Prediction of Scholarship Recipients Using Hybrid Data Mining Method with Combination of K-Means and C4.5 Algorithms

Mardison[a,1*], Sarjon Defit[b,2] Shaza Alturky[c]

[a,b] *Universitas Putra Indonesia YPTK Padang, Jl. Raya Lubuk Begalung, Padang, 25221, Indonesia*
[c]*Aleppo University, Syrian Arab Republic*
[1] *mardison@upiyptk.ac.id*; [2] sarjond@co.uk*
* corresponding author

ARTICLE INFO

ABSTRACT

Obtaining a scholarship is the desire of every student or student who studies, especially those who come from poor families. The scholarship can lighten the burden on parents who pay for these students and can streamline the lecture process. However, students do not know exactly what they have to do to get the scholarship. Aside from that, students naturally want to know what causes and conditions have the greatest impact on achievement. The objective of this research is how to predict which number of students among them are predicted to get a scholarship at the opening of the scholarship acceptance using the K-Means and C4.5 methods. Apart from that, the aim of this research is to discover how the K-Means algorithm conducts data clustering (clustering) of student data to determine if they will succeed or not, as well as how the C4.5 algorithm makes predictions against students who have been clustered together. The Rapid Miner program version 9.7.002 was used to process the data in this report. The results of this study were that out of 100 students, 32 students were not scholarship recipients and 68 students were scholarship recipients. Another result of this research is that out of 100 students it is predicted that 9 (9%) will receive scholarships and 91 (91%) will not receive scholarships.

I. Introduction

Education is the right of all children of the nation in every country in the world, including the Indonesian nation. There are various levels of education in Indonesia ranging from the lowest to the highest. The lowest level of education is kindergarten (TK) followed by an elementary school (SD) than junior high school (SMP) and senior high school (SMA) while the highest education level is tertiary education (PT). Every child who wants to continue to a higher level of education must first complete the previous level of education. For example, a child who will continue his education at the tertiary level must first complete his education at the high school level [1]–[3].

In getting an education, a student certainly needs everything that can support the educational process such as equipment, transportation, communication & internet, meals, expenses, housing, books, and so on. One of the most important supporting elements is the cost of education. The cost of education is several costs (money) paid by students to the educational institutions that they participate in. Besides, costs are also needed to buy equipment, transportation, food, shelter, and others. Of the many things that need to be funded in pursuing education, the most important is the tuition fees paid for educational institutions [4]–[6].

The amount of tuition fees paid by students depends on the educational institution that is being followed. Some have a large fee and some are small. The amount of cost does not determine the quality and also does not depend on the status of educational institutions whether public or private. Among all levels of education, the highest average cost of education in higher education, namely tertiary education [7]–[9]. Students who attend tertiary education are given the name of the student. The

student of course also pays several fees (money) to the college where the student is educated. For students who come from well-to-do families or whose parents have a large income, it will be easy to provide the cost of education, while for students who come from underprivileged families or their parents have a small income, it will be difficult to provide the required tuition fees [10], [11].

Many ways can be done for students who have difficulty in making education costs, one of which is scholarships. Many governments, private or private institutions provide scholarships but the number is limited. Not all available scholarships can be enjoyed by all students. Therefore, educational institutions always make a selection or selection among several students who want to get the scholarships that are offered. The method of dredging, storing, or extracting valuable information from a broad data set is known as data mining. The data mining process often uses statistical, mathematical methods to utilize artificial intelligence technology. Knowledge discovery mining in databases (KDD), knowledge extraction, data/pattern analysis, data archeology, data dredging, information harvesting, business intelligence, and other terms are used to describe data mining. Various algorithms are applied by educational institutions in selecting or selecting students who will receive scholarships. Some do it manually, some do it computerized. In computerization, there are also many algorithms used. Two algorithms that are mostly used by researchers are the K-Means algorithm and the C4.5 algorithm. The combination of the two algorithms will produce a more accurate prediction calculation than using only one algorithm [12], [13].

