

KLASTERISASI TINGKAT KELAYAKAN PROVINSI DALAM PEMBANGUNAN KAWASAN INDUSTRI MENGGUNAKAN ALGORITMA *K-MEANS*

Ibrahim^{1*}, Wendi Usino²

^{1,2} Sistem Informasi, Fakultas Teknologi Informasi, Universitas Budi Luhur, Jakarta, Indonesia

Email: ^{1*}ibrahim.cbshp@gmail.com, ²wendi.usino@budiluhur.ac.id
(* : corresponding author)

Abstrak-Pembangunan atau pengembangan kawasan industri merupakan hal positif bagi pihak industri atau pihak pemerintah, namun masih seringkali keberadaan kawasan industri berdampak negatif bagi masyarakat sekitar dan juga penentuan lokasi yang kontradiksi terhadap kawasan pariwisata, oleh karena itu perlu mempertimbangkan kepadatan penduduk dan sektor pariwisata. Selain itu dalam upaya memaksimalkan segala operasional industri maka aspek infrastruktur transportasi dan aspek tenaga kerja menjadi hal penting untuk dipertimbangkan. Penelitian ini bertujuan untuk mendapatkan informasi terkait tingkat kelayakan provinsi dalam pembangunan Kawasan Peruntukan Industri (KPI) sehingga hasil penelitian ini dapat dimanfaatkan sebagai pertimbangan dalam proses evaluasi Rencana Tata Ruang Wilayah (RTRW) suatu Provinsi khususnya pada sektor KPI. Berdasarkan permasalahan yang ada, dapat diketahui kelompok-kelompok provinsi dengan masing-masing tingkat kelayakannya dengan memanfaatkan Bahasa Pemrograman Python dalam menerapkan pendekatan *data mining* yaitu metode *clustering*, berdasarkan *dataset* yang digunakan dan hasil dari metode evaluasi seperti metode *Elbow*, *Silhouette Score* dan *Davies Bouldin Index (DBI)* yang dilakukan terhadap beberapa algoritma *clustering* seperti algoritma *K-Means*, *Agglomerative*, *BIRCH* dan *Spectral*, maka terpilih algoritma *K-Means* sebagai algoritma terbaik dengan *cluster* berjumlah 6. Berdasarkan identifikasi masalah maka proses perhitungan metode *clustering* dilakukan berdasarkan 7 atribut dengan data pada tahun 2022 yaitu atribut Kepadatan Penduduk, Jumlah Bandara, Jumlah Pelabuhan, Panjang Jalan Negara/ Nasional, Jumlah Angkatan Kerja, Tingkat Pendidikan SMA Sederajat, dan Jumlah Wisatawan Mancanegara. Dengan terbentuknya 6 *cluster* maka terdapat 6 kelompok provinsi dengan tingkat kelayakan yang berbeda, yang pertama terdapat Kelayakan *Cluster* Tingkat 1: 3 Provinsi, Tingkat 2: 11 Provinsi, Tingkat 3: 15 Provinsi, Tingkat 4: 1 Provinsi, Tingkat 5: 1 Provinsi, Tingkat 6: 1 Provinsi.

Kata Kunci: *Data mining*, *Clustering*, *K-Means*, Kawasan Industri

CLUSTERIZATION OF PROVINCIAL FEASIBILITY LEVELS IN THE DEVELOPMENT OF THE INDUSTRIAL AREA USING *K-MEANS* ALGORITHM

Abstract-The development or expansion of industrial areas is a positive thing for the industry or the government, but often the existence of industrial areas has a negative impact on the surrounding community and also the determination of locations that are contradictory to tourism areas, therefore it is necessary to consider population density and the tourism sector. In addition, in an effort to maximize all industrial operations, aspects of transportation infrastructure and labor aspects are important things to consider. This research aims to obtain information related to the level of provincial feasibility in the development of Industrial Areas so that the results of this research can be used as a consideration in the evaluation process of the Regional Spatial Planning Plan of a Province, especially in the Industrial Areas sector. Based on the existing problems, it can be seen that clusters of provinces with their respective levels of feasibility can be identified by utilizing the Python Programming Language in implementing a data mining approach, namely the clustering method, based on the dataset used and the results of evaluation methods such as the Elbow method, Silhouette Score and Davies Bouldin Index (DBI) which are carried out on several clustering algorithms such as the K-Means, Agglomerative, BIRCH and Spectral algorithms, the K-Means algorithm is selected as the best algorithm with 6 clusters. Based on the identification of the problem, the calculation process of the clustering method is carried out based on 7 attributes with data in 2022, namely the Population Density attribute, Number of Airports, Number of Ports, Length of National/State Roads, Number of Workforce, High School Education Level, and Number of Foreign Tourists. With the formation of 6 clusters, there are 6 clusters of provinces with different levels of feasibility, the first is Cluster Feasibility Level 1: 3 Provinces, Level 2: 11 Provinces, Level 3: 15 Provinces, Level 4: 1 Province, Level 5: 1 Province, Level 6: 1 Province.

Keywords: *Data mining*, *Clustering*, *K-Means*, *Industrial Area*

1. PENDAHULUAN

Pengembangan kawasan industri merupakan hal positif bagi pihak industri karena menjadikan perkembangan produktifitas yang menguntungkan namun seringkali dengan adanya kawasan industri, masyarakat sekitar terkena dampak negatif dari keberadaan kawasan industri itu sendiri. Meningkatnya pembangunan industri berkelanjutan, selain memberikan dampak positif terhadap pendapatan daerah dan perekonomian sosial, juga mempunyai sisi negatif terhadap lingkungan dan kehidupan sosial di masa depan [1]. Oleh karena itu dalam pembangunan kawasan industri, kepadatan penduduk merupakan faktor yang perlu dipertimbangkan. Pembagian wilayah yang tidak seimbang menimbulkan masalah pada kepadatan penduduk dan himpitan penduduk di suatu wilayah [2].

