

Clusterisasi Tingkat Pengangguran Terbuka Menurut Provinsi di Indonesia Menggunakan Algoritma K-Medoids

Abdul Karim¹, Shinta Esabella^{2,*}, Kusmanto³, Sudi Suryadi⁴, Erwin Mardinata⁵

¹Fakultas Sain dan Teknologi, Program studi Teknologi Informasi, Universitas Labuhanbatu, Rantauprapat, Indonesia

²Program Studi Informatika, Universitas Teknologi Sumbawa, Sumbawa, Indonesia

³Fakultas Ilmu Komputer, Prodi Teknik Informatika, Universitas Al Washliyah Labuhanbatu, Rantauprapat, Indonesia

⁴Fakultas Sain dan Teknologi, Program studi Sistem Informasi, Universitas Labuhanbatu, Rantauprapat, Indonesia

⁵Program studi Bisnis Digital, Fakultas Ekonomi dan Bisnis, Universitas Teknologi Sumbawa, Sumbawa, Indonesia

Email: ¹Abdkarim6@gmail.com, ^{2,*}shinta.esabella@uts.ac.id, ³kusnabara03@gmail.com,

⁴sudisuryadi28@gmail.com, ⁵erwin.mardinata@uts.ac.id

Email Penulis Korespondensi: shinta.esabella@uts.ac.id

Submitted: 05/11/2024; Accepted: 02/12/2024; Published: 03/12/2024

Abstrak—Tingkat Pengangguran Terbuka (TPT) di Indonesia mengalami penurunan pada Februari 2024 menjadi 4,82%, menunjukkan perbaikan dibandingkan Februari 2023. Meskipun TPT menurun, masih terdapat wilayah dengan TPT mencapai 7,02%, yang berpotensi menimbulkan konsekuensi negatif seperti peningkatan kriminalitas. Upaya untuk mengatasi TPT termasuk meningkatkan pertumbuhan ekonomi, mengembangkan kualitas pendidikan dan pelatihan. Penelitian ini memanfaatkan clusterisasi dalam data mining. Adapun jumlah cluster yang dibentuk sebanyak 3 cluster dengan nilai DBI -1,685. Penelitian ini menggunakan clusterisasi K-Medoids untuk mengelompokkan 38 provinsi berdasarkan TPT. Dari 38 data tersebut, terdapat data yang tidak lengkap sehingga dilakukan preprocessing dengan menggunakan operator "filter example" yang ada didalam rapidminer untuk menghilangkan data yang tidak lengkap sehingga ada 34 data yang akan di gunakan dalam penelitian ini (setelah dilakukan preprocessing). Hasilnya menunjukkan 4 provinsi dengan TPT tertinggi (Kep. Riau, DKI Jakarta, Jawa Barat, dan Banten) dengan persentase 11,76%.

Kata Kunci: Data Mining; Clustering; Algoritma K-Medoids; TPT

Abstract—The Open Unemployment Rate (OER) in Indonesia decreased in February 2024 to 4.82%, showing an improvement compared to February 2023. Despite the decline in TPT, there are still regions with TPT reaching 7.02%, which could potentially lead to negative consequences such as increased crime. Efforts to address TPT include increasing economic growth, developing the quality of education and training. This research utilizes clustering in data mining. The number of clusters formed was 3 clusters with a DBI value of -1.685. This study uses K-Medoids clustering to group 38 provinces based on TPT. Of the 38 data, there is incomplete data so preprocessing is done using the "filter example" operator in rapidminer to eliminate incomplete data so that there are 34 data that will be used in this study (after preprocessing). The results show 4 provinces with the highest TPT (Riau Islands, DKI Jakarta, West Java, and Banten) with a percentage of 11.76%.

Keywords: Data Mining; Clustering; K-Medoids Algorithm; TPT

1. PENDAHULUAN

Pengangguran terbuka menurut Badan Pusat Statistik (BPS) adalah Mereka yang tidak memiliki pekerjaan namun aktif mencari pekerjaan, Mereka yang tidak memiliki pekerjaan namun sedang mempersiapkan usaha dan Mereka yang tidak memiliki pekerjaan dan tidak mencari pekerjaan karena merasa tidak mungkin mendapatkannya. Badan Pusat Statistik (BPS) melaporkan data pengangguran Februari 2024. Jumlah pengangguran terbuka di Indonesia mencapai 4,82 persen, menunjukkan penurunan 0,63 persen poin dibandingkan Februari 2023. Total angkatan kerja mencapai 149,38 juta orang, naik 2,76 juta orang dari tahun sebelumnya. Hal ini menunjukkan peningkatan partisipasi dalam angkatan kerja. Sektor yang mengalami peningkatan terbesar dalam penyerapan tenaga kerja adalah Penyediaan Akomodasi dan Penyediaan Makan Minum, dengan penambahan 0,96 juta orang. Walaupun dari tahun sebelumnya tingkat pengangguran menurun, tapi masih ada wilayah dengan persentase tingkat pengangguran mencapai 7,02% yang menunjukkan jumlah yang terbilang cukup besar dibandingkan rata-rata Jumlah pengangguran terbuka tahun 2024. Tingginya tingkat pengangguran di Indonesia dapat membawa berbagai konsekuensi negatif bagi individu, keluarga, masyarakat, dan bahkan negara secara keseluruhan. Pengangguran dapat meningkatkan risiko kejahatan. Orang-orang yang putus asa karena tidak memiliki pekerjaan mungkin lebih mudah terjerumus ke dalam kegiatan kriminal. Mengingat jumlah kriminal per Februari 2024 sebanyak 31.365 kasus yang dapat disebabkan karena masih tingginya angka pengangguran tersebut.

Mengatasi Tingkat Pengangguran Terbuka yang Tinggi di Indonesia yaitu Meningkatkan Pertumbuhan Ekonomi dengan menciptakan lapangan kerja, seperti industri kreatif, pariwisata, ekonomi digital, dan sektor-sektor padat karya. Selain itu perlu Mengembangkan Kualitas Pendidikan dan Pelatihan juga. Walaupun sekarang sudah dilakukan hal-hal tersebut, akan tetapi masih ada beberapa provinsi yang perlu perhatian lebih dari pemerintah dibanding provinsi lainnya karena jumlah populasi juga dapat mempengaruhi hal tersebut. Untuk melakukan evaluasi tersebut tentunya memerlukan data-data sebelumnya agar dapat dievaluasi lebih mendalam lagi. Data Mining adalah salah satu bidang yang dapat digunakan untuk mencari informasi baru (informasi tersembunyi) dari data-data yang sudah ada. Didalam data mining ada berbagai kelompok yang dapat digunakan sesuai dengan permasalahan yang di teliti, diantaranya Asosiasi yang digunakan untuk pencarian pola, prediksi, clustering untuk melakukan

pengelompokan terhadap beberapa cluster data kedalam cluster yang akan di bentuk, selain itu ada juga klasifikasi yang digunakan untuk pelabelan. Pada penelitian ini, dari beberapa kelompok data mining tersebut maka penelitian ini akan melakukan clusterisasi dimana akan dilakukan pengelompokan tingkat pengangguran terbuka terhadap 38 provinsi yang ada di Indonesia. Didalam clusterisasi tersebut tentunya ada beberapa algoritma yang dapat digunakan salah satunya algoritma K-Medoids.

