

Analisis Persebaran Izin Usaha di Kabupaten Sragen Menggunakan Metode K-Means Clustering

Nasrun Nugroho^{*1}, Wijiyanto², Afu Ichsan Pradana³

^{1,2,3}Teknik Informatika, Fakultas Ilmu Komputer, Universitas Duta Bangsa, Surakarta, Indonesia
Email: ¹230103302@mhs.udb.ac.id, ²wijiyanto@udb.ac.id, ³afu_ichsan@udb.ac.id

Abstrak

Peningkatan perekonomian adalah tantangan yang dihadapi oleh Pemerintah daerah. Dalam membangun perekonomian Kabupaten Sragen memiliki dinas yang berperan dalam keberhasilan Investasi yang menunjang pertumbuhan ekonomi yaitu Dinas Penanaman Modal dan Pelayanan Terpadu Satu Pintu. Dinas ini melaksanakan tugas yang mendukung visi misi Bupati dalam hal peningkatan perekonomian. Dalam melaksanakan tugasnya dinas ini mempunyai program yang terdiri dari kegiatan untuk mencapai target yang sudah ditetapkan. Sasaran program tersebut adalah terlaksananya promosi dan pelayanan perizinan berusaha. Melihat data dari tahun sebelumnya perizinan berusaha di Sragen berkembang sangat pesat, namun persebaran izin usaha tersebut tidak merata, hanya sebagian kecil wilayah saja yang menunjukkan tingginya izin usaha. Hal ini menuntut adanya analisis data yang efektif yang bisa memetakan pola persebaran tersebut. Penelitian ini menerapkan analisis data menggunakan metode K-Means Clustering, data yang dianalisis adalah data izin usaha tahun sebelumnya yang meliputi izin usaha mikro, kecil dan menengah. Analisis ini mengelompokkan wilayah kecamatan yang ada di Kabupaten Sragen menjadi 3 kluster yaitu: (1) Kluster dengan persebaran izin usaha tinggi, (2) kluster persebaran izin usaha sedang dan (3) Kluster persebaran izin usaha rendah. Hasil penelitian diharapkan bisa menjadi acuan bagi dinas terkait untuk merumuskan kebijakan pemerataan izin usaha. Penelitian ini memberikan kontribusi pada pengambilan keputusan berbasis data untuk pemerataan izin usaha di Kabupaten Sragen.

Kata kunci: *Data Mining, Kabupaten Sragen, K-Means Clustering, Persebaran Izin Usaha.*

Analysis of Business License Distribution in Sragen Regency Using the K-Means Clustering Method

Abstract

Economic growth is a challenge faced by the local government. In building the economy, Sragen Regency has an agency that plays a role in the success of investments that support economic growth, namely the Investment and One-Stop Integrated Services Agency. This agency carries out tasks that support the Regent's vision and mission in terms of economic growth. In carrying out its duties, this agency has a program consisting of activities to achieve predetermined targets. The target of the program is the implementation of business licensing promotions and services. Looking at data from the previous year, business licensing in Sragen has grown very rapidly, but the distribution of business licenses is uneven, only a small part of the area shows a high number of business licenses. This requires effective data analysis that can map the distribution pattern. This study applies data analysis using the K-Means Clustering method, the data analyzed is the previous year's business license data which includes micro, small and medium business licenses. This analysis groups the sub-district areas in Sragen Regency into 3 clusters, namely: (1) Cluster with high business permit distribution, (2) cluster with medium business permit distribution and (3) cluster with low business permit distribution. The results of the study are expected to be a reference for related agencies to formulate policies for equalizing business permits. This study contributes to data-based decision making for equalizing business permits in Sragen Regency.

Keywords: *Data Mining, Distribution of Business Licenses, K-Means Clustering, Sragen Regency*

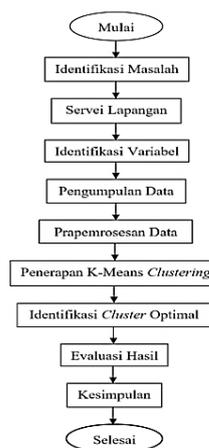
1. PENDAHULUAN

Nilai investasi merupakan salah satu indikator atau tolak ukur perkembangan ekonomi di suatu daerah tak terkecuali di wilayah Kabupaten Sragen. Dalam tiga tahun terakhir Kabupaten Sragen mengalami peningkatan ekonomi yang sangat baik. Salah satu faktor utama yang menunjukkan pertumbuhan tersebut adalah banyaknya

jumlah izin usaha yang dikeluarkan oleh pemerintah daerah. Namun dari data yang dikeluarkan oleh Pemerintah Daerah didapat persebaran izin usaha ini tidak sepenuhnya merata di seluruh wilayah Kabupaten. Permasalahan-permasalahan itu meliputi beberapa aspek yang menghambat pelaku usaha, dan salah satu masalah utamanya adalah masih kurangnya pemahaman dari pelaku usaha mikro, kecil dan menengah terkait perizinan, bagi pelaku usaha proses ini begitu membingungkan dan membutuhkan waktu yang cukup lama dari yang diharapkan oleh pelaku usaha. Dari permasalahan tersebut Pemerintah daerah Kabupaten Sragen lewat Dinas Penanaman Modal Pelayanan terpadu Satu Pintu mencanangkan Program Pelayanan Perizinan dan Program Promosi. Namun dari program dan kegiatan tersebut belum mampu menangani permasalahan dengan baik dikarenakan kegiatan seringkali dilakukan berdasarkan kerangka kerja yang sudah ada tanpa mempertimbangkan data rinci tahun-tahun sebelumnya. Mengatasi masalah ini penulis mencoba mengolah data perizinan tahun sebelumnya untuk dijadikan acuan sasaran kegiatan selanjutnya. Penulis mencoba mengelompokan izin usaha di setiap wilayah Kecamatan dengan menggunakan teknik data mining metode *K-Means Clustering*. Kelebihan *K-Means* terletak pada kemampuannya mengolah data dalam jumlah besar dan melakukan klastering dengan cepat[1], selain itu *K-Means Clustering* dengan menggunakan perangkat lunak R Studio dibangun untuk memetakan masalah[2] sehingga cocok digunakan dalam analisis data izin usaha yang memiliki volume data yang besar[3]. Dengan pengelompokan data tersebut diharapkan akan didapat hasil pengelompokan atau klaster yang bisa dijadikan dasar dalam menentukan wilayah yang sudah tinggi persebaran usahanya sampai yang masih rendah izin usahanya. Penelitian serupa menggunakan *K-Means Clustering* telah dilakukan oleh A. Azzam, A. Irma Purnamasari, and I. Ali[4][5], tetapi fokus pada aspek berbeda seperti analisis UMKM di Jawa Barat. Berbeda dari penelitian sebelumnya, studi ini menerapkan *K-Means* untuk mendukung perencanaan ekonomi berbasis wilayah. Penelitian ini menerapkan konsep data mining dengan menggunakan algoritma *K-Means Clustering* untuk menentukan klaster atau wilayah mana saja yang membutuhkan perhatian lebih dalam hal perizinan usaha, hasil dari pengolahan data tahun-tahun sebelumnya. Dari analisis hasil *clustering* tersebut bisa memetakan data pendukung keputusan persebaran izin usaha yang lebih jelas dan menjadikan dasar kegiatan yang terstruktur[5]. Dan bagi pemerintah daerah diharapkan bisa merumuskan kebijakan yang mendukung pertumbuhan ekonomi yang lebih merata di seluruh Kabupaten Sragen khususnya dalam hal izin usaha mikro, kecil dan menengah.

