IMPLEMENTASI ALGORITMA *K-MEANS CLUSTERING* DALAM MENGELOMPOKKAN KEPADATAN PENDUDUK DI PROVINSI DKI JAKARTA

Nurul Oktaviany, Nana Suarna, Willy Prihartono
Program Studi Teknik Informatika S1, STMIK IKMI Cirebon
Jalan Perjuangan No.10B, Karyamulya, Kec. Kesambi, Kota Cirebon, Jawa Barat 45135
vivioktaviany0510@gmail.com

ABSTRAK

Provinsi DKI Jakarta adalah salah satu wilayah dengan tingkat populasi yang tinggi di Indonesia, dengan populasinya yang sangat beragam, termasuk penduduk asli, pendatang dari berbagai daerah di Indonesia, serta warga negara asing yang tinggal di wilayah DKI Jakarta. Permasalahan kepadatan disebabkan karena faktor urbanisasi serta pekerjaan. Oleh karena itu penerapan algoritma *K-Means Clustering* untuk mengelompokkan tingkat kepadatan penduduk di Provinsi DKI Jakarta merupakan fokus utama penelitian ini. Tujuan dari penelitian ini adalah mengenali pola kepadatan penduduk di berbagai wilayah DKI Jakarta dengan metode algoritma *K-Means Clustering*. Data yang digunakan melibatkan informasi seperti jumlah penduduk, nama kelurahan, nama kecamatan dan variabel terikat lainnya. Pendekatan ini digunakan untuk mengelompokkan wilayah berdasarkan tingkat kepadatan penduduknya, sehingga dapat memberikan pemahaman yang lebih mendalam mengenai sebaran penduduk di seluruh wilayah DKI Jakarta. Melalui *implementasi* algoritma *K-Means* ini, penelitian berhasil mengelompokkan menjadi beberapa kategori kepadatan penduduk dari 2018-2020. Berdasarkan hasil evaluasi menunjukkan dua jumlah kelompok optimal, yaitu kelompok pertama padat terdapat pada kelurahan Kali Anyar kecamatan Tambora dengan kepadatan penduduk 95676.10063 jiwa/km pada tahun 2019 dan kelompok kedua adalah rendah pada kelurahan P.harapan kecamatan Kep.Seribu Utr dengan jumlah kepadatan 1045.276234 jiwa/km pada tahun 2018. Tujuan dari penelitian pengelompokan kepadatan penduduk menggunakan metode algoritma *K-Means* ini dapat memberikan informasi yang bermanfaat bagi perkembangan penduduk di Provinsi DKI Jakarta, serta bisa menjadi landasan bagi pemerintah terkait dalam perkembangan wilayah yang lebih efisien terkait alokasi sumber daya di Provinsi DKI Jakarta.

***Kata Kunci**: Data mining, Algoritma K-Means Clustering, Kepadatan Penduduk*

1. PENDAHULUAN

Provinsi DKI Jakarta adalah salah satu wilayah dengan tingkat populasi yang tinggi di Indonesia, dengan populasinya yang sangat beragam, termasuk penduduk asli, pendatang dari berbagai daerah di Indonesia, serta warga negara asing yang tinggal di sini. Pertumbuhan populasi di provinsi DKI Jakarta juga cukup cepat, terutama karena faktor urbanisasi dan pekerjaan, seringkali menimbulkan tantangan dalam pengelolaan infrastruktur, layanan publik, dan berbagai aspek kehidupan perkotaan lainnya.

Salah satu cara untuk memahami pola kepadatan penduduk adalah melalui bidang teknologi, dengan menggunakan algoritma *Clustering* seperti *K-Means* agar bisa menganalisis data yang lebih canggih dan efisien. Dampaknya bisa mengelompokkan data populasi dengan cara yang lebih tepat dan cepat, membantu dalam pemahaman pola kepadatan penduduk yang lebih akurat, serta memberikan informasi yang bermanfaat bagi perkembangan penduduk di Provinsi DKI Jakarta.

Kajian penelitian terdahulu menurut pendapat Sembering dalam jurnalnya yang berjudul Implementasi Metode *K-Means* Dalam Pengklasteran Daerah Pungutan Liar Di Kabupaten Sukabumi (Studi Kasus: Dinas Kependudukan Dan Pencatatan Sipil), menyimpulkan bahwa teknik *K-Means* digunakan untuk mengklasifikasikan praktik pungutan liar di berbagai kecamatan di Kabupaten Sukabumi dalam urusan kependudukan dan pencatatan sipil. Variabel yang dianalisis dalam penelitian ini meliputi E-KTP, AKTA, dan Kartu Keluarga. Data yang dipakai dalam penelitian berasal dari laporan bulan Januari 2019, dan klaster tingkat tinggi mencakup daerah Cireunghas, Gegerbitung, Kalapanunggal, Kalibunder, Purabaya, Simpenan, Parung Kuda, Sukaraja, Nagrak, Nyalindung, Pelabuhan Ratu, Surade, dan Warungkiara. Terdapat 19 kecamatan dengan tingkat pungutan liar sedang dan 15 kecamatan dengan tingkat pungutan liar rendah. Temuan ini dapat memberikan masukan kepada Dinas Kependudukan dan Pencatatan Sipil, di mana kecamatan dengan tingkat pungutan liar tertinggi menjadi prioritas untuk tindakan lebih lanjut[1]

