
ANALISIS POLA KETENAGAKERJAAN DI KABUPATEN BEKASI MENGGUNAKAN ALGORITMA K-MEANS CLUSTERING

Roki Fatih Musthofa¹⁾, Nono Heryana²⁾, Apriade Voutama³⁾

Program Studi Sistem Informasi, Fakultas Ilmu Komputer

Universitas Singaperbangsa Karawang

Karawang, Indonesia

e-mail: 2210631250029@student.unsika.ac.id¹⁾, nono@staff.unsika.ac.id²⁾, apriade.voutama@staff.unsika.ac.id³⁾

Abstrak

Permasalahan ketenagakerjaan di Kabupaten Bekasi masih ditandai oleh fluktuasi tingkat pengangguran serta ketimpangan penyerapan tenaga kerja antar sektor ekonomi. Kondisi tersebut menunjukkan perlunya pemetaan karakteristik sektor ekonomi untuk mendukung perumusan kebijakan ketenagakerjaan yang lebih tepat sasaran. Penelitian ini bertujuan mengelompokkan sektor lapangan usaha berdasarkan indikator Produk Domestik Regional Bruto (PDRB) sektoral, jumlah tenaga kerja terserap, dan produktivitas tenaga kerja menggunakan algoritma K-Means Clustering dengan metodologi *Knowledge Discovery in Databases* (KDD). Data yang digunakan merupakan data rata-rata periode 2022–2024. Hasil analisis menunjukkan bahwa sektor ekonomi di Kabupaten Bekasi terbagi menjadi tiga kluster, yaitu kluster unggulan dengan kontribusi ekonomi dan produktivitas tinggi, kluster menengah, serta kluster intervensi dengan nilai indikator relatif rendah. Evaluasi model menghasilkan nilai *Davies-Bouldin Index* sebesar 0,353 dan *Silhouette Coefficient* sebesar 0,598 yang menunjukkan kualitas kluster yang cukup baik. Hasil pengelompokan ini memberikan gambaran mengenai sektor ekonomi yang berpotensi menjadi motor pertumbuhan sekaligus sektor yang memerlukan dukungan kebijakan untuk meningkatkan penyerapan tenaga kerja.

Kata kunci : K-Means Clustering; KDD; Ketenagakerjaan; Sektor Ekonomi; Clustering

1. Pendahuluan

Indonesia merupakan negara dengan jumlah penduduk terbesar keempat di dunia dengan total sekitar 277,7 juta jiwa. Pertumbuhan penduduk yang terus meningkat memberikan tekanan terhadap berbagai aspek pembangunan, terutama dalam penyediaan lapangan kerja yang memadai [1]. Apabila peningkatan jumlah angkatan kerja tidak diimbangi dengan ketersediaan kesempatan kerja, maka risiko munculnya permasalahan sosial seperti pengangguran, kemiskinan, dan ketimpangan ekonomi akan semakin besar [2]. Salah satu indikator yang umum digunakan untuk mengukur kondisi ketenagakerjaan suatu wilayah adalah Tingkat Pengangguran Terbuka (TPT), yaitu persentase angkatan kerja yang belum terserap oleh pasar kerja. Tingginya nilai TPT mencerminkan belum optimalnya penyerapan tenaga kerja serta menunjukkan adanya ketidakseimbangan antara kebutuhan tenaga kerja dan kemampuan sektor ekonomi dalam menyerap angkatan kerja [3].

Provinsi Jawa Barat merupakan provinsi dengan jumlah penduduk terbesar di Indonesia dan memiliki aktivitas industri yang sangat berkembang (BPS). Kondisi tersebut menyebabkan tingginya jumlah angkatan kerja sekaligus meningkatkan tekanan terhadap pasar kerja di berbagai wilayah. Kabupaten Bekasi sebagai salah satu pusat kawasan industri terbesar di Jawa Barat memiliki peran strategis dalam pertumbuhan ekonomi regional [4]. Keberadaan berbagai kawasan industri seperti MM2100, EJIP, Jababeka, dan Delta Silicon menjadikan wilayah ini sebagai salah satu pusat manufaktur penting di Indonesia [5]. Meskipun demikian, tingginya aktivitas industri tidak selalu sejalan dengan optimalnya penyerapan tenaga kerja [6], sehingga Kabupaten Bekasi masih menghadapi permasalahan ketenagakerjaan, salah satunya tercermin dari tingkat pengangguran terbuka yang relatif tinggi [7].

Permasalahan tersebut menunjukkan adanya ketidakseimbangan antara kontribusi sektor ekonomi terhadap pertumbuhan daerah dengan kemampuan sektor tersebut dalam menyerap tenaga kerja. Beberapa sektor memiliki kontribusi ekonomi dan produktivitas yang tinggi namun bersifat padat modal sehingga daya serap tenaga kerjanya relatif rendah, sementara sektor lain memiliki jumlah tenaga kerja besar tetapi kontribusinya lebih kecil [8]. Oleh karena itu, diperlukan suatu pendekatan analisis yang mampu mengidentifikasi karakteristik sektor ekonomi berdasarkan indikator ketenagakerjaan sehingga dapat diketahui kelompok sektor yang berpotensi menjadi motor pertumbuhan maupun sektor yang memerlukan perhatian dan intervensi kebijakan.

