

IMPLEMENTASI ALGORITMA K-MEANS CLUSTERING UNTUK ANALISIS PERSEBARAN UMKM DI JAWA BARAT

Ahmad Azzam Arrosyad¹, Ade Irma Purnamasari², Irfan Ali³

^{1,2}Teknik Informatika, STMIK IKMI Cirebon

³Rekayasa Perangkat Lunak, STMIK IKMI Cirebon

Jl. Perjuangan No.10B, Karyamulya, Kec. Kesambi, Kota Cirebon

Azzamar27@gmail.com

ABSTRAK

Usaha Mikro Kecil dan Menengah, atau dikenal sebagai UMKM, adalah usaha produktif yang telah terbukti menciptakan dan memberikan lapangan pekerjaan dan memiliki kontribusi yang besar pada roda perekonomian di Indonesia. Pengembangan Usaha Mikro, Kecil, dan Menengah (UMKM) memainkan peran penting dalam pertumbuhan ekonomi regional, termasuk di Jawa Barat. Namun, tantangan utama yang dihadapi adalah kurangnya pemahaman yang mendalam tentang pola persebaran dan karakteristik UMKM di wilayah tersebut. Oleh karena itu, penelitian ini bertujuan untuk mengisi kesenjangan pengetahuan ini dengan menerapkan Metode Algoritma K-Means Clustering untuk mengidentifikasi variabel-variabel penting yang dapat digunakan dalam mengelompokkan UMKM, serta untuk menganalisis pola-pola persebarannya, sehingga penelitian ini dapat menjadi perhatian bagi pemerintah ataupun lembaga yang terkait dalam meningkatkan perekonomian UMKM di Jawa Barat. Ini karena diperlukan pengembangan potensi dalam pelaksanaan UMKM melalui analisis strategi persebaran UMKM dan peningkatan jumlah UMKM di Jawa Barat. Pada metode Clustering UMKM ini akan dibagi menjadi 3 bagian berdasarkan tingkat persebaran UMKM yaitu cluster 0 atau rendah yang memperoleh hasil 11 Kabupaten/Kota, cluster 1 atau sedang memperoleh hasil 3 Kabupaten/Kota, dan cluster 2 atau tinggi memperoleh hasil 12 Kabupaten/Kota. Hasil evaluasi clustering akan menggunakan metode Davies Bouldin Index (DBI) menghasilkan nilai sebesar -0,471. Dari hasil evaluasi cluster tersebut menunjukkan bahwa hasil cluster yang terbentuk adalah cluster yang terbaik. Hasil yang diperoleh itu dapat memberikan pemahaman yang mendalam tentang pola persebaran UMKM di Jawa Barat. Hasil dari penelitian ini dapat digunakan sebagai pendukung rumusan kebijakan ekonomi regional yang lebih terarah serta strategi pengembangan UMKM yang lebih efektif.

Kata kunci : *Davies Bouldin Index, Jawa Barat, K-Means Clustering, UMKM.*

1. PENDAHULUAN

Definisi dari UMKM berdasarkan Undang-Undang No 20 Tahun 2008 tentang UMKM, menggariskan sebagai berikut, Usaha Mikro adalah usaha produktif milik orang perorangan dan/atau badan usaha perorangan yang memenuhi kriteria Usaha Mikro sebagaimana diatur dalam undang-undang. Usaha Kecil adalah usaha ekonomi produktif yang berdiri sendiri, yang dilakukan oleh orang perorangan atau badan usaha yang bukan merupakan anak perusahaan atau bukan cabang perusahaan yang dimiliki, dikuasai, atau menjadi bagian baik langsung maupun tidak langsung dari Usaha Menengah atau Usaha Besar yang memenuhi kriteria Usaha Kecil sebagaimana dimaksud dalam undang-undang. Usaha Menengah adalah usaha ekonomi produktif yang berdiri sendiri, yang dilakukan oleh orang perorangan atau badan usaha yang bukan merupakan anak perusahaan atau cabang perusahaan yang dimiliki, dikuasai, atau menjadi bagian baik langsung maupun tidak langsung dengan Usaha Kecil atau Usaha Besar dengan jumlah kekayaan bersih atau hasil penjualan tahunan sebagaimana diatur dalam undang-undang [1], [2].

Kesejahteraan masyarakat suatu negara salah satunya bergantung kepada kestabilan perekonomiannya. Ada beberapa hal suatu negara

dikatakan stabil dalam ekonominya, diantaranya nilai mata uang rupiah stabil, laju pertumbuhan ekonomi yang naik, meratanya pendapatan masyarakat, tidak terjadi inflasi maupun deflasi dan lain-lain Contohnya di Jawa Barat, laju ekonominya berada di kuartal I tahun 2021. Dalam hal ini stabilitas keuangan Jawa Barat ini dapat dikatakan masih dalam kondisi baik. Namun, dilihat dari Tingkat Pengangguran Terbuka (TPT) di Jawa Barat ini terbilang cukup tinggi yaitu sebanyak 8,92%.

Hal ini berkaitan dengan kesejahteraan masyarakat yang belum pulih seutuhnya. Pada september 2020, tingkat kemiskinan mengalami peningkatan dibandingkan tahun 2019 menjadi 8,43%. Namun, angka tersebut belum memperlihatkan adanya perbaikan kondisi ekonomi yang terjadi di Jawa Barat sejak tahun 2020 sampai triwulan I 2021. Perekonomian Jawa Barat diprediksi di tahun 2021 ini akan kembali tumbuh positif sesudah berkontraksi dalam akibat dari pandemi covid-19. Untuk dapat mewujudkan prediksi dan meningkatkan laju ekonomi di Jawa Barat salah satunya yaitu dengan melakukan upaya pemerataan pendapatan masyarakat. Upaya ini dapat dilakukan dengan meningkatkan jumlah dan persebaran UMKM di Jawa Barat [1].

Banyak kebijakan, bantuan, pelatihan kegiatan lainnya telah diarahkan untuk mendukung

pengembangan dan peningkatan daya saing dan kapasitas UMKM. Namun terkadang kegiatan ini masih belum optimal antara lain dikarenakan reduksi kegiatan, tidak tepat sasaran program, masih belum berkesinambungnya program sampai dengan kegiatan yang terkadang kurang efektif. Salah satu faktor yang menyebabkan hal ini adalah belum terdatanya dengan baik serta belum adanya pemetaan yang komprehensif terhadap UMKM yang ada. Selain itu, database yang mumpuni terhadap kegiatan dan kebijakan kepada sasaran UMKM juga belum ada. Untuk itulah perlu dilakukan pembuatan kelompok-kelompok UMKM berdasarkan karakteristik yang dimiliki oleh masing-masing unit usaha. Diharapkan dengan pengelompokan (clusterisasi) ini, maka pengambil kebijakan dalam hal ini pemerintah atau lembaga-lembaga yang kompeten dapat memetakan karakteristik ataupun kebutuhan UMKM. Informasi pengelompokan UMKM ini menjadi salah satu informasi yang sangat penting dan solusi untuk mendapatkan informasi ini adalah dengan menggunakan konsep data mining.