The K-Means algorithm is one of the algorithms used in data mining techniques to group data, which is also known as clustering. Although the C4.5 algorithm is one of the algorithms used in data mining techniques to forecast or predict data, it is not the only one. After the data from the clustering results of the K-Means algorithm are obtained, it is then used to forecast data using the C4.5 algorithm [14]–[16]. The use of two types of algorithms in data mining techniques for one data analysis is known as hybrid data mining. In this research, hybrid data mining was carried out by combining the K-Means algorithm with the C4.5 Algorithm.

The objective of this research is how to predict or predict students who will receive scholarships using the K-Means method and the C4.5 algorithm for computer science students. The aim of this research is to learn how the K-Means algorithm clusters data of potential scholarship recipients and how the C4.5 algorithm forecasts among the clustered data [17], [18]. One of the quite large and prominent private universities in Indonesia, especially in the province of West Sumatra, is the Universitas Putra Indonesia YPTK Padang. This university has a faculty that studies computer science called the faculty of computer science (FILKOM). In this study, students who obtained education at the Universitas Putra Indonesia YPTK Padang's computer science faculty in the year 2019 were used as source data.

Many scholarships are available at Universitas Putra Indonesia YPTK Padang's Faculty of Computer Science, including 1st and 2nd-grade winners, Bidik Misi scholarships, BBM scholarships, PPA scholarships, and scholarships from various banks in Indonesia. All available scholarships are distributed evenly to students who are entitled to receive them. For this reason, students who are eligible to receive scholarships are selected. So that students in the future can find out what are the determining factors that determine a scholarship will receive a scholarship, predictive action is needed using data mining techniques, namely a combination of the K-Means and C4.5 algorithms [19].

Previous research that has been conducted and is relevant to this research is by Nurul Rohmawati W et al in 2015 who came from the Singaperbangsa Karawang University, West Java [20]. In this study, the results obtained are the k-means algorithm clustering based on the results of clustering from each different dataset format (partial codification, whole codification, and original data) by measuring the accuracy level of clustering, namely calculating the purity measure value of the results. cluster. The greater the purity value (closer to 1) the better the quality of the clusters produced by an algorithm. The weakness in this study is that the purity value in the partially codified data dataset for the k-means algorithm cluster results is 61.11%.

The second research that was carried out was by Iin Parlina in 2018 from Amik Tunas Bangsa Pematangsiantar, North Sumatra [21] which produced clusters by applying several criteria from the SDP program list using the K-Means algorithm. SDP by conducting an advanced assessment center are those who pass and the classification results of the SDP program that almost pass, the classification results of the SDP program data list that do not pass must improve their disciplinary data for 1 year.

The drawback of this study is that the results of the grouping obtained three groups, namely the group passed, almost passed, and did not just pass.

II. Methodology

A. Research Framework
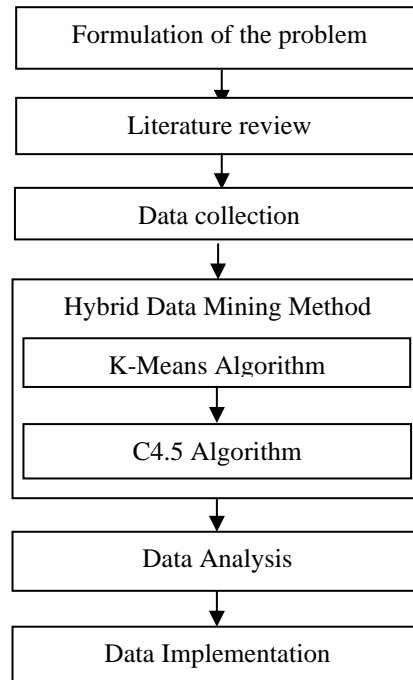

Fig. 1. Research Framework

a. Research Framework. In conducting research, of course, you must follow the rules or a structured and systematic research framework. The research framework can be seen in fig 1.

b. Description of the Research Framework.

- Problem formulation is the process of formulating or determining what are the main and important problems that occur in the field for which a solution must be found in this research.

- Literature study is the process of finding or analyzing relevant previous research and library books with predetermined problem-solving.

