

Best Programming Creator Content Selection with K-Means Clustering Algorithm and MAUT Method

Witari Aryunani, Yeni Setiani*, Indra Purnama

Information Systems, Faculty of Computer Science and Information Technology, Gunadarma University, Depok, Indonesia
E-mail : ¹ witari_aryunani@staff.gunadarma.ac.id, ^{2,*} yeni_setiani@staff.gunadarma.ac.id, ³ indrapumama@staff.gunadarma.ac.id
Correspondence Author Email: yeni_setiani@staff.gunadarma.ac.id
Submitted: **08/05/2025**; Accepted: **04/06/2025**; Published: **05/06/2025**

Abstract—Selecting quality programming content creators on platforms such as YouTube is becoming a complex challenge as digital educational content expands. This research designs a systematic approach by combining K-Means algorithm and MAUT method to objectively evaluate and rank creators. Data from 100 programming channels was analysed using K-Means, resulting in three main clusters based on audience views and interactions. The leading cluster was identified with an average of 335,461 views per video and an engagement rate of 0.31%. The MAUT method then ranked the creators in this cluster, revealing Brackeys as the best programming content creator with an optimal balance between audience reach and participation with a final score of 0.624. The results show that the integration of these two methods is effective in providing a data-driven solution for educational content selection, as well as a reference for creators in improving the quality of the material. The combination of K-Means and MAUT not only answers the need for objectivity in content curation, but also enriches the literacy of multidimensional evaluation methods in the era of online learning.

Kata Kunci: Clustering; K-Means Algorithm; DSS; MAUT Method; Programming Creator Content

1. INTRODUCTION

The digital era has brought about significant changes in the way learning is conducted, especially in the field of programming. Online learning strategies have a significant impact on students' academic performance, with an emphasis on the importance of using appropriate teaching methods to improve learning outcomes[1]. Content creators are now modern educators who provide easily accessible learning materials through various platforms, such as YouTube. YouTube content creators have shown a positive impact in improving the learning experience[2], [3], [4]. However, as the number of content creators increases, new challenges arise in identifying high-quality creators in an objective and structured manner. A systematic selection mechanism plays an important role in enhancing online community engagement[5]. This situation demands an approach that is capable of evaluating and categorizing content creators based on various relevant criteria.

Various methods have been proposed to address these challenges. Data mining and Multi-Criteria Decision Analysis (MCDA) methods have shown their effectiveness in systematically evaluating alternatives[6], [7]. In data mining, K-means Clustering as a solution for optimizing data distribution in machine learning, which is relevant in grouping content creators based on their characteristics[8]. On the other hand, the MAUT (Multi-Attribute Utility Theory) method has been proven effective in multi-criteria decision making in decision support systems. An innovative MAUT framework to improve the Decision-making process, with a comprehensive guide on implementing this method [7][9]. For Clustering optimization, the use of the Elbow method to determine the optimal number of Clusters in K-Means provides a strong methodological foundation[10], [11]. In this study, data was obtained through scraping using the YouTube API to collect information about content creators, such as the number of followers, interaction levels, and content quality. The K-Means algorithm was used to group creators based on these data patterns. Furthermore, the MAUT method was applied to rank the groups that had been formed, taking into account the weight of the relevant criteria[12], [13].

Previous research conducted by Ridha Maya Faza Lubis and her team in 2023 focused on efforts to reduce the increasing death rate from cervical cancer. In this study, they used clustering techniques to group data based on similar characteristics. The K-Means algorithm was implemented with the help of the RapidMiner application. The results showed that cluster 1 had the largest amount of data. Of the 72 cervical cancer data analyzed, 28 were classified as cervical cancer sufferers, while the other 44 data were not included in that category[14].

Ayu Pangestu and Taufik Ridwan in 2022 aimed to group water customers based on the volume of water usage in cubic units. This grouping is done to provide information on customer categories based on water sales data at PAM Kerta Raharja. In this study, the method used is the K-Means algorithm. Water sales data that has been processed using the Weka application is grouped into three categories, namely economical, moderate, and wasteful. The results of the calculation using the K-Means algorithm show that there are three cluster groups. Cluster 0 with a centroid (46.6) is a group of customers who are classified as wasteful, consisting of 9 people. Cluster 1 with a centroid (13.6) is a group of customers with moderate water usage, while cluster 2 with a centroid (25.4) is a group of customers who are economical in water usage[15].

Mohammad Aldinugroho Abdullah and Rima Tamara Aldisa in 2023 highlighted the importance of Decision Support Systems (DSS) in warehouse manager recruitment. The conventional recruitment process carried out by HRD with various requirements often makes prospective candidates feel difficult and reluctant to apply. As a result, HRD also has difficulty finding the right candidate. DSS is present as a solution to overcome this problem. DSS helps HRD

in identifying and analyzing the criteria needed for the warehouse manager position, as well as providing a list of the most suitable candidates. In the study, the application of DSS with the MAUT method and ROC weighting resulted in the best decision, namely a candidate named Dindin Samsudin with a U_i value = 0.761. These results indicate that Dindin Samsudin is the most qualified candidate for the warehouse manager position[16].

Agung Triayudi and his team in 2022 highlighted the importance of making the right and effective decisions in determining superior regional products that need to be prioritized. In the midst of increasingly tight competition, local governments (PEMDA) are required to choose the best products. Decision Support Systems (DSS) can be a solution by helping to identify and analyze priority criteria, as well as producing alternative products that have the potential to be developed. Based on the application of the ROC and MAUT methods, coffee products are ranked first with the highest utility value, making them the most potential superior regional product to be developed[17].

In 2023, Eugenius Kau Suni and Stephen Aprius Sutresno conducted a study, this study focused on developing a decision support system to evaluate the performance of content creators at the Aurora News Agency, which has more than 700 members. The TOPSIS method was used to determine the ranking of the best content creators based on data from 630 content creators with a total of 10,916 contents. The results showed that Anemz Tv was ranked first with a preference value of 0.4368, followed by Talenta.TV in second place with a value of 0.3203. Further analysis revealed differences in strategies between content creators, with some focusing on quantity and others on quality of content. In conclusion, the TOPSIS method can be implemented simply with fast computing and produces diverse preference values[18].

This combined approach integrates the strengths of K-Means in data clustering and MAUT in weighted evaluation, offering significant advantages over single methods. With this approach, the study is expected to provide a systematic solution to help educational platforms curate content, support creators in improving their quality, and make it easier for users to choose the best learning resources.

2. RESEARCH METHODOLOGY

2.1 Research Stages

Good research follows a structured and coherent flow. Each step in this flow contributes to the quality of the final research result. The stages in this research can be likened to a map that guides the researcher from the beginning to the end of the research process. These stages are visualized in Figure 1.

