

## Penerapan Algoritma K-Means Dalam Pengklasteran Hasil Evaluasi Akademik Mahasiswa

Fitri Safnita<sup>1</sup>, Sarjon Defit<sup>2</sup>, Gunadi Widi Nurcahyo<sup>3</sup>

<sup>1,2,3</sup>Fakultas Ilmu Komputer, Universitas Putra Indonesia “YPTK” Padang, Indonesia

E-mail: <sup>1</sup>fitrisafnita01@gmail.com, <sup>2</sup>sarjond@yahoo.co.uk, <sup>3</sup>gunadiwidi@yahoo.co.id

### Abstract

Several institutions that have utilized computer-based information systems for many years certainly have quite large amounts of data. The data generated and stored in a computer system is designed to be fast and accurate in both operation and administration. This data is designed for reporting and analysis that uses that data. It turns out that there is a lot of data available, with so much data we are increasingly faced with the question, "What knowledge can we gain from this data?" The K-Means algorithm is an iterative clustering algorithm that partitions a data set into a number of clusters that are initially determined. The K-Means algorithm is an iterative clustering algorithm that partitions a data set into a number of clusters that are initially determined. The K-Means algorithm is easy to implement and run, relatively fast, easy to adapt, commonly used in practice. The parameter that must be entered when using the K-Means algorithm is the K value. The K value is generally used based on previously known information regarding how many clusters appear in This research aims to group students based on academic evaluation results. The method used to manage student academic data uses the Data Mining method with the K-Means Clustering Algorithm. The dataset processed in this research comes from the Faculty of Engineering, Informatics Engineering Study Program, Islamic University of Riau. The dataset consists of 180 student data starting from semester 1 to semester 4. The results obtained from this research are in the form of grouping students based on the achievement student cluster, there are 104 students with a percentage of 57.72%, the student cluster with potential for achievement is 62 students with a percentage of 34.41%, the potentially problematic student cluster has 10 students with a percentage of 5.55%, and the problematic student cluster has 4 students with a percentage of 2.22%. Therefore, it is hoped that the results of this research will provide new knowledge that can be used as a source of information and function as a reference model for academic planners to monitor and predict the development of each student's academic performance.

**Keywords:** Clustering, K-Means, Evaluation, RapidMiner, Data Mining

### Abstrak

Beberapa institusi yang memanfaatkan sistem informasi berbasis komputer selama bertahun-tahun sudah pasti memiliki jumlah data yang cukup besar pula. Data yang dihasilkan dan disimpan dalam sistem komputer dirancang agar cepat dan akurat baik dalam mengoperasikan maupun administrasinya. Data ini dirancang untuk pelaporan dan analisis yang menggunakan data tersebut. Data yang tersedia ternyata sangat banyak, dengan banyaknya data kita semakin dihadapkan pada pertanyaan, "Pengetahuan apa yang bisa kita peroleh dari data tersebut?". Algoritma K-Means adalah algoritma pengelompokan iterative yang melakukan partisi set data ke dalam sejumlah cluster yang diawal sudah ditetapkan. Algoritma K-Means adalah algoritma pengelompokan iterative yang melakukan partisi set data ke dalam sejumlah cluster yang diawal sudah ditetapkan. Algoritma K-Means mudah diimplementasikan dan dijalankan, relatif cepat, mudah beradaptasi, umum penggunaannya dalam praktek. Parameter yang

