methods

Evaluation is done to measure machine learning performance. Generally, machine learning performance is measured by using a confusion matrix. A confusion matrix combines actual value and predicted value in the classifier. The confusion matrix is True Positive (TP), False Positive (FP), True Negative (TN), and False Negative (FN). The matrices for measuring accuracy, precision, recall, and f1-Score are obtained from the following calculation:

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \quad (19)$$

$$Precision = \frac{TP}{TP + FP} \quad (20)$$

$$Recall = \frac{TP}{TP + FN} \quad (21)$$

$$F1 = 2 \frac{Precision \times Recall}{Precision + Recall} \quad (22)$$

The fundamental concept for calculating the confusion matrix is binary classification [62]. A single comparison is made between two classes in binary classification, while this single comparison becomes irrelevant in multi-class classification [63]. Each class's precision, recall, and f1-score are estimated as micro-averaged and macro-averaged. After that, the metrics are calculated using the one vs. all method [64]. For example, the micro-averaged scores and macro-averaged precision (PRE) scores in the k -class are defined as follows [65]:

$$PRE_{micro} = \frac{TP_1 + \dots + TP_k}{TP_1 + \dots + TP_k + FP_1 + \dots + FP_k} \quad (23)$$

$$PRE_{macro} = \frac{PRE_1 + \dots + PRE_k}{k} \quad (24)$$

4. Results and Discussion

This section reviews the performance evaluation of the proposed method. We actualized the discussion in two parts: the performance of each feature selection method on the psychosocial education dataset and the implementation of ARAS in selecting the best feature selection method. Analysis and evaluation will also be conducted by comparing performance against a single criterion. In this case, accuracy criteria are used as a comparison.

4.1. Performance Analysis of the Feature Selection Method

This section discusses the performance measures of the feature selection method. The feature selection reduces the dimension by eliminating the least important features and retaining the important ones. It is expected that, by reducing the dimensions, the model and computational performance will increase. If the baseline used all 118 features, the other methods only performed the subset features according to the algorithm. Table 2 shows the selected feature for each method.

Table 2. Selected features in each method.

No	Models	Σ Feature	Features Selected
1	Baseline [29,60]	118	All Features
2	ANOVA [35]	11	S2, S3, S4, S5, S7, S8, S10, D3, D28, L1, R8
3	Chi-Square [39,40]	11	S2, S4, S5, S10, D3, D6, D28, C16, L1, R5, R8
4	MI [45,46]	11	S2, S3, S7, S8, S10, D9, D27, D37, C21, L1, L4
5	EFS [47,48]	24	S1, S2, S3, D2, D5, D8, D14, D17, D19, D25, D27, D31, D34, D38, C2, C9, L3, L9, L10, L12, L30, R3, R4, M1
6	ERF [50]	9	S2, S3, S4, S5, S7, S8, S10, D3, L1
7	Lasso [52]	10	S1, S3, S4, S5, S7, S8, S10, M1, M2, M3
8	RFE [54]	11	S2, S3, S8, D2, D18, D21, D26, D28, D38, L17, R5

The performance measure of the feature selection method is carried out to obtain performance parameters that will be used as the criteria for ARAS in the future. The measurement consists of performance models such as accuracy, precision, recall, f1-score, weighted precision, weighted recall, and weighted f1-score. The computation performance consists of train time and inference time. From a series of experiments conducted, what is interesting is that the baseline model requires the longest training time (34.3910 s) compared to other feature selection methods. It is decent because the baseline model used all the features in the psychosocial education datasets. However, the baseline model produced lower results than other models with far fewer selection features in the accuracy metric. Details of performance matrices for each feature selection method can be seen in Table 3.

Table 3. The performance metrics of feature selection methods.

Model	Accuracy	Precision	Recall	F1-Score	Weighted Prec.	Weighted Recall	Weighted F1-Score	Train Time (s)	Inference Time (s)
Baseline	0.9725	0.9581	0.9427	0.9501	0.9721	0.9725	0.9722	34.3910	0.9917
ANOVA	0.9734	0.9585	0.9453	0.9517	0.9730	0.9734	0.9731	3.9885	0.9933
Chi-Square	0.9752	0.9614	0.9484	0.9547	0.9749	0.9752	0.9750	3.9892	0.9950
MI	0.9757	0.9647	0.9479	0.9569	0.9754	0.9757	0.9753	3.9893	0.9636
EFS	0.9265	0.9324	0.9257	0.9267	0.9344	0.9265	0.9273	8.4677	0.9770
ERF	0.9770	0.9719	0.9478	0.9591	0.9769	0.9770	0.9766	2.0112	0.9823
Lasso	0.9770	0.9684	0.9513	0.9597	0.9768	0.9770	0.9768	2.9948	0.9646
RFE	0.9706	0.9537	0.9400	0.9470	0.9703	0.9706	0.9704	3.9878	0.9646

