



## *Comparison of the K-Means and K-Medoids Algorithms for Grouping Student Expenditures*

### **Perbandingan Algoritma K-Means dan K-Medoids untuk Pengelompokan Pengeluaran Mahasiswa**

Sofia Fulvi Intan<sup>1\*</sup>, Winda Elvira<sup>2</sup>, Sherly Rahayu<sup>3</sup>, Nadia Nurfadilla<sup>4</sup>

<sup>1,2,3,4</sup>Program Studi Sistem Informasi, Fakultas Sains dan Teknologi,  
Universitas Islam Negeri Sultan Syarif Kasim Riau, Indonesia

E-Mail: <sup>1</sup>12050322953@students.uin-suska.ac.id, <sup>2</sup>12050325545@students.uin-suska.ac.id,  
<sup>3</sup>12050327417@students.uin-suska.ac.id, <sup>4</sup>12050320416@students.uin-suska.ac.id

*Corresponding Author: Sofia Fulvi Intan*

#### **Abstract**

*Data mining is information processing that can be done hidden from a database. The use of data mining is widely applied in various sectors of life. One of them is in the field of Education, in this study data mining is applied to classify student monthly expenses by a comparison of 2 algorithms. Data processing is performed using the K-Means and K-Medoids algorithms. This research was conducted because of the many problems related to financial management and difficulties in achieving financial goals for students. The existence of this comparison is to measure the best algorithm by producing a grouping of the amount of student spending at least in one month. From the data processed the K-Medoids algorithm is more optimal in grouping student monthly expenditure data compared to K-Means with the smallest number of k located at  $K = 10$  with a DBI value of 0.107, this shows that student spending is the least in that grouping. Meanwhile, clustering in the K-Means algorithm produces the lowest DBI value in cluster 9 with a value of 0.119 indicating a higher number and the amount of student spending is greater than K-Medoids.*

*Keyword: Data Mining, DBI, K-means, K-Medoids, Spending.*

#### **Abstrak**

Data mining merupakan pemrosesan informasi yang dapat dilakukan secara tersembunyi dari suatu database. Penggunaan data mining banyak sekali diterapkan dalam berbagai sector kehidupan. Salah satunya dalam bidang Pendidikan, pada penelitian ini data mining diterapkan untuk mengelompokkan pengeluaran mahasiswa dengan perbandingan 2 algoritma. Pengolahan data yang dilakukan menggunakan algoritma K- Means dan K-Medoids. Penelitian ini dilakukan karena banyaknya permasalahan terkait pengelolaan keuangan dan kesulitan dalam pencapaian tujuan finansial bagi mahasiswa. Adanya perbandingan ini guna mengukur algoritma yang paling baik dengan menghasilkan pengelompokan jumlah pengeluaran mahasiswa paling sedikit dalam satu bulan. Dari data yang di olah algoritma K-Medoids lebih optimal dalam pengelompokkan data pengeluaran mahasiswa dibandingkan K-Means dengan jumlah k paling kecil terletak pada  $K=10$  dengan nilai DBI 0,107, hal ini menunjukkan pengeluaran mahasiswa paling sedikit pada pengelompokan tersebut. Sedangkan pengelompokan pada algoritma K-Means menghasilkan nilai DBI terendah pada cluster 9 dengan nilai 0,119 menunjukkan angka yang lebih tinggi dan jumlah pengeluaran mahasiswa lebih besar dibandingkan K-Medoids.

Kata Kunci: Data Mining, DBI, K-Means, K-Medoids, Pengeluaran.

#### **1. PENDAHULUAN**

Data mining ialah suatu pemrosesan informasi yang dilakukan dari database walau tidak terlihat secara langsung [1]. Data mining dalam penggunaannya dapat dipakai dalam berbagai bidang. Pada bidang pendidikan data mining berfokus pada pengembangan dan penerapan metode dengan tujuan memperkenalkan

penemuan baru [2]. Umumnya teknik analisis digunakan seperti analisis regresi, analisis cluster, teknik estimasi nonlinear, dan lainnya [3]. Dalam pendidikan data mining penggunaannya ditentukan oleh hasil analisis dan lingkup persoalan. Dalam menganalisis dan mengelompokkan data tersebut dibutuhkan metode data mining [4]. Adapun metode yang dapat diterapkan dalam mengelompokkan data tersebut antara lain Classification, Estimation, Clustering, Association dan Prediction [5]. Pada penelitian ini metode yang diterapkan ialah metode clustering.

