



Analisis Faktor Yang Mempengaruhi Penumpang Angkutan Umum Beralih ke Transportasi *Online* Go-Jek menggunakan Metode *K-Means Clustering*

Muhammad Ilyas Sahputra ¹, Agung Triayudi ^{2*}, Albaar Rubhasy ³

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

article info

Article history:

Received 12 June 2021

Received in revised form

29 July 2021

Accepted 26 August 2021

Available *online* January 2022

DOI:

<https://doi.org/10.35870/jtik.v6i1.381>

Keywords:

Online Transportation Go-Jek; Clusters; Data Mining; K-Means.

Kata Kunci:

Transportasi Online Go-Jek; Klustering; Data Mining; K-Means.

abstract

The objectives of this research are; 1) Can find the results of the analysis of switching factors, namely the age data, and 2) Can find the results of the analysis of switching factors, namely the data on the travel time of public transportation, 3) Can find the results of the analysis of switching factors, namely the data on the travel time of Go-jek 4) Can find the results of the analysis of switching factors, namely the tariff data, and 5) Can find the results of the analysis of switching factors, namely the data that is easy to obtain. This study was designed to determine the results of the analysis of factors that influence the shifting of public transportation to Go-jek online transportation using the K-means clustering algorithm. The data collection technique in this study was by means of a questionnaire through the Go-jek community in Indonesia and secondary data taken from the internet media. Based on the results of the analysis that has been carried out on the analysis of factors that influence public transport passengers to switch to Go-jek online transportation using the K-means clustering algorithm, it is hoped that further researchers will test with other clustering algorithms, and the rapidminer software used as research material can be developed further. become more other features.

abstrak

Tujuan Penelitian ini yaitu; 1) Dapat menemukan hasil analisa faktor yang beralih yaitu pada data umur, dan 2) Dapat menemukan hasil analisa faktor yang beralih yaitu pada data jangka tempuh angkutan umum, 3) Dapat menemukan hasil analisa faktor yang beralih yaitu pada data jangka tempuh Go-jek, 4) Dapat menemukan hasil analisa faktor yang beralih yaitu pada data tarif, dan 5) Dapat menemukan hasil analisa faktor yang beralih yaitu pada data mudah di dapat. Penelitian ini di rancang untuk mengetahui hasil analisis faktor yang mempengaruhi penumpang angkutan umum beralih ke transporasi online Go-jek menggunakan algoritma K-means klustering. Teknik pengumpulan data dalam penelitian ini dengan cara kuesioner melalui komuninas Go-jek di Indonesia dan data sekunder yang di ambil dari media internet. Berdasarkan hasil analisis yang telah dilakukan pada analisis faktor yang mempengaruhi penumpang angkutan umum beralih ke transportasi online Go-jek menggunakan algoritma K-means klustering maka diharapkan untuk peneliti selanjutnya untuk menguji dengan algoritma clustering lainnya, serta software rapidminer yang digunakan sebagai bahan penelitian bisa dikembangkan lagi menjadi lebih banyak fitur lainnya.

*Corresponding author. Email: agungtriayudi@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

Penggunaan transportasi yaitu aspek yang sangat penting bagi masyarakat karena dapat membantu pendapatan ekonomi suatu negara. Oleh karena itu, dengan bertambahnya masyarakat dari tahun ke tahun akan menimbulkan kelonjakan penduduk dan akan membuat kemacetan di kota-kota besar Indonesia disebabkan kendaraan masyarakat yang terus bertambah [1]. Apabila fenomena seperti ini terus menerus terjadi dan tidak ada respon yang baik dari pihak Pemerintah setempat maka akan menimbulkan penghambatan lalu lintas kendaraan dan berkurangnya perekonomian di Indonesia.

Faktor perpindahan penggunaan layanan transportasi tidaklah terjadi sendiri, akan tetapi adanya keperluan yang berbeda. Peralihan akan terjadi apabila ada faktor-faktor yang memungkinkan [2]. Pengurangan yang telah menimpa pada transportasi umum yang di sebabkan oleh kelonjakan penduduk dan penggunaan kendaraan pribadi telah terjadi di kota-kota besar Indonesia. maka untuk mengetahui faktor-faktor yang mempengaruhi perpindahan penggunaan layanan transportasi yang di sebabkan oleh kelonjakan penduduk, perusahaan-perusahaan besar membuat layanan transportasi baru yang berbasis *online* yaitu contohnya Go-jek.