4.2. Evaluation Feature Selection Method Using ARAS

At this stage, choosing the best feature selection method is performed. ARAS determines the ranking using performance metrics from each feature selection method. The first step is to initialize the decision-making matrices for each alternative and their respective criteria pairs. By assigning each feature selection method as an alternative, and assignment performance matrices, i.e., accuracy (A), precision (P), recall (R), f1-score (FS), weighted precision (WP), weighted recall (WR), weighted f1-score (WFS), train time (TT), and inference time (IT) as criteria x_n . Based on the analysis, it is determined that the value of the criteria x_1 – x_7 are the benefit, while x_8 and x_9 are as the cost. In addition, it is also determined that the weighted value (w) of criteria x_1 is 0.2 and criteria x_2 – x_9 is 0.1, with the sum of their weighted values of 1. Criteria x_1 gets a higher weight because, in real problems, accuracy is one of the most important performance matrices that is widely used as a benchmark for machine learning [66,67]. The initial decision-making matrix's complete formation with each criterion's weight and optimization is shown in Table 4.

After the initial decision matrix is completed, the next step is to normalize the decision matrix. The step is finding the optimal value of A_0 value. The max operator is used for criteria with the benefit value, and the min operator is used for criteria with the cost value using Equation (10):

Table 4. Initial decision-making matrix X.

Alternative	Criteria								
	A	P	R	FS	WP	WR	WFS	TT	IT
	x_1	x_2	x_3	x_4	x_5	x_6	x_7	x_8	x_9
Optimization	Benefit	Benefit	Benefit	Benefit	Benefit	Benefit	Benefit	Cost	Cost
Weight (w)	0.2	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1
Baseline	0.9725	0.9581	0.9427	0.9501	0.9721	0.9725	0.9722	34.3910	0.9917
ANOVA	0.9734	0.9585	0.9453	0.9517	0.9730	0.9734	0.9731	3.9885	0.9933
Chi-Square	0.9752	0.9614	0.9484	0.9547	0.9749	0.9752	0.9750	3.9892	0.9950
MI	0.9757	0.9647	0.9479	0.9569	0.9754	0.9757	0.9753	3.9893	0.9636
EFS	0.9265	0.9324	0.9257	0.9267	0.9344	0.9265	0.9273	8.4677	0.9770
ERF	0.9770	0.9719	0.9478	0.9591	0.9769	0.9770	0.9766	2.0112	0.9823
Lasso	0.9770	0.9684	0.9513	0.9597	0.9768	0.9770	0.9768	2.9948	0.9646
RFE	0.9706	0.9537	0.9400	0.9470	0.9703	0.9706	0.9704	3.9878	0.9646

After obtaining the value A_{0j} , all of the criteria in the matrix are normalized. The decision matrix is normalized using Equation (12) for benefit and Equation (13) for cost. The formation of the normalization of the decision matrix $\bar{X}$ is shown in detail in Table 5, and for example of calculating the values of $\bar{x}_1 (A_0)$ and $\bar{x}_1 (Baseline)$ are as follows:

$$\bar{x}_1 (A_0) = \frac{0.9770}{0.9770+0.9725+0.9734+0.9752+0.9757+0.9265+0.9770+0.9770+0.9706}$$

$$\bar{x}_1 (A_0) = 0.1120$$

$$\bar{x}_1 (Baseline) = \frac{0.9725}{0.9770+0.9725+0.9734+0.9752+0.9757+0.9265+0.9770+0.9770+0.9706}$$

