

PERBANDINGAN ALGORITMA K-MEANS DAN K-MEDOIDS DALAM KLASTERISASI PRODUK ASURANSI PERUSAHAAN NASIONAL

Frenda Farahdinna¹, Irfan Nurdiansyah², Apriati Suryani³, Arief Wibowo⁴

^{1,2,3,4}Program Studi Magister Ilmu Komputer, Universitas Budi Luhur
Jl.Raya Ciledug, Petukangan Utara, Jakarta Selatan 12260 Indonesia
¹fardinna5@gmail.com, ²irfan.nurdiansyah26@gmail.com,
³suryaniaprianti@gmail.com, ⁴arief.wibowo@budiluhur.ac.id

Abstract

Insurance has an important role in transferring risks, both business and non-business risks. insurance is an economic institution that aims to reduce risk, by combining a number of units that have the same or almost the same risk in sufficient quantities, so that the probability can be channeled to the units at risk. In insurance there is the term premium, where the premium is under the rights of the insurance owner in the payment of the premium. The national insurance company that is the object of this research provides various insurance products that have several variables such as premium value, number of customers and the year of release of the product. A technique is needed that can classify insurance products with the aim of making it easier for companies to see superior products and choose products that suit customer needs. The purpose of this study is to analyze and process data by comparing the K-Means and K-Medoids methods in the clustering of national company insurance products. Based on research that has been done, when compared to the K-Means method with K-Medoids, the K-Means method is more optimal with a Davies Boldin Index (DBI) value of 0.018 with $k=5$.

Keyword: *K-Means, K-Medoids, Clustering, Insurance Products, Davies Bouldin Index*

Abstrak

Asuransi mempunyai peran penting dalam rangka mengalihkan resiko, baik resiko bisnis maupun non bisnis. Asuransi merupakan suatu lembaga ekonomi yang bertujuan mengurangi resiko, dengan menggabungkan sejumlah unit-unit yang mempunyai resiko sama atau hampir sama dalam jumlah memadai, agar probabilitas dapat disalurkan kepada unit yang mengalami resiko. Dalam asuransi terdapat istilah premi, dimana premi tersebut berada pada hak dari pemilik asuransi dalam pembayaran premi tersebut. Perusahaan asuransi nasional yang menjadi obyek penelitian ini menyediakan berbagai produk asuransi yang memiliki beberapa variabel seperti nilai premi, jumlah nasabah dan tahun rilis dari produk. Diperlukan suatu teknik yang dapat mengelompokan produk-produk asuransi dengan tujuan mempermudah perusahaan dalam melihat produk-produk unggulan dan memilih produk yang sesuai dengan kebutuhan nasabah. Tujuan penelitian ini yaitu melakukan analisis dan pengolahan data dengan membandingkan metode K-Means dan K-Medoids dalam klasterisasi produk asuransi perusahaan nasional. Berdasarkan penelitian yang telah dilakukan, jika dibandingkan metode K-Means dengan K-Medoids maka metode K-Means lebih optimal dilihat dari nilai Davies Boldin Index (DBI) terendah 0,018 pada $k=5$.

Kata Kunci: *K-Means, K-Medoids, Klasterisasi, Produk Asuransi, Davies Bouldin Index*

I. Pendahuluan

Dalam masyarakat modern, asuransi memegang peran penting dalam rangka mengalihkan resiko, baik resiko bisnis maupun non bisnis. Karena dalam perjalanan kehidupan yang pasang surut membuat adanya sarana pengamanan bagi para pelakunya, agar kerugian yang diderita karena resiko (kerugian, kehilangan aset, kecelakaan dan lain-lain) dapat dialihkan kepada pihak ketiga sebagai penanggung.

Asuransi adalah suatu lembaga ekonomi yang bertujuan untuk mengurangi resiko, dengan menggabungkan sejumlah unit-unit yang mempunyai resiko yang sama atau hampir sama dalam jumlah memadai, agar probabilitas dapat diramalkan dan dapat disalurkan kepada unit yang mengalami resiko [1].

