PENERAPAN DATA MINING DALAM MENGELOMPOKAN JUMLAH UMKM BERDASARKAN KABUPATEN KOTA MENGGUNAKAN *K-MEANS CLUSTERING*

Iin[1], Rizal Fadila[2], Ade Rizki Rinaldi[3], Fathurrohman[4]
[1, 3, 4] Rekayasa Perangkat Lunak, STMIK IKMI Cirebon
[2] Teknik Informatika, STMIK IKMI Cirebon
Jl. Perjuangan No.10 B Majasem Kec. Kesambi, Kota Cirebon Indonesia
iin@gmail.com

ABSTRAK

Jumlah UMKM di indonesia mengalami permasalahan setiap tahun yakni jumlah pertumbuhannya, jumlah tersebut menunjukan bahwa adanya perkembangan ekonomi. Di pulau jawa, Jawa Barat menjadi salah satu provinsi dengan jumlah unit UMKM terbanyak. Tersebar dari berbagai kabupaten atau kota dengan berbagai jenis usahanya, jumlahnya belum mengalami pemerataan bisa dipengaruhi oleh letak geografis dan demografis wilayah kabupaten atau kota. Maka perlu adanya pengelompokan kabupaten atau kota berdasarkan jumlah UMKM untuk mengetahui berapa jumlah pengelompokan yang terbentuk. Tujuan pada penelitian ini adalah untuk mengelompokan kabupaten atau kota di provinsi Jawa Barat berdasarkan tahun dan jumlah usaha mikro kecil menengah (UMKM) menerapkan teknik data mining menggunakan Algoritma *K-Means clustering*. Data penelitian ini berasal dari opendata.jabarprov.go.id dengan jumlah data sebanyak 216 dan bersumber dari 27 kabupaten atau kota, dari tahun 2016 hingga 2023 di Jawa Barat. Penelitian ini bisa memberikan pemahamam bagi para pelaku UMKM, masyakarat dan pemerintah berkontribusi dalam pengambilan keputusan di lingkup provinsi Jawa Barat. Hasil penelitian ini menunjukan bahwa *cluster* terbaik terdapat pada nilai K=2 dengan nilai DBI 0.619 dari percobaan sebanyak 20 kali, yang membagi data menjadi 2 *cluster* yaitu cluster 0 sebanyak 112 items sebagai cluster dengan jumlah UMKM tinggi dan *cluster* 1 sebanyak 104 items dengan jumlah UMKM sedang dari total data sebanyak 216 items jumlah data UMKM. Proses dilakukan menggunakan *Rapidminer versi 10.2*.

Kata kunci : *UMKM, Data Mining, K-Means clustering, Rapidminer*

1. PENDAHULUAN

Perkembangan teknologi sekarang mempunyai peran yang sangat penting dalam terciptanya kemudahan informasi dan dapat dirasakan manfaatnya oleh semua sektor. Perusahaan industri adalah suatu badan yang melakukan kegiatan ekonomi, mempunyai tujuan yang menghasilkan barang dan jasa, berkedudukan pada suatu bangunan atau lokasi tertentu mempunyai catatan pengelolaan produksi dan stuktur biayanya sendiri, serta mempunyai seorang atau lebih pekerja. Sebuah badan usaha yang tertanggung jawab ats bisnis ini [1]. Banyak sekali jumlah UMKM di provinsi Jawa Barat tersebar di setiap kota atau kabupaten, salah satu jenis usaha yang bergerak yaitu di bidang kuliner, fashion, kreatif dan kerajinan, kelompok budidaya, pedagang dan lainya. Dari proyeksi jumlah yang ada di provinsi Jawa Barat perlu adanya pengelompokan kabupaten atau kota berdasarkan jumlah UMKM yang ada. Karena jumlahnya belum mengalami pemerataan, hal ini bisa dipengaruhi oleh letak geografis dan demografis. Pada penelitian terdahulu yang dibuat oleh Wahyu Sudrajat, Idahm Cholid dan Johannes Petrus "Penerapan Algoritma *K-Means Clustering* pada UMKM kelompok menggunakan Rapidminer" tentang pengelompokan dan pendampingan UMKM menjadi 3 kelompok yaitu mikro, kecil dan menengah berdasarkan omset dan aset menghasilkan kesimpulkan berhenti pada iterasi kedua dan menghasilkan 3 (tiga) cluster. Terbit pada jurnal JUPITER, Vol. 14 No. 2 [2].

