Bulletin of Electrical Engineering and Informatics
Vol. 14, No. 2, April 2025, pp. 1276~1289
ISSN: 2302-9285, DOI: 10.11591/eei.v14i2.8715

Optimizing the best student selection: hybrid K-Means approach and entropy-grey relational analysis

Heni Sulistiani[1], Setiawansyah[2], Pritasari Palupiningsih[3], Ahmad Ferico Octaviansyah Pasaribu[4], Rio Andika[4], Muhammad Hamdan Sobirin[4]
[1]Department of Computer Science, Faculty of Engineering and Computer Science, Universitas Teknokrat Indonesia, Bandar Lampung, Indonesia
[2]Department of Informatics, Faculty of Engineering and Computer Science, Universitas Teknokrat Indonesia, Bandar Lampung, Indonesia
[3]Department of Informatics Engineering, Faculty of Energy Telematics, Institut Teknologi Perusahaan Listrik Negara, Jakarta, Indonesia
[4]Department of Information Technology, Faculty of Engineering and Computer Science, Universitas Teknokrat Indonesia, Bandar Lampung, Indonesia

Article Info

Article history:

Received May 19, 2024
Revised Oct 17, 2024
Accepted Nov 19, 2024

Keywords:

Clustering
Entropy
Grey relational analysis
Hybrid
K-Means clustering

ABSTRACT

The selection of the best students is an important process in recognizing students' achievements and dedication in various fields. Through careful and fair selection, students who stand out in both academic and non-academic terms can be identified and assigned. The purpose of the research on the use of hybrid entropy-grey relational analysis (GRA) and K-Means clustering in the selection of the best students is to develop a more objective, accurate, and comprehensive assessment system. The silhouette score results show that 2 clusters have a value of 0.5733, so in this study 2 clusters are used with the best cluster at cluster 0. Data from cluster 0 will be used in determining the best students using hybrid entropy-GRA. The results of the best student ranking using the hybrid entropy-GRA method, for the first best student with a final score of 0.25 were obtained by Mareta Amelia. The hybrid approach of K-Means and entropy-GRA offers a powerful tool to improve decision-making in the student selection process. The hybrid approach of K-Means grouping and entropy-GRA presents a powerful solution, improving the decision-making process and ensuring that high-achieving students are accurately recognized and rewarded.

Corresponding Author:

Setiawansyah
Department of Informatics, Faculty of Engineering and Computer Science
Universitas Teknokrat Indonesia
Bandar Lampung, Indonesia
Email: setiawansyah@teknokrat.ac.id

1. INTRODUCTION

Education is one of the most important aspects in the development of a country. The quality of a country's education system greatly influences social, economic, and cultural development and progress. In an effort to improve the quality of education, one of the key factors is the selection of the best students to enter the right educational institution. The selection of the best students is not only important for the educational institution itself, but also has a significant impact on students, society, and even the country as a whole [1]. By ensuring that selected students have academic and other development potential, educational institutions can optimize the learning process and ensure high graduation and availability of a quality workforce in the future. In the era of technology and information like now, the approach of optimizing the selection of the best students is increasingly driven by the use of sophisticated methods and technology. A variety of quantitative and qualitative approaches can be used to evaluate student selection criteria, ranging from academic exam results, non-academic achievement, to psychological and personality assessments. In this article, we will explore the different approaches, strategies, and technologies used in the optimization of the selection of the best students. We will highlight the importance of using data and analytics to gain a deep understanding of prospective students, as well as the challenges and opportunities faced in this process. The purpose of this article is to provide a comprehensive insight into how educational institutions can improve effectiveness and efficiency in the selection of the best students, as well as the contribution that technology and innovation can make in this field. The selection of the best students is an important process in recognizing students' achievements and dedication in various fields. Through careful and fair selection, the students who stand out are good. This selection not only strengthens students' motivation to excel, but also gives them proper recognition for their hard work. With this award, the best students become role models for others, encourage the spirit of healthy competition, and create a dynamic learning environment and motivate to achieve higher achievements in the future. The problem that occurs in the selection of the best students is that there are many students who will be selected to be the best students for each batch. In one batch of 471 students who will be selected the three best students for each year, this process takes a long time to get results from the selection of the best students.

Clustering is a data analysis technique used to group a set of objects based on their similarity or characteristics [2]–[4]. In this process, similar data will be grouped in one cluster, while different data will be grouped in another cluster. This method is often used in areas such as marketing, biology, image processing, and machine learning, to find hidden patterns in big data. Clustering helps in identifying natural structures in the data, facilitating better and effective decision making. Common techniques used in clustering include K-Means clustering, K-Means clustering is one of the most popular and simple clustering algorithms in data analysis. This method works by dividing a set of data into a predefined number of k clusters, K-Means clustering is effective in identifying the underlying data structure and is widely used in various applications such as market segmentation, pattern recognition, and data compression [5]–[7]. K-Means clustering has a number of advantages that make it a popular choice in data analysis. The algorithm is relatively simple and easy to implement, which allows it to be widely used by researchers and practitioners in various fields. Its high convergence speed makes K-Means efficient in handling large datasets, making it suitable for large-scale applications such as big data analysis. In addition, the results of K-Means clustering are easy to understand and interpret because they use an intuitive centroid concept [8], [9]. The algorithm is also flexible in a variety of situations and can be easily adapted or extended. Student data will be clustered first to find the best cluster from the selection of the best students, after the best cluster is obtained, it will continue in the decision support system (DSS) to determine the best students.

DSS is a computer-based system designed to assist decision makers in compiling and analyzing information, so as to make better and appropriate decisions [10]–[13]. DSS plays an important role in improving decision quality by providing relevant, timely, and easy-to-understand information, while reducing risk and uncertainty in the decision-making process. The main advantage of DSS lies in its ability to combine multiple data and analytics sources, allowing users to explore different options and scenarios quickly and efficiently [14]–[16]. The system also provides tools that support collaboration, allowing teams to work together on decision-making. DSS not only improves the efficiency and effectiveness of decision making, but also provides a solid foundation for the development of better and more informed strategies in various sectors. Grey relational analysis (GRA) is an analytical method used to address decision-making problems with limited or incomplete information [17]–[20]. Developed as part of grey system theory, GRA focuses on measuring the relationship between a number of factors in situations full of uncertainty. This method is particularly useful in situations where the available data is insufficient for conventional statistical analysis. By calculating the gray relationship coefficient, the GRA can identify and sort alternatives based on their relative proximity to the ideal solution. This makes GRA highly effective in a wide range of applications such as performance evaluation, parameter optimization, and strategy selection in management and engineering. GRA's main advantages are its simplicity in application and its ability to deliver meaningful results even when little or incomplete data is available, making it a valuable tool in decision-making in uncertain environments. In the entropy method, weights are determined using entropy, which measures uncertainty or degree of irregularity in the data. Entropy is used to give weight to each criterion based on variations in that data: the greater the variation or irregularity in the data of a criterion, the higher the weight given to that criterion [21], [22]. This approach ensures that more varied and significant criteria in differentiating student performance will have a greater influence in the assessment process. Determination of weights using entropy provides a more dynamic and accurate evaluation, reflecting the importance of each criterion in the overall context.