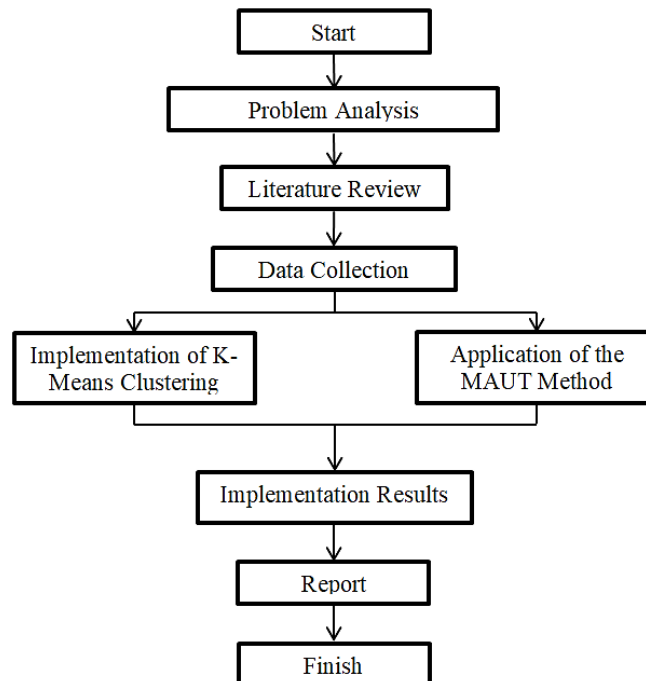


Figure 1. Research Stages

Based on Figure 1, this research begins with a problem analysis which is a crucial stage to identify and understand the problems to be studied. This stage involves collecting relevant information and data. A good problem analysis will determine the direction and focus of the research, so that the research does not lose direction and produces useful results. Furthermore, a literature study is conducted to examine the research topic by collecting, analyzing, and synthesizing information from various trusted sources systematically and in a structured manner. The third stage is data collection to answer the problem formulation. The collected data is then processed using an appropriate algorithm. In this study, the K-Means clustering algorithm is used and evaluated with the RapidMiner platform. One of the

clusters is then selected as a sample for decision making using the MAUT method. The implementation of research is not only a complement, but also an important element that shows the real impact of research on society. The results of the implementation are proof that research can answer needs and provide solutions to problems in the field. Implementation shows the practical application of research and its benefits for life, so that research does not only stop in the academic realm, but also contributes positively to the progress of society. The final stage is writing a research report that presents the findings and results of the research comprehensively. This report must be written in language that is easy to understand and systematically structured.

2.2 Data Mining

Data Mining is the process of identifying valuable patterns and insights from massive data sets. It uses statistical, mathematical, and artificial intelligence techniques to extract useful information from data. Data mining can be applied across a range of sectors, including marketing, finance, healthcare, and e-commerce, to uncover trends, make predictions, and improve decision-making efficiency [19], [20], [21], [22].

2.3 Decision Support System

Decision Support System (DSS) is a computer-based system designed to facilitate decision-making in an organization or company. DSS provides information, data analysis, and various tools that assist in information processing. The main advantage of DSS is its ability to handle semi-structured or unstructured problems. Structured problems have predetermined solutions, while semi-structured and unstructured problems have more complex and situational solutions. In this context, DSS provides comprehensive information to decision makers to help them find the most appropriate solution. DSS uses various methods in decision-making, including PSI, EXPROM II, WASPAS, OCRA, TOPSIS, MAUT, SAW, COPRAS, MOORA, AHP, ARAS, WP, EDAS, and MOOSRA [23], [24], [25], [26], [27], [28], [29], [30], [31].

2.4 K-Means Clustering

K-Means is a method used in data analysis to group data into several groups based on existing attributes. The main goal of the K-Means algorithm is to find the center of the group called the centroid, where data with similar characteristics will be grouped together. The initial process of this algorithm involves determining the desired number of groups (k) and setting the initial position of the centroid randomly. Next, the algorithm will calculate the distance between each data and the centroid and group the data to the closest centroid. After that, the position of the centroid is updated by calculating the average of the data in each group. This process is repeated continuously until there is no change in the placement of the data or reaches a predetermined stopping criterion [32], [33], [34], [35], [36], [37]. The stages of K-Means Clustering start with determining of the number of clusters. After that, in the first iteration calculating of the centroid center by using randomly taken data values as the initial centroid center. After determining the initial centroid, the next step is calculating the shortest distance using the initial centroid, for the formula can be seen in equation 1 below.

$$d_{Euclidean}(X, Y) = \sqrt{\sum_{i=1}^n (X_i - Y_i)^2} \quad (1)$$

$d(x, y)$ is the distance of data x to the center of cluster y , X_i is the i -th data on the n -th data attribute and i is the j -th data on the n -th data attribute. Data that has the closest value will be grouped together in one group (cluster), while data that has a greater distance will be placed in a different group.

Perform the next iteration using the new centroid position, which is determined based on the group with the minimum distance to the data. The formula can be seen in equation 2 below.

$$K_i = \frac{1}{M} \sum_{j=1}^M X_j \quad (2)$$

This process is repeated from the first step until it reaches the last iteration. If there is a change in the group centroid position, then the iteration will continue. However, if there is no change in the centroid position of the group, then the iteration process will be stopped.

2.5 MAUT Method

The Multi-Attribute Utility Theory (MAUT) method is an approach to decision making that involves comparing multiple options based on a set of criteria. Each criterion has a weight that reflects its level of importance in the decision-making process. To determine the best option, the utility value of each option is calculated by multiplying the weight of each criterion by the value of the criterion in question. The option with the highest utility value is considered to be the option that best suits the user's preferences. In the MAUT method, criteria weights are expressed on a scale of 0 to 1, where 0 indicates that the criterion is not important, while 1 indicates that the criterion is very important. The utility value of each option can be directly compared to determine the best option [38], [39], [40], [41], [42], [43], [44], [45], [46], [47], [48], [49], [50], [51], [52], [53]. The initial process that needs to be done in applying the MAUT method is to form a decision matrix that can be seen in the following equation 3 which comes from the sample data in the study.

$$x_{ij} = \begin{bmatrix} x_{11} & x_{12} & x_{1n} \\ x_{21} & x_{22} & x_{2n} \\ x_{m1} & x_{m2} & x_{mn} \end{bmatrix} \tag{3}$$

Where m is the number of candidate alternatives, n is the number of evaluation criteria and X_{ij} is the performance of the alternatives with respect to the performance of j. The next step is to perform normalisation. For benefit criteria use equation 4, while for cost criteria use equation 5.

$$r_{ij}^* = \frac{x_{ij} - \min(x_{ij})}{\max(x_{ij}) - \min(x_{ij})} \tag{4}$$

$$r_{ij}^* = 1 + \left(\frac{\min(x_{ij}) - (x_{ij})}{\max(x_{ij}) - \min(x_{ij})} \right) \tag{5}$$

After the normalisation process, then calculate the Marginal Utility (U_{ij}) value of each alternative using the following equation 6.

$$U_{ij} = \frac{\exp(r_{ij}^*) - 1}{1.71} \tag{6}$$

Calculating the Final Utility Value (U_i) is the last stage in applying the MAUT method. The process of calculating the utility value is done by multiplying the weight of each criterion with the value of the criteria concerned. This utility value reflects how well an option fulfils the desired criteria, for the formula can be seen in equation 7 below.

$$U_i = \sum_j^n = 1 U_{ij} \cdot W_j \tag{7}$$

The utility value (u_j) indicates the level of suitability of each alternative. This value is calculated by combining the attribute weights (w_j) and the actual level (u_{ij}) of each alternative. The alternative with the highest utility value is the most recommended choice.

