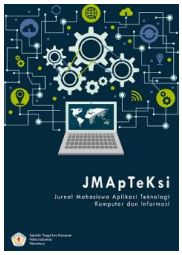


Jurnal Mahasiswa Aplikasi Teknologi Komputer dan Informasi

Vol. 2 No. 2 Thn 2020, Hal 97-101, E-ISSN : 2685-6565

Teknik *Cluster* Untuk Pengelompokan Kelas Akselarasi Menggunakan *K-Means*

Yuki Candra [a], *Deny Jollyta* [b]

[a]*Institut Bisnis dan Teknologi Pelita Indonesia, Jl. Jend. Ahmad Yani No. 78-88 Pekanbaru, yuki.candra@student.pelitaindonesia.ac.id*
[b]*Institut Bisnis dan Teknologi Pelita Indonesia, Jl. Jend. Ahmad Yani No. 78-88 Pekanbaru, deny.jollyta@lecturer.pelitaindonesia.ac.id*

INFORMASI ARTIKEL

Sejarah Artikel:
Diterima Redaksi: 10 Agustus 2020
Revisi Akhir: 30 Desember 2020
Diterbitkan *Online*: 9 Maret 2021

KATA KUNCI

Algoritma *K-Means*, Kelas Akselarasi, *Cluster*, Informasi

KORESPONDENSI

E-mail:
deny.jollyta@lecturer.pelitaindonesia.ac.id

A B S T R A C T

Penyelesaian pendidikan program sarjana tepat waktu menjadi target utama semua pergruan tinggi. Program akaselerasi berupa penyelesaian studi tujuh semester merupakan salah satu upaya yang dilakukan. Namun permasalahan yang muncul adalah mahasiswa yang masuk dalam program ini tidak tepat sasaran karena tidak terpilih secara sistem berdasarkan kriteria yang telah ditentukan sehingga belum mampu memenuhi target selesai teapt waktu. Penelitian ini bertujuan untuk memperoleh informasi tentang pengelompokan minat mahasiswa terhadap kelas akselerasi pada sebuah perguruan tinggi swasta melalui IPK, total sks dan nilai peminatan dengan menggunakan algoritma *K-Means* dalam teknik *cluster* pada data *mining*. Hasil *cluster* yang didapatkan dari tiga *centroid* dan berhenti pada iterasi kedelapan ini menunjukkan bahwa program akselerasi dapat diikuti oleh mahasiswa yang memiliki IPK > 3,3 dengan total sks selesai sebesar 109 dan rata-rata nilai peminatan lebih dari 77.

1. PENDAHULUAN

Data *Mining* merupakan ilmu komputer yang bekerja untuk menggali informasi. Pada buku Konsep Data *Mining* dan Penerapan [1], terdapat berbagai definisi tentang data *mining* yang mengarah pada upaya mengekstraksi informasi penting yang tersembunyi dalam *database* besar. Salah satu teknik penggalian dapat dilakukan melalui pengelompokan *(cluster)* menggunakan algoritma *K-Means*. Sejumlah informasi diperoleh dari penggalian data telah mampu menyelesaikan berbagai masalah, termasuk untuk mengetahui peminatan mahasiswa terhadap kelas akselerasi sebagai upaya penyelesaian studi tepat waktu. Perguruan Tinggi swasta saat ini telah menjalankan program penyelesaian studi selama tujuh semester. Umumnya mahasiswa tertarik untuk mengikuti namun belum diikuti dengan aturan yang menjaga agar mahasiswa pasti selesai dalam tujuh semester dengan mutu lulusan yang tetap terjaga.

Permasalahan ini disolusikan dengan menggali informasi dari data lulusan yang mengikuti program akselerasi sebelumnya berdasarkan beberapa kriteria yang dianggap menentukan, seperti IPK, total sks dan nilai peminatan. Algortima *K-Means* dianggap mampu membantu mendapatkan informasi yang dapat dijadikan acuan bagi mahasiswa dan Perguruan Tinggi pelaksana yang menjalankan program akselerasi tersebut.