2. METODE PENELITIAN

Jenis penelitian yang digunakan dalam penelitian ini tergolong penelitian kuantitatif, karena data yang digunakan pada penelitian ini berupa jumlah izin usaha berdasarkan jenis skala usaha mulai dari mikro, kecil dan menengah yang dikeluarkan oleh Dinas Penanaman Modal Pelayanan Terpadu Satu Pintu Kabupaten Sragen. Data yang diperoleh berdasarkan penelitian ini bersumber dari data primer dimana peneliti mengambil langsung pada pengelola data dan informasi di bidang Pelayanan Terpadu Satu Pintu. Pada penelitian ini objek yang diteliti adalah jenis izin usaha yang diterbitkan, sedangkan variabel yang digunakan adalah total izin usaha yang dikeluarkan oleh Dinas Penanaman Modal dan Pelayanan terpadu Satu Pintu. Adapun tahapan penelitian digambarkan pada Gambar 1 berikut :



Gambar 1. Alur tahapan penelitian

Berdasarkan Gambar 1 alur tahapan penelitian dapat dijelaskan secara rinci sebagai berikut :

1. Tahap identifikasi masalah, merupakan langkah awal dalam proses penelitian ini. Berfokus untuk mengevaluasi objek penelitian, menetapkan cakupan penelitian, dan mengidentifikasi permasalahan yang

perlu dipecahkan[6]. Dalam konteks perizinan berusaha Dinas Penanaman Modal dan Pelayanan terpadu Satu Pintu Kabupaten Sragen, tahap ini akan mencakup identifikasi dan pemetaan permasalahan terkait dengan proses persebaran izin usaha yang kurang optimal. Dengan menargetkan perbaikan pada proses peningkatan ekonomi melalui promosi dan bimbingan kepada masyarakat dan pelaku usaha, penelitian ini bertujuan untuk mengatasi persebaran usaha yang lebih merata di seluruh wilayah Kabupaten Sragen.

2. Tahapan survei lapangan dalam penelitian ini bertujuan untuk memperoleh informasi terkait bidang penelitian serta menjadi langkah dalam pengumpulan data yang diperlukan.
3. Identifikasi variabel, pada tahapan identifikasi variabel berdasarkan tahapan identifikasi masalah dan survei lapangan maka didapatkan solusi dari permasalahan yang diketahui dengan atribut yang telah ditentukan yaitu kelompok wilayah, jumlah izin usaha mulai dari kategori usaha mikro, usaha kecil dan usaha menengah[7].
4. Tahap pengumpulan data dalam penelitian ini dilakukan dengan meminta informasi dari pengelola data dan informasi di bidang Pelayanan Terpadu Satu Pintu untuk dijadikan sebagai bahan analisis. Data yang telah dikumpulkan selanjutnya dilakukan proses pengolahan data guna memperoleh informasi terkait lokasi izin usaha, kategori izin usaha, jumlah izin usaha di wilayah Kabupaten Sragen.
5. Pada tahap prapemrosesan data, dilakukan serangkaian proses termasuk pembersihan, penambahan, pengurangan, dan penyusunan data guna memastikan data yang digunakan dalam proses analisis mining sesuai dengan kebutuhan[8].
6. Penerapan K-Means Clustering, pada tahap ini melakukan proses normalisasi data terlebih dahulu. Normalisasi data diperlukan untuk mempersempit cakupan suatu data. Pada penelitian ini menggunakan normalisasi Min Max yaitu konversi nilai dari sebuah data dengan rentan nilai minimum 0 dan nilai maksimum 1[9]. Untuk persamaannya sebagai berikut:

$$x^1 = \frac{x - \min(x)}{\max(x) - \min(x)} \quad (1)$$

Keterangan :

- X¹ : nilai skala baru
- X : nilai asli
- Min(x) : nilai minimum pada kolom
- Max(x) : nilai maksimum pada kolom

Pengukuran jarak antar objek dengan menggunakan jarak Euclidean, yang dapat dihitung menggunakan persamaan sebagai berikut:

$$d_{ij} = \sqrt{\sum_{p=1}^l (x_{ip} - x_{jp})^2} \quad (2)$$

Keterangan :

- d_{ij} : jarak antara objek ke-i dengan objek ke-j
- l : jumlah variabe
- x_{ip} : nilai objek ke-i pada variabel ke-p
- x_{jp} : nilai objek ke-j pada variabel ke-p

Ukuran kesamaan antar objek tersebut digunakan untuk melihat kemiripan antar objek.[10] Selanjutnya menerapkan K-Means Clustering untuk mendapatkan pembentukan cluster, antara lain sebagai berikut:

- a) Menetapkan nilai k (jumlah klaster) sesuai kebutuhan.
- b) Menetapkan posisi awal pusat klaster (centroid) secara acak sebanyak k.
- c) Menghitung jarak antara setiap data input dengan masing-masing centroid.
- d) Mengelompokkan setiap data ke dalam klaster berdasarkan jarak terdekat dengan centroid.
- e) Memperbaharui posisi centroid baru dengan menghitung rata-rata dari anggota dalam masing-masing klaster.