Studi-studi sebelumnya dalam analisis persebaran Usaha Mikro, Kecil, dan Menengah (UMKM) menggunakan Metode Algoritma K-Means Clustering telah memberikan wawasan yang berharga, namun masih menyisakan beberapa aspek yang belum tergarap sepenuhnya. Sebagai contoh, penelitian yang dilakukan oleh [3] di wilayah Sumatera Utara berhasil mengidentifikasi pola persebaran UMKM dengan menggunakan pendekatan K-Means Clustering. Temuan tersebut menunjukkan bahwa UMKM cenderung berkumpul dalam klaster tertentu, namun penelitian tersebut belum mendalami faktor-faktor eksternal yang memengaruhi pola tersebut, seperti kondisi sosial-ekonomi lokal atau infrastruktur pendukung. Perhitungan algoritma K-Means yang dilakukan menghasilkan 3 cluster. Sedangkan pada penelitian yang dilakukan oleh [4].

Penggunaan algoritma K-Means digunakan untuk pengelompokan koperasi di Jawa barat. Pengelompokan yang dilakukan didasarkan atas aspek modal sendiri, modal luar dan volume usaha, dimana hasil perhitungan yang diperoleh adalah 3 (tiga) cluster yaitu cluster tinggi, sedang dan rendah [6]. Pada penelitian yang dilakukan pengelompokan UMKM dilihat dari aspek asset dan omset yang dimiliki oleh UMKM. Hal ini dipilih mengacu pada undang-undang RI Nomor 20 Tahun 2008 tentang Usaha Mikro, Kecil, dan Menengah. Hasil dari perhitungan algoritma K-Means Clustering dapat dijadikan inputan bagi para pemangku kepentingan terhadap pendampingan, pengembangan dan pengelolaan UMKM yang ada, seperti Dinas Koperasi dan UMKM atau instansi yang lain yang berkenaan dengan hal ini. Dengan harapan UMKM yang ada dapat berkembang lebih baik.

Pada penelitian ini akan menggunakan Metode Algoritma K-Means Clustering sebagai pendekatan utama untuk menganalisis persebaran Usaha Mikro, Kecil, dan Menengah (UMKM) di Jawa Barat.

Pendekatan ini akan melibatkan pengolahan data geografis dan karakteristik bisnis dari UMKM menggunakan teknik analisis cluster untuk mengidentifikasi pola-pola persebaran yang ada. Data akan dikumpulkan dari sumber-sumber yang terpercaya, seperti lembaga pemerintah terkait dan basis data industri, kemudian akan diolah dan dianalisis menggunakan perangkat lunak statistik dan komputasi data.

Penelitian ini diharapkan dapat memberikan kontribusi penting pada pemahaman tentang pola dan faktor-faktor yang memengaruhi persebaran Usaha Mikro, Kecil, dan Menengah (UMKM) di Jawa Barat dalam konteks Informatika. Temuan dari penelitian ini dapat memberikan wawasan yang lebih dalam tentang dinamika UMKM dan bagaimana faktor-faktor tertentu memengaruhi distribusi geografis mereka. Implikasi dari hasil penelitian ini mungkin sangat berharga bagi para praktisi dan pengambil kebijakan di bidang ekonomi dan pengembangan wilayah. Mereka dapat menggunakan informasi yang diperoleh untuk merancang strategi pengembangan UMKM yang lebih efektif, menyesuaikan alokasi sumber daya, dan merumuskan kebijakan yang lebih tepat.

Selain itu, hasil penelitian ini juga dapat menjadi landasan bagi penelitian lebih lanjut dalam mengembangkan model-model prediktif atau integratif yang lebih kompleks dalam analisis UMKM. Para peneliti di bidang Informatika dan analisis data juga dapat menggunakan temuan ini sebagai dasar untuk melanjutkan penelitian mereka dalam mengembangkan teknik-teknik analisis yang lebih canggih dan terarah. Potensi penggunaan algoritma-algoritma baru atau kombinasi metode analisis data yang inovatif dapat dijelajahi untuk meningkatkan pemahaman tentang UMKM dan aplikasi Informatika dalam konteks ini. Hal ini dapat menghasilkan kemajuan signifikan dalam penggunaan teknologi informasi untuk mendukung pertumbuhan sektor UMKM secara lebih efektif.

2. TINJAUAN PUSTAKA

2.1. Penelitian Terdahulu

Menurut Nursyifa dan Resti Noor Fahmi Usaha Mikro Kecil dan Menengah atau UMKM ialah usaha produktif yang sudah teruji membuka lapangan kerja dan menjadi penggerak roda perekonomian di Indonesia. Perlu adanya pengembangan potensi dalam melaksanakan UMKM dengan menganalisis strategi persebaran UMKM dan peningkatan jumlah UMKM. Penelitian ini menjelaskan cara mengimplementasikan Metode Algoritma K-Means Clustering untuk menganalisis persebaran UMKM sehingga diharapkan dapat menjadi perhatian bagi pemerintah atau institusi terkait dalam meningkatkan perekonomian UMKM di Jawa Barat. Clustering UMKM akan dibagi menjadi 3 bagian berdasarkan tingkat persebaran UMKM yaitu cluster 0 atau rendah memperoleh hasil 9 Kabupaten/Kota, cluster 1 atau sedang memperoleh hasil 15 Kabupaten/Kota, dan cluster 2 atau tinggi

memperoleh hasil 3 Kabupaten/Kota. Hasil evaluasi clustering akan dibandingkan dengan 2 metode evaluasi yaitu silhouette coefficient yang menghasilkan nilai sebesar 0,73, sedangkan metode davies bouldin index (DBI) menghasilkan nilai sebesar 0,29. Dari hasil perbandingan kedua evaluasi cluster tersebut menunjukkan hasil cluster dengan menerapkan algoritma kmeans yang terbentuk adalah klaster yang terbaik.

