

Data Mining Strategi Pembangunan Infrastruktur Menggunakan Algoritma K-Means

Mangapul Siahaan^{[1]*}

Program Studi Sistem Informasi^[1]

Fakultas Ilmu Komputer Universitas Internasional Batam

Batam, Indonesia

mangapul.siahaan@uib.ac.id^[1]

Abstract—Batam City is an industrial city and part of the Riau Islands province. The increase in population will have an impact on the availability of infrastructure, so it is necessary to consider equitable infrastructure development. The purpose of this study is to find a strategy for infrastructure development using the K-Means algorithm to group based on population, population growth, and area. The dataset used is from BPS Batam City. The method used in this research is CRISP-DM. Based on the results of the C1 cluster calculation, there are 3 sub-districts (Belakang Padang, Bulang, and Galang). The results of cluster C2 are 4 sub-districts (Sungai Beduk, Sagulung, Lubuk Baja, and Batu Ampar). The results of the C3 cluster are 3 sub-districts (Sekupang, Batu Aji, and Bengkong). While the calculation of cluster C4 there are 2 sub-districts (Nongsa and Batam City). According to the cluster calculations, development should be prioritized in areas with Clusters C3 and C4 because the status of the population is very high, and the growth rate is very high.

Keywords— Data Mining, K-Means, Infrastructure Development

Abstrak—Kota Batam merupakan kota industri dan bagian dari provinsi Kepulauan Riau. Dengan meningkatnya jumlah penduduk akan berdampak terhadap ketersediaan infrastruktur sehingga perlu dipertimbangkan pembangunan infrastruktur yang merata. Tujuan penelitian ini adalah untuk mencari strategi pembangunan infrastruktur dengan menggunakan algoritma K-Means melakukan pengelompokan berdasarkan jumlah penduduk, pertumbuhan penduduk dan luas daerah. Dataset yang digunakan dari BPS Batam Kota. Metode yang digunakan dalam penelitian ini adalah CRISP-DM. Berdasarkan hasil perhitungan *cluster* C1 terdapat 3 kecamatan (Belakang Padang, Bulang dan Galang). Hasil *cluster* C2 terdapat 4 kecamatan (Sungai Beduk, Sagulung, Lubuk Baja dan Batu Ampar). Hasil *cluster* C3 terdapat 3 kecamatan (Sekupang, Batu Aji, dan Bengkong). Sedangkan perhitungan *cluster* C4 terdapat 2 kecamatan (Nongsa dan Batam Kota). Menurut perhitungan *cluster* tersebut sebaiknya pembangunan diutamakan di daerah dengan *Cluster* C3 dan C4 karena status jumlah penduduk yang sangat tinggi, laju pertumbuhan yang sangat tinggi.

Kata Kunci—Data Mining, K-Means, Pembangunan Infrastruktur

I. PENDAHULUAN

Kota Batam merupakan kota bagian dari provinsi Kepulauan Riau yang memiliki luas daratan sebesar 715 km², sedangkan luas wilayah secara keseluruhan 1.575 km². Dengan didukungnya dari pemerintah pusat serta kebijakan kebijakan dari pemerintah, pada tahun 1970 wilayah Batam bertransformasi menjadi Kota Industri. Semenjak ditetapkan Kota Industri para investor berdatangan melakukan investasi untuk menggerakkan ekonomi Batam. Adapun industri yang ada di Kota Batam seperti industri Oil Gas, industri perkapalan, industri elektronik, industri pariwisata, dll.

Kota Batam sendiri penyumbang terbesar dalam pendapatan domestic regional bruto atau dikanal PDRB untuk wilayah Kepulauan Riau, peningkatan pendapatan ini didukung dengan kenaikan nilai ekspor sebesar 10.85%. Efek meningkatnya nilai investasi di Kota Batam berpengaruh signifikan terhadap meluasnya jenis lapangan usaha. Beberapa lapangan usaha yang meningkat pada tahun 2021 yakni dibidang informasi dan komunikasi 10.39% dan bidang jasa kesehatan sebanyak 7.65%. Dengan bertumbuhnya lapangan usaha akan berdampak signifikan terhadap perkembangan wilayah dan masyarakat sehingga tersedianya lapangan pekerjaan bagi masyarakat Kota Batam dan membuat Kota Batam menjadi kawasan yang lebih modern. Pendapatan yang diperoleh dari industri-industri di Kota Batam dipergunakan untuk pembangunan infrastruktur sehingga meningkat kan kesejahteraan masyarakat, perbaikan dan meningkatkan taraf hidup[2].

Dengan berkembangnya industri dan meluasnya lapangan pekerjaan di Kota Batam tidak dipungkiri banyak nya perpindahan penduduk dari suatu Kota menuju ke Kota Batam dan melakukan urbanisasi demi mendapatkan pekerjaan. Perpindahan penduduk tersebut mengakibatkan bertambahnya jumlah penduduk, lajunya pertumbuhan penduduk serta berkurangnya luas daerah. Bertambahnya jumlah penduduk, laju pertumbuhan akan mempengaruhi terhadap infrastruktur yang tersedia di daerah tersebut, seperti dibutuhkan infrastruktur tambahan yakni sekolah, tempat hiburan, tempat ibadah, infrastruktur jalan. Dampak kekurangan infrastruktur dapat menimbulkan lambatnya pertumbuhan ekonomi dan tenaga kerja. Oleh sebab itu infrastruktur kesehatan, infrastruktur pendidikan dan jumlah penduduk berpengaruh

signifikan terhadap pertumbuhan ekonomi[4].