$$\bar{x}_1 (Baseline) = 0.1115$$

$$x_{A_{0j}} = \left\{ \begin{array}{l} [\max(0.9725, 0.9734, 0.9752, 0.9757, 0.9265, 0.9770, 0.9770, 0.9706)], \\ [\max(0.9581, 0.9585, 0.9614, 0.9647, 0.9324, 0.9719, 0.9684, 0.9537)], \\ [\max(0.9427, 0.9453, 0.9484, 0.9479, 0.9257, 0.9478, 0.9513, 0.9400)], \\ [\max(0.9501, 0.9517, 0.9547, 0.9569, 0.9267, 0.9591, 0.9597, 0.9400)], \\ [\max(0.9721, 0.9730, 0.9749, 0.9754, 0.9344, 0.9769, 0.9768, 0.9703)], \\ [\max(0.9725, 0.9734, 0.9752, 0.9757, 0.9265, 0.9770, 0.9770, 0.9706)], \\ [\max(0.9722, 0.9731, 0.9750, 0.9753, 0.9273, 0.9766, 0.9768, 0.9704)], \\ [\min(35.942, 3.9885, 3.9892, 3.9893, 8.4677, 2.0112, 2.9948, 3.9878)], \\ [\min(0.9668, 0.9933, 0.9950, 0.9636, 0.9770, 0.9823, 0.9646, 0.9646)] \end{array} \right\}$$

$$x_{A_{0j}} = \{0.9770, 0.9719, 0.9513, 0.9597, 0.9770, 0.9770, 0.9766, 2.0112, 0.9646\}$$

After the normalization of the matrix $\bar{X}$ is obtained, the next step is to perform the weighted normalization by multiplying the criteria weight by the normalized weighted matrix according to the formula (16). The results of weighted normalization are presented in detail in Table 6.

Table 5. Normalized decision-making matrix $\bar{X}$ on each criterion.

Alternative	Criteria								
	$\bar{x}_1$	$\bar{x}_2$	$\bar{x}_3$	$\bar{x}_4$	$\bar{x}_5$	$\bar{x}_6$	$\bar{x}_7$	$\bar{x}_8$	$\bar{x}_9$
Weight (w)	0.2	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1
A_0	0.1120	0.1125	0.1119	0.1120	0.1119	0.1120	0.1120	0.2007	0.1124
Baseline	0.1115	0.1109	0.1109	0.1109	0.1113	0.1115	0.1114	0.0112	0.1120
ANOVA	0.1116	0.1109	0.1112	0.1111	0.1114	0.1116	0.1116	0.1012	0.1090
Chi-Square	0.1118	0.1113	0.1116	0.1115	0.1117	0.1118	0.1118	0.1012	0.1088
MI	0.1118	0.1116	0.1115	0.1117	0.1117	0.1118	0.1118	0.1012	0.1124
EFS	0.1062	0.1079	0.1089	0.1082	0.1070	0.1062	0.1063	0.0477	0.1108
ERF	0.1120	0.1125	0.1115	0.1120	0.1119	0.1120	0.1120	0.2007	0.1102
Lasso	0.1120	0.1121	0.1119	0.1120	0.1119	0.1120	0.1120	0.1348	0.1122
RFE	0.1112	0.1104	0.1106	0.1106	0.1111	0.1112	0.1112	0.1012	0.1122

Table 6. Normalization-weighted decision-making matrix $\hat{X}$ of each criterion.

Alternative	Criteria								
	$\hat{x}_1$	$\hat{x}_2$	$\hat{x}_3$	$\hat{x}_4$	$\hat{x}_5$	$\hat{x}_6$	$\hat{x}_7$	$\hat{x}_8$	$\hat{x}_9$
A_0	0.0224	0.0113	0.0112	0.0112	0.0112	0.0112	0.0112	0.0201	0.0112
Baseline	0.0223	0.0111	0.0111	0.0111	0.0111	0.0112	0.0111	0.0011	0.0112
ANOVA	0.0223	0.0111	0.0111	0.0111	0.0111	0.0112	0.0112	0.0101	0.0109
Chi-Square	0.0224	0.0111	0.0112	0.0112	0.0112	0.0112	0.0112	0.0101	0.0109
MI	0.0224	0.0112	0.0112	0.0112	0.0112	0.0112	0.0112	0.0101	0.0112
EFS	0.0212	0.0108	0.0109	0.0108	0.0107	0.0106	0.0106	0.0048	0.0111
ERF	0.0224	0.0113	0.0112	0.0112	0.0112	0.0112	0.0112	0.0201	0.0110
Lasso	0.0224	0.0112	0.0112	0.0112	0.0112	0.0112	0.0112	0.0135	0.0112
RFE	0.0223	0.0110	0.0111	0.0111	0.0111	0.0111	0.0111	0.0101	0.0112

Next, the optimal value S_i is calculated, where S_i is the value of the ideal function of alternative i . After that, the criteria K_i is ranked using Equations (17) and (18). Meanwhile, the value K_i is calculated by dividing the value S_i by the value S_0 . For the values S_0 , $K_{Baseline}$ and K_{ANOVA} can be computed as follows:

$$S_0 = 0.0224 + 0.0113 + 0.0112 + 0.0112 + 0.0112 + 0.0112 + 0.0112 + 0.0201 + 0.0112$$

$$S_0 = 0.1210$$

$$K_{(Baseline)} = \frac{0.1013}{0.1210} ; \quad K_{(Baseline)} = 0.8372$$

$$K_{(ANOVA)} = \frac{0.1101}{0.1210} ; \quad K_{(ANOVA)} = 0.9099$$

In detail, the calculation of the optimal value K is presented in Table 7. Then, based on the results of the K_i , the final results of the rankings are shown in Table 8.