Studi penelitian terdahulu menurut pendapat Hidayat dalam jurnalnya yang berjudul *Clustering* daerah rawan stunting di Jawa Barat menggunakan algoritma *k-means*, menyimpulkan bahwa penelitian ini berfokus terhadap stunting di Jawa Barat pada rentang tahun 2019-2021. Metode yang diterapkan dalam penelitian ini adalah *Cross Industry Standard Process for Data mining* (CRISP-DM) dan menggunakan data kejadian stunting beserta faktor risiko stunting dari 27 kabupaten/kota di Jawa Barat, yang diambil dari sumber resmi, dan hasil dari pengelompokkan dievaluasi menggunakan *silhouette*

coefficient dan mendapatkan nilai evaluasi sebesar 0,61. Nilai ini mengindikasikan kualifikasi *Cluster* yang memadai, sesuai dengan struktur dari masing-masing *Cluster* yang terbentuk[2]

Kajian penelitian terdahulu menurut pendapat Sari dalam jurnalnya yang berjudul Penerapan Algoritma *K-Means* Untuk *Clustering* Data Kemiskinan Provinsi Banten Menggunakan Rapidminer, menyimpulkan bahwa jumlah penduduk miskin di Provinsi Banten cenderung rendah secara nasional, terbukti dari persentase penduduk miskin Banten yang mencapai 4,94% pada September 2019, berada di bawah rata-rata nasional pada periode yang sama sebesar 9,22%. Penelitian ini menggunakan teknik *Data mining* dengan menerapkan metode *K-Means Clustering*. Data yang digunakan diambil dari Badan Pusat Statistik (BPS) dari tahun 2015 hingga 2019, melibatkan 8 Kabupaten/Kota dengan 3 variabel, yaitu jumlah penduduk miskin (dalam ribu jiwa), rata-rata lama pendidikan sekolah (dalam tahun), dan pengeluaran per kapita yang disesuaikan (dalam ribu rupiah/tahun). Semua data tersebut diolah menggunakan Rapidminer dengan pembagian menjadi 3 klaster: klaster tingkat sedang (C0), klaster tingkat tinggi (C1), dan klaster tingkat rendah (C2). Hasil analisis Rapidminer menunjukkan bahwa Kabupaten Tangerang, Kota Cilegon, dan Kota Serang termasuk dalam klaster 0, sementara Kabupaten Pandeglang, Kabupaten Lebak, dan Kabupaten Serang berada dalam klaster 1. Kota Tangerang dan Kota Tangerang Selatan ditempatkan dalam klaster 2 [3]

Penelitian terdahulu yang dilakukan oleh Ilmu dan rekan-rekannya membahas penerapan *Data mining* dalam penelitian yang berjudul "Penerapan *Data mining* Untuk Pengelompokan Kepadatan Penduduk Kabupaten Deli Serdang Menggunakan Algoritma *K-Means*." Fokus penelitian ini adalah pada kepadatan penduduk di wilayah Deli Serdang yang mendatangkan sejumlah masalah. Dampak dari peningkatan jumlah penduduk menjadi semakin signifikan, termasuk masalah kemacetan dan beban lalu lintas yang berdampak besar pada kapasitas dan volume lalu lintas di jalan-jalan. Selain itu, permasalahan kualitas air juga muncul akibat kurangnya kesadaran lingkungan di kalangan penduduk, yang seringkali membuang sampah atau limbah ke sungai. Aspek kepadatan populasi juga memainkan peran penting dalam meningkatkan tingkat kejahatan, terutama terkait tingginya tingkat pengangguran dan pekerjaan tidak tetap.Dalam penelitian ini, peneliti menggunakan metode k-means untuk mengelompokkan kepadatan penduduk ke dalam tiga kelompok (klaster): sangat padat (klaster 1), padat (klaster 2), dan sedang (klaster 3). Algoritma yang diadopsi adalah algoritma *K-Means*. Hasil dari penerapan algoritma *K-Means* menunjukkan bahwa dari total 22 kecamatan yang ada, terdapat 3 kecamatan dengan kepadatan sangat tinggi (klaster 1), 4 kecamatan dengan kepadatan tinggi (klaster 2), dan 15 kecamatan dengan kepadatan sedang (klaster 3) [4]

Penelitian kesepuluh pendapat Nasution dkk dengan penelitian yang berjudul Implementasi Algoritma K-Means dalam Pengelompokan Data Penduduk Miskin Menurut Provinsi sedang diperbincangkan. Pembahasan melibatkan penggunaan data Badan Pusat Statistik dari tahun 2012 hingga 2018 dengan memanfaatkan teknik *Data mining*. Dalam proses pengolahan data menggunakan metode K-means, data dikelompokkan ke dalam beberapa kelompok di mana setiap kelompok memiliki karakteristik yang serupa dan berbeda dari kelompok lainnya. Analisis melibatkan 34 provinsi yang dibagi menjadi 2 kelompok: kelompok tinggi dan kelompok rendah.

