SSRG International Journal of Electronics and Communication Engineering
ISSN: 2348-8549/ https://doi.org/10.14445/23488549/IJECE-V10I5P109

Volume 10 Issue 5, 96-104, May 2023

Original Article

Data Mapping using Combining Clustering Methods and C.45 Classification

Robbi Rahim[1], Unik Hanifah Salsabila[2], Akhmad Anwar Dani[3], Eka Maya S.S. Ciptaningsih[4], M. Mohzana[5]

[1]*Sekolah Tinggi Ilmu Manajemen Sukma, Medan, Indonesia*
[2]*Universitas Ahmad Dahlan, Yogyakarta, Indonesia*
[3]*IAIN Surakarta, Surakarta, Indonesia*
[4]*Bina Nusantara University, Jakarta, Indonesia*
[5]*Universitas Hamzanwadi, Lombok, Indonesia*

[1]*Corresponding Author : usurobbi85@zoho.com*

Received: 10 March 2023 Revised: 14 April 2023 Accepted: 10 May 2023 Published: 31 May 2023

Abstract *- School participation is measured by the Pure Participation Rate (APM). This study examines whether data mining can generate new knowledge. The Central Sumatra Statistic Central Agency (BPS-North Sumatra) provided secondary statistics on APM by city/district (2011–2019) for elementary, junior high, high school, and PT. Data mining uses clustering (k-means) and classification (Decision tree). This cluster maps the APM. Mapping clusters are utilized again for categorization. Cluster value ranges indicate classification. C1 was the high APM cluster, and C2 was the low APM cluster. RapidMiner aids processing. The study found 18 high-cluster (C1) cities and 15 low-cluster cities (C2). Based on the clustering results obtained, classification results show that SMA and PT become influential attributes in mapping the area based on the Decision tree method, resulting in 3 rules: if SMA has a percentage <68,085% (high cluster); if SMA has a percentage> 68,085% and PT has a presentation <18,730% (low cluster); and if SMA and PT have a presentation> 18,730%. (high cluster). Classification and clustering have yielded new data.*

Keywords *- Classification, Data mining, North sumatra, Pure participation.*

1. Introduction

To see school participation in an area, several indicators can be used to find out, among others: School Participation Rate (APS), Rough Participation Rate (APK), and Pure Participation Rate (APM). This is very closely related to the success of the development of an area where qualified human resources (from now on abbreviated as HR) are needed[1-4]. Because education is one way to improve the quality of these human resources. One indicator used is the Pure Participation Rate (APM), which compares sure school-age students at the level of education with the appropriate age population and is expressed as a percentage. The APM aims to determine the number of school-age children attending school at the appropriate level. So that the higher the NER in an area means that more school-age children attend school according to the official age at a certain level of education (Ideally 100%)[5]. The Central Statistics Agency for North Sumatra (abbreviated as BPS-North Sumatra) is a non-ministerial government agency where the BPS-North Sumatra website can be accessed https://sumut.bps.go.id/ which provides all the information needed ranging from social and population[6], economic and trade, agriculture and mining. APM data presented by BPS-North Sumatra in a tabular form consisting of 33 districts/cities in North Sumatra where the APM consists of "primary school" APM, "junior high school" APM, "as high school" and APM "tertiary education" Can provide helpful information if processed using computer science techniques [7]. In this case, researchers used mapping (cluster) and mapping classification using data mining techniques.

Data mining is a branch of artificial intelligence that can perform unsupervised learning techniques which do not require training data or test data to make predictions, classifications, and classifications [8-11]. One of the clustering techniques often used is k-means [12]. K-Means has the advantage of being efficient for smaller data sets [13]. Also used for datasets with few attributes [14]–[16]. For these advantages, with a small dataset and few attributes, k-means becomes a solution in mapping compared to k-medoids. One example of research that uses the ability of k-means is [14] about mapping the potential areas of productive rubber farming in northern Sumatra. In this research, the k-means method can be applied with the results of the number of clusters of 3 labels, with the highest group results being 1 item (C1), the middle group there are six items (C2), and the lowest group there are 19 items (C3) with a total number of data 26 records. In addition, the clustering results will be classified by the Decision Tree (C4.5) method. Method C4.5 has the advantage of producing a decision tree that is easily interpreted [17], [18]. The results of the decision tree classification represent the rules, where the rules can be easily understood with natural language. Based on this, research results will provide new information about the North Sumatra region's Pure Participation Rate (APM).