Hybrid K-Means clustering and entropy-GRA are methods used for data analysis in the field of grouping and decision making. Hybrid K-Means clustering combines the K-Means algorithm with other techniques to improve accuracy and efficiency in grouping data into clusters based on the similarity of their attributes. Meanwhile, entropy-GRA is a method used to evaluate the relationship between several variables under conditions of uncertainty and incomplete data as well as the determination of the weight of criteria. By combining these two methods, we can perform more accurate clustering of data and also better analyze complex relationships between variables, providing more comprehensive and reliable results in a variety of applications. The incorporation of hybrid K-Means clustering and entropy-GRA makes it possible to overcome the limitations of each method and improve overall data analysis performance. In its implementation, hybrid K-Means clustering is used first to group data into homogeneous clusters. After the data is grouped, entropy will determine the weight of the criteria used, and GRA is applied to evaluate and analyze the relationship between variables in each cluster, so as to identify the most influential factors and determine priorities in decision making.

Research related to hybrid between K-Means clustering and entropy-GRA has never been done, research conducted by previous research only around clustering the best students and selecting the best students, but has never been done by applying or combining the two techniques. Hybrid between K-means clustering and entropy-GRA is a potential novelty in this study using the advantages of each technique used. In the selection of the best students, hybrid K-Means clustering and entropy-GRA can be used to improve the objectivity and accuracy of assessment. First, hybrid K-Means clustering is applied to group students based on assessment criteria such as academic grades, extracurricular activities, and attendance. This grouping process ensures that students with similar characteristics are grouped together, facilitating further analysis. After students are grouped, entropy will determine the weight of the criteria used, and GRA is used to analyze the relationship and effect of each criterion on student performance in each cluster. With this method, the key factors that determine a student's best performance can be identified more clearly. The results of this analysis provide a solid basis for objectively selecting the best students, considering the various aspects that contribute to their success holistically.

This study examines the impact of student grouping using the K-Means method on the selection of the best students. While previous studies investigated the influence of technique for order of preference by similarity to ideal solution (TOPSIS) and analytic hierarchy process (AHP) evaluation methods in academic performance assessment, the study did not explicitly address the effect of the combination of K-Means clustering and entropy-GRA analysis on the improvement of student selection decisions. This gap demonstrates the need for a hybrid approach that can provide deeper insights into the process of selecting the best students. The research objective of using hybrid GRA and K-Means clustering in the selection of the best students is to develop a more objective, accurate, and comprehensive assessment system. By applying K-Means clustering, this study aims to group students based on various assessment criteria for student academic score data, so that groups of students with similar characteristics can be analyzed more deeply. Furthermore, through the application of GRA, this study aims to evaluate the relative influence of each assessment criterion within each group that has been formed, identifying key factors that contribute to student performance. The ultimate goal of this study is to provide a holistic, data-driven method of selecting the best students, which can help decision makers better understand and appreciate various aspects of student success, as well as increase fairness and transparency in the assessment process.

2. METHOD

Research stages are a series of steps carried out to plan, implement, and analyze a study or research. Research stages provide a structured and organized framework for effectively managing their research projects [23]. By adhering to this process, research can avoid confusion, reduce the risk of errors, and improve the validity and reliability of research results. The research stage is not only a formal process in carrying out research, but also a solid foundation for producing meaningful and sustainable knowledge. Figure 1 is a framework of research concepts conducted in optimizing the selection of the best students.

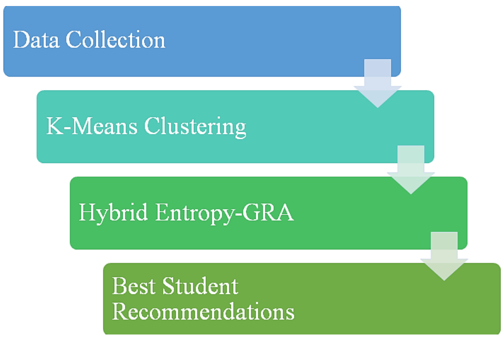

Figure 1. Research concept framework

The research concept framework Figure 1 is a framework for problem-solving concepts in determining the best students, the first stage is collecting data in determining the best students. After the data is collected, it then performs clustering by applying the K-Means clustering algorithm with the aim of identifying hidden patterns or structures in the data, which can help in understanding, analysis, and decision making. The next process applies hybrid entropy and GRA in the process of determining the best students, finally providing recommendations in selecting the best students from the results of clustering and hybrid entropy-GRA.

2.1. Data collection

Data collection is an important process in research that involves various methods to gather relevant and accurate information to answer research questions or test hypotheses. Careful and ethical data collection and analysis is a strong foundation for valid and reliable research, and supports effective and evidence-based decision making. The data collected in the study of student data to be selected as the best student, the student data used amounted to 471 student data collected for grade 12 high school students.

2.2. K-Means clustering

K-Means clustering is one of the data analysis methods used to group a number of data into several groups or clusters based on similar characteristics [24]–[26]. This method works by assigning a random number of k centroids, which are then used as the center of each cluster. Each data point in the dataset will be spaced to each centroid, and the data point will be allocated to the cluster whose centroid is closest. This process is repeated iteratively, with the centroid updated based on the average position of the data points within each cluster, until there is no significant change in the position of the centroid or data point again. K-Means clustering is highly effective in data segmentation and has a wide range of applications in areas such as customer clustering, image analysis, and pattern recognition, although the choice of the right k value and sensitivity to outliers are challenges that need to be overcome to achieve optimal results [27].

