



Analisis *Clustering* Dokumen Tugas Akhir Mahasiswa Sistem Informasi Universitas Nasional menggunakan Metode *K-Means Clustering*

Frankly Sept Genius Zendrato ¹, Agung Triayudi ^{2*}, Endah Tri E ³

^{1,2,3} Program Studi Sistem Informasi, Fakultas Teknologi Komunikasi dan Informatika, Universitas Nasional.

article info

Article history:

Received 11 June 2021

Received in revised form

2 July 2021

Accepted 26 August 2021

Available online January 2022

DOI:

<https://doi.org/10.35870/jti.k.v6i1.389>

Keywords:

Abstract Thesis; k-Means

Clustering; Text mining.

Kata Kunci:

Abstrak Skripsi; K-Means

Klustering; Penambangan

Text.

abstract

The purpose of this study was to determine the results of the analysis of the final project document for students majoring in information systems, National University. The research data is grouped based on the theme, object and research method. In this study, the K-Means Clustering method will be used which in the data includes the type of final project, year of publication and reasons for selecting the data. The data collection technique was chosen from the thesis document. The subjects in this study were part of the document that was processed in the abstract. Based on the results of the Clustering process above using the K-means algorithm method and the rapidminer application, it is concluded; 1) In the three clusters, it shows that the final project data 1 has 3 data, the final project 2 has 5 data, the final project 3 has 3 data and the final project 4 has 5 data., 2) In the three clusters the data is the most years old. group 3, namely in cluster 2 there are 3 data, cluster 3 there is 1 data, 3) In the three clusters, it shows that the data for the selection of 3 at least in cluster 2 there is 1 data.

abstrak

Tujuan penelitian ini untuk mengetahui hasil analisa dokumen tugas akhir mahasiswa dengan jurusan sistem informasi, Universitas Nasional. Data penelitian ini di kelompokkan berdasarkan tema, objek dan metode penelitian. Dalam penelitian ini akan menggunakan metode *K-Means Clustering* yang di dalam data tersebut meliputi jenis tugas tugas akhir, tahun terbit dan alasan pemilihan data. Teknik pengumpulan data dipilih dari dokumen skripsi. Subyek dalam penelitian ini bagian dari dokumen yang diolah secara abstrak. Berdasarkan hasil proses Clustering di atas dengan menggunakan metode algoritma K-means dan aplikasi rapidminer maka disimpulkan; 1) Pada ketiga cluster menunjukkan bahwa data tugas akhir 1 memiliki data sebanyak 3 data, tugas akhir 2 sebanyak 5 data, tugas akhir 3 sebanyak 3 data dan tugas akhir 4 sebanyak 5 data., 2) Pada ketiga cluster tersebut data yang paling banyak tahun golongan 3 yaitu pada cluster 2 terdapat 3 data, cluster 3 terdapat 1 data, 3) Pada ketiga cluster menunjukkan bahwa data pemilihan 3 paling sedikit pada cluster 2 terdapat 1 data.

*Corresponding author. Email: agungtriyudi@civitas.unas.ac.id ².

© E-ISSN: 2580-1643.

Copyright © 2022. Published by Lembaga Otonom Lembaga Informasi dan Riset Indonesia (KITA INFO dan Riset) (<http://creativecommons.org/licenses/by/4.0/>).

1. Latar Belakang

Perkembangan teknologi sekarang ini semakin canggih dan modern. Perkembangan saat ini telah menghasilkan banyak sekali kumpulan-kumpulan data yang datanya tidak dapat diketahui secara manual [1,2]. Pada perkembangan saat ini dibutuhkan sebuah teknologi basis data yang disebut sebagai data mining. Data mining adalah salah satu bagian dari sistem informasi yang mempunyai ide sampai dengan implementasi akhir pengumpulan data [3,4]. Proses ini dapat mengumpulkan suatu data yang tujuannya agar pada data yang tekumpul dapat menemukan pola-pola yang belum diketahui [5,6]. *K-Means* merupakan salah satu algoritma yang sangat efektif untuk dapat menemukan suatu *cluster* dalam sebuah kumpulan data [7,8,9], algoritma tersebut dapat melakukan analisa suatu data ke dalam kelompok berdasarkan nilai rata-rata *means* terdekat [10,11].