Algoritma K-Medoids merupakan teknik pengelompokan data yang populer untuk membagi data menjadi kelompok-kelompok homogen. Berbeda dengan algoritma K-Means yang menggunakan rata-rata sebagai pusat kelompok, K-Medoids memanfaatkan objek-objek dalam data yang paling mewakili karakteristik kelompoknya masing-masing, yang disebut medoids. Medoids ini bukan nilai rata-rata abstrak, melainkan objek data yang sebenarnya, sehingga lebih mudah diinterpretasikan dan tidak terpengaruh oleh outlier (titik data yang jauh dari data lain).

Beberapa peneliti lainnya telah melakukan penelitian yaitu penelitian yang dilakukan oleh Fatchi Arifatul Ulya dan penelitian lainnya pada tahun 2023. Penelitian tersebut mengelompokkan tingkat pengangguran terbuka di Jawa Tengah dengan menggunakan data sekunder dari Badan Pusat Statistik Jawa Tengah yang mencakup 35 kabupaten/kota sebagai sampel datanya. Metode pengelompokan yang digunakan adalah K-Means dengan 3 atribut, yaitu jumlah penduduk, UMK, dan jumlah industri. Jumlah cluster optimal yang ditentukan melalui metode Elbow adalah 4. Hasil analisis menunjukkan bahwa terdapat 4 kelompok kabupaten/kota dengan tingkat pengangguran terbuka yang berbeda-beda. Klaster 1 terdiri dari 9 kabupaten/kota, klaster 2 terdiri dari 8 kabupaten/kota, klaster 3 terdiri dari 12 kabupaten/kota, dan klaster 4 terdiri dari 6 kabupaten/kota. Tingkat akurasi pengelompokan mencapai 66,8% [1].

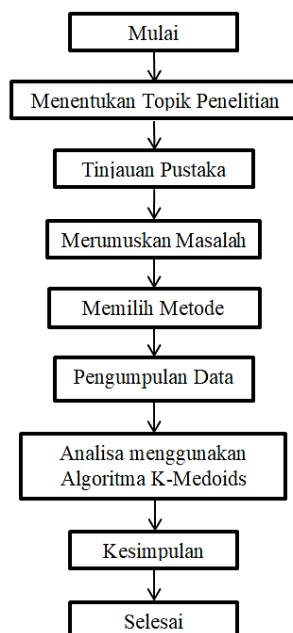
Nita Mirantika, dan rekan penelitian lainnya juga melakukan penelitian pada tahun 2023. Penelitian ini menunjukkan bahwa segmentasi pelanggan dengan algoritma K-Medoids merupakan alat yang efektif untuk memahami karakteristik pelanggan dan merumuskan strategi CRM yang tepat, sehingga meningkatkan keuntungan perusahaan. Melalui segmentasi pelanggan, perusahaan dapat mengelompokkan pelanggan berdasarkan kesamaan atribut, seperti pola pembelian (recency, frequency, monetary). Penelitian ini menggunakan algoritma K-Medoids dalam RStudio untuk mengidentifikasi 3 segmen pelanggan XYZ Store. Dimana pelanggan Loyal terdiri dari 21 individu dengan pembelian terbaru (recency tinggi), frekuensi pembelian tinggi, dan nilai transaksi tinggi (monetary tinggi). Pelanggan Tipikal terdiri dari 31 individu dengan pembelian lama (recency rendah), frekuensi pembelian sedang, dan nilai transaksi sedang (monetary sedang). Sedangkan pelanggan Baru terdiri dari 61 individu dengan pembelian terbaru (recency tinggi), frekuensi pembelian rendah, dan nilai transaksi rendah (monetary rendah) [2].

Pada tahun 2024 Elina Sarastuti dan beberapa penelitian lainnya melakukan sebuah penelitian. Penelitian ini bertujuan mengelompokkan kabupaten/kota di Jawa Barat berdasarkan tingkat kriminalitasnya menggunakan data mining. Data penelitian yang digunakan adalah data kriminalitas di Jawa Barat tahun 2019-2021 dengan 9 jenis tindak kriminalitas. Metode analisisnya menggunakan algoritma K-Medoids dengan tools Rapidminer untuk mengelompokkan pola kriminalitas ke dalam kelompok-kelompok yang memiliki kesamaan karakteristik. Pengelompokan dilakukan berdasarkan 9 jenis tindak kejahatan dan lokasi kejadiannya. Jawa Barat mengalami kenaikan kriminalitas signifikan di tahun 2022. Data BPS menunjukkan Jawa Barat berada di urutan kesembilan terkait tindak kriminalitas di tahun 2021. Kemenko Polhukam mencatat lonjakan 69% pada tahun 2022, menempatkan Jawa Barat di urutan kelima nasional. Hasil penelitian menunjukkan keakuratan yang tinggi, dengan hasil akhir penggunaan tools dan perhitungan manual yang sama. Klasterisasi status kabupaten/kota dibagi menjadi 2 klaster, yaitu aman dan rawan, dengan nilai DBI 0,761. Klaster aman terdiri dari 3 kabupaten/kota, yaitu Kabupaten Kuningan, Kabupaten Pangandaran, dan Kota Banjar. Sedangkan klaster rawan terdiri dari 16 kabupaten/kota, di antaranya Kabupaten Bandung, Kabupaten Bandung Barat, Kabupaten Bekasi, Kabupaten Bogor, dan seterusnya [3].

2. METODOLOGI PENELITIAN

2.1 Tahapan Penelitian

Tahapan penelitian merupakan serangkaian langkah atau prosedur yang wajib dijalankan dalam sebuah penelitian. Tujuannya adalah untuk memastikan kelancaran dan keberhasilan penelitian, serta memastikan semua aspeknya telah dipertimbangkan dengan matang. Langkah-langkah dalam penelitian yang digunakan yaitu pertama, peneliti menentukan topik penelitian. Topik ini harus menarik, penting, dan memiliki nilai ilmiah. Kedua, peneliti melakukan tinjauan pustaka. Tujuannya adalah untuk mempelajari literatur yang relevan dengan topik penelitian, sehingga bisa mendapatkan pemahaman yang lebih mendalam tentang masalah yang dikaji. Ketiga, peneliti merumuskan masalah penelitian. Masalah penelitian ini harus jelas, terukur, dan dapat diuji. Keempat, peneliti memilih metode penelitian yang tepat. Metode penelitian ini harus sesuai dengan jenis penelitian dan sifat data yang akan dikumpulkan. Kelima, peneliti mengumpulkan data. Data ini dapat dikumpulkan melalui berbagai cara, seperti survei, wawancara, observasi, dan eksperimen. Keenam, peneliti menganalisis data. Data ini dianalisis dengan menggunakan metode K-Medoid. Ketujuh, peneliti menarik kesimpulan dan saran. Kesimpulan ini harus berdasarkan pada hasil analisis data, dan saran ini harus ditujukan kepada pihak-pihak yang berkepentingan. Berikut adalah gambar tahapan penelitian yang akan dilakukan.