Salah satu metode yang dapat digunakan untuk menganalisis pola data tersebut adalah teknik clustering dengan algoritma K-Means [9]. Metode ini merupakan salah satu teknik data mining yang berfungsi untuk mengelompokkan data ke dalam beberapa kluster berdasarkan tingkat kemiripan karakteristik tertentu. Dalam penelitian ini, proses pengelompokan dilakukan menggunakan indikator Produk Domestik Regional Bruto (PDRB) sektoral, jumlah penduduk yang bekerja, dan produktivitas tenaga kerja [10]. Melalui proses klusterisasi, sektor ekonomi di Kabupaten Bekasi dapat dikelompokkan ke dalam beberapa kategori dengan karakteristik yang berbeda sehingga memudahkan proses analisis dan interpretasi data [11].

Berdasarkan latar belakang tersebut, penelitian ini bertujuan untuk menganalisis pola ketenagakerjaan di Kabupaten Bekasi melalui pengelompokan sektor ekonomi menggunakan algoritma K-Means. Hasil penelitian diharapkan dapat

memberikan gambaran mengenai karakteristik sektor ekonomi berdasarkan kontribusi ekonomi, penyerapan tenaga kerja, dan produktivitas tenaga kerja, serta menjadi referensi dalam mendukung perumusan kebijakan ketenagakerjaan yang lebih tepat sasaran dan berbasis data.

2. Landasan Teori Clustering

Clustering merupakan salah satu metode dalam data mining yang digunakan untuk mengelompokkan sejumlah data ke dalam beberapa kelompok atau cluster berdasarkan tingkat kemiripan antar data. Objek-objek yang memiliki karakteristik atau nilai yang serupa akan ditempatkan dalam satu cluster yang sama, sementara objek yang berbeda akan berada pada cluster yang lain [12]. Kualitas hasil pengelompokan sangat dipengaruhi oleh metode atau algoritma yang digunakan, karena setiap metode memiliki cara tersendiri dalam mengukur kedekatan dan menentukan batas antar kelompok [13]. Clustering berbeda dengan pengelompokan biasa atau group, karena data dalam satu cluster tidak harus memiliki kondisi yang sama persis, melainkan cukup memiliki tingkat kedekatan tertentu berdasarkan karakteristik yang dianalisis, misalnya melalui perhitungan jarak seperti Euclidean distance [14]. Teknik ini banyak diterapkan dalam berbagai bidang untuk membantu proses identifikasi pola dan pengambilan keputusan, termasuk dalam pengelompokan data berdasarkan nilai hasil belajar [15]. Dalam data mining, salah satu metode yang banyak dikenal dan sering diterapkan adalah teknik clustering, yang memiliki peran penting dalam analisis pola, pengambilan keputusan, pembelajaran mesin, dan berbagai bidang lainnya [16].

Algoritma K-Means

Algoritma K-Means merupakan salah satu metode clustering yang bersifat non-hierarchy yang digunakan untuk melakukan pengelompokan data. Prosesnya dimulai dengan pembentukan partisi awal, kemudian dilakukan perbaikan secara berulang melalui proses iterasi sampai tidak terjadi perubahan yang signifikan pada hasil pengelompokan [17]. K-Means juga termasuk ke dalam algoritma unsupervised learning dimana pengelompokan dilakukan dengan menghitung jarak antara setiap data dengan pusat cluster menggunakan rumus jarak, seperti Euclidean distance, di mana semakin kecil jarak antar data maka semakin tinggi tingkat kemiripannya [18]. Setiap cluster diwakili oleh sebuah titik pusat atau yang biasa disebut centroid dan algoritma ini berfokus untuk meminimalkan jarak total antara data dengan centroid, sehingga menghasilkan kelompok data yang lebih homogen. Karena kemampuannya tersebut, K-Means menjadi salah satu metode clustering yang paling banyak digunakan dalam berbagai analisis data [19].

Metode Elbow

Metode Elbow digunakan untuk menentukan jumlah kluster yang optimal dengan mengevaluasi kurva yang menunjukkan hubungan antara jumlah kluster dan nilai SSE (Sum of Squared Errors), yaitu jumlah kuadrat jarak setiap titik data terhadap pusat kluster terdekatnya [12]. Nilai SSE dihitung untuk berbagai jumlah kluster, dan jumlah kluster optimal ditentukan ketika penurunan SSE tidak lagi signifikan, membentuk titik “siku” pada kurva. Metode ini dinamakan Elbow karena kurva plot yang dihasilkan menyerupai bentuk siku pada bagian elbow-nya [20].