Penelitian yang dilakukan oleh Herwinda Kurnaedi, Rijal Abdul Hakim, Mohamad Jajuli dan Jajam Haeruman Jaman " Pemetaan UMKM dalam Upaya Pengentasan Kemiskinan dan Penyerapan Tenga Kerja Menggunakan Algorima K-Means berdasarkan lokasi, jumlah UMKM, penduduk miskin, dan jumlah pencari kerja" . Menghasilkan cluster tinggi ditemukan terdiri dari enam kecamatan, sedang 8 kecamatan dan rendah 3 kecamatan algoritma K-Means dapat diterapkan untuk mengetahui pengelompokan UMKM di Kabupaten Purwakarta. Terbit pada jurnal Journal of Applied Informatics and Computing(JAIC) Vol.6, No.2[3] .

Penelitian yang dilakukan oleh Elly Muningsih, Ina Maryani dan Vembria Rose Handayani pada tahun 2021 pengujian dilakukan menggunakan aplikasi rapidminer dengan judul "Penerapan K-Means dan Optimasi Jumlah Cluster dengan Index Davies Bouldin untuk Clustering Provinsi berdasarkan Potensi Desa", dan setelah dilakukan pengolahan data DBI dicari nilai minimumnya. Jumlah minimum cluster 0,175 di jumlah cluster 3 Terbit pada jurnal Sains dan manajemen Vol.9, No.1[4].

Tujuan pada penelitian ini adalah untuk mengelompokan kabupaten atau kota di provinsi Jawa Barat berdasarkan tahun dan jumlah usaha mikro kecil menengah (UMKM) dengan menerapkan teknik data mining menggunakan Algoritma *K-Means clustering*.

mengevaluasi hasil dari proses clustering dengan nilai *index davies bouldin*. Implikasi hasil dari penelitian ini memiliki potensi yang baik bagi para pelaku UMKM untuk mengetahui analisis wilayah mereka masuk kedalam kategori kelompok yang mana, terhadap masyarakat umum bisa menambah wawasan atau penegetahuan baru mengenai hasil pengelompokan, dan bagi pemerintahan baik daerah atau pusat bisa memberikan kontribusi dalam pengambilan keputusan terhadap sektor UMKM. Penelitian ini bisa menjadi referensi pemahaman terhadap proses penerapan data mining dengan algoritma *K-Means* menggunakan software *Rapidminer* 10.2. RapidMiner adalah perangkat lunak terbuka yang dapat digunakan untuk melakukan analisis pada data mining, text mining, dan analisis prediktif [5].

2. TINJAUAN PUSTAKA

Metode tinjaun pustaka yang digunakan dalam penelitian ini adalah dengan meninjau pada jurnal yang terkait dengan data mining dengan algoritma *K-Means Clustering* bagi proyeksi jumlah umkm di provinsi Jawa Barat.

2.1. Data Mining

Data mining ialah proses menemukan informasi yang bermanfaat pada repositori data yang besar secara otomatis [6]. Serta memanipulasi suatu data untuk mendapatkan informasi yang berharga [7].

2.2. Clustering

Salah satu teknik yang termasuk dalam data mining adalah teknik clustering. Data yang dikelompokkan berdasarkan teknik clustering dapat mengekstraksi informasi yang tersembunyi dengan lebih baik untuk kebutuhan analisis dan prediksi data daripada data dibandingkan dengan data yang tidak dikelompokkan [8].

2.3. Algortima K-Means

Algoritma K-Means merupakan algoritma yang mengelompokan data dengan mencoba memisahkan data menjadi beberapa kelompok sehingga data yang memiliki beberapa kemiripan dalam kelompok yang sama [9].

