

Klasterisasi Data Penggunaan Layanan Bpjs Kesehatan

Menggunakan Algoritma K-Means

Clustering of Data on the Use of BPJS Health Services Using the K-Means Algorithm

Sandi Salvan N¹, Wahyu Hadikristanto², Edora³

^{1,2,3}Program Studi Teknik Informatika, Fakultas Teknik, Universitas Pelita Bangsa

¹sandi9@mhs.pelitabangsa.ac.id*, ²wahyu.hadikristanto@pelitabangsa.ac.id *, ³edora@pelitabangsa.ac.id

Abstract

BPJS Health, which is organized by the government by upholding the principle of mutual assistance in equalizing public health insurance, many patients use these facilities. The clustering method is processed with the K-Means algorithm, where the results also show a new insight, namely grouping the use of BPJS health services based on 3 clusters. Cluster 1 is a category of health facilities with low or Low use of BPJS health services, which is 353 out of 1000 categories of health facilities based on the number of BPJS health service usages tested, then cluster 2 is a category of health facilities with moderate or Medium use of BPJS health services, which is 474 out of 1000 categories. the name of the health facility based on the number of health BPJS service usage tested, and lastly cluster 3 is the category of health facility name with the use of BPJS health services quite high or High, namely 173 out of 1000 categories of health facility names based on the number of health BPJS service usage tested. Tests using Rapid Miner tools can also produce similar insights, namely each cluster has cluster group members according to manual calculations such as Cluster_0 on Rapid Miner has 474 cluster members representing the Medium cluster, Cluster_1 has 353 cluster group members as the Low cluster representation, and Cluster_2 has 173 cluster members that correspond to the cluster representation of High.

Keywords: Data Mining, K-Means, BPJS Health

Abstrak

BPJS Kesehatan yang diselenggarakan pemerintah dengan menjunjung prinsip gotong royong dalam penyetaraan jaminan kesehatan masyarakat, banyak pasien yang menggunakan fasilitas tersebut. Metode klasterisasi tersebut diproses dengan algoritma K-Means yang dimana hasilnya juga menunjukkan sebuah wawasan baru yaitu pengelompokan penggunaan layanan BPJS kesehatan berdasarkan 3 cluster. Cluster 1 merupakan kategori nama faskes dengan penggunaan layanan BPJS kesehatan rendah atau Low yaitu 353 dari 1000 kategori nama faskes berdasarkan jumlah penggunaan layanan BPJS kesehatan yang diuji, kemudian cluster 2 adalah kategori nama faskes dengan penggunaan layanan BPJS kesehatan sedang atau Medium yaitu 474 dari 1000 kategori nama faskes berdasarkan jumlah penggunaan layanan BPJS kesehatan yang diuji, dan terakhir adalah cluster 3 merupakan kategori nama faskes dengan penggunaan layanan BPJS kesehatan cukup tinggi atau High yaitu 173 dari 1000 kategori nama faskes berdasarkan jumlah penggunaan layanan BPJS kesehatan yang diuji. Pengujian menggunakan tools Rapid Miner juga dapat menghasilkan wawasan yang serupa yaitu masing-masing klaster memiliki anggota kelompok klaster sesuai dengan perhitungan manual seperti Cluster_0 pada Rapid Miner memiliki 474 anggota klaster yang merepresentasikan klaster Medium, Cluster 1 memiliki 353 anggota kelompok klaster sebagai representasi klaster Low, dan Cluster_2 memiliki 173 anggota klaster yang sesuai dengan representasi klaster High.