Kemacetan transportasi lalu lintas sering menjadi isu di daerah perkotaan, hal ini karena banyaknya kendaraan menumpuk disuatu jalan sehingga perlu didukung pengembangan infrastruktur jalan seperti penambahan dan pelebaran jalan dan perluasan area jalan. Bertambahnya jumlah penduduk akan berdampak signifikan terhadap pertumbuhan infrastruktur. Semakin tinggi jumlah penduduk maka semakin banyak ketersediaan infrastruktur[1].

Data mining adalah proses menggunakan teknik statistik, matematika, kecerdasan buatan dan *machine learning* untuk mengidentifikasi informasi yang bermanfaat buat pengetahuan[5]. Data mining merupakan cara menemukan informasi baru dengan melihat pola atau aturan tertentu dari jumlah data dan dapat mengatasi masalah atau kondisi pembangunan infrastruktur secara merata[3]. Dengan penerapan data mining menggunakan data dari Badan Pusat Statistik (BPS) Batam Kota dengan menggunakan variabel jumlah penduduk, laju pertumbuhan dan luas daerah akan membuat *cluster* atau data yang berkelompok sehingga bisa merancang strategi pembangunan infrastruktur di Kota Batam.

Metode yang digunakan dalam *clustering* pengelompokan pertumbuhan penduduk menggunakan metode CRISP-DM dengan memahami proses *business understanding, data understanding, data preparation, modeling evaluation serta deployment*. Algoritma yang digunakan pada *clustering* ini menggunakan K-Means, algoritma ini dipilih karena dengan menggunakan konsep data mining, pengelompokan data berdasarkan kemiripan data maksimum pada satu *cluster*, sehingga didapat hasil pengelompokan data. Berdasarkan pengelompokan data yang dilakukan bisa dirancang strategi pembangunan infrastruktur di Kota Batam, daerah mana jumlah penduduk yang padat, pertumbuhan penduduk tinggi dan luas daerah yang luas, sehingga pemerintah bisa memprioritaskan pengembangan daerah berdasarkan hasil *clustering* K-Means tersebut[6].

Data mining menggunakan algoritma K-Means *clustering* untuk mengetahui potensi penyebaran virus corona di Kota Cirebon, dari hasil penelitian ini didapat 3 *cluster* dengan tingkat penyebaran rendah (C1) dengan jumlah 13 kelurahan, tingkat penyebaran sedang(C2) dengan jumlah 5 kelurahan, kemudian tingkat penyebaran tinggi(C3) dengan jumlah 4 kelurahan [7]. Penerapan algoritma K-Means *clustering* untuk pemetaan kepadatan penduduk berdasarkan jumlah Penduduk Kota Medan, dari hasil penelitian didapat 3 *cluster* dengan jumlah penduduk rendah (C1), sedang (C2) dan padat (C3), dimana penduduk Kota Medan yang padat berjumlah 121 kelurahan dan daerah penduduk sangat padat 30 kelurahan[6]. Penerapan algoritma K-Means dalam pengelompokan jumlah penduduk berdasarkan kelurahan di Kota Pematang Siantar, dari hasil penelitian tersebut didapat 3 *cluster* yaitu dengan jumlah penduduk rendah C1 29 kelurahan, jumlah penduduk sedang terdapat 18 kelurahan, dan 6 kelurahan dengan jumlah penduduk tinggi[11]

Dari penjabaran diatas didapat hasil bahwa pengukuran jumlah penduduk, luas daerah, dan laju pertumbuhan dapat dilakukan dengan menggunakan algoritma K-Means sesuai

dengan permasalahan yang ada. Penelitian ini bertujuan untuk menganalisa pemerataan pembangunan infrastruktur di Kota Batam. Berdasarkan penelitian sebelumnya parameter yang digunakan hanya berdasarkan jumlah penduduk, sedangkan pada penelitian ini menambah parameter seperti luas daerah dan lajunya pertumbuhan penduduk di Kota Batam.

II. TINJAUAN PUSTAKA

A. Penelitian terkait

Beberapa penelitian yang pernah dilakukan sebelumnya yaitu Penerapan Algoritma K-Means *Clustering* Untuk Pemetaan Kepadatan Penduduk Berdasarkan Jumlah Penduduk Kota Medan. Pada penelitian ini melakukan pemetaan kedalam 3 kelompok *cluster* C1 = 21 kecamatan, didapat hasil *cluster* pertama dengan penduduk yang sangat padat, C2 = 10 kecamatan dengan jumlah penduduk yang padat, sedangkan *cluster* C3 tidak ditemukan kecamatan yang jumlah penduduk sedang[6].

Penelitian berikutnya yaitu Dampak Pertumbuhan Penduduk Terhadap Infrastruktur Di Kota Semarang. Hasil dari penelitian ini didapat kesimpulan, dengan bertambahnya jumlah penduduk akan berpengaruh signifikan terhadap pembangunan infrastruktur, perlu memperhatikan ketersediaan infrastruktur bagi daerah-daerah yang kepadatan jumlah penduduk masih rendah agar tidak menimbulkan masalah dikemudian hari seperti kemacetan Kota [1].

B. Data Mining

Data Mining merupakan kegiatan dalam mengumpulkan data, pemakaian data historis, untuk mendapatkan keteraturan, serta pola hubungan dalam himpunan data yang besar. Data Mining sering disebut juga dengan *Knowledge Discovery in Database* atau KDD[7].

Data Mining bekerja dengan mencari pola-pola yang menarik pada data yang digunakan dengan menggunakan teknik atau metode tertentu. Teknik atau algoritma metode yang digunakan bervariasi tergantung dengan tujuan yang akan digunakan.