Penelitian oleh Feni Dwi Anggraeni, Imam Hardjanto dan Ainul Hayat membahas tentang Pengembangan Usaha Mikro, Kecil, dan Menengah (UMKM) Melalui Fasilitasi Pihak Eksternal dan Potensi Internal di Kelurahan Pandanwangi, Kecamatan Blimbing, Kota Malang). Pengembangan secara internal dari kelompok usaha menggunakan modal awal dengan tabungan sendiri, melakukan inovasi untuk hasil produksinya, memperluas jaringan pemasaran, dan melengkapi sarana dan prasarana usahanya. Selain itu, fasilitasi dari pihak eksternal yaitu Dinas Koperasi dan UKM Kota Malang telah memberikan akses UMKM terhadap sumber-sumber permodalan, mengadakan pembinaan dan pelatihan, kegiatan promosi produk, memperluas pemasaran produk, serta menyediakan sarana dan prasarana. Namun beberapa pengusaha terkendala dengan meningkatnya harga bahan baku, keterbatasan sumber daya manusia, memiliki permasalahan dalam permodalan, kurangnya sarana dan prasarana serta kurangnya akses pemasaran produk

Penelitian yang dilakukan oleh Saputra Sejauh ini usaha kecil dan menengah (UMKM) telah menunjukkan kontribusi sebesar 61,7% terhadap PDB atau Rp 8.573,89 triliun. Selain itu, UMKM mampu menyerap sekitar 97% tenaga kerja nasional. Namun pada kenyataannya, UMKM menghadapi banyak permasalahan, salah satu permasalahan yang paling umum adalah kurangnya modal usaha. Salah satu faktor penyebab lambatnya pertumbuhan UMKM di Indonesia adalah kebijakan pembangunan yang tidak tepat sasaran. Hal ini terutama berlaku pada penelitian di Kabupaten Aceh Barat. Setiap UMKM memiliki keunikan sehingga menyulitkan lembaga perbankan untuk menetapkan kebijakan pembiayaan yang konsisten [7].

Penelitian ini bertujuan untuk memetakan karakteristik UMKM dalam bentuk kelompok untuk memudahkan dalam menentukan pola pengambilan kebijakan. Metode Hierarchical Data Mining Clustering dinilai tepat karena memiliki bias yang jauh lebih rendah dibandingkan K-means. Selain itu, cara ini dapat mengurangi kompleksitas data. Berdasarkan hasil pendistribusian data UMKM sektor perikanan dengan menggunakan software Matlab 2016b terdapat 6 cluster, dan hasil pendistribusian data UMKM sektor perdagangan dengan menggunakan software Matlab 2016b terdapat 7 cluster. Setiap klaster memiliki parameter utama yang menjadikan UMKM unggul, seperti lama usaha, kepemilikan modal, proyeksi

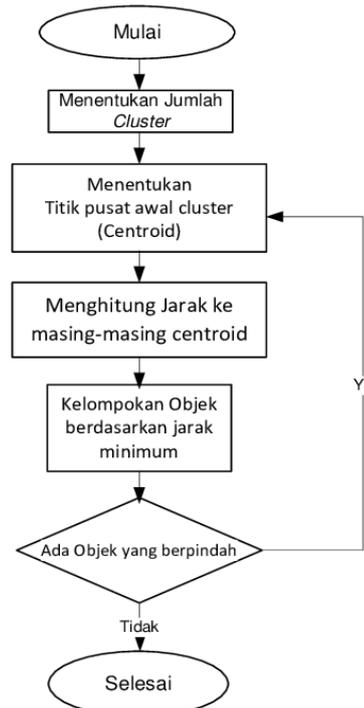
penjualan, dan rata-rata penjualan. Untuk parameter terakhir.

Penelitian (Sudrajat, Cholid dan Petrus) membahas tentang pendampingan UMKM secara tepat dapat meningkatkan kualitas UMKM, namun demikian untuk memberikan pendampingan secara tepat kepada setiap UMKM di perlukan identifikasi dan pengelompokan usaha UMKM tersebut. Pengelompokan usaha di Indonesia dibagi dalam tiga kelompok, yaitu Mikro, Kecil dan Menengah, dimana setiap kelompok didasarkan atas omset dan aset. Tujuan dari penelitian ini adalah pengelompokan UMKM yang ada di Kabupaten di Sumatera Selatan dengan menerapkan algoritma K-Means Clustering. Tahapan penelitian yang dilakukan diantaranya adalah business understanding phase, data understanding phase, data processing phase, Modeling Phase, evaluation phase dan desimination phase. Pada pengujian 15 data usaha, penelitian ini berhasil menerapkan algoritma K-Means Clustering untuk mengidentifikasi dan mengklasifikasi UMKM dengan hasil pengujian sebesar 53% data ke cluster 1 sebanyak 8 data, 40% data ke cluster 2 sebanyak 6 data dan 7% data ke cluster 3 sebanyak 1 data. Hasil perhitungan ini juga telah diuji coba menggunakan software rapidminer dan menghasilkan data yang sama [8].

2.2. Data Mining

Data mining merupakan proses di mana pengolahan data, proses mencari pola atau informasi berdasarkan metode ataupun teknik yang akan digunakan. Pada pemilihan metode yang sesuai digunakan sangat bergantung kepada tujuan dan proses KDD keseluruhan. [9]

2.3. Algoritma K-Means



Gambar 1. Flowchart Algoritma K Means

Algoritma *K-Means*, bagian dari jenis pembelajaran tanpa supervisi, difungsikan untuk mengelompokkan data menjadi beberapa kluster dengan metode pemisahan tertentu. Pembelajaran tanpa supervisi adalah metode dalam data *mining* yang bertujuan untuk mengidentifikasi pola dari berbagai variabel (atribut), tanpa menentukan variabel mana yang akan menjadi label.

K-Means merupakan salah satu teknik dalam data *mining* yang melakukan proses analisis data tanpa supervisi, di mana pendekatannya terfokus pada pengelompokan berdasarkan kesamaan atau pola yang ada. Data-data yang memiliki sifat yang mirip akan dikelompokkan dalam satu *cluster*, sedangkan data yang memiliki perbedaan karakteristik akan ditempatkan dalam *cluster* yang berbeda [10].

