Data Mining Analysis as a Determinant of CPNS Passing Score Utilizing the K-Means Algorithm (Case Study of the BMKG Office in North Sumatra)

Rianto Kaban, Zulham Sitorus

Abstract

This research aims to analyse the application of data mining as a basis for determining the passing grade of prospective civil servants at the BMKG Office in North Sumatra using the K Means algorithm. The background of this research is based on the importance of establishing an objective, measurable passing grade that aligns with the distribution pattern of the participants' abilities. The method used is a quantitative approach with stages of collecting participant score data, data cleaning, normalisation, clustering process using the K Means algorithm, and interpretation of the clustering results. The processed data is grouped based on the level of achievement, resulting in several clusters that represent the participants' ability categories, such as high, medium, and low. The analysis results show that the K Means algorithm can help identify participant score patterns more systematically and provide a more accurate picture in determining the passing grade threshold. Thus, the use of data mining through the K Means algorithm can serve as an alternative decision-support tool in determining the CPNS passing grade more fairly, effectively, and based on data at the BMKG Office in North Sumatra.

***Keywords**:* Data Mining, Passing Grade, CPNS, K Means, Clustering, BMKG North Sumatra

Rianto Kaban[1]
[1]Information Technology, Universitas Pembangunan Panca Budi, Indonesia
e-mail: riantokaban@gmail.com[1]

Zulham Sitorus[2]
[2]Information Technology, Universitas Pembangunan Panca Budi, Indonesia
e-mail: zulhamsitorus@dosen.pancabudi.ac.id[2]
2nd International Conference on Islamic Community Studies (ICICS)
Theme: History of Malay Civilisation and Islamic Human Capacity and Halal Hub in the Globalization Era
https://proceeding.pancabudi.ac.id/index.php/ICIE/index

Introduction

The selection of Civil Servant Candidates is one of the strategic mechanisms in obtaining professional, competent human resources for the apparatus who are capable of supporting the effective administration of government. In the selection process, the determination of the passing grade has a very important role because it serves as the minimum graduation threshold that determines whether a participant is eligible to proceed to the next stage or not. Therefore, the determination of the passing grade must be done objectively, rationally, and based on data so that the selection results can truly reflect the quality of the participants. Setting an inappropriate threshold value has the potential to cause injustice, either because it is too high, thereby reducing the chances of actually competent participants, or too low, thereby lowering the quality standards of the accepted state apparatus candidates.

This research uses a qualitative approach by conducting interviews and document reviews to collect data, which is then analysed to obtain a detailed description of the object being studied. The results of the research that have been presented earlier can be concluded that, in general, the implementation of the Actualisation Learning of the Basic Training for Civil Servant Candidates at the Provincial Human Resource Development Agency of South Sulawesi, starting from the determination of issues, the creation of the actualisation design, the guidance of the actualisation design [1]. Experts have verified the number of diseases in each clustering—Safe 10 diseases, Caution 7 diseases, and Emergency 1 disease—based on the tests carried out in this study. The system and the manual calculation get identical results. Using Confusion Matrix testing, the K-Means approach yields an average cluster accuracy of 88.95% [2]. This research applies the data mining process with clustering techniques and uses the k-means clustering method, K-Means algorithm. Meanwhile, the method used in this research is CRISP-DM (Cross-Industry Standard Process for Data Mining). [3], [4].

In this case, Cluster Analysis plays a significant role in the formation of those segments. Cluster Analysis is a technique used to divide a set of objects into groups so that the values within each group are homogeneous, referring to certain attributes based on specific criteria. The similarity of values (homogeneity) in each segment represented by the cluster reflects the similarity of CPNS behaviour patterns. The higher the clustering accuracy, the clearer the similarity of CPNS behaviour patterns that can be uncovered through the formed clusters. Thus, business practitioners can determine performance strategies more accurately.

In practice, the determination of the passing grade is often still orientated towards a general uniform approach and does not fully consider the distribution pattern of participants' scores in a specific institution. In fact, the characteristics of exam results in each institution can vary depending on the number of participants, the difficulty level of the questions, educational background, and the competency requirements of the position applied for. This condition indicates that the conventional approach to setting passing grades may not necessarily represent the actual data conditions produced from the selection process. Thus, an analytical approach that can read data patterns more deeply is needed so that the determination of the passing grade can be done in a more adaptive manner and based on empirical evidence.

