

PEMETAAN DAERAH BERDASARKAN JENIS USAHA UMKM DENGAN ALGORITMA K-MEANS DI JAWA BARAT

*(Regional Mapping Based on The Type of UMKM Business With K-means Algorithm
in West Java)*

Hendrik^[1], Teady Matius Surya Mulyana^[2]

^{[1][2]}Program Studi Informatika, Fakultas Teknologi dan Desain, Universitas Bunda Mulia
Jl. Lodan Raya No.2, Ancol, Jakarta Utara, INDONESIA

Email: ^[1]HendrikLay0916@gmail.com, ^[2]tmulyana@bundamulia.ac.id

Abstract

Usaha mikro, kecil, dan menengah (UMKM) adalah usaha ekonomi produktif yang dimiliki perorangan, dan jumlah UMKM banyak tersebar di Provinsi Jawa Barat tergolong banyak, dan terdapat banyak jenis usahanya, seperti usaha aksesoris, border, batik, mebel, dan lainnya. Oleh karena itu akan menyebabkan susah pemerintah untuk pengelompokan data-data pelaku usaha mikro, kecil, dan menengah. Ruang lingkup penelitian, yaitu usaha mikro, kecil dan menengah (UMKM) yang terdapat pada Provinsi Jawa Barat. Dan serta tujuan dari penelitian ini yaitu untuk mengetahui nilai k yang optimal dari dataset yang digunakan, serta mengetahui tingkat akurasi dari setiap kluster. Metode dan algoritma yang akan digunakan dalam penelitian kali ini, yaitu algoritma K-means dan metode Elbow dan untuk penentuan cluster dengan parameter jarak antara titik data dan pusat cluster, kenapa menggunakan, dikarenakan algoritma K-means memungkinkan peneliti untuk membagi data menjadi cluster yang telah ditentukan sebelumnya, memfasilitasi identifikasi pola dan kelompok data, algoritma relatif efisien dan dapat diaplikasikan pada dataset yang besar, dan implementasinya yang sederhana membuatnya cocok untuk penelitian yang tidak melibatkan analisis data yang mendalam, hasil dari algoritma k-means berupa pusat cluster, yang mudah diinterpretasikan dan membantu menjelaskan hasil penelitian kepada pihak lain. Dari hasil penerapannya, Metode K-Means dapat mengklusterkan dataset usaha mikro, kecil, dan menengah (UMKM) di Jawa Barat dengan mendapatkan hasil cluster mulai dari C1, C2, C3, C4, C5 dengan jumlah anggota dari 18, 40, 44, 57, 3 anggota. Hasil perhitungan dari metode elbow didapatkan nilai K optimal dari dataset yaitu 5 dilihat dari grafik yang menunjukkan penurunan yang stabil pada cluster 5.

Keywords: Usaha mikro, kecil dan menengah, UMKM, Elbow Method, K-Means, Machine Learning

*Penulis Korespondensi

1. PENDAHULUAN

Usaha Mikro, Kecil, dan Menengah (UMKM) merupakan salah satu bidang usaha yang dapat berkembang dan konsisten dalam perekonomian nasional [1]. UMKM memiliki peran yang sangat penting bagi perekonomian di Indonesia, salah satunya adalah menciptakan serta memperluas kesempatan kerja sehingga pengangguran akibat dari angkata kerja yang tidak terserap dapat berkurang. Di Indonesia unit usaha bertambah setiap tahunnya, dikutip dari website Kementerian Koperasi dan Usaha Kecil Menengah Republik Indonesia jumlah UMKM pada tahun 2015 ada sekitar 59,2 juta unit dan pada tahun 2018 jumlah UMKM ada sekitar 64,2 juta unit [2].

Banyak sekali masalah yang dihadapi para pelaku UMKM di masa pandemi ini, masalah-masalah tersebut

seperti kesulitan mencari bahan baku, penurunan penjualan, modal yang semakin menipis, penurunan produksi dan terhambatnya distribusi. Jika pandemi covid-19 ini berlanjut, dikhawatirkan akan ada separuh atau lebih UMKM di Indonesia yang gulung tikar. Oleh karena itu untuk membangkitkan kembali ekonomi di Indonesia, pemerintah melakukan pemulihan ekonomi. Salah satu program dalam pemulihan ekonomi tersebut adalah program Banpres Produktif Usaha Mikro Kecil Menengah (BPUM) yang diselenggarakan oleh Kementerian Koperasi dan UKM Republik Indonesia dan masih ada banyak program pemerintah untuk membanyak usaha UMKM [2].

Penelitian yang dilakukan ini untuk dapat mendukung dan membantu pemerintah dalam menjalankan program bantuan untuk para pelaku usaha mikro, kecil dan menengah (UMKM). UMKM yang

terdapat di Provinsi Jawa Barat tergolong banyak dan terdapat banyak jenis usaha mikro, kecil, dan menengah di Jawa Barat, oleh karena itu diharap penelitian yang dilakukan dapat membantu pemerintah dalam membagikan bantuan untuk para pelaku usaha mikro, kecil, dan menengah secara merata sesuai dengan program bantuan yang diberikan pemerintah.

Pada penelitian ini akan dibuat sebuah sistem pemetaan usaha mikro, kecil, dan menengah pada setiap kabupaten dan kota yang ada di Provinsi Jawa Barat yang diharapkan dapat membantu pemerintah. Penelitian ini melakukan pemetaan UMKM di Jawa Barat untuk mendapatkan hasil berupa kabupaten dan kota di Jawa Barat bergolong UMKM yang banyak dan tidak banyak. Sistem pemetaan yang dibuat dengan menggunakan algoritma K-Means untuk proses pengelompokan berdasarkan kabupaten dan kota masing-masing yang terdapat pada Provinsi Jawa Barat. Pada penggunaan algoritma K-Means, data yang dapat diolah pada perhitungan merupakan data angka atau numerik. Masing-masing data dihitung dari kedekatan dengan nilai centroid yang sebelumnya telah ditentukan, jarak terkecil antara data dengan tiap-tiap centroid ialah anggota cluster yang terdekat. Dan serta menggunakan Metode Elbow untuk menentukan K optimal yang akan digunakan dalam proses perhitungan dengan algoritma K-means.