- Data collection is the process of collecting data in the field that will be used in problem-solving. In this study, the data came from the Faculty of Computer Science, University of Putra Indonesia YPTK Padang which was obtained from the Vice Dean III (WD III). The number of lines of testing data to be tested that has been collected is 100 rows of data, which are 100 students. In grouping (clustering) the data in this study the authors use the K-Means algorithm from the data of 100 people. Then in making predictions using the C4.5 method also uses the same data that has been grouped. The data collected consists of 6 data attributes as in table 1 below:

Table 1. Data Attributes

No	Data Attributes	Data Types
1	BP Number	Text
2	Name	Text
3	Father's income	Integer
4	Mother's income	Integer
5	GPA	Integer
6	Scholarship Receipt (YES/NO)	Binominal

- Data mining is a system of processes for extracting additional value from a set of data and information that cannot be calculated manually. The term "mining" refers to the process of extracting a few useful products from a large number of simple materials. As a result, data mining has a long history in fields like artificial intelligence, machine learning, statistics, and database management. The process of applying this approach to data in order to discover hidden patterns is known as data mining. Data mining may also be described as the process of extracting patterns from large amounts of data. For translating this data into information, data mining is becoming increasingly relevant. It's commonly used in advertisement, surveillance, fraud detection, and scientific discovery, among other things. There are various algorithms used for data mining including the K-Means algorithm, the C4.5 algorithm, the Support Vector Machine (SVM) algorithm, the a priori algorithm, the expectation-maximization algorithm, and others. In this study, two types of algorithms were combined, namely the K-Means algorithm and the C4.5 algorithm.

- Data Processing Using the K-Means Method

  The data that the authors have collected in the field is processed using the K-Means method. The following are the steps for the K-Means method, namely:

  1) Determining the number of data clusters to be created, the number of data clusters is called the value k

  2) Determine the mean (centroid value) randomly for each predetermined group.

  3) Determine the nearest cluster center on each data row with the centroid value, to determine this value using the formula [22]:

  $$d_{Euclidean}(x, y) = \sqrt{\sum(x_i - y_i)^2} \tag{1}$$

  Information: $d_{Euclidean}(x, y) =$ distance value on each data row with centroid value, $centroid, x_i = x_1, x_2, x_3, \ldots\ldots etc, y_i = y_1, y_2, y_3, \ldots\ldots etc$

  4) Determine the closest group (cluster) for each row of data by comparing the closest distance value obtained in the previous step and updating the group's center value using the formula:

  $$Cluster\ Center = \sum \frac{a_i}{n} \tag{2}$$

  Information: Cluster Center= The cluster center value, a_i= Value on each cluster, n= number of clusters.

  5) Repeating steps 3 to step 5 until there is no transfer of data for each row of data from one group to another.

- Data Processing Using C4.5 Method

  1) The results of data processing using the K-Means method are then processed using the C4.5 method. The steps for the C4.5 method are as follows:

  2) Determine the data attribute that will be used as the root node or prediction in the decision tree and calculate the number of YES and NO values for each row of data.

  3) Determine the branch from the root (root) for each value after determining the root of the decision tree by calculating the Gain value. The gain calculation formula is as follows [23]:

  $$Gain(S, A) = Entropy(S) - \sum_{i=1}^{n} \frac{|S_i|}{|S|} * Entropy(S_i) \tag{3}$$

  Information: $Gain(S, A) =$ total gain value with attributes, $Entropy(S) =$ total entropy value, $Entropy(S_i) =$ Entropy value for each attribute, $n =$ number of clusters.

  The formula for calculating the entropy value is:

  $$Entropy(S) = \sum_{i=1}^{n} - pi * log_2 pi \tag{4}$$

Information: $Entropy(S)$ = total entropy value, pi = proportion of S_i to S

4) Divide cases for each existing branch
5) Repeating steps 2 to step 3 for each branch, before the branch's cases all have the same class.