2.4. Clustering

Clustering adalah teknik yang mencari dan mengelompokkan data yang memiliki kesamaan karakteristik. Fokus utama dari *clustering* adalah mengelompokkan data atau objek ke dalam cluster sehingga setiap *cluster* memuat data yang serupa. Dalam proses ini, clustering berupaya menyatukan objek yang serupa ke dalam satu *cluster* sambil menjaga jarak yang jauh antar *cluster* sebisa mungkin. Hasilnya, objek dalam satu *cluster* saling mirip satu sama lain dan berbeda dengan objek dalam *cluster* lainnya [11]. Metode *clustering* juga digunakan untuk mengelompokkan berbagai entri data ke dalam kategori yang mirip, membentuk kelas objek yang berisi data serupa dan membedakan kelompok satu dari yang lain.

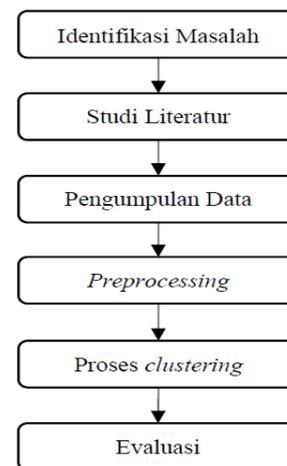
2.5. Rapid Miner

RapidMiner, sebuah perangkat lunak yang dapat diakses oleh siapa saja dan memiliki sifat terbuka (*open source*), telah diadopsi sebagai solusi dalam analisis data. Dalam platform *RapidMiner*, beragam teknik diterapkan, termasuk teknik deskriptif dan prediksi, yang digunakan untuk proses analisis data. Pengoperasian *RapidMiner* menggunakan bahasa pemrograman Java [12].

3. METODE PENELITIAN

3.1. Metode Penelitian

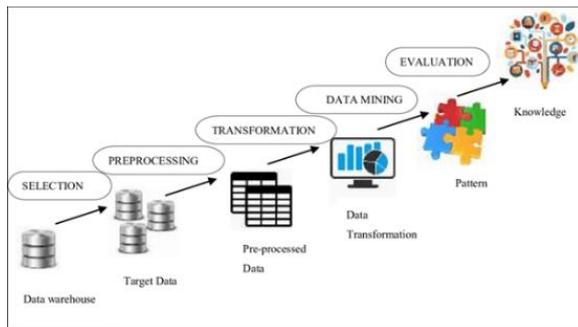
Penelitian ini menerapkan jenis penelitian deskriptif menggunakan pendekatan kuantitatif. Deskriptif kuantitatif berfungsi untuk menentukan jumlah setiap variabel dimana satu atau lebih variabel bebas, tanpa membentuk suatu ikatan atau perbedaan dengan variabel yang lain dengan tujuan memberikan penjelasan yang objektif tentang suatu kondisi. Gambar 1 menunjukkan alur metodologi penelitian yang akan digunakan dalam penelitian ini.



Gambar 2. Metode penelitian

Dari Gambar alur metode penelitian dapat dijelaskan sebagai berikut;

- a. **Identifikasi masalah**
Mengidentifikasi tujuan dan manfaat dari implementasi data mining dalam analisis persebaran UMKM yang ada di Jawa Barat.
- b. **Studi literature review**
Membaca dan menganalisis literatur atau tulisan yang berkaitan dengan topik tertentu dan memfokuskan pada pembacaan jurnal ilmiah yang relevan dengan topik penelitian yang sedang diambil.
- c. **Pengumpulan data**
Pada tahap ini, proses pemilihan, pencarian, dan pengambilan data dilakukan untuk menghasilkan sekumpulan data operasional yang akan digunakan. Dinas Perindustrian dan Perdagangan menerima data dari opendata.jabarprov.go.id, yang dipublikasikan setahun sekali. Data yang digunakan adalah jumlah unit industri kecil, menengah, dan besar yang ada di Jawa Barat dari tahun 2015 hingga 2019. Data ini dikumpulkan berdasarkan Kabupaten/Kota.
- d. **Preprocessing**
Proses pembersihan dan penyederhanaan dengan tujuan dari untuk menghilangkan suara yang ada karena data yang diambil selama proses pengambilan data pasti berupa data yang tidak terstruktur, di mana suara masih ada [13]. Selain itu, proses pembersihan data lainnya termasuk menghilangkan duplikat data, mengoreksi kesalahan, dan meninjau data yang tidak konsisten.
- e. **Proses Clustering/Data Mining**
Merupakan tahap dimana pola atau informasi yang menarik dicari dari data yang telah terpilih menggunakan teknik atau metode tertentu. Pemilihan metode dan algoritma harus dilakukan dengan tepat sesuai dengan tujuan dari *Knowledge Data Discovery (KDD)* secara keseluruhan. Proses KDD divisualisasikan pada gambar dibawah.



Gambar 3. Metode Knowledge Data Discovery (KDD)

f. Evaluation

Pola informasi yang diperoleh dari data mining harus disampaikan secara jelas kepada pihak yang berkepentingan agar mudah dipahami. Langkah ini merupakan bagian dari proses Knowledge Data Discovery (KDD), di mana pengecekan dilakukan untuk memastikan kesesuaian atau ketidaksesuaian pola informasi dengan fakta dan hipotesis yang telah ada sebelumnya [14].

3.2. Sumber Data

Sumber data pada penelitian implementasi metode K-Means Clustering dalam analisis persebaran Usaha Mikro, Kecil, dan Menengah (UMKM) di Jawa Barat menggunakan Data Sekunder, Data ini diambil dari data public melalui situs opendata.jabarprov.go.id diperoleh Dinas Perindustrian dan Perdagangan yang dipublikasikan tiap periode setahun sekali. Data yang digunakan adalah data jumlah unit industri kecil, menengah dan besar yang ada di Jawa Barat berdasarkan Kabupaten/Kota dari tahun 2016 sampai dengan 2023.

3.3. Populasi dan Sampel

Populasi UMKM yang terdapat di Jawa Barat terdapat sekitar 2 juta UMKM yang beroperasi di Jawa Barat, termasuk dalam berbagai sektor seperti pertanian, perdagangan, manufaktur, jasa, dan lainnya. Ini mencakup berbagai skala usaha mulai dari mikro, kecil, hingga menengah. Sampel ini menggunakan metode Stratified random sampling, yaitu teknik yang digunakan untuk pengambilan sampel dengan memperhatikan suatu tingkatan (strata) pada elemen populasi. Elemen populasi dibagi menjadi beberapa tingkatan (stratifikasi) berdasarkan karakter yang melekat padanya. Dalam stratified random sampling elemen populasi dikelompokkan pada tingkatan-tingkatan tertentu dengan tujuan pengambilan sampel akan merata pada seluruh tingkatan dan sampel mewakili karakter seluruh elemen populasi yang heterogen. di mana peneliti membagi UMKM menjadi beberapa kelompok berdasarkan proyeksi jumlah unit usaha mikro kecil menengah.