2. TINJAUAN PUSTAKA

2.1. Tinjauan Pustaka

Sistem pemetaan dengan menggunakan algoritma K-Means sudah pernah dilakukan selama beberapa tahun terakhir akan tetapi dalam penelitian kali ini menggunakan algoritma K-Means, serta menggunakan metode *Elbow* dan topik yang berbeda dengan penelitian sebelumnya. Penelitian sebelumnya yang akan dijadikan rujukan untuk sistem pemetaan ini adalah sebagai berikut:

Penelitian yang dilakukan Mira Suci Yana pada tahun 2018 dengan judul penelitian “Penerapan Metode K-Means dalam Pengelompokan Wilayah Menurut Intensitas Kejadian Bencana Alam di Indonesia Tahun 2013-2018”. Hasil dari penelitian ini yaitu, kluster pertama merupakan wilayah yang rawan terjadinya bencana alam dengan anggotanya adalah Provinsi Jawa Timur, Jawa Barat, dan Jawa Tengah. Sedangkan 31 provinsi lainnya tergolong dalam cluster kedua, yang dalam artian bukan daerah rawan terjadinya bencana alam. Kesimpulan ini diambil berdasarkan nilai rata-rata dari masing-masing cluster, dimana rata-rata setiap variabel pada cluster pertama lebih besar dibandingkan nilai rata-rata setiap variabel pada cluster kedua [3].

Penelitian yang dilakukan Much. Zuyyinal Haqqul Barir pada tahun 2019 dengan judul penelitian “sistem

informasi geografis pemetaan lahan pertanian bawang merah dengan metode k-means clustering berbasis website”. Hasil dari penelitian ini yaitu, sebuah sistem informasi geografis yang dapat memetakan lahan pertanian bawang merah. Pengujian SIG dilakukan pada lahansi 20 kecamatan di kabupaten Nganjuk dengan pertimbangan 3 variabel dan percobaan hitungsebanyak 4 kali. Dari hasil pengujian, diperoleh angka rata-rata kesesuaian penghitungan sebesar 93.51% antara hasil penghitungan manual dan hasil penghitungan menggunakan SIG [4].

Penelitian yang dilakukan Dicky Jordan Aji Putra pada tahun 2021 dengan judul penelitian “Metode K-Means Untuk Pemetaan Persebaran Usaha Mikro Kecil Dan Menengah”. Hasil dari penelitian ini yaitu, bahwa dari 200 dataset yang dimasukkan ke dalam program, diperoleh 176 data UMKM termasuk mikro, 22 data UMKM termasuk kecil, dan 2 data UMKM termasuk menengah sesuai dengan peraturan UU No 20 Tahun 2008 tentang UMKM. Dari 200 data UMKM diambil 40 data untuk diuji dengan metode Shilhouette Coefficient dengan 90% data UMKM yang terstruktur dengan baik. Hal ini menunjukkan bahwa dengan menggunakan metode K-Means dapat mengelompokkan tingkatan usaha maupun pemerataan bantuan UMKM mikro karena hasil perhitungan akan dipetakan ke dalam map Kecamatan Karang malang sesuai cluster masing-masing [5].

Berdasarkan dari tinjauan Pustaka diatas, maka penulis akan merancang dan membuat sebuah sistem pemetaan daerah berdasarkan umkm di Jawa Barat dengan menggunakan algoritma K-means. Kenapa algoritma K-means dipilih sebagai algoritma dalam penelitian ini yaitu, dikarenakan dataset yang dimiliki dalam penelitian tidak memiliki sebuah label tertentu, oleh karena itu membutuhkan sebuah algoritma yang bisa mengelompokkan atau pengklasteran sebuah dataset yang tidak memiliki sebuah label tertentu, dan algoritma K-means memungkinkan peneliti untuk membagi data menjadi cluster yang telah ditentukan sebelumnya, memfasilitasi identifikasi pola dan kelompok data, algoritma relatif efisien dan dapat diaplikasikan pada dataset yang besar, dan implementasinya yang sederhana membuatnya cocok untuk penelitian yang tidak melibatkan analisis data yang mendalam, hasil dari algoritma k-means berupa pusat cluster, yang mudah diinterpretasikan dan membantu menjelaskan hasil penelitian kepada pihak lain, dari penjelasan ini maka Algoritma yang paling cocok untuk menyelesaikan permasalahan ini, yaitu Algoritma K-means, dikarenakan Algoritma K-Means adalah algoritma yang paling cocok untuk pengelompokan atau pengklasteran terhadap dataset yang tidak memiliki sebuah label tertentu, karena Algoritma K-means tidak membutuhkan sebuah label pada dataset yang akan digunakan. Tujuan dari

penelitian ini yaitu untuk menerapkan algoritma K-means ke dalam sistem yang akan dibuat yaitu sistem pemetaan daerah berdasarkan jenis usaha umkm di jawa barat untuk mengetahui daerah-daerah di jawa barat tergolong jumlah umkm yang sangat rendah, rendah, sedang, tinggi, atau sangat tinggi berdasarkan hasil centroid.