Dalam asuransi terdapat istilah premi, dimana premi tersebut berada pada keberhakan pemilik asuransi tersebut dalam pembayaran premi tersebut. Premi merupakan pendapatan suatu perusahaan asuransi, dimana premi tersebut akan digunakan untuk mencapai suatu target perusahaan asuransi, dalam hal tersebut pemilik produk asuransi dapat memanfaatkan nilai premi ketika memiliki hal kepentingan dalam kehidupan.

Data Mining memiliki beberapa pandangan yang seperti *knowlegde discover* maupun *pattern recognition*, yang dimana keduanya memiliki istilah tersebut yang sebenarnya memiliki ketepatan masing-masing, istilah *knowledge discovery* atau penemuan pengetahuan yang digunakan dengan tujuan utama dari *data mining* memang untuk mendapatkan suatu pengetahuan yang masih tersembunyi dalam suatu data [2].

Penelitian ini mengacu pada penelitian sebelumnya oleh Pratiwi,dkk dalam studi yang menerapkan algoritma K-Means *clustering* untuk memprediksi minat nasabah pada PT. Asuransi Jiwa Bersama 1912 Bumiputera Prabumulih, menyebutkan bahwa penelitian ini menggunakan data nasabah yang diolah menggunakan aplikasi *Weka* dengan metode K-Means. Atribut yang digunakan yakni jenis asuransi, pekerjaan, pendapatan, dan alamat, penelitian ini tidak menyertakan tahun rilis pada setiap produknya. Penelitian menghasilkan kelompok nasabah yang berprofesi sebagai petani dengan jenis asuransi mitra beasiswa di wilayah gelombang yang memiliki minat tinggi untuk menjadi nasabah asuransi [3].

Novianto, dkk melakukan penelitiann menggunakan algoritma K-Means *clustering* untuk menganalisis bisnis perusahaan asuransi. Penelitian ini menggunakan metode K-Means namun tidak menyertakan jumlah nasabah pada setiap produk asuransinya. Penelitian ini menghasilkan 3 jenis *cluster* yaitu nilai terendah, sedang dan tertinggi dalam pengelompokan premi, klaim dan nilai pertanggungan [4].

Menurut Taslim dan Fajrizal dalam penelitiannya menggunakan algoritma K-Means untuk *clustering* data obat pada puskesmas Rumba, digunakan 2 atribut saja. Penelitian ini menghasilkan 3 *cluster* untuk kategori obat yaitu klaster pertama merupakan obat yang tidak/kurang ada permintaan, klaster kedua terdiri dari obat yang rata-rata permintaannya > 300 buah per bulan dan klaster ketiga yang isinya adalah daftar obat yang sangat tinggi rata-rata permintaannya >2000 buah per bulan [5].

Peneliti lain, Pramesti, dkk menggunakan metode K-Medoids untuk pengelompokan data potensi kebakaran hutan/lahan berdasarkan persebaran titik panas (*Hotspot*). Penelitian ini menghasilkan 2 jenis *cluster*. *Cluster* 1 area hutan yang termasuk dalam potensi kebakaran yang tinggi dengan hasil rata-rata *brightness* sebesar 344.470K dan *cluster* 2 meliputi area dengan potensi kebakaran sedang dengan hasil rata-rata *brightness* sebesar 318.800K. Penelitian ini menghasilkan rata-rata *confidence* sebesar 58,73% [6].

Anggreini dalam penelitiannya dengan algoritma K-Medoids, menganalisis data pendaftar sebagai strategi promosi di politeknik TEDC Bandung. Penelitian ini menggunakan metode K-Medoids dan menghasilkan klasterisasi berbentuk *chart bar* yang menunjukkan bahwa pendaftar terbanyak di Politeknik TEDC Bandung berasal dari Propinsi Jawa Barat, sedangkan kota Sumedang menjadi pendaftar paling sedikit. Nilai akurasi DBI pada penelitian ini adalah sebesar -1,906 [7].