Kawasan Industri merupakan wilayah yang memiliki kontradiksi terhadap sektor pariwisata. Pariwisata merupakan segala jenis aktivitas wisata dan didukung dengan segala fasilitas serta layanan yang disediakan oleh masyarakat, pengusaha, pemerintah, maupun pemerintah daerah. Selain itu, pariwisata adalah segala sesuatu yang terkait dengan kepariwisataan, termasuk peningkatan daya tarik wisata, serta kegiatan lain yang berhubungan dengan sektor pariwisata, yang dimaksud dengan sasaran wisata adalah objek dan daya tarik wisata [3].

Selain mempertimbangkan kepadatan penduduk dan sektor pariwisata dalam pembangunan kawasan industri, transportasi menjadi faktor yang sangat penting untuk memberikan efisiensi dalam aktifitas pendistribusian. Transportasi adalah kegiatan pemindahan barang (muatan) dan penumpang dari suatu tempat ke tempat lain, transportasi juga dapat dijelaskan sebagai usaha dan kegiatan mengangkut atau membawa barang dan/atau penumpang dari tempat asal ke tempat tujuan untuk keperluan tertentu, dengan alat tertentu pula [4]. Transportasi memegang peranan sangat penting dalam dunia industri, dari proses mendatangkan bahan baku produk sampai proses distribusi produk menuju konsumen, seluruh kegiatan ini memerlukan transportasi sebagai perantaranya. Sehingga dapat dikatakan bahwa transportasi itu sendiri adalah bagian yang tidak terpisahkan dalam aktifitas industri [4]. Perkembangan terhadap dunia industri khususnya pada sektor transportasi, sangat diperlukan pertimbangan yang sesuai dalam mengambil keputusan jenis transportasi yang akan digunakan, baik dari segi jenis barang, waktu pengiriman, maupun biaya. Biaya transportasi bisa mencapai angka 60% dari total biaya logistik perusahaan, oleh karena itu pengelolaan transportasi dalam proses produksi harus diupayakan seefektif dan seefisien mungkin [4]. Beberapa penelitian menemukan dampak positif dari transportasi jalan terhadap perekonomian, khususnya terhadap Produk Domestik Bruto (PDB) suatu wilayah. Terkait infrastruktur transportasi pelabuhan dan bandar udara, beberapa studi menunjukkan adanya dampak yang positif terhadap pertumbuhan ekonomi [5]. pengelolaan sumber daya dan infrastruktur industri merupakan faktor kunci keberhasilan dalam mendorong pertumbuhan industri. Tingkat pertumbuhan industri di suatu daerah akan bergantung pada hubungan internal dan terkait serta sinergi antara berbagai sektor [6].

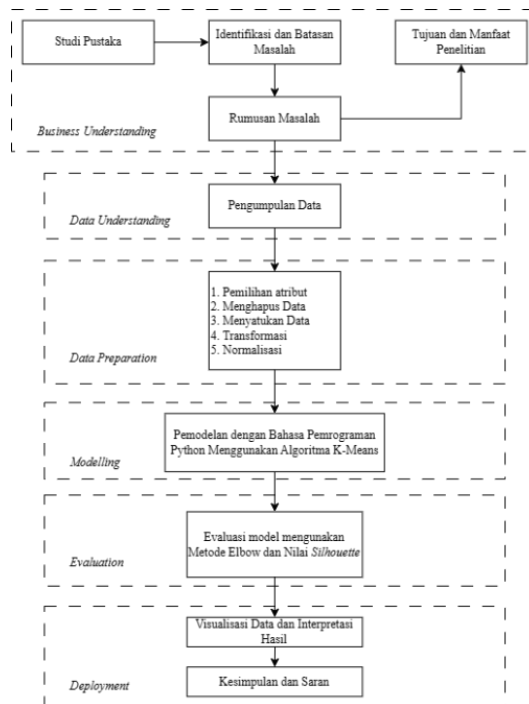
Selain berpengaruh terhadap pendistribusian, transportasi juga menjadi bagian penting terhadap produktivitas tenaga kerja. kekurangan Infrastruktur transportasi menjadi salah satu faktor yang menyebabkan produktivitas tenaga kerja Indonesia tertinggal dari beberapa negara Asia Tenggara lainnya, seperti Singapura, Malaysia, dan Thailand [5]. tenaga kerja merupakan hal yang krusial dalam kegiatan usaha suatu daerah. sebab energi kerja berguna buat mengelola asal daya yang ada buat dihasilkan suatu barang serta jasa. Semakin berkualitas energi kerja tersebut maka akan membentuk produk serta jasa lebih efektif serta efisien [7].

Pemerintah daerah sampai saat ini belum mempunyai solusi yang saling menguntungkan terhadap permasalahan zona industri yang berada di kawasan pemukiman padat penduduk [8]. Oleh karena itu tujuan dari penelitian ini adalah untuk mendapatkan informasi berupa klaster-klaster provinsi berdasarkan tingkat kelayakan dalam pembangunan kawasan industri menggunakan pendekatan *data mining* yaitu metode *clustering* dengan algoritma *K-Means*. *Data mining* merupakan pembelajaran berbasis induksi (*induction based learning*) adalah proses pembentukan definisi-definisi konsep pada umum yang dilakukan dengan cara mengobservasi contoh-contoh spesifik dari konsep yang akan dipelajari [9].