Table 7. The result of optimality value on the feature selection methods.

Alternative	i	S	K	Rank
A_0	0	0.1210	-	-
Baseline	1	0.1013	0.8372	8
ANOVA	2	0.1101	0.9099	5
Chi-Square	3	0.1105	0.9132	4
MI	4	0.1109	0.9165	3
EFS	5	0.1015	0.8388	7
ERF	6	0.1208	0.9983	1
Lasso	7	0.1143	0.9446	2

Table 8. Final results of the ARAS rank for feature selection methods.

Model	Rank
ERF	1
Lasso	2
MI	3
Chi-Square	4
ANOVA	5
RFE	6
EFS	7
Baseline	8

The decision results using ARAS show that the ERF method is top-ranking, and the baseline is at the lowest rank. ERF with 11 features gives better results than the baseline, which uses 118 features. It shows that selecting the best subset features is still relevant to machine learning problems.

We compare the ARAS rank with single machine learning measurements such as accuracy. In that case, the results obtained tend to be the same while on ARAS: ERF > Lasso > MI > Chi-square > Anova > RFE > EFS > Baseline, and on the other hand, using accuracy, the performance order is obtained as follows: ERF > Lasso > MI > Chi-square > Anova > Baseline > RFE > EFS. It happens because the overall performance produced by feature selection methods is mostly stable, so there are no models with cross-dominating criteria. To consider the dominating performance result, Figure 3 shows the comparative performance of every model.

Figure 3. Performance comparison of feature selection methods.

The experiment shows that the machine learning phase accomplished the model's performance analysis. By selecting specific metrics, the aim of the performance of machine learning can be defined. For example, the accuracy metric can be used as a benchmark metric to find the best accuracy model. Nevertheless, a decision model to measure and evaluate the overall performance metrics of feature selection methods is still necessary.

Finally, the goal of the proposed method is to show that the proposed model can resolve the problem formulation. Theoretically, this methodology is relevant and should be proposed. ARAS can perform a fair mapping in ranking the feature selection methods in the psychosocial education domain, especially to identify Colombia's teachers' stress level problems. However, this methodology has not fully demonstrated the significance of performance evaluation in the current dataset case, where several dominant criteria ultimately dictate the ranking results. More experience is necessary to provide a robust comparison and conclusion, and more experience based on a similar dataset might provide better results.

5. Conclusions

ARAS has proven effective and can be implemented as an evaluation model to determine the best feature selection method in the psychosocial education dataset. The evaluation used performance matrices to rank the feature selection methods. From the evaluation that has been accomplished, the determination of weight and optimization value plays an essential role in the ARAS model. Giving subjective weights affects the overall ARAS ranking.

Regarding future research directions, we recommend further investigation on the proposed method on different datasets with conditions where each criterion contradicts and does not predominate the other. The problem associated with imbalanced datasets that show uneven and contradictory performance matrices can be challenging. This problem is expected to measure the extent of ARAS's ability to provide an optimal ranking.

Author Contributions: Conceptualization, F.M. and J.-T.W.; Methodology, J.-T.W. and F.M.; Software, M.M.; Visualization, F.M. and M.M.; Project administration, J.-S.L. and J.-T.W.; Supervision, J.-S.L. and J.-T.W.; Writing—original draft, F.M. and M.M.; Writing—review and editing, J.-T.W. and J.-S.L. All authors have read and agreed to the published version of the manuscript.

Funding: This work was supported by the Ministry of Science and Technology, Taiwan, under Grant MOST-110-2637-E-011-003-.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Not applicable.

Acknowledgments: The authors gratefully acknowledge the support extended by the Ministry of Science and Technology, Taiwan, under Grant MOST-110-2637-E-011-003-.

Conflicts of Interest: The authors declare no conflict of interest.