Davies-Bouldin Index (DBI)

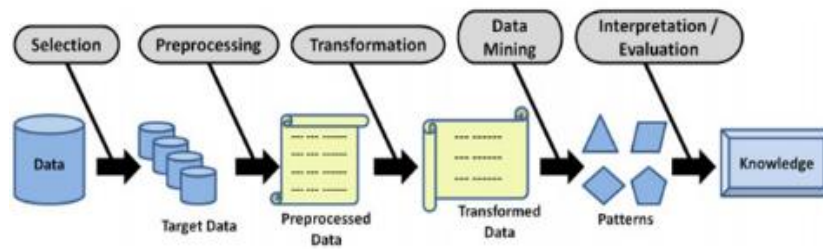
Davies-Bouldin Index (DBI) merupakan salah satu matriks evaluasi yang digunakan untuk menilai kualitas clustering dan menentukan model atau konfigurasi optimal dalam algoritma clustering. Sementara itu, Azran N. (2023) menyatakan bahwa clustering dikatakan optimal apabila nilai DBI mendekati nol, yang menunjukkan bahwa jarak antar titik dalam satu cluster minimal [11].

Silhouette Coefficient

Silhouette coefficient merupakan metode evaluasi yang menggabungkan konsep cohesion dan separation untuk menilai kualitas hasil klusterisasi [21]. Rentang nilai silhouette berkisar dari 1 hingga -1. Semakin kecil jarak antara objek dalam satu kelompok dan semakin besar jarak antara objek dengan objek di kelompok lain, maka semakin baik kualitas kluster tersebut.

3. Metode Penelitian

Knowledge Discovery in Databases (KDD) adalah metodologi yang menggunakan pendekatan sistematis untuk mengevaluasi pola dan struktur data, dengan tujuan mengelompokkan entitas yang memiliki karakteristik serupa ke dalam kelompok tertentu.



Gambar 1. Tahapan Metodologi KDD [22]

Secara umum, metodologi KDD ini terdiri dari 5 proses atau tahapan didalamnya, yang dilakukan secara berurutan sebagai berikut.

Data Selection

Tahap ini berfokus pada pemilihan data yang akan dianalisis, dengan tujuan memperoleh subset data yang relevan dan signifikan. Pemilihan ini bertujuan agar proses analisis dapat dilakukan secara efisien dan hasilnya lebih akurat sesuai dengan tujuan penelitian.

Data Preprocessing/Cleaning

Pada tahap prapemrosesan, data dibersihkan dari kesalahan, missing values, dan noise. Selain itu, dilakukan transformasi awal seperti normalisasi data, scaling, atau reduksi dimensi menggunakan metode seperti Principal Component Analysis (PCA) untuk menyederhanakan data dan feature selection agar analisis berikutnya lebih optimal.

Data Transformation

Tahap transformasi melibatkan perubahan dan pengkodean data ke dalam format yang lebih sesuai untuk analisis. Contohnya, data numerik dapat diubah menjadi kategori, atau data kategori dikonversi menjadi numerik. Transformasi ini juga mencakup penggabungan, agregasi, atau reduksi atribut data sesuai dengan pola informasi yang ingin dianalisis.

Data Mining

Data mining merupakan tahap inti dari KDD, di mana algoritma khusus diterapkan untuk mengekstraksi pola, hubungan, dan informasi tersembunyi dari dataset yang telah dipersiapkan. Tahap ini bertujuan untuk menghasilkan pengetahuan baru yang bermanfaat untuk pengambilan keputusan atau pengembangan strategi berbasis data.

Evaluation/Interpretation

Pada tahap ini, kualitas hasil analisis dievaluasi menggunakan metrik tertentu, seperti SSE (*Sum of Squared Errors*), *Silhouette Coefficient*, dan *Davies Bouldin Index*, untuk mengukur keakuratan dan konsistensi hasil clustering atau model yang diperoleh. Selanjutnya, hasil dianalisis dan diinterpretasikan agar sesuai dengan tujuan penelitian atau hipotesis yang telah ditetapkan sebelumnya.

4. Hasil Penelitian

Tahapan dalam penelitian ini mengikuti metode yang diterapkan, yaitu pendekatan KDD (*Knowledge Discovery in Databases*).

Data Selection

Dataset yang digunakan dalam penelitian ini merupakan data sektoral ketenagakerjaan Kabupaten Bekasi periode 2022–2024 yang diperoleh dari publikasi resmi Dinas Tenaga Kerja Kabupaten Bekasi dan Badan Pusat Statistik (BPS). Pada tahap ini dilakukan proses seleksi data dengan memilih atribut yang relevan dengan tujuan penelitian, yaitu PDRB sektoral, jumlah penduduk yang bekerja (PYB), dan produktivitas tenaga kerja. Atribut yang sesuai dan relevan dengan tujuan penelitian akan diolah lebih lanjut. Adapun hasil akhir data selection dapat dilihat pada Gambar 2.