harus dimasukkan ketika menggunakan algoritma *K-Means* adalah nilai *K*. Nilai *K* umumnya digunakan berdasarkan informasi yang diketahui sebelumnya mengenai berapa banyak *cluster* yang muncul dalam *X*, berapa banyak yang digunakan untuk penerapannya atau jenis *cluster* dicari dengan melakukan percobaan dengan beberapa nilai *K*. Penelitian ini bertujuan untuk mengelompokkan mahasiswa berdasarkan hasil evaluasi akademik. Metode yang digunakan dalam mengelola data akademik mahasiswa menggunakan metode *Data Mining* dengan Algoritma *K-Means Clustering*. Dataset yang diolah dalam penelitian ini bersumber dari Fakultas Teknik Program Studi Teknik Informatika, Universitas Islam Riau. Dataset terdiri dari 180 data mahasiswa mulai dari semester 1 hingga mahasiswa semester 4. Hasil yang diperoleh dari penelitian ini berupa pengelompokan mahasiswa berdasarkan *cluster* mahasiswa berprestasi terdapat 104 orang mahasiswa dengan persentase 57,72%, *cluster* mahasiswa berpotensi berprestasi terdapat 62 orang mahasiswa dengan persentase 34,41%, *cluster* mahasiswa berpotensi bermasalah terdapat 10 orang mahasiswa dengan persentase 5,55%, dan *cluster* mahasiswa bermasalah terdapat 4 orang mahasiswa dengan persentase 2,22%. Oleh karena itu, diharapkan hasil penelitian ini memberikan pengetahuan baru yang bisa dijadikan sumber informasi serta berfungsi sebagai model acuan bagi perencana akademik untuk memonitor dan memprediksi perkembangan performa akademik setiap mahasiswa.

**Kata Kunci:** *Clustering, K-Means, Evaluasi, RapidMiner, Data Mining*

## 1. Pendahuluan

Algoritma *K-Means*, sebagai salah satu algoritma pengelompokan (*clustering*) yang populer, telah menjadi instrumen yang efektif dalam analisis data dan eksplorasi pola. Algoritma ini tergolong dalam keluarga algoritma *unsupervised learning* yang bertujuan untuk mengelompokkan data menjadi beberapa *cluster* berdasarkan kesamaan karakteristik. Kemudahan implementasi dan kinerja yang relatif cepat menjadikan *K-Means* sebagai pilihan yang sangat diandalkan dalam berbagai bidang, mulai dari ilmu data hingga aplikasi praktis di dunia industri. Algoritma *K-Means* berfungsi dengan cara yang sederhana namun efektif. Langkah pertama melibatkan inisialisasi sejumlah *centroid* awal, yang kemudian diatur ulang secara iteratif untuk membentuk *cluster-cluster* yang sesuai. Proses ini melibatkan dua tahap utama: pengelompokan data ke dalam *cluster* yang sesuai dengan *centroid* terdekat dan perhitungan ulang posisi *centroid* berdasarkan data yang termasuk dalam setiap *cluster*. Iterasi dilanjutkan hingga konvergensi, di mana tidak ada perubahan signifikan dalam pengelompokan data.

Hasil penelitian dengan menggunakan perhitungan metode *K-Means Clustering* dapat menghasilkan prediksi kelulusan mahasiswa dengan lulus tepat waktu dan lulus terlambat[1]. Secara keseluruhan, optimasi berbasis rata-rata dari algoritma *K-Means* dinamis meningkatkan kualitas rata-rata *cluster* dibandingkan dengan algoritma *K-Means* lainnya ketika mengevaluasi PC, SSE, dan varians *cluster* dan setiap jumlah iterasi pengujiannya adalah konstan. Hasil pengujian dapat digunakan sebagai salah satu metode terbaik tidak hanya untuk mengevaluasi kinerja akademik mahasiswa, tetapi juga untuk pengambilan keputusan dalam menetapkan kebijakan akademik universitas[2]. Algoritma *K-Means* adalah algoritma pengelompokan berulang yang melakukan partisi set data menjadi kumpulan kluster tertentu[3]. Teknik *Clustering Data Mining* seperti *K-Means* merupakan salah satu algoritma yang banyak diterapkan oleh peneliti di antaranya menggunakan algoritma *clustering* pengelompokan data, pemetaan data, klasifikasi data, dan sebagainya[4]. Tahapan dalam implementasi algoritma *K-Means* terdiri dari penentuan jumlah *cluster*, selanjutnya menentukan pusat awal *cluster* dengan mengambil data secara acak[5].