2. Literature Review

2.1. Data Mapping

Data integration relies on data mapping. According to Zohra, it helps information systems interact by translating data from one format to another [19]. Due to rising data volumes and complexity, Sheth & Larson[20] found that automated data mapping solutions are in demand. In big data situations, these techniques improve efficiency and accuracy[21].

2.2. Clustering

Clustering is a widely utilized unsupervised machine learning methodology that finds application in data mining and statistical data analysis[22]. According to Jain[23], clustering is a method that involves categorizing data points or objects to ensure that those in the same cluster exhibit more remarkable similarity to each other than those in different clusters based on specific predetermined standards. Xu and Wunsch[24]have proposed a novel density-based clustering algorithm that surpasses conventional techniques in effectively handling noise and identifying clusters with varying densities. This is an example of recent progress in clustering algorithms made by researchers.

2.3. C4.5 Algorithm

Decision tree learning uses the C4.5 algorithm extensively. The Quinlan[25] technique uses information entropy to create decision trees from structured training data. Data builds these trees. This method categorizes well. Despite its popularity and efficacy, the approach has limitations, including overfitting and difficulty handling missing data[26]. We have overcome these constraints recently. Mazid[27] optimized decision tree construction using evolutionary techniques to enhance accuracy over C4.5. The latest C4.5 algorithm.

2.4. Intersections and Applications

Data mapping, clustering, and C4.5 converge in many applications, especially data mining. Schaefer[28] analyzed massive healthcare data using multiple methods. Data mapping pre-processed the different types of data obtained.

Data mapping, clustering, and C4.5 work well for data analysis and prediction. In large, complex data sets in healthcare and geographic analysis, using all of them may reveal helpful information[29].

3. Methodology

3.1. Data Mining

Data mining is a process of discovering meaningful patterns, relationships, and new trends by filtering vast amounts of data stored in previously unknown storage [8], [14], [30-32]. Data mining processing consists of predictive modelling, association, classification and classification [12], [33-35].

3.2. K-Means Method

The k-means method is a method that is quite easy to implement and run, relatively fast, easy to adjust and widely used. The main principle of the k-means method is to compile centroids) from a collection of data that begins with the initial cluster formation. then iteratively repairs the cluster until there are no significant changes in the cluster [36, 37].

3.3. Decision Tree Method (C4.5)

Decision trees are a widely recognized classification technique that transforms voluminous data into decision trees that embody rules. Furthermore, decision trees are a valuable tool for data exploration, as they facilitate the identification of latent associations among multiple potential input variables and a target variable[38, 39].

3.4. Data

The data in this study used secondary data obtained from the Central Statistics Agency for the North Sumatra Region (abbreviated as BPS-North Sumatra - https://sumut.bps.go.id/) for the Pure Participation Rate (APM) of the North Sumatra region in 2011-2019, which consisted of from 33 districts/ cities. APM data consists of the percentage of APM at elementary, junior high, high school and university levels for each district/ city. Data will be processed using RapidMiner software assistance.

4. Results and Discussion

At this juncture, the information exhibited in Table 1 shall undergo processing. The initial step involves performing mapping through the utilization of the k-means algorithm. The outcomes of the mapping process will be subjected to classification through the utilization of the C4.5 algorithm. This will enable the visualization of the rules in a decision tree format, facilitating their comprehension straightforwardly and intuitively. During the clustering process utilizing the k-means algorithm, the assigned mapping labels consist of two distinct clusters: the high cluster (C1) and the low cluster (C2). The present study outlines a combination model design that integrates clustering and classification techniques, implemented through RapidMiner software.

The model depicted in Figure 1 utilizes an Excel file to elucidate the input data, which is subsequently linked to the k-means and C4.5 (decision tree) methodologies, resulting in clustering and classification outputs. Furthermore, a performance evaluation examines the correlation between the clusters generated using the Davies-Bouldin Index tool as a benchmark for optimal clustering. The Bregmann Divergences were utilized as the size type for finding the nearest neighbour in the cluster results obtained through the RapidMiner software. Bregman divergences are a class of "closeness" measures that are more general and do not adhere to symmetry or triangle inequality properties.

Following are the complete results of clustering that has been exported from RapidMiner to Excel, as shown in Table 2, where the high cluster (hc_0) and low cluster (lc_1).