3. RESULTS AND DISCUSSION

This study applies a quantitative approach by integrating the Clustering method and multi-criteria ranking. The selection of this approach is based on the need to produce objective and measurable analysis in identifying programming content creator performance patterns. The Clustering method was chosen because of its ability to identify natural structures and groupings that emerge from data based on characteristic similarities, thus providing a more systematic understanding of content creator segmentation in the online programming learning ecosystem. To achieve a comprehensive evaluation, this study implements the Multi-Attribute Utility Theory (MAUT) method in the content creator ranking process. MAUT. The following is the general architecture of the use of both methods in the research used in completing this research.

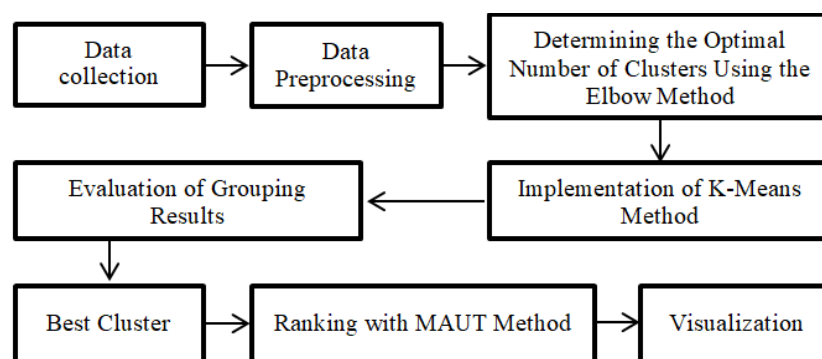


Figure 2. General Architecture

Based on Figure 2, this study carries out a series of analysis processes starting from data collection to visualization of results. The initial stage begins with data collection, followed by data preprocessing to ensure the quality and consistency of the data to be analyzed. Data preprocessing is crucial because it is the foundation for accurate analysis in subsequent stages. After the data is ready to be analyzed, the process continues to determine the optimal number of Clusters using the Elbow Method. This method was chosen because of its ability to determine the right number of Clusters based on the decrease in the Sum of Square Error (SSE) value, thus providing an objective basis for determining the number of Clusters. The results of the Elbow Method are then used as input for the Clustering process using the K-Means algorithm. After the Cluster is formed, an evaluation is carried out to ensure the quality of the resulting grouping. The final stage of the analysis process involves ranking using the MAUT method to provide an assessment of the best content creators in the Cluster. The results of the analysis are then visualized and stored to



facilitate interpretation and use of the research results. This series of processes is designed systematically to ensure comprehensive and scientifically accountable analysis results.

3.1 Data Collection and Preprocessing

In this study, the data collection process was carried out by utilizing the YouTube Data API v3 implemented through Google Apps Script integrated with Google Spreadsheet. This approach was chosen because of its ability to automate the data retrieval process efficiently and in a structured manner. The YouTube API provides access to various important metrics needed for content creator analysis, including Channel statistics and video performance. The data scraping process was designed to collect information from 100 selected programming Channels. The implementation was carried out using several main functions that were integrated with each other, the data collection process (data scraping) from YouTube using Google Apps Script in a spreadsheet. The process begins by receiving input in the form of video duration in ISO 8601 format, the dates of the first and last videos, and the number of videos. The `parseDuration` function is used to convert the video duration from ISO 8601 format to total seconds. Furthermore, the `calculateUploadFrequency` function calculates the frequency of video uploads based on the dates of the first and last videos and the number of videos. Finally, the `scrapeContentCreator` function scrapes data from YouTube. This function opens a spreadsheet, reads a list of content creator names, and retrieves data from the YouTube API for each content creator. The extracted data is then processed and stored in a spreadsheet. This process also addresses potential errors that may occur during data collection. The data collected includes key metrics needed for in-depth analysis such as Channel Statistics (subscriber count, total views, video count), Engagement metrics (likes, comments, views per video), Content consistency (upload frequency, video duration), and Historical performance (Channel age, growth metrics). The following is a sample of data that will be used in this study.

Table 1. Sample Data

channel_title	number_of videos	total_views	average_views_per video	...	average_video duration
Jombo	30	47102748	1570091,6	...	90,47
Programming Tutorials	292	3217408	11018,52	...	159,59
CroatCode	48	55778098	1162043,71	...	450,52
Programming with Mosh	230	224093242	974318,44	...	85,99
Telusko	2013	302619924	150332,8	...	79,15
Programiz	176	12037623	68395,59	...	320,34
Fireship	698	524927114	752044,58	...	20,2
The Coding Sloth	33	12598319	381767,24	...	644,82
w3schools.com	301	8656962	28760,67	...	7,08
ThePrimeTime	1122	175116363	156075,19	...	81,36
Hablu Programmer	719	13624174	18948,78	...	122,63
Coding with Lewis	630	184241311	292446,53	...	61,1
Tim Kim	8	7192333	899041,63	...	585
freeCodeCamp.org	1789	819551736	458106,06	...	508,79
...
...
AdiCodeHub	91	749496	8236,22	...	117,97
TitanTech	130	9542396	73403,05	...	15,25

After data preprocessing is carried out, the data samples that will be used in the grouping process become 76 data from a total of 100 initial data samples.

3.2 Determining the Number of Clusters

After the data is collected and processed, the next step is to determine the most appropriate number of clusters to group content creators using the K-Means algorithm. Choosing the right number of clusters is crucial to ensure that the clustering results produce relevant and easily interpretable groups. In this study, the Elbow method is used, which is based on the calculation of the Within-Cluster Sum of Squares (WCSS) for various values of the number of clusters (k). This method selects a value of k that results in a significant decrease in WCSS, which measures how well the data is grouped within each cluster. This process begins with data pre-processing, where raw data that may contain missing values or anomalies is corrected to ensure clean and representative data. Next, data normalization is performed using `RobustScaler` to address the problem of uneven data distribution, keeping each variable on an equal scale and reducing the impact of anomalies. Without normalization, variables with a larger range of values can dominate the analysis. After the data is normalized, the K-Means algorithm is applied to the data for various values of k ranging from 2 to 10, and WCSS is calculated for each value of k. The results are used to construct an Elbow curve, which visualizes the decrease in WCSS and helps determine the optimal number of clusters. The point at which the decrease in WCSS begins to slow down indicates the best number of clusters, as adding further clusters does not provide significant improvement in data separation. This process ensures that the number of clusters chosen best describes the structure



of the data, allowing for more effective and accurate analysis in subsequent steps. Next, the results of the WCSS calculation for various values of k are used to construct an Elbow curve. This curve visualizes the decrease in WCSS as the number of clusters increases. The point at which the decrease in WCSS begins to slow down (elbow point) indicates the optimal number of clusters chosen. The following is a visualization of the Elbow curve that illustrates the results of this analysis.