Clustering ialah proses satu pola pengelompokan dibagikan pada suatu kumpulan data [6]. Pada penelitian ini menggunakan dua algoritma clustering yaitu algoritma K-Means dan K-Medoids. Algoritma K-Means ialah algoritma data mining yang sederhana dalam pengolahan data, dikarenakan mudah beradaptasi dan relatif cepat [7]. Algoritma K-Medoid menentukan nilai centroidnya dari nilai median, sedangkan K-Means dari nilai rata-rata [8]. Didapatkan varian pengembangan K-Means yaitu K-Medoids. Bertujuan untuk mengurangi partisi yang sensitivitas hasilnya dengan nilai ekstrim yang terdapat di dataset [9]. Algoritma K-Medoids hadir sebagai bentuk penanggulangan dari kelemahan Algoritma K-Means yang outliernya sensitif terhadap substansi menyimpang dari distribusi data dengan nilai yang besar dari sebuah objek [10]. Kedua algoritma tersebut digunakan pada penelitian ini untuk mengelompokkan data pengeluaran mahasiswa.

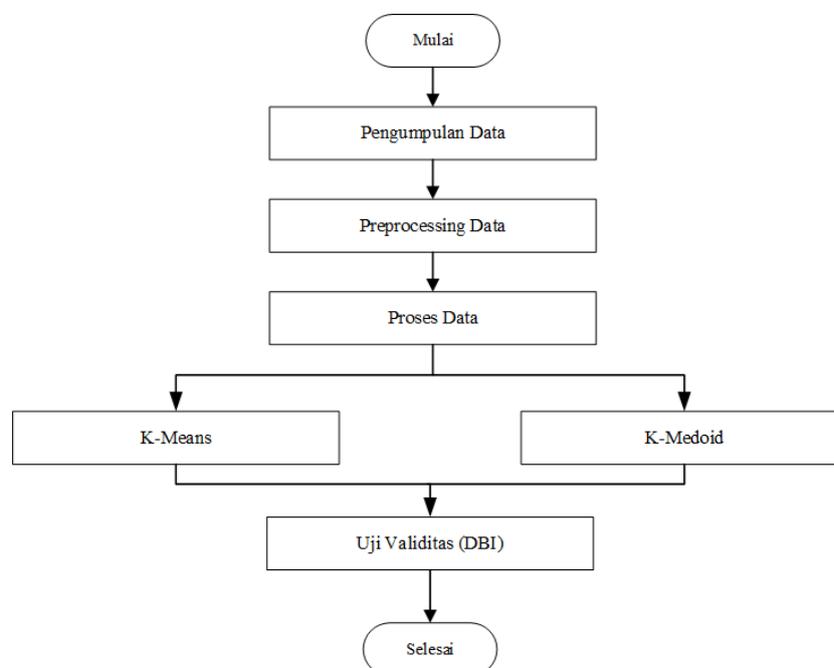
Data Pengeluaran Mahasiswa ini bersumber dari Kaggle yang berjumlah 105 data, dimana data tersebut memiliki 8 atribut yaitu jenis kelamin, umur, tahun belajar, tempat tinggal, beasiswa, kerja paruh waktu, kendaraan dan pengeluaran. Permasalahan dari penelitian ini adalah untuk mengelompokkan data pengeluaran mahasiswa yang mencukupi atau tidak. Berdasarkan permasalahan diatas mengharuskan untuk memanfaatkan teknik clustering sebagai pengolahan data yang tepat.