Kata Kunci: Data Mining, K-Means, BPJS Kesehatan

Pendahuluan

BPJS Kesehatan yang diselenggarakan pemerintah dengan menjunjung prinsip gotong royong dalam penyetaraan jaminan kesehatan masyarakat, banyak pasien yang menggunakan fasilitas tersebut. Pentingnya asuransi kesehatan telah mendorong pemerintah Indonesia untuk membentuk badan Penyelenggara

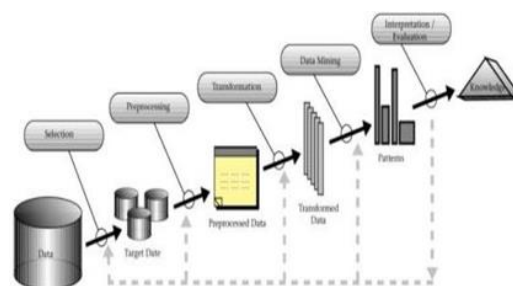
Jaminan Sosial (BPJS) Kesehatan, suatu badan hukum publik yang memiliki tugas untuk menyelenggarakan Jaminan Kesehatan Nasional (JKN) bagi rakyat Indonesia [1][2].

Jaminan Sosial (BPJS) Kesehatan menjadi salah satu layanan penting bagi masyarakat. Karena, BPJS Kesehatan adalah badan hukum yang dibentuk untuk menyelenggarakan program jaminan kesehatan. Oleh karena itu, bagi masyarakat yang mengalami sakit dan telah terdaftar sebagai peserta Jaminan. Kesehatan Nasional (JKN), biaya pengobatannya akan ditanggung oleh pihak BPJS, sehingga penting bagi BPJS untuk mengetahui faskes yang sesuai dengan standar rekredensialing, yaitu suatu kegiatan kualifikasi dan evaluasi yang dilakukan oleh BPJS untuk menyetujui atau menolak faskes untuk bekerjasama kembali dengan berdasarkan aspek administrasi dan teknis pelayanan. [3]Hal ini berguna untuk menghindari penerimaan kerjasama dengan faskes yang tidak bermutu berdasarkan aspek administrasi dan teknis pelayanan. Namun terkadang dalam pengambilan keputusan untuk bekerjasama Kembali dengan BPJS sering terjadi kesulitan karena belum adanya pengelompokan penyedia layanan faskes yang dapat membedakan penyedia layanan faskes yang seharusnya dapat bekerjasama Kembali [4][5].

Berdasarkan uraian masalah diatas maka penelitian ini menggunakan salah satu algoritma k- means untuk menentukan klasterisasi data dan dilihat dari penelitian sebelumnya sehingga memiliki nilai atau novelty sebagai unsur pembaruan dari penelitian ini dapat dilihat dari data atau atribut yang akan diuji sebagai unsur informasi pengetahuan dalam menghasilkan hasil cluster.

Metode Penelitian

Metode yang digunakan dalam penelitian ini adalah metode scientific yang disebut dengan Knowledge Discovery in Database (KDD).[6]



Gambar 1 Knowledge Discovery in Database

Gambar 1 di atas menjelaskan Tahapan metode KDD yaitu *Selection*, *Preprocessing*, *Transformation*, *Data Mining*, *Interpretation/Evaluation*. Tahap *Selection* merupakan menyeleksi data dari multiple data source dengan membersihkan data yang memiliki noise ataupun missing value. Tahap berikutnya yaitu *preprocessing*, pada tahap ini dilakukan penggabungan seluruh data *source* yang telah terkumpul sebelum tahapan dari proses selanjutnya dilakukan. Setelah tahap *preprocessing* [7], Langkah selanjutnya yaitu *transformation*, dimana dilakukan transformasi data ke dalam bentuk yang lebih sesuai untuk dilakukan data mining. Langkah keempat *Data Mining*, tahapan ini diterapkannya metode data mining pada sistem, sehingga dapat menghasilkan pola dari data yang terkumpul. Pemodelan data pada penelitian ini dilakukan dengan metode klasterisasi menggunakan algoritma K-Means. Urutan langkah yang dilakukan dalam menggunakan algoritma K- Means adalah sebagai berikut [8][9][10] :