Permasalahan dalam penelitian ini yaitu faktor yang mempengaruhi para penumpang angkutan umum beralih pada transportasi *online* Go-Jek. Sedangkan batasan pada penelitian ini yaitu; 1) Jumlah penduduk yang berlebihan membuat jasa angkutan umum tidak memadai dan membuat kemacetan sehingga para penumpang berpindah ke jasa transportasi lain yaitu ojek *online* Go-Jek, dan 2) Jangka tempuh mempengaruhi keputusan Penumpang dalam beralih menggunakan jasa transportasi *online* yaitu yang terpilih Go-Jek. Tujuan Penelitian yaitu; 1) Dapat menemukan hasil analisa faktor yang beralih yaitu pada data umur, dan 2) Dapat menemukan hasil analisa faktor yang beralih yaitu pada data jangka tempuh angkutan umum, 3) Dapat menemukan hasil analisa faktor yang beralih yaitu pada data jangka tempuh Go-jek, 4) Dapat menemukan hasil analisa faktor yang beralih yaitu pada data tarif, dan 5) Dapat menemukan hasil analisa faktor yang beralih yaitu pada data mudah di dapat.

Beberapa penelitian yang digunakan sebagai referensi. Peneliti mengacu pada jurnal penelitian pertama yaitu Analisis K-Means *Clustering* Dalam Pengelompokkan Data Imunisasi Campak Balita di Indonesia, dapat digunakan untuk memudahkan proses analisis [3]. Pada penelitian kedua yaitu tentang analisis *cluster* dengan algoritma k-means dan fuzzy c-means *clustering* untuk pengelompokan data obligasi korporasi. Dapat di simpulkan bahwa dengan di terapkannya metode K-means ini dapat membantu dalam analisis data yang dibuat [4]. Pada penelitian ketiga yaitu tentang implementasi algoritma K-means *clustering* untuk mengetahui bidang skripsi mahasiswa multimedia pendidikan teknik informatika dan komputer. Dapat di simpulkan bahwa dengan di terapkannya metode K-means ini dapat membantu dalam analisis data yang di buat [5], dan Pada penelitian ke empat yaitu sistem monitoring dan klaster ketersediaan energi menggunakan metode k-means pada pembangkit listrik tenaga surya Dapat di simpulkan bahwa dengan di terapkannya metode K-means ini dapat membantu pengerjaan dalam penelitian ini [6].

Berikut studi literatur yang berkaitan dengan penelitian ini:

1) K-Means

K-means *Clustering* yaitu teknik yang di gunakan untuk mencari data persamaan atau data yang memiliki kemiripan dan mengelompokkan data antara beberapa data dengan data lainnya. K-means adalah penyatuan atau pengelompokkan data melalui *cluster* suatu pemberhentian yang telah ditetapkan. Metode K-means berupaya menggabungkan data melalui beberapa kelompok yang memiliki ciri khas berbeda dan data yang ada ke dalam kelompok berbeda [7].

2) Data Mining

Data mining yaitu suatu pencarian informasi yang berguna dari beberapa data. Informasi berguna itu di ambil melalui beberapa proses yang sulit contohnya melalui AI atau artificial intelligence, ilmu matematika, machine learning, dan sebagainya. Metode-metode sulit berikut selanjutnya akan mencari tahu dan mencari sari informasi yang berguna dari suatu data besar [8].

3) RapidMiner

Rapidminer yaitu *software* data mining yang tidak ragu untuk di gunakan lagi dan mempunyai sistem *open-source* yang telah di gunakan banyak

ahli dan terkenal di kalangan dunia. Tersuguhkan sebagai *software* yang terbentuk mandiri untuk di pergunakan menganalisis data mining untuk koalisi melalui barang sendiri. RapidMiner Text mining Mirip dengan text analytics, yaitu proses untuk mendapatkan informasi bermutu tinggi dari teks [9].