After the data is processed and generated using the K-Means method and the C4.5 method, then it is analyzed and conclusions are drawn. How much data is the result of grouping and who is included in the group previously processed using the K-Means method then the grouping data is analyzed so that it is known which students are predicted to get scholarships and who are predicted not to get scholarships using the C4.5 method among all the data that has been processed. Data processing using Rapid Miner software version 9.7.002.

Data analysis results and conclusions are implemented in the field which is very useful and helps institutions, especially educational institutions, in predicting their students will get scholarships and not get scholarships.

III. Result and Discussion

A. Results of Clustering Data Processing Using the K-Means Method

The initial data that has been collected consisting of 6 attributes and 100 data records can be seen in Table 2 below:

The result of the research can be presented in the form of tables, graphs or figures. They can be compiled with written text to build a discussion of the findings, that is about the new, the modification or the established theory.

Table 2. Preliminary data

No	BP Number	Name	Number of Dependent Parents (Person)	Father's Income (Rp. / Month)	Mother's Income (Rp. / Month)	Grade Point Average (GPA)
1	17101152600027	Williem Kusnedi T	4	2.000.000	0	3,33
2	18101152600007	Wira Wahyuni	4	800.000	0	2,86
3	18101152600020	Raesa Islamiati	2	1.000.000	0	3,87
4	18101152600022	Yogi Hensyah	3	1.500.000	0	3,42
5	18101152600029	Wirda Jihadita	4	1.500.000	0	3,74
….	…	…	…	…	…	…
100	17101152610444	Admel Brina	5	2.000.000	500.000	3,25

After processing data on testing data which is the initial data after being collected as in table 2 above with the K-Means method using the Rapid Miner application, data clustering is obtained. Data clustering was carried out on all data attributes, namely attributes of father's income (IDR / month), mother's income (IDR / month), cumulative grade point average (GPA). The attributes of the father's income and mother's income are combined into one attribute, namely parental income (PO) by adding the father's and mother's income then dividing it in half to get the average with the aim of simplifying calculations. After combining the attributes of the father's income and the mother's income into the parent's income, the data is converted into values 1, 2 3, and 4 using table 3 below:

Table 3. Initial Data Conversion Guidelines

Category	PO Qualification	Codification
Category 4	≤ 112.000	4
Category 3	113.000 – 512.000	3
Category 2	513.000 – 912.000	2
Category 1	≥ 913.000	1

Below are the results of data conversion based on table 3 above as in table 4 below:

Table 4. Results of Data Conversion

No	BP No	PO	IPK
1	17101152600027	3	3.33
2	18101152600007	3	3.27
3	18101152600020	1	3.35
4	18101152600022	3	3.4
5	18101152600029	3	3.35
….	….	….	….
100	17101152610444	3	3.1

After obtaining the converted data as in table 3 above, the next step is to cluster the data using the K-Means algorithm. The steps for the K-Means algorithm are as follows:

Step 1. Determine the number of data clusters to be created; this is referred to as the value k. The number of data groups determined in this study derived from the data in table 4 is as many as 2 data clusters, namely cluster 0 and cluster 1.

Step 2. Determine the centroid value randomly for each predetermined group. Below is the centroid value for each cluster is shown in table 5 below:

Table 5. Clustering Results: Middle Value (Centroid Value)

No	Attribute	Cluster 0	Cluster 1
1	Parent's Income	1.594	3.147
2	Cumulative Achievement Index (GPA)	3.361	3.322

Step 3. Determine the closest cluster center on each row of data with the centroid value, to determine this value using the formula:

$$d_{Euclidean}(x,y) = \sqrt{\sum (x_i - y_i)^2}$$

$$d_{Euclidean}(x_1, y_1) = \sqrt{(3.33 - 3.1)^2 + (3.27 - 3.1)^2 + \cdots + (3.1 - 3.1)^2} = \sqrt{4.0529} = 2.013$$

….

$$d_{100}(x_{100}, y_{100}) = \sqrt{(3-1)^2 + (3-1)^2 + \cdots + (3-1)^2} = \sqrt{4.0289} = 2.007$$