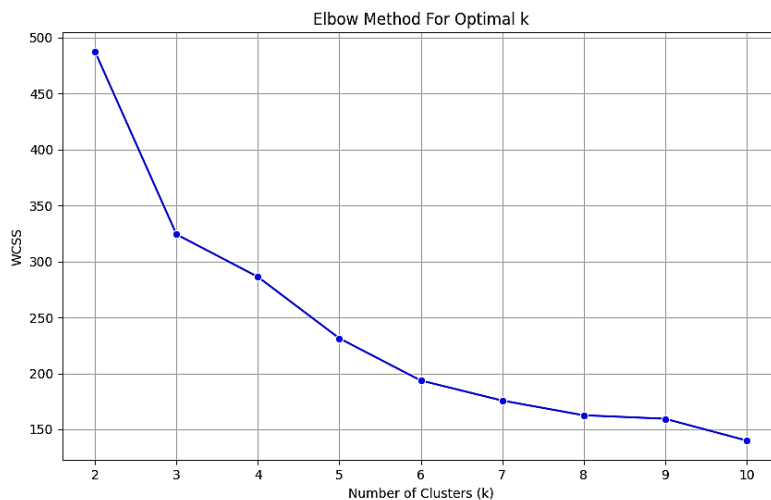


Figure 3. Elbow Method Results

Based on the Elbow method analysis in Figure 3, there is a significant decrease in the Within-Cluster Sum of Squares (WCSS) value until it reaches K = 3. After that point, the decrease in WCSS slows down and forms a kind of 'elbow' on the curve. This indicates that increasing the number of clusters after K = 3 does not provide a significant increase in performance, but rather increases the complexity of the model. Therefore, the optimal number of clusters selected for further analysis is K = 3.

3.3 Implementation of K-Means Method

After determining the optimal number of Clusters, the next step is to apply the K-Means algorithm to group the Channels based on similar characteristics. This process aims to form homogeneous groups, thus facilitating further analysis of patterns and relationships between Channels. The following table 2 contains the final clustering results after applying the K-Means method.

Table 2. Clustering results using the K-Means method

channel_title	number_of videos	total_views	average_views per video	average_vid eo_duration	Cluster
Programming Tutorials	292	3217408	11018,52	159,59	0
Telusko	2013	302619924	150332,8	79,15	0
Programiz	176	12037623	68395,59	320,34	0
w3schools.com	301	8656962	28760,67	7,08	0
ThePrimeTime	1122	175116363	156075,19	81,36	0
Hablu Programmer	719	13624174	18948,78	122,63	0
Coding with Lewis	630	184241311	292446,53	61,1	2
freeCodeCamp.org	1789	819551736	458106,06	508,79	2
Ethio Programming	767	8081349	10536,31	34,35	0
Learn Coding	937	241120641	257332,59	154,16	2
Sahil & Sarra	200	69838869	349194,34	35,99	2
Aaron Jack	122	36680208	300657,44	234,41	1
The StudyTube Project	119	3833897	32217,62	346,51	0
Pentanik IT Solution Park	1369	15250727	11140,05	2,93	0
...
...
AdiCodeHub	91	749496	8236,22	117,97	0
TitanTech	130	9542396	73403,05	15,25	0

After clustering the content creator data, there are 50 content creators grouped into cluster 0, 11 data grouped into cluster 1 and cluster 2 as many as 15 data. Clustering results are obtained without using dimensionality reduction techniques, which means that the data is used in its original form by maintaining two important features, namely



average views per video and engagement rate. After the Clustering process is complete, an evaluation is carried out by analyzing various metrics such as average views, average engagement, and Cluster size. The Composite score is calculated based on the average views and engagement, which provides an overall picture of the performance of each Cluster. Based on this score, the best Cluster for further analysis can be selected. The following is a visualization of the clustering results using the K-Means method.

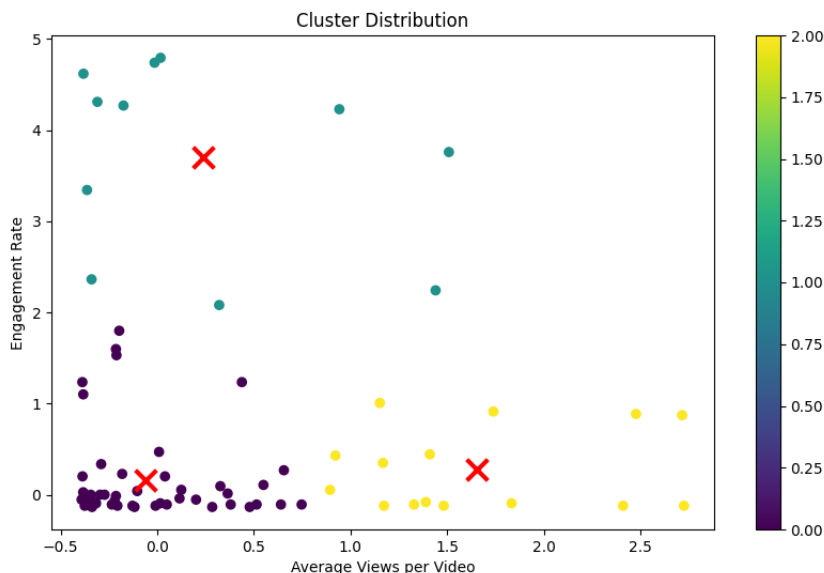


Figure 4. Visualization of Clustering Results with K-Means

Based on Figure 4, the Clustering results obtained from the K-Means analysis show three Clusters with different characteristics in terms of the number of views and engagement levels. Cluster 0, consisting of 50 Channels, has an average of 57,014.91 views per video, but its engagement level is quite low, at 0.22%. This illustrates that although the Channels in this Cluster have a large audience, interaction with that audience tends to be less active. On the other hand, Cluster 1, although it only contains 11 Channels, stands out with a higher average of 105,942.22 views per video, and a very good engagement level, reaching 2.86%. The Channels in this Cluster have a more engaged audience and interact with the content more intensely. Meanwhile, Cluster 2 consists of 15 Channels and has a very high average of 335,461.95 views per video, although its engagement level is slightly lower than Cluster 1, at 0.31%. This cluster shows a wide audience reach and consistency in gaining attention. Based on the composite score evaluation, Cluster 2 was selected as the best Cluster for further analysis using the MAUT method, due to its good balance between high number of views and adequate engagement level. This cluster shows great potential in terms of audience reach and content performance, making it an optimal choice for further development strategies. The following table 5 contains data grouped into cluster 2.