3.4. Teknik Pengumpulan Data

Setelah mendapatkan data sampel UMKM yang mewakili variasi yang ada dalam populasi di Jawa

Barat, berikut adalah teknik pengumpulan data yang dilakukan menggunakan teknik analisis data sekunder dengan memanfaatkan data yang tersedia dari data public melalui situs opendata.jabarprov.go.id diperoleh Dinas Perindustrian dan Perdagangan. Data sekunder ini memberikan informasi tentang jumlah unit UMKM di Jawa Barat. Data yang akurat dan bermanfaat untuk dilakukan analisis persebaran UMKM menggunakan metode K-Means Clustering.

3.5. Teknik Analisis Data

Teknik analisa data untuk penelitian menggunakan proses Knowledge Discovery In Database (KDD) dan menggunakan algoritma K-means dalam pengelompokkannya. Penggunaan proses KDD berfungsi agar tahapan yang dilakukan berjalan secara sistematis. Proses ini untuk mencari secara sistematis suatu ikatan baru di dalam hubungan antara setiap variabel. Data mining serta KDD sering digunakan secara bergantian berfungsi untuk menggambarkan proses menggali data tersembunyi dari dasar informasi yang besar. Salah satu tahapan dalam proses KDD merupakan data mining. Kedua hal itu memiliki konsep yang berbeda, tetapi memiliki kaitan satu sama lain. Tahapan Knowledge Discovery in Databases (KDD) terdiri dari Selection, Preprocessing, Transformation, Data mining, Evaluation, dan Knowledge.

4. HASIL DAN PEMBAHASAN

4.1. Hasil Penelitian

4.1.1. Data Selection

Pada tahap data selection dilakukan proses pencarian data dan pengambilan data yang akan digunakan. Pengambilan data diambil dari data public melalui situs opendata.jabarprov.go.id diperoleh Dinas Perindustrian dan Perdagangan yang dipublikasikan tiap periode setahun sekali. Data yang digunakan adalah data jumlah unit industri kecil, menengah, dan besar yang ada di Jawa Barat berdasarkan Kabupaten/Kota dari tahun 2016 sampai dengan 2023.

4.1.2. Preprocessing (Data Cleaning)

Tabel 1. Dataset

id	kode provinsi	nama provinsi	kode kabupaten kota	nama kabupaten kota	proyeksi jumlah UMKM	satuan	tahun
1	32	JAWA BARAT	3201	KABUPATEN BOGOR	375048	UNIT	2016
2	32	JAWA BARAT	3202	KABUPATEN SUKABUMI	269002	UNIT	2016
3	32	JAWA BARAT	3203	KABUPATEN CIANJUR	250808	UNIT	2016
4	32	JAWA BARAT	3204	KABUPATEN BANDUNG	353277	UNIT	2016
...
213	32	JAWA BARAT	3276	KOTA DEPOK	247207	UNIT	2023
214	32	JAWA BARAT	3277	KOTA CIMAH	86635	UNIT	2023
215	32	JAWA BARAT	3278	KOTA TASIKMALAYA	138703	UNIT	2023
216	32	JAWA BARAT	3279	KOTA BANJAR	39422	UNIT	2023

Preprocessing adalah proses pembersihan dan perubahan data mentah menjadi data yang siap digunakan pada tahap berikutnya. Dataset yang didapatkan berupa data yang terdiri dari 216 record. Pada data di bawah terdapat data id, kode provinsi, nama provinsi, kode kabupaten kota, nama kabupaten kota, jumlah unit, satuan serta tahun. Kode provinsi

untuk wilayah Provinsi Jawa Barat ialah 32. Berikut detail isi dataset awal yang terdapat pada tabel 1.

Tabel 1 memperlihatkan *dataset* jumlah UMKM yang ada di Jawa Barat dengan terdapat beberapa atribut data. Pada penelitian ini atribut yang digunakan hanya nama kabupaten kota, jumlah unit, dan tahun. Maka dari itu pada proses *preprocessing* akan dilakukan penghapusan atribut yang tidak dibutuhkan, dapat dilihat pada tabel 2 berikut.

Tabel 2. hasil penghapusan atribut

Id	nama_kabupaten_kota	proveksi_jumlah_umkm	tahun
1	KABUPATEN BOGOR	375048	2016
2	KABUPATEN SUKABUMI	269002	2016
3	KABUPATEN CIANJUR	250808	2016
4	KABUPATEN BANDUNG	353277	2016
5	KABUPATEN GARUT	259141	2016
6	KABUPATEN TASIKMALAYA	188068	2016
7	KABUPATEN CIAMIS	139719	2016
8	KABUPATEN KUNINGAN	94885	2016
9	KABUPATEN CIREBON	252604	2016
10	KABUPATEN MAJALENGKA	156841	2016
...
213	KOTA DEPOK	247207	2023
214	KOTA CIMAHI	86635	2023
215	KOTA TASIKMALAYA	138703	2023
216	KOTA BANJAR	39422	2023

Proses menghapus atribut yang tidak dibutuhkan sehingga hanya terdapat atribut id, nama_kabupaten_kota, jumlah_unit, dan tahun. Di mana data tersebut digunakan untuk proses selanjutnya. Perubahan data dilakukan dengan mengubah yang tadinya data dalam satu kolom menunjukkan jumlah UMKM tiap tahunnya, pada tahap ini mengubah jumlah UMKM tersebut menjadi beberapa kolom yang dikelompokan berdasarkan tahun. Kemudian menambahkan atribut jumlah yang menunjukkan jumlah UMKM selama tahun 2019 s.d. 2023.

4.1.3. Transformasi

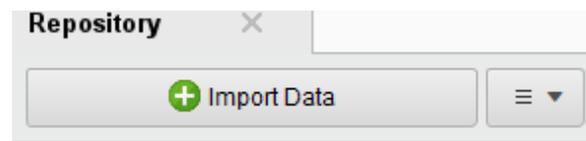
Pada proses ini menghasilkan 27 data yang berisi nama kabupaten/kota di Jawa Barat yang terdapat pada kolom Kabupaten/Kota. Pada setiap kabupaten/kota terdapat data jumlah UMKM sesuai tahunnya dari tahun 2019 hingga 2023 yang terletak pada kolom 2019, 2020, 2021, 2022 dan 2023.