2.3. Hybrid entropy-grey relational analysis

Hybrid entropy-GRA is a multi-criteria analysis method that combines entropy-GRA to improve accuracy and reliability in decision making. This method utilizes entropy to determine the objective weight of each criterion based on the degree of uncertainty and variation of the data, thereby reducing subjective judgment in the process of determining weights [28], [29]. Once weights are assigned, the GRA is used to evaluate and sort alternatives based on their relative similarity to the ideal solution. Hybrid entropy-GRA is particularly effective in situations where there are many criteria to consider, and each criterion has a different level of importance. This method is often applied in many fields, including project management, performance evaluation, and supplier selection, due to its ability to provide more balanced and informative results than traditional methods that rely on only one analytical approach. The first stage using entropy is to make a decision matrix made using (1).

$$X = \begin{bmatrix} x_{11} & \cdots & x_{n1} \\ \vdots & \ddots & \vdots \\ x_{1m} & \cdots & x_{nm} \end{bmatrix} \tag{1}$$

The second stage, normalizing the decision matrix, is an important step in decision analysis that aims to overcome variations in the scale or range of criteria values used in the decision matrix. The normalization of the decision matrix in the entropy method is calculated using (2).

$$k_{ij} = \frac{x_{ij}}{\max x_{ij}} \tag{2}$$

The third stage, calculating the value of the criterion matrix, is a numerical representation of the weight or relative importance of each criterion used in the decision-making process. The value of the criterion matrix in the entropy method is calculated using (3).

$$a_{ij} = \frac{k_{ij}}{\sum_{i=1}^{k} k_{ij}} \tag{3}$$

The fourth stage is to calculate the entropy value to evaluate the level of clarity or certainty of a distribution. The entropy value is calculated using (4).

$$E_j = \left[\frac{-1}{\ln m}\right] \sum_{i=1}^{n} \left[a_{ij} * \ln\left(a_{ij}\right)\right] \tag{4}$$

The fifth stage, calculating the dispersion value, can help decision makers identify the level of consistency or variation between weights or criteria values used. The dispersion value in the entropy method is calculated using (5).

$$D_j = 1 - E_j \quad (5)$$

The sixth stage, which calculates the value of the dispersion of criteria, provides an idea of the extent of variation in the weight of the criterion, and can help identify the degree of consistency or variation of preferences between criteria used in an evaluation. The value of the dispersion criterion in the entropy method is calculated using (6).

$$W_j = \frac{D_j}{\sum D_j} \quad (6)$$

The seventh stage in solving problems using the GRA method is normalization using (7).

$$X_{norm} = \frac{X_{ij} - X_{min}}{X_{max} - X_{min}} \quad (7)$$

The eighth stage is to calculate the relative weight of each alternative using (8).

$$V_{ij} = x_{i,j} * w_j \quad (8)$$

The final step is to calculate the value of the gray relation calculated for each alternative using (9).

$$GRG_i = \frac{1}{n}\sum_{j=1}^{n} V_{ij} \quad (9)$$

The hybrid entropy-GRA method is an effective approach in multi-criteria decision making because it combines the advantages of the entropy weight and GRA methods. By using entropy weight, this method is able to reduce subjectivity in determining the weight of criteria by measuring the uncertainty or diversity of information from each criterion. The GRA is used to evaluate the relative performance of alternatives, enabling a more objective and comprehensive assessment. This combination provides more robust and accurate results in the ranking and selection of alternatives, and provides a holistic and reliable framework for making better decisions based on a variety of relevant criteria.

2.4. Best student recommendations

The best student recommendations using the hybrid entropy-GRA method result in a more objective and accurate evaluation by considering various relevant criteria. This method combines entropy to determine the objective weight of each criterion based on the degree of data uncertainty, and GRA to measure the proximity relationship between the evaluated alternative and the ideal solution. With this approach, the resulting decision is more comprehensive and reliable because it combines quantitative analysis of entropy and relative quality assessment of GRA. As a result, the hybrid entropy-GRA method is able to provide the best student recommendations by considering various academic and non-academic aspects holistically.

3. RESULTS AND DISCUSSION

Optimizing best student selection: enhancing decision making with a hybrid K-Means clustering and entropy-GRA approach results in a more effective and efficient selection system. In this study, K-Means clustering was used to group students based on similar characteristics, thus facilitating a more focused and structured data analysis. Furthermore, entropy is used to determine the objective weight of each criterion, while GRA is used to evaluate the relationship between available alternatives and ideal solutions. The combination of these two methods increases accuracy and objectivity in the determination of the best students, as it combines clustering capabilities for handling large and heterogeneous data with detailed analysis of entropy-GRA. The results of this approach show improvements in consistency and transparency of decisions, ultimately supporting a more fair and thorough selection of the best students.

3.1. Data collection

Data collection is a systematic process of collecting and measuring information from various sources in order to get an accurate and comprehensive picture of a phenomenon or condition. Data collection is an important stage in research or analysis, as the quality and accuracy of the data collected will greatly affect the results and conclusions that can be drawn from the study. Table 1 is the result of data collection carried out in determining the best students.

Table 1. Best student assessment data

Number	NISN	Student name	Average value	Attitude	Achievement	Activeness
1	1227294	Adinda Putri Utami	83.59	89	93	91
2	1227295	Ai Pirnawati	84.34	89	93	91
3	1227296	Akbar Azura	81.22	86	90	88
4	1227297	Akmaludin Muzaki	82.1	87	91	89
5	1227298	Alvin Cahya Imanda	81.96	87	91	89
6	1227299	Andreas	83.32	88	92	90
7	1227300	Arjun Surya Maulana	82.64	88	92	90
8	1227301	Auliya Hafidah	81.44	86	90	88
9	1227302	Bagus Hilal Akbar	79.97	85	89	87
10	1227303	Desta Syuda Wirat	84.27	89	93	91
…	…	…	…	…	…	…
471	1227765	Tomy Andrian	80.93	86	90	88

3.2. K-Means clustering algorithm

Clustering students using the K-Means method is an effective technique in educational data analysis to identify groups of students based on their academic performance. With K-Means, we can divide student data into multiple clusters, where each cluster represents a group of students with similar performance characteristics. This process begins by determining the desired number of clusters (k), then the K-Means algorithm allocates each student to the nearest cluster based on the Euclidean distance from the cluster center (centroid). After that, the centroid is updated and the grouping process is repeated until convergence is achieved. The end result is grouping that can assist educators in identifying the best students and designing more targeted learning strategies to support their academic achievement. The use of K-Means clustering in education analysis not only helps in grouping students based on their performance, but also provides deeper insights to improve the overall quality of education. The first stage is pre-processing data, pre-processing the best student data involves several important steps to ensure the data is ready for clustering analysis using K-Means. First, relevant data such as average scores, attitudes, achievements, and activeness. Next, the data is cleaned by looking for the missing value of the data used. Figure 2 is the result of searching for missing values using Python.

```
df_segmentation[df_segmentation.isnull().any(axis=1)]

      Average Score  Attitude  Achievement  Activeness
NIS
```

Figure 2. Missing value

Figure 2 is the search result of the missing value of the data used, from the missing value results no data was found that had a null value. The next stage forms a correlation matrix which is a statistical tool used to measure and display relationships between various variables in a dataset. With visualizations such as heatmaps, correlation matrices make it easier for data analysts to recognize significant relationship patterns, which can be used to make better decisions in managing and improving students' academic performance. Figure 3 is the result of a visualization of the correlation matrix.