C. Clustering

Clustering merupakan salah satu Teknik dalam Data Mining. Algoritma *clustering* sering disebut juga pengelompokan sejumlah data menjadi kelompok-kelompok data tertentu[7]. Proses pengelompokan data berdasarkan pada kemiripan data yang maksimum pada satu *cluster* dan kemiripan data yang minimum pada *cluster* yang lain[10].

D. K-Means

K-means sendiri merupakan salah satu algoritma dalam melakukan pengelompokan atau *clustering*. K-means bagian dari *Unsupervised Learning* yang berfungsi membagikan data menjadi beberapa kelompok dengan system partisi. Algoritma ini menerima masukan berupa data label. Di setiap *cluster*

didapat titik pusat atau disebut centroid yang mempresentasikan cluster tersebut. secara sederhana algoritma K-Means digunakan untuk proses pengelompokan[7]. Pengelompokan data mining dengan K-Means menggunakan partisi berbasis titik dengan proses komputasi dilakukan cepat dan efisien[9].

Langkah-langkah dalam melakukan clustering dengan metode algoritma K-Means yaitu sebagai berikut

1. Menentukan jumlah cluster (K) pada dataset.
2. Menentukan titik pusat (centroid), penentuan centroid awal dilakukan secara acak dan pada tahap iterasi selanjutnya digunakan rumus.

$$V_{ij} = \frac{1}{N_i} \sum_{k=0}^{N_i} X_{kj}$$

Keterangan:

V_{ij} = titik pusat(centroid) rata-rata Cluster ke-i untuk variable ke-j

N_i = Jumlah Cluster ke-i

i,k = Indeks dari Cluster

j = Indeks dari variable

X_{kj} = Nilai data ke-k variable ke-j pada cluster tersebut.

3. Hitung jarak terdekat dengan centroid pada masing-masing record. Jarak centroid yang digunakan yakni Euclidean Distance dengan rumus

$$De = \sqrt{(x_i - s_i)^2 + (y_i - t_i)^2}$$

Keterangan:

De = Euclidean Distance

i = banyaknya objek

(x, y) = Koordinat objek

(s,t) = Koordinat Centroid

4. Kelompokkan objek berdasarkan jarak centroid terdekat.
5. Ulangi Langkah ke-2, lakukan iterasi hingga centroid menghasilkan nilai optimal

III. METODOLOGI PENELITIAN

Metode yang digunakan pada penelitian ini yaitu menggunakan metode CRISP-DM dimana terdiri dari 6 fase yaitu *business understanding*, *data understanding*, *data preparation*, *modeling* *evaluation* serta *deployment*.



Fig. 1. CHRISP-DM

A. Business Understanding

Pemahaman dari kegiatan data mining yang akan dilakukan dari sisi perspektif bisnis atau disebut dengan pemahaman bisnis. Pemahaman konsep bisnis pada penelitian ini mengacu kepada didasari terhadap latar belakang, tujuan, substansi.

Tujuan bisnis pada penelitian ini adalah untuk mencari pola atau strategi pembangunan infrastruktur Kota Batam dengan mengacu data dari BPS Kota Batam mengenai jumlah penduduk, luas daerah dan laju pertumbuhan penduduk dengan menggunakan algoritma K-Means. Tujuan penelitian ini untuk menggali pengetahuan tentang pola sehingga diketahui pengelompokan jumlah penduduk, luas daerah dan laju pertumbuhan di Kota Batam.

B. Data Understanding

Data understanding merupakan fase pengumpulan data dan mempelajari agar bisa mengenal data yang akan digunakan, mengidentifikasi data yang akan dipakai dan melihat kualitas data yang digunakan. Pada penelitian ini digunakan dataset dari batamkota.bps.go.id dimana data jumlah penduduk, luas daerah dan laju pertumbuhan penduduk Kota Batam di unduh.

TABLE I. TABEL AWAL PERTUMBUHAN PENDUDUK, LUAS DAERAH DAN LAJU PERTUMBUHAN (SUMBER: BATAMKOTA.BPS.GO.ID)

| No | Kecamatan | JP | LD | LP |
|----|-----------------|--------|---------|------|
| 1. | Belakang Padang | 22666 | 601.54 | 0.76 |
| 2. | Bulang | 10872 | 463.14 | 1.15 |
| 3. | Galang | 17267 | 2018.49 | 1.22 |
| 4. | Sungai Beduk | 98282 | 120.67 | 1.62 |
| 5. | Nongsa | 217175 | 63.86 | 2.78 |
| 6. | Sagulung | 85766 | 290.36 | 3.95 |
| 7. | Batam Kota | 203370 | 46.81 | 1.79 |

| | | | | |
|-----|------------|--------|--------|------|
| 8. | Sekupang | 163832 | 106.78 | 3.6 |
| 9. | Batu Aji | 141312 | 61.94 | 0.97 |
| 10. | Lubuk Baja | 85252 | 36.12 | 0.64 |
| 11. | Batu Ampar | 60865 | 39.99 | 0.51 |
| 12. | Bengkong | 123438 | 19.27 | 2.25 |

C. Data Preparation

Pada data preparation yaitu proses persiapan data untuk menyesuaikan dataset agar sesuai dengan kebutuhan yang akan digunakan pada tahap pemodelan.

- *Data Selection*

Dataset pertumbuhan penduduk Kota Batam yang diunduh dari batamkota.bps.go.id terdiri dari 12 kecamatan dengan variable yang digunakan yaitu jumlah penduduk, luas daerah serta laju pertumbuhan penduduk. Pada table dataset awal, semua data pertumbuhan penduduk perkecamatan diguakan dan tidak ada data yang dibuang.