Gambar 1. Tahapan Penelitian

2.2 Data Mining

Data mining, perpaduan ilmu statistik, basis data, pencarian informasi, dan pembelajaran mesin, bagaikan mantra untuk membuka gerbang pengetahuan baru dari lautan data. Layaknya proses KDD (Knowledge Discovery in Database), data mining menjelajahi dan menganalisis data dengan berbagai metode untuk tujuan yang beragam. Inti dari data mining adalah ekstraksi informasi berharga dari kumpulan data raksasa, bagaikan mencari mutiara tersembunyi di lautan luas, dengan bantuan algoritma dan teknik dari statistika dan manajemen basis data[4]–[9]

Data mining memiliki beberapa kelompok diantaranya Prediksi, Prediksi merupakan proses memperkirakan nilai atau kejadian di masa depan berdasarkan data yang ada. Hal ini bertujuan untuk menghasilkan perkiraan yang logis atau memiliki probabilitas tinggi tentang apa yang mungkin terjadi di masa depan. Dalam data mining, prediksi menggunakan teknik statistik, machine learning, atau algoritma prediktif lainnya untuk menemukan pola atau tren dalam data yang dapat dipakai untuk memprediksi nilai atau kejadian di masa depan. Algoritma umum yang digunakan dalam prediksi termasuk KNN, Naive Bayes, C4.5, Rough Set, SVM, dan lainnya[10]–[13].

Selain prediksi, ada juga Clustering atau dikenal sebagai Pengelompokan dalam dunia data mining, merupakan suatu teknik canggih untuk mengurai data menjadi kelompok-kelompok yang bermakna. Tanpa label atau kategori yang telah ditentukan, metode ini mampu mengidentifikasi pola alami dan struktur tersembunyi yang terkandung dalam data. Proses pengelompokan ini bagaikan memilah data menjadi tumpukan-tumpukan berdasarkan kesamaan dan perbedaannya. Algoritma cerdas berperan sebagai penentu, mengelompokkan data dengan tingkat kemiripan tinggi ke dalam satu tumpukan, dan memisahkannya dari data lain yang memiliki perbedaan signifikan. Beberapa metode populer dalam pengelompokan data antara lain K-Means, AHC, dan K-Medoids. Masing-masing metode memiliki keunggulan dan kelemahannya sendiri, dan pilihan metode yang tepat bergantung pada karakteristik data dan tujuan pengelompokan[14]–[16]. Selain kedua kelompok tersebut ada kelompok lainnya lagi seperti klasifikasi, asosiasi dan estimasi[17]–[23].

2.5 Algoritma K-Medoids

Algoritma k-medoids adalah sebuah teknik dalam analisis kluster yang bertujuan untuk mengelompokkan data berdasarkan tingkat kemiripan atau jarak di antara mereka. Fokus utama dari k-medoids adalah untuk menemukan k titik data yang mewakili setiap kluster, yang disebut sebagai medoid. Proses dimulai dengan memilih secara acak k titik data awal sebagai medoid. Selanjutnya, data yang tidak menjadi medoid akan diberi label dan ditempatkan ke dalam kluster yang memiliki medoid terdekat, berdasarkan jarak yang telah ditentukan sebelumnya. Medoid untuk setiap kluster akan diperbarui dengan mencari titik data baru yang memiliki total jarak minimum dengan seluruh data dalam kluster tersebut. Iterasi ini dilanjutkan hingga medoid tidak berubah atau mencapai batas iterasi maksimum yang ditentukan. Algoritma k-medoids memiliki keunggulan dalam menangani outlier dan data yang tidak terdistribusi normal. Selain itu, k-medoids lebih efisien dalam menangani dataset besar dibandingkan dengan k-means. Namun, kelemahannya termasuk sensitivitas terhadap pemilihan medoid awal yang acak serta kompleksitas komputasi yang tinggi. Dengan menerapkan algoritma k-medoids, data dapat dikelompokkan berdasarkan kemiripan mereka, memfasilitasi analisis data, penemuan pola, dan pengambilan keputusan yang lebih baik[24][25][26][27]. Berikut proses penyelesaian masalah dengan menerapkan metode K-Medoids:

1. Tentukan jumlah cluster yang akan dibentuk (K)



2. Tentukan pusat cluster dengan cara acak (sesuai dengan jumlah k yang akan dibentuk)
3. Menghitung jarak (cost) dengan menggunakan persamaan euclidian distance berikut

$$d_{Euclidean}(X, Y) = \sqrt{\sum_{i=1}^n (X_i - Y_i)^2} \tag{1}$$

4. Tentukan letak cluster berdasarkan nilai minimum dari jarak cost (C1 hingga Cn)
5. Hitung total jarak minimum antar medoid (medoid wal)
6. Tentukan medoids yang baru (non medoid) diambil dari sampel data secara acak untuk iterasi selanjutnya (pilih data yang bukan medoids)
7. Menghitung jarak (cost) dengan menggunakan persamaan euclidian distance seperti langkah 3 dan lakukan juga langkah ke4 dan 5
8. Menghitung total simpangan (S)

$$S = \text{Medoid Cost Baru} - \text{Medoid Cost Lama} \tag{2}$$

Jika nilai $S < 0$ maka ulangi langkah 6 hingga diperoleh nilai simpangan (S) > 0 . Jika nilai $S > 0$ maka proses perhitungan selesai.