Fokus penelitian ini adalah mengidentifikasi provinsi-provinsi dengan tingkat kemiskinan tertinggi dan terendah. Hasilnya menunjukkan adanya 8 kelompok dengan tingkat kemiskinan tinggi dan 26 kelompok dengan tingkat kemiskinan rendah. Diharapkan temuan dari penelitian ini dapat memberikan rekomendasi kepada pemerintah untuk memberikan perhatian lebih pada provinsi-provinsi yang mengalami tingkat kemiskinan yang tinggi. [5]

2. TINJAUAN PUSTAKA

2.1. Pengertian Penduduk

Menurut ketentuan Pasal 26 ayat 2 Undang-Undang Dasar 1945, penduduk merujuk kepada warga negara Indonesia dan individu asing yang tinggal di Indonesia. Definisi penduduk suatu negara atau daerah dapat dibagi menjadi dua aspek, yakni individu yang tinggal di wilayah tersebut dan mereka yang secara hukum memiliki hak tinggal di wilayah tersebut. Dalam perspektif sosiologi, penduduk diartikan sebagai kelompok manusia yang mendiami suatu wilayah geografis dan ruang tertentu. [6]

2.2. K-Means

K-Means merupakan salah satu teknik pengelompokan data non-hirarki yang melibatkan proses pembagian data ke dalam satu atau lebih kelompok atau *Cluster*. Prinsipnya adalah mengelompokkan data dengan karakteristik yang serupa ke dalam satu *Cluster*, sedangkan data dengan karakteristik yang berbeda akan ditempatkan dalam kelompok lainnya. Algoritma ini menggunakan metode pengelompokan berdasarkan jarak, di mana data dikelompokkan ke dalam sejumlah *Cluster*. *K-Means* bekerja khususnya pada atribut yang memiliki nilai numerik. Sebagai metode partitioning *Clustering*, algoritma ini memecah data menjadi k daerah bagian yang terpisah. Dikembangkan oleh Macqueen pada tahun 1967, *K-Means* adalah salah satu algoritma dalam bidang pengelompokan data. Fungsinya melibatkan analisis data tanpa supervisi dan merupakan metode *Data mining* yang mampu mengelompokkan data ke dalam beberapa kelompok.[7]

Tujuan dari Algoritma Pengelompokan (*K-Means*) adalah mengurangi fungsi objektif yang telah ditentukan selama proses pengelompokan. Langkah ini dilakukan dengan cara mengurangi variasi data di dalam suatu *Cluster* dan, sebaliknya, meningkatkan variasi data di *Cluster* yang berbeda. [8]. Langkah-langkah dalam algoritma *K-means Clustering* adalah :

a. Menentukan jumlah *Cluster*
b. Menentukan nilai *centroid* untuk awal iterasi, nilai awal *centroid* akan dilakukan secara acak. Sedangkan jika menentukan nilai *centroid* yang merupakan tahap dari iterasi.

$$\nabla_{ij} = \frac{1}{N_i} \sum_{K=0}^{N_i} X_{kj} \quad (1)$$

Dimana:

vij : *centroid*/ rata-rata *Cluster* ke-I untuk variabel ke-j
Ni : jumlah data yang menjadi anggota *Cluster* ke-i
i,k : *indeks* dari *Cluster*
j : *indeks* dari variabel
xkj : nilai data ke-k yang ada di dalam *Cluster* tersebut untuk variabel ke-j

2.3. *Davies-Bouldin Index*

Davies-Bouldin Index(DBI) merupakan metrik yang digunakan untuk menentukan jumlah *Cluster* optimal setelah selesainya proses pengelompokan. Pendekatan DBI ini bertujuan untuk maksimalkan jarak antara satu *Cluster* dengan *Cluster* lainnya, sambil berupaya meminimalkan jarak antar objek di dalam suatu *Cluster*. Semakin kecil nilai DBI yang diperoleh (non-negatif≥0), maka hasil pengelompokan model *Clustering* yang digunakan dianggap semakin baik [9]. Menghitung *Indeks Davies-Bouldin* (DBI) dengan maksud mendapatkan nilai K yang paling optimal. Dengan demikian, dapat ditarik kesimpulan apakah pengelompokan data rumah sakit telah mencapai tujuan yang diharapkan [10]

$$DBI = \frac{1}{K} \sum_{i}^{k} = 1 \, max_{i \neq j}(R_{ij}) \quad (2)$$

Keterangan :

K : jumlah *Cluster* yang digunakan
I : banyaknya objek

Semakin kecil angka DBI yang diperoleh dari hasil percobaan dengan angka yang bukan negative , yaitu angka yang mendekati 0 maka hasilnya akan semakin baik untuk jumlah *Cluster* yang sudah dikelompokkan.