Melakukan iterasi dari langkah 3 hingga 5, terus-menerus, hingga tidak ada perubahan pada anggota setiap cluster.

7. Identifikasi klaster optimal, pada tahapan ini akan ditentukan banyaknya kelompok K-Means yang terbentuk dengan menggunakan metode Elbow, metode Shilhouette, Gap Statistik dan menggunakan Dunn Index untuk mengetahui kelompok klaster yang optimal. Dunn Index menghitung rasio antar jarak klaster minimum sebagai separation (pemisah) dan ukuran klaster maksimum sebagai compactness (kepadatan). Jumlah klaster terbaik dapat dilihat dari hasil Dunn Index yang semakin besar[11]. Adapun rumus perhitungan Dunn Index adalah sebagai berikut:

$$Dunn\ Index = \frac{\min d(c_i, c_j)}{\max(diam(c_k))} \quad (3)$$

Dimana $\min d(C_i, C_j)$ mengukur jarak minimum antar kluster dan $\max(diam(C_k))$ merupakan diameter kluster maksimum.

Tahap terakhir dalam proses data mining adalah evaluasi hasil yang diperoleh. Evaluasi dilakukan terhadap hasil pengelompokan dengan membandingkan dari hasil kluster optimal.

3. HASIL DAN PEMBAHASAN

3.1. Proses Normalisasi

Pada tahap normalisasi, data yang digunakan dalam penelitian ini telah melalui proses transformasi menggunakan metode Min-Max Scaling untuk mempersempit rentang nilai. Tujuan utama dari normalisasi ini adalah untuk mengkonversi data ke dalam skala yang seragam, sehingga setiap atribut atau variabel memiliki kontribusi yang setara dalam analisis lebih lanjut, terutama pada algoritma yang sensitif terhadap perbedaan skala antar fitur, seperti klusterisasi k-means[12]. Untuk hasil normalisasi bisa dilihat pada Gambar 2 berikut.

```
> print(head(norma))
      Usaha.Mikro Usaha.Kecil Usaha.Menengah
Gemolong      0.4438861  0.19148936  0.0000000
Gesir          0.0000000  0.06382979  0.0000000
Gondang       0.2671692  0.12765957  0.0000000
Jenar         0.2068677  0.06382979  0.0000000
Kalijambe     0.2453936  0.29787234  0.1428571
Karangmalang  0.7152429  0.78723404  0.5714286
> summary(norma)
      Usaha.Mikro      Usaha.Kecil      Usaha.Menengah
Min.   :0.0000   Min.   :0.0000   Min.   :0.0000
1st Qu.:0.1880   1st Qu.:0.1117   1st Qu.:0.0000
Median :0.3124   Median :0.1702   Median :0.0000
Mean   :0.3572   Mean   :0.2628   Mean   :0.1286
3rd Qu.:0.5205   3rd Qu.:0.2500   3rd Qu.:0.1429
Max.   :1.0000   Max.   :1.0000   Max.   :1.0000
```

Gambar 2. Hasil normalisasi Min Max.

Hasil normalisasi ini memberikan dasar yang lebih baik untuk tahap analisis selanjutnya, memastikan bahwa data yang digunakan untuk klusterisasi memiliki keseragaman skala yang dapat meningkatkan akurasi hasil klusterisasi yang diperoleh.[13].

3.2. Penerapan Algoritma K-Means

Dalam penerapan algoritma K-Means, penulis melakukan pengujian dengan menetapkan nilai k sebesar 3, yang berarti data akan dikelompokkan ke dalam tiga kluster yang berbeda. Setelah proses iterasi dan perhitungan dilakukan menggunakan Rstudio diperoleh kluster 1 sebanyak 3 data, kluster 2 sebanyak 7 data dan kluster 3 sebanyak 10 data.

```
K-means clustering with 3 clusters of sizes 3, 7, 10
Cluster means:
      Usaha.Mikro Usaha.Kecil Usaha.Menengah
1  0.7730318  0.8723404  0.6666667
2  0.4777459  0.1975684  0.04081633
3  0.1481575  0.1255319  0.02857143
Clustering vector:
      Gemolong      Gesir      Gondang      Jenar      Kalijambe Karangmalang      Kedawung
      2          3          3          3          3          1          2
      Masaran      Miri      Mondokan      Ngrampal      Plupuh      Sambirejo Sambungmacan
      2          3          3          2          3          3          2
      Sidoharjo      Sragen      Sukodono Sumberlawang      Tangen      Tanon
      2          1          3          1          3          2
```