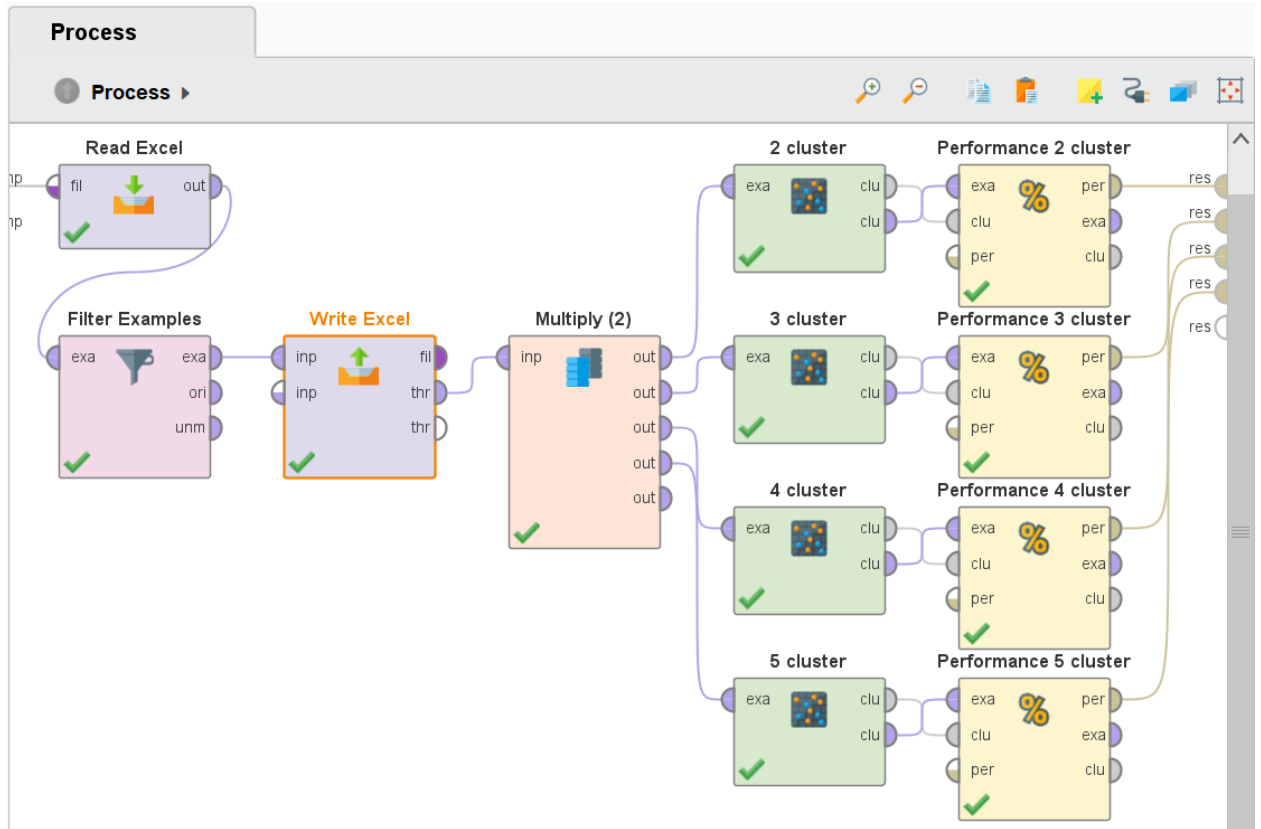
3. HASIL DAN PEMBAHASAN

Penelitian ini akan melakukan pengelompokkan terhadap tingkat pengangguran terbuka berdasarkan provinsi di Indonesia, pengelompokkan tersebut dibentuk berdasarkan data Tingkat Pengangguran Terbuka Menurut Provinsi dari tahun 2020 hingga tahun 2024. Data tersebut diambil dari data BPS yang terdiri dari 38 provinsi serta data pertahun itu ada dua jenis yaitu data bulan februari dan data bulan agustus. Sehingga sebanyak 9 atribut yang akan digunakan dalam proses pengelompokkan dengan algoritma K-Medoids. Berikut tabel sampel data yang akan dikelompokkan pada penelitian ini.

Tabel 1. Data Tingkat Pengangguran Terbuka Menurut Provinsi, 2020-2024

Provinsi	Februari 2020	Agustus 2020	Februari 2021	Agustus 2021	...	Februari 2024
Aceh	5,4	6,59	6,3	6,3	...	5,56
Sumatera_Utara	4,71	6,91	6,01	6,33	...	5,1
Sumatera_Barat	5,25	6,88	6,67	6,52	...	5,79
Riau	4,92	6,32	4,96	4,42	...	3,85
Jambi	4,26	5,13	4,76	5,09	...	4,45
Sumatera_Selatan	3,9	5,51	5,17	4,98	...	3,97
Bengkulu	3,08	4,07	3,72	3,65	...	3,17
Lampung	4,26	4,67	4,54	4,69	...	4,12
Kep._Bangka_Beli tung	3,35	5,25	5,04	5,03	...	3,85
Kep._Riau	5,98	10,34	10,12	9,91	...	6,94
Dki_Jakarta	5,15	10,95	8,51	8,5	...	6,03
Jawa_Barat	7,71	10,46	8,92	9,82	...	6,91
Jawa_Tengah	4,2	6,48	5,96	5,95	...	4,39
Di_Yogyakarta	3,38	4,57	4,28	4,56	...	3,24
Jawa_Timur	3,6	5,84	5,17	5,74	...	3,74
...
Papua_Tengah	-	-	-	-	...	2,49
Papua_Pegununga n	-	-	-	-	...	1,18

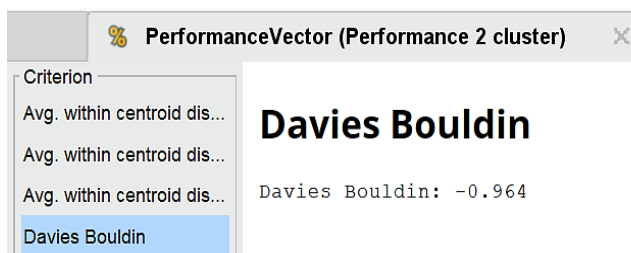
Tabel 1 terlihat bahwa ada sebagian data yang tidak ada (data tidak lengkap) sehingga perlu dilakukan preprocessing data, dalam melakukan preprocessing peneliti memanfaatkan tools rapidminer agar lebih mudah dan lebih cepat prosesnya. Selain itu, peneliti juga memanfaatkan tools rapidminer untuk menentukan jumlah cluster (kelompok) yang sesuai dengan data yang dikelompokkan. Berikut gambar 2 yang merupakan proses preprocessing dan proses penentuan jumlah cluster yang akan di bentuk.



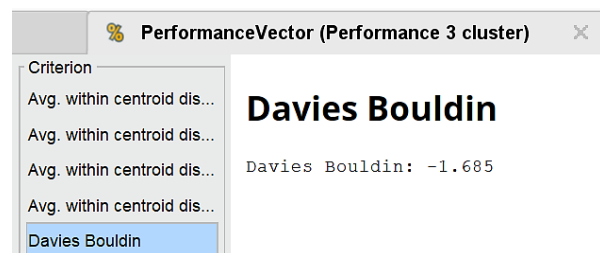
Gambar 2. Proses Preprocessing dan Penentuan Jumlah Cluster

Gambar 2 diatas merupakan proses preprocessing serta penentuan jumlah cluster yang sesuai berdasarkan sampel data serta algoritma yang digunakan. Dari proses tersebut ada beberapa operator yang di gunakan, dimulai dari read excel untuk memasukkan dataset yang bertype excel, kemudian menggunakan operator filter example untuk menghilangkan data yang tidak lengkap (dalam kasus ini ada beberapa data yang tidak ada dikarenakan ada beberapa provinsi yang mengalami pemekaran sehingga data sebelum pemekaran tidak ada), parameter yang digunakan dalam memfilter dataset yang akan digunakan dengan memilih “no missing attributes” yang akan menghilangkan semua data kosong atau tidak lengkap (Setelah selesai proses preprocessing maka selanjutnya menyimpan dataset yang telah dilakukan preprocessing dengan menggunakan operator write excel dengan membuat nama file baru serta lokasi penyimpanannya. Setelah itu maka data yang baru disimpan otomatis akan dapat digunakan dalam proses selanjutnya yaitu menentukan jumlah cluster yang akan di bentuk.