Hasil penelitian yang didapat dari 250 data pasien ISPA berhasil dikumpulkan dan diproses menggunakan software *Rapid Miner* dan algoritma *K-Means* digunakan untuk mengelompokkan gejala ISPA secara manual[6]. Berdasarkan perhitungan *Sum of Squares* diperoleh jumlah *cluster* optimal dengan  $k=2$ . Hasil *clustering* menghasilkan *cluster* 1 dengan nilai *centroid* sebanyak 2.686 dan jumlah anggota sebanyak 15 orang dan *cluster* 2 dengan nilai *centroid* sebanyak 3.245 dan jumlah anggota sebanyak 40 orang. Dari penelitian ini dapat disimpulkan bahwa anggota *cluster* 2 merupakan kelompok mata kuliah yang mempunyai nilai rata-rata tinggi[7]. Teknologi analisis *clustering K-Means* digunakan untuk menganalisis data evaluasi siswa secara komprehensif dalam makalah ini, dengan tujuan menentukan keteraturan implikasi data, mendiagnosis masalah pembelajaran secara akurat, dan memberikan landasan untuk mengembangkan strategi pengelolaan siswa yang efektif[8]. Hasil penelitian ini menunjukkan tingkat penjualan produk didominasi oleh produk-produk yang masuk kedalam kategori kurang laris, dari pengetahuan tersebut pihak perusahaan dapat mengambil keputusan terkait strategi penjualan maupun inovasi pada produk-produk yang kurang laris dengan harapan dapat meningkatkan penjualan[9]. Dengan menerapkan algoritma *K-Means Clustering*, dapat diusulkan sebagai pendekatan yang dapat digunakan untuk memprediksi pemilihan siswa berprestasi[10]. Penggunaan properti *Euclidean* dari fungsi tujuan metode ini membuatnya lebih cepat dibandingkan teknik pengelompokan pohon yang ada, sehingga lebih cocok untuk menganalisis kumpulan data evolusioner yang besar[11]. *K-Means* merupakan metode relokasi yang paling banyak digunakan, dimana pusat *cluster* awal dipilih secara acak dari  $k$  titik sampel, berdasarkan jumlah *cluster* ( $k$ ) yang diinginkan dalam solusi akhir[12].

Tujuan dari penelitian ini adalah untuk menganalisa data akademik mahasiswa, membangun pola pengelompokan menggunakan metode *K-Means Clustering* sehingga dapat menghasilkan pengetahuan atau *knowledge*. Dan menguji data yang sudah diolah dengan menggunakan *tools Rapid Miner*.

## 2. Metodologi Penelitian

Metodologi penelitian merupakan suatu cara atau teknik untuk mendapatkan informasi dan sumber data yang akan digunakan dalam penelitian. Untuk memperoleh data primer, peneliti dapat mengumpulkan data secara langsung. Teknik yang digunakan peneliti untuk mengumpulkan data primer adalah dengan melakukan observasi dan wawancara. Data yang digunakan untuk penelitian ini adalah data akademik mahasiswa dengan jumlah data sebanyak 180 data mahasiswa.



**Gambar 1.** Proses Metode *K-means Clustering*

Tahapan proses perhitungan data akademik mahasiswa dengan algoritma *K-Means* sebagai berikut:

### 2.1. Menentukan Jumlah $k$

Untuk menentukan berapa banyak  $cluster$   $k$  dilakukan dengan beberapa pertimbangan seperti pertimbangan teoritis dan konseptual yang mungkin diusulkan untuk menentukan beberapa banyak  $cluster$ .

### 2.2. Menentukan Pusat $Cluster$ Awal ( $Centroid$ )

Untuk menentukan  $centroid$  awal dilakukan secara random / acak dari objek-objek yang tersedia sebanyak  $k$   $cluster$ , kemudian untuk menghitung  $centroid$   $cluster$  ke-1 berikutnya, digunakan rumus sebagai berikut :

$$v = \frac{\sum_{i=1}^n xi}{n} ; i = 1,2,3,...n \tag{1}$$

Di mana  $\Leftrightarrow$   $v$  :  $centroid$  pada  $cluster$ .  
 $xi$  : objek ke- $i$ .  
 $n$  : banyaknya objek/jumlah objek yang menjadi anggota  $cluster$ .