In table 2 it can be explained the results of the mapping in the form of clusters of pure participation rates by district/city obtained high cluster results (C1) consisting of 18 districts/cities (N i a s, Mandailing Natal, South Tapanuli, A s a h a n, Simalungun, Deli Serdang, Langkat, South Nias, Serdang Bedagai, Batu Bara, Padang Lawas Utara, Labuhanbatu Utara, Pematangsiantar, Tebing Tinggi, M e d a n, B i n j a I, Padangsidimpuan, Gunungsitoli) and low cluster results (C2) consisting of 15 districts/cities (Central Tapanuli, North Tapanuli, Toba, Labuhanbatu, Dairi, Karo, Humbang Hasundutan, Pakpak Barat, Samosir, North Padang Lawas, Labuhanbatu Selatan, North Nias, West Nias, S i b o l g a, Tanjungbalai). Based on the mapping results, the percentage of high (C1) and low (C2) clusters still has a mean value that is not much different, namely 54.5% (C1) and 45% (C2).

Table 1. Research data

Initial	Regency / City	Elementary school	Middle School	High school	PT
01	N i a s	98.78	78.11	62.52	5.85
02	Mandailing Natal	99.22	83.14	62.57	13.09
03	South Tapanuli	98.57	82.07	66.93	15.34
04	Central Tapanuli	99.07	88.26	70.66	15.63
05	North Tapanuli	99.35	88.2	78.23	17.71
06	Toba	98.52	89.22	82.07	2.57
07	Labuhanbatu	99.37	86.94	68.22	10.82
08	Asahan	99.82	81.91	60.93	15.2
09	Simalungun	98.64	77.48	63.7	20.07
10	Dairi	99.36	90.53	80.65	9.48
11	Karo	98.64	83.15	73.18	10.53
12	Deli Serdang	95.03	70.82	67.81	19.77
13	Langkat	98.93	78.86	64.43	13.13
14	South Nias	95.91	70.74	66.73	10.86
15	Humbang Hasundutan	99.29	92.94	86.51	10.3
16	Pakpak Barat	99.05	88.02	80.01	8.27
17	Samosir	99.57	91.58	81.78	3.6
18	Serdang Bedagai	99.14	77.67	67.95	9.67
19	Batu Bara	99.2	74.37	60.19	10.89
20	North Padang Lawas	98.78	83.28	69.18	7.1
21	Padang Lawas Utara	98.95	82.82	62.62	10.93
22	Labuhanbatu Selatan	98.8	84.44	71.75	11.02
23	Labuhanbatu Utara	99.8	74.57	64.27	12.51
24	North Nias	98.61	80.79	74.03	7.81
25	West Nias	99.52	82.58	78.85	5.94
26	Sibolga	99.14	87.91	74.27	9.23
27	Tanjungbalai	98.38	81.82	71.75	10.04
28	Pematangsiantar	99.58	81.55	75.78	23.11
29	Tebing Tinggi	98.04	82.62	67.17	11.81
30	Medan	93.47	80	61.43	33.54
31	Binjai	99.26	83.43	72.62	20.4
32	Padangsidimpuan	99.64	84.38	77.46	29.64
33	Gunungsitoli	98.83	82.87	75.07	19.75

source: BPS-Sumut

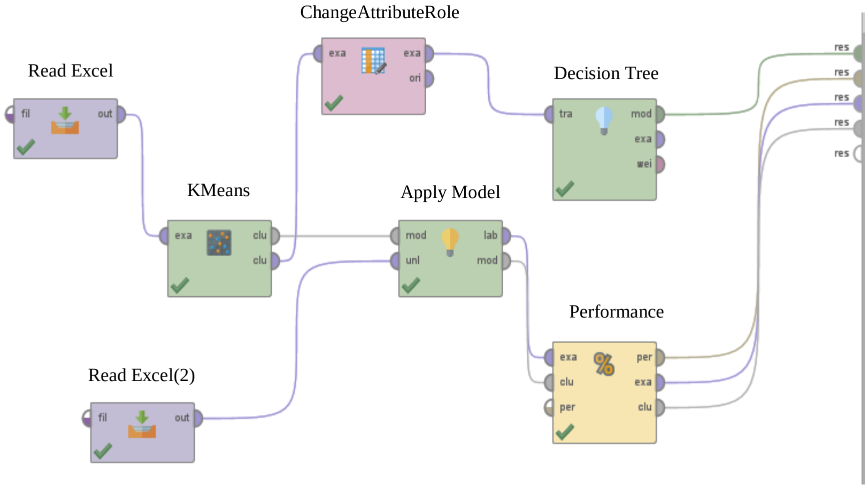


Fig. 1 The RapidMiner model on APM is based on districts/cities

Cluster Model

Cluster 0: 18 items

Cluster 1: 15 items

Total number of items: 33

Fig. 2 Results of clustering with k-means

Fig. 3 The high cluster (C1)