1. Menentukan jumlah cluster yang akan dipakai, dalam penelitian ini dipakai jenis cluster, antara lain *Low Demand* (C1) untuk tingkat produksi rendah, *Medium Demand* (C2) untuk tingkat sedang, dan *High Demand* (C3) untuk tingkat tinggi.
2. Menentukan nilai centroid pada tahap awal untuk iterasi ke- 0, dilakukan secara random dengan rumus menentukan target awal k-means untuk mendapatkan target data atau jarak antar kelompok, menggunakan rumus berikut:

$$V_{ij} = \frac{1}{N_i} \sum_{k=0}^{N_i} \chi^{kj}$$

Tahap terakhir yaitu *Interpretation/Evaluation*. Interpretasi dan evaluasi dari pola yang didapatkan, sehingga dapat diidentifikasi apakah pola tersebut sudah dapat mewakili knowledge yang ingin dicapai [11].

Evaluasi. Melakukan pengecekan terhadap setiap nilai atribut dan model yang sudah dibangun. Kemudian melakukan evaluasi dengan cara mengamati dan menganalisa hasil dari algoritma yang digunakan untuk memastikan bahwa hasil pengujian menghasilkan nilai C1,C2,C3 benar dan sesuai hasil pembahasan, pengujian dilakukan untuk mengukur hasil dari tiap model yang diusulkan [12].

Data mining adalah proses untuk menemukan korelasi, pola, dan tren baru yang bermakna dengan memilah-milah data dalam jumlah besar yang disimpan di dalam repositori, menggunakan teknologi pengenalan pola serta teknik statistik dan matematika [13]. Data mining adalah proses menemukan pola dan pengetahuan menarik dari data dalam jumlah yang besar [14].

Clustering mengacu pada pengelompokan seperti record, pengamatan, atau memperhatikan dan membentuk kelas objek-objek yang memiliki kemiripan. Cluster adalah kumpulan dari record yang memiliki kemiripan satu sama lain, dan berbeda dengan record di klaster lain. Clustering mencoba untuk membagi seluruh kumpulan data menjadi kelompok-kelompok yang relatif memiliki kemiripan, di mana kemiripan record dalam satu kelompok akan bernilai maksimal, sedangkan kemiripan dengan record dalam kelompok lain akan bernilai minimal [15]. Clustering dalam data mining berguna untuk menemukan pola distribusi di dalam sebuah data set yang berguna untuk proses analisa data. Kesamaan objek biasanya diperoleh dari kedekatan nilai-nilai atribut yang menjelaskan objek-objek data, sedangkan objek-objek data biasanya direpresentasikan sebagai sebuah titik dalam ruang multidimensi [16].

Hasil dan Pembahasan

Penelitian ini menggunakan algoritma k-means untuk menganalisa pengelompokan penggunaan layanan BPJS kesehatan. Sumber data sebagai objek pada penelitian ini adalah data yang diambil dari situs Kaggle.com. Data yang digunakan dalam penelitian ini terdiri dari atribut atau variabel seperti Nama Faskes, NoLink, dan NoData [12].

Tabel 1 Dataset Penggunaan Layanan BPJS Kesehatan

Nama Faskes	NoLink	NoData
RSUD Gayo Lues	1	1
Pintu Rime Uring	1	1
Rerebe	1	2
Da Bu Glang	1	3
Blang Jerango	1	4
Cinta Maju	1	5
Gumpang	1	6
Blang Kejeren Galus	1	7
Pinding	1	8
Kuta Panjang	1	9

Setelah mendata pengguna layanan BPJS Kesehatan (Tabel 1). Selanjutnya yaitu mengimplementasi algoritma K-Means dengan menggunakan dataset yang akan diolah yaitu sebanyak 1000 yang akan dikelompokkan kedalam tiga cluster yaitu low, medium, dan high. Dengan pemodelan yang sudah ditetapkan sebelumnya, maka berikut adalah contoh untuk perhitungan Euclidean Distance pada record ke 1 dan 1000 dari proses iterasi pertama.