References

- Hoti, A.H.; Heinzmann, S.; Müller, M.; Buholzer, A. Psychosocial Adaptation and School Success of Italian, Portuguese and Albanian Students in Switzerland: Disentangling Migration Background, Acculturation and the School Context. *J. Int. Migr. Integr.* **2015**, *18*, 85–106. [\[CrossRef\]](#)
- Wong, R.S.M.; Ho, F.; Wong, W.H.S.; Tung, K.T.S.; Chow, C.B.; Rao, N.; Chan, K.L.; Ip, P. Parental Involvement in Primary School Education: Its Relationship with Children's Academic Performance and Psychosocial Competence through Engaging Children with School. *J. Child Fam. Stud.* **2018**, *27*, 1544–1555. [\[CrossRef\]](#)
- Raskind, I.G.; Haardörfer, R.; Berg, C.J. Food insecurity, psychosocial health and academic performance among college and university students in Georgia, USA. *Public Health Nutr.* **2019**, *22*, 476–485. [\[CrossRef\]](#)
- Sierra-Díaz, M.J.; González-Villora, S.; Pastor-Vicedo, J.C.; Sánchez, G.F.L. Can We Motivate Students to Practice Physical Activities and Sports Through Models-Based Practice? A Systematic Review and Meta-Analysis of Psychosocial Factors Related to Physical Education. *Front. Psychol.* **2019**, *10*, 2115. [\[CrossRef\]](#)
- Souravlas, S.; Anastasiadou, S. Pipelined Dynamic Scheduling of Big Data Streams. *Appl. Sci.* **2020**, *10*, 4796. [\[CrossRef\]](#)
- López-Belmonte, J.; Segura-Robles, A.; Moreno-Guerrero, A.-J.; Parra-González, M.E. Machine Learning and Big Data in the Impact Literature. A Bibliometric Review with Scientific Mapping in Web of Science. *Symmetry* **2020**, *12*, 495. [\[CrossRef\]](#)
- Al-Jarrah, O.Y.; Yoo, P.; Muhaidat, S.; Karagiannidis, G.K.; Taha, K. Efficient Machine Learning for Big Data: A Review. *Big Data Res.* **2015**, *2*, 87–93. [\[CrossRef\]](#)
- Altman, N.; Krzywinski, M. The curse(s) of dimensionality. *Nat. Methods* **2018**, *15*, 399–400. [\[CrossRef\]](#)