No	Tahun	KLUI	Sektor Lapangan Usaha	PDRB	PYB	Produktivitas
1	2022	A	Pertanian, Kehutanan dan Perikanan	2593003.4	79459	32.63
2	2023	A	Pertanian, Kehutanan dan Perikanan	2651590.9	63475	41.77
3	2024	A	Pertanian, Kehutanan dan Perikanan	2692499.7	50677	53.13
4	2022	B	Pertambangan dan Penggalian	1939164.4	6074	319.26
5	2023	B	Pertambangan dan Penggalian	1900229.8	5978	317.89
6	2024	B	Pertambangan dan Penggalian	1827415.1	3973	459.96
7	2022	C	Industri Pengolahan	206837718	491173	421.11
8	2023	C	Industri Pengolahan	218988737	449479	487.21
9	2024	C	Industri Pengolahan	229793214	416636	551.54
10	2022	D	Pengadaan listrik dan gas	1234212.1	2180	566.15
11	2023	D	Pengadaan listrik dan gas	1261112.5	6728	187.44
12	2024	D	Pengadaan listrik dan gas	1247539.2	8290	150.49
13	2022	E	Pengadaan air, pengelolaan sampah, limbah dan daur ulang	80308.8	18633	4.31
14	2023	E	Pengadaan air, pengelolaan sampah, limbah dan daur ulang	86191.7	25140	3.43
15	2024	E	Pengadaan air, pengelolaan sampah, limbah dan daur ulang	90346.47	10378	8.71
16	2022	F	Konstruksi	16082985	94705	169.82
17	2023	F	Konstruksi	16066696	80163	200.43
18	2024	F	Konstruksi	17374695	88249	196.88
19	2022	G	Perdagangan besar dan eceran; Resparasi mobil dan sepeda motor	15288908	417818	36.59
20	2023	G	Perdagangan besar dan eceran; Resparasi mobil dan sepeda motor	15783330	301776	52.3
21	2024	G	Perdagangan besar dan eceran; Resparasi mobil dan sepeda motor	16157068	310831	51.98
22	2022	H	Transportasi dan pergudangan	2968203.9	143263	20.72
23	2023	H	Transportasi dan pergudangan	3288408.7	95292	34.51
24	2024	H	Transportasi dan pergudangan	3736964.5	120718	30.96
25	2022	I	Penyediaan akomodasi dan makan minum	1478512.7	150566	9.82
26	2023	I	Penyediaan akomodasi dan makan minum	1606347.3	124508	12.9
27	2024	I	Penyediaan akomodasi dan makan minum	1733231.3	146664	11.82
28	2022	J	Informasi dan Komunikasi	5298853.3	19880	266.54
29	2023	J	Informasi dan Komunikasi	5653867.9	7635	740.39
30	2024	J	Informasi dan Komunikasi	6107865.1	13348	457.59
31	2022	K	Jasa keuangan dan asuransi	2292656.3	21202	108.13
32	2023	K	Jasa keuangan dan asuransi	2268675.3	13465	168.49
33	2024	K	Jasa keuangan dan asuransi	2374868.5	17043	139.35
34	2022	L	Real Estate	1726782.7	27259	63.35
35	2023	L	Real Estate	1862027.3	17918	103.92
36	2024	L	Real Estate	1970361.8	24058	81.9
37	2022	M,N	Jasa Perusahaan	326899.8	52998	6.17
38	2023	M,N	Jasa Perusahaan	344891.2	46216	7.46
39	2024	M,N	Jasa Perusahaan	355160.43	41557	8.55
40	2022	O	Administrasi Pemerintah, pertahanan dan jaminan sosial wajib	1658346.6	37194	44.59
41	2023	O	Administrasi Pemerintah, pertahanan dan jaminan sosial wajib	1677398.3	33251	50.45
42	2024	O	Administrasi Pemerintah, pertahanan dan jaminan sosial wajib	1861782.7	33314	55.89
43	2022	P	Jasa Pendidikan	2898445	79421	36.49
44	2023	P	Jasa Pendidikan	3147364.7	67804	46.42
45	2024	P	Jasa Pendidikan	3505727.5	74013	47.37
46	2022	Q	Jasa Kesehatan dan kegiatan sosial	700833	30635	22.88
47	2023	Q	Jasa Kesehatan dan kegiatan sosial	752732.4	29293	25.7
48	2024	Q	Jasa Kesehatan dan kegiatan sosial	787138.21	23146	34.01
49	2022	R,S,T,U	Jasa lainnya	1724989.5	127208	13.56
50	2023	R,S,T,U	Jasa lainnya	1886201.9	83501	22.59
51	2024	R,S,T,U	Jasa lainnya	2048912.7	94183	21.75

Gambar 2. Hasil Data Selection

Preprocessing Data/Cleaning Data

Pada tahap ini bertujuan untuk memastikan data berada dalam kondisi baik sebelum dilakukan analisis. Proses yang dilakukan meliputi pengecekan *missing value* serta identifikasi data duplikat pada seluruh atribut penelitian. Pemeriksaan ini dilakukan untuk menjaga keakuratan dan konsistensi data yang digunakan. Berdasarkan hasil pengecekan, tidak ditemukan nilai kosong maupun data ganda pada dataset.