Sebagai algoritma *clustering* yang dikenal, sejumlah penelitian telah membuktikan bahwa *K-Means* bekerja dengan baik pada data numerik dengan jumlah yang banyak, seperti [2], [3] dan [4] yang membahas tentang penggunaan *K-Means* untuk mengelompokkan penyakit dan konsumen hingga penentuan *cluster* optimal. *K-Means* juga digunakan untuk mendapatkan informasi berkenaan dengan penempatan mahasiswa pada jurusan berdasarkan sejumlah kriteria [5] dan [6].

Penelitian ini disertai dengan sistem yang mempermudah pengolahan data mahasiswa menggunakan algoritma *K-Means*. Pengelompokkan dibentuk dalam tiga *centroid* dengan jumlah data 68.

2. TINJAUAN PUSTAKA

2.1 Konsep Dasar Sistem Informasi

Berbagai definisi tentang sistem terdapat pada sejumlah buku dan artikel. Sistem adalah suatu jaringan kerja dari prosedur–prosedur yang saling berhubungan, berkumpul bersama–sama untuk melakukan suatu kegiatan atau untuk menyelesaikan suatu

sasaran tertentu [7]. Pada sumber lain disebutkan bahwa sistem adalah sekumpulan komponen yang mengimplementasi model dan fungsionalitas yang dibutuhkan. Komponen-komponen tersebut saling berinteraksi di dalam sistem guna mentransformasi *input* yang diberikan kepada sistem tersebut menjadi *output* yang berguna dan bernilai bagi *actor*-nya [8]. Untuk informasi, sejumlah definisi direalisasikan dalam banyak penelitian, seperti [9], [10] dan [11]. Dengan demikian sistem informasi merupakan kesatuan elemen-elemen yang saling berinteraksi secara sistematis dan teratur untuk menciptakan dan membentuk aliran informasi [12].

2.2. Clustering dan Algoritma K-Means

Clustering merupakan salah satu bagian dari teknik data *mining* yaitu sekumpulan objek yang mempunyai "kesamaan" diantara anggotanya dan memiliki "ketidaksamaan" dengan objek lain pada *cluster* lainnya, dengan kata lain sebuah *cluster* adalah sekumpulan objek yang digabung bersama karena persamaan atau kedekatannya. Algoritma *K-Means* merupakan satu dari sekian banyak algoritma yang mendukung prinsip kerja teknik ini.

Algoritma *K-means* sangat terkenal karena kemudahan dan kemampuannya untuk mengklasifikasi data besar dan *outlier* dengan sangat cepat. Berikut ini adalah langkah-langkah algoritma *K-means* [1] :

1. Pilih jumlah *cluster* K.
2. Inisialisasi K pusat *cluster* ini bisa dilakukan dengan berbagai cara. Namun yang paling sering dilakukan adalah dengan cara acak *(random)*. Pusat *cluster* diberi nilai awal dengan angka acak.
3. Alokasikan semua data / objek ke *cluster* terdekat. Kedekatan dua objek ditentukan berdasarkan jarak kedua objek tersebut. Demikian juga kedekatan suatu data ke *cluster* tertentu ditentukan jarak antara data dengan pusat *cluster*. Dalam tahap ini perlu dihitung jarak tiap data ke tiap pusat *cluster*. Jarak antara satu data dengan satu cluster tertentu akan menentukan suatu data masuk dalam *cluster* mana.
4. Hitung kembali pusat *cluster* dengan keanggotaan *cluster* yang sekarang. Pusat *cluster* adalah rata-rata dari semua data/ objek dalam *cluster* tertentu. Jika dikehendaki bisa juga menggunakan median dari *cluster* tersebut. Jadi rata-rata *(mean)* bukan satu-satunya ukuran yang bisa dipakai.
5. Tugaskan lagi setiap objek memakai pusat *cluster* yang baru. Jika pusat *cluster* tidak berubah lagi maka proses *clustering* selesai. Atau, kembali ke langkah nomor 3 sampai pusat *cluster* tidak berubah lagi.