In the context of student data analysis, this matrix helps identify the extent to which variables such as test scores, grade point average, attitude, achievement, and activeness are related. Each element in the matrix shows the correlation coefficient between the two variables, with values ranging from -1 to 1. A value of +1 indicates a perfect positive correlation, -1 indicates a perfect negative correlation, and 0 indicates no correlation. The results of the correlation matrix Figure 3 show a good correlation of the variables to be used.

The next stage of determining the optimal number of clusters is an important step in clustering analysis to ensure that the results obtained are representative and useful. Determining the optimal number of clusters in clustering analysis is an important process that ensures accurate and useful results, determining the optimal number of clusters using the Elbow method is a popular and intuitive approach in clustering analysis. The visualizations generated by the Elbow method help provide an intuitive picture of the trade-off between the number of clusters and the explanation of variability in the data, making decision-making easier for data analysts. Figure 4 is the result of visualizing the number of clusters using an elbow.

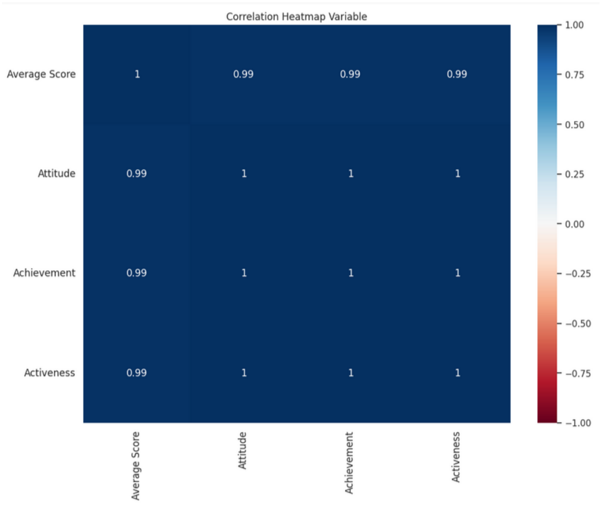

Figure 3. Correlation matrix

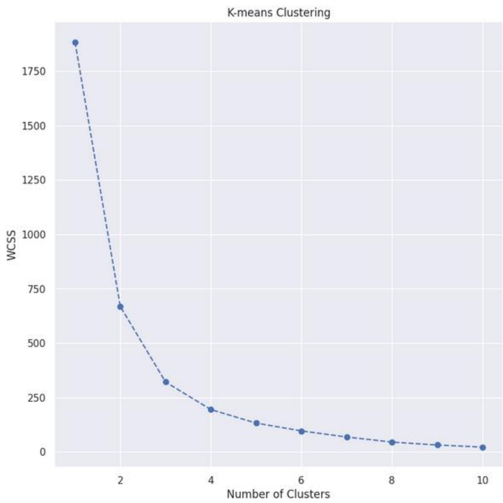

Figure 4. Number of clusters

After obtaining the number of clusters, then testing the existing cluster values using the silhouette score. Silhouette score is an evaluation metric used in clustering analysis to evaluate how well each data point matches the cluster it is in. This score describes how close the point is to the cluster it is in compared to other adjacent clusters. Figure 5 is the result of visualizing the silhouette score of each cluster.

```
from sklearn.metrics import silhouette_score
from sklearn.cluster import KMeans

for n_cluster in range(2, 5):
    kmeans = KMeans(n_clusters=n_cluster).fit(X)
    label = kmeans.labels_
    sil_coeff = silhouette_score(X, label, metric='euclidean')
    print("For n_clusters={}, The Silhouette Coefficient is {}".format(n_cluster, sil_coeff))

For n_clusters=2, The Silhouette Coefficient is 0.5733409050801208
For n_clusters=3, The Silhouette Coefficient is 0.5692149481198364
For n_clusters=4, The Silhouette Coefficient is 0.5689583498480574
```

Figure 5. Silhouette score value

The results of the silhouette score of Figure 5 show cluster 2 has a value of 0.5733, cluster 3 has a value of 0.5692, and cluster 3 has a value of 0.5689. This shows that the best score is owned by cluster 2, in determining the best students will use 2 clusters based on the highest silhouette score. The next stage of clustering using the K-Means algorithm shows the separation of data into several clusters based on patterns in the dataset. The K-Means clustering process begins with the random initialization of the cluster center point (centroid) within the data feature space. After initialization, the algorithm repeats two main steps: placement of data points into clusters that have a nearest center, and updating the centroid position of the cluster based on the average of the data points in the corresponding cluster. This iteration continues until there is no change in the placement of data points or the centroid position of the cluster, indicating the convergence of the algorithms. This process seeks to minimize the sum of squares of distances between each data point and its cluster center, thus ensuring that each point belongs to the cluster closest to it. Figure 6 is the result of clustering carried out with the best student data.

	Average Score	Attitude	Achievement	Activeness	Cluster
NIS					
1227294	83.59	89	93	91	1
1227295	84.34	89	93	91	1
1227296	81.22	86	90	88	0
1227297	82.10	87	91	89	0
1227298	81.96	87	91	89	0
...	...	...	...	...	...
1227761	81.43	86	90	88	0
1227762	81.00	86	90	88	0
1227763	86.14	91	95	93	1
1227764	78.86	84	88	86	0
1227765	80.93	86	90	88	0

471 rows × 5 columns

Figure 6. Clustering result

The results of the cluster Figure 6 show the results of clustering student data using the K-Means algorithm. The cluster labels result from the K-Means algorithm, which divides students into groups based on similar patterns from the data with labels 0 and 1. In K-Means clustering, the means and standard deviations of each cluster can provide valuable insights into the characteristics and variability within each cluster. The average of each cluster represents the center or average position of the data points in that cluster within the feature space. It serves as a representative point for the cluster. Understanding the mean and standard deviation of clusters helps in interpreting the characteristics of each cluster, identifying outliers or anomalies, as well as evaluating the density or distribution of data within the cluster. These statistics can help in making informed decisions and gaining applicable insights from clustering results. Figure 7(a) shows the results of the means and standard deviations of each cluster. The results of Figure 7(b) show that the best cluster is in cluster 0 because it has the highest value of means and standard deviations with the amount of data as much as 198 student data in cluster 0. Data from cluster 0 will be used in determining the best students using hybrid entropy-GRA.

```
# calculate the mean and standard deviation of each feature for each cluster
cluster_means = df_scaled_df.groupby(cluster_labels).mean()
cluster_stds = df_scaled_df.groupby(cluster_labels).std()

# print the cluster means and standard deviations
print(cluster_means)
print(cluster_stds)
```

```
          0         1         2         3
0  0.940346  0.944018  0.944018  0.944018
1 -0.682009 -0.684672 -0.684672 -0.684672
          0         1         2         3
0  0.726068  0.719785  0.719785  0.719785
1  0.489157  0.486996  0.486996  0.486996
```