Penelitian terdahulu membandingkan algoritma K-Medoids dan K-Means dalam Pengelompokan Kelas Siswa Tuna grahita menunjukan K-Means lebih efektif dalam mengelompokkan kelas siswa karena menghasilkan nilai DBI yang lebih kecil [11]. Kemudian penelitian yang menerapkan algoritma yang sama pada studi kasus di provinsi Jawa Barat yaitu mengelompokkan Daerah Rawan Tahan Longsor menyatakan algoritma K-Means lebih optimal terlelak pada cluster 6 dalam penelitian tersebut [12].

Berdasarkan latar belakang penelitian diatas diangkat judul “Perbandingan Algoritma K-Means dan K-Medoids untuk Pengelompokan Pengeluaran Mahasiswa”. Adanya perbandingan ini bertujuan untuk menghasilkan jumlah pengeluaran mahasiswa paling sedikit dengan tingkat cluster terbaik menggunakan algoritma K-Means dan K-Medoids.

## 2. METODOLOGI PENELITIAN

### 2.1 Tahapan Penelitian



**Gambar 1.** Metodologi Penelitian

Gambar 1 merupakan tahapan penelitian yang dilakukan yaitu pengumpulan data, preprocessing data, hasil dan pembahasan, serta uji validitas. Dalam melakukan perbandingan data diproses menggunakan dua algoritma cluster dengan pengujian validitas DBI agar bisa mengetahui mana cluster dan algoritma terbaik untuk pengelompokan data yang diuji.

## 2.2 K-Means

K-Means termasuk dalam aplikasi clustering data mining [13]. Algoritma ini mengelompokkan data menggunakan pembatas dan membagi data menjadi blok-blok, adapun langkah-langkah, yaitu:

1. Tentukan nilai k untuk jumlah cluster yang akan dibentuk dan sebagai penentu pusat cluster [14].
2. Menentukan Jarak, menggunakan persamaan Euclidean untuk menghitung setiap jarak ke pusatcluster. Centroid digunakan untuk mencari jarak terpendek dari data.
3. Menghitung Pusat Cluster, perhitungan pusat cluster berasal dari jumlah rata-rata data didalam cluster [15].

Umumnya nilai k wajib ditetapkan terlebih dahulu saat melakukan klasterisasi. Biasanya user telah sudah mempunyai data awal untuk objek yang akan dipelajari [16].

## 2.3 K-Medoid

K-Medoid adalah bagian algoritma clustering, dalam dataset untuk menghitung jarak semua kelompok medoids dalam cluster terkecil, dan jarak titik antar cluster besar [17]. Pada persamaan 3 euclidian data akan dikelompokkan pada cluster terdekat [18].

## 2.4 Davies Bouldin Index (DBI)

DBI digunakan untuk mengumpulkan suatu algoritma clustering pada dataset yang digunakan berdasarkan tingkat densitas data centroid, keterpisahan antara cluster dan rasio antar cluster, hasil perhitungan ketiga nilai [14]. Jika keadaan maksimal pada jarak cluster, maka menghasilkan sangat kecil karakter cluster sehingga perbedaannya dapat dilihat sangat jelas. Kriteria DBI dapat dilihat berdasarkan jarak pada cluster dan rasio yang ada [19]. Jika setiap objek tingkat kesamaan cluster tinggi karakteristiknya berarti jarak intra clusternya minimum [13]. DBI berfungsi sebagai evaluasi cluster [20].

## 3. HASIL DAN PEMBAHASAN

Penelitian yang dilakukan dengan membandingkan 2 algoritma, dengan melakukan normalisasi data terlebih dahulu, hal ini guna memastikan record data tetap konsisten. Kemudian mengimplementasikannya ke algoritma K-Means dan K-Medoid untuk mendapatkan cluster terbaik dan jumlah pengeluaran mahasiswa paling sedikit.

### 3.1 Pengumpulan dan Preprocessing Data

dari 5 atribut yaitu data tempat tinggal, beasiswa, kerja paruh waktu, kendaraan, dan pengeluaran bulanan. Pada penelitian ini, data yang dipakai berjumlah 105 data. Data yang didapat dinormalisasikan dengan excel, pengolahan data merupakan sekelompok kegiatan yang dilakukan setelah data dikumpulkan. Kemudian data normalisasi diolah di Rapidminer.