Dilihat dari penelitian sebelumnya, yang membedakan studi ini adalah digunakannya 2 metode yaitu sebagai perbandingan klasterisasi antara metode K-Means dan K-Medoids dalam pengelompokkan produk asuransi perusahaan nasional. Penelitian ini menggunakan 3 atribut yaitu premi, jumlah nasabah dan tahun rilis pada setiap produk.

Penelitian ini mencakup produk-produk asuransi yang memiliki keterkaitan kekurangan dan kelebihan benefit setiap produk. Perusahaan ingin mengetahui produk-produk apa saja yang memiliki keunggulan. Produk-produk tersebut memiliki nilai premi dan tahun rilis produk. Peminatan suatu produk tidak semata diukur dari tahun produk tersebut dirilis, terkadang produk yang baru saja muncul menjadi produk yang lebih unggul dalam hal penjualan. Ada beberapa faktor yang mempengaruhi kondisi tersebut, salah satunya dari sisi pemasaran produk yang dilakukan saat produk tersebut diluncurkan. Ada beberapa unit yang terlibat dalam strategi pemasaran produk, antara lain *agency*, *bancassurance* dan *telemarketing*. Masing-masing unit menawarkan produk-produk terhadap calon nasabah dengan penawaran premi yang murah disertai *benefit* sesuai yang nasabah inginkan. Namun demikian

perusahaan kerap kali masih kesulitan dalam hal mengelompokkan produk-produk unggulan yang banyak diminati nasabah.

Berdasarkan latar belakang kondisi tersebut, dibutuhkan suatu teknik yang dapat mengelompokkan produk-produk asuransi dengan tujuan mempermudah pengelompokkan produk-produk unggulan dan mana kategori produk yang sesuai dengan kebutuhan nasabah. Tujuan penelitian ini adalah melakukan analisis dan pengolahan data dengan membandingkan metode K-Means dan K-Medoids dalam proses klasterisasi produk asuransi perusahaan nasional.

II. Metodologi Penelitian

Data mining merupakan serangkaian proses dalam pencarian pola, hubungan, penggalian nilai tambah dari data dan informasi yang berukuran besar berupa pengetahuan dengan tujuan menemukan hubungan dan menyederhanakan data agar diperoleh informasi yang dapat dipahami dan bermanfaat dengan bantuan ilmu statistik dan matematika [8].

Clustering merupakan pekerjaan yang memisahkan data atau vektor ke dalam sejumlah kelompok (*cluster*) menurut karakteristiknya masing-masing. Data-data yang mempunyai kemiripan karakteristik akan berkumpul dalam *cluster* yang sama, dan data-data dengan karakteristik berbeda akan terpisah dalam *cluster* yang berbeda. Tidak diperlukan label kelas untuk setiap data yang diproses dalam suatu *clustering*, karena nantinya label baru bisa diberikan ketika *cluster* sudah terbentuk. Karena tidak adanya target label kelas untuk setiap data, maka *clustering* sering disebut juga *unsupervised learning* [9].

Algoritma K-Means merupakan algoritma yang bekerja dengan cara mempartisi data ke dalam kluster, dengan demikian data yang memiliki kemiripan erada pada satu *cluster* yang sama dan data yang memiliki ketidaksamaan berada pada *cluster* yang lain [10].

Berikut ini langkah-langkah yang terdapat pada algoritma K-Means:

1. Tentukan jumlah kluster (k), tetapkan pusat kluster secara acak.
2. Hitung jarak setiap data ke pusat kluster
3. Kelompokkan data ke dalam kluster dengan jarak yang paling pendek.
4. Hitung pusat kluster baru.
5. Ulangi langkah 2 (dua) sampai 4 (empat) sehingga tidak ada lagi data yang pindah ke *cluster* lain.

Proses *clustering* dimulai dengan mengidentifikasi data yang diklasterisasi, digunakan rumus formula *Euclidean Distance* seperti yang terlihat pada rumus persamaan (1).

$$d_{ij} = \sqrt{(X_{1i} - X_{1j})^2 + (X_{2i} - X_{2j})^2 + \dots + (X_{ki} - X_{kj})^2} \quad (1)$$

Keterangan:

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

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

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

$$C = \frac{\sum m}{n} \quad (2)$$

Dimana C merupakan centroid data, m adalah anggota data yang termasuk ke dalam centroid tertentu dan n adalah jumlah data yang menjadi anggota centroid tertentu.