2.4. UMKM

Usaha Mikro Kecil Menengah (UMKM) merupakan usaha yang potensial bagi perkembangan ekonomi di indonesia [10],melakukan kegiatan ekonomi serta bertujuan menghasilkan barang atau jasa[11] .

2.5. Rapidminer

Rapidminer yang merupakan software independen untuk menganalisa data dan mesin data yang terintegrasi dengan berbagai bahasa pemrograman secara mudah [12].

2.6. Hasil literature

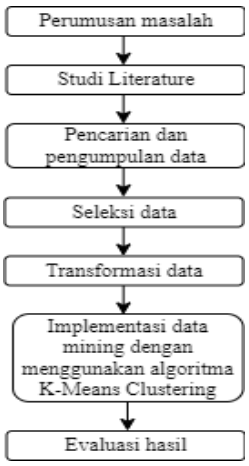
Hasil literature review yang telah dilakukan pada jurnal-jurnal dapat dijabarkan sebagai berikut :

1) Paper 1 penelitian yang dilakukan oleh membahas mengenai "Data mining Klastering Pada Perusahaan Industri Mikro di Indonesia". Menggunakan metode clustering dengan hasil penelitian yang diperoleh dari pengelompokan dari 34 provinsi adalah cluster jumlah perusahaan industrinya besar (C1) yaitu terbagi di jawa barat, jawa tengah dan jawa timur. Dan 31 provinsi lainnya dengan perusahaan industri yang relatif sedikit (C2) [11].
2) Paper 2 membahas perlunya pengelompokan jumlah UKM agar potensi penyerapan tenaga kerja yang dimiliki kelompok industri ini dapat terus dikembangkan dan dioptimalkan. pengelompokkan mengungkapkan tiga cluster prioritas tinggi, prioritas rendah, dan prioritas menengah. Hasil pengelompokkan menggunakan silhouette coefficient kemudian dievaluasi menggunakan perpustakaan Sklearn dan divisualisasikan menggunakan library matplotlib.pyplot dan geopandas, mengguanakan bahasa pemrograman python [13].
3) Paper 3 membahas cara meggunakan K-Means untuk mengelompokan kecamatan di pematangsiantar dengan usaha kecil/mikro berdasarkan industri. Data penelitian yang digunakan di peroleh dari BPS dan dibagi menjadi dua kelompok kecil dan mikro. Hasil (C1) Cluster Tertinggi dengan jumlah 8 kelompok UMKM , (C2) Cluster Sedang dengan jumlah 6 kelompok UMKM, (C3) Cluster Terendah dengan jumlah 39 kelompok UMKM pada tahun 2021 [14].

3. METODE PENELITIAN

3.1. Knowledge Discovery in Database

Adapun untuk menganalisis data sebagai bagian dari aplikasi data mining, gunakan proses langkah knowledge discovery in databases (KDD) [15].



Sumber : Penulis (2023)

Gambar 1. Metode Penelitian

Berikut merupakan penjelasan tahapanya yang akan dilakukan dalam penelitian ini :

Table 1. Penjelasan Tahapan

Tahapan	Aktivitas	Deskripsi Aktivitas
Perumusan masalah	Melakukan perumusan masalah	Merumuskan masalah berdasarkan pada latar belakang, kemudian menentukan Batasan masalah sehingga penelitian lebih terfokus pada satu masalah, setelah itu mencari studi- studi literatur terkait topik yang akan diteliti.
Studi Literature	Melakukan	Melakukan pengumpulan data dari sumber terpercaya yakni dari website open data jabar provinsi jawa barat
Pencarian dan pengumpulan data	Melakukan pengumpulan data	Data proyeksi jumlah UMKM diambil dari tahun 2016 sampai 2023. Data yang digunakan sebanyak 216 record.
Data Selection	Melakukan Seleksi Data	Menyeleksi data yang akan digunakan untuk menganalisis atribut pada data proyeksi jumlah UMKM yang sesuai dan relevan untuk dilakukan metode data mining.
Transformasi	Melakukan Transformasi data	Mengubah data kedalam sebuah format yang tepat dalam penelitian ini mengubah data menjadi numeric dari teks untuk untuk dilakukan proses data mining.
Data Mining	Melakukan pengimplementasian algoritma k-Means Clustering	Menerapkan system pengelompokan yaitu clustering dan algoritma K-Means, dengan menggunakan tools Rapidminer 10.2. Dimana hasil data yang telah dibersihkan dianalisis untuk mengungkap pola atau informasi yang berguna.
Evaluasi	Melakukan evaluasi dari analisis data	Mengetahui hasil evaluasi penerapan K-*Means* tersebut berupa nilai dari index Davies Bouldin untuk menentukan apakah hasil tersebut dapat dicapai.