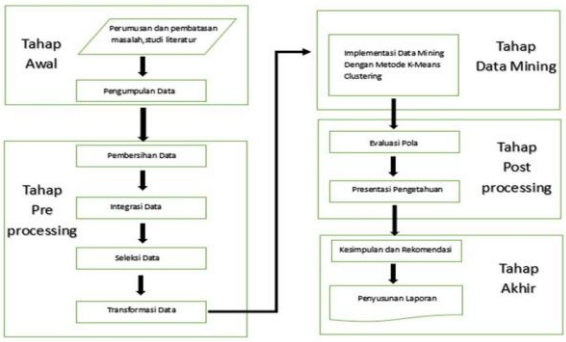
2.4. *Clustering*

Cluster adalah sekelompok entitas data yang serupa di dalam satu kelompok, sementara berbeda dengan entitas di kelompok lain. Objek-objek tersebut dikelompokkan ke dalam satu atau lebih *Cluster*, sehingga entitas yang berada dalam satu *Cluster* memiliki tingkat kesamaan yang tinggi satu sama lain. Dengan menerapkan teknik *Clustering*, kita dapat mengelompokkan berbagai jenis data rumah sakit di Jawa Barat, mengidentifikasi pola distribusi secara komprehensif, dan menemukan hubungan menarik di antara atribut-atribut data. Proses *Clustering* ini melibatkan pembentukan kelompok data yang sebelumnya tidak diketahui, serta menentukan kategori taksonomi atau batiriologi serta menganalisis topologi dari data yang sudah ada [11]

3. METODE PENELITIAN

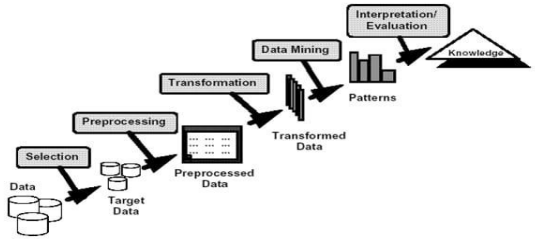
Pada penelitian ini menggunakan metode penelitian *K-Means Clustering* dengan pendekatan kuantitatif yang dimana pada hasilnya akan menampilkan angka statistik dalam bentuk visualisasi grafik. Metode *K-Means Clustering* merupakan suatu bentuk pengelompokkan sebuah nilai pada subjek yang diambil nilai berdekatan kemudian dijadikan sebagai satu kelompok, dan apabila nilai tersebut saling berjauhan maka akan dibuat kelompok baru. Penentuan kelompok tergantung oleh seseorang yang akan disaring menjadi beberapa kelompok, umumnya sebuah angka statistik dinilai akurat dengan membuat menjadi dua kelompok.

Tahapan Penelitian ini ditampilkan pada Gambar 1



Gambar 1. Tahapan Penelitian

Pada metode penelitian ini juga menggunakan tahapan KDD *(Knowledge Discovery Database).* KDD merupakan sebuah Langkah praktis dalam melakukan tahapan *Data mining* untuk mencari, mengidentifikasi pola pada data, dimana pola yang ditemukan bersifat sah dan bermanfaat dan mudah dimengerti. Dalam tahapan KDD terdapat beberapa tahapan diantaranya pembersihan data, integrasi data, pemilihan data, seleksi data, transformasi data. Tahapan– tahapan dalam KDD *(Knowledge Discovery Database)* dapat ditampilkan pada Gambar 2



Gambar 2. Tahapan KDD

Dalam tahap perancang atau penelitian, penulis menggunakan tahapan *knowledge discovery in database* (KDD), tahapannya sebagai berikut:

3.1. *Data Selection*

Data selection adalah proses pengambilan data yang berhubungan dengan analisis dari basis data. Pada tahapan ini dilakukan teknik perolehan sebuah pengurangan representasi dari data dan meminimalkan hilangnya informasi data. Hal ini meliputi metode pengurangan atribut dan kompresi data. Dari data tersebut terdapat 7 *field* atau atribut yaitu tahun, nama provinsi, nama kabupaten/kota, nama kecamatan, nama kelurahan, jumlah luas wilayah dan jumlah kepadatan penduduk.