Algoritma K-Medoids atau *Partitioning Around Method* (PAM) merupakan bagian dari *partitioning clustering*. Metode k -medoids cukup efisien untuk dataset yang kecil. Langkah awal k -medoids adalah mencari titik yang paling representatif (medoids) dalam sebuah dataset dengan menghitung jarak dalam kelompok dari semua kemungkinan kombinasi dari medoids sehingga jarak antar titik dalam suatu *cluster* kecil sedangkan jarak titik antar *cluster* besar [11].

K-Medoids menggunakan objek representatif sebagai titik acuan, bukan mengambil nilai rata-rata dari objek dalam setiap kluster. Algoritma mengambil parameter input k , jumlah kluster yang akan dipartisi di antara satu set n objek [12].

Tahapan dari K-Medoids adalah:

- Tahap 1: Pilih titik awal K. Titik ini adalah medoid yang berfungsi sebagai kandidat dan dimaksudkan untuk menjadi poin paling sentral dari kelompok tersebut.
- Tahap 2: Kemudian setiap objek yang tersisa dikelompokkan dengan objek representatif yang paling mirip.
- Tahap 3: Kemudian mengganti salah satu medoid dengan medoid lain yang secara kualitas lebih baik dan dilakukan secara *iterative*.

David L. Davies dan Donald W. Bouldin memperkenalkan sebuah metode yang diberi nama dengan nama mereka berdua, yaitu *Davies-Bouldin Index (DBI)* yang digunakan untuk mengevaluasi *cluster* [13]. DBI adalah suatu metode yang digunakan untuk mengukur validitas *cluster* dalam metode *clustering*, kohesi didefinisikan sebagai jumlah dari kedekatan data terhadap titik pusat *cluster* dari *cluster* yang diikuti. Sedangkan separasi didasarkan pada jarak antar titik pusat *cluster* terhadap klasternya. Pengukuran dengan DBI dapat memaksimalkan jarak antara *cluster* Ci dan Cj dan pada saat yang sama mencoba untuk meminimalkan jarak antara titik-titik dalam *cluster*. Jika jarak antar *cluster* adalah maksimum, perbedaan signifikan pada setiap *cluster*, maka perbedaan kecil antara *cluster* lebih jelas. Jika jarak intra-*cluster* minimal, itu berarti setiap objek dalam *cluster* memiliki tingkat karakteristik penting yang tinggi [14]. Formula perhitungan DBI adalah sebagai berikut:

$$DBI = \frac{1}{k} \sum_{i=1}^k Ri, qt \quad (3)$$

Dari persamaan tersebut, k adalah banyaknya *cluster* yang digunakan. Jika nilai DBI yang diperoleh semakin kecil (non-negatif >=0), maka *clustering* yang diperoleh semakin baik [12].

III. Hasil dan Pembahasan

Pra-pemrosesan Data

Data pada penelitian ini bersumber dari perusahaan asuransi berskala nasional sejak tahun 1996 hingga 2019. Data tersebut melalui tahap pra-pemrosesan data sebelum digunakan untuk melakukan proses klasterisasi pada beberapa atribut. Dari 7 atribut yang digunakan proses perhitungan yaitu jenis produk, kode produk, nama produk, tahun rilis, premi dan jumlah nasabah. Atribut yang dipilih untuk dijadikan acuan dalam proses pengelompokan yaitu atribut jenis produk dan id produk.

Cleansing data merupakan untuk mengurangi *noise* yang dapat mempengaruhi perhitungan. Dalam proses *cleansing data*, data yang memiliki premi sangat besar dan jumlah nasabah yang sangat sedikit maka tidak digunakan, maka data yang digunakan yaitu 102 produk dari 133 produk yang digunakan. Data yang telah di proses *cleansing data* dapat dilihat pada Tabel 1.