Gambar 3. Hasil klusterisasi dengan 3 kluster

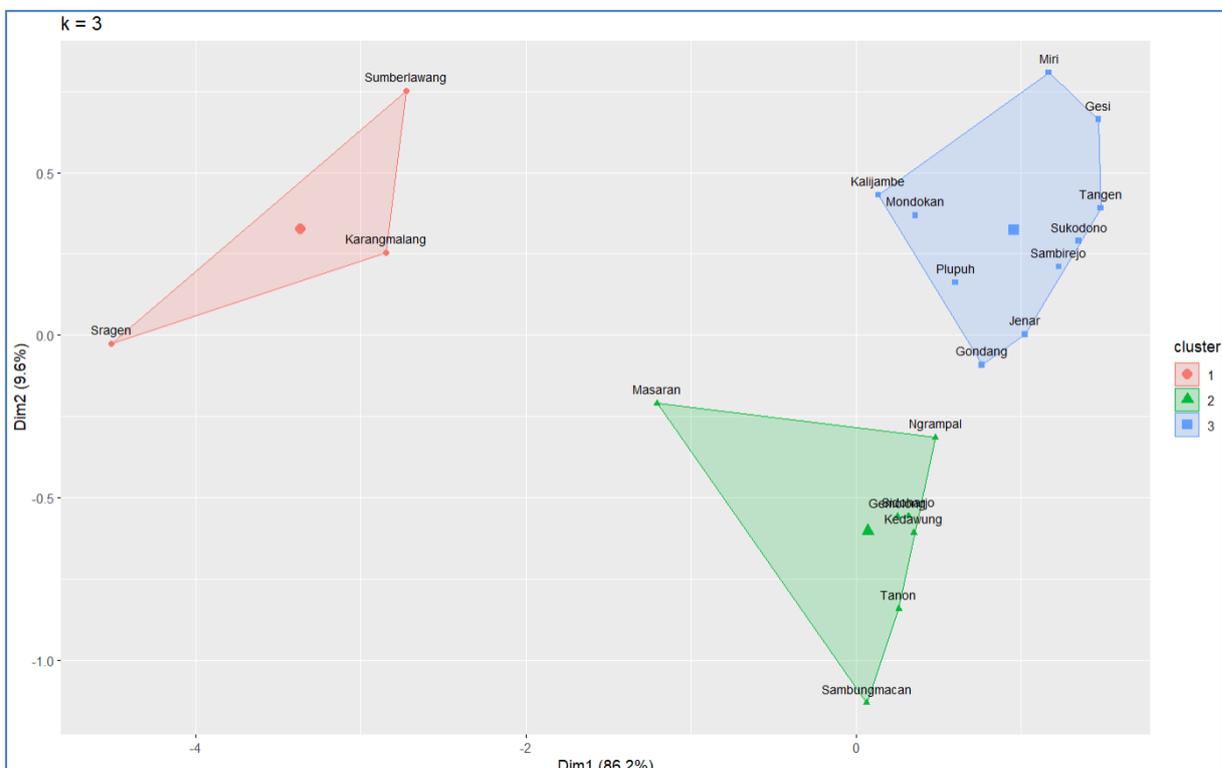
Pada klasifikasi ini diketahui "Within-Cluster Sum of Squares" (WCSS) dan "Between-Cluster Sum of Squares" (BCSS) sebagai berikut :

```

within cluster sum of squares by cluster:
[1] 0.2856973 0.2192968 0.1883728
(between_SS / total_SS = 83.1 %)
    
```

Gambar 4. WCSS dan BCSS dari 3 klaster

Nilai [0.2856973, 0.2192968, 0.1883728] menunjukkan WCSS untuk masing-masing dari tiga klaster, WCSS mengukur variasi titik data dalam setiap klaster, di mana nilai yang lebih rendah menunjukkan bahwa titik-titik dalam klaster tersebut lebih dekat dengan pusat klaster. Klaster 1 memiliki WCSS sebesar 0.2856973, yang menunjukkan bahwa titik-titik dalam klaster ini memiliki penyebaran yang moderat di sekitar pusat klaster. Klaster 2 memiliki WCSS sebesar 0.2192968, yang menunjukkan bahwa klaster ini sedikit lebih kompak dibandingkan Klaster 1. Klaster 3 memiliki WCSS sebesar 0.1883728, yang berarti titik-titik dalam Klaster 3 lebih kompak lagi dibandingkan dengan klaster 1 dan klaster 2. Sedangkan rasio *Between-Cluster Sum of Squares* (BCSS) terhadap *Total Sum of Squares* (TSS) menunjukkan bahwa 83.1% dari variasi total data dijelaskan oleh pemisahan antar-klaster. Rasio yang tinggi (mendekati 100%) menunjukkan bahwa klaster-klaster tersebut terpisah dengan baik, sementara rasio yang rendah menunjukkan bahwa klaster-klaster tersebut tidak terpisah dengan jelas. Dalam hal ini, rasio 83.1% menunjukkan pemisahan yang relatif baik antara klaster-klaster tersebut. Hasil plot untuk pembagian data ke dalam tiga klaster dapat dilihat pada gambar di bawah ini. Plot ini memberikan representasi visual mengenai bagaimana data dikelompokkan berdasarkan karakteristik yang serupa di dalam dataset. Setiap klaster direpresentasikan dengan warna yang berbeda untuk memudahkan identifikasi dan membedakan antara kelompok-kelompok data.



Gambar 5. Visualisasi Plot dari 3 klaster

Jika disimpulkan hasil klastering dari 3 klaster ini terdiri dari:

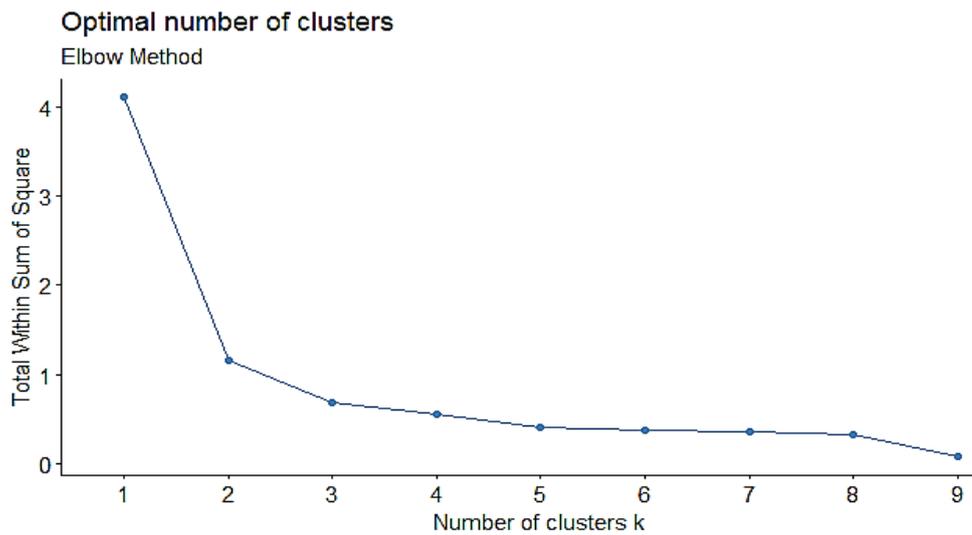
Klaster 1 : Sragen, Sumberlawang, Karangmalang

Klaster 2 : Gemolong, Masaran, Sidoharjo, Kedawung, Ngrampal, Sambungmacan, Tanon.