**Tabel 1.** Normalisasi Data

No	Jenis Kelamin	Umur	Tahun Belajar	Tempat Tinggal	Beasiswa	Kerja Paruh Waktu	Kendaraan	Pengeluaran Bulanan
1	0.000	0.500	0.333	1.000	0.000	0.000	0.000	0.013
2	1.000	1.000	0.667	0.000	0.000	1.000	0.500	0.105
3	1.000	0.750	0.333	1.000	1.000	0.000	0.000	0.000
4	1.000	0.250	0.667	0.000	0.000	0.000	0.500	0.079
5	0.000	0.250	0.333	1.000	0.000	0.000	0.500	0.211
6	1.000	0.625	0.667	0.000	0.000	1.000	1.000	0.474
7	0.000	0.500	0.333	1.000	1.000	0.000	0.000	0.079
8	1.000	0.625	0.667	0.000	0.000	1.000	0.500	0.539
9	0.000	0.125	0.000	1.000	1.000	0.000	0.000	0.013
10	1.000	0.250	0.000	1.000	0.000	0.000	0.000	0.026
11	1.000	0.625	0.667	1.000	0.000	0.000	1.000	0.276
12	1.000	0.125	0.333	0.000	0.000	0.000	0.500	0.053
13	0.000	0.125	0.333	1.000	0.000	0.000	0.000	0.000
14	1.000	0.250	0.333	1.000	0.000	1.000	1.000	0.605
15	0.000	0.625	1.000	0.000	1.000	1.000	1.000	0.211

Hasil dari normalisasi data menggunakan rumus max-min pada Tabel 1 berfungsi untuk mengetahui data sudah bersih dan dependensi data.

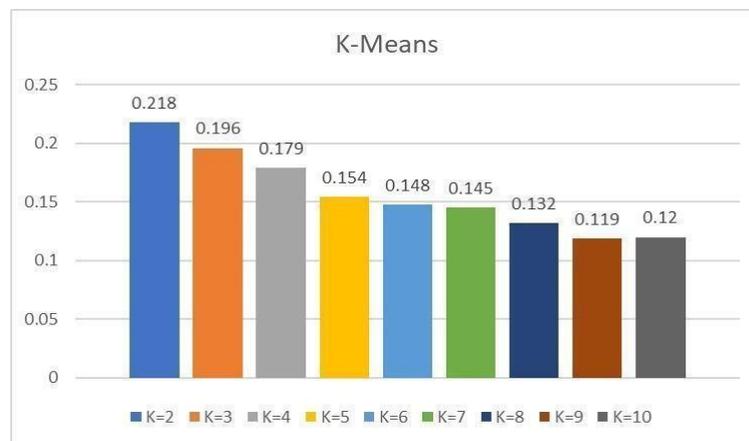
### 3.2 K-Means

K-Mean diterapkan pada pengelompokan pengeluaran mahasiswa dengan 9 cluster, dimana cluster 2 sampai cluster 10. Perhitungan cluster K-Means di Rapidminer dapat dilihat pada Tabel 2.

**Tabel 2.** Hasil Klasterisasi K-Means

Percobaan	K-Means									
	1	2	3	4	5	6	7	8	9	10
K=2	33	72	-	-	-	-	-	-	-	-
K=3	15	36	54	-	-	-	-	-	-	-
K=4	10	30	37	28	-	-	-	-	-	-
K=5	23	10	38	12	22	-	-	-	-	-
K=6	11	9	37	16	18	14	-	-	-	-
K=7	29	23	3	10	9	15	16	-	-	-
K=8	14	9	15	17	12	5	9	24	-	-
K=9	10	19	15	14	14	9	7	12	5	-
K=10	12	6	10	14	9	7	24	10	7	6

Pada Tabel 2 terdapat hasil dari perhitungan 9 cluster K-Means. Dimana cluster terbaik dari K-Means ditentukan dari nilai DBI yang dapat ditunjukkan pada Gambar 2.