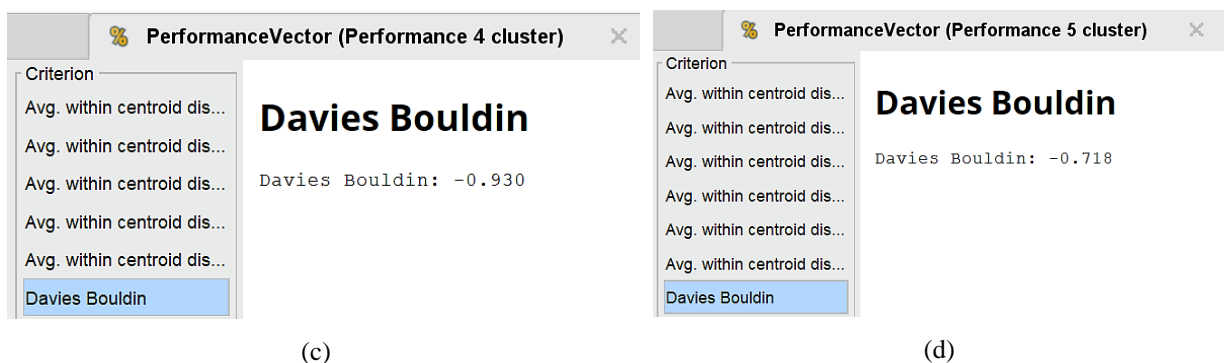
Dalam pembentukan jumlah cluster pada gambar 2 diatas, ada 3 operator tambahan setelah dilakukan preprocessing yaitu operator multiply untuk menghubungkan satu port data dengan beberapa percobaan penentuan jumlah cluster. Operator cluster yang digunakan adalah K-Medoid dengan masing-masing ketentuan jumlah clusternya, kemudian menggunakan operator Cluster Distance Performance untuk menilai performa dari masing-masing jumlah cluster yang dibentuk. Standarisasi penilaian digunakan adalah DBI (Davies-Bouldin Index), dimana semakin rendah nilainya maka pengelompokan yang dihasilkan akan lebih baik. Untuk hasil pengujian performa dapat dilihat pada gambar dibawah ini.



(a)



(b)



Gambar 3. Performa DBI (a,b,c dan d)

Berdasarkan gambar 3 diatas, performance untuk 2 cluster memperoleh nilai DBI sebesar -0,964, performance untuk 3 cluster sebesar -1,687 sedangkan untuk 4 cluster nilai DBI sebesar -0,930 dan 5 cluster nilai DBInya sebesar -0,718. Dari hasil pengujian performa tersebut maka dalam kasus ini jumlah cluster yang cocok adalah 3 cluster dikarenakan nilai DBI yang diperoleh lebih rendah dibanding yang lain. Berikut sampel data yang telah dilakukan preprocessing.

Tabel 2. Data yang telah di lakukan preprocessing

Provinsi	Februari2020	Agustus2020	Februari2021	Agustus2021	...	Februari2024
Aceh	5,4	6,6	6,3	6,3	...	5,6
Sumatera_Utara	4,7	6,9	6,0	6,3	...	5,1
Sumatera_Barat	5,3	6,9	6,7	6,5	...	5,8
Riau	4,9	6,3	5,0	4,4	...	3,9
Jambi	4,3	5,1	4,8	5,1	...	4,5
Sumatera_Selatan	3,9	5,5	5,2	5,0	...	4,0
Bengkulu	3,1	4,1	3,7	3,7	...	3,2
Lampung	4,3	4,7	4,5	4,7	...	4,1
Kep._Bangka_Belitung	3,4	5,3	5,0	5,0	...	3,9
Kep._Riau	6,0	10,3	10,1	9,9	...	6,9
Dki_Jakarta	5,2	11,0	8,5	8,5	...	6,0
Jawa_Barat	7,7	10,5	8,9	9,8	...	6,9
Jawa_Tengah	4,2	6,5	6,0	6,0	...	4,4
Di_Yogyakarta	3,4	4,6	4,3	4,6	...	3,2
Jawa_Timur	3,6	5,8	5,2	5,7	...	3,7
...
Papua_Barat	6,8	6,8	6,2	5,8	...	4,3
Papua	3,4	4,3	3,8	3,3	...	5,8

Berdasarkan tabel 2 diatas, ada perubahan jumlah sampel data dikarenakan proses preprocessing sehingga jumlah sampel data awal sebanyak 38 provinsi setelah dilakukan preprocessing berubah menjadi 34 provinsi. Setelah selesai dilakukan preprocessing serta penentuan jumlah cluster maka berikut proses untuk pengelompokkan dengan menggunakan algoritma K-Medoid

3.1 Penerapan Algoritma K-Medoids

Dalam mengelompokkan tingkat penangguran terbuka di Indonesia, algoritma K-Medoids dapat digunakan. Pertama, tentukan jumlah kelompok (klaster) yang ingin dibuat. Lalu, pilih secara acak medoid sebanyak jumlah klaster yang ditentukan. Hitung jarak (biaya) antar data menggunakan rumus jarak Euclidean dan hitung total simpangan. Berikut penjelasan langkah-langkahnya:

1. Jumlah cluster yang akan dibentuk sebanyak 3 cluster
2. Menentukan pusat 3 cluster setiap medoids dilakukan dengan cara acak (memilih 3 sampel data secara acak)

Tabel 3. Medoid Awal

Provinsi	Februari2020	Agustus2020	Februari2021	Agustus2021	...	Februari2024
Bengkulu	3,08	4,07	3,72	3,65	...	3,17
Di_Yogyakarta	3,38	4,57	4,28	4,56	...	3,24
Banten	7,99	10,64	9,01	8,98	...	7,02

3. Menghitung jarak (cost) dengan menggunakan persamaan euclidian distance sebagai berikut



$$d_{Euclidean}(X, Y) = \sqrt{\sum_{i=1}^n (X_i - Y_i)^2}$$

Data 1: Aceh

$$C_1 = \frac{\sqrt{(5,40 - 3,08)^2 + (6,59 - 4,07)^2 + (6,30 - 3,72)^2 + (6,30 - 3,65)^2 + (5,97 - 3,39)^2 + (6,17 - 3,59)^2 + (5,75 - 3,21)^2 + (6,03 - 3,42)^2 + (5,56 - 3,17)^2}}{7,60}$$

$$C_2 = \frac{\sqrt{(5,40 - 3,38)^2 + (6,59 - 4,57)^2 + (6,30 - 4,28)^2 + (6,30 - 4,56)^2 + (5,97 - 3,73)^2 + (6,17 - 4,06)^2 + (5,75 - 3,58)^2 + (6,03 - 3,69)^2 + (5,56 - 3,24)^2}}{6,35}$$

$$C_3 = \frac{\sqrt{(5,40 - 7,99)^2 + (6,59 - 10,64)^2 + (6,30 - 9,01)^2 + (6,30 - 8,98)^2 + (5,97 - 8,53)^2 + (6,17 - 8,09)^2 + (5,75 - 7,97)^2 + (6,03 - 7,52)^2 + (5,56 - 7,02)^2}}{7,56}$$

Untuk data ke-2 (Sumatera_Utara) hingga data ke-34 (Papua) lakukan pencarian jarak (cost) sesuai dengan perhitungan Data 1(Aceh). Hasil keseluruhan provinsi setelah dihitung jarak (cost) dapat dilihat pada tabel berikut.