2.2. Dasar Teori

Teori-teori yang akan digunakan dalam penelitian ini adalah sebagai berikut:

2.2.1. Algoritma K-Means

Metode k-means clustering merupakan salah satu metode data clustering non hierarchical yang berusaha mempartisi data-data yang ada ke dalam bentuk satu atau lebih cluster atau kelompok sehingga data yang memiliki karakteristik yang sama dikelompokkan ke dalam satu cluster yang sama dan data yang mempunyai karakteristik berbeda, dikelompokkan ke dalam kelompok lainnya. Kemiripan anggota terhadap cluster diukur dengan kedekatan objek terhadap nilai mean pada cluster atau dapat disebut sebagai centroid cluster atau pusat massa [6]. Dalam Metode K-means hal pertama yang dilakukan adalah Inisialisasi K pusat cluster ini bisa dilakukan dengan berbagai cara. Namun yang paling sering dilakukan adalah dengan cara random [7]. K dalam Metode *K-means* sebenarnya merupakan sebuah nilai dari sebuah pusat cluster (Centroid). Nilai pusat/ Centroid yang telah dipilih merupakan factor dari hasil aktif dari Metode K-means [5]. Dasar dari metode klasifikasi ini merupakan pengelompokan data berdasarkan jarak dari setiap nilai data yang ada. Dalam metode ini apabila semakin dekat jarak antar data maka dipastikan semakin tinggi tingkat kemiripannya, dan sebaliknya apabila semakin besar jarak antar data dipastikan semakin besar tingkat ketidak miripannya data [8]. Setelah menentukan jumlah cluster k yang dibuat, selanjutnya menentukan nilai dari pusat/centroid awal yang ditentukan secara acak atau penentuan sendiri, dan setelah mendapatkan nilai titik pusat cluster Langkah selanjutnya adalah menghitung jarak setiap data ke masing-masing centroid dengan menggunakan rumus korelasi antar dua objek, dapat menggunakan rumus *Euclidean Distance* yang diperagakan pada rumus (1) [6].

$$ED = \sqrt{(X_i - W_i)^2 + (Y_i - Z_i)^2} \quad (1)$$

Keterangan:

ED: *Euclidean Distance*

i: Jumlah Objek

X, Y: Koordinat pada Objek

W, Z: Koordinat pada Centroid

Setelah mendapatkan hasil dari perhitungan jarak setiap data, Langkah selanjutnya adalah menggabungkan setiap data yang dilihat berdasarkan jarak terdekat antara data dan cluster. Langkah selanjutnya menentukan titik pusat centroid baru dengan menghitung nilai rata-rata berdasarkan data pada cluster yang serupa, untuk menentukan pusat centroid baru untuk menghitung jarak antara data dan centroid dapat menggunakan rumus (2)[6].

$$C_{kl} = \frac{X_{1l} + X_{2l} + \dots + X_{pl}}{P} \quad (2)$$

Keterangan:

Ckl: Nilai *Centroid* ke-k pada *Variabel* ke-l.

P : Banyaknya data/Jumlah data.

Sesudah semua perhitungan dengan rumus yang ada, dan pusat cluster tidak mengalami adanya perubahan maka proses cluster selesai/berhasil, dan apabila belum maka akan kembali ke proses yang bisa di tunjukkan dengan Rumus (1).

2.2.2. Metode Elbow

Metode Elbow merupakan suatu metode yang digunakan untuk menghasilkan informasi dalam menentukan jumlah cluster terbaik atau jumlah cluster yang optimal dengan cara melihat persentase hasil perbandingan antara jumlah *cluster* (k) yang akan membentuk siku pada suatu titik[9]. Hasil persentase yang berbeda dari setiap nilai cluster dapat ditunjukkan dengan menggunakan grafik sebagai sumber informasinya. Jika nilai cluster pertama dengan nilai cluster kedua memberikansudut dalam grafik atau nilainya mengalami penurunan paling besar maka nilai cluster tersebut yang terbaik [10]. Untuk nnendapatkan perbandingannya dengan menggunakan rumus SSE (Sum of Square Error) untuk menghitung masing-masing nilai cluster. Karena semakin besar jumlah cluster K maka nilai SSE akan semakin kecil. Berikut adalah rumus SSE yang diperagakan pada rumus (3)

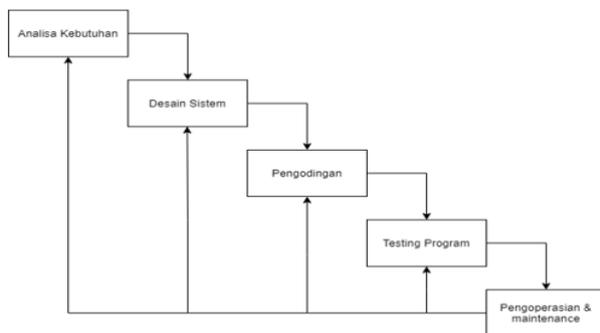
$$SSE = \sum_{k=1}^k \sum_{x_1 e S_k} \| X_i - C_k \|^2 \quad (3)$$

3. METODE PENELITIAN

3.1. Metode Pengembangan Sistem

Model waterfall adalah model yang paling banyak digunakan untuk tahap pengembangan. Model waterfall ini juga dikenal dengan nama model tradisional atau model klasik. Model air terjun

(*waterfall*) sering juga disebut model sekuensial linier (*sequential linear*) atau alur hidup klasik (*Classic cycle*)”[11]. Dalam perancangan sistem informasi ini dengan menggunakan model *Waterfal*. *Waterfall* merupakan metode yang handal dalam perencanaan yang dilakukan dengan iterasi yang tidak berulang. *Waterfall* juga melaksanakan perancangan dengan metode yang sistematis sehingga perencanaan dilakukan dengan detail dan lengkap. Sehingga metode ini akan menunjang dalam melakukan analisis yang tepat untuk pembuatan sistem informasi [12]. Tahapan dalam model *Waterfall* terdapat 5 Tahapan yaitu, sebagai berikut:



Gambar 1. Model Waterfall

a. Analisis Kebutuhan

Tahapan metode *waterfall* yang pertama adalah mempersiapkan dan menganalisa kebutuhan dari aplikasi yang akan dikerjakan. Informasi dan insight yang diperoleh dapat berupa dari hasil wawancara, survei, studi literatur, *observasi*, hingga diskusi.