Sederhana-nya data ini dapat disimpulkan sebagai proses menguraikan atau mengidentifikasi sebuah data yang data tersebut ada dalam suatu *database* besar. Data ini juga bisa digunakan sebagai bahan penelitian mahasiswa untuk mencai pola atau nilai dalam suatu basis data yang sangat besar. Dalam mengelola berbagai proses tersebut diperlukan sebuah penelitian yang dapat mempermudah pengelolaan atau dapat befungsi sebagai hasil uji penelitian tersebut. Informasi tersebut diharapkan dapat membantu dosen dalam menemukan pola-pola baru, dan mempunyai nilai yang berguna serta dapa memahami data yang ada. Jadi, untuk dapat mengetahui hasil data dibutuhkan suata data yang disebut sebagai data mining.

Pengelompokan data penelitian yang umumnya berbentuk teks karena biasanya data tersebut di dapat dari suatu dokumen [12,13]. Tujuannya mendapatkan informasi yang bisa digunakan dari sekumpulan dokumen yang ada [14]. Jadi, untuk dapat menganalisa dan mencari pola teks yang bermanfaat untuk tujuan tertentu digunakan teknik dalam *text mining* yang merupakan penerapan konsep. Adapun kegunaan *text mining* yaitu mengelomppokan data.

Salah satu metode yang digunakan *text mining* adalah *Clustering*. *Clustering* adalah metode analisa data yang disebut sebagai pengelompokkan data [15].

Pengelompokkan data yang dimaksud yaitu suatu objek data yang ada dalam satu *cluster* memiliki tingkat kemiripan antar satu data dengan data lainnya. Hasil *Clustering* tersebut dapat menghasilkan kesamaan data yang tinggi dan kesamaan data yang rendah.

Adapun rumusan masalah yaitu bagaimana metode *K-means Clustering* menganalisa dokumen tugas akhir mahasiswa?, dalam penelitian ini peneliti membatasi permasalahan yaitu; 1) Data yang digunakan adalah data abstrak daridokumen tugas akhir prodi sistem informasi, Universitas Nasional, dan 2) Data yang digunakan adalah data mahasiswa sistem informasi tahun 2019/2021. Tujuan dari penelitian ini adalah untuk menghasilkan analisa metode *K-Means Clustering* pada pengelompokan dokumen tugas akhir sistem informasi.

2. Metode Penelitian

Dalam metode penelitian ini dilakukan tahapan-tahapan yang berfungsi sebagai pedoman dalam penelitian, tahapan tersebut sebagai berikut:

Dataset

Data penelitian ini dipilih dari data skripsi yang diterbitkan pada tahun 2019/2021.

Tabel 1. Contoh Tabel Data

Jenis Tugas Akhir	Tahun	Pemilihan
2	5	1
4	5	2
3	5	3
2	1	1
1	1	4
4	1	2
2	2	1
4	2	2
1	2	4
3	4	3
2	4	1
4	4	2
2	3	1
4	3	2
3	3	3
1	3	4

Pra-Proses Data

Proses data ini merupakan proses untuk pengumpulan data mentah sebelum dilakukan proses lainnya. Dalam proses ini diperlukan inisialisasi untuk mempermudah proses *Clustering* pada algoritma *K-means*. Data yang digunakan ini bersifat numerik, maka data nominal seperti pemilihan harus diubah dalam bentuk nilai numerik.

Tabel 2. Sampel inisialisasi data Tugas Akhir

Data Mining	1
Sistem Pakar	2
Perancangan Sistem Informasi	3
Sistem Pendukung Keputusan	4

Pada tabel 2 dilakukan inisialisasi data tugas akhir yaitu Data Mining diberi inisialisasi 1, Sistem Pakar diberi inisialisasi 2, Perancangan Sistem Informasi diberi inisialisasi 3, dan Sistem Pendukung kPK Keputusan diberi inisialisasi 4.