9. Köppen, M. The curse of dimensionality. In Proceedings of the 5th Online World Conference on Soft Computing in Industrial Applications (WSC5), Online, 4–8 September 2000; Volume 1, pp. 4–8.
10. Khaire, U.M.; Dhanalakshmi, R. Stability of feature selection algorithm: A review. *J. King Saud Univ. Comput. Inf. Sci.* **2019**, *34*. [\[CrossRef\]](#)
11. Jović, A.; Brkić, K.; Bogunović, N. A review of feature selection methods with applications. In Proceedings of the 2015 38th International Convention on Information and Communication Technology, Electronics and Microelectronics (MIPRO), Opatija, Croatia, 25–29 May 2015; pp. 1200–1205.
12. Chandrashekar, G.; Sahin, F. A survey on feature selection methods. *Comput. Electr. Eng.* **2014**, *40*, 16–28. [\[CrossRef\]](#)
13. Cai, J.; Luo, J.; Wang, S.; Yang, S. Feature selection in machine learning: A new perspective. *Neurocomputing* **2018**, *300*, 70–79. [\[CrossRef\]](#)
14. Moorthy, U.; Gandhi, U.D. A novel optimal feature selection technique for medical data classification using ANOVA based whale optimization. *J. Ambient. Intell. Humaniz. Comput.* **2020**, *12*, 3527–3538. [\[CrossRef\]](#)
15. Ding, H.; Li, D. Identification of mitochondrial proteins of malaria parasite using analysis of variance. *Amino Acids* **2015**, *47*, 329–333. [\[CrossRef\]](#)
16. Utama, H. Sentiment analysis in airline tweets using mutual information for feature selection. In Proceedings of the 2019 4th International Conference on Information Technology, Information Systems and Electrical Engineering (ICITISEE), Yogyakarta, Indonesia, 20–21 November 2019; pp. 295–300.
17. Richhariya, B.; Tanveer, M.; Rashid, A. Diagnosis of Alzheimer’s disease using universum support vector machine based recursive feature elimination (USVM-RFE). *Biomed. Signal Process. Control.* **2020**, *59*, 101903. [\[CrossRef\]](#)
18. Park, D.; Lee, M.; Park, S.E.; Seong, J.-K.; Youn, I. Determination of Optimal Heart Rate Variability Features Based on SVM-Recursive Feature Elimination for Cumulative Stress Monitoring Using ECG Sensor. *Sensors* **2018**, *18*, 2387. [\[CrossRef\]](#)
19. ZLiu, Z.; Thapa, N.; Shaver, A.; Roy, K.; Siddula, M.; Yuan, X.; Yu, A. Using Embedded Feature Selection and CNN for Classification on CCD-INID-V1—A New IoT Dataset. *Sensors* **2021**, *21*, 4834.
20. Loscalzo, S.; Wright, R.; Acunto, K.; Yu, L. Sample aware embedded feature selection for reinforcement learning. In Proceedings of the 14th Annual Conference on Genetic and Evolutionary Computation, Philadelphia, PA, USA, 7–11 July 2012; pp. 887–894.
21. Liu, H.; Zhou, M.; Liu, Q. An embedded feature selection method for imbalanced data classification. *IEEE/CAA J. Autom. Sin.* **2019**, *6*, 703–715. [\[CrossRef\]](#)
22. Kou, G.; Yang, P.; Peng, Y.; Xiao, F.; Chen, Y.; Alsaadi, F.E. Evaluation of feature selection methods for text classification with small datasets using multiple criteria decision-making methods. *Appl. Soft Comput.* **2020**, *86*, 10583. [\[CrossRef\]](#)
23. Hashemi, A.; Dowlatshahi, M.B.; Nezamabadi-Pour, H. Ensemble of feature selection algorithms: A multi-criteria decision-making approach. *Int. J. Mach. Learn. Cybern.* **2021**, 1–21. [\[CrossRef\]](#)
24. Singh, R.; Kumar, H.; Singla, R.K. TOPSIS based multi-criteria decision making of feature selection techniques for network traffic dataset. *Int. J. Eng. Technol.* **2014**, *5*, 4598–4604.
25. Souravlas, S.; Anastasiadou, S.; Katsavounis, S. A Survey on the Recent Advances of Deep Community Detection. *Appl. Sci.* **2021**, *11*, 7179. [\[CrossRef\]](#)
26. Acosta, D.; Fujii, Y.; Joyce-Beaulieu, D.; Jacobs, K.D.; Maurelli, A.T.; Nelson, E.J.; McKune, S.L. Psychosocial Health of K-12 Students Engaged in Emergency Remote Education and In-Person Schooling: A Cross-Sectional Study. *Int. J. Environ. Res. Public Health* **2021**, *18*, 8564. [\[CrossRef\]](#)
27. Carreon, A.D.V.; Manansala, M.M. Addressing the psychosocial needs of students attending online classes during this COVID-19 pandemic. *J. Public Health* **2021**, *43*, e385–e386. [\[CrossRef\]](#)
28. Mahapatra, A.; Sharma, P. Education in times of COVID-19 pandemic: Academic stress and its psychosocial impact on children and adolescents in India. *Int. J. Soc. Psychiatry* **2021**, *67*, 397–399. [\[CrossRef\]](#)
29. Navarro, R.M.; Castrillón, O.D.; Osorio, L.P.; Oliveira, T.; Novais, P.; Valencia, J.F. Improving classification based on physical surface tension-neural net for the prediction of psychosocial-risk level in public school teachers. *PeerJ. Comput. Sci.* **2021**, *7*, e511. [\[CrossRef\]](#)
30. Kira, K.; Rendell, L.A. A practical approach to feature selection. In *Machine Learning Proceedings 1992*; Sleeman, D., Edwards, P., Eds.; Morgan Kaufmann: San Francisco, CA, USA, 1992; pp. 249–256.
31. Dash, M.; Liu, H. Feature selection for classification. *Intell. Data Anal.* **1997**, *1*, 131–156. [\[CrossRef\]](#)
32. Urbanowicz, R.J.; Meeker, M.; La Cava, W.; Olson, R.S.; Moore, J.H. Relief-based feature selection: Introduction and review. *J. Biomed. Inform.* **2018**, *85*, 189–203. [\[CrossRef\]](#) [\[PubMed\]](#)
33. Bommert, A.; Sun, X.; Bischl, B.; Rahnenführer, J.; Lang, M. Benchmark for filter methods for feature selection in high-dimensional classification data. *Comput. Stat. Data Anal.* **2020**, *143*, 106839. [\[CrossRef\]](#)
34. Ashik, M.; Jyothish, A.; Anandaram, S.; Vinod, P.; Mercaldo, F.; Martinelli, F.; Santone, A. Detection of Malicious Software by Analyzing Distinct Artifacts Using Machine Learning and Deep Learning Algorithms. *Electronics* **2021**, *10*, 1694. [\[CrossRef\]](#)
35. Johnson, K.J.; Synovec, E.R. Pattern recognition of jet fuels: Comprehensive GC×GC with ANOVA-based feature selection and principal component analysis. *Chemom. Intell. Lab. Syst.* **2002**, *60*, 225–237. [\[CrossRef\]](#)
36. Vora, S.; Yang, H. A comprehensive study of eleven feature selection algorithms and their impact on text classification. In Proceedings of the 2017 Computing Conference, London, UK, 18–20 July 2017; pp. 440–449.