- Data Ke- 1
 $D1(c1) = \sqrt{(1-26)^2 + (1-7)^2} = 26$
 $D1(c2) = \sqrt{(1-14)^2 + (1-7)^2} = 14$
 $D1(c3) = \sqrt{(1-5)^2 + (1-8)^2} = 8$
- Data Ke- 1000
 $D1000(c1) = \sqrt{(30-26)^2 + (10-7)^2} = 5$
 $D1000(c2) = \sqrt{(30-14)^2 + (10-7)^2} = 16$
 $D1000(c3) = \sqrt{(30-5)^2 + (10-8)^2} = 25$

Tabel 2 berikut ini yaitu hasil dari perhitungan *Euclidean Distance* pada tahap iterasi pertama.

Tabel 2 Euclidean Distance Iterasi 1

KC1	C2	C3	Jarak Terpendek
26	14	8	8
26	14	8	8
26	14	7	7
26	13	6	6
26	13	6	6
25	13	5	5
25	13	5	5
25	13	4	4
25	13	4	4
25	13	4	4
....
4	16	25	4
4	16	25	4
4	16	25	4
4	16	25	4
5	16	25	5

Selanjutnya jarak dari hasil perhitungan akan dilakukan perbandingan diantara 3 kluster data terdekat dengan pusat kluster dengan mengambil nilai terkecil, jarak ini menunjukkan bahwa data tersebut berada dalam 1 kelompok dengan pusat kluster terdekat. Berikut adalah matriks kelompok kluster yang terjadi pada iterasi ke-1, dilambangkan dengan simbol "T".

Tabel 3 Matriks Kelompok Data Iterasi ke-1

C1	C2	C3
		T
		T
		T
		T
		T
		T
		T
		T
		T
		T
....
T		
T		
T		
T		
T		
337	310	353

Langkah berikutnya perlu ditentukan kembali titik kluster baru yang dihitung dengan mencari nilai rata-rata berdasarkan dari data anggota masing-masing kelompok kluster. Berikut adalah perhitungan untuk penentuan nilai titik kluster baru dari proses iterasi pertama:

$$\begin{aligned}
 \text{C1 Baru ke 2 (NoLink)} &= \frac{8845}{337} = 26 \\
 \text{C1 Baru ke 2 (NoData)} &= \frac{2517}{337} = 7 \\
 \text{C2 Baru ke 2 (NoLink)} &= \frac{4350}{310} = 14 \\
 \text{C2 Baru ke 2 (NoData)} &= \frac{1564}{310} = 5 \\
 \text{C3 Baru ke 2 (NoLink)} &= \frac{1947}{353} = 6 \\
 \text{C3 Baru ke 2 (NoData)} &= \frac{3175}{353} = 9
 \end{aligned}$$

Lakukan perhitungan seperti di atas sampai posisi objek sudah tidak ada yang berubah. dari perhitungan diatas, untuk memenuhi syarat indeks kelompok masing-masing cluster memiliki kemiripan jumlah objek seperti pada iterasi sebelumnya, maka perhitungan iterasi berhenti di iterasi ke-11. Untuk lebih jelasnya perbandingan indeks jumlah objek data pada masing-masing cluster dapat dilihat pada table berikut:

Tabel 4 Perbandingan Jumlah Objek Data dari Iterasi ke 10 & ke 11

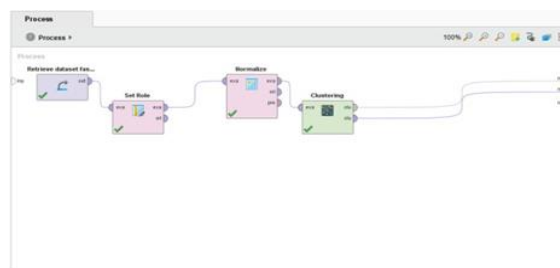
Proses Iterasi	Jml_Objek C1	Jml_Objek C2	Jml_Objek C3
Iterasi ke 10	353	474	173
Iterasi ke 11	353	474	173