Klaster 3 : Gesi, Jenar, Gondang, Kalijambe, Miri, Sukodono, Tangen, Sambirejo, Plupuh, Mondokan.

3.3. Identifikasi Cluster Optimal

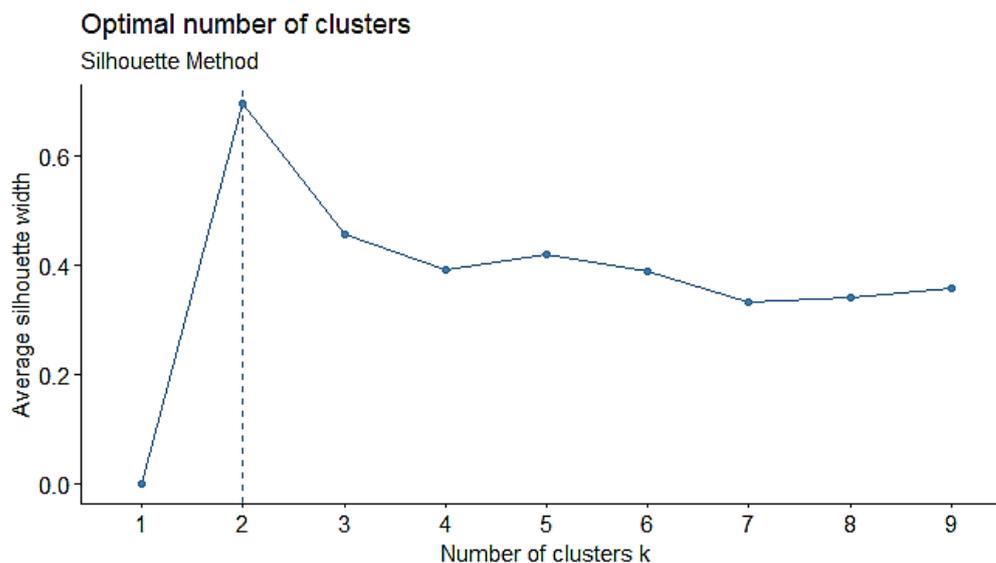
a) Metode Elbow



Gambar 6. Metode Elbow

Berdasarkan plot yang ditampilkan di atas, terlihat bahwa titik yang menyebabkan grafik membentuk siku atau "elbow" adalah titik kedua. Hal ini menunjukkan adanya perubahan signifikan dalam penurunan Within-Cluster Sum of Squares (WCSS) hingga titik tersebut, dan setelahnya, penurunan WCSS cenderung melambat. Metode elbow memiliki tujuan untuk menentukan nilai k yang paling kecil dan memiliki nilai within yang rendah[14]. Metode Elbow ini bekerja dengan prinsip mengidentifikasi titik di mana penambahan jumlah klaster tidak lagi memberikan pengurangan WCSS yang signifikan[15], sehingga titik ini dianggap sebagai jumlah klaster optimal. Dalam kasus ini, titik siku pada posisi ke dua mengindikasikan bahwa dua klaster merupakan pilihan optimal.

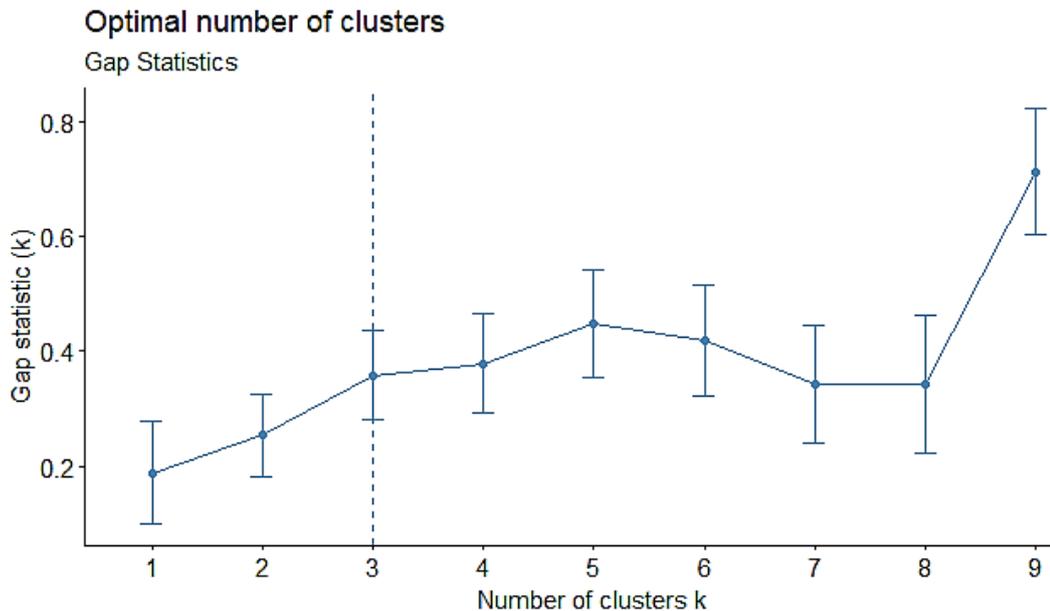
b) Metode Silhouette



Gambar 7. Metode Silhouette

Berdasarkan plot diatas, titik yang berada di puncak garis yaitu pada saat titik ke dua. Dalam kasus ini, titik puncak pada grafik Silhouette menunjukkan bahwa pembagian data ke dalam dua kluster menghasilkan kualitas klasterisasi yang terbaik dibandingkan jumlah kluster lainnya. Hal ini mengindikasikan bahwa dengan dua kluster, data dalam masing-masing kelompok lebih homogen (serupa), dan jarak antara kluster yang berbeda cukup besar, menciptakan pemisahan yang jelas[16].