Step 4. Determine the closest cluster for each row of data by comparing the closest distance values that have been obtained in the previous process then updating the center value of the group using the formula:

$$Cluster\ Center = \sum \frac{a_i}{n} = \frac{171.559}{100} = 1.72$$

Step 5. Repeating steps 3 to step 5 until there is no transfer of data for each row of data from one group to another. The following table 6 shows the results of grouping data using the K-Means algorithm:

Table 6. Data Clustering Results

No	Cluster	Item	Named
1	Cluster 0	32 Items	Not Receipt Scolarship
2	Cluster 1	68 Items	Receipt Scolarship
3	Total Item	100 Items	-

In order to produce data as shown in table 6 above requires a data processing block design using the Rapid Miner application. Fig. 2 illustrates the block architecture:
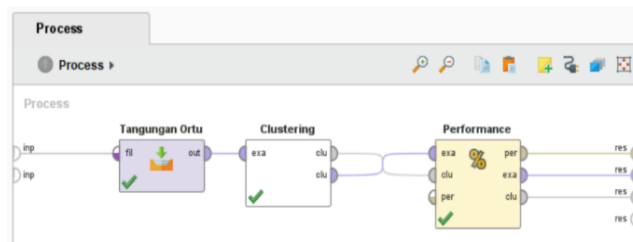

Fig. 2. Block Architecture of K-Means Algorithm in Rapid Miner

So that the data can be seen visually in the form of clustering results so that we can find out what kind of data grouping has been generated, a visualization view is needed as in Fig. 3 below:

Fig. 3. Visualization View

B. Prediction Data Processing Results Using The C4.5 Method

The data that the researchers collected were 100 rows of data or 100 students who received scholarships and did not receive scholarships in the previous year. Data that has been grouped into clusters are given names that describe the type of cluster. The following are the steps for the C4.5 algorithm:

Step 1. Determine the data attribute that will be used as the root node or prediction in the decision tree and calculate the number of YES and NO values for each row of data. Below, in table 7 the data that have been named YES and NO and the addition of the attributes of receiving a scholarship.

Table 7. Naming Data According to Clusters

No	PO	Clustering	IPK	Clustering	Parents Status	Receipt Scholarship
1	3	Low	3.33	Low	Present	Yes
2	3	Low	3.27	Low	Present	Yes
3	1	Hight	3.35	Low	Present	No
4	3	Low	3.4	Low	Present	Yes
5	3	Low	3.35	Low	Present	Yes
….	….	….	….	….	….	….
100	3	Low	3.1	Low	Present	Yes

Step 2. Determine the branch from the root (root) for each value after determining the root of the decision tree by calculating the Gain value. The gain calculation formula is as follows:

$$Gain(S, A) = Entropy(S) - \sum_{i=1}^{n} \frac{|S_i|}{|S|} * Entropy(S_i)$$

$Gain(PO, A) = 0.905 - (0.77 * 0.806) = 0.2844$
$Gain(IPK, A) = 0.905 - (0.94 * 0.925) = 0.0355$
$Gain(Status, A) = 0.905 - (0.78 * 0.977) = 0.1429$

Step 3. Divide the cases for each existing branch.

Step 4. For each branch, repeat steps 2–3 until all cases on the branch have the same class.

Because the calculation of the C4.5 algorithm is repeated with quite several iterations, it requires data mining software to speed up and simplify calculations. The application used is the Rapid Miner application version 9.7.002. Table 8 shows the results of prediction calculations using the C4.5 algorithm.

Table 8. Prediction Result Data

No		True (YES)	True (NO)	Class Precision
1	pred. YES	5	4	55.56%
2	pred. NO	0	1	100.00%
3	class recall	100.00%	20.00%	
4	pred. YES	5	4	55.56%

Accuracy: 60.00%

The prediction result data above can be presented in graphical form (plot view) so that you can easily see the prediction results as shown in Fig. 4 below:

Fig. 4. Plot View of Predicted Data

Processing data on the Rapid Miner application requires a block design that describes the sequence of data processing. To produce data as seen in table 7, table 8, and fig. 4, the data processing block design using the Rapid Miner application can be seen in fig. 5 below:
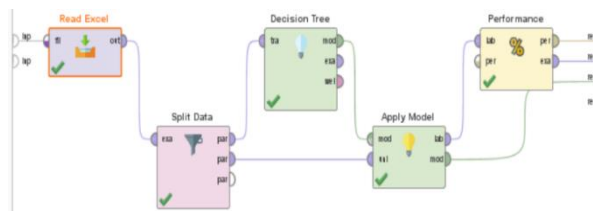

Fig. 5. Block Design C4.5 Algorithm in Rapid Miner

The outcomes of the Table 8 predictions may take the form of statistics or data conclusions. The data is presented in the form of a summary decision tree text view below.