3.2. Sumber Data

Sumber data yang digunakan dalam penelitian ini adalah data sekunder dari situs opendata jabar Provinsi Jawa Barat tahun 2023. Berikut link sumber https://opendata.jabarprov.go.id/.

3.3. Teknik pengumpulan

Teknik pengumpulan data pada penelitian ini melakukan pencarian dataset mengunjungi situs resmi Open Data Jawa Barat. Sedangkan pencarian literature dilakukan dengan cara mencari referensi penelitian melalui buku, jurnal, internet dan lain-lain, yaitu dengan mengambil dari ferefensi jurnal penelitian terdahulu dan proses penelitian [15].

3.4. Teknik analisis

Pada tahap data selection dilakukan pemilihan atribut mencakup nama kabupaten atau kota, jumlah proyeksi, dan tahun. Analisis dengan algoritma K-Means terhadap pengelompokan jumlah usaha mikro kecil menengah di provinsi Jawa Barat tahun 2023. Langkah untuk melakukan analisis pertama yaitu membersihkan data yang tidak valid, menghapus outlier dan mengubah format jika di perlukan. Langkah berikutnya memiilih fitur-fitur yang paling relevan untuk digunakan, mentukan jumlah (K) Cluster, selanjutnya memasukan data pada aplikasi Rapidminer, Gunakan algoritma K-Means untuk melakukan clustering pada jumlah usaha mikro kecil menengah. K-Means clustering digunakan dalam mengelompokan kabupaten atau kota di Jawa Barat berdasarkan fitur yang telah dipilih. Hasil akhir yaitu memperoleh nilai Index Davied Bouldin.

4. HASIL DAN PEMBAHASAN

4.1. Dataset

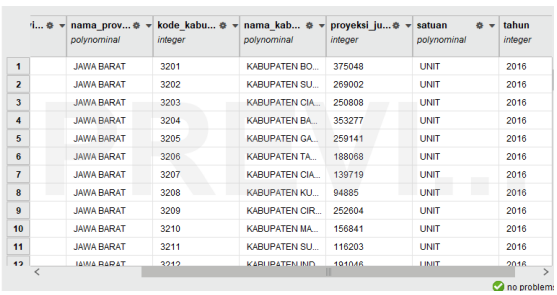
Data yang digunakan peneliti ialah proyeksi jumlah UMKM di Provinsi Jawa Barat dari tahun 2016 sampai 2023. Data sebanyak 216 items.

Table 2. Variabel dataset

Variabel	Tipe Data
kode_provinsi	Numerik
nama_provinsi	Teks
kode_kabupaten_kota	Numerik
nama_kabupaten_kota	Teks
proyeksi_jumlah_umkm	Numerik
Satuan	Teks
Tahun	Numerik

4.2. Data selection

Operator digunakan untuk memasukan data yang berbentuk file excel dari perangkat komputer kedalam Rapidminer. Setelah menemukan operator Read Excel tarik operator kedalam halaman desain dan cari data penelitian.