- *Data Preprocessing*

Data *Preprocessing* disini yaitu persiapan data mentah agar siap menjadi data yang digunakan untuk pemodelan. Tahap *preprocessing* merupakan tahap pembersihan data yang tidak valid dan data *missing value*. Pada penelitian ini dataset nya lengkap sehingga tidak ditemukan data *missing value* maupun data yang tidak konsisten dan dapat digunakan pada tahap transformation.

- *Data Transformation*

Data transformation adalah proses mentransformasi data atau perubahan data misalnya merubah standard angka data menjadi 0-1. Pada penelitian ini tidak dilakukan data *transformation* karena data yang di unduh dari bps batam kota sudah lengkap sehingga tidak perlu melakukan *transformation*.

D. Modeling

Tahap modeling merupakan pemodelan dengan menggunakan Teknik *clustering* untuk mendapatkan pola data mining. Teknik pemodelan pada penelitian ini menggunakan algoritma K-Means. Sebelum melakukan pemodelan, langkah yang pertama dilakukan pada penelitian ini adalah menentukan jumlah *cluster*. Penentuan jumlah *cluster* yang populer dalam menentukan jumlah K optimal dalam penelitian ini menggunakan metode *elbow*.

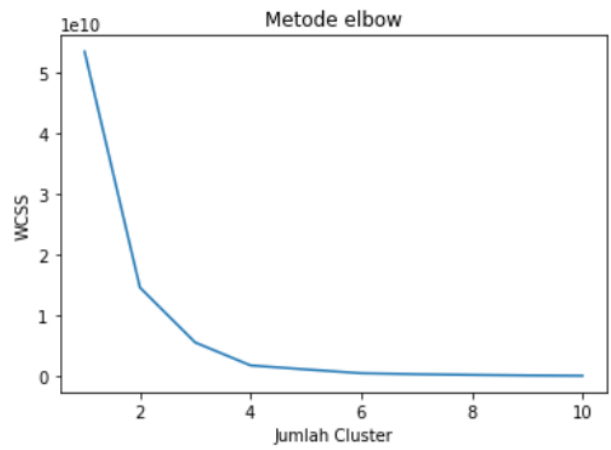


Fig. 2. Jumlah Cluster

Berdasarkan gambar diatas garis mengalami patahan yang membentuk elbow pada saat K=4, maka dengan metode elbow didapat nilai K optimal yaitu K=4.

```

## Menggunakan metode elbow untuk menentukan angka cluster
wcss = []
for i in range(1,11):
    kmeans = KMeans(n_clusters = i, init = "k-means++", random_state = 42)
    kmeans.fit(x)
    wcss.append(kmeans.inertia_)

plt.plot(range(1,11), wcss)
plt.title('Metode elbow')
plt.xlabel('Jumlah Cluster')
plt.ylabel('WCSS')
plt.show()
    
```

Fig. 3. Proses pemodelan K-Means

E. Evaluation

Setelah melakukan tahap pemodelan, Langkah selanjutnya adalah melakukan tahap evaluasi. Evaluasi yang digunakan pada Teknik ini adalah *shilhouette score*, dengan melihat nilai hasil *shilhouette* yang paling tinggi.

```

for n_clusters in range(2,7):
    clusterer = KMeans(n_clusters=n_clusters, max_iter=1000, random_state=10)
    cluster_labels = clusterer.fit_predict(x)

    silhouette_avg = silhouette_score(x, cluster_labels)
    print("n_cluster = ",n_clusters, "Nilai avg : ", silhouette_avg )

n_cluster = 2 Nilai avg : 0.5563925174101233
n_cluster = 3 Nilai avg : 0.5414724720580272
n_cluster = 4 Nilai avg : 0.6596801081890215
n_cluster = 5 Nilai avg : 0.5813093738543107
n_cluster = 6 Nilai avg : 0.5500448133362295
    
```

Fig. 4. Nilai silhouette score

Berdasarkan gambar diatas dengan mencari nilai *shilhouetter score*, maka didapat kluster K=4 memiliki nilai yang tertinggi sehingga didapat nilai optimal K = 4.

F. Deployment

Setelah tahap evaluasi selesai, langkah terakhir adakah melakukan *deployment* yaitu impelementasi dari keseluruhan

model yang sudah dibangun. Pada penelitian ini impelementasi akan diterapkan dengan dibuat sebuah laporan yang berisi hasil yang sesuai dengan metode CRISP-DM.

IV. HASIL DAN PEMBAHASAN

Pengolahan dataset yang digunakan pada penelitian ini diunduh dari website bps Kota Batam dengan data tahun 2022. Atribut pada penelitian ini adalah jumlah penduduk (JP), luas daerah Km² (LD), dan laju pertumbuhan penduduk pertahun (LP). Berdasarkan hasil *shilhouette score*, didapat nilai K optimal adalah 4, sehingga *Cluster* yang akan dicari menjadi empat bagian, yaitu *cluster* C1 = Jumlah penduduk rendah dengan luas daerah rendah, *cluster* C2 = dengan jumlah penduduk rata-rata dan luas daerah rata-rata, dan *cluster* C3 = dengan jumlah penduduk padat dan luas daerah yang besar dan C4 dengan jumlah penduduk yang sangat padat dengan luas daerah masih memungkinkan.