b. Desain Sistem

Tahap yang selanjutnya adalah pembuatan desain aplikasi sebelum masuk pada proses coding. Tujuan dari tahap ini, supaya mempunyai gambaran jelas mengenai tampilan dan antarmuka aplikasi yang kemudian akan dieksekusi oleh tim programmer.

c. Pengodingan

Tahapan metode *waterfall* yang berikutnya adalah implementasi kode program dengan menggunakan berbagai tools dan bahasa pemrograman sesuai dengan kebutuhan dalam pengembangan aplikasi. Jadi, pada tahap implementasi ini lebih berfokus pada hal teknis, dimana hasil dari desain perangkat lunak akan diterjemahkan ke dalam bahasa pemrograman melalui tim programmer atau developer.

d. Testing Program

Pada tahap yang keempat, masuk dalam proses integrasi dan pengujian sistem. Pada tahap ini, akan dilakukan penggabungan modul yang sudah dibuat pada

tahap sebelumnya. Setelah proses integrasi sistem telah selesai, berikutnya masuk pada pengujian modul.

e. Pengoperasian & Maintenance

Tahapan metode *waterfall* yang terakhir adalah pengoperasian dan perbaikan dari aplikasi. Setelah dilakukan pengujian sistem, maka akan masuk pada tahap produk dan pemakaian aplikasi oleh pengguna (*user*). Untuk proses pemeliharaan, memungkinkan pengembang untuk melakukan perbaikan terhadap kesalahan yang ditemukan pada aplikasi setelah digunakan oleh user.

3.2. Analisis Kebutuhan Sistem

Analisis kebutuhan sistem akan menjelaskan kebutuhan fungsional dan non fungsional yang dibutuhkan dalam pengembangan sistem, sebagai berikut:

3.2.1. Kebutuhan Fungsional

kebutuhan fungsional merupakan sebuah proses atau layanan yang disediakan oleh system yang dibuat. Mencakup bagaimana sistem harus bereaksi pada input tertentu dan bagaimana perilaku sistem pada situasi tertentu.

- User dapat mengisi username dan password yang sudah disediakan.
- User dapat menginput dataset format excel (xlsx) pada menu input dataset.
- User dapat mengakses menu map, K-means Elbow(perhitungan).
- User dapat melihat hasil dari perhitungan dengan algoritma K-means dan metode Elbow.

3.2.2. Kebutuhan Non-Fungsional

Pada kebutuhan non-fungsional pada dasarnya digunakan untuk mengetahui spesifikasi yang dibutuhkan untuk membuat sistem. Spesifikasi tersebut biasanya berupa hardware atau software.

a. Perangkat Keras (Hardware)

Hardware yang dipakai dalam merancang dan mengembangkan system ini, sebagai berikut:

- Processor: AMD Ryzen 5 5600H
- Memory Ram 18 GB
- Grafik Card (GPU): Nvidia GTX 1650
- Storage: 500 GB

b. Perangkat Lunak (Software)

Software yang dipakai dalam merancang dan mengembangkan system ini, sebagai berikut:

- Sistem Operasi: minimal Windows 10
- Text Editor: Virtual Studio Code

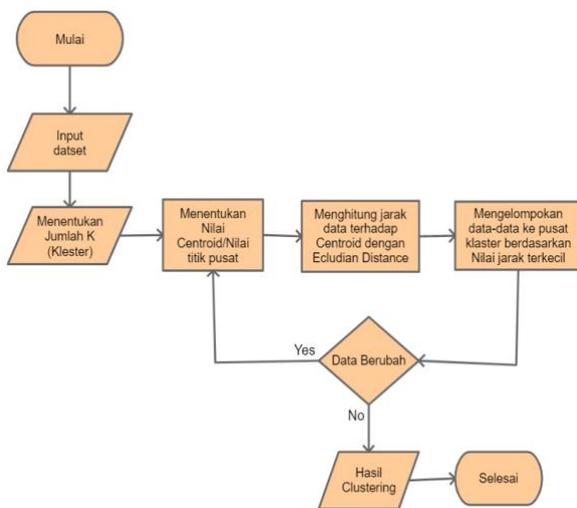
3. Chrome Version 100.0.4896.127 (Official Build) (64-bit)
4. XAMPP v3.3.0

3.3. Perancangan Sistem

Pada perancangan sistem yang dilakukan yaitu merancang apa saja yang dibutuhkan dalam pembuatan sistem untuk penelitian yang akan dilakukan.

3.3.1. Perancangan Proses

Pada sistem yang dibuat pada penelitian ini, ada satu algoritma yang dipakai dan satu metode yang dipakai, yaitu algoritma K-means dan metode Elbow. Berikut merupakan flowchart dari algoritma K-means, metode Elbow, dan flowchart dari sistem aplikasi:

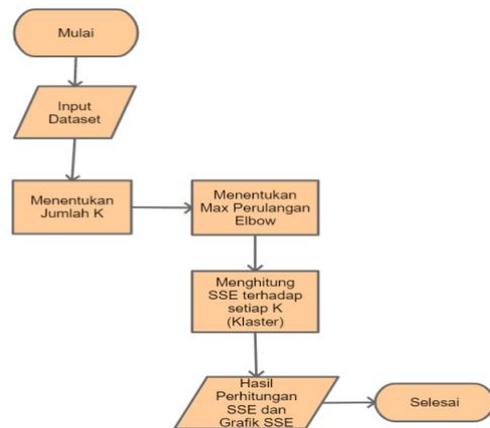


Gambar 2. Flowchart Diagram Algoritma K-Means

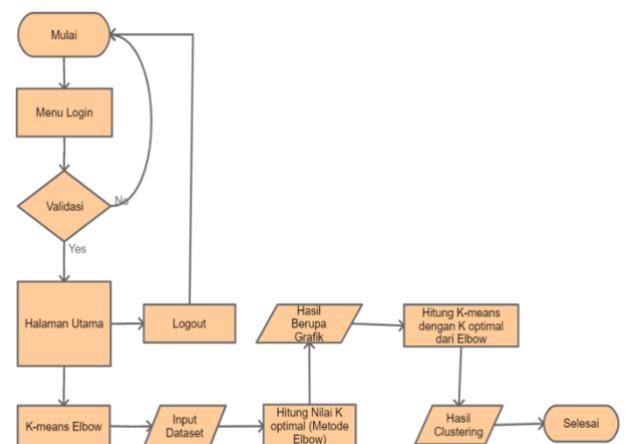
Gambar 2 menjelaskan menjelaskan alur proses dari metode K-Means. Mula-mula data diinput, kemudian menentukan jumlah cluster yang diinginkan dan menentukan titik pusat awal secara acak. Setelah itu menghitung jarak data terhadap titik pusat cluster menggunakan rumus *Ecludian Distance*. Setelah mendapatkan hasil dari perhitungan, dilakukan pengelompokan tiap data berdasarkan jarak terkecil terhadap tiap titik pusat cluster. Setelah semua data sudah di kelompokkan di lakukan pengecekan apabila data berpindah dari kelompok awal maka akan di lakukan perulangan (iterasi), yang dimulai dari tahap penentuan titik pusat cluster. Apabila data tidak berpindah kelompok berarti proses pengklasteran sudah selesai.

Gambar 3 menjelaskan alur proses yang terdapat pada metode *elbow* yang diterapkan pada aplikasi. Pertama-tama user harus menginput dataset, sesudah dataset di input langsung akan dilakukan perhitungan,

akan tetapi sebelum itu user harus menentukan jumlah nilai K yang akan dihitung. Sesudah user menentukan jumlah nilai k yang akan dihitung, maka nilai pada setiap k akan dihitung dengan menggunakan rumus SSE, untuk mendapatkan hasil darisetiap selisih nilai k. dan sudah mendapatkan nilai selisih dari setiap k, akan langsung dibuat sebuah grafik dari hasil perhitungan SSE dan user bisa menentukan nilai K optimal dengan melihat grafik dimana yang membentuk siku itulah nilai k yang optimal.



Gambar 3. Flowchart Diagram Metode Elbow



Gambar 4. Flowchart Diagram Sistem

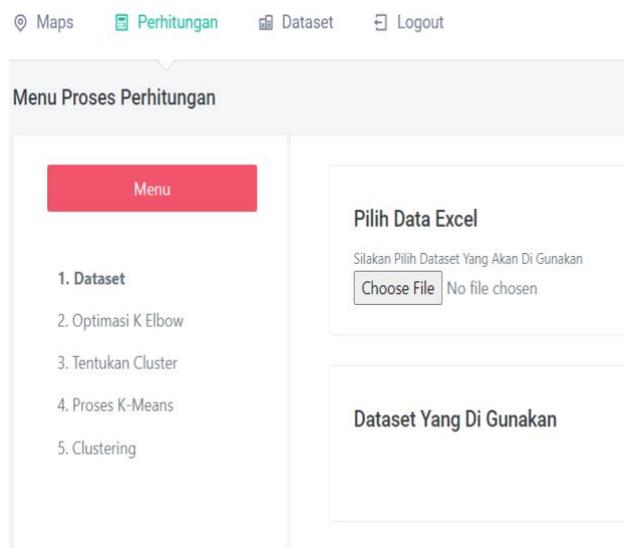
Bisa dilihat pada gambar 4 *Flowchart Diagram* pada aplikasi akan menjelaskan alur proses yang berjalann di aplikasi yang dibuat. Pertama user terlebih dahulu harus mempunyai username dan *password* yang sudah disediakan untuk melakukan login pada menu login, sesudah user memasukkan username dan *password system* akan melakukan validasi data apa sudah benar atau tidak, kalua tidak benar akan langsung Kembali dan disuruh isi username dan password yang benar dan sebaliknya apabila username dan password sudah benar akan langsung masuk ke halaman utama. Setelah masuk di dalam halaman utama user bisa

langsung melakukan input dataset yang akan dihitung. Sesudah dataset di input bisa langsung melakukan perhitungan untuk menentukan nilai K yang optimal, dan hasil dari metode *elbow* ini akan berbentuk data dan grafik yang menunjukkan hasil nilai k yang optimal. Sesudah nilai k yang optimal sudah di tentukan langsung akan melakukan perhitungan dengan metode *k-means* dengan memasukan nilai K yang optimal, dan akan memunculkan hasil pengelompokan data-datanya.

4. HASIL DAN PEMBAHASAN

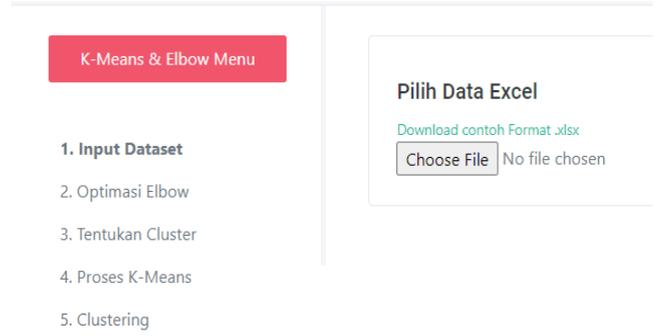
4.1. Implementasi Aplikasi

Hasil dari penelitian ini berupa sebuah aplikasi system pemetaan daerah berdasarkan jenis usaha umkm di jawa barat. Adapun tampilan utama pada aplikasi yang dibuat dalam penelitian ini, dapat dilihat pada gambar 5.