37. Ghosh, M.; Sanyal, G. Performance Assessment of Multiple Classifiers Based on Ensemble Feature Selection Scheme for Sentiment Analysis. *Appl. Comput. Intell. Soft Comput.* **2018**, *2018*, 8909357. [\[CrossRef\]](#)
38. Alazab, M. Automated Malware Detection in Mobile App Stores Based on Robust Feature Generation. *Electronics* **2020**, *9*, 435. [\[CrossRef\]](#)
39. Cilia, N.D.; De Stefano, C.; Fontanella, F.; di Freca, A.S. A ranking-based feature selection approach for handwritten character recognition. *Pattern Recognit. Lett.* **2019**, *121*, 77–86. [\[CrossRef\]](#)
40. Bahassine, S.; Madani, A.; Al-Sarem, M.; Kissi, M. Feature selection using an improved Chi-square for Arabic text classification. *J. King Saud Univ. Comput. Inf. Sci.* **2020**, *32*, 225–231. [\[CrossRef\]](#)
41. Thejas, G.S.; Joshi, S.R.; Iyengar, S.S.; Sunitha, N.R.; Badrinath, P. Mini-Batch Normalized Mutual Information: A Hybrid Feature Selection Method. *IEEE Access* **2019**, *7*, 116875–116885. [\[CrossRef\]](#)
42. Macedo, F.; Oliveira, M.R.; Pacheco, A.; Valadas, R. Theoretical foundations of forward feature selection methods based on mutual information. *Neurocomputing* **2019**, *325*, 67–89. [\[CrossRef\]](#)
43. Guyon, I.; Elisseeff, A. An Introduction to Variable and Feature Selection. *J. Mach. Learn. Res.* **2003**, *3*, 1157–1182.
44. Peng, H.; Long, F.; Ding, C. Feature selection based on mutual information criteria of max-dependency, max-relevance, and min-redundancy. *IEEE Trans. Pattern Anal. Mach. Intell.* **2005**, *27*, 1226–1238. [\[CrossRef\]](#) [\[PubMed\]](#)
45. Gonzalez-Lopez, J.; Ventura, S.; Cano, A. Distributed multi-label feature selection using individual mutual information measures. *Knowl.-Based Syst.* **2020**, *188*, 105052. [\[CrossRef\]](#)
46. Zhou, H.; Zhang, Y.; Zhang, Y.; Liu, H. Feature selection based on conditional mutual information: Minimum conditional relevance and minimum conditional redundancy. *Appl. Intell.* **2019**, *49*, 883–896. [\[CrossRef\]](#)
47. Ruggieri, S. Complete Search for Feature Selection in Decision Trees. *J. Mach. Learn. Res.* **2019**, *20*, 1–34.
48. Igarashi, Y.; Ichikawa, H.; Nakanishi-Ohno, Y.; Takenaka, H.; Kawabata, D.; Eifuku, S.; Tamura, R.; Nagata, K.; Okada, M. ES-DoS: Exhaustive search and density-of-states estimation as a general framework for sparse variable selection. *J. Phys. Conf. Ser.* **2018**, *1036*, 012001. [\[CrossRef\]](#)
49. Lee, C.-Y.; Chen, B.-S. Mutually-exclusive-and-collectively-exhaustive feature selection scheme. *Appl. Soft Comput.* **2018**, *68*, 961–971. [\[CrossRef\]](#)
50. Granitto, P.; Furlanello, C.; Biasioli, F.; Gasperi, F. Recursive feature elimination with random forest for PTR-MS analysis of agroindustrial products. *Chemom. Intell. Lab. Syst.* **2006**, *83*, 83–90. [\[CrossRef\]](#)
51. Tibshirani, R. Regression Shrinkage and Selection Via the Lasso. *J. R. Stat. Soc. Ser. B Methodol.* **1996**, *58*, 267–288. [\[CrossRef\]](#)
52. Hesterberg, T.; Choi, N.H.; Meier, L.; Fraley, C. Least angle and ℓ_1 penalized regression: A review. *Stat. Surv.* **2008**, *2*, 61–93. [\[CrossRef\]](#)