3.2. *Pre-processing/Cleaning*

a. Pemprosesan pendahuluan dan pembersihan data merupakan operasi dasar seperti dilakukannya penghapusan noise.
b. Sebelum proses *Data mining* dapat dilaksanakan, perlu dilakukan proses cleaning pada data yang menjadi fokus KDD.
c. Proses *cleaning* mencakup antara lain membuang duplikasi data, memeriksa data yang inkonsisten, dan memperbaiki kesalahan pada data.
d. Dilakukan proses enrichment, yaitu proses memperkaya data yang sudah ada dengan data atau informasi lain *(eksternal).*

3.3. *Data Transformation*

Proses integrasi pada data yang sudah dipilih, sehingga data sesuai untuk proses *Data mining*. Merupakan proses yang sangat bergantung pada jenis atau pola informasi yang dicari dalam basis data, perubahan data tipe data yang di ubah nominal menjadi numerik (0-1) *Coding* adalah proses transformasi pada data yang telah dipilih, sehingga data tersebut sesuai untuk proses *Data mining*. Proses *coding* dalam KDD merupakan proses kreatif dan sangat tergantung pada jenis atau pola informasi yang akan dicari dalam basis data.

3.4. *Data mining*

Data mining adalah proses mencari pola atau informasi menarik dalam data terpilih dengan menggunakan teknik atau metode tertentu. Metode *k-means* atau algoritma dalam *Data mining* sangat bervariasi., seperti *karakterisasi*, *klasifikasi*, regresi, *Clustering*, asosiasi, dan lain sebagainya. Pemilihan teknik, metode atau algoritma yang tepat sangat bergantung pada tujuan dan proses KDD secara keseluruhan.

3.5. Penerapan *K-Means Clustering*

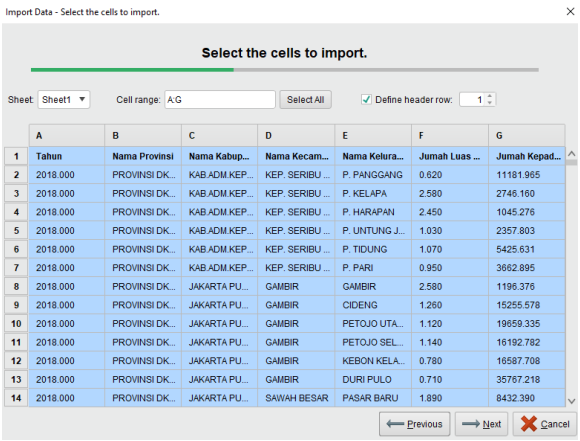
Clustering Algoritma (*K-Means*) memiliki tujuan untuk meminimalkan fungsi objective yang telah di set dalam proses *Clustering*. Tujuan tersebut dilakukan dengan cara meminimalkan variasi data yang ada di dalam *Cluster* dan memaksimalkan variasi data yang ada di *Cluster* lainnya.

4. HASIL DAN PEMBAHASAN

Data yang digunakan dalam penelitian ini sebanyak 801 data dari periode 2018-2020 yang diperoleh melalui sumber *https://data.jakarta.go.id/* yang disajikan dalam format terstruktur. Data ini mencakup atribut penting seperti tahun, nama kabupaten/kota, nama kecamatan, nama kelurahan, luas wilayah, dan jumlah kepadatan.

4.1. *Data Selection*

Data dimasukkan ke dalam proses seleksi untuk menghilangkan data yang tidak bertentangan atau duplikat dan memperbaiki kesalahan pada data yang sudah ada. Data ini memiliki 7 *field* atau atribut yaitu, tahun, nama kabupaten/kota, nama kecamatan, nama kelurahan, luas wilayah, dan jumlah kepadatan, Berikut adalah *import* data yang dapat ditampilkan pada Gambar 3.

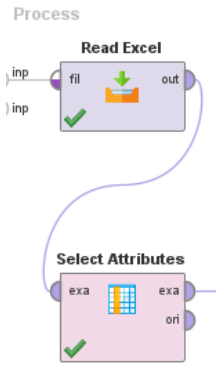


	A	B	C	D	E	F	G
1	Tahun	Nama Provinsi	Nama Kabup...	Nama Kecam...	Nama Kelura...	Jumah Luas ...	Jumah Kepad...
2	2018.000	PROVINSI DK...	KAB.ADM.KEP...	KEP. SERIBU ...	P. PANGGANG	0.620	11181.965
3	2018.000	PROVINSI DK...	KAB.ADM.KEP...	KEP. SERIBU ...	P. KELAPA	2.580	2746.160
4	2018.000	PROVINSI DK...	KAB.ADM.KEP...	KEP. SERIBU ...	P. HARAPAN	2.450	1045.276
5	2018.000	PROVINSI DK...	KAB.ADM.KEP...	KEP. SERIBU ...	P. UNTUNG J...	1.030	2357.803
6	2018.000	PROVINSI DK...	KAB.ADM.KEP...	KEP. SERIBU ...	P. TIDUNG	1.070	5425.631
7	2018.000	PROVINSI DK...	KAB.ADM.KEP...	KEP. SERIBU ...	P. PARI	0.950	3662.895
8	2018.000	PROVINSI DK...	JAKARTA PU...	GAMBIR	GAMBIR	2.580	1196.376
9	2018.000	PROVINSI DK...	JAKARTA PU...	GAMBIR	CIDENG	1.260	15255.578
10	2018.000	PROVINSI DK...	JAKARTA PU...	GAMBIR	PETOJO UTA...	1.120	19659.335
11	2018.000	PROVINSI DK...	JAKARTA PU...	GAMBIR	PETOJO SEL...	1.140	16192.782
12	2018.000	PROVINSI DK...	JAKARTA PU...	GAMBIR	KEBON KELA...	0.780	16587.708
13	2018.000	PROVINSI DK...	JAKARTA PU...	GAMBIR	DURI PULO	0.710	35767.218
14	2018.000	PROVINSI DK...	JAKARTA PU...	SAWAH BESAR	PASAR BARU	1.890	8432.390