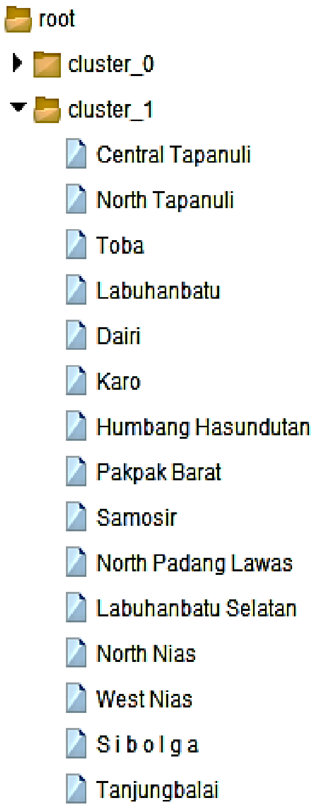
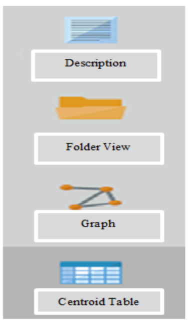


Fig. 4 The low cluster (C2)

Table 2. The results of the Rapid Miner export file to excel

Regency / City	Elementary school	Middle School	High school	PT	cluster
N i a s	98.8	78.1	62.5	5.9	hc_0
Mandailing Natal	99.2	83.1	62.6	13.1	hc_0
South Tapanuli	98.6	82.1	66.9	15.3	hc_0
Central Tapanuli	99.1	88.3	70.7	15.6	lc_1
North Tapanuli	99.4	88.2	78.2	17.7	lc_1
Toba	98.5	89.2	82.1	2.6	lc_1
Labuhanbatu	99.4	86.9	68.2	10.8	lc_1
A s a h a n	99.8	81.9	60.9	15.2	hc_0
Simalungun	98.6	77.5	63.7	20.1	hc_0
Dairi	99.4	90.5	80.7	9.5	lc_1
Karo	98.6	83.2	73.2	10.5	lc_1
Deli Serdang	95.0	70.8	67.8	19.8	hc_0
Langkat	98.9	78.9	64.4	13.1	hc_0
South Nias	95.9	70.7	66.7	10.9	hc_0
Humbang Hasundutan	99.3	92.9	86.5	10.3	lc_1
Pakpak Barat	99.1	88.0	80.0	8.3	lc_1
Samosir	99.6	91.6	81.8	3.6	lc_1
Serdang Bedagai	99.1	77.7	68.0	9.7	hc_0
Batu Bara	99.2	74.4	60.2	10.9	hc_0
North Padang Lawas	98.8	83.3	69.2	7.1	lc_1
Padang Lawas Utara	99.0	82.8	62.6	10.9	hc_0
Labuhanbatu Selatan	98.8	84.4	71.8	11.0	lc_1
Labuhanbatu Utara	99.8	74.6	64.3	12.5	hc_0
North Nias	98.6	80.8	74.0	7.8	lc_1
West Nias	99.5	82.6	78.9	5.9	lc_1
S i b o l g a	99.1	87.9	74.3	9.2	lc_1
Tanjungbalai	98.4	81.8	71.8	10.0	lc_1
Pematangsiantar	99.6	81.6	75.8	23.1	hc_0
Tebing Tinggi	98.0	82.6	67.2	11.8	hc_0
M e d a n	93.5	80.0	61.4	33.5	hc_0
B i n j a i	99.3	83.4	72.6	20.4	hc_0
Padangsidimpuan	99.6	84.4	77.5	29.6	hc_0
Gunungsitoli	98.8	82.9	75.1	19.8	hc_0

Here are the high-end cluster centroids (hc_0) and low cluster (lc_1) as in the following image:



Attribute	Cluster_0	Cluster_1
Elementary School	98.378	99.030
Middle School	79.301	86.644
High School	66.677	76.076
PT	16.420	9.337

Fig. 5 The final centroid results

The mapping results using clustering (k-means) will be input and processed using classification (C4.5) with the help of RapiMiner software. Here are the results of the classification of Pure Participation Rates: From Figure 7, it can be explained that the rules resulting from the decision tree are three rules, namely if the SMA has a percentage <68,085% (high cluster); if SMA has a percentage> 68,085% and PT has a presentation <18,730% (low cluster) and if SMA has a percentage> 68,085%. PT has a presentation of> 18,730% (high cluster). Following are the full results of the decision tree. In Figure 8, the validity test is performed with the Davies-Bouldin Index (DBI). Using DBI, the cluster results obtained are optimal with the number of clusters (k = 2) 1,148. So that in the process of mapping in the form of clusters of pure participant numbers by district/city in North Sumatra, there are two clusters.