2. METODE PENELITIAN

2.1 *Cross Industry Standard Process for Data mining (CRISP-DM)*

Metode *CRISP-DM* merupakan proses *data mining* yang sudah populer di lingkup data science karena telah menjadi standar industri. Pada tahun 2020 di Indonesia, *CRISP-DM* menjadi acuan Standar Kompetensi Kerja Nasional berdasarkan Keputusan Menteri Ketenagakerjaan No. 299 tahun 2020 [10]. Pada Gambar 1 digambarkan tahapan atau proses penelitian yang diterapkan sebagai berikut :



Gambar 1. Tahapan atau Proses Penelitian

- a. *Business Understanding*
Business understanding yaitu tahap pemahaman tujuan, kebutuhan, batasan dan sudut pandang bisnis, kemudian menjadikan pemahaman tersebut kedalam definisi masalah, menentukan strategi yang akan dicapai pada *data mining* [10].
- b. *Data Understanding*
Data understanding yaitu tahapan mengumpulkan data, pemahaman data, mendeskripsikan data, mengidentifikasi data, dan mengevaluasi kualitas data yang akan digunakan untuk proses selanjutnya [10].
- c. *Data Preparation*
Proses data *preparation* yaitu pengolahan data mentah menjadi *dataset* akhir. Ada beberapa hal yang dilakukan seperti pembersihan data, pemilihan data, memeriksa atribut-atribut, memilih variabel yang sesuai untuk dianalisis, memeriksa distribusi data serta melakukan transformasi terhadap data [10].
- d. *Modelling*
Pada tahap *modelling*, yaitu menerapkan teknik pemodelan yang sesuai, mengidentifikasi dan menampilkan pola. Pada tahap *modelling* yang telah menggunakan *machine learning*, yaitu membagi data dan menentukan algoritma yang cocok untuk data serta mengkalibrasi aturan model agar mendapatkan hasil yang optimal [10].
- e. *Evaluation*
Tahap dalam proses ini yaitu melakukan evaluasi terhadap satu atau lebih model yang akan digunakan, memastikan permasalahan penting diatasi dengan baik, memastikan dan melakukan perhitungan model sesuai dengan tahap awal, melihat pola yang di hasilkan dari algoritma, melihat parameter yang akan digunakan untuk dilakukan evaluasi komparasi algoritma [10].
- f. *Deployment*
Tahap *deployment* yaitu membuat laporan hasil dari seluruh data yang diolah, dikembangkan dan divisualisasikan [10].

2.2 K-Means Clustering

Algoritma *K-Means* melakukan pembagian data menjadi beberapa kelompok sehingga data yang memiliki karakteristik yang sama berada pada satu *cluster* dan data dengan karakteristik yang berbeda berada pada *cluster* yang berbeda [11]. Secara lebih spesifik, algoritma *K-Means* dapat dijelaskan sebagai berikut [11]:

- a. Menetapkan *K* sebagai jumlah *cluster* yang optimal.
- b. Menentukan atau menghasilkan nilai *random* untuk pusat *cluster* awal (*centroid*) sebanyak *k*.

- c. Menggunakan rumus jarak setiap data *input* terhadap masing-masing *centroid* menggunakan rumus jarak *Euclidean (Euclidean Distance)* hingga ditemukan jarak yang paling pendek dari setiap data dengan *centroid*. persamaan *Euclidean Distance* antara lain :

$$d(x_i, \mu_j) = \sqrt{\sum (x_i - \mu_j)^2} \quad (1)$$

Keterangan:

x_i : data kriteria

μ_j : *centroid* pada *cluster* ke- j

- d. Mengklasifikasikan setiap data berdasarkan kedekatannya dengan *centroid* (jarak terdekat). Memperbaharui nilai *centroid*. Menurut Rahman *et al.* (2017) dalam Amalina, Pramana dan Sari (2022) nilai *centroid* baru diperoleh dari rata-rata *cluster* yang bersangkutan dengan menggunakan rumus :

$$C_k = \frac{1}{n_k} \sum d_i \quad (2)$$

Keterangan:

n_k : Jumlah data dalam *cluster* k

d_i :: Jumlah nilai jarak yang masuk dalam masing-masing *cluster*.

- e. Melakukan perulangan dari langkah 2 hingga 4 sampai anggota tiap *cluster* tidak ada yang berubah.
 f. Jika langkah terakhir telah terpenuhi, maka nilai pusat *cluster* (μ_j) pada iterasi terakhir akan digunakan sebagai parameter untuk menentukan klasifikasi data.

3. HASIL DAN PEMBAHASAN

3.1 *Business Understanding*

Dapat dipahami permasalahan dari penelitian ini berdasarkan penjelasan sebagai berikut:

- a. Identifikasi Masalah

Berdasarkan apa yang sudah dijelaskan di pendahuluan, dapat diuraikan permasalahan dalam penelitian ini, diantaranya yaitu :

1. Pentingnya mempertimbangkan Kepadatan Penduduk dalam penentuan wilayah pembangunan, khususnya terhadap kawasan industri.
2. Keberadaan kawasan industri yang kontradiksi terhadap sektor pariwisata.
3. Infrastruktur dan moda transportasi yang menjadi bagian penting dalam dunia industri.
4. Pentingnya tenaga kerja berkualitas dalam pengembangan atau kegiatan operasional.

Berdasarkan keempat permasalahan di atas maka belum diketahui tingkat kelayakan provinsi dalam pembangunan Kawasan Peruntukan Industri (KPI).