### 2.3. Menghitung Jarak Dengan Pusat $Cluster$

Untuk menghitung jarak antara objek dengan  $centroid$  dapat menggunakan  $Euclidian Distance$ :

$$d(x, y) = ||x - y|| = \sqrt{\sum_{i=1}^n (xi - yi)^2} ; I = 1,2,3,...n \tag{2}$$

Di mana  $\Leftrightarrow$   $xi$  : objek  $x$  ke- $i$ .  
 $yi$  : daya  $y$  ke- $i$ .  
 $n$  : banyaknya objek.

### 2.4. Pengelompokan Data

Jarak yang dihitung dibandingkan dan jarak terdekat antara data dan pusat  $cluster$  dipilih. Jarak ini menunjukkan bahwa data berada dalam kelompok yang sama dengan pusat cluster terdekat.

### 2.5. Menentukan Pusat $Cluster$ Baru

Untuk memperoleh pusat  $cluster$  baru dapat dihitung dari nilai rata-rata anggota  $cluster$  dan pusat  $cluster$ . Jika hasil yang diperoleh tidak sama, maka digunakan pusat  $cluster$  baru untuk melakukan iterasi berikutnya. Proses iterasi berhenti ketika jumlah maksimum iterasi yang dimasukkan oleh user tercapai atau ketika hasil yang diperoleh sudah sama (pusat  $cluster$  baru sama dengan pusat  $cluster$  lama).

## 3. Hasil dan Pembahasan

Data merupakan informasi yang diperoleh melalui pengamatan atau pencarian dari beberapa sumber. Sumber data penelitian ini diperoleh dari Program Studi Teknik Informatika Universitas Islam Riau sebanyak 180 data dan dapat dilihat pada Tabel 1.

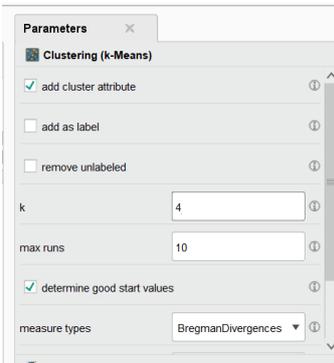
**Tabel 1.** Data Mahasiswa

No	Mahasiswa	Indeks Prestasi				Rata-rata IPK
		Ipk 1	Ipk 2	Ipk 3	Ipk 4	
1	Mahasiswa 1	3,48	3,30	3,48	3,00	3,315
2	Mahasiswa 2	3,85	3,62	4,00	3,79	3,815
3	Mahasiswa 3	3,85	3,83	3,98	3,79	3,863
4	Mahasiswa 4	3,45	3,30	3,10	3,61	3,365
5	Mahasiswa 5	1,75	3,25	2,96	3,37	2,833
6	Mahasiswa 6	3,33	3,30	3,42	3,71	3,440
7	Mahasiswa 7	3,92	4,00	4,00	3,96	3,970
8	Mahasiswa 8	3,55	3,65	3,43	3,46	3,523
9	Mahasiswa 9	3,55	3,67	3,67	3,46	3,588

No	Mahasiswa	Indeks Prestasi			Rata-rata
10	Mahasiswa 10	3,77	3,75	3,91	3,82
....	....	....	....	....	....
180	Mahasiswa 180	3,42	3,38	3,55	3,08

### 3.1. Menentukan nilai K

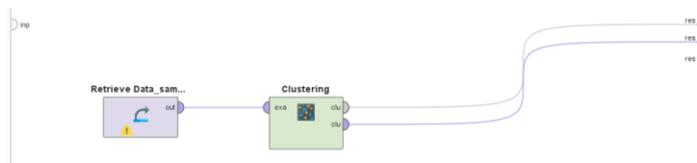
Dalam algoritma *K-Means*, nilai *k* ditentukan berdasarkan kebutuhan yang akan menjadi acuan untuk mendapatkan hasil dalam *cluster*. Dalam penelitian ini nilai *k* yang digunakan adalah  $k=4$ .