(a)

```
df_segm_kmeans['Cluster'].value_counts()
```

```
Cluster
1    273
0    198
Name: count, dtype: int64
```

(b)

Figure 7. Clustering results for: (a) means and standard deviations and (b) the amount of data from each cluster

3.2. Hybrid entropy-grey relational analysis

Hybrid entropy-GRA is a combined method that combines the concepts of entropy weight and GRA. Basically, entropy weight is used to determine the relative weight of each criterion in the selection process, while GRA is used to calculate the relative proximity between alternatives and ideal solutions in the criteria space. In this context, entropy weight helps to reduce subjectivity in the determination of the weight of criteria by considering the degree of uncertainty or information diversity of each criterion, while GRA allows to evaluate the relative performance of alternatives in different criteria as well as determine their ranking. The combination of these two methods provides a more holistic approach to multi-criteria decision making by leveraging the strengths of each method. The data to be used in both methods are based on data from cluster 0 using K-Means, Table 2 is the result of data from cluster 0 in determining the best students. The first stage of normalizing the decision matrix in the entropy method is calculated using (2), Table 3 is the result of matrix normalization in the entropy method. The second stage of calculating the value of the criterion matrix in the entropy method is calculated using (3), Table 4 is the result of the value of the criterion matrix in the entropy method.

Table 2. Clustering results data

Number	NISN	Student name	Average value	Attitude	Achievement	Liveliness
1	1227294	Adinda Putri Utami	83.59	89	93	91
2	1227295	Ai Pirnawati	84.34	89	93	91
3	1227297	Akmaludin Muzaki	82.1	87	91	89
4	1227298	Alvin Cahya Imanda	81.96	87	91	89
5	1227299	Andreas	83.32	88	92	90
6	1227300	Arjun Surya Maulana	82.64	88	92	90
7	1227303	Desta Syuda Wirat	84.27	89	93	91
8	1227304	Ekta Imelda	83.22	88	92	90
9	1227305	Farhan	81.66	87	91	89
10	1227306	Fitri Anggraeni	82.41	87	91	89
…	…	…	…	…	…	…
198	1227763	Wulan Fitriyani	86.14	91	95	93

Table 3. Matrix normalization entropy method

Number	NISN	Student name	Average value	Attitude	Achievement	Liveliness
1	1227294	Adinda Putri Utami	0.943	0.947	0.949	0.948
2	1227295	Ai Pirnawati	0.951	0.947	0.949	0.948
3	1227297	Akmaludin Muzaki	0.926	0.926	0.929	0.927
4	1227298	Alvin Cahya Imanda	0.925	0.926	0.929	0.927
5	1227299	Andreas	0.940	0.936	0.939	0.938
6	1227300	Arjun Surya Maulana	0.932	0.936	0.939	0.938
7	1227303	Desta Syuda Wirat	0.951	0.947	0.949	0.948
8	1227304	Ekta Imelda	0.939	0.936	0.939	0.938
9	1227305	Farhan	0.921	0.926	0.929	0.927
10	1227306	Fitri Anggraeni	0.930	0.926	0.929	0.927
…	…	…	…	…	…	…
198	1227763	Wulan Fitriyani	0.972	0.968	0.969	0.969

Table 4. Entropy method criteria matrix value

Number	NISN	Student name	Average value	Attitude	Achievement	Liveliness
1	1227294	Adinda Putri Utami	0.00506	0.00508	0.00508	0.00508
2	1227295	Ai Pirnawati	0.00511	0.00508	0.00508	0.00508
3	1227297	Akmaludin Muzaki	0.00497	0.00497	0.00497	0.00497
4	1227298	Alvin Cahya Imanda	0.00496	0.00497	0.00497	0.00497
5	1227299	Andreas	0.00505	0.00503	0.00503	0.00503
6	1227300	Arjun Surya Maulana	0.00501	0.00503	0.00503	0.00503
7	1227303	Desta Syuda Wirat	0.00510	0.00508	0.00508	0.00508
8	1227304	Ekta Imelda	0.00504	0.00503	0.00503	0.00503
9	1227305	Farhan	0.00495	0.00497	0.00497	0.00497
10	1227306	Fitri Anggraeni	0.00499	0.00497	0.00497	0.00497
…	…	…	…	…	…	…
198	1227763	Wulan Fitriyani	0.00522	0.00520	0.00519	0.00519

The third stage of calculating the entropy value in the entropy method is calculated using (4), Table 5 is the result of the criteria matrix value in the entropy method. The fourth stage of calculating the dispersion value in the entropy method is calculated using (5), Table 6 is the result of the dispersion value in the entropy method. The fifth stage of calculating the value of the weight of the criterion in the entropy method is calculated using (6), Table 7 is the result of the value of the weight of the criterion in the entropy method.

Table 5. Value entropy criterion

Criteria	Average value	Attitude	Achievement	Liveliness
Entropy value	0.99996836	0.99997132	0.99997374	0.99997257

Table 6. Dispersion value

Criteria	Average value	Attitude	Achievement	Liveliness
Dispersion value	0.00003164	0.00002868	0.00002626	0.00002743

Table 7. Criterion weight value

Criteria	Average value	Attitude	Achievement	Liveliness
Weight value	0.00003164	0.00002868	0.00002626	0.00002743

The weighted results of Table 7 criteria will be the weights that will be used in determining the best students using the GRA method. The sixth stage of normalizing the decision matrix in the GRA method is calculated using (7), Table 8 is the result of matrix normalization in the GRA method. The seventh stage of calculating alternative relative weights in the GRA method is calculated using (8), Table 9 is the result of alternative relative weights in the GRA method.