Tabel 1. Hasil Pembersihan Data

Jenis Produk	Kode Produk	Tahun Rilis	Premi	Jumlah Nasabah
MJI	BLOS	2007	10.000.000	631
UNIT LINK	BSPR	2016	10.000.000	792
TRADISIONAL	BPED	2006	25.000.000	727
KESEHATAN	MCBM	2009	12.000	155.701
CRITICAL ILLNES	BCI	2008	1216	287
PERSONAL ACCIDENT	BMDP	2019	1200	133

Setelah dilakukan *cleansing data*, selanjutnya dilakukan pengelompokan produk-produk tersebut dari berdasarkan jenis produk, pemetaan produk dari berdasarkan jenis produk akan memudahkan dalam perhitungan yang untuk proses selanjutnya. Pemetaan perubahan data kedalam bentuk per-jenis kelompok dapat dilihat pada tabel 2.

Tabel 2. Hasil Pengelompokan Data

No	Jenis Produk	Kode Produk	Tahun Rilis	Premi	Jumlah Nasabah
30	MJI	BLOS	2007	10.000.000	631
99	WAIVER OF PREMIUM	BLWTPD	2010	1.041	3.318
11	FUTURE PLAN	BLTP	2009	200.000	248
94	UNIT LINK	BSPR	2016	10.000.000	792
73	TRADISIONAL	BPET	2010	25.000.000	271
16	KESEHATAN	BLHI	2011	25.365	7.232
33	PAYOR BENEFIT	BLPB	2007	1.212	5.905
8	CRITICAL ILLNES	BLPICS	2018	3.300	513
4	ANUITAS	BUP	2006	188.100	142
96	UNIVERSAL	SPDP	2016	25.000.000	171
44	PERSONAL ACCIDENT	AMPP	2018	20.000	2.484

Implementasi Algoritma

Penelitian ini bertujuan untuk membandingkan kelompok mana yang paling optimal diantara metode K-Means dan K-Medoids dalam pengelompokan data produk asuransi perusahaan nasional pada *cluster* 1 sampai *cluster* 10 dengan menggunakan metode K-Means pada tabel 3 dan metode K-Medoids pada tabel 4.

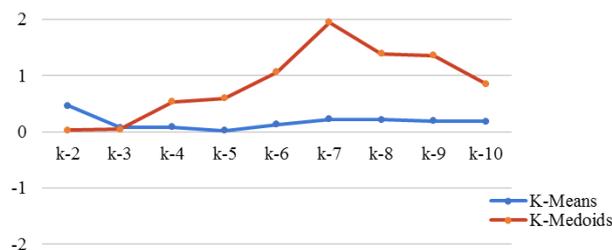
Tabel 3. Hasil Pemodelan K-Means

Kelompok	1	2	3	4	5	6	7	8	9	10
Jumlah data k=2	93	9	-	-	-	-	-	-	-	-
Jumlah data k=3	92	4	6	-	-	-	-	-	-	-
Jumlah data k=4	91	4	6	1	-	-	-	-	-	-
Jumlah data k=5	91	4	5	1	1	-	-	-	-	-
Jumlah data k=6	88	4	5	1	1	3	-	-	-	-
Jumlah data k=7	60	4	5	1	1	3	28	-	-	-
Jumlah data k=8	55	4	5	1	1	28	3	5	-	-
Jumlah data k=9	55	4	5	1	1	2	28	1	5	-
Jumlah data k=10	52	4	5	1	1	2	15	16	1	5

Tabel 4. Hasil Pemodelan K-Medoids

Kelompok	1	2	3	4	5	6	7	8	9	10
Jumlah data k=2	98	4	-	-	-	-	-	-	-	-
Jumlah data k=3	4	92	6	-	-	-	-	-	-	-
Jumlah data k=4	6	69	4	23	-	-	-	-	-	-
Jumlah data k=5	23	6	5	64	4	-	-	-	-	-
Jumlah data k=6	22	38	4	7	6	25	-	-	-	-
Jumlah data k=7	6	4	16	36	16	5	19	-	-	-
Jumlah data k=8	19	6	13	13	5	15	4	27	-	-
Jumlah data k=9	5	26	9	13	4	6	5	15	19	-
Jumlah data k=10	14	27	5	1	13	18	4	12	2	6