Tree

```
PO = LOW
|   STATUS = PRESENT
|   |   GPA = LOW: YES {YES=2, NO=0}
|   |   GPA = HIGHT: YES {YES=41, NO=15}
|   STATUS = NO: YES {YES=11, NO=0}
```

Mardison et.al (Prediction of Scholarship Recipients Using Hybrid Data Mining Method with Combination of K-Means and C4.5 Algorithms)

```
PO = VERY LOW: YES {YES=9, NO=0}
PO = HIGHT: NO {YES=0, NO=12}
```

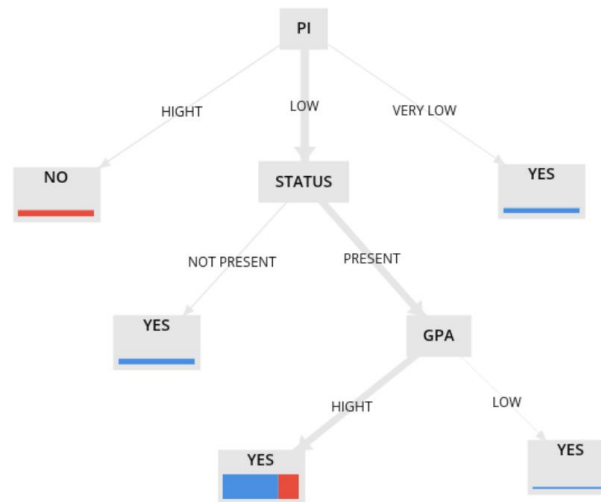The shape of the decision tree from the prediction results can be seen in fig. 6 below:

Fig. 6. Decision Tree Prediction Method

C. Discussion

Based on the results of the K-Means method clustering data processing using the Rapid Miner application according to the data in table 2 it can be seen that from a total of 100 student data it is made into two clusters, namely cluster 0 and cluster 1. Cluster 0 is defined as a scholarship recipient and cluster 1 is not a scholarship recipient. It can be seen from the output that 32 students did not receive scholarships and 68 students did receive scholarships.

Figure 2 is a design view, which is a block design view of clustering data processing for each attribute. Each data attribute is clustered using the K-Means method in the Rapid Miner application. From the far left is the read excel block which functions as a block for inputting excel data which will be processed then the clustering block which functions to perform data grouping (clustering) which is the performance block that functions to see the data processing capability.

Figure 3 is a visualization view, which is a visual display of data that has been grouped (clustering) using the K-Means method which is visualized in the form of colored dots. The blue dots represent the cluster 1 data group and the green dots are the 0 cluster data groups. The number of points on the graph illustrates the amount of data or students included in the data collected.

Based on the results of predictive data processing using the C4.5 method using the Rapid Miner application, according to the data in table 7, it can be seen that from a total of 100 student data processed, 9 people received a YES score and 91 received a NO.

Figure 4 is a plot view, which shows a point in the form of a scholarship recipient's prediction. The blue plot is the student who is predicted to receive a scholarship and the green plot is the student who is predicted NOT to receive the scholarship. The number of points is the number of scholarship recipients. In the plot, it can be seen that the predicted YES scholarship recipient is 9 people and the predicted NOT the scholarship recipient is 1 person.

In Figure 5 is the C4.5 Method block design in the RapidMiner application. In the block design, it can be seen that the RapidMiner application requires data input in the first block then the data that has been inputted is split into 2 blocks, namely the decision tree block and the application model block which are then combined back into the final block, namely the performance block to determine the accuracy of the data.