- b. Batasan Masalah

Supaya cakupan penelitian tidak meluas yang mengakibatkan substansi menjadi tidak jelas dan supaya hasil yang disajikan lebih optimal maka penelitian ini dibatasi sebagai berikut:

1. Menggunakan data populasi penduduk menurut provinsi pada tahun 2022 dan dilakukan pembagian dengan data luas provinsi yang bersumber dari Badan Pusat Statistik (BPS).
2. Menggunakan data jumlah pelabuhan menurut provinsi yang bersumber dari Sistem Informasi Pelabuhan, Direktorat Kepelabuhanan.
3. Menggunakan data jumlah bandara menurut provinsi pada tahun 2022 yang bersumber dari Portal Data Kementerian Perhubungan.
4. Berdasarkan poin 2 dan 3 di atas, maka moda transportasi yang dipertimbangkan yaitu kapal laut dan pesawat.
5. Menggunakan data panjang jalan nasional pada tahun 2022 untuk moda transportasi truk yang bersumber dari BPS.
6. Terkait tenaga kerja maka data yang digunakan yaitu data Jumlah Angkatan Kerja dan Tingkat Pendidikan SMA Sederajat pada tahun 2022 yang bersumber dari BPS.
7. Menggunakan data Jumlah Wisatawan Mancanegara pada tahun 2022 yang bersumber dari Kementerian Pariwisata dan Ekonomi Kreatif.
8. Menggunakan 32 provinsi karena 4 provinsi baru dan 2 provinsi yang berkaitan dengan provinsi baru, data yang diperlukan dalam penelitian ini belum tersedia secara resmi.

- c. Rumusan Masalah

Ditentukan rumusan masalah yaitu dengan pendekatan *data mining* dapat menggunakan metode *clustering* dengan algoritma *K-Means* untuk melakukan klusterisasi provinsi menurut tingkat kelayakan

dalam pembangunan kawasan industri berdasarkan Data Kepadatan Penduduk, Jumlah Bandara, Jumlah Pelabuhan, Panjang Jalan Negara/ Nasional, Jumlah Angkatan Kerja, Tingkat Pendidikan SMA Sederajat, dan Jumlah Wisatawan Mancanegara.

Perlu dipahami bahwa hasil dari penelitian ini dapat digunakan oleh pihak pemerintah Indonesia dalam mengoptimalkan evaluasi Rencana Tata Ruang Wilayah (RTRW) khususnya pada sektor Kawasan Peruntukan Industri (KPI) atau bermanfaat untuk pihak mitra industri dalam menentukan provinsi yang akan dijadikan lokasi pembangunan.

3.2 Data Understanding

Data yang diperlukan berasal dari beberapa situs resmi, data yang didapatkan antara lain :

- Kepadatan penduduk tahun 2022 menurut provinsi, data ini diperoleh dari perhitungan jumlah populasi dibagi oleh luas wilayah, data didapatkan melalui situs BPS. Data ini dibutuhkan karena pertimbangan kepadatan penduduk.
- Jumlah Wisatawan Mancanegara Tahun 2022 menurut provinsi, data ini didapatkan melalui situs Kementerian Pariwisata dan Ekonomi Kreatif. Data ini dibutuhkan karena terdapat kontradiksi antara keberadaan wilayah pariwisata dengan keberadaan kawasan industri.
- Rasio Panjang Jalan Nasional tahun 2022 menurut provinsi, data ini diperoleh dari perhitungan luas wilayah dibagi oleh panjang jalan, data didapatkan melalui situs BPS dan khusus data panjang jalan tol dilakukan pengukuran secara mandiri melalui Google Maps. Data ini dibutuhkan karena pertimbangan moda transportasi truk yang menjadi transportasi jalur darat.
- Jumlah Bandara tahun 2022 menurut provinsi, data ini didapatkan melalui situs Portal Data Kementerian Perhubungan. Data ini dibutuhkan karena pertimbangan moda transportasi pesawat yang menjadi transportasi jalur udara.
- Jumlah Pelabuhan tahun 2022 menurut provinsi, data ini didapatkan melalui situs Sistem Informasi Pelabuhan Direktorat Kepelabuhanan. Data ini dibutuhkan karena pertimbangan moda transportasi Kapal Laut yang menjadi transportasi jalur laut.
- Jumlah Angkatan Kerja tahun 2022 menurut provinsi, data ini didapatkan melalui situs BPS. Data ini dibutuhkan karena pentingnya keseimbangan dalam penyerapan jumlah angkatan kerja.
- Tingkat Lulusan SMA Sederajat tahun 2022 menurut provinsi, data ini didapatkan melalui situs BPS. Data ini dibutuhkan karena pentingnya pemilihan kualitas tenaga kerja.