Data Transformation

Pada tahap ini, nilai atau isi dalam dataset akan diubah kedalam format yang sesuai, sehingga data dapat diolah dengan baik.

```

... <class 'pandas.core.frame.DataFrame'>
RangeIndex: 51 entries, 0 to 50
Data columns (total 4 columns):
#   Column                Non-Null Count  Dtype
---  -
0   Sektor Lapangan Usaha  51 non-null     object
1   PDRB                   51 non-null     float64
2   PYB                    51 non-null     float64
3   Produktivitas          51 non-null     float64
dtypes: float64(3), object(1)
memory usage: 1.7+ KB

```

Gambar 3. Hasil Data Transformation

Dalam penelitian ini, data yang bersifat numerik akan diubah ke dalam bentuk float, sedangkan data non-numerik akan diubah kedalam tipe data object. Maka, hasil yang dapat dilihat pada gambar 3.

Data Mining

Tahap ini merupakan proses data mining dengan menerapkan algoritma K-Means clustering untuk mengelompokkan sektor lapangan usaha berdasarkan kemiripan karakteristik indikator ketenagakerjaan. Sebelum proses klusterisasi dilakukan, terlebih dahulu ditentukan jumlah kluster yang optimal menggunakan metode Elbow, yaitu dengan menganalisis pola penurunan nilai *Sum of Squared Errors* (SSE) pada beberapa jumlah kluster.

Jumlah kluster optimal ditentukan pada titik perubahan kurva yang menyerupai bentuk siku (*elbow*), yang menunjukkan bahwa penambahan kluster selanjutnya tidak lagi memberikan penurunan nilai SSE yang signifikan.

Berdasarkan hasil analisis metode Elbow, diperoleh jumlah kluster optimal sebanyak tiga kluster. Selanjutnya algoritma K-Means dijalankan untuk mengelompokkan setiap sektor ke dalam kluster yang sesuai, dan hasil pengelompokan divisualisasikan dalam bentuk scatter plot untuk mempermudah interpretasi karakteristik masing-masing kluster.

Nilai SSE dihitung untuk nilai kluster 1 – 10, dan didapatkan hasil sebagai berikut.

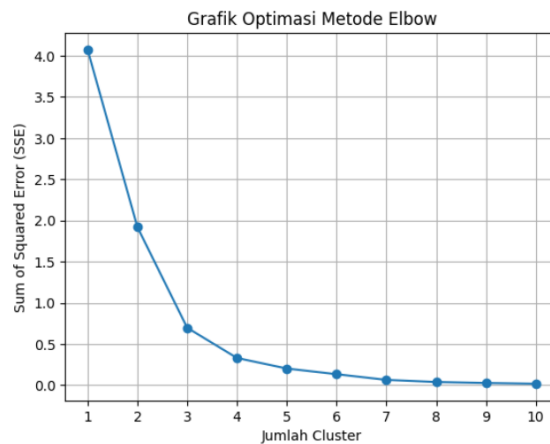
```

... K = 1 SSE = 4.073719930411013
    K = 2 SSE = 1.9203204248494892
    K = 3 SSE = 0.6981884030029055
    K = 4 SSE = 0.3314828864998438
    K = 5 SSE = 0.20360951476426725
    K = 6 SSE = 0.13462527069725663
    K = 7 SSE = 0.06574376116607936
    K = 8 SSE = 0.03998295237518292
    K = 9 SSE = 0.0279062740332384
    K = 10 SSE = 0.019016928143578353

```

Gambar 4. Hasil Perhitungan Nilai SSE

Setelah diketahui nilai SSE, selanjutnya dibuat grafik Elbow untuk memvisualisasikan hubungan antara jumlah kluster yang berbeda dengan nilai SSE.



Gambar 5. Grafik Metode Elbow

Pada Gambar 5., titik yang membentuk siku ada pada cluster 3. Hal ini mengindikasikan bahwa jumlah optimal cluster adalah 3 cluster. Selanjutnya, masuk kedalam proses K-Means Clustering dengan langkah awalnya adalah menghitung nilai centroid untuk masing-masing kluster berdasarkan data yang tersedia.

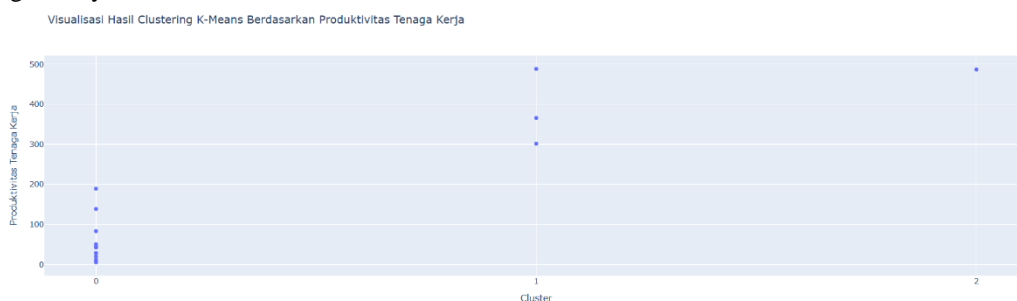
```

... [[0.01791051 0.17713895 0.09922385]
     [0.01307104 0.00646431 0.78641686]
     [1.         1.         0.99678192]]
array([0, 0, 2, 1, 0, 0, 0, 0, 0, 0, 1, 0, 0, 1, 0, 0], dtype=int32)

```

Gambar 6. Hasil Centroid

Dalam Gambar 6., Tiga baris awal adalah nilai centroid, sementara baris terakhir menampilkan label kluster untuk setiap sampel data. Angka dalam label kluster (0,1, dan 2) mengindikasikan kluster mana setiap sampel data termasuk di dalamnya. Untuk dapat memahami lebih mudah, maka dari hasil proses K-Means Clustering, dibuatlah visualisasi grafiknya.