Tabel 3. Sampel inisialisasi data tahun terbit

Ganjil 2019	1
Genap 2019	2
Ganjil 2020	3
Genap 2020	4
Ganjil 2021	5

Pada tabel 3 adalah inisialisasi data tahun terbit yaitu pada Ganjil 2019 diberi inisialisasi 1, pada Genap 2019 diberi inisialisasi 2, pada Ganjil 2020 diberi inisialisasi 3, pada Genap 2020 diberi inisialisasi 4, pada Ganjil 2021 diberi inisialisasi 5.

Tabel 4. Sampel inisialisasi data pemilihan

Menghimpun data dalam jumlah terbesar	1
Memberikan informasi untuk mengambil keputusan	2
Memberi gambaran dan rancang bangun yang lengkap	3
Data yang digunakan untuk menghasilkan informasi	4

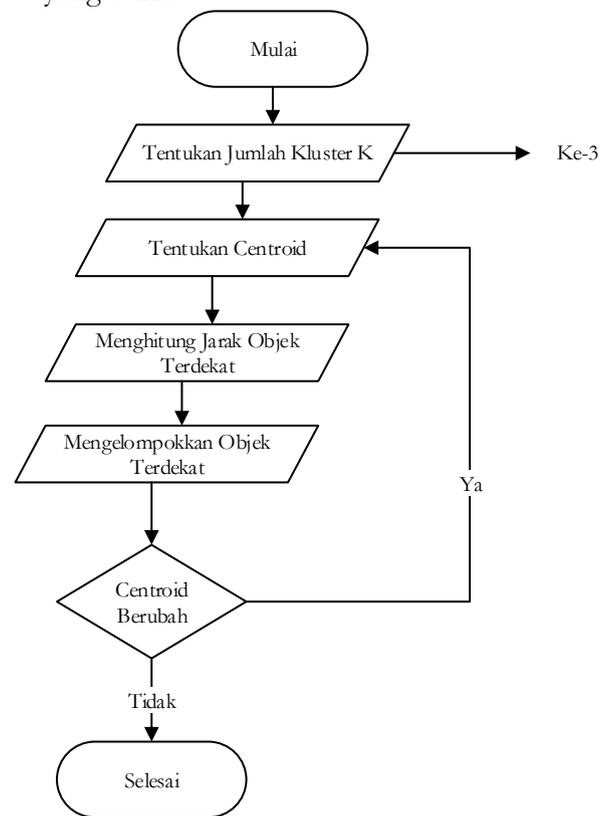
Pada tabel 4 dilakukan inisialisasi data pemilihan yaitu Menghimpun data dalam jumlah terbesar diberi inisialisasi 1, Memberikan informasi untuk mengambil keputusan diberi inisialisasi 2, Memberi

gambaran dan rancang bangun yang lengkap diberi inisialisasi 3, Data yang digunakan untuk menghasilkan informasi diberi inisialisasi 4.

Alur proses Clustering

Alur proses *Clustering* ini dengan menggunakan metode *K-Means Clustering* terdapat langkah-langkah yang dilakukan dengan algoritma dasar *K-Means*, sebagai berikut:

- 1) Jumlah *cluster* yang akan digunakan.
- 2) Memilih *cluster* secara acak dan menentukan *centroid* baru.
- 3) Menghitung data dengan menggunakan *centroid*.
- 4) Dan jika terdapat data yang berpindah *cluster* ke *cluster* yang lain maka harus menemukan *centroid* baru supaya hasil *cluster* sampai memiliki hasil data yang sama.



Gambar 1. Flowchart algoritma metode *K-means Clustering*

3. Hasil dan Pembahasan

Dalam pengolahan data ini untuk mendapatkan informasi tentang dokumen tugas akhir dengan cara perhitungan manual. Dalam proses ini terdapat 3 kluster yang digunakan dan atribut data yang ada pada tabel di bawah yaitu jenis tugas akhir, tahun dan pemilihan. Pada tabel 5, penulis memilih data secara

acak untuk mendapatkan *centroid* data yang berjumlah 16 data.

Tabel 5. Sampel data.