3.3 Data Preparation

3.3.1 Dataset Asli

Pada tahap ini dilakukan pembentukan *dataset* dengan menyatukan data/ atribut yang dipilih yaitu atribut Kepadatan Penduduk, Rasio Panjang Jalan Nasional, Jumlah Pelabuhan, Jumlah Bandara, Jumlah Angkatan Kerja, Tingkat Lulusan SMA Sederajat dan Jumlah Wisman yang disatukan berdasarkan Nama Provinsi. Susunan *Dataset* dapat dilihat pada Tabel 1 Berikut :

Tabel 1. *Dataset Asli (sample data)*

Provinsi	Kepadatan Penduduk	Rasio Panjang Jalan Nasional	Jumlah Pelabuhan	Jumlah Bandara	Jumlah AK	Tingkat Lulusan SMA Sederajat	Jumlah Wisman
Aceh	93,31044	26,98301	7	11	2552562	70,67	1886
Sumatera Utara	207,1108	26,49739	10	7	7669870	77,16	74498
Sumatera Barat	134,2588	29,52417	4	3	2868270	65,96	4142
Riau	76,00692	61,43222	8	6	3313818	66,91	3985
Jambi	72,53762	37,9516	3	3	1884278	65,85	0

3.3.2 Dataset Transformasi

Pada tahap ini dilakukan perubahan data terhadap angka positif menjadi angka negatif yang terdapat pada atribut Kepadatan Penduduk, Rasio Panjang Jalan Nasional dan Jumlah Wisman perubahan data ini perlu dilakukan karena setiap atribut harus memiliki karakter data yang sama yaitu angka semakin besar maka angka tersebut menjadi lebih baik. Perubahan data dapat dilihat pada Tabel 2 berikut :

Tabel 2. *Dataset Transformasi (sample data)*

Provinsi	Kepadatan Penduduk	Rasio Panjang Jalan Nasional	Jumlah Pelabuhan	Jumlah Bandara	Jumlah AK	Tingkat Lulusan SMA Sederajat	Jumlah Wisman
Aceh	-93,31044	-26,98301	7	11	2552562	70,67	-1886
Sumatera Utara	-207,1108	-26,49739	10	7	7669870	77,16	-74498
Sumatera Barat	-134,2588	-29,52417	4	3	2868270	65,96	-4142
Riau	-76,00692	-61,43222	8	6	3313818	66,91	-3985
Jambi	-72,53762	-37,9516	3	3	1884278	65,85	0

3.3.3 Dataset Normalisasi

Tahap ini merupakan tahap terakhir dalam *Data Preparation* yaitu melakukan normalisasi *dataset* dengan merubah data ke dalam angka yang lebih kecil yaitu angka 0 sampai dengan 1 sehingga bobot angka pada masing-masing data menjadi lebih terukur dan klasterisasi memiliki dimensi yang ideal. Tahap ini perlu dilakukan supaya proses klasterisasi mampu menghasilkan pengelompokan data yang optimal. Hasil normalisasi data dapat dilihat pada Tabel 3 berikut :

Tabel 3. *Dataset Normalisasi (sample data)*

Provinsi	Kepadatan Penduduk	Rasio Panjang Jalan Nasional	Jumlah Pelabuhan	Jumlah Bandara	Jumlah AK	Tingkat Lulusan SMA Sederajat	Jumlah Wisman
Aceh	0,994795	0,782169	0,466667	0,769231	0,086581	0,651163	0,999125
Sumatera Utara	0,987715	0,78638	0,666667	0,461538	0,289583	0,782406	0,965442
Sumatera Barat	0,992248	0,760134	0,266667	0,153846	0,099105	0,555915	0,998079
Riau	0,995872	0,483452	0,533333	0,384615	0,11678	0,575126	0,998151
Jambi	0,996087	0,687058	0,2	0,153846	0,06007	0,553691	1

3.4 Modelling

Pada tahap ini dilakukan proses untuk mendapatkan model terbaik. Proses komparasi model dilakukan dalam perbandingan tingkat akurasi yang dihasilkan oleh beberapa algoritma *data mining* supaya dapat menentukan algoritma terbaik, algoritma yang digunakan sebagai perbandingan yaitu algoritma *K-Means*, algoritma *Agglomerative*, algoritma *BIRCH* dan algoritma *Spectral*. Proses komparasi model dilakukan dengan menggunakan library Bahasa Pemrograman Python yaitu library Scikit Learn. Berikut pada Tabel 4 merupakan hasil dari proses algoritma *K-Means* yang telah dibandingkan nilai *Davies Bouldin Index (DBI)* pada perbedaan jumlah *cluster*:

Tabel 4. *Davies Bouldin Index Algoritma K-Means*

Jumlah Cluster	Davies Bouldin Index
2 Cluster	1,338152
3 Cluster	0,992594
4 Cluster	0,806343
5 Cluster	0,676747
6 Cluster	0,585901
7 Cluster	0,826347
8 Cluster	1,026165
9 Cluster	0,939513

Data pada Tabel 4 merupakan angka *DBI* yang dapat dibandingkan dengan hasil evaluasi menggunakan metode *Elbow* dan *Silhouette Score*. Berdasarkan hasil evaluasi menggunakan metode *Elbow* pada Gambar 2 yaitu grafik yang membuat sudut siku pada 6 *cluster* dan hasil evaluasi menggunakan perhitungan *Silhouette Score* pada Tabel 6 menunjukkan bahwa jumlah *cluster* 6 yang memiliki nilai tertinggi.

Oleh karena itu *cluster* berjumlah 6 dengan angka *DBI* 0,585901 menjadi yang terbaik pada algoritma *K-Means* karena pada grafik *Elbow*, nilai *DBI* dan hasil *Silhouette Score* menunjukkan jumlah *cluster* terbaik pada jumlah yang sama, yaitu 6 *Cluster*.