Tabel 4. Hasil iterasi 1

Provinsi	C1	C2	C3	Jarak terdekat	Cluster
Aceh	7,60	6,35	7,56	6,35	2
Sumatera_Utara	6,93	5,64	8,22	5,64	2
Sumatera_Barat	8,06	6,80	7,09	6,80	2
Riau	3,80	2,75	11,44	2,75	2
Jambi	3,59	2,42	11,49	2,42	2
Sumatera_Selatan	3,52	2,25	11,53	2,25	2
Bengkulu	0,00	1,43	14,99	0,00	1
Lampung	2,78	1,69	12,33	1,69	2
Kep._Bangka_Belitung	3,06	1,79	12,14	1,79	2
Kep._Riau	14,69	13,35	2,68	2,68	3
Dki_Jakarta	13,00	11,70	3,45	3,45	3
Jawa_Barat	15,04	13,72	0,96	0,96	3
Jawa_Tengah	5,95	4,65	9,18	4,65	2
Di_Yogyakarta	1,43	0,00	13,69	0,00	2
Jawa_Timur	4,38	3,05	10,91	3,05	2
...
Papua_Barat	7,17	5,96	8,18	5,96	2
Papua	2,91	3,32	14,76	2,91	1
Medoid Cost				98,23	

- Menentukan 3 pusat cluster medoids yang baru (non medoid) diambil dari sampel data secara acak untuk iterasi 2 dan dapat dilihat pada tabel berikut.

Tabel 5. Medoid Baru

Cluster	Provinsi	Februari202 0	Agustus202 0	Februari202 1	Agustus202 1	...	Februari202 4
C1	Bali	2,64	4,28	3,38	3,77	...	3,17
	Kalimantan_Utar						
C2	a	4,47	5,81	5,73	5,82	...	4,2
C3	Kep._Riau	5,98	10,34	10,12	9,91	...	6,94

- Lakukan perhitungan jarak (cost) dengan menggunakan persamaan euclidian distance sesuai dengan langkah ke-3. Berikut hasil jarak cost yang diperoleh setelah dilakukan perhitungan terhadap seluruh provinsi.

Tabel 6. Hasil iterasi 2

Provinsi	C1	C2	C3	Jarak terdekat	Cluster
Aceh	7,94	2,95	7,51	2,95	3
Sumatera_Utara	7,22	2,26	7,98	2,26	3
Sumatera_Barat	8,38	3,38	6,94	3,38	3
Riau	4,14	2,15	11,38	2,15	3
Jambi	3,95	1,62	11,29	1,62	3
Sumatera_Selatan	3,83	1,62	11,21	1,62	3
Bengkulu	0,68	4,86	14,69	0,68	1
Lampung	3,19	2,34	12,14	2,34	2
Kep._Bangka_Belitung	3,34	2,00	11,73	2,00	2

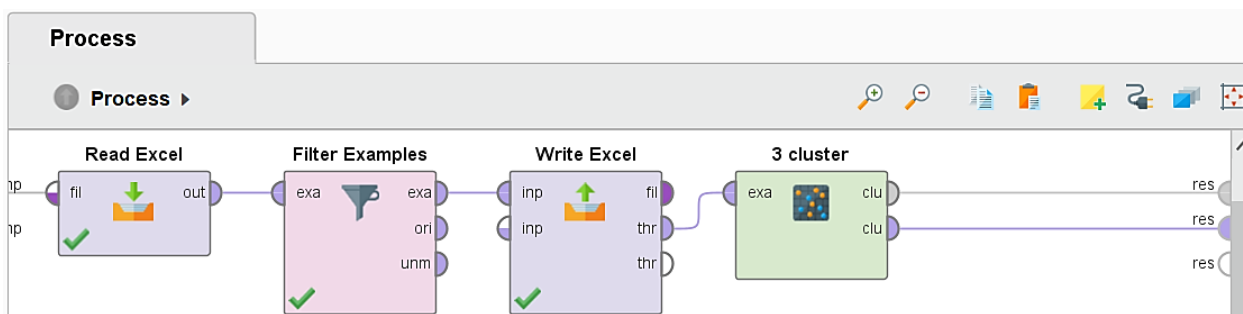
Provinsi	C1	C2	C3	Jarak terdekat	Cluster
Kep_Riau	14,94	9,93	0,00	0,00	3
Dki_Jakarta	13,21	8,41	2,77	2,77	3
Jawa_Barat	15,33	10,25	2,25	2,25	3
Jawa_Tengah	6,23	1,47	8,84	1,47	2
Di_Yogyakarta	1,75	3,53	13,35	1,75	1
Jawa_Timur	4,62	1,23	10,50	1,23	2
...
Papua_Barat	7,57	2,93	8,47	2,93	2
Papua	2,98	5,44	14,52	2,98	1
Medoid Cost Baru	73,26				

6. Menghitung total simpangan (S)

$$S = \text{Medoid Cost Baru} - \text{Medoid Cost Lama} = 73,26 - 98,23 = -24,97$$

Nilai $S < 0$ sehingga dihitung lagi proses iterasi selanjutnya dengan mengulangi langkah 2 hingga diperoleh nilai simpangan $(S) > 0$. Jika nilai $S > 0$ maka proses perhitungan selesai.

3.2 Penerapan Algoritma K-Medoids dengan Rapidminer



Gambar 4. Proses Pengelompokkan dengan menggunakan algoritma K-Medoids

Gambar 4 terlihat bahwa operator yang digunakan hanya sesuai dengan kebutuhan dalam pengelompokkan setelah di lakukan proses preprocessing dan penentuan jumlah cluster pada gambar 2 dan gambar 3 diatas, yaitu data yang telah di preprocessing serta operator algoritma K-Medoids. Setelah proses tersebut di jalankan, maka akan muncul (ditampilkan) sebuah model dari ketiga cluster tersebut yang dapat dilihat pada gambar 5 berikut ini.

Cluster Model

```
Cluster 0: 15 items
Cluster 1: 15 items
Cluster 2: 4 items
Total number of items: 34
```

Gambar 5. Cluster Model

Berdasarkan gambar 5, terlihat bahwa cluster pertama (cluster 0) memiliki anggota kelompok sebanyak 15 item, dimana jumlah anggota kelompoknya memiliki jumlah yang sama juga dengan cluster kedua (cluster 1). Sedangkan cluster ketiga (cluster 2) hanya memiliki anggota 4 provinsi saja. Untuk detail kelompok dari masing-masing cluster tersebut dapat dilihat pada tabel 7 berikut.