**Gambar 2.** Nilai DBI K-Means

Gambar 2 terlihat pada cluster 9 merupakan cluster terbaik. Nilai DBI diperoleh rata-rata 0,156 dari algoritma K-Mean. Percobaan K=9 nilai DBI 0,119 dengan cara membagi data tempat tinggal, beasiswa, kerja paruh waktu, kendaraan, dan pengeluaran menjadi 9 cluster.

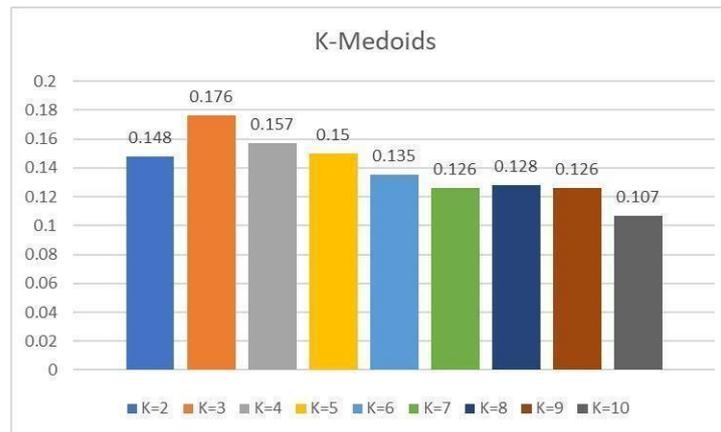
### 3.3 K-Medoid

K-Medoids dalam percobaan pengujian 9 cluster di Rapidminer, dari klaster 2 hingga 10. Pada tabel 3 dapat dilihat hasil klasterisasi.

**Tabel 3.** Hasil Klasterisasi K-Medoid

Percobaan	K-Medoids									
	1	2	3	4	5	6	7	8	9	10
K=2	55	50	-	-	-	-	-	-	-	-
K=3	15	36	54	-	-	-	-	-	-	-
K=4	54	27	13	11	-	-	-	-	-	-
K=5	21	11	53	13	7	-	-	-	-	-
K=6	6	12	34	10	25	18	-	-	-	-
K=7	10	15	8	14	37	5	16	-	-	-
K=8	11	10	3	37	14	7	8	15	-	-
K=9	8	6	37	5	14	10	14	6	5	-
K=10	10	8	11	5	2	6	12	32	5	14

Tabel 3 Menunjukkan data yang telah diklasterisasi dari atribut tempat tinggal, beasiswa, kerja paruh waktu, kendaraan, dan pengeluaran bulanan. Mengetahui cluster terbaik, dapat ditunjukkan dari gambar dibawah ini.

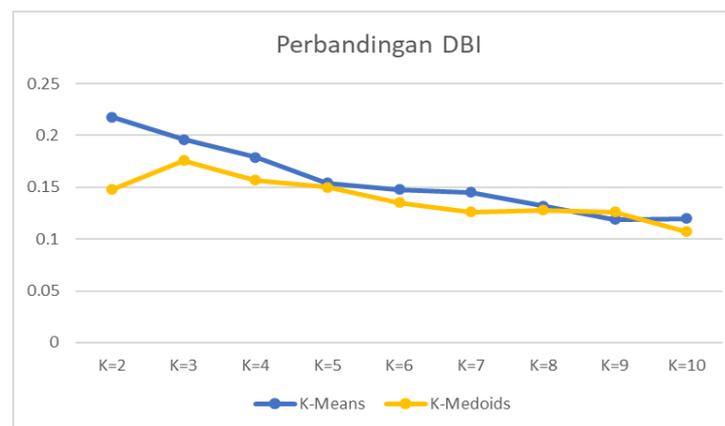


**Gambar 3.** Nilai DBI K-Medoid

Berdasarkan Gambar 4 diperoleh rata-rata 0,139 nilai DBI K-Medoids pada tiap cluster dimana percobaan K=10 merupakan cluster terbaik dalam membagi data menjadi 10 cluster.