- 4) *Centroid*
Centroid adalah nilai rata-rata aritmetik dari sebuah bentuk objek dari seluruh titik dalam objek tersebut [10].

2. Metode Penelitian

Pada penelitian ini memiliki langkah-langkah sebagai berikut:

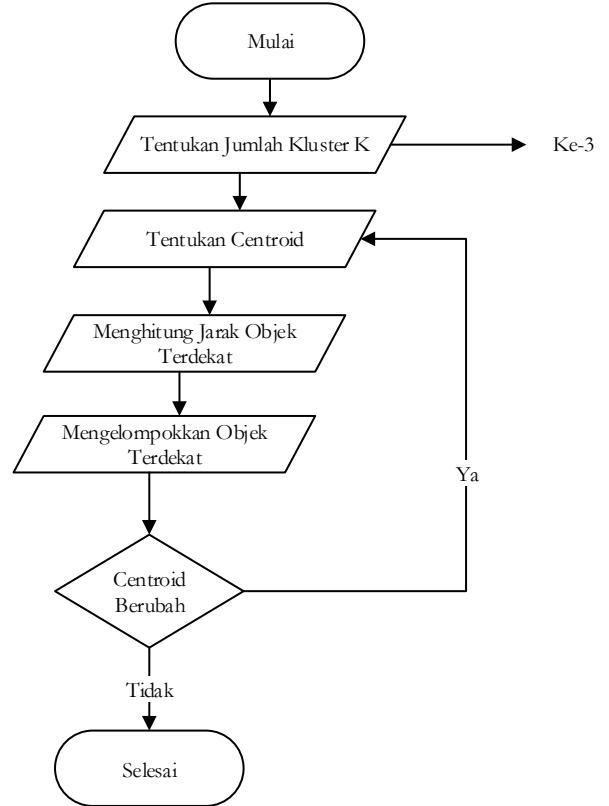
- 1) Pengumpulan Data
Pada langkah ini melakukan tahapan pengumpulan data pada beberapa orang yang beralih ke transportasi *online* Go-Jek yang bersumber melalui hasil ulasan pengguna melalui aplikasi Go-Jek di *play store* dan *website* detik.com.
- 2) pengurutan Data
Pada bagian ini penulis mengambil langkah dengan cara mengurutkan data yang akan di gunakan. Contohnya seperti menggunakan data yang telah di kumpulkan.
- 3) Menyiapkan Dataset
Pada bagian ini penulis menggunakan data yang di perlukan melalui data pelanggan angkutan umum dan pelanggan Go-jek yang di dapat melalui hasil koesioner di base came Go-jek *online* dan teman terdekat serta tambahan *website* detik.com.

Tabel 1. Data

Nama	Umur	Jangka tempuh Angkutan Umum	Jangka Tempuh Go-Jek	Tarif	Mudah didapat
Ilyas	23	6	3	3	8
Dody	22	10	6	3	10
Akmal	18	12	4	4	8
Ilham	19	6	8	4	4
Yeni	24	14	4	6	9
Inka	23	12	3	4	8
Saiba	21	10	4	4	9
Putra	19	10	3	3	3
Desy	23	4	6	3	6
Fikri	20	8	7	5	10

Berikut pada tabel 1 berisi data yang telah di dapat yaitu data umur, data tempuh angkutan umum, data tempuh Go-jek, data tarik, dan data mudah di dapat. Pada bagian ini data yang telah di dapat akan di gunakan untuk proses penelitian.

- 4) Proses Data
Pada tahapan ini lebih diperlukan karna dalam proses data mining sangat terikat kebutuhannya untuk persediaan dan pembangunan pada penetapan data awal. Initialization data dipersiapkan agar lebih gampang untuk proses klustering, karna data yang telah di siapkan berbentuk angka maka akan mempermudah dalam proses algoritma K-means.
- 5) Algoritma K-means
Melalui tahapan ini k-means yaitu metode klustering non-hirarki yang mengelompokkan data melalui beberapa kluster.