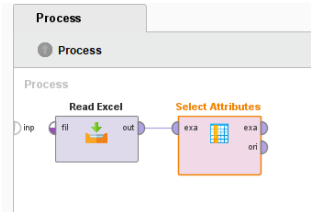
	nama_prov... polynominal	kode_kabu... integer	nama_kab... polynominal	proyeksi_ju... integer	satuan polynominal	tahun integer
1	JAWA BARAT	3201	KABUPATEN BO...	375048	UNIT	2016
2	JAWA BARAT	3202	KABUPATEN SU...	269002	UNIT	2016
3	JAWA BARAT	3203	KABUPATEN CIA...	250808	UNIT	2016
4	JAWA BARAT	3204	KABUPATEN BA...	353277	UNIT	2016
5	JAWA BARAT	3205	KABUPATEN GA...	259141	UNIT	2016
6	JAWA BARAT	3206	KABUPATEN TA...	188068	UNIT	2016
7	JAWA BARAT	3207	KABUPATEN CIA...	139719	UNIT	2016
8	JAWA BARAT	3208	KABUPATEN KU...	94885	UNIT	2016
9	JAWA BARAT	3209	KABUPATEN CIR...	252604	UNIT	2016
10	JAWA BARAT	3210	KABUPATEN MA...	156841	UNIT	2016
11	JAWA BARAT	3211	KABUPATEN SU...	116203	UNIT	2016
12	JAWA BARAT	3212	KABUPATEN IND...	191046	UNIT	2016

Sumber : Penulis (2023)

Gambar 2. Data Selection

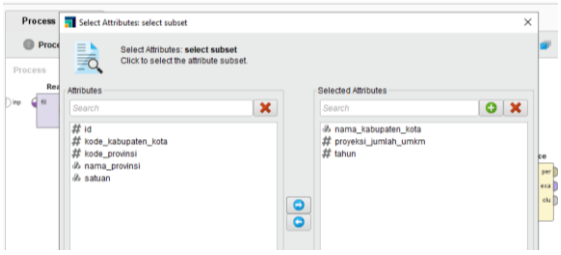
4.3. Data preprocessing

Tahap yang kedua setelah membaca file exel, yaitu dengan memilih attribute yang paling relevan untuk di analisis. Setelah itu pilih menu a subset.



Sumber : Penulis (2023)
Gambar 3. Preprocessing

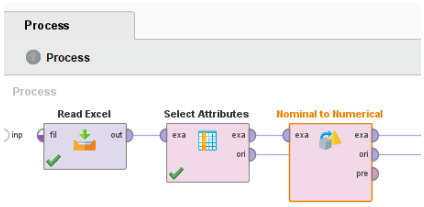
Attribute yang dipilih dalam analisis proyeksi jumlah UMKM di provinsi Jawa Barat di dalam menu select a subset yaitu terdapat 8 atribut maka yang akan di proses yaitu atribut nama kabupaten kota, proyeksi jumlah umkm, dan tahun.



Sumber : Penulis (2023)
Gambar 4. Pemilihan Atribut

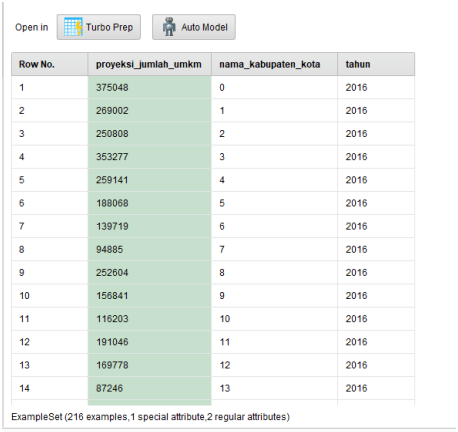
4.4. Data transformasi

Langkah transformasi mengubah data polinominal menjadi angka sehingga dapat mengolah data berdasarkan jarak. Untuk mengolah data polinominal menjadi data numerik menggunakan operatoe Nominal to Numerical [16].



Sumber : Penulis (2023)
Gambar 5. Transformasi

Parameter transformasi nominal to numerical menggunakan atribut filter type : subset, atribut yang di pindahkan yaitu nama kabupaten/kota ,coding type : unique. Setelah proses tersebut maka akan muncul hasil transformasi pada atribut nama kabupaten kota menjadi numeric dari 0-26.