Setelah melakukan pengelompokan dengan metode K-Means dan K-Medoids, tahap selanjutnya penelitian ini menentukan jumlah *cluster* yang paling optimal dengan *Rapidminer* dilihat dari *Davies-Bouldin Index* (DBI). Proses tersebut dilakukan untuk mengetahui nilai DBI dari masing-masing metode di setiap *cluster* nya. Pengujian dilakukan dari *cluster* k=2 sampai k=10. Berikut hasil gambaran perbandingan DBI dapat dilihat pada gambar 1.



Gambar 1. Perbandingan DBI Algoritma K-Means dan K-Medoids

Untuk melihat hasil pengujian setiap *cluster* pada metode K-Means dan K-Medoids dapat dilihat pada tabel 5.

Tabel 5. Hasil Perbandingan DBI

Cluster	K-Means	K-Medoids
k-2	0,464	0,027
k-3	0,072	0,045
k-4	0,081	0,532
k-5	0,018	0,599
k-6	0,13	1,064
k-7	0,221	1,947
k-8	0,213	1,39
k-9	0,192	1,355
k-10	0,185	0,855
Nilai DBI terkecil	0,018 (k=5)	0,027 (k=2)

Setelah melakukan perbandingan DBI dengan metode K-Means dan K-Medoids dari k=2 sampai k=10 yang terdapat pada tabel 5, maka nilai DBI terkecil pada K-Means terletak pada *cluster* K-Means ke-5 dengan nilai DBI sebesar 0,018, sedangkan nilai DBI terkecil pada K-Medoids terletak pada *cluster* K-Medoids ke-2 dengan nilai DBI sebesar 0,027. Karena nilai DBI pada K-Means (k=5) lebih kecil, maka diketahui bahwa pengelompokan menggunakan metode K-Means lebih baik dibandingkan dengan metode K-Medoids dalam klusterisasi produk asuransi perusahaan nasional. Berdasarkan kelompok yang paling optimal selanjutnya dilakukan interpretasi *clustering* terhadap data produk asuransi yang hasilnya dapat dilihat pada tabel 6.

Tabel 6. Hasil *Cluster* Optimal dengan Metode K-Means (k=5)

Kelompok	Jumlah Anggota	Tahun Rilis	Rata-rata Premi	Rata-rata Nasabah
1	91	1997 – 2019	69.210	12.231
2	4	2006 – 2016	25.000.000	321
3	5	2006 – 2016	10.000.000	5.822
4	1	2008	2.500.000	404
5	1	2004	7.500.000	392

IV. Kesimpulan

Setelah melakukan analisis dan pengolahan data dengan membandingkan metode K-Means dan K-Medoids dalam klusterisasi produk asuransi, maka dapat disimpulkan bahwa nilai DBI yang diperoleh dari metode K-Means dan K-Medoid dengan eksperimen pembentukan *cluster* sebanyak sembilan menghasilkan nilai DBI terkecil pada metode K-Means adalah pada nilai k=5 yaitu 0,018. Sebagai pembanding kinerja, pembentukan *cluster* dengan metode K-Medoids memiliki nilai DBI terkecil pada k=2 yaitu sebesar 0,027. Dengan demikian maka pembentukan *cluster* yang paling optimal dalam klusterisasi produk asuransi perusahaan nasional adalah menggunakan metode K-Means.