Tabel 6. Hasil Pembentukan Cluster Akhir

Provinsi	Februari2020	Agustus2020	Februari2021	Agustus2021	...	Februari2024	Cluster
Aceh	5.4	6.59	6.3	6.3	...	5.56	1
Sumatera_Utara	4.71	6.91	6.01	6.33	...	5.1	1
Sumatera_Barat	5.25	6.88	6.67	6.52	...	5.79	1
Riau	4.92	6.32	4.96	4.42	...	3.85	1
Jambi	4.26	5.13	4.76	5.09	...	4.45	1

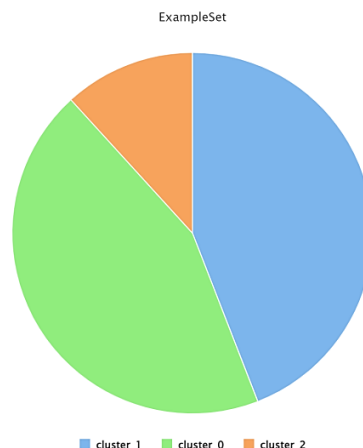
Provinsi	Februari2020	Agustus2020	Februari2021	Agustus2021	...	Februari2024	Cluster
Sumatera_Selatan	3.9	5.51	5.17	4.98	...	3.97	1
Bengkulu	3.08	4.07	3.72	3.65	...	3.17	0
Lampung	4.26	4.67	4.54	4.69	...	4.12	0
Kep._Bangka_Belitung	3.35	5.25	5.04	5.03	...	3.85	0
Kep._Riau	5.98	10.34	10.12	9.91	...	6.94	2
Dki_Jakarta	5.15	10.95	8.51	8.5	...	6.03	2
Jawa_Barat	7.71	10.46	8.92	9.82	...	6.91	2
Jawa_Tengah	4.2	6.48	5.96	5.95	...	4.39	1
Di_Yogyakarta	3.38	4.57	4.28	4.56	...	3.24	0
Jawa_Timur	3.6	5.84	5.17	5.74	...	3.74	1
...
Papua_Barat	6.78	6.8	6.18	5.84	...	4.31	1
Papua	3.42	4.28	3.77	3.33	...	5.81	0

Berdasarkan tabel 3, dari ketiga cluster tersebut tidak di ketahui mana cluster yang memiliki tingkat pengangguran terbuka tertinggi dan mana tingkat pengangguran terbuka yang terendah, tentunya perlu ditentukan cluster mana yang memiliki tingkat pengangguran terbuka tertinggi sehingga cluster itu akan di jadikan fokus utama pemerintah dalam menanggulangi pengangguran tersebut. Dalam menentukan hal tersebut maka di perlukan centroid medoid yang dapat dilihat ada gambar 6 berikut ini.

Attribute	cluster_0	cluster_1	cluster_2
Februari2020	3.420	6.780	7.990
Agustus2020	4.280	6.800	10.640
Februari2021	3.770	6.180	9.010
Agustus2021	3.330	5.840	8.980
Februari2022	3.600	5.780	8.530
Agustus2022	2.830	5.370	8.090
Februari2023	3.490	5.530	7.970
Agustus2023	2.670	5.380	7.520
Februari2024	5.810	4.310	7.020

Gambar 6. Centroid Akhir

Dari nilai centroid medoid pada gambar 6 diatas, maka provinsi yang dikelompokkan kedalam cluster_2 dinyatakan sebagai kelompok provinsi yang memiliki tingkat pengangguran terbuka tertinggi dari semua provinsi. Provinsi tersebut diantaranya Kep. Riau, Dki Jakarta, Jawa Barat dan Banten. Sedangkan provinsi dengan tingkat pengangguran terbuka terendah diantaranya Bengkulu, Lampung, Kep. Bangka Belitung, DI Yogyakarta, Bali, NTB, NTT, Kalimantan Tengah, Kalimantan Selatan, Sulawesi Tengah, Sulawesi Tenggara, Gorontalo, Sulawesi Barat, Maluku Utara dan Papua. Untuk persentase tingkat pengangguran terbuka tersebut dapat dilihat berdasarkan visualisasi gambar 7 berikut.



Gambar 7. Visualisasi Cluster K-medoids

Gambar 7 terlihat bahwa cluster_0 (hijau) dan cluster_1 (biru) masing-masing terdiri dari 15 provinsi didalam satu kelompok sehingga persentasenya sebesar 44,12% dan cluster yang lebih sedikit hanya ada 4 provinsi dengan jumlah persentase hanya 11,76% yaitu cluster_2 (orange).

4. KESIMPULAN

Kesimpulan yang dapat diambil setelah diterapkan algoritma K-Medoid dalam pengelompokan tingkat pengangguran terbuka di Indonesia menunjukkan bahwa Tingginya TPT memiliki konsekuensi negatif, seperti peningkatan kriminalitas. Upaya untuk mengatasi TPT termasuk meningkatkan pertumbuhan ekonomi, mengembangkan kualitas pendidikan dan pelatihan, dan melakukan evaluasi dengan data mining. Dengan menggunakan clusterisasi K-Medoids untuk mengelompokkan 38 provinsi berdasarkan TPT. Hasilnya menunjukkan 4 provinsi dengan TPT tertinggi (Kep. Riau, DKI Jakarta, Jawa Barat, dan Banten) dan 16 provinsi dengan TPT terendah (Bengkulu, Lampung, dll). Informasi ini penting untuk membantu pemerintah dalam merumuskan kebijakan yang tepat untuk mengatasi TPT di berbagai wilayah di Indonesia.