The development of information technology has encouraged the use of data in various decision-making processes, including in the field of human resource management and employee selection. One of the approaches that can be used is data mining, which is the process of extracting hidden information, patterns, and knowledge from a large set of data. In the context of CPNS selection, data mining can be utilised to identify groups of participants' abilities based on the similarity of the obtained score characteristics. Thus, the results of data processing not only serve as administrative archives but can also form the basis for analysis in formulating more accurate threshold value policies.

One of the algorithms widely used in data mining is the K Means algorithm. This algorithm works by grouping data into several clusters based on the degree of proximity or similarity between the data. In this study, K Means is considered relevant because it can categorise the scores of CPNS participants into specific groups, such as high, medium, and low score groups. Through this

categorisation, researchers can observe the distribution patterns of scores more systematically and determine the score thresholds deemed representative as the passing grade. The use of the K Means algorithm also has advantages in terms of process simplicity, calculation efficiency, and its ability to process numerical data objectively.

The North Sumatra BMKG Office, as one of the government agencies that requires personnel with specialised competencies, serves as an important context in this research. As an institution engaged in the fields of meteorology, climatology, and geophysics, BMKG requires employees who not only meet administrative requirements but also possess high intellectual capacity and meticulousness. Therefore, the CPNS selection process within BMKG needs to be supported by an accurate assessment system to effectively filter candidates who truly meet the institution's needs. Analysis of the selection data at the BMKG Office in North Sumatra is important to see whether the participants' score patterns can be used as a basis for determining a more measurable passing grade.

Academically, this research has strong relevance as it lies at the intersection of information systems, data analysis, and public sector human resource management. This research not only positions data mining as a technical tool but also as a scientific approach to support more objective decision-making. Furthermore, studies on the application of the K Means algorithm in determining the passing grade for CPNS are still relatively limited, particularly in case studies of regional or vertical government agencies such as BMKG North Sumatra. Customers are grouped in this study according to financial and demographic characteristics, including age, gender, income, marital status, and ownership of banking assets, using the K-Means clustering method, The results of this research can be grouped into 3 clusters using K-Means [5], [6]. The K-Means Clustering technique is used in data mining to assess student learning results [7]. Their study uses the K-Means clustering algorithm to group the PYMES data from 2021–2023 according to investment value, production volume, and production value [8].

Thus, this research is expected to provide theoretical contributions to the development of clustering method applications in the field of employee selection, while also offering practical contributions to institutions in formulating more data-driven selection policies.

Based on the description, it can be understood that the determination of the passing grade for CPNS requires a method that not only relies on general provisions but also takes into account the empirical conditions of the country.

.

Literature Review

2.1 Data Mining

Data mining is a term used to describe the discovery of knowledge within a database. Data mining is a process that employs statistical techniques, mathematics, artificial intelligence, and machine learning to extract and identify useful information and related knowledge from various large databases [9]. The general definition of data mining itself is the process of searching for hidden patterns in the form of knowledge that was previously unknown from a set of data, which can be found in databases, data warehouses, or other information storage media. Important aspects related to data mining are:

1. Data mining is an automatic process on existing data.
2. The data to be processed is very large data.
3. The goal of data mining is to obtain relationships or patterns that may provide useful indications [10].

Data mining is performed using specialised tools to execute data operations that have been defined based on an analysis model. Data mining is the process of analysing data with an emphasis on discovering hidden information within a large amount of data stored when running the company's business. The remarkable and continuous advancements in the field of data mining are driven by several factors, including:

1. Rapid growth in data sets.

2. Data storage in data warehouses, allowing the entire company access to a reliable database.
3. Increased data access through web navigation and the internet.
4. Business competition pressure to enhance market dominance in economic globalisation.