c) **Gap Statistik**



Gambar 8. Metode Gap Statistik

Berdasarkan plot yang ditampilkan di atas, terlihat bahwa garis putus-putus yang menjadi penanda optimalitas klasterisasi terdapat pada titik ketiga. Hal ini menunjukkan bahwa nilai Gap Statistic mencapai nilai tertinggi atau signifikan pada saat jumlah kluster adalah tiga. Pada kasus ini, garis putus-putus yang berada di titik ketiga menandakan bahwa jumlah kluster tiga memberikan pemisahan terbaik antara kluster pada data asli dibandingkan dengan data acak. Ini berarti bahwa dengan tiga kluster, model klasterisasi dapat menjelaskan struktur data secara signifikan lebih baik dibandingkan jumlah kluster lainnya, dan metode inilah yang dipilih oleh penulis dalam menentukan jumlah kluster dalam proses penelitiannya.

d) **Dunn Index**

Dari hasil proses Rstudio didapatkan hasil nilai dunn index sebagai berikut:

Tabel 1. Dunn Index

K	Dunn Index
1	N/A
2	0,6212307
3	0,1396416
4	0,2200888

Dalam kasus ini, nilai Dunn Index k=1 tidak bisa dihitung karena hanya ada satu kluster sehingga jarak antar kluster tidak valid, untuk hasil perhitungan tertinggi dicapai ketika jumlah kluster adalah dua kluster. Hal ini mengindikasikan bahwa dengan dua kluster, struktur klasterisasi mampu memberikan keseimbangan terbaik antara pemisahan antar kluster (*inter-cluster separation*) dan kepadatan di dalam kluster (*intra-cluster compactness*)[17].

Dikarenakan nilai dari Dunn Index berbeda dengan yang lainnya maka penulis perlu membandingkan hasil dari k=2 dengan k=3. Untuk hasil dari k=2 didapat kluster 1 berjumlah 3 data dan kluster 2 sebanyak 17 data.

```

K-means clustering with 2 clusters of sizes 3, 17

Cluster means:
  Usaha.Mikro Usaha.Kecil Usaha.Menengah
1  0.7730318  0.8723404  0.66666667
2  0.2838703  0.1551940  0.03361345

Clustering vector:
  Gemolong      Gesi      Gondang      Jenar      Kalijambe Karangmalang      Kedawang
           2           2           2           2           2           1           2
  Masaran      Miri      Mondokan      Ngrampal      Plupuh      Sambirejo Sambungmacan
           2           2           2           2           2           2           2
  Sidoharjo      Sragen      Sukodono Sumberlawang      Tangen      Tanon
           2           1           2           1           2           2
    
```

Gambar 9. Hasil klasterisasi dengan 2 klaster

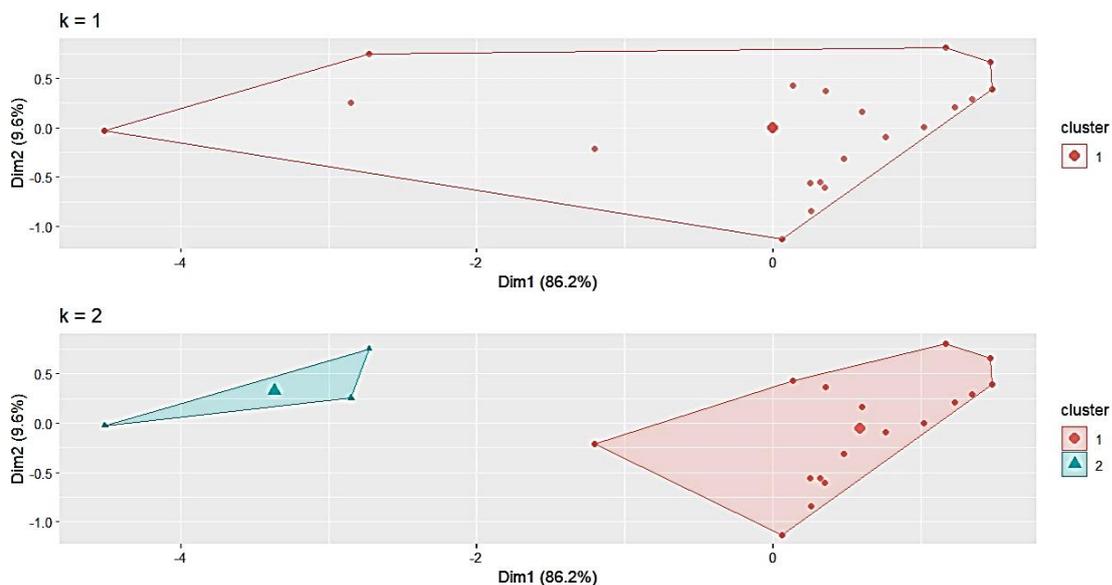
Pada klasifikasi ini didapatkan "Within-Cluster Sum of Squares" (WCSS) dan "Between-Cluster Sum of Squares" (BCSS) sebagai berikut :

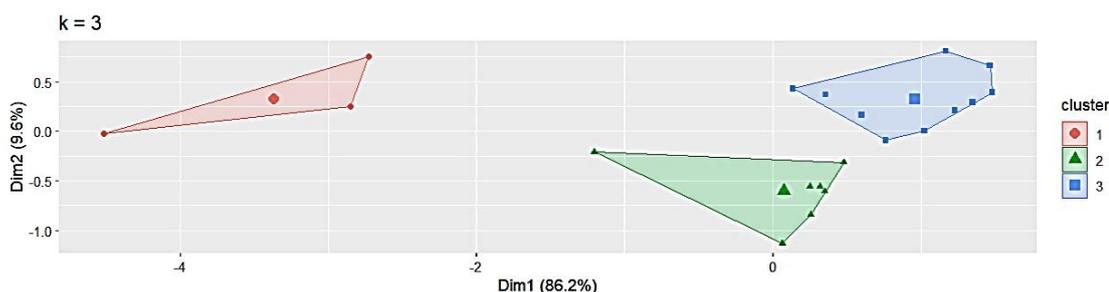
```

Within cluster sum of squares by cluster:
[1] 0.2856973 0.8769484
(between_SS / total_SS = 71.7 %)
    
```