No	Jenis Tugas Akhir	Tahun	Pemilihan
1	2	5	1
2	4	5	2
3	3	5	3
4	2	1	1
5	1	1	4
6	4	1	2
7	2	2	1
8	4	2	2
9	1	2	4
10	3	4	3
11	2	4	1
12	4	4	2
13	2	3	1
14	4	3	2
15	3	3	3
16	1	3	4

Setelah itu memilih 3 data secara acak dan sekaligus dapat menemukan sample *centroid* data yang di inginkan.

Tabel 6. Sampel *centroid*.

Cluster 1	4	5	2
Cluster 2	3	3	3
Cluster 3	2	3	1

Setelah itu dilakukan perhitungan dari setiap data ke pusat kluster dengan perhitungan jarak *Euclidean* dan mendapatkan matriks perhitungan jarak tersebut sebagai berikut:

$$D(e, f) = \sqrt{(e1 - e2)^2 + (f1 - f2)^2} \dots (1)$$

Kemudian hasil antara data pertama sampai dengan data ke 16 data *centroid* akan menghasilkan perhitungan sebagai berikut:

$$d = \sqrt{(2 + 4)^2 + (5 + 5)^2 + (1 + 2)^2}$$

$$d = 2,23$$

Untuk mendapatkan hasil C1 terhadap data dapat menggunakan rumus di atas, dan hasil perhitungan

sesuai dengan data yang terdapat pada tabel 7.

$$d = \sqrt{(2 + 3)^2 + (5 + 3)^2 + (1 + 3)^2}$$

$$d = 3$$

Untuk mendapatkan hasil C2 terhadap data dapat menggunakan rumus di atas, dan hasil perhitungan sesuai dengan data yang terdapat pada tabel 7.

$$d = \sqrt{(2 + 2)^2 + (5 + 3)^2 + (1 + 1)^2}$$

$$d = 2$$

Untuk mendapatkan hasil C3 terhadap data dapat menggunakan rumus di atas, dan hasil perhitungan sesuai dengan data yang terdapat pada tabel 7.

Tabel 7. Hasil literasi pertama

No	C1	C2	C3	JARAK TER	CLUSTER
1	2,236067977	3	2	2	3
2	0	2,449489743	3	0	1
3	1,414213562	2	3	1,414213562	1
4	4,582575695	3	2	2	3
5	5,385164807	3	3,741657387	3	2
6	4	2,449489743	3	2,449489743	2
7	3,741657387	2,449489743	1	1	3
8	3	1,732050808	2,449489743	1,732050808	2
9	4,69041576	2,449489743	3,31662479	2,449489743	2
10	1,732050808	1	2,449489743	1	2
11	2,449489743	2,449489743	1	1	3
12	1	1,732050808	2,449489743	1	1
13	3	2,236067977	0	0	3
14	2	1,414213562	2,236067977	1,414213562	2
15	2,449489743	0	2,236067977	0	2
16	4,123105626	2,236067977	3,16227766	2,236067977	2

Setelah itu menghitung jarak dari data *cluster* ke pusat *cluster* kemudian mendapatkan *centroid* baru dari hasil perhitungan nilai data setiap *cluster*nya, *centroid* yang baru yang telah didapatkan, sebagai berikut:

Tabel 8. *Centroid* Baru

C2	3,666	4,6666	2,333
C2	2,625	2,375	3
C3	2	3	1

Setelah itu penulis melakukan proses perhitungan ulang kembali menggunakan *centroid* baru kemudian mendapatkan hasil literasi yang kedua pada tabel di bawah, sebagai berikut:

Tabel 9. Hasil Literasi Dua

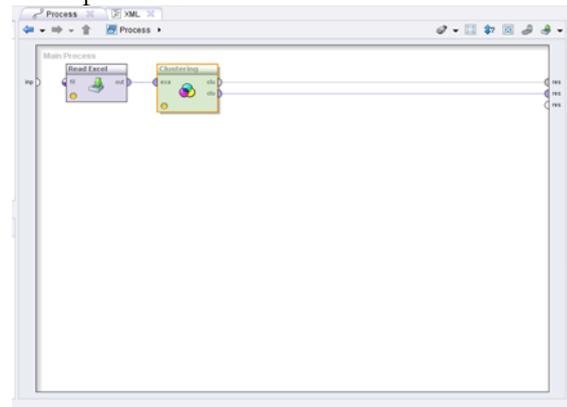
No	C1	C2	C3	JARAK TER	CLUSTER
1	2,159537117	3,35875721	2	2	3
2	0,577581648	3,127499	3	0,577581648	1
3	0,99980026	2,65165043	3	0,99980026	1
4	4,242216468	2,50624221	2	2	3
5	4,830155335	2,35186097	3,741657	2,351860965	2
6	3,696809511	2,18660696	3	2,186606961	2
7	3,415142832	2,1286733	1	1	3
8	2,707988287	1,74104853	2,44949	1,741048535	2
9	4,12276613	1,94454365	3,316625	1,944543648	2
10	1,154469818	1,66770801	2,44949	1,154469818	1
11	2,235352446	2,65165043	1	1	3
12	0,816578569	2,35186097	2,44949	0,816578569	1
13	2,707397378	2,18660696	0	0	3
14	1,732050969	1,81142209	2,236068	1,732050969	1
15	1,914680276	0,72886899	2,236068	0,728868987	2
16	3,558651509	2,0077973	3,162278	2,007797301	2

Setelah mendapatkan hasil perhitungan dari literasi kedua, penulis akan membandingkan hasil literasi pertama dan literasi kedua jika hasil data kluster atau kelompok data tidak berubah ataupun data tidak berpindah dari satu kluster ke kluster yang lain maka perhitungan tersebut akan berhenti pada hasil literasi pertama dan hasil literasi kedua maka hasil kluster data tersebut dianggap sudah konvergen. Tetapi penulis dalam hal ini akan melakukan perhitungan sampai literasi ketiga agar mendapatkan *cluster* yang sama untuk mencapai titik konvergen.

Tabel 10. Hasil literasi ketiga

No	C1	C2	C3	JARAK TER	CLUSTER
1	2,271563338	3,715164196	2	2	3
2	0,979795897	3,624699298	3	0,979795897	1
3	1,166190379	3,077733744	3	1,166190379	1
4	3,841874542	2,408826478	2	2	3
5	4,422668877	1,863449758	3,741657387	1,863449758	2
6	3,249615362	2,266813843	3	2,266813843	2
7	3,059411708	2,191448151	1	1	3
8	2,271563338	2,034316839	2,449489743	2,034316839	2
9	3,762977544	1,57240103	3,31662479	1,57240103	2
10	0,871779789	2,114815595	2,449489743	0,871779789	1
11	2,13541565	2,966891471	1	1	3
12	0,6	2,852795997	2,449489743	0,6	1
13	2,441311123	2,408826478	0	0	3
14	1,326649916	2,266813843	2,236067977	1,326649916	1
15	1,469693846	1,213443447	2,236067977	1,213443447	2
16	3,280243893	1,863449758	3,16227766	1,863449758	2

Setelah itu penulis melakukan klustering data dengan perangkat lunak aplikasi rapidminer. Di bawah ini terdapat hasil pengolahan data dengan bantuan aplikasi rapidminer.



Gambar 2. Pemodelan K-means menggunakan aplikasi rapidminer

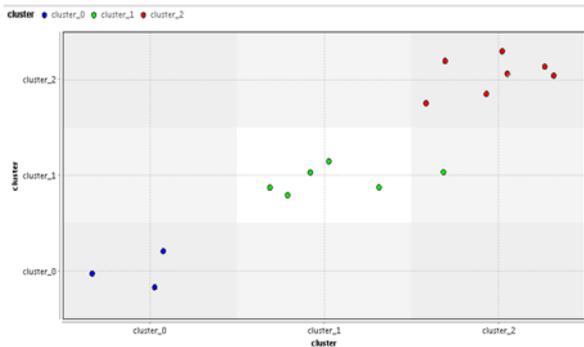
Row No.	id	cluster	Jenis Tuga...	Tahun	Pemilihan
1	1	cluster_0	2	5	1
2	2	cluster_0	4	5	2
3	3	cluster_0	3	5	3
4	4	cluster_1	2	1	1
5	5	cluster_2	1	1	4
6	6	cluster_1	4	1	2
7	7	cluster_1	2	2	1
8	8	cluster_1	4	2	2
9	9	cluster_2	1	2	4
10	10	cluster_0	3	4	3
11	11	cluster_0	2	4	1
12	12	cluster_0	4	4	2
13	13	cluster_1	2	3	1
14	14	cluster_1	4	3	2
15	15	cluster_1	3	3	3
16	16	cluster_2	1	3	4