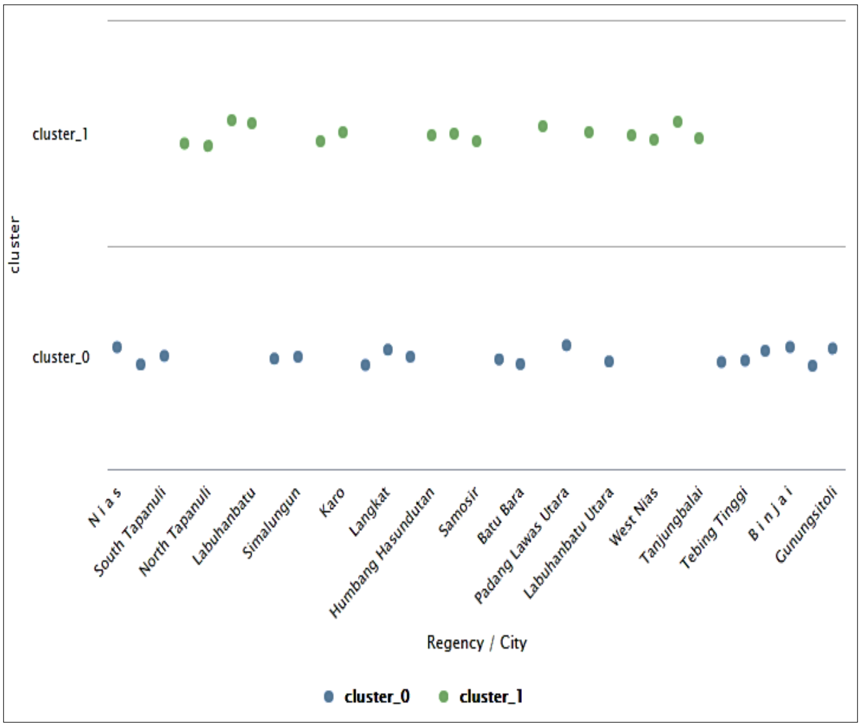


Fig. 6 The present study employs a scatter plotter to visualize the outcomes of clustering analysis

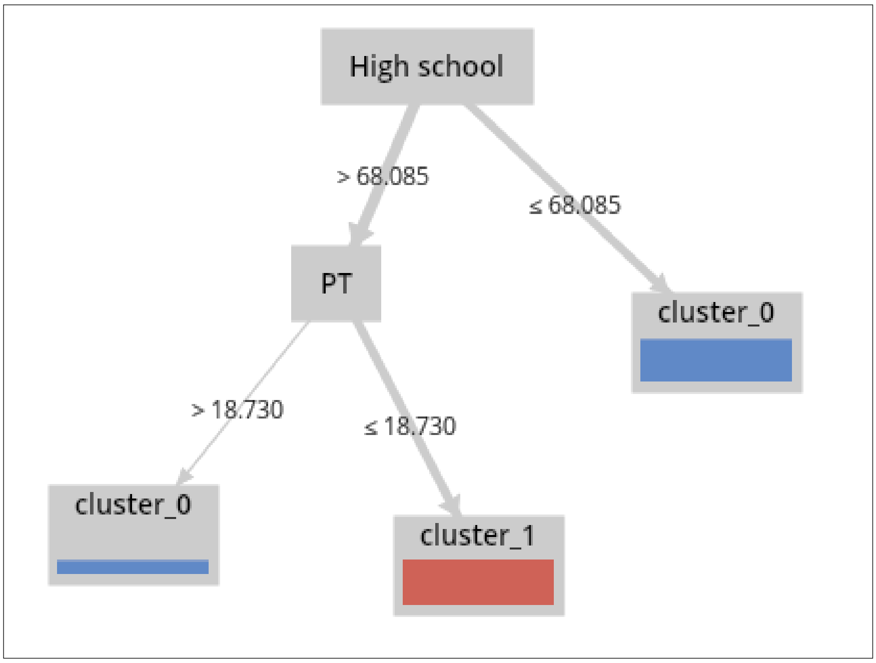


Fig. 7 Decision tree results from pure participation rates

Tree
High school > 68.085
| PT > 18.730: hc_0 {hc_0=4, lc_1=0}
| PT ≤ 18.730: lc_1 {hc_0=0, lc_1=15}
High school ≤ 68.085: hc_0 {hc_0=14, lc_1=0}