In Figure 6, the decision tree can be seen that in predicting the process flow is by training the parents first (PO), namely low, very low, high. If the parents 'income is low, then look at the status of the students' parents whether they still exist or not. If the parents are still there, then see if the GPA is high or low. If high, the student will receive a scholarship.

IV. Conclusions and Recomendations

Based on the findings of the study and the discussion that follows, it can be concluded that data clustering will yield good results if the k value (number of clusters) is chosen in the appropriate amount. If the number of clusters is too large, the results of the clustering will not be good. In making data predictions, the data should be converted into polynomial data or given a name first according to the data group (cluster) so that the resulting decision tree is easy to see and analyze. Also, out of 100 students whose data were processed, 9 (9%) were predicted to receive scholarships and 91 (91%) did not receive scholarships.

Based on the conclusions obtained, several things are suggested for future research, namely: We recommend that you use more attributes than we have done so that the results of the clustering and prediction that are carried out are better. We recommend that the data attribute for clustering with the K-Means method is in the form of numerical data and then the results are used as polynomial data so that it can be used for the prediction of the C4.5 method.

After the text edit has been completed, the paper is ready for the template. Duplicate the template file by using the Save As command, and use the naming convention prescribed by your conference for the name of your paper. In this newly created file, highlight all of the contents and import your prepared text file. You are now ready to style your paper; use the scroll-down window on the left of the MS Word Formatting toolbar.

References

[1]   B. Mardiyati and R. Yuniawati, "Differences in Career Adaptability in terms of Type of School (SMA dan SMK)," *Empathy J. Fak. Psikol.*, vol. 3, no. 1, pp. 31–41, 2015.

[2]   K. Winanti, Yuliyani, and A. Agoestanto, "Cultivating the Character of Students of SMAN 13 Semarang through Kanjusera," *J. Profesi Kegur.*, vol. 3, no. 2, pp. 16–24, 2017.

[3]   M. Zammi, E. Susilaningsih, and K. I. Supardi, "Increasing Motivation and Learning Outcomes of Circular Motion Physics with the Assisted Bicycle Bike Learning Method for Class X MIPA2 Students of SMA Negeri 1 Gemuh" *J. Profesi Kegur.*, vol. 4, no. 1, pp. 37–41, 2018.

[4]   L. T. Antika and L. Ibana, "Correlation between identification of evolution theory misconceptions and learning outcomes of Biology education students" *J. Pendidik. dan Kebud.*, vol. 3, no. 2, pp. 117–128, 2018.

[5]   A. Putra, J. S. Ardiwinata, and V. R. Hasanah, "Learning Components of the Cultural Literacy Program at Eco Bambu Cipaku" *J. Pendidik. dan Kebud.*, vol. 3, no. 2, p. 141, 2018, doi: 10.24832/jpnk.v3i2.921.

[6]   S. B. Raharjo, L. Yuliana, and Y. H. Yudha, "Achievement of National Education Standards as Predictors of School Quality" *J. Pendidik. dan Kebud.*, vol. 3, no. 2, p. 129, 2018, doi: 10.24832/jpnk.v3i2.750.

[7]   F. D. Indonesia, "Journal of higher education systems analysis" *Anal. Sist. Pendidik. Tinggi*, vol. 2, no. 1, pp. 1–60, 2018.

[8]   S. A. Nulhaqim, D. H. Heryadi, R. Pancasilawan, and M. Ferdryansyah, "The Role of Higher Education in Improving the Quality of Education in Indonesia to Face the 2015 Asean Community Case Study: University of Indonesia, Padjadjaran University, Bandung Institute of Technology" *Share   Soc.   Work J.*, vol. 6, no. 2, p. 197, 2016, doi: 10.24198/share.v6i2.13209.

[9]   U. Rahardja, I. Handayani, and R. Wijaya, "Application of the Technomedia Journal Viewboard using the iLearning Journal Center system in Higher Education" *Technomedia J.*,

vol. 2, no. 2, pp. 81–93, 2018, doi: 10.33050/tmj.v2i2.326.