TABLE II. DATASET PERTUMBUHAN PENDUDUK KOTA BATAM, LUAS DAERAH, DAN LAJU PERTUMBUHAN 2022 (SUMBER: BATAMKOTA.BPS.GO.ID)

| No | Kecamatan | JP | LD | LP |
|-----|-----------------|--------|---------|------|
| 1. | Belakang Padang | 22666 | 601.54 | 0.76 |
| 2. | Bulang | 10872 | 463.14 | 1.15 |
| 3. | Galang | 17267 | 2018.49 | 1.22 |
| 4. | Sungai Beduk | 98282 | 120.67 | 1.62 |
| 5. | Nongsa | 217175 | 63.86 | 2.78 |
| 6. | Sagulung | 85766 | 290.36 | 3.95 |
| 7. | Batam Kota | 203370 | 46.81 | 1.79 |
| 8. | Sekupang | 163832 | 106.78 | 3.6 |
| 9. | Batu Aji | 141312 | 61.94 | 0.97 |
| 10. | Lubuk Baja | 85252 | 36.12 | 0.64 |
| 11. | Batu Ampar | 60865 | 39.99 | 0.51 |
| 12. | Bengkong | 123438 | 19.27 | 2.25 |

A. Analisa dengan Perhitungan manual dengan K-Means
Perhitungan manual dengan menggunakan metode K-Means pertama menentukan titik iterasi.

1. Iterasi 1:
 - i. Penentuan centroid atau titik pusat

TABLE III. CENTROID AWAL CLUSTER

| No | Kecamatan | JP | LD | LP | Cluster |
|----|-----------|-------|--------|------|---------|
| 1 | Belakang | 22666 | 601.54 | 0.76 | C1 |

| | Padang | | | | |
|----|------------|--------|-------|------|----|
| 11 | Batu Ampar | 60865 | 39.99 | 0.51 | C2 |
| 12 | Bengkong | 123438 | 19.27 | 2.25 | C3 |
| 5 | Nongsa | 217175 | 63.86 | 2.78 | C4 |

- ii. Menghitung jarak terdekat (*centroid*) terhadap masing-masing pusat *cluster*

Perhitungan jarak terdekat terhadap masing masing pusat *cluster* menggunakan persamaan *euclidean distance*

$$d_{ij} = \sqrt{\sum_{j=1}^m (x_{ij} - c_{kj})^2}$$

$$d_{11} = \sqrt{(22666 - 22666)^2 + (601.54 - 601.54)^2 + (0.76 - 0.76)^2} = 0$$

$$d_{12} = \sqrt{(10872 - 22666)^2 + (463.14 - 601.54)^2 + (1.15 - 0.76)^2} = 11,794.81$$

$$d_{13} = \sqrt{(17267 - 22666)^2 + (2018.49 - 601.54)^2 + (1.22 - 0.76)^2} = 5,581.84$$

$$\dots$$

$$d_{112} = \sqrt{(123438 - 22666)^2 + (19.27 - 601.54)^2 + (2.25 - 0.76)^2} = 100,773.68$$

$$d_{21} = \sqrt{(22666 - 60865)^2 + (601.54 - 39.99)^2 + (0.76 - 0.51)^2} = 38,203.13$$

$$d_{22} = \sqrt{(10872 - 60865)^2 + (463.14 - 39.99)^2 + (1.15 - 0.51)^2} = 49,994.79$$

$$d_{23} = \sqrt{(17267 - 60865)^2 + (2018.49 - 39.99)^2 + (1.22 - 0.51)^2} = 43,642.87$$

$$\dots$$

$$d_{212} = \sqrt{(123438 - 60865)^2 + (19.27 - 39.99)^2 + (2.25 - 0.51)^2} = 62,573$$

$$d_{31} = \sqrt{(22666 - 123438)^2 + (601.54 - 19.27)^2 + (0.76 - 2.25)^2} = 100,773.68$$

$$d_{32} = \sqrt{(10872 - 123438)^2 + (463.14 - 19.27)^2 + (1.15 - 2.25)^2} = 112,566.88$$

$$d_{33} = \sqrt{(17267 - 123438)^2 + (2018.49 - 19.27)^2 + (1.22 - 2.25)^2} = 106,189.82$$

$$\dots$$

$$d_{312} = \sqrt{(123438 - 123438)^2 + (19.27 - 19.27)^2 + (2.25 - 2.25)^2} = 0$$

$$d_{41} = \sqrt{(22666 - 217175)^2 + (601.54 - 63.86)^2 + (0.76 - 2.78)^2} = 194,509.74$$

$$d_{42} = \sqrt{(10872 - 217175)^2 + (463.14 - 63.86)^2 + (1.15 - 2.78)^2} = 206,303.39$$

$$d_{43} = \sqrt{(17267 - 217175)^2 + (2018.49 - 63.86)^2 + (1.22 - 2.78)^2} = 199,917.56$$

...

$$d_{412} = \sqrt{(123438 - 217175)^2 + (19.27 - 63.86)^2 + (2.25 - 2.78)^2} = 93,737.01$$

iii. Pengelompokan berdasarkan cluster.