Table 8. GRA method matrix normalization

Number	NISN	Student name	Average value	Attitude	Achievement	Liveliness
1	1227294	Adinda Putri Utami	0.2927	0.2857	0.2857	0.2857
2	1227295	Ai Pirnawati	0.3978	0.2857	0.2857	0.2857
3	1227297	Akmaludin Muzaki	0.084	0	0	0
4	1227298	Alvin Cahya Imanda	0.0644	0	0	0
5	1227299	Andreas	0.2549	0.1429	0.1429	0.1429
6	1227300	Arjun Surya Maulana	0.1597	0.1429	0.1429	0.1429
7	1227303	Desta Syuda Wirat	0.3880	0.2857	0.2857	0.2857
8	1227304	Ekta Imelda	0.2409	0.1429	0.1429	0.1429
9	1227305	Farhan	0.0224	0	0	0
10	1227306	Fitri Anggraeni	0.1275	0	0	0
…	…	…	…	…	…	…
198	1227763	Wulan Fitriyani	0.6499	0.5714	0.5714	0.5714

Table 9. GRA method matrix normalization

Number	NISN	Student name	Average value	Attitude	Achievement	Liveliness
1	1227294	Adinda Putri Utami	0.2927	0.2857	0.2857	0.2857
2	1227295	Ai Pirnawati	0.3978	0.2857	0.2857	0.2857
3	1227297	Akmaludin Muzaki	0.084	0	0	0
4	1227298	Alvin Cahya Imanda	0.0644	0	0	0
5	1227299	Andreas	0.2549	0.1429	0.1429	0.1429
6	1227300	Arjun Surya Maulana	0.1597	0.1429	0.1429	0.1429
7	1227303	Desta Syuda Wirat	0.3880	0.2857	0.2857	0.2857
8	1227304	Ekta Imelda	0.2409	0.1429	0.1429	0.1429
9	1227305	Farhan	0.0224	0	0	0
10	1227306	Fitri Anggraeni	0.1275	0	0	0
…	…	…	…	…	…	…
198	1227763	Wulan Fitriyani	0.6499	0.5714	0.5714	0.5714

The stages of depality calculating alternative grey relational values in the GRA method are calculated using (9), Table 10 is the result of alternative grey relational values in the GRA method. The final grade results of GRA Table 10 are the final results of calculations in the selection of the best students, Figure 8 is the result of visualization of the best student ranking based on the final score of GRA. The results of Figure 8 show the results of ranking the best students using the hybrid entropy-GRA method, for the first best student with a final grade of 0.25 obtained by Mareta Amelia, the second best student with a final grade of 0. 2486373 was obtained by Taufiq Qulmanan, and the third best student with a final grade of 0. 1866506 was obtained by Princess Handayani.

Table 10. Alternative grey relational values

Number	NISN	Student name	GRA value
1	1227294	Adinda Putri Utami	0.0719153
2	1227295	Ai Pirnawati	0.0792157
3	1227297	Akmaludin Muzaki	0.0058403
4	1227298	Alvin Cahya Imanda	0.0044776
5	1227299	Andreas	0.0435014
6	1227300	Arjun Surya Maulana	0.0368824
7	1227303	Desta Syuda Wirat	0.0785343
8	1227304	Ekta Imelda	0.0425280
9	1227305	Farhan	0.0015574
10	1227306	Fitri Anggraeni	0.0088578
…	…	…	…
198	1227763	Wulan Fitriyani	0

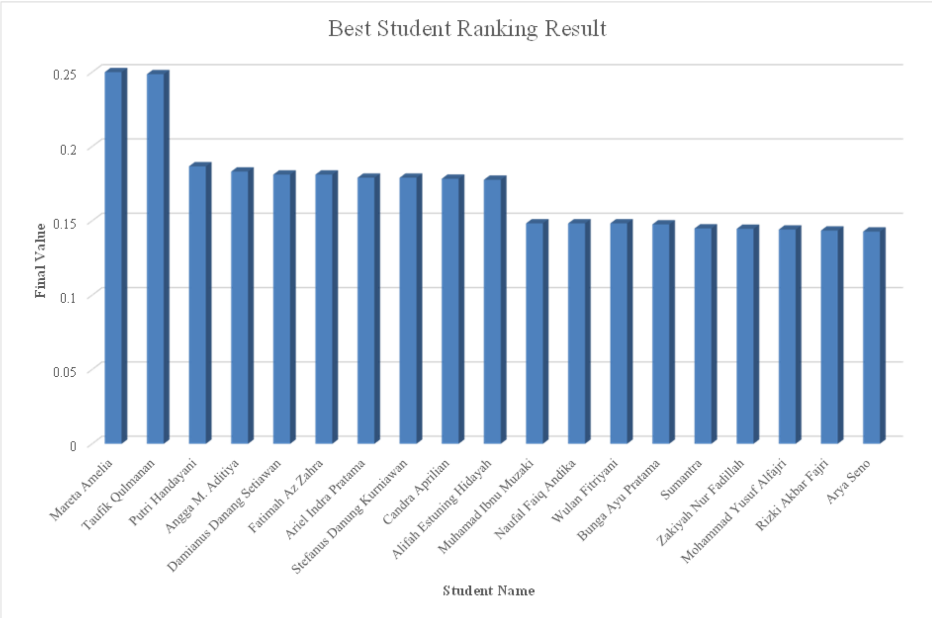

Figure 8. Best student ranking results

3.3. Discussion

The selection of the best students is an important task that has an impact on the recognition of academic excellence as well as other awards. Traditional methods often rely on a limited number of criteria, so they may overlook a comprehensive evaluation. A hybrid approach combining K-Means clustering and entropy-GRA offers a robust solution by integrating quantitative data analysis with multi-criteria decision-making techniques. K-Means clustering is an unsupervised learning algorithm that divides data into k different groups based on feature similarity. In the context of student selection, this algorithm groups students based on performance metrics, such as average grade, attitude, achievement, and activeness. These groupings help identify patterns and group students who have similar academic and non-academic profiles.

Entropy-GRA improves grouping outcomes by evaluating and ranking students within each group. Entropy measures the amount of irregularity or uncertainty in the data, assigning weight to each criterion based on its variability. GRA then uses these weights to determine the relationship between individual student performance and the ideal best student profile. The hybrid approach leverages the power of both methods K-Means clustering provides a structured way to categorize students into manageable cohorts based on performance metrics. Within each group, entropy-GRA refines the selection by ranking students based on several criteria, ensuring a fair and comprehensive evaluation.

The hybrid approach of K-Means and entropy-GRA offers powerful tools to improve decision-making in the student selection process. By integrating grouping and multi-criteria analysis, this method ensures a more fair and comprehensive evaluation. Future work could focus on automating parameter selection for K-Means and developing more efficient algorithms to handle large datasets. In addition, incorporating feedback from education stakeholders can further refine the selection criteria and improve the overall process. Optimizing the selection of the best students requires a method that can handle a variety of performance metrics and provide a balanced evaluation. The hybrid approach of K-Means clustering and entropy-GRA presents a robust solution, improving decision-making processes and ensuring that high-achieving students are accurately recognized and rewarded. Our research shows that the hybrid approach of K-Means clustering and entropy-GRA is more effective in selecting the best students than the traditional method of using only methods from DSSs. Future research may explore the application of this method in different contexts, such as the selection of students for scholarship programs, as well as develop practical methods for integrating artificial intelligence (AI)-based technologies in this decision-making process.