[10] Q. Aini, U. Rahardja, A. H. Arribathi, and N. P. L. Santoso, "Application of Cloud Accounting in Support" vol. 4, no. 1, pp. 60–64, 2019.

[11] N. I. Kurniati, A. Rahmatulloh, and D. Rahmawati, "Comparison of the Performance of the Ant Colony Algorithm and the Genetic Algorithm - Taboo Search in Lecture Scheduling" *Comput. Eng. Sci. Syst. J.*, vol. 4, no. 1, p. 17, 2019, doi: 10.24114/cess.v4i1.11387.

[12] M. Kavitha, M. Suganthy, and R. Srinivasan, "Intelligent learning system using data mining-ilsdm," *Int. J. Innov. Technol. Explor. Eng.*, vol. 8, no. 9, pp. 505–508, 2019, doi: 10.35940/ijitee.i7705.078919.

[13] S. A. Safdar, T. Yue, S. Ali, and H. Lu, *Using multi-objective search and machine learning to infer rules constraining product configurations*, vol. 27, no. 1–2. Springer US, 2020.

[14] Adiwijaya, U. N. Wisesty, E. Lisnawati, A. Aditsania, and D. S. Kusumo, "Dimensionality reduction using Principal Component Analysis for cancer detection based on microarray data classification," *J. Comput. Sci.*, vol. 14, no. 11, pp. 1521–1530, 2018, doi: 10.3844/jcssp.2018.1521.1530.

[15] A. Alos and Z. Dahrouj, "Using multiple deep neural networks platform to detect different types of potential faults in unmanned aerial vehicles," *J. Aerosp. Technol. Manag.*, vol. 13, pp. 1–12, 2021, doi: 10.1590/jatm.v13.1186.

[16] A. R. Damanik *et al.*, "Implementation of the K-Means Method in Comparison of Unemployment Rates and Workers by Province," *Pros. Semin. Nas. Ris. Dan Inf. Sci.*, vol. 2, no. 0, pp. 24–31, 2020, [Online]. Available: https://www.bps.go.id.

[17] J. Jaroji, D. Danuri, and F. P. Putra, "K-Means To Determine Prospective Bidik Misi Scholarship Recipients in Polbeng" *INOVTEK Polbeng - Seri Inform.*, vol. 1, no. 1, p. 87, 2016, doi: 10.35314/isi.v1i1.129.

[18] B. M. Metisen and H. L. Sari, "Clustering analysis uses the K-Means method in grouping product sales at Fadhila Supermarkets" *J. Media Infotama*, vol. 11, no. 2, pp. 110–118, 2015.

[19] H. Hendri, H. Awal, and Mardison, "Implementation of Solar-Cell for Free Tourist Facilities and Tourism Promotion Media for the City of Padang" *STMIK Indones. Padang*, vol. 8, no. 2, p. 121, 2019.

[20] M. J. Nurul Rohmawati, Sofi Defiyanti, "Implementation of the K-Means Algorithm in Clustering Scholarship Applicants" *J. Ilm. Teknol. Inf. Terap.*, vol. I, no. 2, pp. 62–68, 2015.

[21] M. R. L. Iin Parlina, Agus Perdana Windarto, Anjar Wanto, "Utilizing the K-Means Algorithm in Determining Eligible Employees to Participate in the Assessment Center" *J. Comput. Eng. Syst. Sci.*, vol. 3, no. 1, pp. 87–93, 2018.

[22] L. Maulida, "Application of Datamining in Grouping Tourist Visits to Leading Tourist Attractions in Prov. Dki Jakarta with K-Means" *JISKA (Jurnal Inform. Sunan Kalijaga)*, vol. 2, no. 3, p. 167, 2018, doi: 10.14421/jiska.2018.23-06.

[23] R. A. Asroni, "Application of the K-Means Method for Student Clustering Based on Academic Values with Weka Interface Case Studies at the Department of Informatics, UMM Magelang" *Ilm. Semesta Tek.*, vol. 18, no. 1, pp. 76–82, 2015, doi: 10.1038/hdy.2009.180.