TABLE IV. JARAK KE TITIK PUSAT AWAL

| No | Kecamatan | Jarak ke centroid | | | | Hasil |
|----|-----------------|-------------------|------------|------------|------------|-------|
| | | C1 | C2 | C3 | C4 | |
| 1 | Belakang Padang | 0 | 38,203.13 | 100,773.68 | 194,509.74 | C1 |
| 2 | Bulang | 11,794.81 | 49,994.79 | 112,566.88 | 206,303.3 | C1 |
| 3 | Galang | 5,581.84 | 43,642.87 | 106,189.82 | 199,917.56 | C1 |
| 4 | Sungai Beduk | 75,617.53 | 37,417.09 | 25,156.20 | 118,893.01 | C3 |
| 5 | Nongsa | 194,509.74 | 156,310.00 | 93,737.01 | 0 | C4 |
| 6 | Sagulung | 63,100.77 | 24,902.26 | 37,672.98 | 131,409.20 | C2 |
| 7 | Batam Kota | 180,704.85 | 142,505.00 | 79,932.00 | 13,805.01 | C4 |
| 8 | Sekupang | 141,166.87 | 102,967.02 | 40,394.09 | 53,343.02 | C3 |
| 9 | Batu Aji | 118,647.23 | 80,447.00 | 17,874.05 | 75,863.00 | C3 |
| 10 | Lubuk Baja | 62,588.55 | 24,387.00 | 38,186.00 | 131,923.00 | C2 |
| 11 | Batu Ampar | 38,203.13 | 0 | 62,573.00 | 156,310.00 | C2 |
| 12 | Bengkong | 100,773.68 | 62,573.00 | 0 | 93,737.01 | C3 |

Hasil perhitungan iterasi pertama menghasilkan perhitungan $C_1 = 3$, $C_2 = 3$, $C_3 = 4$, dan $C_4 = 2$. Kemudian melakukan titik pusat yang baru dengan melakukan iterasi ke dua.

2. Iterasi 2

i. Menentukan titik pusat awal baru, dengan mencari nilai centroid baru dari tiap-tiap cluster.

$$C_{1JP} = \frac{JP1+JP2+JP3}{3} = 16935$$

$$C_{1LD} = \frac{LD1+LD2+LD3}{3} = 1027.72$$

$$C_{1LP} = \frac{LP1+LP2+LP3}{3} = 1.04$$

$$C_{2JP} = \frac{JP6+JP10+JP11}{3} = 77294.33$$

$$C_{2LD} = \frac{LD6+LD10+LD11}{3} = 122.16$$

$$C_{2LP} = \frac{LP6+LP10+LP11}{3} = 1.70$$

$$C_{3JP} = \frac{JP4+JP8+JP9+JP12}{4} = 131716$$

$$C_{3LD} = \frac{LD4+LD8+LD9+LD12}{4} = 77.17$$

$$C_{3LP} = \frac{LP4+LP8+LP9+LP12}{4} = 2.11$$

$$C_{4JP} = \frac{JP5+JP7}{2} = 210272.50$$

$$C_{4LD} = \frac{LD5+LD7}{2} = 31.51$$

$$C_{4LP} = \frac{LP5+LP7}{2} = 1.46$$

Berikut hasil centroid atau titik awal baru

TABLE V. HASIL CENTROID ITERASI-2

| Cluster | JP | LD | LP |
|---------|-----------|---------|------|
| C1 | 16935.00 | 1027.72 | 1.04 |
| C2 | 77294.33 | 122.16 | 1.70 |
| C3 | 131716.00 | 77.17 | 2.11 |
| C4 | 210272.50 | 31.51 | 1.46 |

- ii. Menghitung kembali jarak terdekat terhadap masing-masing pusat cluster, seperti langkah ke 2 diatas.
- iii. Pengelompokan berdasarkan cluster, sehingga diperoleh hasil sebagai berikut

TABLE VI. JARAK KE TITIK PUSAT ITERASI 2

| No | Kecamatan | Jarak ke centroid | | | | Hasil |
|----|-----------------|-------------------|-----------|------------|------------|-------|
| | | C1 | C2 | C3 | C4 | |
| 1 | Belakang Padang | 5,746.82 | 54,630.43 | 109,051.26 | 187,607.37 | C1 |

| | | | | | | |
|----|--------------|-----------------|------------------|------------------|-----------------|----|
| 2 | Bulang | 6,089.23 | 66,423.21 | 120,844.62 | 199,400.97 | C1 |
| 3 | Galang | 1,044.92 | 60,057.28 | 114,465.46 | 193,015.73 | C1 |
| 4 | Sungai Beduk | 81,352.06 | 20,987.67 | 33,434.03 | 111,990.54 | C2 |
| 5 | Nongsa | 200,242.32 | 139,880.68 | 85,459.00 | 6,902.58 | C4 |
| 6 | Sagulung | 68,834.95 | 8,473.34 | 45,950.49 | 124,506.77 | C2 |
| 7 | Batam Kota | 186,437.58 | 126,075.69 | 71,654.01 | 6,902.52 | C4 |
| 8 | Sekupang | 146,899.89 | 86,537.67 | 32,116.01 | 46,440.56 | C3 |
| 9 | Batu Aji | 124,380.75 | 64,017.70 | 9,596.01 | 68,960.51 | C3 |
| 10 | Lubuk Baja | 68,324.20 | 7,958.14 | 46,464.02 | 125,020.50 | C2 |
| 11 | Batu Ampar | 43,941.10 | 16,429.54 | 70,851.01 | 149,407.50 | C2 |
| 12 | Bengkong | 106,507.77 | 46,143.78 | 8,278.20 | 86,834.50 | C3 |

$$C_{4JP} = \frac{JP5+JP7}{2} = 210272.50$$

$$C_{4LD} = \frac{LD5+LD7}{2} = 55.34$$

$$C_{4LP} = \frac{LP5+LP7}{2} = 2.29$$

TABLE VII. HASIL CENTROID ITERASI-3

| Cluster | JP | LD | LP |
|---------|-----------|---------|------|
| C1 | 16935.00 | 1027.72 | 1.04 |
| C2 | 82541.25 | 121.79 | 1.68 |
| C3 | 142860.67 | 62.66 | 2.27 |
| C4 | 210272.50 | 55.34 | 2.29 |

Hasil perhitungan iterasi ke 2 menghasilkan perhitungan C1 = 3, C2 = 4, C3 = 3 dan C4=2. Kemudian melakukan titik pusat yang baru dengan melakukan iterasi ke 3.