Gambar 3. Data Selection

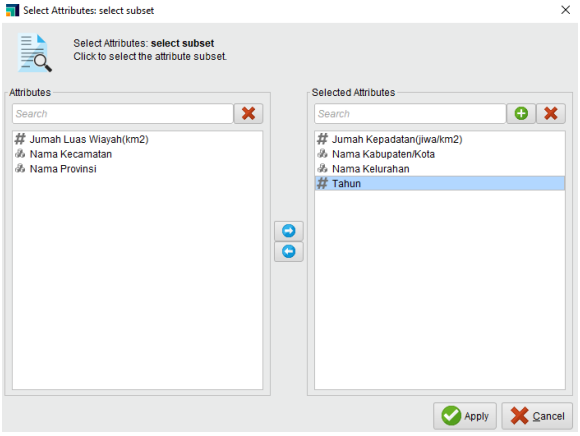
4.2. *Data Preproccesing*

Setelah memilih memasukan data, langkah berikutnya dengan melakukan *select attribute.* Pada tahap ini dilakukan pemilihan atribut mana saja yang akan dipakai nantinya untuk pengolahan data, dengan memastikan atribut tersebut memiliki nilai yang penting karena akan mempengaruhi pada hasil untuk *Clustering*. Berikut adalah proses yang dapat ditampilkan pada Gambar 4.



Gambar 4 Select Attibutes

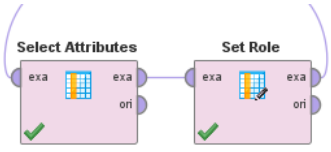
Kemudian jika sudah berada dalam *box select* atribut, dilakukan pemilihan data atribut mana saja yang akan dipakai. Untuk mengetahui data kepadatan penduduk maka atribut yang akan dipakai meliputi jumlah kepadatan(jiwa/km2), nama kabupaten/kota, nama kelurahan dan tahun. dan untuk data atribut yang tidak terpakai yaitu, jumlah luas wilayah, nama kecamatan dan nama provinsi. Berikut proses yang dapat ditampilkan pada Gambar 5.



Gambar 5 Data Processing

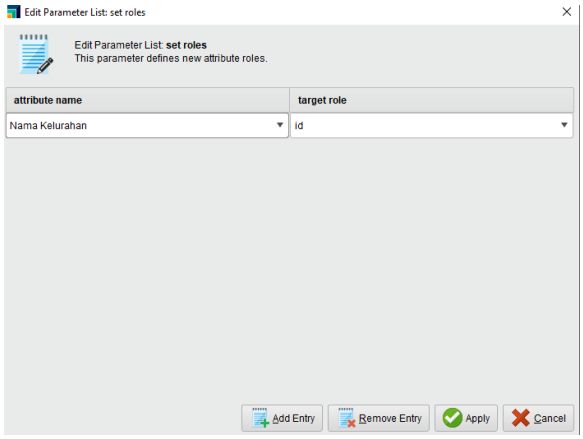
4.3. *Data Transformation*

Setelah menentukan atribut mana saja yang dipilih selanjutnya adalah dengan menentukan id pada atribut, hal ini bertujuan agar *visualisasi statistic* dapat mengambarkan secara mudah dianalisis. Dalam hal ini atribut yang akan dijadikan id yaitu nama kelurahan. Untuk mengubah atribut menjadi *id* dengan rapidminer dapat menggunakan operator *set role*. Berikut proses yang dapat ditampilkan pada Gambar 6.



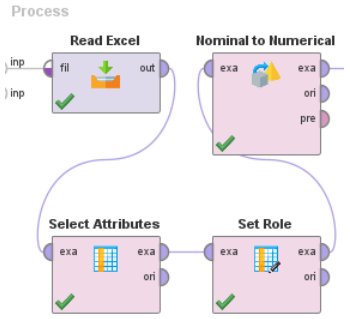
Gambar 6 Set Role

Setelah menambahkan operator *set role* kedalam process rapidminer, berikutnya dengan mengisi atribut name dan target role seperti Gambar 7.