Gambar 7. Hasil Visualisasi Clustering

Pada visualisasi tersebut, kluster 1 dan kluster 2 memiliki posisi yang relatif lebih tinggi dibandingkan kluster 0. Hal ini menunjukkan bahwa sektor lapangan usaha yang termasuk dalam kluster 1 dan kluster 2 cenderung memiliki tingkat produktivitas tenaga kerja yang lebih baik. Meskipun terdapat satu sektor pada kluster 1 yang nilainya sedikit lebih tinggi dibandingkan kluster 2, secara umum kedua kluster tersebut tetap merepresentasikan kelompok dengan produktivitas yang lebih tinggi. Sementara itu, kluster 0 berada pada rentang nilai yang lebih rendah, sehingga mencerminkan sektor dengan tingkat produktivitas tenaga kerja yang masih perlu ditingkatkan.

Evaluation/Interpretation

Pada tahap evaluasi ini, kluster yang telah terbentuk selanjutnya dianalisis untuk mengetahui kualitasnya. Metode Davies-Bouldin Index (DBI) digunakan untuk mengukur seberapa baik tingkat pemisahan dan kekompakan antar kluster, sehingga dapat menunjukkan seberapa optimal hasil klusterisasi yang dihasilkan.

... Davies-Bouldin Index: 0.353270032891489

Gambar 8. Hasil DBI

Dari perhitungan DBI diatas, nilai yang didapat sebesar 0.353. Dalam aturan DBI, semakin nilai DBI mendekati 0 atau semakin rendah nilainya, maka semakin baik pula cluster yang dihasilkan. Untuk memperkuat hasil evaluasi ini, maka dalam penelitian ini juga memakai metode *Silhouette Coefficient*.

... Silhouette Score: 0.598

Gambar 9. Hasil Silhoutte Coefficient

Perhitungan Nilai Silhouette Nilai Silhouette yang didapat adalah 0.598., nilai tersebut masuk kedalam rentang nilai Silhouette Coefficient, yakni berkisar antara -1 sampai dengan 1.

Selanjutnya untuk memudahkan dalam memahami dan melihat hasil pengelompokkan, maka dibuatlah suatu tabel sebagai berikut dengan memasukkan sektor lapangan usaha, cluster, dan kriteria.

Tabel 1. Penyebaran Kelompok Cluster

NO	SEKTOR LAPANGAN USAHA	CLUSTER	KRITERIA
1	Pertanian, Kehutanan dan Perikanan	0	Perlu Intervensi
2	Pertambangan dan Penggalian	1	Potensial
3	Industri Pengolahan	2	Unggulan
4	Pengadaan listrik dan gas	1	Potensial
5	Pengadaan air, pengelolaan sampah, limbah dan daur ulang	0	Perlu Intervensi
6	Konstruksi	0	Perlu Intervensi
7	Perdagangan besar dan eceran; Resparasi mobil dan sepeda motor	0	Perlu Intervensi
8	Transportasi dan pergudangan	0	Perlu Intervensi
9	Penyediaan akomodasi dan makan minum	0	Perlu Intervensi
10	Informasi dan Komunikasi	1	Potensial
11	Jasa keuangan dan asuransi	0	Perlu Intervensi
12	Real Estate	0	Perlu Intervensi
13	Jasa Perusahaan	0	Perlu Intervensi
14	Administrasi Pemerintah, pertahanan dan jaminan sosial wajib	0	Perlu Intervensi
15	Jasa Pendidikan	0	Perlu Intervensi
16	Jasa Kesehatan dan kegiatan sosial	0	Perlu Intervensi
17	Jasa lainnya	0	Perlu Intervensi

Hasil pengelompokan menunjukkan adanya perbedaan karakteristik antar sektor lapangan usaha berdasarkan nilai PDRB, PYB, dan produktivitas tenaga kerja. Klaster unggulan ditandai dengan nilai ketiga indikator yang relatif tinggi, sehingga berperan sebagai motor pertumbuhan ekonomi daerah. Namun demikian, sektor dalam klaster ini cenderung bersifat padat modal, sehingga peningkatan output ekonomi tidak selalu diikuti dengan penyerapan tenaga kerja yang optimal. Sebaliknya, klaster intervensi memiliki kontribusi ekonomi dan produktivitas yang lebih rendah, tetapi berpotensi dalam menyerap tenaga kerja apabila didukung melalui peningkatan keterampilan, penguatan UMKM, serta modernisasi proses usaha. Oleh karena itu, hasil klasterisasi ini memberikan gambaran bahwa strategi pembangunan perlu diarahkan secara berbeda pada tiap klaster, dengan mendorong ekspansi pada sektor unggulan sekaligus memperkuat sektor yang masih tertinggal agar mampu meningkatkan penyerapan tenaga kerja secara lebih merata.