2.2 Cross Industry Standart Process for Data Mining

Cross-Industry Standard Process for Data Mining (CRISP-DM) is a consortium of companies developed in 1996 by analysts from several industries such as Daimler Chrysler, SPSS, and NCR. CRISP-DM provides a standard data mining process as a general problem-solving strategy for businesses or research units. In CRISP-DM, a data mining project has a process divided into six phases. The entire sequence of phases is adaptive. The next phase in the sequence depends on the output of the previous phase. The following is Figure 2.1, the development life cycle process of CRISP-DM:

Figure 1. Data Mining Phase Model in CRISP-DM [9]

2.3 Classification

Classification is the process of finding a model or function that describes and distinguishes classes or data concepts. The function of classification is to classify a target class into the chosen categories [11]. There are many methods to build classification models such as naive Bayesian classification, k-means, and k-nearest neighbour classification.

2.4 K- Means

Data Clustering is one of the Data Mining methods that is unsupervised. There are two types of data clustering that are often used in the data grouping process, namely hierarchical data clustering and non-hierarchical data clustering. K-Means is one of the non-hierarchical data clustering methods that attempts to partition existing data into one or more clusters/groups. This method partitions data into clusters/groups so that data with similar characteristics are grouped into the same cluster, and data with different characteristics are grouped into different clusters. In addition, it serves as a Decision Support System and Data Mining, such as market segmentation, area mapping, marketing management, etc [12].

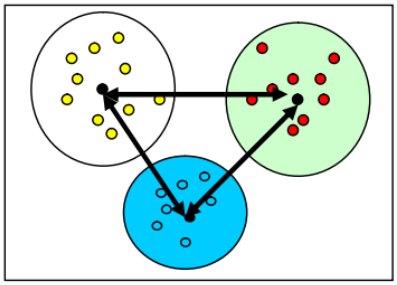

Figure 2. Clustering

Characteristics of K-means:

1. K-means is very fast in the clustering process.
2. K-means is very sensitive to the random generation of initial centroids.

3. It allows for a cluster to have no members.
4. The clustering results with K-means are unique (always changing, sometimes good, sometimes bad).

Distance Space for Calculating the Distance Between Data and Centroid Several distance spaces have been implemented to calculate the distance (distance between data and centroid), including L1 (Manhattan/City Block) distance space, L2 (Euclidean) distance space, and Lp (Minkowski) distance space. The distance between two points x1 and x2 in the Manhattan/City Block distance space is calculated using the following formula [12].

$$D_{L_1}(x_2, x_1) = \|x_2 - x_1\|_1 = \sum_{j=1}^{p} |x_{2j} - x_{1j}|$$

Meanwhile, for the L2 (Euclidean) distance space, the distance between two points is calculated using the following formula:

$$D_{L_2}(x_2, x_1) = \|x_2 - x_1\|_2 = \sqrt{\sum_{j=1}^{p} (x_{2j} - x_{1j})^2}$$

Prasetyo Eko (2013) states that the K-Means clustering algorithm is a distance-based clustering technique that attempts to partition data into several clusters. This method partitions data into clusters according to the characteristics possessed by each data point, with data points that share the same characteristics grouped into the same cluster, and data points with different characteristics grouped into different clusters.

In this algorithm, the center of the cluster is called the centroid. The centroid is a random value from the entire dataset selected at the initial stage. Then, K-Means selects each component from the entire data and separates the data into one of the previously described centroids based on the closest distance between the data components and the center of each centroid, with the condition that no more data moves between groups. The K-means data clustering algorithm is as follows (Eko, P., 2013).

In step 3 of the above Algorithm, the location of the centroid (center point) of each group, which is taken from the average (mean) of all data values for each feature, must be recalculated. If M represents the number of data points in a cluster, i represents the i-th feature in a cluster, and p represents the data dimension, then to calculate the centroid of the i-th feature.

$$C_i = \frac{1}{M} \sum_{j=1}^{M} x$$

The formula is performed for p dimensions, where i ranges from 1 to p. The method of measuring the distance of data to the cluster center uses Euclidean (Bezdek, 1981).

$$D(x_2, x_1) = \|x_2 - x_1\|_2 = \sqrt{\sum_{j=1}^{p} |x_{2j} - x_{1j}|^2}$$

Explanation
D = the distance between data x2 and x1, and | . | is the absolute value.
P = Dimension of data
X_{2j} = Coordinates of object i in dimension k
X_{1j} = Coordinates of object j in dimension k

In step 4 of the Algorithm, the reallocation of data into each group in the K-means method is based on the comparison of the distance between the data and the centroid of each existing group. The data is strictly reallocated to the group with the centroid closest to the data. The allocation of data to clusters uses equation 2.3 (MacQueen, 1967):

$$a_{il} = \begin{cases} 1 & d = \min\{D(x_i, C_l)\} \\ 0 & lainnya \end{cases}$$

Where ail is the membership value of point xi to the cluster center Cl, d is the shortest distance from data xi to K clusters after comparison, and Cl is the centroid (cluster center) of the l-th cluster.