Gambar 10. WCSS dan BCSS dari 2 klaster

Nilai WCSS yang diberikan adalah [0.2856973, 0.8769484], yang menggambarkan variasi titik data dalam masing-masing klaster. Klaster 1 memiliki WCSS sebesar 0.2856973, yang menunjukkan bahwa titik data dalam klaster ini tersebar cukup jauh dari centroid klaster (titik data tidak terpusat). Klaster 2 memiliki WCSS sebesar 0.8769484, yang berarti secara keseluruhan persebaran titik dalam klaster ini lebih jauh dibandingkan dengan klaster 1, menunjukkan kepadatan klaster yang lebih renggang (klaster ini tidak terlalu kompak). (between_SS / total_SS = 71.7%): Rasio ini menunjukkan bahwa sekitar 71.7% dari variasi total data dijelaskan oleh pemisahan antara klaster-klaster. Dengan rasio 71.7%, dapat disimpulkan bahwa klaster-klaster ini memiliki pemisahan yang moderat. Pemisahan antar-klaster tidak terlalu besar, dan sebagian besar variasi data masih ada dalam klaster itu sendiri (dapat dilihat dari nilai WCSS yang cukup besar pada Klaster 2). Untuk visualisasi plot dari ketiga k bisa digambarkan sebagai berikut:





Gambar 11. WCSS dan BCSS dari 3 kluster

Dari plot tersebut, dapat disimpulkan bahwa:

K = 1: terdapat satu kluster dengan satu titik pusat dari keseluruhan titik yang tersebar.

K = 2: terdapat dua kluster yang memiliki ciri khas berbeda. Kluster 1 memiliki bentuk yang lebih memanjang ke arah kanan, sementara kluster 2 lebih tersebar.

K = 3: terdapat tiga kluster yang menunjukkan perbedaan yang lebih jelas. Kluster 1 memiliki bentuk yang memanjang ke arah kiri dan sedikit ke atas, cluster 2 memiliki bentuk yang lebih rapat dan cenderung berada di tengah, sementara kluster 3 memiliki bentuk yang lebih terkonsentrasi di sebelah kanan meskipun ada satu titik yang terlihat sedikit lebih menjauh.

Secara keseluruhan, plot tersebut menunjukkan bahwa nilai k yang berbeda dapat menghasilkan kluster yang berbeda pula. Pada kasus ini, K = 3 menghasilkan kluster yang lebih terdefinisi dan mudah dibedakan dibandingkan dengan K = 2.

Jika ditransformasikan ke dalam bentuk tabel maka hasil klasteringnya sebagai berikut:

Tabel 2. Hasil Klastering

No	Kecamatan	K=1			K=2			K=3		
		Klaster			Klaster			Klaster		
		1	2	3	1	2	3	1	2	3
1	Kecamatan Gemolong	T			S					R
2	Kecamatan Gesi	T			S			S		
3	Kecamatan Gondang	T			S			S		
4	Kecamatan Jenar	T			S			S		
5	Kecamatan Kalijambe	T			S			S		
6	Kecamatan Karangmalang	T		T				T		
7	Kecamatan Kedawaung	T			S					R
8	Kecamatan Masaran	T			S					R
9	Kecamatan Miri	T			S			S		
10	Kecamatan Mondokan	T			S			S		
11	Kecamatan Ngrampal	T			S					R
12	Kecamatan Plupuh	T			S			S		
13	Kecamatan Sambirejo	T			S			S		
14	Kecamatan Sambungmacan	T			S					R
15	Kecamatan Sidoharjo	T			S					R
16	Kecamatan Sragen	T		T				T		
17	Kecamatan Sukodono	T			S			S		
18	Kecamatan Sumberlawang	T		T				T		
19	Kecamatan Tangen	T			S			S		
20	Kecamatan Tanon	T			S					R

Keterangan:

T : Tinggi

S : Sedang

R : Rendah.

3.4. Diskusi

Hasil analisis kluster menunjukkan konsisten dengan penelitian oleh R. Adhitama, A. Burhanuddin, and A. Febriani yang menunjukkan efektivitas K-Means untuk analisis wilayah. Berdasarkan visualisasi pada Gambar 3, teridentifikasi bahwa kluster dengan nilai tinggi mencakup Kecamatan Sragen, Sumberlawang dan Karangmalang. Temuan ini dapat menjadi panduan yang strategis bagi pemerintah daerah dalam menentukan prioritas wilayah untuk promosi usaha.

4. KESIMPULAN

Pada klusterisasi dengan nilai $K = 3$, hasil menunjukkan adanya tiga kelompok wilayah Kecamatan terhadap persebaran izin usaha sesuai data dari Dinas Dinas Penanaman Modal Dan Pelayanan Terpadu Satu Pintu Kabupaten Sragen: kategori yang "Tinggi" (T) seperti Kecamatan Karangmalang, Sragen, Sumberlawang, kategori wilayah "Sedang" (S) seperti Kecamatan Gemolong, Kedawung, Masaran, Ngrampal, Sambungmacan, Sidoharjo dan Tanon, serta kategori yang "Rendah" (R) yaitu Kecamatan Gesi, Gondang, Jenar, Kalijambe, Miri, Mondokan, Plupuh, Sambirejo, Sukodono, Tangen. Dengan segmentasi ini, perencanaan program atau kegiatan yang menunjang pertumbuhan ekonomi dari sektor usaha mikro, kecil dan menengah dapat lebih tepat karena bisa disesuaikan dengan kondisi wilayah yang sudah diidentifikasi, sehingga alokasi anggaran dapat difokuskan pada Kecamatan yang membutuhkan perhatian lebih. Penelitian ini berkontribusi pada pengembangan pendekatan berbasis data untuk pemerataan izin usaha di Kabupaten Sragen. Penelitian lanjutan dapat menggunakan data tambahan seperti aksesibilitas wilayah atau faktor demografis untuk meningkatkan akurasi analisis. Dengan demikian, Dinas Penanaman Modal dan Pelayanan Terpadu Satu dapat terus menyesuaikan program secara dinamis dan memenuhi kebutuhan masyarakat dalam hal upaya meningkatkan perokominan dari sektor usaha mikro kecil menengah dengan lebih optimal.