Gambar 3. Hasil implementasi data dengan aplikasi rapidminer

Kemudian dilakukan proses pemodelan K-means seperti gambar diatas, dan dalam proses ini terdapat 3 kluster sesuai dengan yang diinginkan penulis yaitu, kluster_0 sebanyak 3 items, kluster_1 sebanyak 6 items, kluster_2 sebanyak 7 items, dengan jumlah keseluruhan item yaitu 16 item.

Tabel 11. Cluster model menggunakan aplikasi rapidminer

Cluster Model
Cluster 0: 3 items
Cluster 1: 6 items
Cluster 2: 7 items
Total number of items: 16



Gambar 4. Tampilan plot view menggunakan rapidminer

Pola penyebaran antar kluster terlihat tidak begitu jauh berbeda karena masing-masing dari kluster tersebut tidak berselisih jauh dalam jumlah anggota. Hasil pengelompokan kluster data yang menggunakan aplikasi rapidminer dapat dilihat di tabel berikut ini:

Tabel 12. Hasil pengelompokan Clustering Data

HASIL CLUSTER 1	HASIL CLUSTER 2	HASIL CLUSTER 3
TUGAS AKHIR 2	2	TUGAS AKHIR 2
TUGAS AKHIR 3	2	TUGAS AKHIR 3
TUGAS AKHIR 4	2	TUGAS AKHIR 4
TAHUN	GOL4=3	TAHUN
	GOL5=3	GOL1=2
PEMILIHAN 1	2	GOL2=2
PEMILIHAN 2	2	GOL3=3
PEMILIHAN 3	2	PEMILIHAN 1
PEMILIHAN 4	0	PEMILIHAN 2
		PEMILIHAN 3
		PEMILIHAN 4

Pada hasil *cluster* 1 data tugas akhir 2 memiliki 2 data, tugas akhir 3 memiliki 2 data, tugas akhir 4 memiliki 2 data. Dan tahun golongan 4 terdapat 3 data, golongan 5 terdapat 3 data. Dan data pemilihan 1 ada 2 data, pemilihan 2 ada 2 data, pemilihan 3 ada 2 data sedangkan di pemilihan 4 tidak terdapat data sama sekali. Pada hasil *cluster* 2 data tugas akhir 2 memiliki 3 data, tugas akhir 3 memiliki 1 data, tugas akhir 4 memiliki 3 data. Dan tahun golongan 1 memiliki 2 data, golongan 2 memiliki 2 data, golongan 3 memiliki 3 data. Dan data pemilihan 1 ada 3 data, pemilihan 2 ada 3 data, pemilihan 3 ada 1 data sedangkan di pemilihan 4 tidak terdapat data sama sekali.

Pada hasil *cluster* 3 data tugas akhir 1 memiliki 3 data. Dan tahun golongan 1 memiliki 1 data, golongan 2 memiliki 1 data, golongan 3 memiliki 1 data. Dan data pemilihan 1, pemilihan 2, pemilihan 3 tidak terdapat data, sedangkan di pemilihan 4 ada 3 data.

4. Kesimpulan

Berdasarkan hasil proses *Clustering* di atas dengan menggunakan metode algoritma *K-means* dan aplikasi rapidminer maka disimpulkan:

- 1) Pada ketiga *cluster* menunjukkan bahwa data tugas akhir 1 memiliki data sebanyak 3 data, tugas akhir 2 sebanyak 5 data, tugas akhir 3 sebanyak 3 data dan tugas akhir 4 sebanyak 5 data.
- 2) Pada ketiga *cluster* tersebut data yang paling banyak tahun golongan 3 yaitu pada *cluster* 2 terdapat 3 data, *cluster* 3 terdapat 1 data.
- 3) Pada ketiga *cluster* menunjukkan bahwa data pemilihan 3 paling sedikit pada *cluster* 2 terdapat 1 data.