### 3.4 Perbandingan Validitas Cluster (DBI)

Data yang sudah diolah di RapiMiner berfungsi untuk mengetahui nilai DBI dari kedua algoritma. Nilai DBI dari perbandingan dua algoritma ditunjukkan pada gambar 4.



**Gambar 4.** Perbandingan Nilai DBI algoritma K-Means dan K-medoids

Gambar 4 menampilkan dengan jelas perbandingan nilai DBI dari kedua algoritma. K-Medoid hasil nilai DBInya lebih kecil, yaitu 0,107 daripada K-Mean. Dapat diketahui K-Medoid lebih baik dibandingkan K-Means dalam pengelompokan pengeluaran mahasiswa.

## 4. KESIMPULAN

Berdasarkan penelitian yang dilakukan terkait data pengeluaran mahasiswa, menunjukkan algoritma K-Medoid lebih baik dalam pengelompokan pengeluaran mahasiswa daripada K-Means. Karena nilai DBI K-Medoids lebih kecil yaitu 0,107 yang terletak di cluster 10 dimana pengeluaran yang mencukupi sekitar 150 hingga 350 ribu perbulan dengan memiliki tempat tinggal dan kendaraan. Sedangkan K-Means memiliki pengeluaran lebih banyak yaitu sekitar 650 hingga 800 ribu perbulan hanya memiliki tempat tinggal saja, yang terletak di cluster 9 dengan nilai DBI 0,119.

## REFERENSI

- [1] Mustika *et al.*, *Data Mining dan Aplikasinya*. 2021.
- [2] D. Syaputri, P. H. Noprita, and S. Romelah, "Implementasi Algoritma K-Means untuk Pengelompokan Distribusi Sosial Ekonomi Masyarakat Berdasarkan Demografi Kependudukan," *Indones. J. Mach. Learn. Comput. Sci.*, vol. 1, no. 1, pp. 1–6, 2021.
- [3] I. Kamila, U. Khairunnisa, and M. Mustakim, "Perbandingan Algoritma K-Means dan K-Medoids untuk Pengelompokan Data Transaksi Bongkar Muat di Provinsi Riau," *J. Ilm. Rekayasa dan Manaj. Sist. Inf.*,