DAFTAR PUSTAKA

- [1] N. Badriyah, Hozairi, and M. Walid, "Penentuan Bidang Minat Tugas Akhir Mahasiswa Teknik Informatika Universitas Islam Madura Menggunakan Metode K-Means," *J. Inform. Teknol. dan Sains*, vol. 5, no. 4, pp. 566–572, 2023, doi: 10.51401/jinteks.v5i4.2782.
- [2] S. Sumarlinda and W. Lestari, "Journal of Intelligent Decision Support System (IDSS) Decision support system for lecturer publication mapping using k-means clustering method," *J. Intell. Decis. Support Syst.*, vol. 5, no. 4, pp. 140–145, 2022.
- [3] I. Romli, "Penerapan Data Mining Menggunakan Algoritma K-Means Untuk Klasifikasi Penyakit Ispa," *Indones. J. Bus. Intell.*, vol. 4, no. 1, p. 10, 2021, doi: 10.21927/ijubi.v4i1.1727.
- [4] A. Azzam, A. Irma Purnamasari, and I. Ali, "Implementasi Algoritma K-Means Clustering Untuk Analisis Persebaran Umkm Di Jawa Barat," *JATI (Jurnal Mhs. Tek. Inform.)*, vol. 8, no. 3, pp. 3062–3070, 2024, doi: 10.36040/jati.v8i3.8450.
- [5] R. Adhitama, A. Burhanuddin, and A. Febriani, "Penerapan X Means Clustering Pada UMKM Kab Banyumas Yang Mendukung Mega Shifting Consumer Behavior Akibat Covid-19," *J. Informatics, Inf. Syst. Softw. Eng. Appl.*, vol. 4, no. 1, pp. 71–80, 2022, doi: 10.20895/inista.v4i1.429.
- [6] Normah, B. Rifai, S. Vambudi, and R. Maulana, "Analisa Sentimen Perkembangan Vtuber Dengan Metode Support Vector Machine Berbasis SMOTE," *J. Tek. Komput. AMIK BSI*, vol. 8, no. 2, pp. 174–180, 2022, doi: 10.31294/jtk.v4i2.
- [7] N. Syifa and R. N. Fahmi, "Implementasi Metode K-Means Clustering dalam Analisis Persebaran UMKM di Jawa Barat," *JOINS (Journal Inf. Syst.)*, vol. 6, no. 2, pp. 211–220, 2021, doi: 10.33633/joins.v6i2.5310.
- [8] S. A. Rahmah, "Review Terbaru Tentang Klusterisasi Data Mining Menggunakan Metode K-Means: Tantangan Dan Aplikasi," *Djtechno J. Teknol. Inf.*, vol. 5, no. 2, pp. 297–303, 2024, doi: 10.46576/djtechno.v5i2.4723.
- [9] 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, p. 96, 2021, [Online]. Available: www.bps.go.id
- [10] R. Supardi and I. Kanedi, "Implementasi Metode Algoritma K-Means Clustering pada Toko Eidelweis," *J. Teknol. Inf.*, vol. 4, no. 2, pp. 270–277, 2020, doi: 10.36294/jurti.v4i2.1444.
- [11] N. T. Hartanti, "Jurnal Nasional Teknologi dan Sistem Informasi Metode Elbow dan K-Means Guna

-
- Mengukur Kesiapan Siswa SMK Dalam Ujian Nasional,” *J. Nas. Teknol. dan Sist. Inf.*, vol. 06, no. 02, pp. 82–89, 2020, [Online]. Available: <https://teknosi.fti.unand.ac.id/index.php/teknosi/article/view/1499/pdf>
- [12] N. Ameliana, N. Suarna, and W. Prihartono, “Analisis Data Mining Pengelompokan Umkm Menggunakan Algoritma K-Means Clustering Di Provinsi Jawa Barat,” *JATI (Jurnal Mhs. Tek. Inform.*, vol. 8, no. 3, pp. 3261–3268, 2024, doi: 10.36040/jati.v8i3.9655.
- [13] N. Mirantika, “Penerapan Algoritma K-Means Clustering Untuk Pengelompokan Penyebaran Covid-19 di Provinsi Jawa Barat,” *Nuansa Inform.*, vol. 15, no. 2, pp. 92–98, 2021, doi: 10.25134/nuansa.v15i2.4321.
- [14] A. Sulistiyawati and E. Supriyanto, “Implementasi Algoritma K-means Clustering dalam Penentuan Siswa Kelas Unggulan,” *J. Tekno Kompak*, vol. 15, no. 2, p. 25, 2021, doi: 10.33365/jtk.v15i2.1162.
- [15] I. Wahyudi, M. B. Sulthan, and L. Suhartini, “Analisa Penentuan Cluster Terbaik Pada Metode K-Means Menggunakan Elbow Terhadap Sentra Industri Produksi Di Pamekasan,” *J. Apl. Teknol. Inf. dan Manaj.*, vol. 2, no. 2, pp. 72–81, 2021, doi: 10.31102/jatim.v2i2.1274.
- [16] B. Arifitama and A. Syahputra, “Analisis Data Mining Pada Klasterisasi UMKM Dengan Menggunakan Algoritma K-Means,” *J. Ind. Kreat. dan Inform.*, vol. 2, no. 2, pp. 66–72, 2022.
- [17] A. Nugraha, O. Nurdiawan, and G. Dwilestari, “Penerapan Data Mining Metode K-Means Clustering Untuk Analisa Penjualan Pada Toko Yana Sport,” *JATI (Jurnal Mhs. Tek. Inform.*, vol. 6, no. 2, pp. 849–855, 2022, doi: 10.36040/jati.v6i2.5755.