Performance Vector
Performance Vector:
Avg. within centroid distance: -77.396
Avg. within centroid distance_cluster_0: -95.607
Avg. within centroid distance_cluster_1: -55.542
Davies Bouldin: -1.148

Fig. 8 Performance vector results

5. Conclusion

Based on the results of the study, it can be realized that the combination of the clustering method (k-means) and classification (C4.5) can be applied to the case of pure participation rates with the assessment attribute is the percentage of elementary and junior high, high school and university-based on districts/cities in North Sumatra. The clustering results obtained 18 districts/cities in the high cluster (hc_0) and 15 districts/cities in the low cluster (lc_1). The classification results were obtained by three rules where SMA and PT become attributes that influence the pure participation rate (APM) in North Sumatra. The higher the percentage of the APM, the better the quality of human resources.

References

[1] Bambang Raditya Purnomoet et al., "Education and Socialization of Mentally Challenged People and Similarities and Differences," *International Journal of Psychosocial Rehabilitation*, vol. 24, no. 8, pp. 1524–1534, 2020. [Google Scholar] [Publisher Link]

[2] Mohzana Mohzana et al., "A Management Model for Character Education in Higher Education," *Talent Development & Excellence*, vol. 12, no. 3, pp. 1596–1601, 2020. [Google Scholar] [Publisher Link]

[3] Harisa Mardiana, "Lecturers' Reasoning in using Digital Technology: A Cognitive Approach in Learning Process," *Journal of Social, Culture and Society*, vol. 1, no. 2, pp. 33–42, 2023. [CrossRef] [Google Scholar] [Publisher Link]

[4] wim winowatan, and Jeanny Pricilia Anneke W, "Effectiveness of Teaching and Learning Process Based on Competency Curriculum and Influence on Student Skills Activation: Study on the Makassar Tourism Polytechnic Students," *Athena: Journal of Social, Culture and Society*, vol. 1, no. 2, pp. 67-74, 2023. [CrossRef] [Publisher Link]

[5] Muh. Fahrurrozi, and Mohzana, "The Development of Android-Based Economic Teaching Materials for Student Independence," *International Journal of Innovation, Creativity and Change*, vol. 5, no. 6, pp. 468–482, 2019. [Google Scholar] [Publisher Link]

[6] Zainul Arifin, and Mohamada Zaky Tatsar, "Application of Quizzz to Pascal Law Material in Increasing Student Response," *SAGA: Journal of Technology and Information Systems*, vol. 1, no. 1, pp. 9–11, 2023. [CrossRef] [Google Scholar] [Publisher Link]

[7] B. Supriyadi et al., "Classification of Natural Disaster-Prone Areas in Indonesia using K-Means," *International Journal of Grid and Distributed Computing*, vol. 11, no. 8, pp. 87–98, 2018. [CrossRef] [Google Scholar] [Publisher Link]

[8] Fathur Rahman et al., "Application of Data Mining Technique using K-Medoids in the Case of Export of Crude Petroleum Materials to the Destination Country," *IOP Conference Series: Materials Science and Engineering*, vol. 835, p. 012058, 2020. [CrossRef] [Google Scholar] [Publisher Link]

[9] Khairunnissa Fanny Irnanda et al., "The Selection Of Calcium Milk Products that are Appropriate for Advanced Age using Promethee II Algorithm," *Journal of Physics: Conference Series*, vol. 1381, no. 1, p. 012070, 2019. [CrossRef] [Google Scholar] [Publisher Link]

[10] Anjar Wanto *et al.*, "Model of Artificial Neural Networks in Predictions of Corn Productivity in an Effort to Overcome Imports in Indonesia," *Journal of Physics: Conference Series*, vol. 1339, no. 1, p. 012057, 2019. [CrossRef] [Google Scholar] [Publisher Link]

[11] Sudirman, Agus Perdana Windarto, and Anjar Wanto, "Data Mining Tools | Rapidminer: K-Means Method on Clustering of Rice Crops by Province as Efforts to Stabilize Food Crops in Indonesia," *IOP Conference Series: Materials Science and Engineering*, vol. 420, p. 012089, 2018. [CrossRef] [Google Scholar] [Publisher Link]