Tabel 3. Transformasi data

Row No.	NAMA KABUPATEN/KOTA	2019.0	2020.0	2021.0	2022.0	2023.0	JUMLAH
1	KABUPATEN BOGOR	449059	476844	506347	537676	570943	2540869
2	KABUPATEN SUKABUMI	322086	342015	363176	385646	409507	1822430
3	KABUPATEN CIANJUR	300302	318882	338612	359563	381810	1699169
4	KABUPATEN BANDUNG	422992	449164	476954	506465	537801	2303376
5	KABUPATEN GARUT	310279	329477	348863	371510	394496	1755625
6	KABUPATEN TASIKMALAYA	225181	239114	253908	269618	286300	1274121
7	KABUPATEN CIAMIS	167291	177642	188633	200304	212697	946567
8	KABUPATEN KUNINGAN	113609	120639	128103	136029	144445	642925
9	KABUPATEN CIREBON	302452	321166	341037	362138	384544	1711337
10	KABUPATEN MAJALENGKA	187792	199411	211749	224850	238762	1062564
11	KABUPATEN SUMEDANG	139134	147743	156884	166591	176898	787250
12	KABUPATEN INDRAMAYU	226747	242900	257929	273887	290833	1294296
13	KABUPATEN SUBAHG	203282	215859	229215	243397	258457	1150210
14	KABUPATEN PURWAKARTA	104463	110926	117790	125078	132816	591073

4.1.4. Data Mining

Tahapan ini yaitu proses pengolahan data dengan metode yang digunakan algoritma *k-means*. Data diolah menggunakan software Rapid Miner, Di mana tahapan awal pada proses *clustering k-means* yaitu menentukan jumlah *cluster* yang akan dibentuk. Langkah awal yaitu dengan Import data. Pada gambar data dari file Excel di masukan ke dalam aliran kerja (workflow) Rapid Miner untuk diolah lebih lanjut dalam proses data mining.



Gambar 4. Import data

Pada gambar diatas menampilkan proses setelah data di masukan kedalam rapid miner untuk di pilih cell range nya. Kemudian langkah selanjutnya adalah menambahkan operator retrieve untuk membaca data set yang telah kita masukan sebelumnya, terlihat pada gambar 4 tampilan operator setelah masuk kedalam lembar kerja rapid miner.



Gambar 5. Retrieve

Pada gambar operator selanjutnya yang akan di masukan kedalam lembar kerja rapid miner adalah multiply untuk mengambil objek dari port input atau objek dari operator Read Excel dan mengirimkan salinannya.



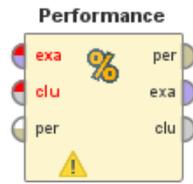
Gambar 6. Multiply

Setelah itu kemudian operator selanjutnya yang akan kita panggil adalah K-Means yang akan mengelompokan atau membuat cluster.



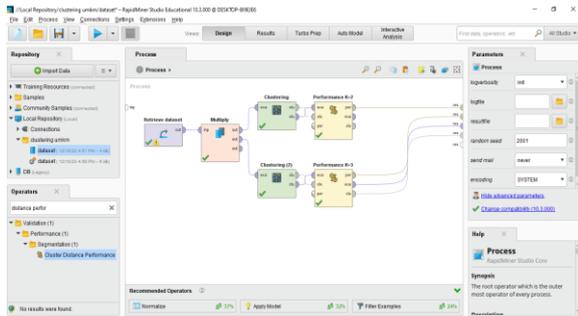
Gambar 7. Clustering

Pada gambar ini terlihat operator performance untuk melihat hasil akhir cluster.



Gambar 8. Cluster distance performance

Setelah menjalankan projek maka akan tampil hasil dari rapid miner seperti pada gambar 8.



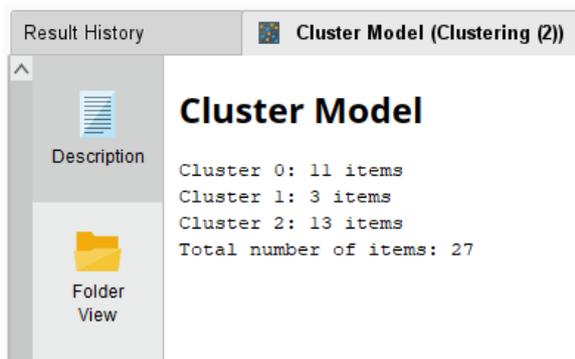
Gambar 8. Project Rapid Miner

Setelah semua oprator di masukan kedalam lembar kerja rapid miner, langkah selanjutnya adalah menjalankan projek yang telah kita susun opratornya. Pada gambar menampilkan gambar runing atau menjalankan projek yang kita buat

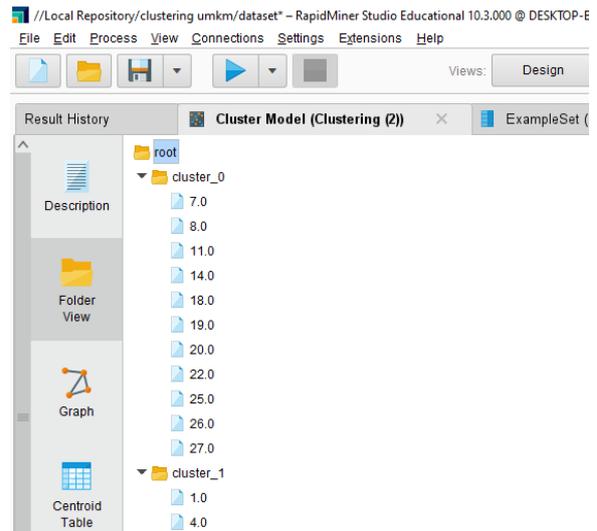


Gambar 10. Run

Setelah proyek di jalankan maka rapid miner akan menunjukan hasil seperti pada gambar



Gambar 11. Cluster Model



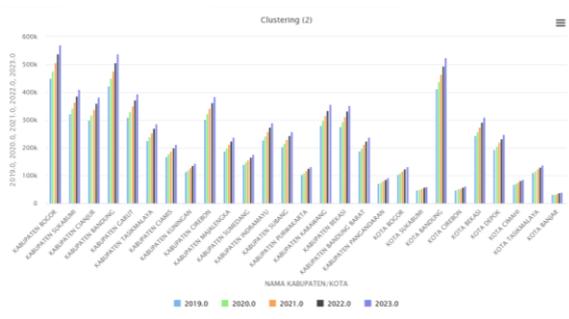
Gambar 12. Hasil Cluster

Penentuan jumlah *cluster* yang dilakukan menggunakan operator cluster distance performance untuk menghasilkan nilai perbandingan dari nilai Davis Bouldin. Hasil operasi tersebut menghasilkan nilai -0.571 untuk kelompok 2 klaster dan -0.437 untuk kelompok 3 klaster sehingga *Cluster* yang dibentuk yaitu 3 *cluster*, diantaranya *cluster* rendah (C0) yang menunjukkan sebaran UMKM rendah. *Cluster* sedang (C1) menunjukkan sebaran UMKM sedang, dan *cluster* tinggi (C2) menunjukkan sebaran UMKM tinggi. Langkah selanjutnya dalam proses *clustering k-means* yaitu menentukan centroid awal yang ditentukan berdasarkan nilai terkecil untuk C0, nilai rata rata untuk C1, dan nilai maksimum untuk C2. Atribut yang menjadi acuan dalam pengklasteran yaitu hanya atribut jumlah. Tabel berikut ini merupakan *centroid* data pada dataset.