Dari proses pra-processing data didapatkan 1000 Data yang akan diproses menggunakan algoritma K-Means. Melalui beberapa tahapan seperti pada bagian pemodelan data, didapatkan hasil bahwa proses clustering dengan algoritma K-Means berhenti pada iterasi ke-11, karena posisi objek dari masing – masing cluster sudah tidak berubah dan mendapatkan nilai yang optimal. Berikut adalah bentuk cluster yang didapat:

1. Cluster pertama memiliki pusat klaster (26, 8), sehingga dapat diartikan pada cluster ini adalah kelompok yang memiliki penggunaan layanan BPJS kesehatan dengan tingkat rendah atau berkategori Low. Ada 353 jumlah penggunaan layanan BPJS kesehatan dari kategori nama faskes yang masuk dalam klaster pertama.
2. Cluster kedua memiliki pusat klaster (10, 4), sehingga dapat diartikan pada cluster ini adalah kelompok yang memiliki penggunaan layanan BPJS kesehatan dengan tingkat sedang dan cenderung stabil atau berkategori Medium. Ada 474 jumlah penggunaan layanan BPJS kesehatan dari kategori nama faskes yang masuk dalam klaster kedua.
3. Cluster ketiga memiliki pusat (7, 15), yang dimana dapat diartikan pada cluster ini adalah kelompok yang memiliki penggunaan layanan BPJS kesehatan dengan tingkat tinggi atau berkategori High. Ada 173 jumlah penggunaan layanan BPJS kesehatan dari kategori nama faskes yang masuk dalam klaster ketiga.

Proses pengujian pada penelitian ini menggunakan *tools rapidminer*. Hasil pengujian yang didapat dengan menggunakan *tools rapidminer* adalah dengan tahapan langkah - langkah sebagai berikut:

1. Melakukan import data yang diperlukan untuk proses pada tools rapidminer. Pada aplikasi rapid miner pilih dan klik Import Data, kemudian pilih data yang akan dipakai serta kemudian menentukan attribut dan label yang akan digunakan.
2. Klik menu Design, pada tampilan proses, tambahkan dataset pada folder ke layar tampilan proses. Pada menu Names & Roles cari fungsi Set Role yang nantinya akan dipakai untuk mengatur role attribut, kemudian drag ke layar tampilan proses.
3. Selanjutnya pada menu Normalization pilih Normalize dan drag ke layar tampilan proses, melalui fungsi ini dapat mengatur normalisasi data yang akan dilakukan dari dataset yang dipakai pada proses ini.
4. Kemudian pada menu Modelling, dalam submenu Segmentation, pilih fungsi k-Means, untuk menerapkan algoritma k-means terhadap proses klasterisasi yang akan dilakukan.
5. Koneksikan semua perintah tersebut sehingga pada layar tampilan proses terlihat alur sebagai berikut:



Gambar 2 Proses RapidMiner Klasterisasi K- Means

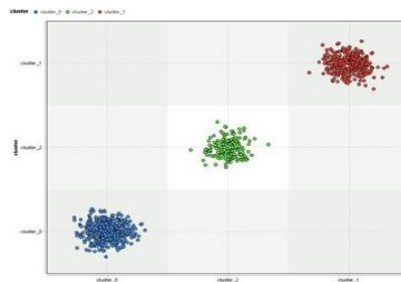
6. Setelah dilakukan Running Process pada tools rapidminer, didapatkan hasil pengelompokan klaster terhadap 1000 record data yang diproses dan dapat dilihat pada gambar berikut:

Cluster Model

Cluster 0: 474 items
 Cluster 1: 353 items
 Cluster 2: 173 items
 Total number of items: 1000

Gambar 3 Cluster Model dari 1000 Record Data

7. Dapat dilihat melalui hasil cluster model yang terbentuk terhadap 3 cluster yang sudah ditentukan sebelumnya, tiap masing-masing cluster memiliki anggota cluster (wilayah) yang sesuai dengan kelompoknya.
8. Adapun bentuk visualisasi menggunakan scatter plot untuk masing – masing cluster juga dapat dilihat pada gambar berikut:



Gambar 4 Grafik Scatter Plot dari Klaster yang Terbentuk

Setelah melakukan tahapan dalam mencari kluster penggunaan layanan BPJS kesehatan melalui metode klusterisasi, pemanfaatan algoritma K-Means yang digunakan menghasilkan suatu pengelompokan kluster terhadap masing-masing nama faskes. Melalui beberapa tahapan didapatkan hasil bahwa proses clustering dengan algoritma K-Means berhenti pada iterasi ke-11, karena posisi objek dari masing – masing cluster sudah tidak berubah dan mendapatkan nilai yang optimal. Berikut adalah bentuk cluster yang didapat yaitu:

1. Cluster pertama memiliki pusat kluster (26, 8), dan terdiri dari 353 jumlah penggunaan layanan BPJS kesehatan dari kategori nama faskes yang masuk dalam kluster pertama.
2. Cluster kedua memiliki pusat kluster (10, 4), dari hasil perhitungan data yang digunakan, ada 474 jumlah penggunaan layanan BPJS kesehatan dari kategori nama faskes yang masuk pada kluster ini.
3. Cluster ketiga memiliki pusat (7, 15), pada kluster ini memiliki jumlah penggunaan layanan BPJS kesehatan dari kategori yang paling sedikit diantara 2 kluster lainnya yaitu ada 173 jumlah penggunaan layanan BPJS kesehatan dari kategori nama faskes.

Pada pengujian dengan tools Rapid Miner juga didapatkan hasil yang serupa. Dalam Rapid Miner ditentukan sebanyak 3 kluster untuk pembentukan kelompok kluster penggunaan layanan BPJS kesehatan yang ada dari dataset sebanyak 1000 data yang digunakan. Untuk Rapid Miner sendiri penamaan kluster dimulai dari Cluster_0, Cluster_1, dan Cluster_2. Anggota dari masing-masing cluster juga memiliki kemiripan dengan perhitungan manual yang dilakukan. Hanya saja dalam proses menggunakan tools Rapid Miner tidak ditentukan nilai cluster awal sebagaimana yang dilakukan dalam proses perhitungan manual. Namun untuk hasil yang didapatkan tidak jauh berbeda yaitu sebagai berikut:

1. Cluster_0 terdiri dari 474 anggota kelompok kluster penggunaan layanan BPJS kesehatan yang dimana ini sesuai dengan kelompok kluster Medium pada tahap perhitungan manual.
2. Cluster_1 terdiri dari 353 anggota kelompok kluster penggunaan layanan BPJS kesehatan yang dimana ini sesuai dengan kelompok kluster Low pada tahap perhitungan manual.
3. Cluster_2 terdiri dari 173 anggota kelompok kluster penggunaan layanan BPJS kesehatan yang dimana ini sesuai dengan kelompok kluster High pada tahap perhitungan manual.

Kesimpulan

Dari penelitian di atas maka dapat disimpulkan bahwa hasil analisa pengguna BPJS dapat menghasilkan Cluster 1 merupakan kategori nama faskes dengan penggunaan layanan BPJS kesehatan rendah atau Low yaitu 353 dari 1000 kategori nama faskes berdasarkan jumlah penggunaan layanan BPJS kesehatan yang diuji, kemudian cluster 2 adalah kategori nama faskes dengan penggunaan layanan BPJS kesehatan sedang atau Medium yaitu 474 dari 1000 kategori nama faskes berdasarkan jumlah penggunaan layanan BPJS kesehatan

yang diuji, dan terakhir adalah cluster 3 merupakan kategori nama faskes dengan penggunaan layanan BPJS kesehatan cukup tinggi atau High yaitu 173 dari 1000 kategori nama faskes berdasarkan jumlah penggunaan layanan BPJS kesehatan yang diuji. Pengujian menggunakan algoritma k- means dan tools Rapid Miner juga dapat menghasilkan wawasan yang serupa yaitu masing-masing kluster memiliki anggota kelompok kluster sesuai dengan perhitungan manual seperti Cluster_0 pada Rapid Miner memiliki 474 anggota kluster yang merepresentasikan kluster Medium, Cluster_1 memiliki 353 anggota kelompok kluster sebagai representasi kluster Low, dan Cluster_2 memiliki 173 anggota kluster yang sesuai dengan representasi kluster Hig