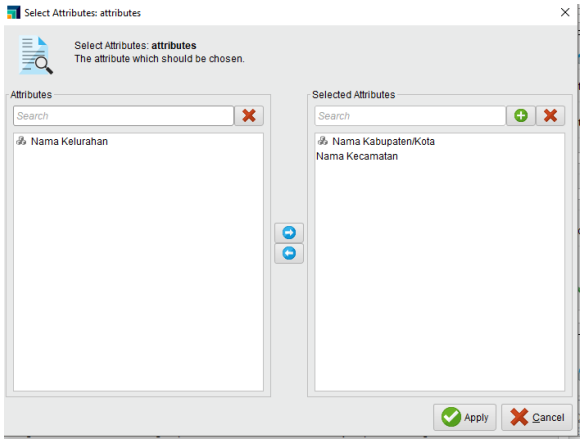
Gambar 7 *Data Transformation*

Setelah merubah atribut nama rumah kelurahan menjadi id, langkah selanjutnya adalah dengan memasukan Operator yang bernama *Nominal to Numeric* ke dalam lembar proses, dan fungsi dari Operator yaitu *Nominal to Numeric* untuk mengubah tipe dari atribut non-numerik menjadi tipe atribut numerik , Pada Parameter di samping lembar proses dari Atribut *filter type* yang digunakan yaitu *all* dan *Coding type* yang digunakan yaitu *Unique integers*, Berikut adalah Memasukan Operator *Nominal to Numeric*. Proses ini dapat ditampilkan pada Gambar 8.



Gambar 8 *Nominal to Numerical*

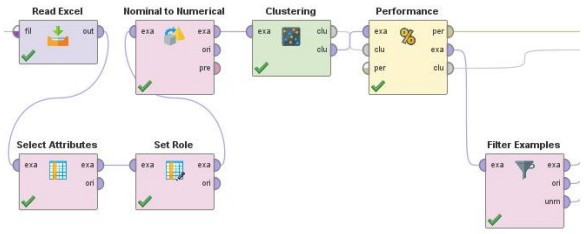
Pada atribut filter *type* pilih subset, kemudian *select atribut* yang akan dijadikan Sebagai *numerical* yaitu nama kelurahan, dan untuk opsi *coding type* dengan memilih *unique integers*. Proses ini dapat ditampilkan pada Gambar 9



Gambar 9 Data *Nominal to Numerical*

4.4. *Data Mining*

Pada tahap ini dilakukan penerapan algoritma dan metode *Data Mining*, Adapun algoritma yang digunakan adalah algoritma *K-Means* dan menggunakan metode *Clustering*. Di tahap ini langkah-langkah algoritma *K-Means* diterapkan untuk melihat hasil akhir yang telah diproses oleh *Data Mining* sesuai dengan atribut pada data yang telah di *transformasi*. Proses ini dapat ditampilkan pada Gambar 10.



Gambar 10 *Data Mining*

Setelah di lakukan proses pengujian dan mendapatkan hasil Performance *Vektor* maka hasil *dari Cluster* Model juga muncul bahwa model *Cluster* yang di hasilkan adalah K=2, K=3 dan K=4 . Berikut adalah hasil dari K=2 dengan hasil *Cluster*_0 dengan 636 Items, *Cluster*_1 dengan 165 Items, Total dari semua dataset yang di uji adalah 801 Items. Hasil tersebut akan ditampilkan pada Gambar 11.

Cluster Model

```
Cluster 0: 636 items
Cluster 1: 165 items
Total number of items: 801
```

Gambar 11 *Hasil Model K=2*

Berikut adalah hasil dari K=3 dengan hasil *Cluster*_0 dengan 482 Items, *Cluster*_1 dengan 93 Items, *Cluster_2* dengan 226 Items. Total dari semua dataset yang di uji adalah 801 Items. Hasil tersebut akan ditampilkan pada Gambar 12.

Cluster Model

```
Cluster 0: 482 items
Cluster 1: 93 items
Cluster 2: 226 items
Total number of items: 801
```

Gambar 12 Hasil Model K=3

Berikut adalah hasil dari K=4 dengan hasil *Cluster*_0 dengan 243 Items, *Cluster*_1 dengan 30 Items, *Cluster_2* dengan 100 Items, *Cluster_3* dengan 428 Items. Total dari semua dataset yang di uji adalah 801 Items. Hasil tersebut akan ditampilkan pada Gambar 13.

Cluster Model

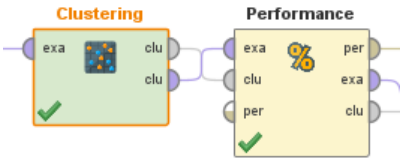
```
Cluster 0: 243 items
Cluster 1: 30 items
Cluster 2: 100 items
Cluster 3: 428 items
Total number of items: 801
```

Gambar 13 Hasil Model K=4

Dari hasil jumlah 4 *Cluster* yang dilakukan untuk mengetahui jumlah kepadatan penduduk yang optimal maka akan dilakukan evaluasi penentuan nilai k.