Tabel 4. Centroid

Centroid data	
Min (C0)	175440
Average (C1)	1162956,259
Max (C2)	2540869

Nilai terkecil atau min (C0) adalah 175440, nilai rata-rata atau average (C1) adalah 1162956,259, dan nilai maksimum atau max (C2) adalah 2540869, semuanya ditunjukkan dalam tabel diatas. Clustering data dapat dilakukan setelah menentukan centroid awal. Ini dilakukan dengan mengambil jarak terdekat dari setiap data. Tabel berikut menunjukkan hasil clustering yang didapatkan dengan alat Rapid Miner.



Gambar 13. Visualisasi Cluster Pada Rapid Miner

NAMA KABUPATEN/KOTA	CLUSTER	KETERANGAN
KABUPATEN BOGOR	1	SEDANG
KABUPATEN SUKABUMI	2	TINGGI
KABUPATEN CIANJUR	2	TINGGI
KABUPATEN BANDUNG	1	SEDANG
KABUPATEN GARUT	2	TINGGI
KABUPATEN TASIKMALAYA	2	TINGGI
KABUPATEN CIAMIS	0	RENDAH
KABUPATEN KUNINGAN	0	RENDAH
KABUPATEN CIREBON	2	TINGGI
KABUPATEN MAJALENGA	2	TINGGI
KABUPATEN SUMEDANG	0	RENDAH
KABUPATEN INDRAMAYU	2	TINGGI
KABUPATEN SUBANG	2	TINGGI
KABUPATEN PURWAKARTA	0	RENDAH
KABUPATEN KARAWANG	2	TINGGI
KABUPATEN BEKASI	2	TINGGI
KABUPATEN BANDUNG BARAT	2	TINGGI
KABUPATEN PANGANDARAN	0	RENDAH
KOTA BOGOR	0	RENDAH
KOTA SUKABUMI	0	RENDAH
KOTA BANDUNG	1	SEDANG
KOTA CIREBON	0	RENDAH
KOTA BEKASI	2	TINGGI
KOTA DEPOK	2	TINGGI
KOTA CIMAH	0	RENDAH
KOTA TASIKMALAYA	0	RENDAH
KOTA BANJAR	0	RENDAH

Tabel 5. Hasil Cluster UMKM Pada Rapid Miner

Pengelompokan data yang didapatkan hasil bahwa yang masuk ke dalam *cluster* rendah (C0) dengan penyebaran UMKM yang rendah diantaranya Kabupaten Ciamis, Kabupaten Kuningan, Kabupaten Sumedang, Kabupaten Purwakarta, Kabupaten Pangandaran, Kota Bogor, Kota Sukabumi, Kota Cirebon, Kota Cimahi, Kota Tasikmalaya, Kota Banjar. Untuk *cluster* sedang (C1) dengan penyebaran UMKM yang sedang diantaranya Kabupaten Bogor, Kabupaten Bandung dan Kota Bandung. Untuk *cluster* tinggi (C2) dengan penyebaran UMKM yang tinggi diantaranya Kabupaten Sukabumi, Kabupaten Cianjur, Kabupaten Garut, Kabupaten Tasikmalaya, Kabupaten Cirebon, Kabupaten Majalengka, Kabupaten Indramayu, Kabupaten Subang, Kabupaten Karawang, Kabupaten Bekasi, Kabupaten Bandung Barat, Kota Bekasi Kota Depok.

4.1.5. Evaluasi

Untuk menguji hasil klasterisasi guna mengevaluasi sejauh mana kualitas klaster yang dihasilkan. Davies Bouldin Index (DBI). Tujuan utamanya adalah untuk menentukan keefektifan metode evaluasi tersebut dalam menilai hasil

klasterisasi dari data mining. Nilai dari perhitungan Davis Bouldin adalah sebagai berikut.

PerformanceVector

```
PerformanceVector:
Avg. within centroid distance: -27075639793.218
Avg. within centroid distance_cluster_0: -28697432757.560
Avg. within centroid distance_cluster_1: -26121643931.841
Davies Bouldin: -0.571
```

Gambar 14. hasil cluster distance performance 2 klaster

PerformanceVector

```
PerformanceVector:
Avg. within centroid distance: -11492392579.397
Avg. within centroid distance_cluster_0: -10249327291.355
Avg. within centroid distance_cluster_1: -1570738099.556
Avg. within centroid distance_cluster_2: -14833829626.166
Davies Bouldin: -0.437
```

Gambar 13. hasil cluster distance performance 3 klaster

Tabel 6. perbandingan nilai DBI

Percobaan	Jumlah Cluster	Nilai Davies Bouldin Index
1	2	-0,571
2	3	-0,437

4.2. Pembahasan

4.2.1. Penerapan K-Means clustering pada UMKM di Jawa Barat

Usaha mikro kecil menengah di Jawa Barat mengalami Peningkatan yang signifikan, Hasil dari pengelompokan dengan algoritma K-Means Clustering berdasarkan proyeksi jumlah unit UMKM yang terdapat di Jawa Barat menghasilkan 3 (tiga) cluster. Dimana Pengelompokan data yang didapatkan hasil bahwa yang masuk ke dalam *cluster* rendah diantaranya Kabupaten Ciamis, Kabupaten Kuningan, Kabupaten Sumedang, Kabupaten Purwakarta, Kabupaten Pangandaran, Kota Bogor, Kota Sukabumi, Kota Cirebon, Kota Cimahi, Kota Tasikmalaya, Kota Banjar. Untuk yang tergolong *cluster* dengan penyebaran UMKM yang sedang diantaranya Kabupaten Bogor, Kabupaten Bandung dan Kota Bandung. Sedangkan untuk *cluster* dengan penyebaran UMKM yang tinggi diantaranya Kabupaten Sukabumi, Kabupaten Cianjur, Kabupaten Garut, Kabupaten Tasikmalaya, Kabupaten Cirebon, Kabupaten Majalengka, Kabupaten Indramayu, Kabupaten Subang, Kabupaten Karawang, Kabupaten Bekasi, Kabupaten Bandung Barat, Kota Bekasi Kota Depok.