Daftar Pustaka

- [1] A. Khairani and I. Irwansyah, "Pemanfaatan Big Data Untuk Percepatan Proses Underwriting Sebagai Strategi Komunikasi Marketing Terpadu Perusahaan Asuransi Jiwa," *J. Nomosleca*, vol. 4, no. 2, pp. 785–

- 793, 2018, doi: 10.26905/nomosleca.v4i2.2546.
- [2] A. Bastian *et al.*, “Penerapan Algoritma K-Means Clustering Analysis Pada Penyakit Menular Manusia (Studi Kasus Kabupaten Majalengka),” *Jsi.Cs.Ui.Ac.Id*, vol. 14, no. 1, pp. 26–32, 2018.
- [3] L. M. Pratiwi, Diana, and E. P. Agustin, “Penerapan K-Means Clustering Untuk Memprediksi Minat Nasabah Pada Pt . Asuransi Jiwa Bersama 1912 Bumiputera Prabumulih,” *Univ. Bina Darma*, pp. 1–16, 2016.
- [4] R. L. Novianto, and Goeirmento, “Penerapan Data Mining menggunakan Algoritma K-Means Clustering untuk Menganalisa Bisnis Perusahaan Asuransi,” *JATISI (Jurnal Tek. Inform. dan Sist. Informasi)*, vol. 6, no. 1, pp. 85–95, 2019, doi: 10.35957/jatisi.v6i1.150.
- [5] T. Taslim and F. Fajrizal, “Penerapan algoritma k-mean untuk clustering data obat pada puskesmas rumbai,” *Digit. Zo. J. Teknol. Inf. dan Komun.*, vol. 7, no. 2, pp. 108–114, 2016, doi: 10.31849/digitalzone.v7i2.602.
- [6] D. F. Pramesti, M. T. Furqon, and C. Dewi, “Implementasi Metode K-Medoids Clustering Untuk Pengelompokan Data Potensi Kebakaran Hutan / Lahan Berdasarkan Persebaran Titik Panas (Hotspot),” *J-ptiik*, vol. 1, no. 9, pp. 723–732, 2017.
- [7] N. L. Anggraeni, “Teknik Clustering Dengan Algoritma K-Medoids Untuk Menangani Strategi Promosi Di Politeknik TEDC Bandung,” *Jurnal Teknologi Informasi dan Pendidikan*, vol. 12 no. 2 pp. 1-7, 2019.
- [8] G. Abdurrahman, “Clustering Data Ujian Tengah Semester (UTS) Data Mining Menggunakan Algoritma K-Means,” *J. Sist. dan Teknol. Inf. Indones.*, vol. 1, no. 2, pp. 71–79, 2016, doi: 10.32528/justindo.v1i2.566.
- [9] Y. Lase and E. Panggabean, “Implementasi Metode K-Means Clustering Dalam Sistem Pemilihan Jurusan Di SMK Swasta Harapan Baru,” *J. Penelit. Tek. Inform.*, vol. 2, no. 2, pp. 43–47, 2019.
- [10] M. Nurul, Rohmawati Defiyanti, Sofi Jajuli, “Implementasi Algoritma K-Means Dalam Pengklasteraan Mahasiswa Pelamar Beasiswa,” *Jitter 2015*, vol. I, no. 2, pp. 62–68, 2015.
- [11] H. Zayuka, S. M. Nasution, and Y. Purwanto, “Perancangan Dan Analisis Clustering Data Menggunakan Metode K-Medoids Untuk Berita Berbahasa Inggris Design and Analysis of Data Clustering Using K-Medoids Method For English News,” *e-Proceeding Eng.*, vol. 4, no. 2, pp. 2182–2190, 2017.
- [12] R. D. Ramadhani and D. J. Ak, “Evaluasi K-Means dan K-Medoids pada Dataset Kecil,” *Semin. Nas. Inform. dan Apl.*, no. September, pp. 20–24, 2017.
- [13] S. Nawrin, M. Rahatur, and S. Akhter, “Exploring K-Means with Internal Validity Indexes for Data Clustering in Traffic Management System,” *Int. J. Adv. Comput. Sci. Appl.*, vol. 8, no. 3, pp. 264–272, 2017, doi: 10.14569/ijacsa.2017.080337.
- [14] M. Arif Wani and R. Riyaz, “A novel point density based validity index for clustering gene expression datasets,” *Int. J. Data Min. Bioinform.*, vol. 17, no. 1, pp. 66–84, 2017, doi: 10.1504/IJDMB.2017.084027.