[12] Agus Perdana Windarto, "Implementation of Data Mining on Rice Imports by Major Country of Origin using Algorithm Using K-Means Clustering Method," *International Journal Of Artificial Intelligence Research*, vol. 1, no. 2, pp. 26–33, 2017. [CrossRef] [Google Scholar] [Publisher Link]

[13] Insanul Kamila, Ulya Khairunnisa, and Mustakim Mustakim, "Perbandingan Algoritma K-Means Dan K-Medoids Untuk Pengelompokan Data Transaksi Bongkar Muat di Provinsi Riau," *Jurnal Ilmiah Rekayasa dan Manajemen Sistem Informasi*, vol. 5, no. 1, pp. 119–125, 2019. [CrossRef] [Google Scholar] [Publisher Link]

[14] Putrama Alkhairi, and Agus Perdana Windarto, "Penerapan K-Means Cluster Pada Daerah Potensi Pertanian Karet Produktif di Sumatera Utara," *Seminar Nasional Teknologi Komputer & Sains (SAINTEKS)*, pp. 762–767, 2019. [Google Scholar] [Publisher Link]

[15] Yosi Pahala et al., "The Influence of Load Factor, Headway, and Travel Time on Total Fleet Requirements and Its Implications for Public Transportation Maintenance Management on Transjakarta," *Review of International Geographical Education*, vol. 11, no. 5, pp. 3422–3436, 2021. [Google Scholar] [Publisher Link]

[16] Raden Didiet Rachmat Hidayat et al., "Study of the Formation of National Logistics Cluster for Disaster Management (KLASNASLOG PB) by National Disaster Management Authority (BNPB) to Streamline Transport for Disaster Management in Indonesia," *MATEC Web Conference*, vol. 229, 2018. [CrossRef] [Google Scholar] [Publisher Link]

[17] Meilin Widyastuti et al., "Classification Model C.45 on Determining the Quality of Custumer Service in Bank BTN Pematangsiantar Branch," *Journal of Physics: Conference Series*, vol. 1255, no. 012002, pp. 1–6, 2019. [CrossRef] [Google Scholar] [Publisher Link]

[18] Wiwiek Katrina et al., "C.45 Classification Rules Model for Determining Students Level of Understanding of the Subject," *Journal of Physics: Conference Series*, vol. 1255, no. 012005, pp. 1–7, 2019. [CrossRef] [Google Scholar] [Publisher Link]

[19] Zohra Bellahsene, Angela Bonifati, and Erhard Rahm, *Schema Matching and Mapping*, 2011. [CrossRef] [Publisher Link]

[20] Amit P. Sheth, and James A. Larson, "Federated Database Systems for Managing Distributed, Heterogeneous, and Autonomous Databases," *ACM Computing Surveys*, vol. 22, no. 3, pp. 183–236, 1990. [CrossRef] [Google Scholar] [Publisher Link]

[21] AguguoIhechukwu.C, Matthias Daniel, and E.O Bennett, "Big Data Mining for Interesting Pattern using MapReduced Technique," *SSRG International Journal of Computer Science and Engineering*, vol. 7, no. 7, pp. 26-33, 2020. [CrossRef] [Publisher Link]

[22] Gunisetti Tirupathi Rao, and Rajendra Gupta, "An approach of Clustering and Analysis of Unstructured Data," *SSRG International Journal of Computer Science and Engineering*, vol. 6, no. 11, pp. 64-69, 2019. [CrossRef] [Publisher Link]

[23] Anil K. Jain, "Data Clustering: 50 Years Beyond K-Means," *Pattern Recognition Letters*, vol. 31, no. 8, pp. 651–666, 2010. [CrossRef] [Google Scholar] [Publisher Link]

[24] Rui Xu, and D. Wunsch, "Survey of Clustering Algorithms," *IEEE Transactions on Neural Networks*, vol. 16, no. 3, pp. 645–678, 2005. [CrossRef] [Google Scholar] [Publisher Link]

[25] Steven L. Salzberg, "C4.5: Programs for Machine Learning by J. Ross Quinlan. Morgan Kaufmann Publishers, Inc., 1993," *Machine Learning*, vol. 16, no. 3, pp. 235–240, 1994. [CrossRef] [Google Scholar] [Publisher Link]

[26] S. B. Kotsiantis, "Decision Trees: A Recent Overview," *Artificial Intelligence Review*, vol. 39, no. 4, pp. 261–283, 2013. [CrossRef] [Google Scholar] [Publisher Link]