Table 5. Content Creator Data in Cluster 2 (Best Cluster)

channel_title	average_views_per_video	subscriber_count	upload_frequency_(videos/month)	total_likes	engagement_rate	Cluster
Coding with Lewis	292446,53	629000	131,38	67445	0,04	2
freeCodeCamp.org	458106,06	10300000	523,61	100298	0,01	2
Learn Coding	257332,59	2220000	165,27	13914	0,01	2
Sahil & Sarra	349194,34	778000	24,95	541911	0,78	2
Low Level	295789,64	774000	105,11	325580	0,43	2
Brackeys	468926,08	1850000	7,41	1555073	0,76	2
Alex Lee	216548,38	421000	3,19	112100	0,42	2
Kevin Stratvert	509215,45	3370000	299,56	68013	0,01	2
Ishan Sharma	212019,62	1640000	918,65	437631	0,14	2
Error Makes Clever	256421,42	899000	57,65	229997	0,36	2
CodeWithHarry	364455,84	7000000	634,38	276225	0,03	2
Codeiyapa	307460,84	46800	6,8	2632	0,01	2

channel_title	average_views_per_video	subscriber_count	upload_frequency_(videos/month)	total_likes	engagement_rate	Cluster
Jenny's Lectures CS IT	282517,31	1810000	158,25	45566	0,02	2
Scratch Team	253827,69	244000	2,13	237958	0,85	2
TechWorld with Nana	507667,42	1190000	2,93	462957	0,75	2

3.4 Ranking with MAUT Method

The final stage of analysis involves ranking the Channels in the best Cluster, namely in table 5 using the MAUT method. Only 5 criteria or attributes are used in this ranking because there are several criteria with similar assessments such as the number of views and average views, so it is sufficient to use one of the two criteria, as well as for the other criteria. The following Table 6 contains the criteria used in the assessment of the data sample

Table 6. Types of Criteria

Criteria Code	Description	Criteria Weight
C1	average_views_per_video	0,0125
C2	subscriber_count	0,05
C3	upload_frequency_(videos/month)	0,5875
C4	total_likes	0,2375
C5	engagement_rate	0,025

Based on Table 6, the five criteria are of the benefit type and it is also seen that criterion C3 is more prioritized than other criteria. After determining the level of importance of the criteria in table 6, the ranking process then begins by determining the evaluation matrix that considers several options in decision making until the overall evaluation process is carried out to obtain the best ranking of the content creator. More clearly, the following are the ranking stages using the MUAT method.

a. Evaluation matrix for considering multiple options when making decisions (Xij).

$$X_{ij} = \begin{vmatrix} 292446,53 & 629000 & 131,38 & 67445 & 0,04 \\ 458106,06 & 1030000 & 523,61 & 100298 & 0,01 \\ 257332,59 & 2220000 & 165,27 & 13914 & 0,01 \\ 349194,34 & 778000 & 24,95 & 541911 & 0,78 \\ 295789,64 & 774000 & 105,11 & 325580 & 0,43 \\ 468926,08 & 1850000 & 7,41 & 1555073 & 0,76 \\ 216548,38 & 421000 & 3,19 & 112100 & 0,42 \\ 509215,45 & 3370000 & 299,56 & 68013 & 0,01 \\ 212019,62 & 1640000 & 918,65 & 437631 & 0,14 \\ 256421,42 & 899000 & 57,65 & 229997 & 0,36 \\ 364455,84 & 7000000 & 634,38 & 276225 & 0,03 \\ 307460,84 & 46800 & 6,8 & 2632 & 0,01 \\ 282517,31 & 1810000 & 158,25 & 45566 & 0,02 \\ 253827,69 & 244000 & 2,13 & 237958 & 0,85 \\ 507667,42 & 1190000 & 2,93 & 462957 & 0,75 \end{vmatrix}$$

b. Calculate the current matrix normalization (r_{ij}^{*}).

$$r_{ij}^* = \frac{x_{ij} - \min(r_{ij})}{\max(r_{ij}) - \min(r_{ij})}$$

$$r_{11}^* = \frac{292446,53 - 212019,62}{509215,45 - 212019,62} = 0,2706$$

$$r_{21}^* = \frac{458106,06 - 212019,62}{509215,45 - 212019,62} = 0,8280$$

$$r_{31}^* = \frac{257332,59 - 212019,62}{509215,45 - 212019,62} = 0,1525$$

Calculate the normalization values C1 to C5 on the 4th to 15th data, after calculating the normalization search for all criteria, then calculate the marginal utility.

c. Calculate Marginal Utility (u_{ij}): $u_{ij} = \frac{e^{(r_{ij}^*)^2} - 1}{1,71}$

$$U_{11} = \frac{e^{(0,2706)^2} - 1}{1,71} = 0,0444$$



$$U_{21} = \frac{e^{(0,8280)^2} - 1}{1.71} = 0,5760$$

$$U_{31} = \frac{e^{(0,1525)^2} - 1}{1.71} = 0,0138$$

Calculate Marginal Utility on data 4 to data 15 by following the same steps as those done in the marginal utility search criteria for data 1 to data 3 above. After obtaining the overall marginal utility value, the next process is to calculate the overall evaluation value (uj) using equation formula 7, where the wj value used is in accordance with the criteria weight in table 6. The following is table 7 containing the results of the MAUT method calculation.

Table 7. Final Results (Ranking)

Channel Title	Utility (Uj)	Rank
Coding with Lewis	0,127	12
freeCodeCamp.org	0,543	2
Learn Coding	0,116	14
Sahil & Sarra	0,392	6
Low Level	0,247	9
Brackeys	0,624	1
Alex Lee	0,120	13
Kevin Stratvert	0,420	5
Ishan Sharma	0,254	8
Error Makes Clever	0,176	10
CodeWithHarry	0,424	4
Codeiyapa	0,097	15
Jenny's Lectures CS IT	0,138	11
Scratch Team	0,269	7
TechWorld with Nana	0,542	3

Table 7 presents the ranking data and utility values (Uj) of various programming-related YouTube channels. The "Brackeys" channel is ranked first with the highest utility value, which is 0.624, followed by "freeCodeCamp.org" in second place with a value of 0.543, and "TechWorld with Nana" in third place with a value of 0.542. The "CodeWithHarry" and "Kevin Stratvert" channels are ranked fourth and fifth with utility values of 0.424 and 0.420, respectively. Several other channels such as "Sahil & Sarra", "Scratch Team", "Ishan Sharma", "Low Level", and "Error Makes Clever" are also included in the list with varying utility values. The channel with the lowest utility value is "Codeiyapa" with a value of 0.097. The following figure 5 provides an overview of the popularity and quality of content from various programming YouTube channels.