The objective function is based on the distance and membership value of data within the cluster.

$$J = \sum_{i=1}^{N} \sum_{l=1}^{K} a_{ic} D(x_i, C_l)^2$$

Where N is the number of data points, K is the number of clusters, ail is the membership value of the data point xi to the center of cluster Cl, Cl is the center of the k-th cluster, D(xi,Cl) is the distance from the point xi to the followed cluster Cl. For a to have a value of 0 or 1. If a data point is a member of a group, then the value of ail = 1; if not, then the value of ail = 0.

Research Methodology

By using the clustering method with the k-means algorithm, the obtained data can be grouped into several clusters based on the similarity of the data, so that data with similar characteristics are grouped into one cluster and data with different characteristics are grouped into another cluster with the same characteristics. With the clustering of data like this, it is hoped that the marketing department can carry out marketing with the right strategy to attract new civil servant candidates.

1. Research Approach
   This research uses a quantitative approach with descriptive and experimental designs. The quantitative approach was chosen to analyse the data from the CPNS selection obtained from the BMKG North Sumatra Office. The collected data will be analysed using data mining methods with the K Means algorithm to identify patterns and clusters of CPNS selection participants' scores.
2. Type of Research
   The type of research used in this study is applied research. This research aims to apply the K Means algorithm in the context of CPNS selection at the BMKG Office in North Sumatra. In addition, this research also aims to produce recommendations that can be used as a basis for determining a more accurate and objective passing grade.
3. Populasi dan Sampel
   The population in this study consists of all CPNS selection participants at the BMKG Office in North Sumatra for the current year. The data used in this study includes the exam scores of all CPNS selection participants who took the exam during that period. The research sample consists of participants who took the CPNS selection at the BMKG Office in North Sumatra and passed the administrative selection stage, as well as having complete and analysable exam result data. The sample size will be adjusted according to the available data and the number of participants who took the exam that year.

Results

In this study, a data processing process is carried out as an analysis stage, requiring data in the form of civil servant candidates (CPNS) who have graduated from institutions in 2024. The data contains personal information of the graduated CPNS, but in this study, only a few data attributes are used, such as CPNS, the city of the institution, and the last education.

Table 1. Example Data of CPNS Who Have Passed

Name Cpns	Hometown	Institution	IP,K
xxxx	Deli Serdang	BMKG	3,39
xxxx	Medan	BMKG	3,18
xxxx	Medan	BMKG	3,25

xxxx	Padang	BMKG	3,35
xxxx	Asahan	BMKG	3,03
xxxx	Binjai	BMKG	3,49

In order for the above data to be processed using the k-means clustering method, nominal data such as the city of origin and institution must first be initialised in numerical form. To initialise the origin cities, the following steps are taken:

1. The origin city data is first divided into several regions, namely:
   a. North Sumatra region and the cities of Aceh Besar, Medan, Binjai, Banda Aceh, Lhokseumawe,
   b. West Sumatra region consisting of the cities of Padang, Bukit Tinggi
2. Then, these regions are sorted from largest to smallest based on the frequency of CPNS originating from those regions.