[27] Mohammed M Mazid, A B M Shawkat Ali, and Kevin S Tickle, "Improved C 4. 5 Algorithm for Rule Based Classification," *Recent Advances in Artificial Intelligence, Knowledge Engineering and Data Bases*, pp. 296-301, 2010. [Google Scholar] [Publisher Link]

[28] Camilla Schaefer, and Ana Makatsaria, "Framework of Data Analytics and Integrating Knowledge Management," *International Journal of Intelligent Networks*, vol. 2, pp. 156–165, 2021. [CrossRef] [Google Scholar] [Publisher Link]

[29] Abaka B.H. Mobio et al., "Mapping Spatial and Temporal Variability of Rainfall in Côte D'Ivoire using TRMM Data," *SSRG International Journal of Geoinformatics and Geological Science*, vol. 4, no. 3, pp. 13-20, 2017. [CrossRef] [Google Scholar] [Publisher Link]

[30] Heny Pratiwi et al., "Sigmoid Activation Function in Selecting the Best Model of Artificial Neural Networks," *Journal of Physics: Conference Series*, vol. 1471, no. 1, p. 012010, 2020. [CrossRef] [Google Scholar] [Publisher Link]

[31] Nurul Rofiqo, Agus Perdana Windarto, and Dedy Hartama, "Penerapan Clustering Pada Penduduk Yang Mempunyai Keluhan Kesehatan Dengan Datamining K-Means," *Konferensi Nasional Teknologi Informasi dan Komputer (KOMIK)*, vol. 2, no. 1, pp. 216–223, 2018. [CrossRef] [Google Scholar] [Publisher Link]

[32] Wisdalia Maya Sariet al., "Improving the Quality of Management with the Concept of Decision Support Systems in Determining Factors for Choosing a Cafe based on Consumers," *Journal of Physics: Conference Series*, vol. 1471, no. 1, p. 012009, 2020. [CrossRef] [Google Scholar] [Publisher Link]

[33] Agus Perdana Windarto et al., "Analysis of the K-Means Algorithm on Clean Water Customers Based on the Province," *Journal of Physics: Conference Series*, vol. 1255, no. 1, p. 012001, 2019. [CrossRef] [Google Scholar] [Publisher Link]

[34] Mhd Gading Sadewo, Agus Perdana Windarto, and Dedy Hartama, "Penerapan Datamining Pada Populasi Daging Ayam Ras Pedaging Di Indonesia Berdasarkan Provinsi Menggunakan K-Means Clustering," *InfoTekJar : Jurnal Nasional Informatika dan Teknologi Jaringan*,vol. 2, no. 1, pp. 60–67, 2017. [Google Scholar] [Publisher Link]

[35] Agus Perdana Windarto, "Penerapan Data Mining Pada Ekspor Buah-Buahan Menurut Negara Tujuan Menggunakan K-Means Clustering Method," *Jurnal Techno.Com*, vol. 16, no. 4, pp. 348–357, 2017. [Google Scholar] [Publisher Link]

[36] Riyani Wulan Sari, Anjar Wanto, and Agus Perdana Windarto, "Implementasi Rapidminer Dengan Metode K-Means (Study Kasus: Imunisasi Campak Pada Balita Berdasarkan Provinsi)," *KOMIK (Konferensi Nasional Teknologi Informasi dan Komputer)*, vol. 2, no. 1, pp. 224–230, 2018. [CrossRef] [Google Scholar] [Publisher Link]

[37] Trupti M Kodinariya, and Prashant Makwana, "Review on Determining Number of Cluster in K-Means Clustering," *International Journal of Advance Research in Computer Science and Management Studies*, vol. 1, no. 6, pp. 90-95, 2013. [Google Scholar] [Publisher Link]

[38] Dini Rizky Sitorus P et al., "Penerapan Klasifikasi C4.5 Dalam Meningkatkan Sistem Pembelajaran Mahasiswa," *KOMIK (Konferensi Nasional Teknologi Informasi dan Komputer)*, vol. 3, pp. 593–597, 2020. [CrossRef] [Google Scholar] [Publisher Link]

[39] Anjar Wanto et al., *Data Mining : Algoritma dan Implementasi*, 2020. [Publisher Link]

[40] Ronanki Umarao, and Behara Vineela, "An Optimized Fuzzy Means Clustering Algorithm for Grouping of Social Media Data," *SSRG International Journal of Computer Science and Engineering*, vol. 4, no. 5, pp. 1-4, 2017. [CrossRef] [Publisher Link]