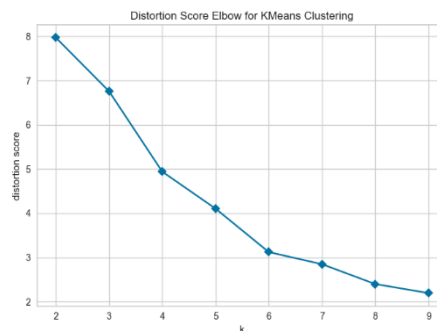
Tabel 5. Hasil Komparasi Model Pada setiap Algoritma

Algoritma	Jumlah Cluster	Davies Bouldin Index
<i>K-Means</i>	6 Cluster	0,585901
<i>Agglomerative</i>	7 Cluster	0,694578
<i>BIRCH</i>	7 Cluster	0,694578
<i>Spectral</i>	8 Cluster	0,806401

Dapat dilihat pada Tabel 5, algoritma terbaik adalah algoritma *K-Means* karena ditentukan berdasarkan model yang memiliki angka *Davies Bouldin Index* paling rendah terhadap hasil evaluasi yaitu perbandingan antara jumlah cluster terpilih pada metode *Elbow* dan *Silhouette Score*.

3.5 Evaluation

Untuk memaksimalkan proses evaluasi dalam pemilihan jumlah cluster pada masing-masing algoritma yang menjadi pertimbangan dalam proses perbandingan maka selain menggunakan metode *Elbow* tahap evaluasi juga menggunakan perhitungan *Silhouette Score*.


Gambar 2. Metode *Elbow* Algoritma *K-Means*

Gambar 2 adalah metode *Elbow* dari algoritma *K-Means*, visualisasi grafik yang menunjukkan perbedaan angka *Distortion score* terhadap jumlah cluster yang berbeda. Bisa dilihat bahwa dari jumlah cluster 5 menuju ke 6 angka *Distortion score* mengalami penurunan yang signifikan dan dari jumlah cluster 6 menuju ke 7 mengalami penurunan yang tidak signifikan. Oleh karena itu jumlah cluster 6 memiliki sudut yang membentuk siku. Namun jika dilihat secara keseluruhan sudut yang terbentuk pada cluster 6 tidak signifikan, oleh karena itu pemilihan cluster pada algoritma *K-Means* perlu dilakukan perbandingan terhadap hasil *Silhouette Score* dan *DBI*.

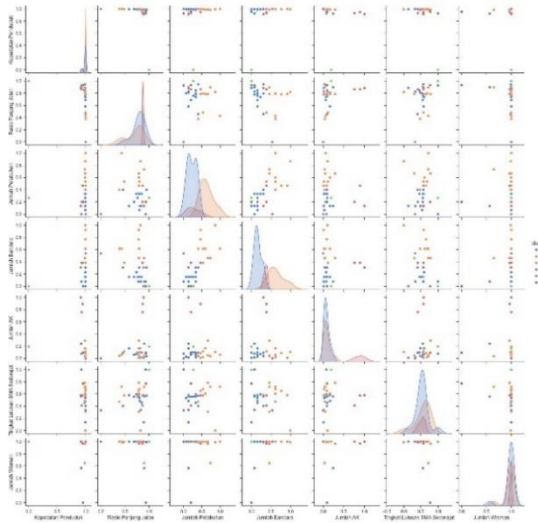
Tabel 6. *Silhouette Score* Algoritma *K-Means*

Jumlah Cluster	<i>Silhouette Score</i>
2 Cluster	0,252628
3 Cluster	0,263769
4 Cluster	0,345877
5 Cluster	0,354598
6 Cluster	0,358952
7 Cluster	0,207272
8 Cluster	0,178108
9 Cluster	0,184359

Berdasarkan data pada Tabel 6 maka angka *Silhouette Score* paling besar yaitu 0,358952 yang menandakan cluster berjumlah 6 menjadi yang terbaik dalam perhitungan *Silhouette Score* terhadap algoritma *K-Means*.

3.6 Deployment

Pada tahap ini dijelaskan informasi terkait hasil klusterisasi yang akan divisualisasikan ke dalam bentuk grafik seperti informasi pemetaan cluster pada setiap atribut, tingkat kelayakan cluster dan jarak data pada centroid yang menjadi indikator tingkat kelayakan pada masing-masing data.



Gambar 3. Grafik *Scatterplot Matrix*

Pada Gambar 3 dapat dilihat grafik pemetaan dari keenam *cluster* pada masing-masing atribut. Cara membaca grafik tersebut yaitu dengan melihat satu per satu dari atribut yang terlibat pada sumbu x dan sumbu y.

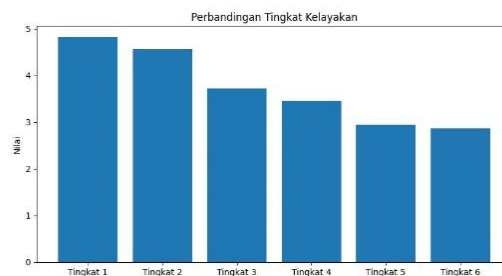
Tabel 7. Data *Centroid* dan Tingkat Kelayakan *Cluster*

<i>Cluster</i>	Kepadatan Penduduk	Rasio Panjang Jalan Nasional	Jumlah Pelabuhan	Jumlah Bandara	Jumlah AK	Tingkat Lulusan SMA Sederajat	Jumlah Wisman	Total <i>Centroid</i>	Kelayakan
0	0,983492	0,782283	0,222222	0,158974	0,081018	0,525891	0,968778	3,722659	Tingkat 3
1	0,994206	0,703108	0,630303	0,629371	0,088997	0,553507	0,963642	4,563133	Tingkat 2
2	0	1	0,266667	0	0,193683	0,995753	1	3,456103	Tingkat 4
3	0,930378	0,870586	0,266667	0,358974	0,883472	0,520795	0,989504	4,820376	Tingkat 1
4	1	0	0	0,538462	0	0,330233	1	2,868694	Tingkat 6
5	0,95308	0,932621	0,2	0	0,093959	0,77088	0	2,950539	Tingkat 5

Tabel 7 merupakan data *centroid* pada setiap *cluster* dan setiap atribut. Berdasarkan data tersebut maka dilakukan penjumlahan seluruh *centroid* pada masing-masing *cluster* supaya mengetahui tingkat kelayakan pada setiap *cluster*.