REFERENCES

- [1] F. A. Ulya, A. N. Abdullah, T. A. Hanan, and I. M. Nur, "Pengelompokan Tingkat Pengangguran Terbuka Di Jawa Tengah Menggunakan Metode K-Means Clustering: Grouping Open Unemployment Rates in Central Java Using the K-Means Clustering Method," *J. Data Insights*, vol. 1, no. 2, pp. 71–80, 2023.
- [2] N. Mirantika, T. S. Syamfithriani, and R. Trisudarmo, "Implementasi Algoritma K-Medoids Clustering Untuk Menentukan Segmentasi Pelanggan," *NUANSA Inform.*, vol. 17, no. 1, pp. 196–204, 2023.
- [3] E. Sarastuti, D. Mahdiana, and N. Kusumawardhany, "Klasterisasi Tindak Kriminalitas di Provinsi Jawa Barat dengan Menggunakan Algoritma K-Medoids," *Bit (Fakultas Teknol. Inf. Univ. Budi Luhur)*, vol. 21, no. 1, pp. 84–91, 2024.
- [4] R. H. Sukarna and Y. Ansori, "Implementasi Data Mining Menggunakan Metode Naive Bayes Dengan Feature Selection Untuk Prediksi Kelulusan Mahasiswa Tepat Waktu," *J. Ilm. Sains dan Teknol.*, vol. 6, no. 1, pp. 50–61, 2022, doi: 10.47080/saintek.v6i1.1467.
- [5] F. O. Lusiana, I. Fatma, and A. P. Windarto, "Estimasi Laju Pertumbuhan Penduduk Menggunakan Metode Regresi Linier Berganda Pada BPS Simalungun," *J. Informatics Manag. Inf. Technol.*, vol. 1, no. 2, pp. 79–84, 2021.
- [6] Z. Nabila, A. Rahman Isnain, and Z. Abidin, "Analisis Data Mining Untuk Clustering Kasus Covid-19 Di Provinsi Lampung Dengan Algoritma K-Means," *J. Teknol. dan Sist. Inf.*, vol. 2, no. 2, p. 100, 2021.
- [7] A. Z. Siregar, "Implementasi Metode Regresi Linier Berganda Dalam Estimasi Tingkat Pendaftaran Mahasiswa Baru," *Kesatria J. Penerapan Sist. Inf. (Komputer dan Manajemen)*, vol. 2, no. 3, pp. 133–137, 2021.
- [8] S. S. S. A. T. Purba, V. Marudut, M. Siregar, T. Komputer, and P. B. Indonesia, "SISTEM PENDUKUNG KEPUTUSAN KELAYAKAN PEMBERIAN PINJAMAN," vol. 3, pp. 25–30, 2020, doi: 10.37600/teknikom.v3i1.131.
- [9] A. Karim, S. Esabella, K. Kusmanto, M. Hidayatullah, and S. Suryadi, "Penerapan Data Mining Untuk Pengelompokan Terhadap Kualitas Kinerja Karyawan Dengan Menggunakan Algoritma K-Medoids Clustering," *J. Media Inform. Budidarma*, vol. 8, no. 2, p. 1001, 2024, doi: 10.30865/mib.v8i2.7445.
- [10] M. M. Effendi, "Menentukan Prediksi Kelulusan Siswa Dengan Membandingkan Algoritma C4. 5 Dan Naive Bayes Studi Kasus SMKN. 1 Cikarang Selatan," *J. SIGMA*, vol. 11, no. 3, pp. 143–148, 2020.
- [11] S. Widaningsih, "Perbandingan Metode Data Mining Untuk Prediksi Nilai Dan Waktu Kelulusan Mahasiswa Prodi Teknik Informatika Dengan Algoritma C4,5, Naive Bayes, Knn Dan Svm," *J. Tekno Insentif*, vol. 13, no. 1, pp. 16–25, 2019, doi: 10.36787/jti.v13i1.78.
- [12] S. U. Putri, E. Irawan, and F. Rizky, "Implementasi Data Mining Untuk Prediksi Penyakit Diabetes Dengan Algoritma C4. 5," *Kesatria J. Penerapan Sist. Inf. (Komputer dan Manajemen)*, vol. 2, no. 1, pp. 39–46, 2021.
- [13] B. Bangun and A. K. Karim, "Pengembalian Data Yang Hilang Pada Dataset Dengan Menggunakan Algoritma K-Nearest Neighbor Imputation Data Mining," *J. Media Inform. Budidarma*, vol. 8, no. 3, p. 1706, 2024, doi: 10.30865/mib.v8i3.8014.
- [14] F. Harahap, "Perbandingan Algoritma K Means dan K Medoids Untuk Clustering Kelas Siswa Tunagrahita," *TIN Terap. Inform. Nusant.*, vol. 2, no. 4, pp. 191–197, 2021.
- [15] M. A. Rofiq, A. Qoiriah, S. Kom, and M. Kom, "Pengelompokan Kategori Buku Berdasarkan Judul Menggunakan Algoritma Agglomerative Hierarchical Clustering Dan K-Medoids," *J. Informatics Comput. Sci.*, vol. 2, no. 03, pp. 220–227, 2021.
- [16] B. Harli Trimulya Suandi As and L. Zahrotun, "PENERAPAN DATA MINING DALAM MENGELOMPOKKAN DATA RIWAYAT AKADEMIK SEBELUM KULIAH DAN DATA KELULUSAN MAHASISWA MENGGUNAKAN METODE AGGLOMERATIVE HIERARCHICAL CLUSTERING (Implementation Of Data Mining In Grouping Academic History Data Before Students And Stud)," *J. Teknol. Informasi, Komput. dan Apl.*, vol. 3, no. 1, pp. 62–71, 2021.
- [17] M. Azhari, Z. Situmorang, and R. Rosnelly, "Perbandingan Akurasi, Recall, dan Presisi Klasifikasi pada Algoritma C4.5, Random Forest, SVM dan Naive Bayes," *J. Media Inform. Budidarma*, vol. 5, no. 2, p. 640, 2021, doi: 10.30865/mib.v5i2.2937.
- [18] A. Damuri, U. Riyanto, H. Rusdianto, and M. Aminudin, "Implementasi Data Mining dengan Algoritma Naive Bayes Untuk Klasifikasi Kelayakan Penerima Bantuan Sembako," *JURIKOM (Jurnal Ris. Komputer)*, vol. 8, no. 6, pp. 219–225, 2021.
- [19] I. A. Nikmatun and I. Waspada, "Implementasi Data Mining untuk Klasifikasi Masa Studi Mahasiswa Menggunakan Algoritma K-Nearest Neighbor," *Simetris J. Tek. Mesin, Elektro dan Ilmu Komput.*, vol. 10, no. 2, pp. 421–432, 2019.
- [20] H. Hozairi, A. Anwar, and S. Alim, "Implementasi Orange Data Mining Untuk Klasifikasi Kelulusan Mahasiswa Dengan Model K-Nearest Neighbor, Decision Tree Serta Naive Bayes," *Netw. Eng. Res. Oper.*, vol. 6, no. 2, pp. 133–144, 2021.
- [21] K. Erwansyah, B. Andika, and R. Gunawan, "Implementasi Data Mining Menggunakan Asosiasi Dengan Algoritma Apriori Untuk Mendapatkan Pola Rekomendasi Belanja Produk Pada Toko Avis Mobile," *J. Teknol. Sist. Inf. dan Sist. Komput. TGD*, vol. 4, no. 1, pp. 148–161, 2021.



- [22] A. S. L. T. T. H. Hafizah, “Data Mining Estimasi Biaya Produksi Ikan Kembang Rebus Dengan Regresi Linier Berganda,” *J. Sist. Inf. Triguna Dharma (JURSI TGD)*, no. Vol 1, No 6 (2022): EDISI NOVEMBER 2022, pp. 888–897, 2022.
- [23] Y. L. Nainel, E. Buulolo, and I. Lubis, “Penerapan Data Mining Untuk Estimasi Penjualan Obat Berdasarkan Pengaruh Brand Image Dengan Algoritma Expectation Maximization (Studi Kasus: PT. Pyridam Farma Tbk),” *JURIKOM (Jurnal Ris. Komputer)*, vol. 7, no. 2, p. 214, 2020, doi: 10.30865/jurikom.v7i2.2097.
- [24] A. Supriyadi, A. Triayudi, and I. D. Sholihati, “Perbandingan algoritma k-means dengan k-medoids pada pengelompokan armada kendaraan truk berdasarkan produktivitas,” *JUPI (Jurnal Ilm. Penelit. dan Pembelajaran Inform.)*, vol. 6, no. 2, pp. 229–240, 2021.
- [25] B. Produktivitas, “Perbandingan algoritma k-means dengan k-medoids pada pengelompokan armada kendaraan truk berdasarkan produktivitas,” vol. 06, pp. 229–240, 2021.
- [26] D. Wahyuli, I. Parlina, A. P. Windarto, and D. Suhendro, “Mengelompokkan Garis Kemiskinan Menurut Provinsi Menggunakan Algoritma K-Medoids,” no. September, pp. 452–461, 2019.
- [27] S. R. Ningsih, I. S. Damanik, A. P. Windarto, and H. Satria, “Analisis K-Medoids Dalam Pengelompokkan Penduduk Buta Huruf Menurut Provinsi,” no. September, pp. 721–730, 2019.