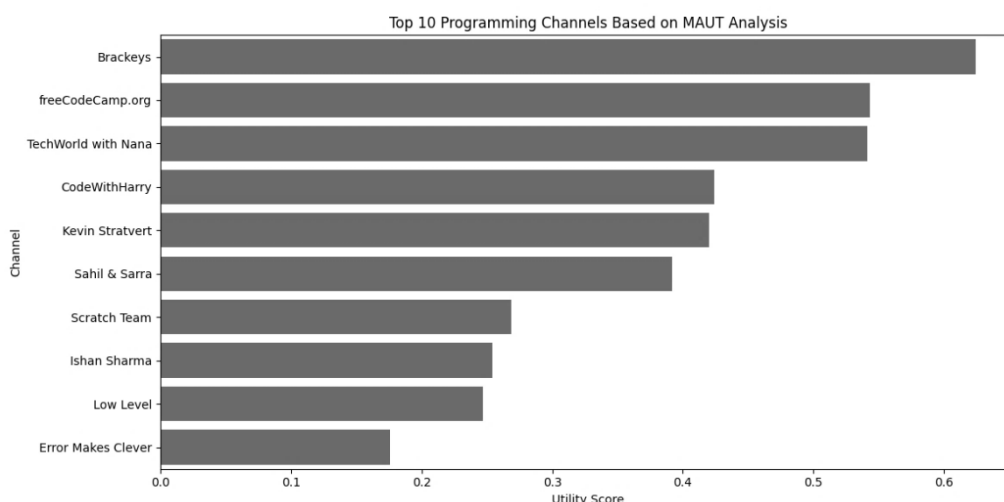


Figure 5. Visualization of MAUT ranking results

Figure 5 presents a list of the top 10 programming channels based on the MAUT analysis. Each channel is assigned a utility score that is visualized as a horizontal bar graph. The longer the bar, the higher the utility score of the channel, which means the better the channel performs in meeting the various criteria assessed in the MAUT analysis. Based on the figure, the channel "Brackeys" is ranked first with the highest utility score, followed by "freeCodeCamp.org" in second place, and so on.

4. CONCLUSION

This study successfully integrates the K-Means algorithm and the MAUT method to systematically identify the best programming content creators. The results of chapter 3 show that this approach is able to provide a data-based solution that supports the argument in chapter 1, namely the importance of an objective selection mechanism in selecting quality creators in the digital era. By utilizing the theories and methods discussed in chapter 2, this study proves the effectiveness of the combination of K-Means in data clustering and MAUT in multi-criteria evaluation. Through the analysis process, data from 100 content creators were collected using the YouTube API and processed to ensure their quality and consistency, after going through the preprocessing stage, the data used became 76. The K-Means algorithm was used to group creators based on characteristics such as average views per video and interaction level. Determining the optimal number of Clusters was done using the Elbow method, which produced relevant and measurable groupings. Furthermore, the MAUT method was applied to rank creators in each Cluster based on a composite score that reflects overall performance. The results showed that the Cluster with the highest composite score included the best performing creators, both in terms of engagement rate and average views. Therefore, the Channel "Brackeys" is ranked first with the highest utility value of 0.624, followed by "freeCodeCamp.org" in second place with a value of 0.543, and "TechWorld with Nana" in third place with a value of 0.542. This approach not only supports the theoretical argument about the effectiveness of MCDA but also provides empirical evidence that the combination of K-Means and MAUT is a powerful method for evaluating creator content performance.

REFERENCES

- [1] R. Zekaj, "The Impact of Online Learning Strategies on Students' Academic Performance: A Systematic Literature Review," *Int. J. Learn. Teach. Educ. Res.*, vol. 22, no. 2, pp. 148–164, 2023, doi: 10.26803/ijlter.22.2.9.
- [2] M. Bal, A. G. K. Aydemir, and M. Coşkun, "Exploring YouTube content creators' perspectives on generative AI in language learning: Insights through opinion mining and sentiment analysis," *PLoS One*, vol. 19, no. 9 September, pp. 1–30, 2024, doi: 10.1371/journal.pone.0308096.
- [3] S. K. Roy, "YouTube's influential factors for academic achievement: A two-stage approach," *Telemat. Informatics Reports*, vol. 10, no. February, p. 100060, 2023, doi: 10.1016/j.teler.2023.100060.
- [4] H. S. Kasman, "Perbandingan Efektifitas Algoritma K-Means Clustering -Topsis dan K-Medoids Clustering -Topsis dalam Menentukan Karyawan IT dengan Kinerja Terbaik A . Peneliti Terdahulu K-Means merupakan metode clustering non-hierarkis yang berusaha membagi data ke dalam," vol. 4, no. 2, 2024.
- [5] J. K. Hsieh, Y. H. Fang, and C. H. Liao, "The power of choice: Examining how selection mechanisms shape decision-making in online community engagement," *Decis. Support Syst.*, vol. 182, no. May, p. 114250, 2024, doi: 10.1016/j.dss.2024.114250.
- [6] M. Cinelli, P. Burgherr, M. Kadziński, and R. Słowiński, "Proper and improper uses of MCDA methods in energy systems analysis," *Decis. Support Syst.*, vol. 163, no. January, 2022, doi: 10.1016/j.dss.2022.113848.
- [7] M. Madanchian and H. Taherdoost, "A comprehensive guide to the TOPSIS method for multi-criteria decision making," *Sustain. Soc. Dev.*, vol. 1, no. 1, pp. 1–6, 2023, doi: 10.54517/ssd.v1i1.2220.
- [8] V. Yfantis, A. Wagner, and M. Ruskowski, "Federated K-means clustering via dual decomposition-based distributed optimization," *Franklin Open*, vol. 10, no. December 2023, p. 100204, 2025, doi: 10.1016/j.fraope.2024.100204.
- [9] S. Rahman, A. S. Alali, N. Baro, S. Ali, and P. Kakati, "A Novel TOPSIS Framework for Multi-Criteria Decision Making with Random Hypergraphs: Enhancing Decision Processes," *Symmetry (Basel)*, vol. 16, no. 12, 2024, doi: 10.3390/sym16121602.
- [10] N. A. Maori and E. Evanita, "Metode Elbow dalam Optimasi Jumlah Cluster pada K-Means Clustering," *Simetris J. Tek. Mesin, Elektro dan Ilmu Komput.*, vol. 14, no. 2, pp. 277–288, 2023, doi: 10.24127/simet.v14i2.9630.
- [11] D. Imantika *et al.*, "Penerapan metode k-means clustering dan analytical hierarchy process (ahp) untuk pengelompokan kinerja guru dan karyawan pada sma brawijaya smart school," *J. Pengemb. Teknol. Inf. dan Ilmu Komput. J-PTIHK*, vol. 3, no. 8, pp. 7382–7390, 2019.
- [12] S. B. H. Sakur, M. Silangen, and D. Tuwohingide, "Penerapan Algoritma K-Means Cluster dan Metode TOPSIS pada Pemilihan Mahasiswa kunjungan Industri," *Jutisi J. Ilm. Tek. Inform. dan Sist. Inf.*, vol. 11, no. 3, p. 851, 2022, doi: 10.35889/jutisi.v11i3.1045.
- [13] R. A. Siregar, "Seleksi Penyerang Utama Menggunakan K-Means Clustering Dan Sistem Pendukung Keputusan Metode Topsis," *Technomedia J.*, vol. 2, no. 1, pp. 37–48, 2017, doi: 10.33050/tmj.v2i1.314.
- [14] R. M. F. Lubis, J.-P. Huang, P.-C. Wang, K. Khoifin, M. Sigiuro, and J. Panjaitan, "Data Clustering Mining Applying the K-Means Algorithm, Cervical Cancer Behavior Risk," *J. MEDIA Inform. BUDIDARMA*, vol. 7, no. 2, pp. 819–827, 2023.
- [15] A. Pangestu and T. Ridwan, "Penerapan Data Mining Menggunakan Algoritma K-Means Pengelompokan Pelanggan Berdasarkan Kubikasi Air Terjual Menggunakan Weka," *JUST IT J. Sist. Informasi, Teknol. Inf. dan Komput.*, vol. 12, no. 3, pp. 67–71, 2022.
- [16] M. A. Abdullah and R. T. Aldisa, "Penerapan Metode MAUT Dalam Rekrutmen Kepala Gudang di Perusahaan Dengan Pembobotan ROC," vol. 4, no. 2, pp. 1073–1079, 2023, doi: 10.30865/klik.v4i2.1323.
- [17] A. Triayudi, J. D. Rajagukguk, and Mesran, "Implementasi Metode MAUT Dalam Menentukan Prioritas Produk Unggulan Daerah Dengan Menerapkan Pembobotan ROC," *J. Comput. Syst. Informatics*, vol. 3, no. 4, pp. 452–460, 2022, doi: 10.47065/josyc.v3i4.2216.
- [18] E. K. Suni and S. A. Sutresno, "Rekomendasi Content Creator Terbaik sebagai Pendukung Keputusan Penilaian pada Agensi Menggunakan Metode TOPSIS," vol. 5, no. 2, pp. 414–422, 2023, doi: 10.47065/bits.v5i2.4082.
- [19] A. S. L. T. H. Hafizah, "Data Mining Estimasi Biaya Produksi Ikan Kembung Rebus Dengan Regresi Linier Berganda," *J. Sist. Inf. Triguna Dharma (JURSI TGD)*, no. Vol 1, No 6 (2022): EDISI NOVEMBER 2022, pp. 888–897, 2022.
- [20] Y. L. Nainel, E. Bulololo, and I. Lubis, "Penerapan Data Mining Untuk Estimasi Penjualan Obat Berdasarkan Pengaruh Brand