Karakter data pada setiap atribut yaitu angka semakin besar maka menjadi semakin baik, oleh karena itu penjumlahan *centroid* yang ada pada kolom *Total Centroid* merupakan indikasi tingkat kelayakan pada setiap *cluster*, angka *Total* semakin besar maka *cluster* tersebut semakin baik.

Berdasarkan penjelasan di atas maka *cluster* 3 merupakan *cluster* terbaik (kelayakan tingkat 1) dan *cluster* 4 merupakan *cluster* terburuk (kelayakan tingkat 6).

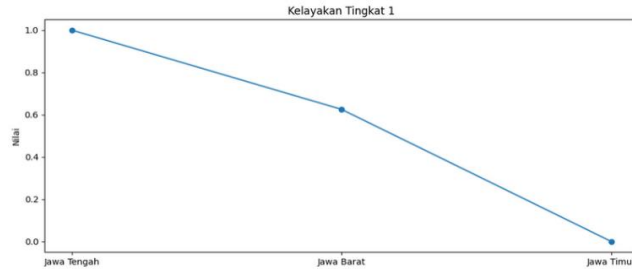


Gambar 4. Perbandingan Tingkat Kelayakan *Cluster*

Pada Gambar 4 merupakan grafik terkait perbandingan Tingkat Kelayakan *Cluster*. Dari grafik tersebut dapat disimpulkan bahwa semua *cluster* layak untuk kawasan industri, namun dari tingkat kelayakan dapat dilihat bahwa perbandingannya membantu dalam menentukan hasil evaluasi Rencana Tata Ruang Wilayah (RTRW) bagi pemerintah provinsi atau membantu dalam pemilihan Kawasan Peruntukan Industri (KPI) bagi pihak mitra.

Berikut visualisasi data terkait perbandingan pada setiap data di masing-masing *cluster* :

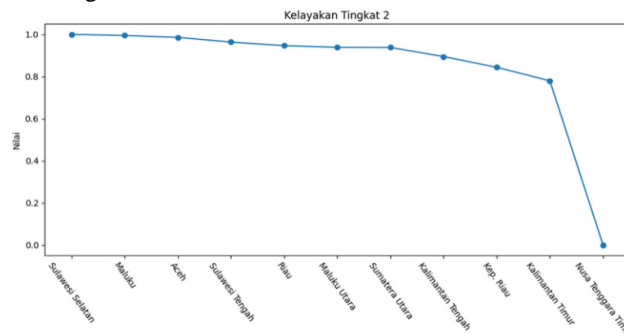
a. Visualisasi Data Kelayakan Tingkat 1



Gambar 5. Grafik Kelayakan Tingkat 1

Pada Gambar 5 dapat dilihat Grafik Kelayakan Tingkat 1 yang menunjukkan bahwa nilai yang terdapat pada sumbu y memiliki skala angka dari 0 sampai 1 yang menandakan data terbaik memiliki nilai 1 dan yang terburuk memiliki nilai 0. Pada grafik tersebut terdapat 3 Provinsi yang dimana Jawa Tengah menjadi yang terbaik dan Jawa Timur menjadi yang terburuk.

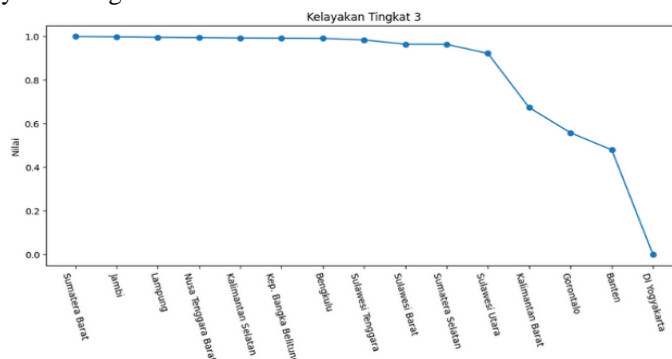
b. Visualisasi Data Kelayakan Tingkat 2



Gambar 6. Grafik Kelayakan Tingkat 2

Pada Gambar 6 dapat dilihat Grafik Kelayakan Tingkat 2 yang menunjukkan bahwa nilai yang terdapat pada sumbu y memiliki skala angka dari 0 sampai 1 yang menandakan data terbaik memiliki nilai 1 dan yang terburuk memiliki nilai 0. Pada grafik tersebut terdapat 11 Provinsi yang dimana Sulawesi Selatan menjadi yang terbaik dan Nusa Tenggara Timur menjadi yang terburuk.

c. Visualisasi Data Kelayakan Tingkat 3



Gambar 7. Grafik Kelayakan Tingkat 3

Pada Gambar 7 dapat dilihat Grafik Kelayakan Tingkat 3 yang menunjukkan bahwa nilai yang terdapat pada sumbu y memiliki skala angka dari 0 sampai 1 yang menandakan data terbaik memiliki nilai 1 dan yang terburuk memiliki nilai 0. Pada grafik tersebut terdapat 15 Provinsi yang dimana Sumatera Barat menjadi yang terbaik dan DI Yogyakarta menjadi yang terburuk.

d. Visualisasi Data Kelayakan Tingkat 4, 5 dan 6

Pada Kelayakan Tingkat 4, 5 dan 6 tidak perlu divisualisasikan dalam bentuk grafik karena hanya terdapat 1 data provinsi yaitu pada kelayakan tingkat 4 terdapat provinsi DKI Jakarta, tingkat 5 terdapat provinsi Bali, dan tingkat 6 terdapat provinsi Kalimantan Utara.