Table 2. Initialisation of Origin City Region Data

Region	Frequency	Initials
North Sumatra	5	1
West Sumatra	3	2

To be able to cluster the data into several clusters, several steps need to be taken, namely:

1. Determine the desired number of clusters. In this study, the existing data will be grouped into three clusters.
2. Determine the initial center point of each cluster. In this study, the initial center points are determined randomly, and the selected center points are data points 3, 4, and 6.
3. Calculate the distance of each data point from the center point of each cluster. For example, the distance from the first cpns data to the first cluster center will be calculated:

   a. Calculation of the distance from data 1 to the cluster center

   b. $A1,1 = \sqrt{(2-1)^2 + (1-2)^2 + (3{,}39-3{,}18)^2} = 1{,}021$

   c. $A1,2 = \sqrt{(2-3)^2 + (1-2)^2 + (3{,}39-3{,}35)^2} = 1{,}414$

   d. $A1,3 = \sqrt{(2-1)^2 + (1-1)^2 + (3{,}39-3{,}49)^2} = 1{,}004$

   b. Calculation of the distance from the second data point to the cluster center

   $A2,1 = \sqrt{(2-1)^2 + (1-1)^2 + (3{,}18-3{,}18)^2} = 1$

   $A2,1 = \sqrt{(2-3)^2 + (1-2)^2 + (3{,}18-3{,}35)^2} = 1{,}424$

   $A2,3 = \sqrt{(2-1)^2 + (1-1)^2 + (3{,}18-3{,}49)^2} = 1{,}046$

   c. Calculation of the distance from the 3rd data point to the cluster center

   $A3,1 = \sqrt{(1-1)^2 + (2-1)^2 + (3{,}25-3{,}18)^2} = 1{,}002$

   $A3,2 = \sqrt{(1-3)^2 + (2-2)^2 + (3{,}25-3{,}35)^2} = 2{,}002$

   $A3,3 = \sqrt{(1-1)^2 + (2-1)^2 + (3{,}25-3{,}49)^2} = 0{,}758$

   d. Calculation of the Distance from Data 4 to the Cluster Center

   $A4,1 = \sqrt{(3-1)^2 + (2-1)^2 + (3{,}35-3{,}18)^2} = 2{,}242$

   $A4,2 = \sqrt{(3-3)^2 + (2-2)^2 + (3{,}35-3{,}35)^2} = 0$

   $A4,3 = \sqrt{(3-1)^2 + (2-1)^2 + (3{,}35-3{,}49)^2} = 1{,}964$

e. Calculation of the Distance from data point 5 to the cluster center

$$A5,1 = \sqrt{(2-1)^2 + (2-1)^2 + (3{,}03-3{,}18)^2} = 1{,}422$$

$$A5,2 = \sqrt{(2-3)^2 + (2-2)^2 + (3{,}03-3{,}35)^2} = 1{,}04$$

$$A5,3 = \sqrt{(2-1)^2 + (2-1)^2 + (3{,}03-3{,}49)^2} = 1{,}10$$

f. Calculation of the Distance from data point 6 to the cluster center

$$A6,1 = \sqrt{(1-1)^2 + (1-1)^2 + (3{,}49-3{,}18)^2} = 0{,}31$$

$$A6,2 = \sqrt{(1-3)^2 + (1-2)^2 + (3{,}49-3{,}.35)^2} = 2{,}29$$

$$A6,3 = \sqrt{(1-1)^2 + (1-1)^2 + (3{,}49-3{,}49)^2} = 0$$

g. Calculation of the Distance from data point 7 to the cluster center

$$A7,1 = \sqrt{(1-1)^2 + (1-1)^2 + (3{,}55-3{,}18)^2} = 0{,}37$$

$$A7,2 = \sqrt{(1-3)^2 + (1-2)^2 + (3{,}55-3{,}.35)^2} = 2{,}24$$

$$A7,3 = \sqrt{(1-1)^2 + (1-1)^2 + (3{,}55-3{,}49)^2} = 0{,}036$$

h. Calculation of the Distance from data point 8 to the cluster center

$$A8,1 = \sqrt{(1-1)^2 + (1-1)^2 + (3{,}45-3{,}18)^2} = 0{,}27$$

$$A8,2 = \sqrt{(1-3)^2 + (1-2)^2 + (3{,}45-3{,}.35)^2} = 2{,}23$$

$$A8,3 = \sqrt{(1-1)^2 + (1-1)^2 + (3{,}45-3{,}49)^2} = 0{,}04$$

The complete calculation results for the 8 CPNS data can be seen in the table Calculation Results for Each Data to Each Cluster.