Row No.	proyeksi_jumlah_umkm	nama_kabupaten_kota	tahun
1	375048	0	2016
2	269002	1	2016
3	250808	2	2016
4	353277	3	2016
5	259141	4	2016
6	188068	5	2016
7	139719	6	2016
8	94885	7	2016
9	252604	8	2016
10	156841	9	2016
11	116203	10	2016
12	191046	11	2016
13	169778	12	2016
14	87246	13	2016

Sumber : Penulis (2023)
Gambar 6. Hasil transformasi

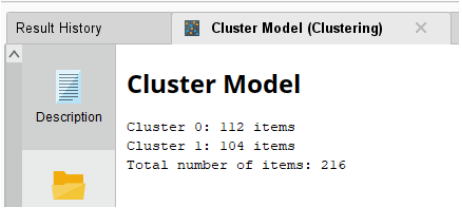
4.5. Data mining

Memasukan Algoritma *K-Means* yang tersedia pada menu Operator di *Rapidminer* Kemudian Pada Parameters terdapat K artinya *Clustering* akan dikelompokan ke beberapa *cluster* yang di inginkan yakni kluster tinggi dan kluster sedang dari semua data dengan proses max runs 10 kali putaran dengan *measure types Numerical Divergences* dan menggunakan *Euclidean Distance* dengan *max optimization* steps nya yang berjumlah 100. Grouping adalah salah satu teknik dalam data mining yang digunakan untuk menganalisis data dengan tujuan memecahkan masalah dalam pengelompokan data, dan lebih khusus lagi membagi kumpulan data menjadi beberapa subset [17]. Penentuan K terbaik ialah K=2 mengacu kepada hasil dari nilai Davies Bouldin setelah melakukan percobaan sebanyak 20 kali. Maka nilai 0,619 ditentukan sebagai nilai terbaik di percoban K=2.

Sumber : Penulis (2023)
Gambar 7. Proses Mining

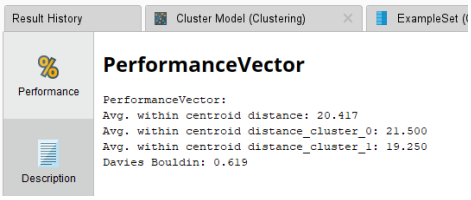
4.6. Evaluasi

Pada tahapan Evaluation ini pada tolls Rapidminner dapat menggunakan Operator Performance[18].



Sumber : Penulis (2023)
Gambar 8. *Cluster Model*

Dalam cluster model ini memiliki 2 cluster yaitu cluster 0 dengan 1. Untuk cluster 1 ditentukan sebagai kategori yang memiliki jumlah tertinggi dan untuk cluster 2 ditetapkan sebagai cluster dengan ketentuan jumlah sedang dari hasil semua data jumlah sebanayak 216 items teridentifikasi sebanyak 112 items di cluster 0 dan 104 items pada cluster 1. Untuk hasil performamceVektornya seperti gambar berikut :



Sumber : Penulis (2023)
Gambar 9. *performace Vector*

Dari hasil pengklasteran yang dilakukan mendapat nilai Davies Bouldin index 0.619 rata-rata pada jarak pusat : 20.417 rata-rata pada jarak pusat 0 : 21.500 dan rata-rata pada jarak pusat 1 : 19.250.

5. KESIMPULAN DAN SARAN

Dari hasil penelitian yang telah dijelaskanm sebelumnya, bisa disimpulkan yakni proses data mining proyeksi jumlah UMKM di provinsi Jawa Barat dengan data yang bersumber dari website open data jabar jumlah data sebanyak 216 item dan terbentuk Cluster 1 memiliki jumlah items sebanyak 112 sedangkan cluster 2 memiliki jumlah items sebanyak 104.

DAFTAR PUSTAKA

[1] E. Muningsih, I. Maryani, and ..., "Penerapan Metode K-Means dan Optimasi Jumlah Cluster dengan Index Davies Bouldin untuk Clustering Propinsi Berdasarkan Potensi Desa," *Evolusi J. Sains dan …*, 2021.

[2] I. Cholid and J. Petrus, "Penerapan Algoritma K-Means Clustering untuk Pengelompokan UMKM Menggunakan Rapidminer," *JUPITER (Jurnal Penelit. Ilmu dan Tek. …*, 2022.