4. KESIMPULAN

Berdasarkan proses dan hasil penelitian dengan pendekatan *CRISP-DM* yang memanfaatkan Bahasa Pemrograman Python dan menggunakan metode *clustering*, maka dapat disimpulkan bahwa pada proses komparasi angka *DBI* antara algoritma *K-Means*, *Agglomerative*, *BIRCH*, dan *Spectral* maka terpilih algoritma *K-Means* dengan 6 *cluster*. Pemilihan jumlah *cluster* pada setiap algoritma menggunakan 3 metode evaluasi yaitu metode *Elbow*, *Silhouette Score*, dan *DBI*. Penelitian ini menunjukkan bahwa 32 provinsi terbagi menjadi 6 *cluster* yang berarti terbentuknya 6 tingkat kelayakan. Keenam tingkat kelayakan terbentuk berdasarkan perhitungan metode *clustering* yang menggunakan 7 atribut. *Cluster* dengan Kelayakan Tingkat 1: 3 Provinsi, Tingkat 2: 11 Provinsi, Tingkat 3: 15 Provinsi, Tingkat 4, 5 dan 6: 1 Provinsi. Hasil penelitian ini dapat direkomendasikan untuk menjadi bagian pertimbangan dalam proses evaluasi Rencana Tata Ruang Wilayah (RTRW) suatu Provinsi khususnya pada sektor Kawasan Peruntukan Industri (KPI). Berdasarkan hasil penelitian ini dapat dilakukan penelitian lebih lanjut dengan melakukan klusterisasi semua Kota/ Kabupaten yang terdapat di semua provinsi dalam tingkat kelayakan yang sama dan dengan pemilihan atribut yang relevan dan dibutuhkan sehingga dapat dijadikan bahan pertimbangan dalam evaluasi RTRW suatu Kota/ Kabupaten khususnya pada sektor KPI.

DAFTAR PUSTAKA

- [1] Y. I. Astuty, M. Noer, D. Stevany, B. Arham, B. M. R, and A. 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.
- [2] D. Gultom, I. Gunawan, I. Purnamasari, S. R. Andani, and Z. A. Siregar, "Penerapan Algoritma *K-Means* Dalam Pengelompokan Kepadatan Penduduk Menurut Kecamatan di Kabupaten Simalungun," vol. 2, no. 10, pp. 622–628, 2022.
- [3] I. M. Hasibuan, S. Mutthaqin, R. Erianto, and I. Harahap, "Kontribusi Sektor Pariwisata Terhadap Perekonomian Nasional," *J. Masharif al-Syariah J. Ekon. dan Perbank. Syariah*, vol. 8, no. 30, pp. 1177–1217, 2023.
- [4] N. P. D. Arwini and I. M. Juniastra, "Peran Transportasi Dalam Dunia Industri," *VASTUWIDYA*, vol. 6, no. 1, pp. 70–77, 2023.
- [5] D. Pratama and Khoirunurrofik, "Peran Infrastruktur Transportasi Terhadap Produktivitas Tenaga Kerja Industri Manufaktur Di Indonesia," *J. Ekon. dan Pembang.*, vol. 31, no. 1, pp. 101–116, 2023.
- [6] N. A. P. Harahap, F. Al Qadri, D. I. Y. Harahap, M. Situmorang, and S. Wulandari, "Analisis Perkembangan Industri Manufaktur Indonesia," *J. Kaji. Ekon. dan Bisnis Islam*, vol. 4, no. 6, pp. 1444–1450, 2023.
- [7] A. A. Pasaribu, A. A. Tanjung, Sukardi, and Paidi, "Analisis Tenaga Kerja Di Sektor Industri Di Indonesia Menggunakan Pendekatan Gmmm," *J. Masharif al-Syariah J. Ekon. dan Perbank. Syariah*, vol. 8, no. 1, pp. 378–390, 2023.
- [8] L. Hakim, E. Rochima, and S. Wyantuti, "Implementasi Kebijakan Dan Realisasi Rencana Tata Ruang Kec. Garut Kota Di Kab. Garut: Studi Analisis Kebijakan," *J. Ekon. Kebijak. Publik*, vol. 12, no. 2, pp. 163–175, 2021.
- [9] I. A. Darmawan, M. F. Randy, I. Yuniyanto, M. M. Mutoffar, and M. T. P. Salis, "Penerapan *Data Mining* Menggunakan Algoritma Apriori Untuk Menentukan Pola Golongan Penyandang Masalah Kesejahteraan Sosial," *Sebatik*, vol. 26, no. 1, pp. 223–230, 2022.
- [10] Y. Yudianta, A. Y. Agustina, and N. Khofifah, "Prediksi Customer Churn Menggunakan Metode *CRISP-DM* Pada Industri Telekomunikasi Sebagai Implementasi Mempertahankan Pelanggan," *Indones. J. Islam. Econ. Bus.*, vol. 8, no. 1, pp. 1–20, 2023.
- [11] T. Amalina, D. B. A. Pramana, and B. N. Sari, "Metode *K-Means Clustering* Dalam Pengelompokan Penjualan Produk Frozen Food," *J. Ilm. Wahana Pendidik.*, vol. 8, no. 15, pp. 574–583, 2022.