Gambar 1. Flowchart algoritma K-means

Berikut ini menggambarkan urutan-urutan dari algoritma K-means yaitu pilih *Centroid* secara acak lalu kelompokkan data *cluster* dengan titik cendtroid yang telah di pilih setelah selesai perbaharui nilai titik *Centroid* dan ulangi langkah-langkah tersebut sehingga nilai dari titik *Centroid* tidak berubah.

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

Berikut rumus persamaan jarak *Euclidean* yang berguna untuk mengecilkan antar keseluruhan *Euclidean* diberbagai tiap pemberhentian e1 dan jarak yang tidak jauh dari kluster.

3. Hasil dan Pembahasan

Pada tahapan ini penulis memakai 3 kluster serta data yang di pakai adalah data umur, data tempuh angkutan umum, jangka tempuh Go-jek, tarif dan mudah di dapat. Pada gambar tabel tentukan secara acak agar mendapatkan *Centroid*. Tahapan ini yaitu pengolahan data untuk menghasilkan informasi tentang faktor penumpang angkutan umum beralih ke transfortasi *online* Go-jek dengan cara perhitungan manual.

Tabel 2. Data Sampel

Nama	Umur	Jangka tempuh Angkutan Umum	Jangka Tempuh Go-Jek	Tarif	Mudah didapat
Ilyas	23	6	3	3	8
Dody	22	10	6	3	10
Akmal	18	12	4	4	8
Ilham	19	6	8	4	4
Yeni	24	14	4	6	9
Inka	23	12	3	4	8
Saiba	21	10	4	4	9
Putra	19	10	3	3	3
Desy	23	4	6	3	6
Fikri	20	8	7	5	10

Pada gambar sampel data yang pertu dilakukan adalah memilih jumlah *cluster* yang akan di gunakan, setelah itu tentukan *Centroid* secara acak.

Tabel 3. *Centroid* 1

C1	30	10	4	4	9
C2	26	14	4	6	9
C2	18	12	4	4	8

Pada tabel 3 di atas yaitu *Centroid* yang sudah di pilih secara acak. Untuk tahap selanjutnya yaitu hitung langkah dari tiap data ke pusat *cluster*.

$$D(e, f) = \sqrt{(25 - 30)^2 + (6 - 10)^2 + (3 - 4)^2 + (3 - 4)^2 + (8 - 9)^2} = 10$$

Perhitungan di atas adalah untuk mendapatkan hasil C1 dan hasil perhitungan persamaan di atas yang di hasilkan yaitu 10.

$$D(e, f) = \sqrt{(25 - 26)^2 + (6 - 14)^2 + (3 - 4)^2 + (3 - 6)^2 + (8 - 9)^2} = 10$$

Perhitungan di atas adalah untuk mendapatkan hasil C2 dan hasil perhitungan persamaan di atas yang di hasilkan yaitu 10.

$$D(e, f) = \sqrt{(25 - 18)^2 + (6 - 12)^2 + (3 - 4)^2 + (3 - 4)^2 + (8 - 8)^2} = 6,85$$

Perhitungan di atas adalah untuk mendapatkan hasil C3 dan hasil perhitungan persamaan di atas yang di hasilkan yaitu 6,85.

Tabel 4. Hasil perhitungan iterasi pertama

C1	C2	C3	Jarak Terdekat	Kelompok Data
10	10	6.8556546	6.8556546	C3
8.366600	6.7823299	5.3851648	5.3851648	C3
27	83	07	07	
12.20655	8.5440037	0	0	C3
56	45			
13.34166	12.569805	8.3066238	8.3066238	C3
41	09	63	63	
6	0	8.5440037	0	C2
		45		
7.416198	4.3588989	5.0990195	4.3588989	C2
49	44	14	44	
0	6	12.206555	0	C1
		62		
12.60952	10.535653	5.6585424	5.6568542	C3
02	75	9	49	
7.681145	11.090536	12.409673	7.6811457	C1
75	51	65	48	
10.72380	9.1104335	5.8309518	5.8309518	C3
53	79	95	95	

Berdasarkan hasil penjumlahan maka mendapatkan hasil yang terdekat yaitu pada C1 2 data yang terdapat pada urutan no 7 dan 9. Untuk hasil yang di dapat pada C2 yaitu 2 data yang terdapat pada urutan no 5 dan 6. Untuk hasil yang di dapat pada C3 yaitu 6 data yang terdapat pada urutan no 1,2,3,4,8,10.