Untuk menghitung jarak antar data, digunakan pengukuran *Euclidean Distance*. Adapun persamaan *Euclidean* adalah sebagai berikut [13]:

$$d(i,j)=\sqrt{(x_{1i}-x_{1j})^2+(x_{2i}-x_{2j})^2+\ldots+(x_{ki}-x_{kj})^2} \quad (1)$$

Dimana:

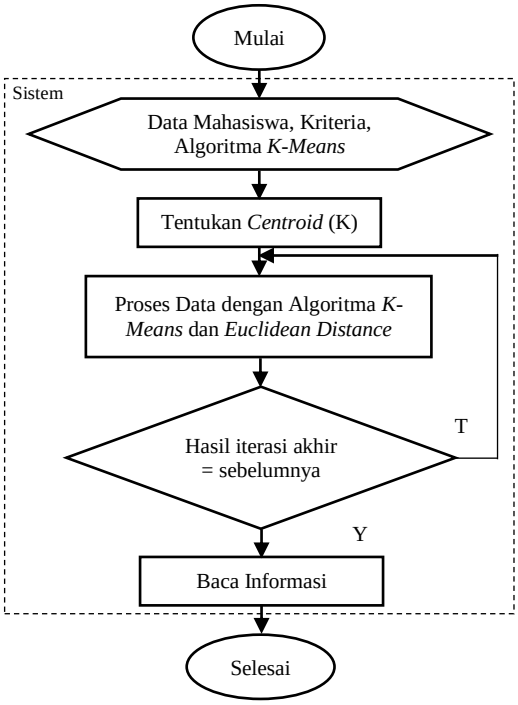
$d(i,j)$ = Jarak data ke i ke pusat *cluster* j

x_{ki} = Data ke i pada atribut data ke k

x_{kj} = Titik pusat ke j pada atribut ke k

3. METODE PENELITIAN

Untuk memudahkan proses penelitian, disusun langkah yang sistematis seperti pada Gambar 1.



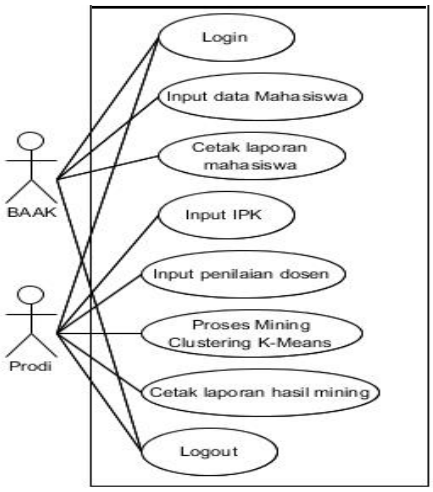
Gambar 1. Tahapan Penelitian

Penelitian diawali dengan mempersiapkan data mahasiswa akselerasi sebelumnya dari sebuah Perguruan Tinggi Swasta sebagai data *training*. Data dipersiapkan dengan tiga kriteria, yakni IPK, total sks dan nilai peminatan. Setelah menentukan jumlah *centroid* (K) yang diinginkan, data diolah menggunakan algoritma *K-Means* dan *Euclidean Distance*. Pengolahan membutuhkan sejumlah iterasi. Jika hasil iterasi akhir sama dengan sebelumnya, maka proses dihentikan.

4. HASIL DAN PEMBAHASAN

4.1. Perancangan Sistem

Berikut adalah rancangan sistem yang dibangun.



Gambar 2. Rancangan Sistem

Sistem yang dibangun adalah sistem untuk mengolah data *training* menjadi *cluster* dan menghasilkan informasi yang diinginkan. Pada Perguruan Tinggi yang dijadikan contoh, sistem melibatkan dua aktor, yakni Biro Akademik Administrasi Mahasiswa (BAAK) dan Program Studi (Prodi).