4. CONCLUSION

The research objective of using hybrid GRA and K-Means clustering in the selection of the best students is to develop a more objective, accurate, and comprehensive assessment system. The results of the silhouette score showed that 2 clusters had a value of 0.5733, so in this study using 2 clusters. The results of clustering using the K-Means algorithm show that the best cluster is in cluster 0 because it has the highest value of means and standard deviations with the amount of data as much as 198 student data in cluster 0. Data from cluster 0 will be used in determining the best students using hybrid entropy-GRA. The results of ranking the best students using the hybrid entropy-GRA method, for the first best student with a final grade of 0.25 were obtained by Mareta Amelia, the second best student with a final grade of 0. 2486373 was obtained by Taufiq Qulmanan, and the third best student with a final grade of 0. 1866506 was obtained by Princess Handayani.

The hybrid approach of K-Means and entropy-GRA offers powerful tools to improve decision-making in the student selection process. By integrating grouping and multi-criteria analysis, this method ensures a more fair and comprehensive evaluation. Future work could focus on automating parameter selection for K-Means and developing more efficient algorithms to handle large datasets. In addition, incorporating feedback from education stakeholders can further refine the selection criteria and improve the overall process. Optimizing the selection of the best students requires a method that can handle a variety of performance metrics and provide a balanced evaluation. The hybrid approach of K-Means clustering and entropy-GRA presents a robust solution, improving decision-making processes and ensuring that high-achieving students are accurately recognized and rewarded. Further research can also explore the integration of machine learning technology to improve prediction accuracy in the selection of the best students. Recent observations show that the hybrid approach of K-Means clustering and entropy-GRA significantly improves accuracy in the selection of the best students. Our findings provide clear evidence that this increase in accuracy is associated with more efficient and systematic data processing, rather than relying solely on traditional evaluation methods that may be less comprehensive.

ACKNOWLEDGEMENTS

Thank you to the Universitas Teknokrat Indonesia for funding this research with research contract number 023/UTI/LPPM/E.1.1/IV/2023.

REFERENCES

[1] F. E. Palominos, S. K. Palominos, F. M. Córdova, and H. Díaz, "Study of the achievement of university students and their relationship with the selection process," *Procedia Computer Science*, vol. 122, pp. 985–992, 2017, doi: 10.1016/j.procs.2017.11.464.

[2] J. Li, R. Ma, M. Deng, X. Cao, X. Wang, and X. Wang, "A comparative study of clustering algorithms for intermittent heating demand considering time series," *Applied Energy*, vol. 353, p. 122046, Jan. 2024, doi: 10.1016/j.apenergy.2023.122046.

[3] S. Ding *et al.*, "A sampling-based density peaks clustering algorithm for large-scale data," *Pattern Recognition*, vol. 136, p. 109238, Apr. 2023, doi: 10.1016/j.patcog.2022.109238.

[4] M. Li *et al.*, "High-performance plant pest and disease detection based on model ensemble with inception module and cluster algorithm," *Plants*, vol. 12, no. 1, p. 200, Jan. 2023, doi: 10.3390/plants12010200.

[5] M.-S. Yang and I. Hussain, "Unsupervised multi-view k-means clustering algorithm," *IEEE Access*, vol. 11, pp. 13574–13593, 2023, doi: 10.1109/ACCESS.2023.3243133.

[6] S. Zhang, R. Wang, and Z. Lin, "Subzone division optimization with probability analysis-based K-Means clustering for coupled control of non-uniform thermal environments and individual thermal preferences," *Building and Environment*, vol. 249, p. 111155, Feb. 2024, doi: 10.1016/j.buildenv.2023.111155.

[7] A. Qtaish, M. Braik, D. Albashish, M. T. Alshammari, A. Alreshidi, and E. J. Alreshidi, "Optimization of K-Means clustering method using hybrid capuchin search algorithm," *The Journal of Supercomputing*, vol. 80, no. 2, pp. 1728–1787, Jan. 2024, doi: 10.1007/s11227-023-05540-5.

[8] D. Festa *et al.*, "Unsupervised detection of insar time series patterns based on PCA and K-Means clustering," *International Journal of Applied Earth Observation and Geoinformation*, vol. 118, p. 103276, Apr. 2023, doi: 10.1016/j.jag.2023.103276.

[9] A. M. Bueno, I. Mendes da Luz, I. L. Niza, and E. E. Broday, "Hierarchical and k-means clustering to assess thermal dissatisfaction and productivity in university classrooms," *Building and Environment*, vol. 233, p. 110097, Apr. 2023, doi:

10.1016/j.buildenv.2023.110097.
[10] S. Tiwari *et al.*, "A smart decision support system to diagnose arrhythmia using ensembled convnet and convnet-LSTM model," *Expert Systems with Applications*, vol. 213, p. 118933, 2023, doi: 10.1016/j.eswa.2022.118933.
[11] L. K. Singh, M. Khanna, and R. Singh, "Artificial intelligence based medical decision support system for early and accurate breast cancer prediction," *Advances in Engineering Software*, vol. 175, Jan. 2023, doi: 10.1016/j.advengsoft.2022.103338.
[12] A. Frisinger and P. Papachristou, "The voice of healthcare: introducing digital decision support systems into clinical practice - a qualitative study," *BMC Primary Care*, vol. 24, no. 1, p. 67, Mar. 2023, doi: 10.1186/s12875-023-02024-6.
[13] H. Sulistiani, Setiawansyah, P. Palupiningsih, F. Hamidy, P. L. Sari, and Y. Khairunnisa, "Employee performance evaluation using multi-attribute utility theory (MAUT) with piprecia-s weighting: a case study in education institution," in *2023 International Conference on Informatics, Multimedia, Cyber and Informations System (ICIMCIS)*, 2023, pp. 369–373, doi: 10.1109/ICIMCIS60089.2023.10349017.
[14] S. Guo and Y. Wang, "Investigating predictors of juvenile traditional and/or cyber offense using machine learning by constructing a decision support system," *Computers in Human Behavior*, vol. 152, p. 108079, Mar. 2024, doi: 10.1016/j.chb.2023.108079.
[15] Q. Ma and H. Li, "A decision support system for supplier quality evaluation based on MCDM-aggregation and machine learning," *Expert Systems with Applications*, vol. 242, p. 122746, May 2024, doi: 10.1016/j.eswa.2023.122746.
[16] H. Yousefi, S. Moradi, R. Zahedi, and Z. Ranjbar, "Developed analytic hierarchy process and multi criteria decision support system for wind farm site selection using gis: a regional-scale application with environmental responsibility," *Energy Conversion and Management: X*, vol. 22, p. 100594, 2024, doi: 10.1016/j.ecmx.2024.100594.
[17] H. Liu and Z. Chang, "Multi-objective optimization of temperature uniformity in the immersion liquid cooling cabinet with taguchi-based grey relational analysis," *International Communications in Heat and Mass Transfer*, vol. 154, p. 107395, May 2024, doi: 10.1016/j.icheatmasstransfer.2024.107395.
[18] B. Zhang *et al.*, "Assessment of the impact of pyrolysis conditions on char reactivity through orthogonal experimental-based grey relational analysis," *Journal of Analytical and Applied Pyrolysis*, vol. 179, p. 106426, May 2024, doi: 10.1016/j.jaap.2024.106426.
[19] R. Hasanzadeh, P. Mojaver, T. Azdast, S. Khalilarya, A. Chitsaz, and M. A. Rosen, "Decision analysis for plastic waste gasification considering energy, exergy, and environmental criteria using topsis and grey relational analysis," *Process Safety and Environmental Protection*, vol. 174, pp. 414–423, Jun. 2023, doi: 10.1016/j.psep.2023.04.028.
[20] S. A. Javed, A. Gunasekaran, and A. Mahmoudi, "DGRA: multi-sourcing and supplier classification through dynamic grey relational analysis method," *Computers & Industrial Engineering*, vol. 173, p. 108674, Nov. 2022, doi: 10.1016/j.cie.2022.108674.
[21] D. Zhao, C. Li, Q. Wang, and J. Yuan, "Comprehensive evaluation of national electric power development based on cloud model and entropy method and topsis: a case study in 11 countries," *Journal of Cleaner Production*, vol. 277, p. 123190, Dec. 2020, doi: 10.1016/j.jclepro.2020.123190.
[22] S. J. Siriyapuraju, V. S. Gowri, S. Balla, M. K. Vanika, and A. Gandhi, "DoS and DDOS attack detection using mathematical and entropy methods," in *2023 2nd International Conference on Paradigm Shifts in Communications Embedded Systems, Machine Learning and Signal Processing (PCEMS)*, IEEE, Apr. 2023, pp. 1–6, doi: 10.1109/PCEMS58491.2023.10136042.
[23] Setiawansyah, A. A. Aldino, P. Palupiningsih, G. F. Laxmi, E. D. Mega, and I. Septiana, "Determining best graduates using topsis with surrogate weighting procedures approach," in *2023 International Conference on Networking, Electrical Engineering, Computer Science, and Technology (IConNECT)*, 2023, pp. 60–64, doi: 10.1109/IConNECT56593.2023.10327119.
[24] H. Vijayan, S. M, and S. K, "A-MKMC: an effective adaptive-based multilevel k-means clustering with optimal centroid selection using hybrid heuristic approach for handling the incomplete data," *Data & Knowledge Engineering*, vol. 150, p. 102243, Mar. 2024, doi: 10.1016/j.datak.2023.102243.
[25] A. Poggiali, A. Berti, A. Bernasconi, G. M. Del Corso, and R. Guidotti, "Quantum clustering with K-Means: a hybrid approach," *Theoretical Computer Science*, vol. 992, p. 114466, Apr. 2024, doi: 10.1016/j.tcs.2024.114466.
[26] A. A. Aldino, D. Darwis, A. T. Prastowo, and C. Sujana, "Implementation of k-means algorithm for clustering corn planting feasibility area in South Lampung regency," in *Journal of Physics: Conference Series*, IOP Publishing, 2021, p. 12038.
[27] Parjito, Z. Abidin, A. Pambudi, D. A. Tyas, E. R. Susanto, and R. R. Suryono, "Prediction of telecommunication company stock price using multiple linear regression," in *2023 3rd International Conference on Smart Cities, Automation & Intelligent Computing Systems (ICON-SONICS)*, IEEE, Dec. 2023, pp. 147–152, doi: 10.1109/ICON-SONICS59898.2023.10435135.
[28] C. Zhong, Q. Yang, J. Liang, and H. Ma, "Fuzzy comprehensive evaluation with ahp and entropy methods and health risk assessment of groundwater in Yinchuan Basin, Northwest China," *Environmental Research*, vol. 204, p. 111956, Mar. 2022, doi: 10.1016/j.envres.2021.111956.
[29] A. I. Petrov, "Entropy method of road safety management: case study of the Russian federation," *Entropy*, vol. 24, no. 2, p. 177, Jan. 2022, doi: 10.3390/e24020177.

BIOGRAPHIES OF AUTHORS

Heni Sulistiani is a Lecturer at the Faculty of Engineering and Computer Science, Universitas Teknokrat Indonesia, Lampung, Indonesia. She completed of informatics engineering in STMIK Teknokrat in 2012. She holds a master of information systems from Institut Teknologi Sepuluh Nopember (ITS), Surabaya in 2016. She holds a Doctoral of Computer Science from Lampung University, Indonesia in 2024. She mainly researches data mining, machine learning, and software engineering. She can be contacted at email: henisulistiani@teknokrat.ac.id.

Setiawansyah is a Lecturer of Informatics Study Program, Universitas Teknokrat Indonesia. He completed his information systems education at STMIK Teknokrat in 2013. He earned a master's degree in computer science from Budi Luhur University in 2018. He mainly researches decision support systems and data mining. He can be contacted at email: setiawansyah@teknokrat.ac.id.

Pritasari Palupiningsih is a Lecturer of Informatics Engineering Study Program, Institut Teknologi Perusahaan Listrik Negara. Her research focuses on computer science. She can be contacted at email: pritasari@itpln.ac.id.

Ahmad Ferico Octaviansyah Pasaribu is a Lecturer of Information Technology Study Program, Universitas Teknokrat Indonesia. He mainly researches decision support systems, artificial intelligence and data mining. He can be contacted at email: fericopasaribu@teknokrat.ac.id.

Rio Andika is a student of Information Technology Study Program, Universitas Teknokrat Indonesia. He mainly researches decision support systems, artificial intelligence and data mining. He can be contacted at email: rio_andika@teknokrat.ac.id.

Muhammad Hamdan Sobirin is a student of Information Technology Study Program, Universitas Teknokrat Indonesia. He mainly researches decision support systems, artificial intelligence and data mining. He can be contacted at email: muhammad_hamdan_sobirin@teknokrat.ac.id.