Gambar 2. Penentuan Jumlah Cluster

### 3.2. Centroid Data

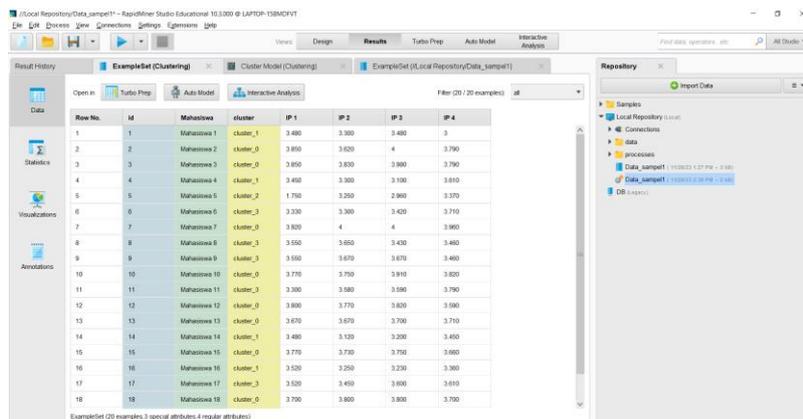
Menemukan *centroid* data merupakan prasyarat untuk melakukan *clustering* dengan menggunakan algoritma *K-Means*. Pengujian 180 data menggunakan *RapidMiner* dapat dilihat pada Gambar 3 berikut:



Gambar 3. Implementasi Data Dengan RapidMiner

### 3.3. Pengelompokan Data

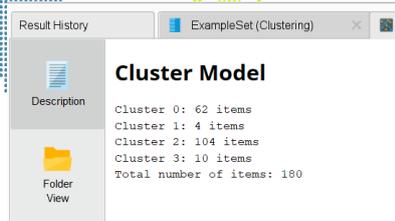
Setelah garis *res* dan *clu* dihubungkan seperti Gambar 4, maka jalankan program dengan cara klik ikon atau *Run* pada bagian atas layar untuk mengetahui hasil *cluster* yang dilakukan pada *RapidMiner*.



Row No	Id	Mahasiswa	cluster	IP1	IP2	IP3	IP4
1	1	Mahasiswa 1	cluster_1	3,480	3,300	3,480	3
2	2	Mahasiswa 2	cluster_0	3,650	3,620	4	3,790
3	3	Mahasiswa 3	cluster_0	3,650	3,620	3,960	3,790
4	4	Mahasiswa 4	cluster_1	3,480	3,300	3,100	3,610
5	5	Mahasiswa 5	cluster_2	1,750	3,250	2,960	3,370
6	6	Mahasiswa 6	cluster_3	3,320	3,300	3,420	3,710
7	7	Mahasiswa 7	cluster_0	3,620	4	4	3,960
8	8	Mahasiswa 8	cluster_3	3,550	3,650	3,420	3,460
9	9	Mahasiswa 9	cluster_3	3,550	3,670	3,670	3,460
10	10	Mahasiswa 10	cluster_0	3,770	3,710	3,810	3,820
11	11	Mahasiswa 11	cluster_3	3,360	3,360	3,590	3,790
12	12	Mahasiswa 12	cluster_0	3,860	3,770	3,820	3,580
13	13	Mahasiswa 13	cluster_0	2,670	3,670	3,700	3,710
14	14	Mahasiswa 14	cluster_1	3,480	3,120	3,200	3,450
15	15	Mahasiswa 15	cluster_0	3,770	3,720	3,750	3,660
16	16	Mahasiswa 16	cluster_1	3,520	3,250	3,220	3,360
17	17	Mahasiswa 17	cluster_3	3,520	3,450	3,600	3,610
18	18	Mahasiswa 18	cluster_0	3,790	3,800	3,800	3,700

Gambar 4. Tampilan Hasil Cluster (Data View)

Selanjutnya grafik hasil pengelompokan atau *cluster* semua data dapat dilihat pada *Plot View* seperti Gambar 5.