Tahap selanjutnya yaitu melakukan pengulangan kembali dalam perhitungan agar memperoleh *Centroid* baru yang di dapat pada tabel 4 di atas dengan cara menghitung rata-rata melalui masing-masing anggota *clusternya* agar mendapatkan hasil seperti gambar di bawah ini:

Tabel 5. Centroid 2

C2	28.5	7	5	3.5	7.5
C2	24.5	13	3.5	5	8.5
C3	19.8333	8.66666667	5.16666667	3.66666667	7.16666667

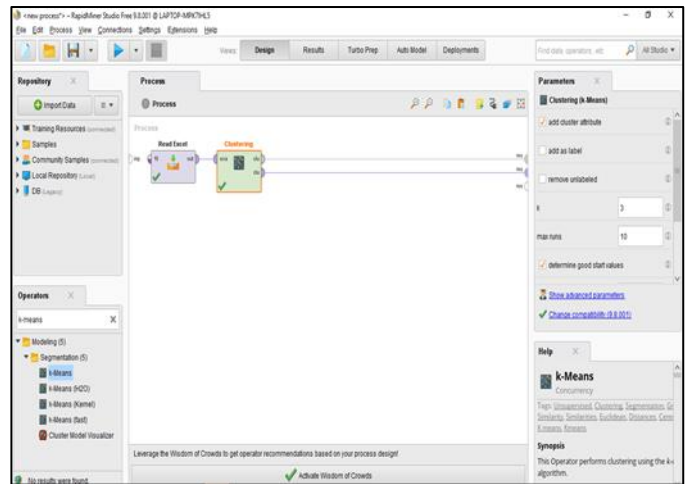
Setelah mendapatkan Centroid baru, tahap selanjutnya melakukan tahapan perhitungan kembali melalui data pada gambar table di atas yang memperoleh hasil sebagai gambar berikut:

Tabel 6. Hasil perhitungan iterasi ke dua

C1	C2	C3	Jarak Terdekat	Kelompok Data
7.858116	8.108637	3.782268	3.782268	C3
82	37	72	573	
7.664854	5.267826	3.954603	3.954603	C3
86	876	506	506	
11.69401	6.689544	4.079079	4.079079	C3
56	08	417	417	
10.61838	10.98863	5.096294	5.096249	C3
03	049	95	5	
8.046738	2.179449	8.754363	2.179449	C2
47	472	991	472	
7.729812	2.179449	5.161287	2.179449	C2
42	472	522	472	
3.840572	6.383572	10.48676	3.840572	C1
87	667	414	874	
11.12429	8.487782	4.997221	4.997221	C3
77	019	45	45	
3.840578	10.18577	8.697062	3.840572	C1
7	439	084	874	
9.260129	7.729812	3.693086	3.693086	C3
59	417	634	634	

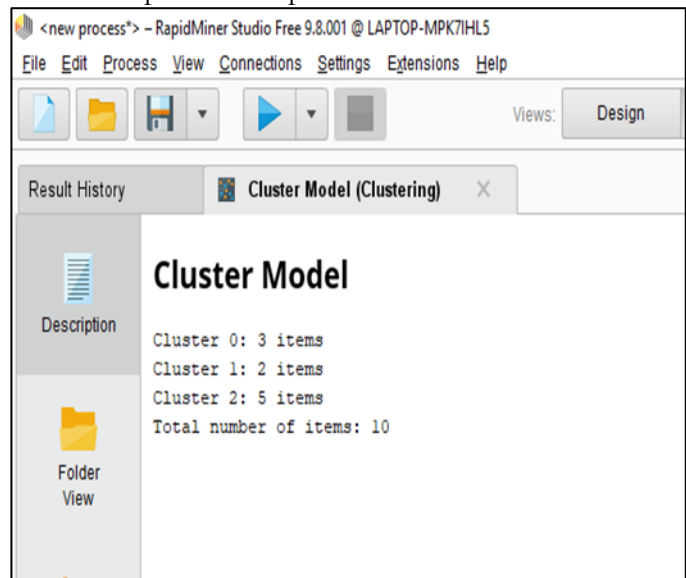
Berdasarkan hasil penjumlahan maka mendapatkan hasil yang sama dengan iterasi pertama yaitu pada C1 2 data yang terdapat pada urutan no 7 dan 9. Untuk hasil yang di dapat pada C2 yaitu 2 data yang terdapat pada urutan no 5 dan 6. Untuk hasil yang di dapat pada C3 yaitu 6 data yang terdapat pada urutan no 1,2,3,4,8,10. Pada tahapan ini peneliti menggunakan iterasi sejumlah 2 tahap. Karena hasil yang di dapat pada iterasi ke 2 tetap sama dengan iterasi ke 1 atau tidak ada data yang berubah oleh karena itu maka perhitungan akan di stop atau di sudahi.