3. Iterasi 3

- i. Menentukan titik pusat awal baru, dengan mencari nilai centroid baru dari tiap-tiap cluster.

$$C_{1JP} = \frac{JP1+JP2+JP3}{3} = 16935.00$$

$$C_{1LD} = \frac{LD1+LD2+LD3}{3} = 1027.72$$

$$C_{1LP} = \frac{LP1+LP2+LP3}{3} = 1.04$$

$$C_{2JP} = \frac{JP4+JP6+JP10+JP11}{4} = 82541.25$$

$$C_{2LD} = \frac{LD4+LD6+LD10+LD11}{4} = 121.79$$

$$C_{2LP} = \frac{LP4+LP6+LP10+LP11}{4} = 1.68$$

$$C_{3JP} = \frac{JP8+JP9+JP12}{3} = 142860.67$$

$$C_{3LD} = \frac{LD8+LD9+LD12}{3} = 62.66$$

$$C_{3LP} = \frac{LP8+LP9+LP12}{3} = 2.27$$

- ii. Menghitung kembali jarak terdekat terhadap masing-masing pusat cluster, seperti langkah ke 2 diatas.
- iii. Pengelompokan berdasarkan cluster, sehingga diperoleh hasil jarak titik pusat iterasi ke 3

TABLE VIII. JARAK KE TITIK PUSAT ITERASI 3

| No | Kecamatan | Jarak ke centroid | | | | Hasil |
|----|-----------------|-------------------|------------------|------------------|-----------------|-------|
| | | C1 | C2 | C3 | C4 | |
| 1 | Belakang Padang | 5,746.82 | 59,877.17 | 120,195.88 | 187,607.37 | C1 |
| 2 | Bulang | 6,089.23 | 71,670.06 | 131,989.28 | 199,400.97 | C1 |
| 3 | Galang | 1,044.92 | 65,301.80 | 125,608.90 | 193,015.73 | C1 |
| 4 | Sungai Beduk | 81,352.06 | 15,740.75 | 44,578.71 | 111,990.54 | C2 |
| 5 | Nongsa | 200,242.32 | 134,633.76 | 74,314.33 | 6,902.58 | C4 |
| 6 | Sagulung | 68,834.95 | 3,229.15 | 57,095.12 | 124,506.77 | C2 |
| 7 | Batam Kota | 186,437.58 | 120,828.77 | 60,509.33 | 6,902.52 | C4 |
| 8 | Sekupang | 146,899.89 | 81,290.75 | 20,971.38 | 46,440.56 | C3 |
| 9 | Batu Aji | 124,380.75 | 58,770.78 | 1,548.67 | 68,960.51 | C3 |
| 10 | Lubuk Baja | 68,324.20 | 2,712.10 | 57,608.68 | 125,020.50 | C2 |
| 11 | Batu Ampar | 43,941.10 | 21,676.40 | 81,995.67 | 149,407.50 | C2 |
| 12 | Bengkong | 106,507.77 | 40,896.88 | 19,422.72 | 86,834.50 | C3 |

Hasil perhitungan iterasi ke 3 menghasilkan perhitungan C1 = 3, C2 = 4, C3 = 3 dan C4=2, karena hasil nilai centroid tidak berubah pada iterasi ke 3, maka proses perhitungan cluster selesai hanya sampai pada iterasi ke 3. Sehingga Cluster 1 dengan jumlah penduduk rendah dan luas daerah rendah diperoleh sebanyak 3 kecamatan, cluster C2 dengan jumlah penduduk rata-rata dan luas daerah rata menghasilkan 4

kecamatan, dan cluster ke 3 dengan jumlah penduduk besar dan luas daerah besar diperoleh 3 kecamatan, sedangkan jumlah penduduk yang sangat padat dengan luas daerah yang besar menghasilkan 2 kecamatan.

B. Analisa dengan RapidMiner

Hasil analisa dengan menggunakan aplikasi Rapidminer dilakukan dengan mengimport dataset dalam bentuk excel.

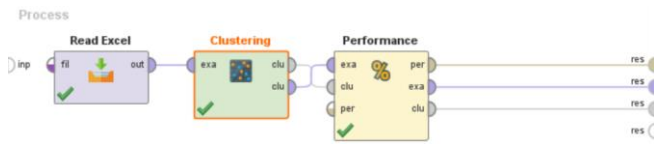


Fig. 5. Proses Analisa dengan RapidMiner

Parameter dalam penentuan jumlah cluster nya yaitu dengan K=4

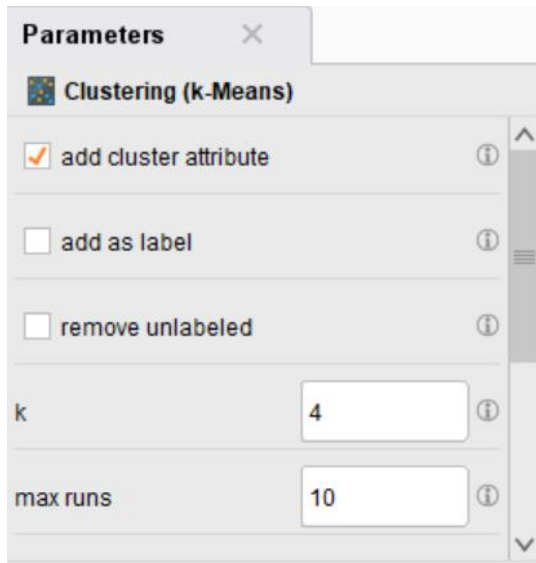


Fig. 6. Setting properties cluster k-means

Setelah selesai menentukan nilai K=4, maka didapat hasil perhitungan dari Rapidminer yaitu, C1 = 4 items, dan C2 = 3 items, C3 sebanyak 2 items, dan C4 sebanyak 3 items.