[3] H. Kurniadewi, R. A. Hakim, M. Jajuli, and ..., "Pemetaan UMKM dalam Upaya Pengentasan Kemiskinan dan Penyerapan Tenaga Kerja Menggunakan Algoritma K-Means," *J. Appl. …*, 2022.

[4] E. Muningsih, I. Maryani, and V. R. Handayani, "Penerapan Metode K-Means dan Optimasi Jumlah Cluster dengan Index Davies Bouldin untuk Clustering Propinsi Berdasarkan Potensi Desa," *J. Sains dan Manaj.*, vol. 9, no. 1, pp. 95–100, 2021.

[5] F. Firmansyah and O. Nurdiawan, "Penerapan Data Mining Menggunakan Algoritma Frequent Pattern - Growth Untuk Menentukan Pola Pembelian Produk Chemicals," *JATI (Jurnal Mhs. Tek. Inform.*, vol. 7, no. 1, pp. 547–551, 2023.

[6] M. K. Hidayat and R. Fitriana, "Implementasi K-Means dan K-Medoids dalam Pengelompokan Wilayah Potensial Produksi Daging Ayam," *J. Teknol. Ind. Pertan.*, 2022.

[7] D. Astuti, "Penentuan Strategi Promosi Usaha Mikro Kecil Dan Menengah (UMKM) Menggunakan Metode CRISP-DM dengan Algoritma K-Means Clustering," *J. Informatics, Inf. Syst. Softw. Eng. Appl.*, vol. 1, no. 2, pp. 60–72, 2019.

[8] M. Murdiaty, A. Angela, and C. Sylvia, "Pengelompokkan Data Bencana Alam Berdasarkan Wilayah, Waktu, Jumlah Korban dan Kerusakan Fasilitas Dengan Algoritma K-Means," *J. Media …*, 2020.

[9] T. Hardiani, "Analisis Clustering Kasus Covid 19 di Indonesia Menggunakan Algoritma K-Means," *J. Nas. Pendidik. Tek. Inform. …*, 2022.

[10] P. Puntoriza and C. Fibriani, "Analisis Persebaran UMKM Kota Malang Menggunakan Cluster K-means," *JOINS (Journal Inf. Syst.*, vol. 5, no. 1, pp. 86–94, 2020.

[11] D. N. Batubara, A. P. Windarto, A. Wanto, and ..., "Penerapan Datamining Klastering Pada Perusahaan Industri Mikro di Indonesia," *… Komput. &Sains …*, 2020.

[12] M. Nasution and A. A. Ritonga, "IMPLEMENTASI RAPIDMINER DALAM MENGKLASIFIKASIKAN INDEKS DEMOKRASI," *J. Comput. Sci. …*, 2022.

[13] F. Fania, A. P. Windarto, and ..., "Penerapan K-Means Clustering Dalam Menentukan Banyaknya Desa/Kelurahan Menurut Keberadaan dan Jenis Industri Kecil dan Mikro (Desa)," *Bull. Inf. …*, 2022.

[14] M. Arifullah, J. T. Hardinata, and Y. P. Purba, "Implementasi Metode K-Means Dalam Klasifikasi Desa/Kelurahan Menurut Jenis Industri Kecil Mikro," *J. Informatics …*, 2022.

[15] O. Nurdiawan, A. Irma Purnamasari, and I. Ali, "Analisa Penjualan Mobil Dengan Menggunakan Algoritma K-Means Di PT. Mulya Putra Kencana," *J. Data Sci. dan Inform.*, vol. 1, no. 2, pp. 32–35, 2021.

[16] M. R. Sulistio, N. Suarna, and ..., "Analisa Penerapan Metode Clustering X-Means Dalam Pengelompokan Penjualan Barang," *J. …*, 2023.

[17] R. R. Khaerullah, N. Suarna, and ..., "Analisa Pengelompokan Dataset Komputer Menggunakan Algoritma X-Means," *J. …*, 2023.

[18] I. Ali, D. A. Kurnia, M. A. Pratama, and ..., "Klasifikasi Status Stunting Balita Di Desa Slangit Menggunakan Metode K-Nearest Neighbor," *… Sci. J. …*, 2021.