5. Kesimpulan

Penelitian ini menganalisis pola ketenagakerjaan di Kabupaten Bekasi menggunakan algoritma K-Means Clustering. Berdasarkan hasil pengolahan data melalui tahapan metodologi *Knowledge Discovery in Databases* (KDD) diperoleh bahwa perhitungan jumlah klaster optimal menggunakan metode Elbow menunjukkan bahwa jumlah klaster yang paling sesuai adalah sebanyak 3 klaster. Selanjutnya, hasil evaluasi menggunakan metode *Davies-Bouldin Index* (DBI) dan *Silhouette Coefficient* menunjukkan bahwa klaster yang terbentuk memiliki kualitas yang cukup baik, sehingga mampu merepresentasikan perbedaan karakteristik antar sektor secara jelas. Berdasarkan hasil pengelompokan, klaster 0 didominasi oleh sektor dengan tingkat produktivitas tenaga kerja yang relatif lebih rendah, sehingga memerlukan perhatian dan intervensi lebih lanjut. Sementara itu, klaster 1 dan klaster 2 terdiri dari sektor-sektor dengan tingkat produktivitas yang lebih tinggi, sehingga berperan sebagai kelompok unggulan dalam mendorong pertumbuhan ekonomi. Perbedaan ini menunjukkan bahwa setiap klaster memiliki karakteristik dan potensi yang berbeda, sehingga memerlukan strategi pengembangan yang disesuaikan untuk meningkatkan kontribusi ekonomi dan penyerapan tenaga kerja.

6. Daftar Pustaka

- [1] E. Sinambela, I. Harahap, and Z. M. Nawawi, "Pertumbuhan Penduduk dan Kualitas Manusia dalam Meningkatkan Ekonomi Syariah di Indonesia (Literatur Review)," *Ekon. J. Ilmu Ekon. dan Stud. Pembang.*, vol. 24, no. 2, pp. 301–308, 2024, [Online]. Available: <https://jurnal.umsu.ac.id/index.php/ekawan/article/view/18186>
- [2] U. Syafiyah, D. P. Puspitasari, I. Asrafi, B. Wicaksono, and F. M. Sirait, "Analisis Perbandingan Hierarchical dan Non-Hierarchical Clustering Pada Data Indikator Ketenagakerjaan di Jawa Barat Tahun 2020," *Semin. Nas. Off. Stat.*, vol. 2022, no. 1, pp. 803–812, 2022, doi: 10.34123/semnasoffstat.v2022i1.1221.
- [3] A. Aziz *et al.*, "Klasterisasi Data Tenaga Kerja Terbuka Menurut Provinsi Dengan Penggunaan Algoritma K - Means 1,2," vol. 10, no. 3, pp. 444–456, 2023.
- [4] Y. I. Astuty, Marwah Noer, Demi Stevany, Brenda Arham, Brigita Maria R, and Adi Wibowo, "Evaluasi Kesesuaian Kawasan Peruntukan Industri Menggunakan Model Spasial (Studi Kasus : Kabupaten Bekasi)," *J. Pendidik. Geogr. Undiksha*, vol. 11, no. 2, pp. 123–132, 2023, doi: 10.23887/jjpp.v11i2.61536.
- [5] M. T. Irma Damayanti, Chiara Zahra Maulani, Bugi Alifya, Aulia Rofika Dewi, Prima Junior Seven M, Ir. Tri Mulyono, "PENGARUH KAWASAN INDUSTRI CIKARANG SEBAGAI HINTERLAND DENGAN KONEKTIVITAS PELABUHAN YANG OPTIMAL," 2024, [Online]. Available: https://www.mendeley.com/catalogue/c7361518-c4a8-3ae7-8630-b21ab372684e/?utm_source=desktop&utm_medium=1.19.4&utm_campaign=open_catalog&userDocumentId=%7Bb775f4b3-805a-44f7-a539-123cb7f8830d%7D
- [6] D. Safira, E. Dwi Lestari, M. Iffa, and S. Annisa, "Pengelompokan Jumlah Penduduk Sumatera Barat," *J. Ilm. Rekayasa dan Manaj. Sist. Inf.*, vol. 6, no. 1, pp. 26–31, 2020.
- [7] J. Juswadi and P. Sumarna, "Produktivitas Tenaga Kerja Sektor Pertanian dan Korelasinya dengan Usia Petani di Jawa Barat," *Paspalum J. Ilm. Pertan.*, vol. 11, no. 2, p. 361, 2023, doi: 10.35138/paspalum.v11i2.630.
- [8] B. U. Afrida M, Ahmad Rafli, Atif Naufal Y, Masturi, Rudiansyah, "Jurnal Manajemen, Akuntansi dan Pendidikan (JAMAPEDIK)," *J. Manajemen, Akunt. dan Pendidik.*, vol. 1, no. 2, pp. 316–325, 2024, doi: 10.59971/jamapedik.v1i2.65.
- [9] M. Rachma and C. Budy Santoso, "Klasterisasi Data Pengangguran Di Pulau Jawa Menggunakan Algoritma K-Means Dalam Penanggulangan Pengangguran Tahun 2020-2023," *IDEALIS Indones. J. Inf. Syst.*, vol. 8, no. 2, pp. 202–209, 2025, doi: 10.36080/idealisis.v8i2.3522.
- [10] V. Herlinda and D. Darwis, "Analisis Clustering Untuk Recredesialing Fasilitas Kesehatan Menggunakan Metode Fuzzy C-Means," *Darwis, Dartono*, vol. 2, no. 2, pp. 94–99, 2021, [Online]. Available: <http://jim.teknokrat.ac.id/index.php/JTSI>
- [11] 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
- [12] Noviyya Adawiyah, Nina Sulistiyowat, and ohamad Jajuli, "Klasterisasi Kasus Kekerasan Terhadap Anak dan