5. Daftar Pustaka

- [1] Putong, I., 2015. *Teori Ekonomi Mikro: Konvensional dan Syariah* (Vol. 1). Buku&Artikel Karya Iskandar Putong.
- [2] Romindo, R., Muttaqin, M., Rasinus, R., Israwan, L.F., Yuswardi, Y., Karim, A., Sari, A.N., Putri, E.E. and Samosir, K., 2021. *Sistem Informasi*. Yayasan Kita Menulis.
- [3] Putri, S.U., Irawan, E. and Rizky, F., 2021. Implementasi Data Mining Untuk Prediksi Penyakit Diabetes Dengan Algoritma C4.5. *Kesatria: Jurnal Penerapan Sistem Informasi (Komputer dan Manajemen)*, 2(1), pp.39-46.
- [4] Ginantra, N.L.W.S.R., Arifah, F.N., Wijaya, A.H., Septarini, R.S., Ahmad, N., Ardiana, D.P.Y., Effendy, F., Iskandar, A., Hazriani, H., Sari, I.Y. and Gustiana, Z., 2021. *Data Mining dan Penerapan Algoritma*. Yayasan Kita Menulis.
- [5] Jabat, J.T. and Murdani, M., 2019. Penerapan Data Mining Pada Penjualan Produk Retail Menggunakan Metode Clustering. *Pelita Informatika: Informasi dan Informatika*, 8(1), pp.26-32.

- [6] Sibuea, M.L. and Safta, A., 2017. Pemetaan Siswa Berprestasi Menggunakan Metode K-Means Clustering. *JURTEKSI (Jurnal Teknologi dan Sistem Informasi)*, 4(1), pp.85-92.
- [7] Sani, A., 2018. Penerapan metode k-means clustering pada perusahaan. *Jurnal Ilmiah Teknologi Informasi*, 353, pp.1-7.
- [8] Amri, M.A., Windarto, A.P., Wanto, A. and Damanik, I.S., 2019. Analisis Metode K-Means Pada Pengelompokan Perguruan Tinggi Menurut Provinsi Berdasarkan Fasilitas Yang Dimiliki Desa. *KOMIK (Konferensi Nasional Teknologi Informasi dan Komputer)*, 3(1).
- [9] Dewi, S.M., Windarto, A.P., Damanik, I.S. and Satria, H., 2019, August. Analisa Metode K-Means pada Pengelompokan Kriminalitas Menurut Wilayah. In *Seminar Nasional Sains dan Teknologi Informasi (SENSASI)* (Vol. 2, No. 1).
- [10] Nurzahputra, A., Muslim, M.A. and Khusniati, M., 2017. Penerapan algoritma K-Means untuk clustering penilaian dosen berdasarkan indeks kepuasan mahasiswa. *Techno. Com*, 16(1), pp.17-24.
- [11] Khanmohammadi, S., Adibeig, N. and Shanehbandy, S., 2017. An improved overlapping k-means clustering method for medical applications. *Expert Systems with Applications*, 67, pp.12-18.
- [12] Yu, S.S., Chu, S.W., Wang, C.M., Chan, Y.K. and Chang, T.C., 2018. Two improved k-means algorithms. *Applied Soft Computing*, 68, pp.747-755.
- [13] Chi, D., 2021, January. Research on the Application of K-Means Clustering Algorithm in Student Achievement. In *2021 IEEE International Conference on Consumer Electronics and Computer Engineering (ICCECE)* (pp. 435-438). IEEE.
- [14] Fabregas, A.C., Gerardo, B.D. and Tanguilig III, B.T., 2017. Enhanced initial centroids for k-means algorithm. *International Journal of Information Technology and Computer Science*, 1, pp.26-33.
- [15] Mohd, W.M.W., Beg, A.H., Herawan, T., Noraziah, A. and Chiroma, H., 2019. Multi-dimensional K-Means Algorithm for Student Clustering. In *Proceedings of the International Conference on Data Engineering 2015 (DaEng-2015)* (pp. 119-128). Springer, Singapore.