4.2. Proses Clustering

Proses *cluster* dilakukan dengan menggunakan persamaan (1). Jarak setiap data dihitung. Berikut ditampilkan perhitungan yang dimaksud dengan mengambil lima data sebagai contoh:

Jumlah cluster : 3
Jumlah data : 5
Jumlah atribut : 3

Tabel 1 : Data Mahasiswa

NIM	IPK	SKS	Nilai Dosen
1510307057001	3.46	109	70
1510307057002	3.23	102	80
⋮			
1510307057043	2.08	100	60
1510307057044	2.73	128	70
1510307057045	2.3	100	80

Iterasi ke-1

1. Penentuan pusat awal *cluster*

Pusat awal *cluster* atau *centroid* didapatkan secara acak, untuk penentuan awal *cluster* diasumsikan :
Pusat *Cluster* 1: (2.3,100,80)
Pusat *Cluster* 2: (2.73,128,70)
Pusat *Cluster* 3: (2.08,100,60)

2. Perhitungan jarak pusat *cluster*

Untuk mengukur jarak antara data (lima data) dengan pusat *cluster* digunakan persamaan (1), kemudian dilakukan penghitungan jarak dengan pusat *cluster* yang dimisalkan dengan **M(a,b,c),** dimana **a** merupakan nilai IPK, **b** total sks, dan **c** nilai peminatan dari dosen.
M1 = (3.46,109,70)
M2 = (3.23,102,80)
M44 = (2.08,100,60)
M45 = (2.73,128,70)
M46 = (2.3,100,80)
Dari hasil penghitungan Euclidean, kita dapat membandingkan :

Tabel 2 : Hasil Iterasi 1

NIM	C1	C2	C3
M1 (1510307057001)	13.50 $(((3.46-2.03)^2) + ((109-100)^2) + ((70-80)^2)^2)$	19.01 $(((3.46-2.73)^2) + ((109-128)^2) + ((70-70)^2)^2)$	10.388 $(((3.46-2.88)^2) + ((109-100)^2) + ((70-60)^2)^2)$
	Ket: $(((IPK\ Mhs - Min.\ IPK)^2) + ((SKS\ Mhs - Min.\ SKS)^2) + ((Nilai\ Mhs - Nilai\ Min)^2)^2)$		
M2 (1510307057002)	20.12	27.86	15.17
⋮			
M44 (1510307057044)	20	29.74	15
M45 (1510307057045)	29.74	0	28.45
M46 (1510307057046)	0	29.74	5

M6 : anggota C1 (Mahasiswa Berprestasi dan Rekomendasi MC)
M5 : anggota C2 (Mahasiswa Pertimbangan Rekomendasi MC)
M1, M2, M4 : anggota C3 (Bukan Mahasiswa Berprestasi dan Bukan Rekomendasi MC)

Iterasi ke-2

1. Hitung titik pusat baru

Tentukan posisi *centroid* baru (Ck) dengan cara menghitung nilai rata-rata dari data-data yang ada pada *centroid* yang sama.

$$C_k = \left(\frac{1}{n_k}\right)\sum d_i \qquad (2)$$

Dimana nk adalah jumlah dokumen dalam *cluster* k dan di adalah dokumen dalam *cluster* k. Sehingga didapatkan titik pusat atau *centroid* yang baru yaitu :
C1=(2.52,48,35.4)
C2=(2.85,140.25,72.5)
C3=(3.27,107.17,75.21)

2. Perhitungan jarak pusat *cluster*

Perhitungan yang sama dengan iterasi 1 kembali dilakukan dengan hasil sebagai berikut:

Tabel 3 : Hasil Iterasi 2

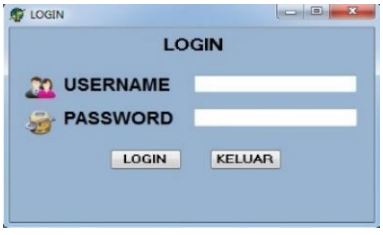
	C1	C2	C3
M1 (1510307057001)	70.12	31.35	5.5
M2 (1510307057002)	70.03	38.98	7.05
⋮			
M44 (1510307057044)	57.52	42.15	16.86
M45 (1510307057045)	87.15	12.503	21.47
M46 (1510307057046)	68.49	40.94	8.68

M45 : anggota C2 (Mahasiswa Pertimbangan Rekomendasi MC)
M46,M44,M1,M2 : anggota C3 (Bukan Mahasiswa Berprestasi dan Bukan Rekomendasi MC)

Karena pada Iterasi ke-2 posisi *cluster* tidak berubah/sama dengan posisi *cluster* pada iterasi pertama maka iterasi dihentikan.