- Perempuan Berdasarkan Algoritma K-Means,” *Gener. J.*, vol. 5, no. 2, pp. 69–80, 2021.
- [13] Ayu Pangestu, “Penerapan Data Mining Menggunakan Algoritma K-Means Pengelompokan Pelanggan Berdasarkan Kubikasi Air Terjual Menggunakan Weka,” *JUST IT J. Sist. Informasi, Teknol. Inf. dan Komput.*, vol. 11, no. 3, pp. 67–71, 2022, [Online]. Available: <https://jurnal.umj.ac.id/index.php/just-it/article/view/11591>
- [14] A. D. Kuswanto, A. N. Fadhila, P. T. Setiawan, M. K. Setiawan, and D. R. Syahputra, “Penerapan K-Means Clustering Untuk Menentukan Jumlah Pengangguran Berdasarkan Umur (Studi Kasus di Badan Statistik Provinsi DKI Jakarta 2020-2022),” *Repeater Publ. Tek. Inform. dan Jar.*, vol. 2, no. 3, pp. 135–146, 2024.
- [15] A. P. Riani, A. Voutama, and T. Ridwan, “Jurnal Teknologi Sistem Informasi dan Sistem Komputer TGD Penerapan K-Means Clustering Dalam Pengelompokan Hasil Belajar Peserta Didik Dengan Metode Elbow Jurnal Teknologi Sistem Informasi dan Sistem Komputer TGD,” *J. Teknol. Sist. Inf. dan Sist. Komput. TGD*, vol. 6, no. 1, pp. 164–172, 2023.
- [16] S. Paembonan and H. Abduh, “Penerapan Metode Silhouette Coefficient untuk Evaluasi Clustering Obat,” *PENA Tek. J. Ilm. Ilmu-Ilmu Tek.*, vol. 6, no. 2, pp. 48–54, 2021, doi: 10.51557/pt_jiit.v6i2.659.
- [17] A. A. H. Tasyifa Nafsiah Muthmainah, Apriade Voutama, “KLASTERISASI PENERIMA BANTUAN LANGSUNG TUNAI (BLT) DANA DESA MENGGUNAKAN ALGORITMA K-MEANS,” *JATI (Jurnal Mhs. Tek. Inform.)*, vol. 8, no. 3, pp. 3916–3919, 2024.
- [18] A. J. Diana Hidayati, Yahya, “Implementasi Data Mining Menggunakan Algoritma K-Means untuk Pengelompokan Obat di Puskesmas Kerongkong Kecamatan Suralaga,” *J. Pengemb. Rekayasa Inform. dan Komput.*, vol. 1, no. 2, pp. 78–88, 2023, [Online]. Available: <https://doi.org/10.29408/jprinter.v1i2.22006>
- [19] D. P. S. Dewi, “Analisis Clustering Penyebaran Hiv Di Karawang Berdasarkan Kecamatan Dengan Algoritma K-Means,” *J. Inform. dan Tek. Elektro Terap.*, vol. 12, no. 3, 2024, doi: 10.23960/jitet.v12i3.4878.
- [20] A. R. Lashiyanti, I. R. Munthe, and F. A. Nasution, “Optimisasi Klasterisasi Nilai Ujian Nasional dengan Pendekatan Algoritma,” *J. Ilmu Komput. dan Sist. Inf.*, vol. 6, no. 1, pp. 14–20, 2023, [Online]. Available: <https://ejournal.sisfokomtek.org/index.php/jikom/article/view/1550>
- [21] T. A. Novia, D. Swanjaya, and R. Wulaningrum, “Analisa Clustering Aplikasi Laporan Kasus Kekerasan,” *J. Borneo Inform. dan Tek. Komput.*, vol. 2, no. 1, pp. 17–30, 2022, doi: 10.35334/jbit.v2i1.2689.
- [22] B. N. S. Afifa Atira, “Penerapan Silhouette Coefficient, Elbow Method dan Gap Statistics untuk Penentuan Cluster Optimum dalam Pengelompokan Provinsi di Indonesia Berdasarkan Indeks Kebahagiaan,” *Ilmiah, J. Pendidikan, Wahana*, vol. 9, no. September, pp. 76–86, 2023.