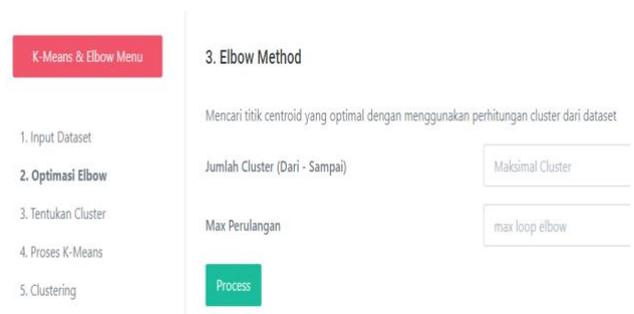
Gambar 5. Tampilan Halaman Utama Aplikasi

Pada gambar 5 menunjukkan tampilan halaman utama pada aplikasi yang terdapat beberapa menu yang bisa digunakan oleh user. Menu yang tersedia menu map dan hasil, Proses *K-means & Elbow*, dan *Logout*. Terdapat 5 menu didalam menu proses *k-means & elbow* untuk melakukan proses perhitungan dengan metode *k-means* dan *elbow*.



Gambar 6. Tampilan Halaman Input dataset

Pada gambar 6 menunjukkan tampilan pada menu input dataset, dimana user akan diminta untuk menginput dataset yang sudah ada pada aplikasi melalui menu ini.



Gambar 7. Tampilan Halaman Metode Elbow

Pada gambar 7 menunjukkan tampilan menu optimal *Elbow*, dimana user akan mengisi jumlah *cluster* yang mau dihitung berapa dan max perulangan untuk mendapatkan K optimalnya. Hasil dari metode *elbow* akan berbentuk graph.



Gambar 8. Tampilan Halaman Tentukan Cluster

Pada gambar 8 menunjukkan tampilan menu tentukan cluster, dimana user diminta input jumlah cluster dan mac perulangan untuk diteruskan pada proses perhitungan dengan algoritma *K-means*. Dibagian jumlah cluster ini kita dapat dari hasil metode *elbow* untuk mencari K yang optimal.

Perulangan 1 - Penentuan Centroid

Centroid	0.332589680987	0.332589680987	0.2762015495067	0.3660000982995	0.2511166289749	0.3601602924095	0.1
Centroid 1							
Centroid 2	0.3313408071749	0.3313408071749	0.2371076231839	0.3314608629707	0.34602646470207	0.33144071100434	0
Centroid 3	0.35644618834081	0.35644618834081	0.25374531598507	0.3566358581045	0.37067552173804	0.356619041581018	0.1
Centroid 4	0.3674607623184	0.3674607623184	0.2637278026906	0.36753504956688	0.38133843484203	0.36751643691054	0.1
Centroid 5	0.38886573463466	0.38886573463466	0.28567660247956	0.38904135274466	0.41215638887531	0.38902205473759	0.1

Gambar 9. Tampilan Halaman Proses K-means

Pada gambar 9 ini akan menampilkan proses perhitungan centroid, perhitungan Euclidean Distance, dan hasil pengelompokan dataset. Diaman hasil proses k-means menunjukkan daerah di Jawa Barat termasuk dalam cluster apa saja dari 1 sampai 5 cluster.

Hasil Cluster KMeans

nama_lokasi	absorpsi	batik	banda	craft	selusin	komoditi	saluran	makanan	rekomendasi	jenis_bahasa	Cluster
Kabupaten Bogor - 2016	727	727	232	3312	3134	1739	147584	142732	18347	44130	1
Kabupaten Bogor - 2017	927	927	132	33111	32316	19734	142580	146232	19337	40330	1
Kota Bandung - 2021	1081	1081	154	28625	37679	23089	166150	121530	25346	52503	1
Kabupaten Bandung - 2021	1110	1110	159	28644	38792	23034	179689	124830	20158	53829	1
Kabupaten Sukabumi - 2021	845	845	121	30184	29470	17796	129956	90511	17633	41064	1
Kabupaten Bogor - 2021	1179	1179	168	42087	41087	20596	181187	132523	34385	57322	1
Kota Bandung - 2020	1018	1018	145	30356	35483	21968	156476	114469	21232	49444	1
Kabupaten Bandung - 2020	1046	1046	149	37343	36447	22257	160725	117556	21888	50787	1
Kabupaten Bogor - 2020	1110	1110	159	28644	38959	23028	179630	124830	20152	53916	1
Kota Bandung - 2019	959	959	137	34238	32416	20486	147339	107780	19995	46562	1

Gambar 10. Tampilan Halaman Hasil Clustering

Pada gambar 10 menampilkan hasil dari proses perhitungan dengan algoritma k-means dan mendapatkan hasil, diaman daerah di Jawa Barat tergolong cluster apa dari 1 sampai 5 kategori cluster.

4.2. Pengujian Algoritma dan Metode

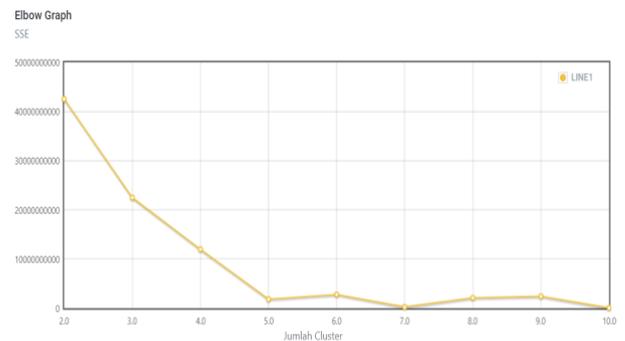
Pengujian pada algoritma dan metode yang dilakukan di penelitian ini, dilakukan dengan menjalankan aplikasi dan mencoba melakukan perhitungan dengan system aplikasi apakah bisa menampilkan hasil perhitungan dengan algoritma dan metode yang diterapkan.