4.3. Sistem Pengelompokan Kelas Akselerasi

Selanjutnya ditampilkan sistem pengelompokan kelas akselerasi yang telah dijalankan dan berhasil dengan baik.

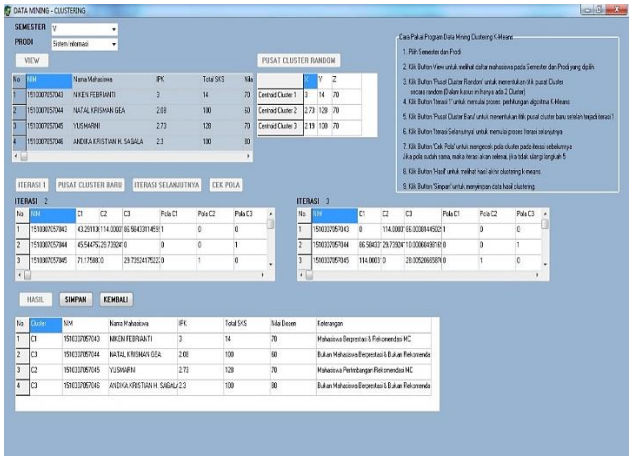


Gambar 3. ***Form Login***

Gambar 3 merupakan form login yang disiapkan untuk BAAK dan Prodi. Setelah login, maka masuk ke menu utama sistem seperti Gambar 4.

Gambar 4. Menu Utama

Menu Utama menyediakan sejumlah menu, termasuk menu untuk input data mahasiswa, IPK dan nilai peminatan. Proses input dilakukan oleh aktor yang sesuai dengan Gambar 2. Setelah semua data masuk, maka dilakukan proses *cluster* dengan tampilan berikut:



Gambar 5. Proses Clustering

4.4. Hasil Cluster

Hasil *cluster* adalah berupa informasi yang disampaikan dalam bentuk laporan dan diserahkan kepada pihak yang membutuhkan. Hasil mencapai konvergen pada iterasi kedelapan. Berdasarkan informasi *cluster* yang dihasilkan sistem, maka pengelompokannya adalah sebagai berikut.