Cluster Model

```
Cluster 0: 4 items
Cluster 1: 3 items
Cluster 2: 2 items
Cluster 3: 3 items
Total number of items: 12
```

Fig. 7. Hasil cluster K-Means

V. PENUTUP

Berdasarkan hasil pengelompokan dengan menggunakan algoritma K-Means pada strategi pembangunan infrastruktur dengan melihat variabel jumlah penduduk (JP), luas daerah(LD) dan laju pertumbuhan(LP) dengan menggunakan nilai K=4 maka didapat kesimpulan bahwa, didapat cluster pertama dengan C1=3, dimana jumlah penduduk rendah, luas daerah rendah dan laju pertumbuhan rendah terdapat 3 kecamatan yaitu kecamatan Belakang Padang, Bulang, dan Galang. pengelompokan cluster C2=4 dimana hasil pengelompokan jumlah penduduk, luas daerah dan laju pertumbuhan rata-rata, didapat 4 kecamatan yaitu Sungai Beduk, Sagulung, Lubuk Baja dan Batu Ampar. Hasil pengelompokan C3 dengan jumlah penduduk, luas daerah dan laju pertumbuhan yang tinggi didapat C3 = 3 kecamatan yakni, Sekupang, Batu Aji dan Bengkong. Sedangkan hasil pengelompokan C4 dengan jumlah penduduk yang tinggi, luas daerah yang besar dan laju pertumbuhan yang ditinggi didapat hasil 2 kecamatan, yaitu kecamatan Nongsa, dan Batam Kota. Berdasarkan hasil penelitian dengan jumlah penduduk yang tinggi maupun jumlah penduduk sangat tinggi dan luas daerah yang tinggi maka perlu dipertimbangkan pembangunan di wilayah yang jumlah penduduk yang tinggi yakni pada cluster C3 dan C4.

REFERENCES

- [1] Arsandi, Arga Satria, Ismiyati Ismiyati, and Ferry Hermawan. "Dampak Pertumbuhan Penduduk Terhadap Infrastruktur Di Kota Semarang." *Jurnal Karya Teknik Sipil* 6.4 (2017): 01-14.
- [2] <https://bpbatam.go.id/batam-kota-industri/>, (25 Maret 2022), "Batam Kota Industri dan Perkembangannya Di Masa Mendatang"
- [3] Purwadi, Purwadi, Puji Sari Ramadhan, and Nurdianti Safitri. "Penerapan Data Mining Untuk Mengestimasi Laju Pertumbuhan Penduduk Menggunakan Metode Regresi Linier Berganda Pada BPS Deli Serdang." *Jurnal SAINTIKOM (Jurnal Sains Manajemen Informatika dan Komputer)* 18.1 (2019): 55-61.
- [4] Pane, Nurbariya, Sri Devi Br Sembiring, and Ikhwan Unsa. "PENGARUH PEMBANGUNAN INFRASTRUKTUR KESEHATAN, PENDIDIKAN DAN JUMLAH PENDUDUK TERHADAP PERTUMBUHAN EKONOMI DI SUMATERA UTARA." *JS (JURNAL SEKOLAH)* 4.2 (2020): 172-182.
- [5] Windarto, Agus Perdana. "Penerapan Datamining Pada Ekspor Buah-Buahan Menurut Negara Tujuan Menggunakan K-Means Clustering Method." *Techno. Com* 16.4 (2017): 348-357.
- [6] Marpaung, Preddy, and R. Fanry Siahaan. "Penerapan Algoritma K-Means Clustering Untuk Pemetaan Kepadatan Penduduk Berdasarkan Jumlah Penduduk Kota Medan." *J-SAKTI (Jurnal Sains Komputer dan*

- Informatika* 5.1 (2021): 503-521.
- [7] Gunawan, Harry, and Vega Purwayoga. "Data Mining Menggunakan Algoritma K-Means Clustering Untuk Mengetahui Potensi Penyebaran Virus Corona di Kota Cirebon." *Jurnal Sisfokom (Sistem Informasi dan Komputer)* 11.1 (2022): 1-8.
- [8] Dhewayani, F. N., et al. "Implementasi K-Means Clustering untuk Pengelompokan Daerah Rawan Bencana Kebakaran Menggunakan Model CRISP-DM." *Jurnal Teknologi dan Informasi* 12.1 (2022): 64-77.
- [9] Yuliani, Ria. "Penerapan Data Mining untuk Mengcluster Data Penduduk Miskin Menggunakan Algoritma K-Means di Dusun Bagik Endep Sukamulia Timur." *Infotek: Jurnal Informatika dan Teknologi* 4.1 (2021): 39-50.
- [10] Fatmawati, Kiki, and Agus Perdana Windarto. "Data Mining: Penerapan rapidminer dengan K-means cluster pada daerah terjangkau demam berdarah dengue (DBD) berdasarkan provinsi." *CESS (Journal of Computer Engineering, System and Science)* 3.2 (2018): 173-178.
- [11] Hutabarat, Luvita Yolanda, et al. "Penerapan Algoritma K-Means Dalam Pengelompokan Jumlah Penduduk Berdasarkan Kelurahan Di Kota Pematangsiantar." *Jurnal Ilmu Komputer dan Teknologi* 2.2 (2021): 20-26