4.2.1. Pengujian Metode Elbow

Dalam pengujian keberhasilan metode elbow untuk menentukan k optimal terhadap total dataset sebanyak 162, dan disertai dengan jumlah dari K=1 sampai K=10 dan dalam pengujian data akan dibagi menjadi 3 bagian untuk melakukan pengujian yaitu 54 data, 108 data, dan 162 untuk mendapatkan k yang optimal dari dataset.

TABEL I. Pengujian Dengan 54 Dataset

Cluster	SSE (Sum of Square Error)	Selisih
2	13.225483084201	13.225483084201
3	7.5454874183365	5.6799956658641
4	4.4319806117249	3.1135068066116
5	3.7302610566238	0.70171955510111
6	2.7675896971299	0.96267135949388
7	2.8376183047284	0.070028607598547
8	2.4970015492363	0.34061675549214
9	2.5238276435859	0.026826094349646
10	2.1918121086677	0.33201553491827
2	13.225483084201	13.225483084201

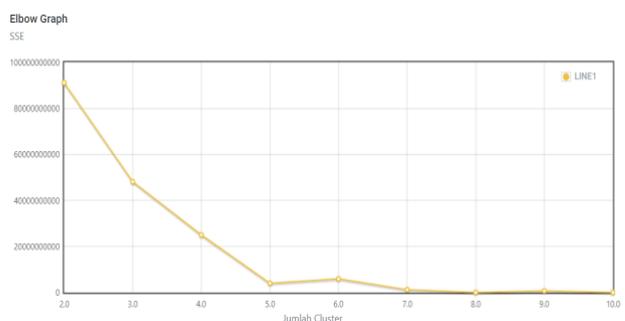


Gambar 11. Graph dari 54 Dataset

Pada gambar 11 dapat dilihat graph menunjukkan terjadi penurunan yang setabil dari cluster 5, jadi k optimal adalah 5.

TABEL II. Pengujian Dengan 108 Dataset

Cluster	SSE (Sum of Square Error)	Selisih
2	22.344170010081	22.344170010081
3	11.278235334367	11.065934675714
4	6.32719244518	4.9510428891872
5	6.1053460871758	0.22184635800418
6	3.8029412502101	2.3024048369658
7	3.351738604789	0.45120264542103
8	2.9107733816722	0.44096522311683
9	2.7050338942243	0.2057394874479
10	2.5597728763725	0.14526101785178
2	22.344170010081	22.344170010081

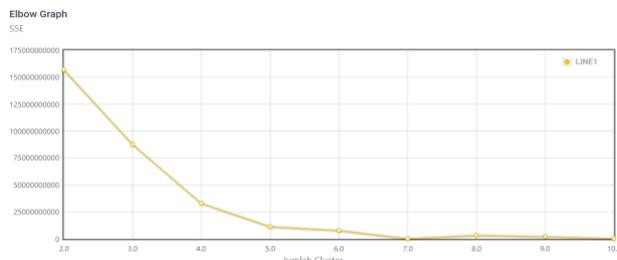


Gambar 11. Graph dari 108 Dataset

Pada gambar 11 menunjukkan graph yang diaman terjadi penurunan yang stabil pada cluster 5, tidak beda dengan pengujian dengan 54 dataset. Jadi k optimal pada pengujian 108 dataset adalah 5.

TABEL III. Pengujian Dengan 162 Dataset

Cluster	SSE (Sum of Square Error)	Selisih
2	22.344170010081	22.344170010081
3	11.278235334367	11.065934675714
4	6.32719244518	4.9510428891872
5	6.1053460871758	0.22184635800418
6	3.8029412502101	2.3024048369658
7	3.351738604789	0.45120264542103
8	2.9107733816722	0.44096522311683
9	2.7050338942243	0.2057394874479
10	2.5597728763725	0.14526101785178
2	22.344170010081	22.344170010081



Gambar 12. Graph dari 162 Dataset

Pada gambar 12 juga menunjukkan penurunan yang stabil pada cluster 5, tidak beda jauh dengan hasil dari pengujian dengan 54 dataset, 108 dataset.

Dari 3 kali pengujian dengan menggunakan 54 dataset, 108 dataset, dan 162 dataset didapat hasil graph menunjukkan terjadi penurunan yang stabil pada cluster 5 dalam 3 kali pengujian, jadi sudah bisa dipastikan kalau k optimal adalah cluster 5

4.2.2. Pengujian Algoritma K-means

Pada pengujian algoritma k-means menggunakan 162 dataset yang akan di proses perhitungannya dengan sistem aplikasi untuk mendapatkan range umkm berdasar atributnya.

TABEL IV. Range UMKM

No	Atribut	C1	C2	C3	C4	C5
1	Aksesoris	224-477	64-414	120-588	522-814	727-1179
2	Batik	224-477	64-414	120-588	522-814	727-1179
3	Bordir	75-91	9-45	46-76	46-116	121-232
4	Craft	21624-23301	2286-12435	11259-19057	19880-32364	30194-42097

5	Fashion	14908-22766	2231-11039	1364-12684	19403-28389	29470-41087
6	Konveksi	11292-12291	1363-7452	6214-12684	11197-17336	14588-25090
7	Kuliner	82764-94982	9640-52089	41154-82020	85562-125192	129956-181187
8	Makanan	62923-71203	7197-36830	31295-59991	62581-94586	95051-132523
9	Minuman	12024-15044	1335-7104	5891-11469	11610-16987	172114-24585
10	Jasa_Lainnya	24048-31329	3109-14815	14831-25938	27036-42295	41064-57252

Pada table IV menunjukkan range umkm berdasar atribut yang ada, dan dari proses perhitungan dengan program didapatkan jumlah anggota C1 sebanyak 18 anggota, C2 sebanyak 40 anggota, C3 sebanyak 44 anggota, C4 sebanyak 57 anggota, dan C5 sebanyak 3 anggota.