Tabel 4 : Hasil *Cluster*

NO	NAMA MAHASISWA	IPK	SKS	NILAI
Cluster 1				
1	Alvin	3.58	110	90
2	Aulya Handayani	3.52	110	80
3	Beniardi	2.89	103	70
4	Billy Chandra	3.44	110	80
5	Boedi Maryoto	3.25	110	75
6	Chandra Putra Dinata	3.77	110	85
7	Andre Jhansen	3.22	110	80
8	Hendrei Tan	3.84	110	80
9	Herry	3.6	110	90
10	Johansen	3.26	110	80
11	Mega Utama	3.22	103	81
12	Mikhael Lauda	3.75	110	75
13	Ryan Putra Agung	3.48	103	70
14	Selly Fransisca	3.48	110	80
15	Sofian Winardi Hartopo	3.24	110	70
16	Vivian Zhen	3.85	110	70
17	Mario	3.07	110	80
18	Fatchurrohman	2.49	120	70
19	Arifin Kurniawan	3.63	110	80
20	Awie Prayogi	3.09	106	75
21	Azizie Muhammad	3.33	110	80
22	Betti Marito Bagariang	2.97	106	65
23	Calvin Rizky Utama Chia	3.76	110	75
24	Deli Zakaria	3.56	110	80
25	Digda Ari Syahputra	3.46	108	70
26	Firlan Riyadi Putra	3.1	108	65
27	Hermanto	3.7	110	90
28	Hermitha	3.85	110	90
29	Irfan Nugraha	3.42	99	60
30	Michael A R Sibarani	3.05	110	80
31	Muhammad Zia Ilham	3	106	75
32	Riko Thenardo	3.8	110	80
33	Rora Lumbantoruan	3.42	110	85
34	Salizar Ramadhan	3.3	105	80
35	Silvia Angelia Gozali	3.19	110	75
36	Thomas Malthus	3.55	110	90
37	Wesley Winardi	3.47	110	90
38	Yohanes	3.19	110	85
39	Yosua Pri	3.1	110	70
40	Yulhandre	2.77	103	65
41	Febri Fella Bella	3.11	98	55
42	Rahayu	3.38	110	80
43	Yoga Setyo Putra	3.68	110	85
44	Rusli	2.97	107	75
45	Jhon Shons	3.05	110	80
46	Randy Andika	3.2	110	75
47	Sherly Sutandy	3.27	110	80
48	Darmawan	3.23	104	75
Cluster 2				
1	Febryan Matra Saputra	3.55	65	50
2	Rudi	3.2	41	55
3	Said Handra Pramana	0.86	14	20
4	Fahrul Aziz	0.4	20	20
5	Andi Leonard Vernando	0.3	20	20
6	Bayu Sapriyanto	3.3	20	20
7	Dani Febrian Pardede	2.35	20	20
8	Debora Jesica Sitompul	2.15	20	20
9	Desman Setya Jaya	1.95	20	20
10	Jevry Anton Lt	2.98	65	40
11	Rahmad Dasril R.P	2.75	20	20
12	Sovanri Purba	0.8	20	10
13	Totiek Cahyadi	2.25	65	30
14	Sandy Irawan	1.68	41	20
15	Daniel Invades	1.85	65	20
Cluster 3				
1	Rio Riyanto	2.31	150	50
2	Sarimanto	3.38	144	80
3	Eveline	3.35	132	80
4	Andri Rizki Putra	2.74	150	70
5	Herman Faobazatulo	2.75	150	70

Cluster 1 berisikan data mahasiswa yang memiliki rata-rata IPK di atas 3,00 dengan jumlah sks 110. Anggota *cluster* ini berjumlah 48 dengan nilai peminatan rata-rata di atas 77. Untuk *Cluster* 2, beranggotakan data dengan rata-rata IPK di bawah 3,00. Kelompok ini memiliki minta yang kecil untuk mengikuti program akselerasi. *Cluster* 3 menjelaskan bahwa mahasiswa yang berminat ikut program akselerasi, tidak bisa dari mahasiswa yang berstatus pindahan (dari Peruruan Tinggi lain) karena menjadi tidak objektif jika ditinjau dari sisi IPK. Hal ini karena mahasiswa tersebut belum terhitung menyelesaikan studi minimal lima semester di Perguruan Tinggi saat ini.

5. KESIMPULAN DAN SARAN

Algoritma *K-Means* dengan rumus jarak *Euclidean* telah mampu memberikan informasi tentang kelas akselerasi berdasarkan tiga kriteria. Berdasarkan hasil yang diperoleh dari menjalankan sistem pengelompokan kelas akselerasi, terdapat tiga kelompok data yang diuji sebanyak delapan kali. Setiap kelompok mengandung informasi yang dapat dijadikan acuan Perguruan Tinggi dalam melaksanakan kelas akselerasi. Untuk mencapai penyelesaian studi tepat waktu, kelas akselerasi idealnya diikuti oleh mahasiswa regular, yakni mahasiswa yang mengikuti perkuliahan dari semester satu secara bertahap dan kontinu,

dengan IPK di atas 3,30 dan memiliki nilai peminatan di atas 77. Hal ini bisa diterima mengingat semester terakhir adalah tujuh sehingga mahasiswa seharusnya telah menyelesaikan total sks minimal 109 agar tujuan penyelesaian studi tepat waktu dapat direalisasikan.

Walaupun terdapat anggota *cluster* yang tidak sesuai sehingga memberikan informasi berbeda, dapat dimaklumi karena proses *cluster* menggunakan *centroid* yang ditentukan secara acak di awal proses. Untuk itu, proses *cluster* masih dapat dikembangkan pada jumlah kriteria yang lebih banyak dan dapat diujikan pada *centroid* yang berbeda hingga mendapatkan hasil yang benar-benar sesuai.