**Gambar 5.** Tampilan Hasil *Cluster* (Text View 180 Data)

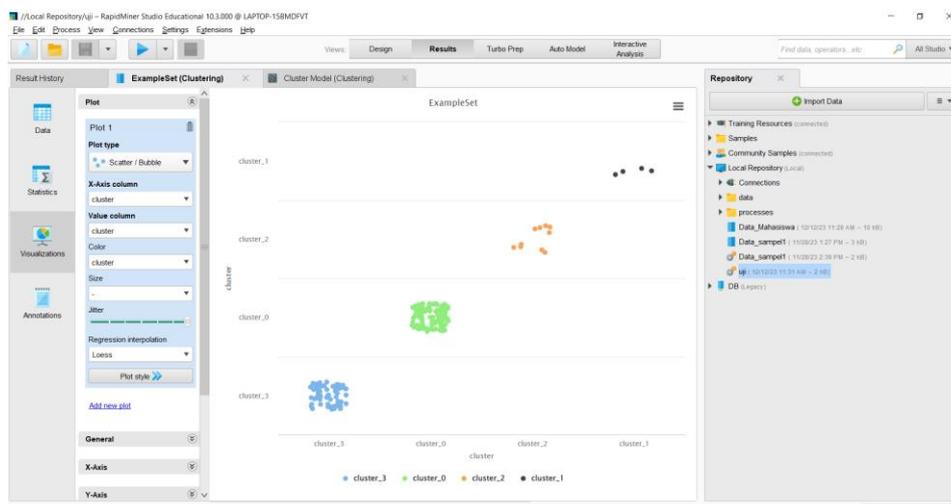
Berdasarkan Gambar 5, dapat dilihat jumlah anggota dari masing-masing *cluster* yaitu *cluster* 0 memiliki anggota sebanyak 62 orang, *cluster* 1 memiliki anggota sebanyak 4 orang, *cluster* 2 memiliki anggota sebanyak 104 orang, dan *cluster* 3 memiliki anggota sebanyak 10 orang. Selanjutnya *centroid* tabel yang merupakan tampilan hasil perhitungan centroid akhir *K-Means*. Tampilan *centroid* dapat dilihat pada Gambar 6.



Attribute	cluster_0	cluster_1	cluster_2	cluster_3
ipsemester1	3.368	1.942	3.679	2.290
ipsemester2	3.143	1.797	3.688	2.506
ipsemester3	3.251	1.238	3.728	2.655
ipsemester4	3.296	2.055	3.659	2.767

**Gambar 6.** Tampilan Hasil *Cluster* (Centroid Table 180 Data)

Berdasarkan Gambar 6, dapat dilihat bahwa *cluster* 0 merupakan kelompok mahasiswa berpotensi berprestasi, *cluster* 1 merupakan kelompok mahasiswa bermasalah, *cluster* 2 merupakan kelompok mahasiswa berprestasi, dan *cluster* 3 merupakan kelompok mahasiswa berpotensi bermasalah. Selanjutnya grafik hasil pengelompokan atau *cluster* semua data dapat dilihat pada *Plot View* seperti Gambar 7.



**Gambar 7.** Tampilan Hasil *Cluster* (Plot View)

Penelitian ini menggunakan *Tools RapidMiner* untuk mengcluster total 180 *dataset*, yang terbagi menjadi 160 data *training* dan 20 data *testing*. Dari data yang diuji, didapatkan 4 *cluster* dengan hasil sebagai berikut:

- a) *Cluster 0* merupakan kelompok mahasiswa berpotensi berprestasi dengan jumlah mahasiswa sebanyak 62 orang mahasiswa dengan persentase 34,41%.
- b) *Cluster 1* merupakan kelompok mahasiswa bermasalah dengan jumlah mahasiswa sebanyak 4 orang mahasiswa dengan persentase 2,22%.
- c) *Cluster 2* merupakan kelompok mahasiswa berprestasi dengan jumlah mahasiswa sebanyak 104 orang mahasiswa dengan persentase 57,72%.
- d) *Cluster 3* merupakan kelompok mahasiswa berpotensi bermasalah dengan jumlah mahasiswa sebanyak 10 orang mahasiswa dengan persentase 5,55%.

#### 4. Kesimpulan

Berdasarkan pengujian dengan 180 data akademik mahasiswa Teknik Informatika menggunakan *software RapidMiner* hasilnya adalah *cluster 0* memiliki anggota sebanyak 62 orang mahasiswa yang berpotensi berprestasi dengan persentase 34,41%, *cluster 1* memiliki anggota sebanyak 4 orang mahasiswa yang bermasalah dengan persentase 2,22%, *cluster 2* memiliki anggota sebanyak 104 orang mahasiswa yang berprestasi dengan persentase 57,72%, dan *cluster 3* memiliki anggota sebanyak 10 orang mahasiswa yang berpotensi bermasalah dengan persentase 5,55%. Hasil proses pengelompokan (*clustering*) terhadap 180 *dataset* mahasiswa menggunakan aplikasi *RapidMiner* mendapatkan pengetahuan baru yang bisa dijadikan sumber informasi serta berfungsi sebagai model acuan bagi perencanaan akademik untuk memonitor dan memprediksi perkembangan performa akademik setiap mahasiswa.