4.2.3. Pengujian dengan Confusion Matrix

Tabel Confusion Matrix dapat dilihat pada Tabel V. Hasil dari perhitungan akurasi dengan confusion matrix dengan data uji benar sebanyak 108 data dan data salah sebanyak 54 data, dan dilakukan perhitungan dengan rumus akurasi dari confusion matrix didapatkan hasil akurasi dari pengklasteran sebesar 66,7%. Alasan hasil pengujian akurasi cluster rendah dikarenakan data yang digunakan dalam penelitian ini kurang banyak.

TABEL V. Confusion Matrix

		Aktual Class	
		True	False
Predictive Class	True	108	54
	False	54	108

5. KESIMPULAN DAN SARAN

5.1. Kesimpulan

Berdasarkan hasil implementasi dan pengujian yang telah dilakukan, maka dapat disimpulkan hasil penelitian ini sebagai berikut.

- Metode Elbow bisa diterapkan untuk mencari nilai k yang optimal yang akan dipakai untuk proses algoritma K-means.
- Berdasarkan pengujian sebanyak 3 kali dengan metode Elbow didapatkan k optimal adalah cluster 5.
- Dari hasil proses algoritma K-means didapatkan C1 memiliki 18 anggota, C2 memiliki 40 anggota, C3 memiliki 44 anggota, C4 memiliki 57 anggota, dan C5 memiliki 3 anggota.
- Akurasi pengklasteran dengan algoritma k-means yaitu, 66,7% dengan data yang digunakan sebanyak 54 data.

5.2. Saran

Berdasarkan hasil penelitian yang didapatkan, saran yang dapat diberikan penulis adalah sebagai berikut.

- Sebaiknya untuk penelitian selanjutnya, bisa menambahkan fitur search, fitur daftar akun untuk mendapatkan username dan password.
- Untuk penelitian selanjutnya, dapat menggunakan dataset yang lebih banyak dari penelitian sekarang.
- Sebaiknya penelitian selanjutnya bisa menggabungkan dengan metode lainnya.
- Untuk penelitian selanjutnya bisa menggunakan data yang lebih banyak/besar agar algoritma k-means lebih efektif.

DAFTAR PUSTAKA

[1] A. Primatami And N. Hidayati, "Perkembangan Usaha Mikro Kecil (Umk) Di Provinsi Jawa Barat Tahun 2006 – 2016," *Jurnal Pengembangan Wiraswasta*, Vol. 21, No. 3, P. 203, Dec. 2019, Doi: 10.33370/Jpw.V21i3.350.

- [2] E. Syarifah, S. Purnamasari, A. Purnomo, P. Studi Ekonomi Syariah, F. Studi Islam, And U. Islam Kalimantan Muhammad Arsyad Al Banjari, "Efektivitas Penyaluran Dana Banpres Produktif Usaha Mikro (Bpum) Untuk Modal Kerja Dalam Meningkatkan Kesejahteraan Pelaku Umkm (Studi Kasus Pada Pelaku Umkm Di Desa Melayu)." [Online]. Available: <https://www.kemenkopukm.go.id/data-umkm>
- [3] M. S. Yana *Et Al.*, "Penerapan Metode K-Means Dalam Pengelompokan Wilayah Menurut Intensitas Kejadian Bencana Alam Di Indonesia Tahun 2013-2018," 2018.
- [4] H. Barir And S. Informasi, "Sistem Informasi Geografis Pemetaan Lahan Pertanian Bawang Merah Dengan Metode K-Means Clustering Berbasis Website (Studi Kasus Di Kabupaten Nganjuk) Achmad Imam Agung Chamdan Mashuri," 2019.
- [5] D. Remawati, D. J. Aji Putra, And T. Irawati, "Metode K-Means Untuk Pemetaan Persebaran Usaha Mikro Kecil Dan Menengah," *Jurnal Teknologi Informasi Dan Komunikasi (Tikomsin)*, Vol. 9, No. 2, P. 39, Nov. 2021, Doi: 10.30646/Tikomsin.V9i2.574.
- [6] Zulrahmadi, S. Defit, And Y. Yunus, "Pemetaan Wilayah Potensial Terhadap Penjualan Sepeda Motor Honda Menggunakan K-Means Clustering," *Jurnal Informatika Ekonomi Bisnis*, Pp. 53–59, Jun. 2020, Doi: 10.37034/InfEb.V2i2.41.
- [7] N. Azizah And N. A. Widiastuti, "Teknologi Geolocation Berbasis Android Dengan Metode K-Means untuk Pemetaan Umkm Di Kabupaten Jepara," *Sistem Informasi Bisnis*, Vol. 02, No. 10, 2018, Doi: 10.21456/Vol8iss2pp214-225.
- [8] "Klasifikasi E-Book Berbahasa Inggris Dengan Menggunakan Metode K-Means Clustering Studi Kasus Perpustakaan Stiki Malang."
- [9] A. Winarta And W. J. Kurniawan, "Optimasi Cluster K-Means Menggunakan Metode Elbow Pada Data Pengguna Narkoba Dengan Pemrograman Python," *Jurnal Teknik Informatika Kaputama (Jtik)*, Vol. 5, No. 1, 2021.
- [10] E. Muningsih And S. Kiswati, "Sistem Aplikasi Berbasis Optimasi Metode Elbow Untuk Penentuan Clustering Pelanggan," 2018.
- [11] M. Susilo And R. Kurniati, "Rancang Bangun Website Toko Online Menggunakan Metode Waterfall," 2018.
- [12] N. Agitha, Y. Lestari, And E. Anjarwani, "Rancang Bangun Sistem Informasi Tracer Study Bagi Alumni Fakultas Teknik Universitas Mataram Berbasis Web." [Online]. Available: <http://jtika.if.unram.ac.id/index.php/jtika/>