DAFTAR PUSTAKA

[1] D. Jollyta, W. Ramdhan, and M. Zarlis, *Konsep Data Mining dan Penerapan*, Pertama. Yogyakarta: Deepublish, 2020.

[2] P. D. . Silitonga, "Clustering of Patient Disease Data by Using K-Means Clustering," *Int. J. Comput. Sci. Inf. Secur. IJCSIS*, vol. 15, no. 7, pp. 219–233, 2017, [Online]. Available: https://sites.google.com/site/ijcsis/%5Cnhttp://sites.google.com/site/ijcsis/.

[3] T. . Kodinariya and P. . Makwana, "Review on determining number of Cluster in K-Means Clustering," *Int. J. Adv. Res. Comput. Sci. Manag. Stud.*, vol. 1, no. 6, pp. 90–95, 2013, [Online]. Available: www.ijarcsms.com.

[4] P. C. Ezenkwu, S. Ozuomba, and C. Kalu, "Application of K-Means Algorithm for Efficient Customer Segmentation: A Strategy for Targeted Customer Services," *Int. J. Adv. Res. Artif. Intell.*, vol. 4, no. 10, pp. 40–44, 2015, doi: 10.14569/ijarai.2015.041007.

[5] S. P. Borgavakar and A. Shrivastava, "Ev aluating Student ' s Performance u sing K-Means Clustering," *Int. J. Eng. Res. Technol.*, vol. 6, no. 05, pp. 114–116, 2017.

[6] E. E. Phyo and E. E. Myat, "Efficient K-Means Clustering Algorithm for Predicting of Students ' Academic Performance," *Int. J. Eng. Trends Appl.*, vol. 5, no. 6, pp. 15–18, 2018, [Online]. Available: http://www.ijetajournal.org/.

[7] C. P. Cordon, "System Theories: An Overview of Various System Theories and Its Application in Healthcare," *Am. J. Syst. Sci.*, vol. 2, no. 1, pp. 13–22, 2013, doi: 10.5923/j.ajss.20130201.03.

[8] S. K. Boell and D. Cecez-Kecmanovic, "What is an information system?," *Proc. Annu. Hawaii Int. Conf. Syst. Sci.*, vol. 2015-March, no. March, pp. 4959–4968, 2015, doi: 10.1109/HICSS.2015.587.

[9] D. Sukrianto and D. Oktarina, "Pemanfaatan Teknologi Barcode Pada Sistem Informasi Perpustakaan Di Smk Muhammadiyah 3 Pekanbaru," *JOISIE (Journal Inf. Syst. Informatics Eng.*, vol. 1, no. 2, pp. 136–143, 2019, doi: 10.35145/joisie.v1i2.216.

[10] W. J. Kurniawan, "Sistem Informasi Pengelolaan Laboratorium Komputer UPI-YPTK Padang," *J. Edik Inform.*, vol. 2, no. 1, pp. 95–101, 2015.

[11] S. Aswati, A. U. Firmansyah, W. Ramdhan, and S. Suhendra, "Analisis dan Perancangan Sistem Informasi Data Siswa Pada Sekolah Menengah Kejuruan (SMK) PGRI 8 Medan dengan Zachman Framework," *J. Sisfo*, vol. 06, no. 03, pp. 309–318, 2017.

[12] N. Novriyenni, E. Manik, and R. D. Andayani, "Perancangan Sistem Informasi Pembelian Tunai (Studi Kasus Koperasi Usaha Tani Mekar Jaya)," *J. Manaj. Inform. Komputerisasi Akunt.*, vol. 1, no. 1, pp. 33–38, 2017, doi: 10.37753/indikator.v1i1.18.

[13] W. Gie and D. Jollyta, "Perbandingan Euclidean dan Manhattan Untuk Optimasi Cluster Menggunakan Davies Bouldin Index : Status Covid-19 Wilayah Riau," in *Prosiding Seminar Nasional Riset Dan Information Science (SENARIS) 2020*, 2020, vol. 2, no. 2020, pp. 187–191.