#### Daftar Pustaka

- [1] M. R. A. Fernanda, P. Sokibi, And R. Fahrudin, "Sistem Prediksi Ketepatan Kelulusan Mahasiswa Berdasarkan Data Akademik Dan Non Akademik Menggunakan Metode K-Means (Studi Kasus: Universitas Catur Insan Cendekia)," *J. Digit*, Vol. 11, No. 1, P. 89, 2021.
- [2] K. Ariasa, I. G. A. Gunadi, And I. M. Candiasa, "Optimasi Algoritma Klaster Dinamis Pada K-Means Dalam Pengelompokan Kinerja Akademik Mahasiswa (Studi Kasus: Universitas Pendidikan Ganesha)," *J. Nas. Pendidik. Tek. Inform. Janapati*, Vol. 9, No. 2, Pp. 181–193, 2020.
- [3] Y. S. Adoe And J. Sembodo, "Peranan Keluarga Menurut Amsal 22:6 Dalam Pembentukan Karakter Anak," *Miktab J. Teol. Dan Pelayanan Kristiani*, Vol. 1, No. 1, P. 52, 2021.
- [4] M. Harahap, A. W. D. R. Zamili, M. A. Arvansyah, E. F. Saragih, S. Rajen, And A. M. Husein, "K-Means Clustering Algorithm Approach In Clustering Data On Cocoa Production Results In The Sumatra Region," *J. Resti (Rekayasa Sist. Dan Teknol. Informasi)*, Vol. 6, No. 6, Pp. 905–910, 2022.
- [5] A. Wibowo, Moh Makruf, Inge Virdyna, And Farah Chikita Venna, "Penentuan Klaster Koridor Transjakarta Dengan Metode Majority Voting Pada Algoritma Data Mining," *J. Resti (Rekayasa Sist. Dan Teknol. Informasi)*, Vol. 5, No. 3, Pp. 565–575, 2021.
- [6] D. Ariyanto, "Data Mining Menggunakan Algoritma K-Means Untuk Klasifikasi Penyakit Infeksi Saluran Pernafasan Akut," *J. Sistim Inf. Dan Teknol.*, Vol. 4, Pp. 13–18, 2022.
- [7] I. Jurnal, I. Bidang, T. Informasi, M. Al Ghifari, W. Trisari, And H. Putri, "Pengelompokan Mata Kuliah Berdasarkan Nilai Siswa Menggunakan K-Means Algoritma Dengan Metode Elbow Untuk Penentuan Centroid Machine Translated By Google," Vol. 8, No. 1, Pp. 42–46, 2023.
- [8] R. Liu, "Artikel Penelitian Analisis Data Evaluasi Pendidikan Menggunakan Metode K- Means Clustering," Vol. 2022, 2022.
- [9] "4. Klasterisasi Tingkat Penjualan Pada Startup Panak.Id Dengan Algoritma K-Means (S4).Pdf."

- [10] P. S. Informatika, U. Bhayangkara, J. Raya, And K. H. Dewantara, “Implementasi Algoritma K-Means Sebagai Clustering Metode Pemilihan Siswa Berprestasi Berdasarkan Kelas Akademik Machine Translated By Google,” Pp. 199–204.
- [11] N. Tahri And B. Ficht, “Membangun Pohon Konsensus Alternatif Dan Pohon Super Menggunakan K-Means Dan Jarak Robinson Dan Foulds,” Vol. 38, No. 13, Pp. 3367–3376, 2022.
- [12] B. J. Frasson And J. C. Rocha, “Studi Kasus Dalam Bahan Konstruksi Gangue Batubara Kaolinitik,” Vol. 18, Pp. 1–16, 2023.