- Image Dengan Algoritma Expectation Maximization (Studi Kasus: PT. Pyridam Farma Tbk),” *JURIKOM (Jurnal Ris. Komputer)*, vol. 7, no. 2, p. 214, 2020, doi: 10.30865/jurikom.v7i2.2097.
- [21] M. Azhari, Z. Situmorang, and R. Rosnelly, “Perbandingan Akurasi, Recall, dan Presisi Klasifikasi pada Algoritma C4.5, Random Forest, SVM dan Naive Bayes,” *J. Media Inform. Budidarma*, vol. 5, no. 2, p. 640, 2021, doi: 10.30865/mib.v5i2.2937.
- [22] H. Maulidiya and A. Jananto, “Asosiasi Data Mining Menggunakan Algoritma Apriori dan FP-Growth sebagai Dasar Pertimbangan Penentuan Paket Sembako,” *Proceeding SENDIU 2020*, vol. 6, pp. 36–42, 2020.
- [23] J. Afriany, K. Tampubolon, and R. Fadillah, “Penerapan Metode TOPSIS Penentuan Pemberian Mikro Faedah Bank Syariah Indonesia (BSI),” *TIN Terap. Inform. Nusant.*, vol. 2, no. 3, pp. 129–137, 2021.
- [24] R. Fadillah, A. N. Purnama, W. P. Mahardika, and M. Mesran, “Sistem Pendukung Keputusan Penentuan Siaran Edukasi di Televisi Menggunakan Metode WASPAS (Weight Aggregated Sum Product Assesment),” *JIKTEKS J. Ilmu Komput. dan Teknol. Inf.*, vol. 1, no. 3, pp. 8–16, 2023.
- [25] M. Yanto, “Sistem Penunjang Keputusan Dengan Menggunakan Metode Ahp Dalam Seleksi Produk,” *J. Teknol. Dan Sist. Inf. Bisnis*, vol. 3, no. 1, pp. 167–174, 2021.
- [26] D. O. Wibowo and A. Thyo Priandika, “Sistem Pendukung Keputusan Pemilihan Gedung Pernikahan Pada Wilayah Bandar Lampung Menggunakan Metode Topsis,” *J. Inform. dan Rekayasa Perangkat Lunak*, vol. 2, no. 1, p. page-page. xx~xx, 2021.
- [27] D. P. Tarigan, A. Wantoro, and S. Setiawansyah, “Sistem Pendukung Keputusan Pemberian Kredit Mobil Dengan Fuzzy Tsukamoto (Studi Kasus: Pt Clipan Finance),” *TELEFORTECH J. Telemat. Inf. Technol.*, vol. 1, no. 1, pp. 32–37, 2020.
- [28] I. Alfansyah, J. Sibagariang, R. Fadillah, and D. Assarani, “Sistem Pendukung Keputusan Seleksi Dosen Non Komputer Terbaik Menerapkan Metode SAW,” *J. Decis. Support Syst. Res.*, vol. 1, no. 1, pp. 30–36, 2023.
- [29] M. Mesran, A. Triayudi, D. Nofrisa, and R. Fadillah, “Penerapan Metode EXPROM II Dalam Menentukan Tempat Wisata Pantai Terbaik,” *J. Sist. Komput. dan Inform.*, vol. 5, no. 2, pp. 337–346, 2023.
- [30] A. A. Nasution, R. T. Aldisa, M. Mesran, and R. Fadillah, “Penerapan Metode Multi Objective Optimization on The Basis of Ratio Analysis (MOORA) dalam Penentuan Pembimbing Skripsi Terbaik,” *J. Inf. Syst. Res.*, vol. 5, no. 2, pp. 614–620, 2024.
- [31] A. Ardinsah, M. Mesran, and A. Triayudi, “Sistem Pendukung Keputusan Menentukan Aplikasi Chat Terbaik Menggunakan Metode OCRA Dengan Pembobotan ROC,” *J. Comput. Syst. Informatics*, vol. 4, no. 4, pp. 891–898, 2023.
- [32] M. Y. Nurzaman, “Implementasi K-Means Clustering Dalam Pengelompokan Banyaknya Jumlah Petani Berdasarkan Kecamatan Di Provinsi Jawa Barat,” *JATISI (Jurnal Tek. Inform. dan Sist. Informasi)*, vol. 10, no. 3, pp. 131–144, 2023.
- [33] M. Masjunedi, N. Suarna, and Y. A. Wijaya, “Analisa Penerapan Metode Clustering K-Means Untuk Pengelompokan Data Transaksi Konsumen,” *JATI (Jurnal Mhs. Tek. Inform.)*, vol. 7, no. 2, pp. 1322–1328, 2023.
- [34] P. Marpaung, I. Pebrian, and W. Putri, “Penerapan Data Mining Untuk Pengelompokan Kepadatan Penduduk Kabupaten Deli Serdang Menggunakan Algoritma K-Means,” *J. Ilmu Komput. dan Sist. Inf.*, vol. 6, no. 2, pp. 64–70, 2023.
- [35] M. A. Rofiq, A. Qoiriah, S. Kom, and M. Kom, “Pengelompokan Kategori Buku Berdasarkan Judul Menggunakan Algoritma Agglomerative Hierarchical Clustering Dan K-Medoids,” *J. Informatics Comput. Sci.*, vol. 2, no. 03, pp. 220–227, 2021.
- [36] B. Harli Trimulya Suandi As and L. Zahrotun, “PENERAPAN DATA MINING DALAM MENGELOMPOKKAN DATA RIWAYAT AKADEMIK SEBELUM KULIAH DAN DATA KELULUSAN MAHASISWA MENGGUNAKAN METODE AGGLOMERATIVE HIERARCHICAL CLUSTERING (Implementation Of Data Mining In Grouping Academic History Data Before Students And Stud,” *J. Teknol. Informasi, Komput. dan Apl.*, vol. 3, no. 1, pp. 62–71, 2021.