Daftar Rujukan

- [1] D. Fithaloka, M. A. Budiman, and D. Rachmawati, (2017), "Perbandingan Algoritma Greedy dan Hill Climbing Untuk Menentukan Fasilitas Kesehatan Tingkat Pertama (FKTP) Terdekat Bagi Peserta BPJS Kesehatan," vol. 1, no. 2, pp. 13–24.
- [2] S. S. Helma, R. R. R, and E. Normala, (2019), "Clustering pada Data Fasilitas Pelayanan Kesehatan Kota Pekanbaru Menggunakan Algoritma K - Means," no. November, pp. 131– 137.
- [3] J. Wandana, and S. Defit, (2020), "Klasterisasi Data Rekam Medis Pasien Pengguna Layanan BPJS Kesehatan Menggunakan Metode K-Means," *J. Informatika*, vol. 2, pp. 4– 9, doi: 10.37034/jidt.v2i4.73.
- [4] J. Tamaela, E. Sediyo, and A. Setiawan, (2018), "Implementasi Metode Association Rule untuk Menganalisis Data Twitter tentang Badan Penyelenggara Jaminan Sosial dengan Algoritma Frequent Pattern - Growth," vol. 01, pp. 25–33.
- [5] G. R. P, A. P. Windarto, E. Irawan, and W. Saputra, (2020), "Penerapan Data Mining Menggunakan Algoritma C4 . 5 Dalam Mengukur Tingkat Kepuasan Pasien BPJS," vol. 2, pp. 376–385.
- [6] G. Widi N. Dicky Nofriansyah, (2015), Algoritma Data Mining Dan pengujian. *Yogyakarta: Cv Budi Utama*.
- [7] K. D. R. Sianipar, S. W. Siahaan, M. Siregar, and P. P. P. A. N. W. F. I. R. H. Zer, (2020), "Penerapan Algoritma K-Means Dalam Menentukan Tingkat Kepuasan Pembelajaran Online Pada Masa Pandemi Covid-19," vol. 4, no. 1, pp. 101–105.
- [8] Link Data BPJS Di Kaggle.com (<https://www.kaggle.com/israhabibi/list-faskes-bpjs-indonesia>).
- [9] Retno Tri vulandari, (2017), Data Mining. *Yogyakarta: Gava Media*.
- [10] Romli, I. (2021). Penerapan Data Mining Menggunakan Algoritma K-Means Untuk Klasifikasi Penyakit Ispa. *Indonesian Journal of Business Intelligence (IJUBI)*, 4(1), 10-15.
- [11] T. M. Tamtelahitu, 2020, "Komparasi Algoritma Clustering Dengan Dataset Penyebaran ISSN : 1979- 8415," *P. S. Informatika, U. Kristen. Maluku* vol. 13, no. 1, pp. 27–34.
- [12] Suyanto, (2017), Data Mining. *Yogyakarta: Informatika*.
- [13] D. T. Larose And C. D. Larose, (2014), Discovering Knowledge In Data An Introduction To Data Mining.
- [14] S. Susanto And D. Suryadi, (2010), Pengantar Data Mining Menggali Pengetahuan Bongkahan Data.
- [15] A. F. Muhammad, (2015), "Klasterisasi Proses Seleksi Pemain Menggunakan Algoritma K-Means," Pp. 1–5,
- [16] A. Wanto Et Al., (2020), Data Mining Algoritma Dan Implementasi.