Table 3. Calculation Results for Each Data to Each Cluster

No	Name	Institution	Hometown	IPK	C1	C2	C3	Results
1	xxxxxx	2	1	3,39	1,021	1,414	1,004	C3
2	xxxxxx	2	1	3,18	1	1,424	1,046	C1
3	xxxxxx	1	2	3,25	1,002	2,002	0,758	C3
4	xxxxxx	3	2	3,35	2,42	0	1,964	C2
5	xxxxxx	2	2	3,03	1,422	1,04	1,10	C2
6	xxxxxx	1	1	3,49	0,31	2,24	0	C3

After obtaining the new centroid of each cluster, repeat from step three until the centroid of each cluster no longer changes and no data moves from one cluster to another. Repeat step two until the data positions do not change.

a. Calculation of the Distance from data point 1 to the cluster center

$$A1,1 = \sqrt{(2-2)^2 + (1-1)^2 + (3{,}39-3{,}18)^2} = 0{,}21$$

$$A1,2 = \sqrt{(2-2{,}5)^2 + (1-2)^2 + (3{,}39-3{,}19)^2} = 1{,}135$$

$$A1,3 = \sqrt{(2-1,2)^2 + (1-1,2)^2 + (3,39-3,4158)^2} = 0,824$$

b. Calculation of the Distance from the 2nd data to the cluster center

$$A2,1 = \sqrt{(2-2)^2 + (1-1)^2 + (3,18-3,18)^2} = 0$$

$$A2,1 = \sqrt{(2-2,5)^2 + (1-2)^2 + (3,18-3,19)^2} = 1,118$$

$$A2,3 = \sqrt{(2-1,2)^2 + (1-1,2)^2 + (3,18-3,4158)^2} = 0,735$$

c. Perhitungan jarak dari data ke 3 terhadap pusat cluster

$$A3,1 = \sqrt{(1-2)^2 + (2-1^2 + (3,25-3,18)^2} = 1,415$$

$$A3,2 = \sqrt{(1-2,5)^2 + (2-2)^2 + (3,25-3,19)^2} = 1,501$$

$$A3,3 = \sqrt{(1-1,2)^2 + (2-1,2)^2 + (3,25-3,4158)^2} = 0,840$$

d. Calculation of the distance from data point 3 to the cluster center

$$A4,1 = \sqrt{(3-2)^2 + (2-1)^2 + (3,35-3,18)^2} = 1,424$$

$$A4,2 = \sqrt{(3-2,5)^2 + (2-2)^2 + (3,35-3,19)^2} = 0,507$$

$$A4,3 = \sqrt{(3-1,2)^2 + (2-1,2)^2 + (3,35-3,4158)^2} = 3,88$$

e. Calculation of the Distance from data point 5 to the cluster center

$$A5,1 = \sqrt{(2-2)^2 + (2-1)^2 + (3,03-3,18)^2} = 1,011$$

$$A5,2 = \sqrt{(2-2,5)^2 + (2-2)^2 + (3,03-3,19)^2} = 0,524$$

$$A5,3 = \sqrt{(2-1,2)^2 + (2-1,2)^2 + (3,03-3,4158)^2} = 1,428$$

f. Calculation of the Distance from data point 6 to the cluster center

$$A6,1 = \sqrt{(1-2)^2 + (1-1)^2 + (3,49-3,18)^2} = 1,046$$

$$A6,2 = \sqrt{(1-2,5)^2 + (1-2)^2 + (3,49-3,19)^2} = 1,82$$

$$A6,3 = \sqrt{(1-1,2)^2 + (1-1,2)^2 + (3,49-3,4158)^2} = 0,291$$

g. Calculation of the Distance from data point 7 to the cluster center

$$A7,1 = \sqrt{(1-2)^2 + (1-1)^2 + (3,55-3,18)^2} = 1,066$$

$$A7,2 = \sqrt{(1-2,5)^2 + (1-2)^2 + (3,55-3,.19)^2} = 1,838$$

$$A7,3 = \sqrt{(1-1,2)^2 + (1-1,2)^2 + (3,55-3,4158)^2} = 0,313$$

h. Calculation of the Distance from data point 8 to the cluster center

$$A8,1 = \sqrt{(1-2)^2 + (1-1)^2 + (3,45-3,18)^2} = 1,035$$

$$A8,2 = \sqrt{(1-2,5)^2 + (1-2)^2 + (3,45-3,.19)^2} = 1,821$$

$$A8,3 = \sqrt{(1-1,2)^2 + (1-1,2)^2 + (3,45-3,4158)^2} = 0,811$$

Based on the results of data clustering using the k-means clustering method, clustering results were obtained up to the 2nd iteration, where the centroids no longer changed and no data moved between clusters.