Tahapan Selanjutnya penulis melakukan clustering menggunakan software rapidminer. Berikut seperti gambar di bawah ini:



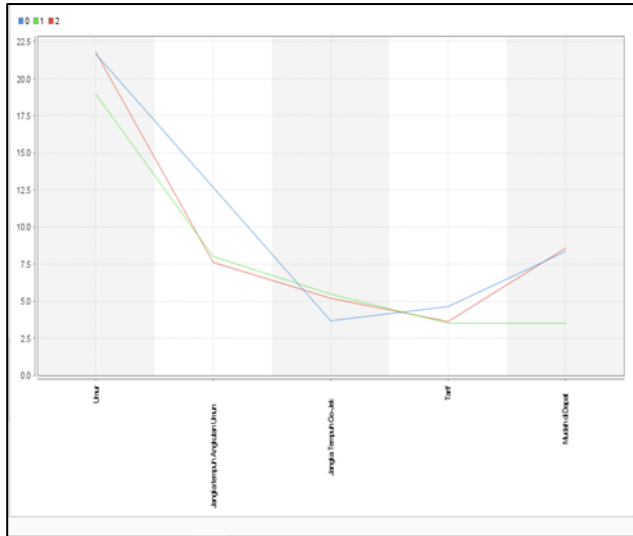
Gambar 2. Pemodelan K-means dalam Rapidminer

Pada gambar di atas penulis melakukan tahapan memasukan data yang sudah dibentuk dalam microsoft excel lalu menambahkan clustering K-means dan dalam parameters penulis memilih 3 cluster.



Gambar 3. Hasil implementasi K-means clustering RapidMiner

Dari hasil implementasi K-means clustering rapidminer yaitu pada cluster 0 terdapat 3 items, untuk cluster 1 terdapat 2 item dan cluster 2 terdapat 5 item dengan total number item yaitu 10.



Gambar 4. Plot view pada rapidminer

Pada gambar plot view di atas menunjukkan faktor peralihan umur memiliki nilai tinggi, sedangkan paling kecil di dapat pada faktor tarif. Statistik penyebaran antara kluster terbukti berbeda dikarenakan tiap kluster berbeda dalam jumlah *clusternya*.

4. Kesimpulan dan Saran

Berdasarkan hasil penelitian yang di lakukan maka dapat di simpulkan bahwa analisis faktor yang mempengaruhi penumpang angkutan umum beralih ke transportasi *online* Go-jek menggunakan metode K-means *clustering* menghasilkan perhitungan secara manual dan melalui aplikasi rapidminer tidak jauh berbeda. Berdasarkan pengujian terhadap hasil perhitungan K-means *clustering* yaitu pada ketiga kluster faktor peralihan umur memiliki nilai paling tinggi untuk faktor peralihan jangka tempuh angkutan umum yaitu di posisi ke tiga, faktor peralihan jangka tempuh Go-jek di posisi ke 5, faktor peralihan tarif berada di posisi 4 dan faktor peralihan mudah di dapat berada pada posisi ke 2.