4.2.2. Hasil evaluasi Davis Bouldin Index

Pada hasil perhitungan menunjukkan hasil evaluasi dengan menggunakan DBI yang di mana di dapatkan nilai sebesar -0,571 dan -0,431. Dimana jika nilai semakin mendekati nol maka klister yang di bentuk dapat dikatakan baik. Sehingga hal tersebut

menyatakan bahwa evaluasi kluster k-means dengan DBI termasuk dalam kategori baik.

5. KESIMPULAN DAN SARAN

Temuan dari penelitian ini menunjukkan bahwa penerapan teknik pengklasteran K-Means mampu membagi UMKM di Jawa Barat menjadi tiga kluster berbeda. Dari hasil klustering tersebut, ditemukan bahwa 11 Kabupaten/Kota termasuk ke dalam kluster 0 dengan tingkat penyebaran UMKM yang rendah, sementara 3 Kabupaten/Kota masuk ke kluster 1 dengan tingkat penyebaran UMKM yang sedang, dan 12 Kabupaten/Kota termasuk ke dalam kluster 2 dengan tingkat penyebaran UMKM yang tinggi. Hal ini menyoroti pentingnya perhatian khusus dari pemerintah atau lembaga terkait terhadap strategi penyebaran dan peningkatan jumlah UMKM di wilayah Jawa Barat.

Dalam penelitian ini, evaluasi klustering dilakukan metode Bouldin Index (DBI). Hasil evaluasi menunjukkan bahwa nilai Davies Bouldin Index (DBI) sebesar -0,437 menunjukkan kategori evaluasi yang baik dibandingkan dengan -0,571. Pengimplementasian algoritma *k-means* ini sudah menunjukkan hasil yang cukup baik untuk pengelompokan data ke dalam *cluster*. Pada penelitian ini masih memiliki banyak kekurangan sehingga diperlukan rencana pengembangan di penelitian selanjutnya. Penelitian ini perlu dilakukan beberapa metode evaluasi lainnya guna mengetahui apakah masih menghasilkan hasil cukup baik atau tidak.

DAFTAR PUSTAKA

- [1] N. Syifa and R. N. Fahmi, "Implementasi Metode K-Means Clustering dalam Analisis Persebaran UMKM di Jawa Barat," *JOINS (Journal of Information System)*, vol. 6, no. 2, pp. 211–220, Dec. 2021, doi: 10.33633/joins.v6i2.5310.
- [2] J. Administrasi Publik, "Kecamatan Blimbing, Kota Malang) Feni Dwi Anggraeni, Imam Hardjanto, Ainul Hayat," UMKM.
- [3] P. Puntoriza and C. Fibriani, "Analisis Persebaran UMKM Kota Malang Menggunakan Cluster K-means," *JOINS (Journal of Information System)*, vol. 5, no. 1, pp. 86–94, May 2020, doi: 10.33633/joins.v5i1.3469.
- [4] L. Magdalena and R. Fahrudin, "Penerapan Data Mining Untuk Koperasi Se-Jawa Barat Menggunakan Metode Clustering pada Kementerian Koperasi dan UKM," 2019.
- [5] Z. Nabila, A. Rahman Isnain, and Z. Abidin, "ANALISIS DATA MINING UNTUK CLUSTERING KASUS COVID-19 DI PROVINSI LAMPUNG DENGAN ALGORITMA K-MEANS," *Jurnal Teknologi dan Sistem Informasi (JTSI)*, vol. 2, no. 2, p. 100, 2021, [Online]. Available: <http://jim.teknokrat.ac.id/index.php/JTSI>
- [6] J. Administrasi Publik, "Kecamatan Blimbing, Kota Malang) Feni Dwi Anggraeni, Imam Hardjanto, Ainul Hayat," UMKM.
- [7] A. Saputra and R. A. Sahputra, "Penentuan Kluster UMKM Sektor Perdagangan dan Perikanan Melalui Pendekatan Metode Clustering Data Mining di Kabupaten Aceh Barat," 2023.
- [8] W. Sudrajat, I. Cholid, and J. Petrus, "Wahyu Sudrajat et al, Penerapan Algoritma K-Means Untuk."
- [9] L. Magdalena and R. Fahrudin, "Penerapan Data Mining Untuk Koperasi Se-Jawa Barat Menggunakan Metode Clustering pada Kementerian Koperasi dan UKM," 2019.
- [10] L. Darwan Ali, A. Maulana, and K. Nur Akbar, "Penerapan Clustering Menggunakan Algoritma K-Means Sebagai Analisis Produksi Komoditas Perikanan Provinsi di Indonesia," *EJECTS: E-Journal Computer, Technology and Informations System*, vol. 01, no. 01, pp. 1–6, 2021.
- [11] M. Anjelita, A. P. Windarto, A. Wanto, and I. Sudahri, "Pengembangan Datamining Klustering Pada Kasus Pencemaran Lingkungan Hidup," *Seminar Nasional Teknologi Komputer & Sains (SAINTEKS)*, pp. 309–313, 2020.
- [12] Y. R. Sari, A. Sudewa, D. A. Lestari, and T. I. Jaya, "Penerapan Algoritma K-Means Untuk Clustering Data Kemiskinan Provinsi Banten Menggunakan Rapidminer," *CESS (Journal of Computer Engineering, System and Science)*, vol. 5, no. 2, p. 192, 2020, doi: 10.24114/cess.v5i2.18519.
- [13] R. R. Putra and C. Wadisman, "IMPLEMENTASI DATA MINING PEMILIHAN PELANGGAN POTENSIAL MENGGUNAKAN ALGORITMA K-MEANS IMPLEMENTATION OF DATA MINING FOR POTENTIAL CUSTOMER SELECTION USING K-MEANS ALGORITHM," *Journal of Information Technology and Computer Science (INTECOMS)*, vol. 1, no. 1, 2018.
- [14] P. R. Wulandhari, N. Rahaningsih, I. Ali, and C. L. Rohmat, "IMPLEMENTASI DATA MINING DALAM PERENCANAAN PERSEDIAAN OBAT," vol. 7, no. 1, 2023