- vol. 5, no. 1, p. 119, 2019, doi: 10.24014/rmsi.v5i1.7381.
- [4] E. Prasetyowati, *DATA MINING Pengelompokan Data untuk Informasi dan Evaluasi*, vol. 29. Duta Media Publishing, 2017.
  - [5] U. R. Gurning and Mustakim, "Penerapan Algoritma K-Means dan K-Medoid untuk Pengelompokan Data Pasien Covid-19," *Build. Informatics, Technol. Sci.*, vol. 3, no. 1, p. 48–55, 2021, doi: 10.47065/bits.v3i1.1003.
  - [6] N. Wakhidah, "Clustering menggunakan k-means algorithm," *J. Transform.*, vol. 8, no. 1, pp. 33–39, 2010.
  - [7] M. D. R. Siahaan, "Pengaruh Literasi Keuangan Terhadap Perilaku Pengelolaan Keuangan Pada Mahasiswa Perguruan Tinggi Di Surabaya," *Artik. Ilm.*, vol. 1, pp. 1–15, 2013.
  - [8] S. D. Nirwana, M. I. Jambak, and A. Bardadi, "Perbandingan Algoritma K-Means Dan K-Medoids Dalam Clustering Rata-Rata Penambahan Kasus Covid-19 Berdasarkan Kota/Kabupaten Di Provinsi Sumatera Selatan," *JSiI (Jurnal Sist. Informasi)*, vol. 9, no. 2, pp. 126–131, 2022, doi: 10.30656/jsii.v9i2.5127.
  - [9] A. Fira, C. Rozikin, and G. Garno, "Komparasi Algoritma K-Means dan K-Medoids Untuk Pengelompokan Penyebaran Covid-19 di Indonesia," *J. Appl. Informatics Comput.*, vol. 5, no. 2, pp. 133–138, 2021, doi: 10.30871/jaic.v5i2.3286.
  - [10] W. A. Triyanto, "Algoritma K-Medoids Untuk Penentuan Strategi Pemasaran Produk," *Simetris J. Tek. Mesin, Elektro dan Ilmu Komput.*, vol. 6, no. 1, p. 183, 2015, doi: 10.24176/simet.v6i1.254.
  - [11] F. Harahap, "Perbandingan Algoritma K Means dan K Medoids Untuk Clustering Kelas Siswa Tunagrahita," *TIN Terap. Inform. Nusantara*, vol. 2, no. 4, pp. 191–197, 2021.
  - [12] M. Herviany, S. P. Delima, T. Nurhidayah, and Kasini, "Perbandingan Algoritma K-Means dan K-Medoids untuk Pengelompokan Daerah Rawan Tanah Longsor di Provinsi Jawa Barat," *MALCOM Indones. J. Mach. Learn. Comput. Sci.*, vol. 1, no. 1, pp. 34–40, 2021, [Online]. Available: <https://journal.irpi.or.id/index.php/malcom/article/view/60>
  - [13] W. M. P. Dhuhita, "Clustering Menggunakan Metode K-Means Untuk," *J. Inform.*, vol. 15, no. 2, pp. 160–174, 2016.
  - [14] R. K. Dinata, H. Novriando, N. Hasdyna, and S. Retno, "Reduksi Atribut Menggunakan Information Gain," vol. 6, no. 1, pp. 48–53, 2020.
  - [15] W. A. Wahyuni and S. Saepudin, "Penerapan Data Mining Clustering Untuk Mengelompokkan Berbagai Jenis Merk Mesin Cuci," *Semin. Nas. Sist. ...*, pp. 306–313, 2021, [Online]. Available: <https://sismatik.nusaputra.ac.id/index.php/sismatik/article/view/35%0Ahttps://sismatik.nusaputra.ac.id/index.php/sismatik/article/download/35/31>
  - [16] N. T. Luchia, H. Handayani, F. S. Hamdi, D. Erlangga, and S. Fitri Octavia, "Perbandingan K-Means dan K-Medoids Pada Pengelompokan Data Miskin di Indonesia," *MALCOM Indones. J. Mach. Learn. Comput. Sci.*, vol. 2, no. 2, pp. 35–41, 2022.
  - [17] D. Marlina, N. Lina, A. Fernando, and A. Ramadhan, "Implementasi Algoritma K-Medoids dan K-Means untuk Pengelompokan Wilayah Sebaran Cacat pada Anak," *J. CoreIT J. Has. Penelit. Ilmu Komput. dan Teknol. Inf.*, vol. 4, no. 2, p. 64, 2018, doi: 10.24014/coreit.v4i2.4498.
  - [18] H. Syukron, M. F. Fayyad, and F. J. Fauzan, "Comparison K-Means K-Medoids and Fuzzy C-Means for Clustering Customer Data with LRFM Model " Perbandingan K-Means K-Medoids dan Fuzzy C-Means untuk Pengelompokan Data Pelanggan dengan Model LRFM," vol. 2, no. October, pp. 76–83, 2022.
  - [19] B. Jumadi Dehotman Sitompul, O. Salim Sitompul, and P. Sihombing, "Enhancement Clustering Evaluation Result of Davies-Bouldin Index with Determining Initial Centroid of K-Means Algorithm," *J. Phys. Conf. Ser.*, vol. 1235, no. 1, 2019, doi: 10.1088/1742-6596/1235/1/012015.
  - [20] F. Farahdinna, I. Nurdiansyah, A. Suryani, and A. Wibowo, "Perbandingan Algoritma K-Means Dan K-Medoids Dalam Klasterisasi Produk Asuransi Perusahaan Nasional," *J. Ilm. FIFO*, vol. 11, no. 2, p. 208, 2019, doi: 10.22441/fifo.2019.v11i2.010.