- [37] H. N. Amaliya and Y. Yamasari, “Clustering Bidang Keilmuan Menggunakan Kombinasi Metode Topsis dan Algoritma K Means,” *JINACS J. Informatics adn Comput. Sci.*, vol. 04, pp. 405–413, 2023.
- [38] D. Aldo, “Sistem Pendukung Keputusan Penilaian Kinerja Dosen Dengan Menggunakan Metode Multi Attribute Utility Theory (Maut),” *Jursima*, vol. 7, no. 2, p. 76, 2019, doi: 10.47024/js.v7i2.180.
- [39] N. Rivaldy, D. Irmayanti, and M. Defriani, “Perancangan SPK Kelayakan Pinjaman Koperasi Karyawan Perum Jasa Tirta Karya Bhakti Raharja dengan Metode MAUT,” *J-SAKTI (Jurnal Sains Komput. dan Inform.)*, vol. 7, no. 1, pp. 172–181, 2023.
- [40] J. Faran and R. T. Aldisa, “Implementasi Metode MAUT dengan Menerapkan Pembobotan ROC Dalam Pemilihan Ketua Himpunan Mahasiswa,” *J. MEDIA Inform. BUDIDARMA*, vol. 7, no. 3, pp. 1315–1322, 2023.
- [41] M. Mesran, A. Harahap, and F. Nugroho, “Sistem Pendukung Keputusan Pemilihan Aplikasi Chat Terbaik Dalam Mendukung Pembelajaran Daring di Masa Pandemi Covid Menggunakan Metode Multi Attribute Utility Theory,” *J. Sist. Komput. dan Inform.*, vol. 4, no. 3, pp. 474–484, 2023.
- [42] G. T. Mareti and A. T. Ayunda, “Komparasi Metode Maut dan Moora dalam Pemilihan Sunscreen untuk Kulit Menggunakan Pembobotan ROC,” *Build. Informatics, Technol. Sci.*, vol. 5, no. 2, pp. 553–564, 2023.
- [43] I. Oktaria, “Kombinasi Metode Multi-Attribute Utility Theory (MAUT) dan Rank Order Centroid (ROC) dalam Pemilihan Kegiatan Ekstrakurikuler,” *J. Ilm. Inform. Dan Ilmu Komput.*, vol. 2, no. 1, pp. 1–11, 2023.
- [44] M. Z. Lubis, R. Fadillah, and R. M. F. Lubis, “Decision Support System for Determining New Branch Locations Applying the Multi Attribute Utility Theory (MAUT) Method,” *Int. J. Informatics Data Sci.*, vol. 1, no. 1, pp. 36–45, 2023.
- [45] A. Azhari, P. Pahrudin, and Y. Yunita, “Penerapan Metode Topsis Pada Sistem Pendukung Keputusan Kelayakan Penerima Dana Bantuan Operasional Sekolah,” *Build. Informatics, Technol. Sci.*, vol. 4, no. 3, pp. 1688–1696, 2022, doi: 10.47065/bits.v4i3.2290.
- [46] H.-S. Shih and D. L. Olson, “TOPSIS and its Extensions: A Distance-Based MCDM Approach,” vol. 447, p. 230, 2022.
- [47] S. Chakraborty, “TOPSIS and Modified TOPSIS: A comparative analysis,” *Decis. Anal. J.*, vol. 2, no. December 2021, p. 100021, 2022, doi: 10.1016/j.dajour.2021.100021.
- [48] P. Pristiwanto, H. Sunandar, and B. Nadeak, “Penerapan Metode MAUT Terhadap Perkembangan Metaverse Untuk Media Pembelajaran Daring Dengan Pembobotan ROC,” *KOMIK (Konferensi Nas. Teknol. Inf. dan Komputer)*, vol. 6, no. 1, pp. 100–107, 2023.
- [49] R. W. Dari, S. Sapriadi, N. A. Rahmi, and P. A. W. Purnama, “Metode Multi Attribute Utility Theory (MAUT) untuk Sistem Pendukung Keputusan Pemilihan Mobil Bekas,” *J. KomtekInfo*, pp. 73–79, 2023.
- [50] K. D. Putri, A. P. Juledi, and I. R. Munthe, “Perancangan Aplikasi SPK Penerimaan Tenaga Kerja Honorer dengan Menggunakan Metode MAUT,” *MEANS (Media Inf. Anal. dan Sist.)*, pp. 48–56, 2023.
- [51] B. Okmarizal and S. Defit, “Implementasi Metode AHP Dan Maut untuk Rekomendasi Produk Tupperware Terlaris,” *J.*



KomtekInfo, pp. 109–115, 2023.

- [52] E. Y. Dewasmita and H. Hendry, “Perbandingan Metode SAW, MAUT, ORESTE, TOPSIS dalam Pendukung Keputusan Pembangunan Supermarket di Kabupaten Pati,” *J-SAKTI (Jurnal Sains Komput. dan Inform.*, vol. 7, no. 2, pp. 555–569, 2023.
- [53] W. K. Murti, A. Triayudi, and M. Mesran, “Penentuan Mahasiswa Berprestasi dengan Menerapkan Metode Multi Attribute Utility Theory (MAUT),” *J. Sist. Komput. dan Inform.*, vol. 5, no. 1, pp. 122–130, 2023.