Conclusion

Based on the results of this study, it can be concluded that the application of the K Means algorithm in the analysis of CPNS selection data at the BMKG Office in North Sumatra has proven effective in grouping participants based on their exam scores. This algorithm is capable of identifying a more systematic and objective pattern of score distribution, thereby providing a clearer picture of the different ability groups of the participants. Through the clustering process, participants can be grouped into categories that reflect their ability levels, such as high, medium,

and low score groups. The use of data mining with the K Means algorithm provides significant benefits in determining a more accurate passing grade. Previously, the process of setting the threshold value was often done with a uniform approach without considering the distribution of participants' scores. By using this approach, BMKG North Sumatra can set the passing grade based on more concrete data and in accordance with the participants' score patterns.

Overall, this research shows that the K Means algorithm can not only be used in academic data clustering but can also be applied to support decision-making in civil servant selection, particularly in determining a more equitable and data-driven passing threshold. Therefore, the application of this method is expected to serve as a useful alternative in improving the CPNS selection system in the future, thereby enhancing the quality and transparency of the state apparatus recruitment process.

.

References

[1] M. Yamin, "Implementasi Pembelajaran Aktualisasi Latsar CPNS pada Badan Pengembangan Sumber Daya Manusia Provinsi Sulawesi Selatan," no. 101, 2015.

[2] D. Sioyong, "IMPLEMENTASI DATA MINING MENGGUNAKAN ALGORITMA K-MEANS CLUSTERING PENYAKIT PASIEN RAWAT JALAN PADA KLINIK DR . ATIRAH," vol. 7, no. 4, pp. 2144–2151, 2023.

[3] W. I. Rahayu, S. F. Pane, L. A. Frimanda, and K. Kunci, "IMPLEMENTASI DATA MINING DENGAN METODE K-MEANS CLUSTERING UNTUK MENENTUKAN IKLAN AUDIO BERDASARKAN USER BEHAVIORS PADA APLIKASI AUDIO SOCIAL MEDIA SVARA DI PT . ZAMRUD KHATULISTIWA TECHNOLOGY," vol. 10, no. 2, pp. 13–19, 2018.

[4] D. I. Wilayah and K. Depok, "Implementasi algoritma k-means untuk clustering kasus kriminal di wilayah kota depok 123," vol. 2, no. 1, pp. 408–415, 2025.

[5] F. P. Dewi, P. S. Aryni, and Y. Umaidah, "Implementasi Algoritma K-Means Clustering Seleksi Siswa Berprestasi Berdasarkan Keaktifan dalam Proses Pembelajaran," vol. 7, no. 2, pp. 111–121, 2022.

[6] D. Irawan, G. Wijaya, and T. T. Warisaji, "Penerapan Algoritma K-Means Clustering untuk Segmentasi Nasabah Bank," vol. 6, no. 1, 2025.

[7] N. Hendrastuty, "Penerapan Data Mining Menggunakan Algoritma K-Means Clustering Dalam Evaluasi Hasil Pembelajaran Siswa," vol. 3, pp. 46–56, 2024.

[8] B. Ariansah *et al.*, "Penerapan K-Means Clustering untuk Pengelompokan Data Industri Kecil Menengah di Provinsi Jambi," vol. 6, no. September, pp. 359–371, 2025, doi: 10.35957/jtsi.v6i2.13553.

[9] I. Kurikulum, "Implementasi kurikulum 2013," 2013.

[10] P. P. Desain, "PARADIGMA PENDIDIKAN DESAIN".

[11] "No Title".

[12] M. Terkait, "K-Means – Penerapan, Permasalahan dan Metode Terkait," vol. 3, no. Pebruari, pp. 47–60, 2007.