Berdasarkan hasil analisis yang telah dilakukan pada analisis faktor yang mempengaruhi penumpang angkutan umum beralih ke transportasi *online* Go-jek menggunakan algoritma K-means klustering maka diharapkan untuk peneliti selanjutnya untuk menguji dengan algoritma *clustering* lainnya, serta *software* rapidminer yang digunakan sebagai bahan penelitian bisa dikembangkan lagi menjadi lebih banyak fitur lainnya.

5. Daftar Pustaka

- [1] Rendy, Y., 2018. Analisis Faktor-Faktor Yang Mempengaruhi Permintaan terhadap Ojek *Online* (Studi Kasus Pada Go-Jek di Kota Malang). *Jurnal Ilmiah Mahasiswa FEB*, 7(1).
- [2] Lestari, A.D., 2018. *Analisis Multivariat Clustering K-Means Pada Kabupaten/Kota Di Provinsi Jawa Tengah Berdasarkan Indeks Pembangunan Manusia Tahun 2017 Dengan Bantuan Software SPSS* (Doctoral dissertation, Universitas Negeri Semarang).
- [3] Sundari, S., Damanik, I.S., Windarto, A.P., Tambunan, H.S., Jalaluddin, J. and Wanto, A., 2019, September. Analisis K-Medoids Clustering Dalam Pengelompokan Data Imunisasi Campak Balita Di Indonesia. In *Prosiding Seminar Nasional Riset Information Science (SENARIS)* (Vol. 1, pp. 687-696).
- [4] Ningrat, D.R., Di Asih, I.M. and Wuryandari, T., 2016. Analisis Cluster Dengan Algoritma K-Means Dan Fuzzy C-Means Clustering Untuk Pengelompokan Data Obligasi Korporasi. *Jurnal Gaussian*, 5(4), pp.641-650.
- [5] Ramdani, A.L. and Firmansyah, H.B., 2018. Clustering Application for UKT Determination Using Pillar K-Means Clustering Algorithm and Flask Web Framework. *Indonesian Journal of Artificial Intelligence and Data Mining*, 1(2), pp.53-59.
- [6] Haris, A. and Hendrian, E., 2019. Sistem Monitoring dan Klaster Ketersediaan Energi Menggunakan Metode K-Means pada Pembangkit Listrik Tenaga Surya. *CESS (Journal of Computer Engineering, System and Science)*, 4(2), pp.266-271.
- [7] Mulaki, S.F., Setiyawati, N. and Wijaya, A.F., 2018. Analisis Data Mahasiswa Menggunakan Algoritma K-Means Clustering sebagai Dasar Pelaksana Promosi. *JBASE-Journal of Business and Audit Information Systems*, 1(2).

- [8] Bootupacademyai, 2021. Lengkap, Data Mining Adalah? Pengertian Hingga Belajar *Clustering*, bootup.ai, 2019. <https://bootup.ai/blog/data-mining-adalah/> (accessed Jun. 04, 2021).
- [9] Devipursitasari, 2021. Eksplorasi Data Mining Software (Rapid Miner), iMe (iLearning Media), 2013. <https://ilearning.me/2013/10/08/eksplorasi-data-mining-software-rapid-miner/> (accessed Jun. 04, 2021).
- [10] Fadhilah, A.M., Wahyuddin, M.I. and Hidayatullah, D., 2020. Analisis Faktor yang Mempengaruhi Perokok Beralih ke Produk Alternatif Tembakau (VAPE) menggunakan Metode K-Means Clustering. *Jurnal JTIIK (Jurnal Teknologi Informasi dan Komunikasi)*, 5(2), pp.219-225.
- [11] Atmaja, E.H.S., 2019. Implementation of k-Medoids Clustering Algorithm to Cluster Crime Patterns in Yogyakarta. *International Journal of Applied Sciences and Smart Technologies*, 1(1), pp.33-44.
- [12] M, B, Tomy J., A., U, & Jacob, p., 2011. K-Means *Clustering* Student Data to Characterize Performance Patterns. *International Journal of Advanced Computer Science and Applications*, 1(3), 138–140
- [13] Durairaj, M. and Vijitha, C., 2014. Educational data mining for prediction of student performance using clustering algorithms. *International Journal of Computer Science and Information Technologies*, 5(4), pp.5987-5991.