53. Abdulsalam, S.O.; Mohammed, A.A.; Ajao, J.F.; Babatunde, R.S.; Ogundokun, R.O.; Nnodim, C.T.; Arowolo, M.O. Performance Evaluation of ANOVA and RFE Algorithms for Classifying Microarray Dataset Using SVM. *Lect. Notes Bus. Inf. Process.* **2020**, *480*–492. [\[CrossRef\]](#)
54. Guyon, I.; Weston, J.; Barnhill, S.; Vapnik, V. Gene Selection for Cancer Classification using Support Vector Machines. *Mach. Learn.* **2002**, *46*, 389–422. [\[CrossRef\]](#)
55. Zavadskas, E.K.; Turskis, Z. A new additive ratio assessment (ARAS) method in multi-criteria decision-making. *Technol. Econ. Dev. Econ.* **2010**, *16*, 159–172. [\[CrossRef\]](#)
56. Radović, D.; Stević, Ž.; Pamučar, D.; Zavadskas, E.K.; Badi, I.; Antuchevičienė, J.; Turskis, Z. Measuring Performance in Transportation Companies in Developing Countries: A Novel Rough ARAS Model. *Symmetry* **2018**, *10*, 434. [\[CrossRef\]](#)
57. Maulana, C.; Hendrawan, A.; Pinem, A.P.R. Pemodelan Penentuan Kredit Simpan Pinjam Menggunakan Metode Additive Ratio Assessment (Aras). *J. Pengemb. Rekayasa Teknol.* **2019**, *15*, 7–11. [\[CrossRef\]](#)
58. García, S.; Luengo, J.; Herrera, F. Data preparation basic models. In *Data Preprocessing in Data Mining*; International Publishing; Springer: Cham, Switzerland, 2015; pp. 39–57.
59. Kotsiantis, S.B.; Kanellopoulos, D.; Pintelas, P.E. Data preprocessing for supervised learning. *Int. J. Comput. Sci.* **2006**, *1*, 111–117.
60. Mosquera, R.; Castrillón, O.D.; Parra, L. Prediction of Psychosocial Risks in Colombian Teachers of Public Schools using Machine Learning Techniques. *Inf. Tecnol.* **2018**, *29*, 267–280. [\[CrossRef\]](#)
61. Newman, M.E.J. Power laws, Pareto distributions and Zipf's law. *Contemp. Phys.* **2005**, *46*, 323–351. [\[CrossRef\]](#)
62. Luque, A.; Carrasco, A.; Martín, A.; de las Heras, A. The impact of class imbalance in classification performance metrics based on the binary confusion matrix. *Pattern Recognit.* **2019**, *91*, 216–231. [\[CrossRef\]](#)
63. Takahashi, K.; Yamamoto, K.; Kuchiba, A.; Koyama, T. Confidence interval for micro-averaged F1 and macro-averaged F1 scores. *Appl. Intell.* **2021**, 1–12. [\[CrossRef\]](#)
64. Pillai, L.; Fumera, G.; Roli, F. F-measure optimisation in multi-label classifiers. In Proceedings of the 21st International Conference on Pattern Recognition (ICPR2012), Tsukuba, Japan, 11–15 November 2012; pp. 2424–2427.
65. Van Asch, V. Macro- and Micro-Averaged Evaluation Measures. 2013, pp. 1–27. Available online: <https://www.semanticscholar.org/paper/Macro-and-micro-averaged-evaluation-measures-%5B-%5B-%5D-Asch/1d106a2730801b6210a67f7622e4d192bb309303> (accessed on 14 November 2021).

-
66. Sokolova, M.; Japkowicz, N.; Szpakowicz, S. Beyond accuracy, F-score and ROC: A family of discriminant measures for performance evaluation. In *AI 2006: Advances in Artificial Intelligence*; Springer: Berlin/Heidelberg, Germany, 2006; pp. 1015–1021.
  67. Yin, M.; Vaughan, J.W.; Wallach, H. Understanding the effect of accuracy on trust in machine learning models. In Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems, Glasgow, UK, 4–9 May 2019; pp. 1–12.