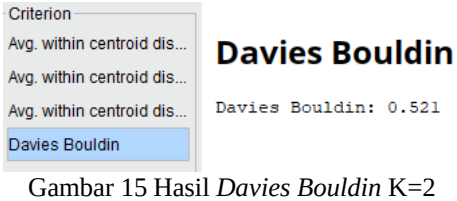
4.5. *Evaluation*

Pada tahap evaluasi dilakukan setelah melakukan tahap *Data Mining*. Tahapan ini untuk menentukan hasil keakuratan yang didapatkan oleh proses *Data Mining*, untuk menentukan seberapa akurat dan ketepatan dari hasil yang diperoleh dengan menyesuaikan dengan nilai performa pada *Data Mining*. Semakin kecil hasil performa pada *Data Mining* maka semakin baik dan akurat. Proses ini dapat ditampilkan pada Gambar 14.



Gambar 14 *Evaluation*

Setelah melakukan proses pengujian dan menghasilkan *Performance Vektor* maka hasil dari *Davies Bouldin Index* dari beberapa *Cluster* tersebut dapat ditampilkan pada Gambar 15.

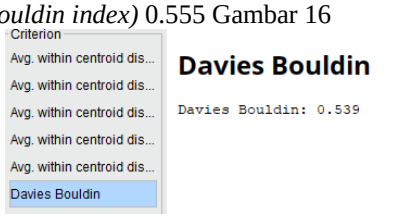


Gambar 15 Hasil *Davies Bouldin* K=2

Hasil *Davies Bouldin Index* dengan jumlah k=2 didapatkan hasil pada gambar, dengan nilai DBI (*davies bouldin index)* 0.521 Gambar 15

Gambar 16 Hasil *Davies Bouldin* K=3

Hasil *Davies Bouldin Index* dengan jumlah k=3 didapatkan hasil pada gambar, dengan nilai DBI (*davies bouldin index)* 0.555 Gambar 16



Gambar 17 Hasil *Davies Bouldin* K=4

Hasil *Davies Bouldin Index* dengan jumlah k=3 didapatkan hasil pada gambar, dengan nilai DBI (*davies bouldin index)* 0.539 Gambar 16. Setelah dilakukan 3 kali percobaan untuk menentukan hasil nilai k, Pada percobaan ke 3 dengan menggunakan nilai k yang ditentukan adalah dinilai k 2 menghasilkan nilai *Performance Vektor* (*Davies Bouldin Index*) yang optimal dibandingkan dengan percobaan , kedua dan ketiga kali proses *run*, Nilai yang dihasilkan adalah nilai yang mendekati 0 dengan nilai 0.521. Berikut hasil uji coba nilai k dapat ditampilkan pada Tabel 1.

Tabel 1 Hasil uji coba nilai K

Nilai K	***Davies Bouldin Index (DBI)***	***Cluster***
2	0.521	*Cluster*_0 : 636 *items* *Cluster*_1 : 165 *items*
3	0.555	*Cluster*_0 : 482 *items* *Cluster*_1 : 93 *items* *Cluster*_2 : 226 *items*
4	0.539	*Cluster*_0 : 243 *items* *Cluster*_1 : 30 *items* *Cluster*_2 : 100 *items* *Cluster*_3 : 428 *items*

Berdasarkan hasil pada masing-masing nilai DBI *(Davies Bouldin Index)* hasil yang paling rendah diantara masing-masing *Cluster* adalah dengan jumlah k=2 dengan nilai DBI sebesar 0.521, maka *Cluster* dengan jumlah k=2 lebih baik dibandingkan dengan *Cluster* jumlah k=3 dan k=4.

Dari hasil penelitian metode *Clustering* yang telah dilakukan menggunakan rapid miner serta penilaian performance menggunakan davies bouldin index. Pada hasil *Cluster* dengan jumlah k=2 didapatkan hasil menjadi 2 *Cluster* seperti pada yang akan ditampilkan pada Tabel 2.

Tabel 2 Hasil *Cluster*

CLUSTER	**JUMLAH**
Cluster 0	636 *items*
Cluster 1	116 *items*

Dari hasil data *Cluster* tersebut menggambarkan bahwa, *Cluster*_0 adalah nilai *Cluster* tertinggi dengan 636 items untuk jumlah kelurahan dengan kepadatan tertinggi di Provinsi DKI Jakarta dan untuk nilai *Cluster*_1 dengan 116 *items* untuk jumlah kelurahan dengan kepadatan terendah di Provinsi DKI Jakarta. Dari hasil beberapa *Cluster* diatas dapat disimpulkan bahwa data kepadatan penduduk di Provinsi DKI Jakarta dari 2018-2020 terdapat 2 wilayah kelurahan yang kepadatan penduduknya terpadat dan terendah. Berikut hasil akan ditampilkan pada Tabel 3.

Tabel 3 Hasil *Clustering*

Nama Kelurahan	**Tahun**	***Cluster***	**Jumlah Kepadatan(Jiwa/Km)**
Kali Anyar	2019	Padat	95676.10063
P.Harapan	2018	Rendah	1045.276234

5. KESIMPULAN DAN SARAN

Dari hasil analisis data pengelompokan kepadatan penduduk di Provinsi DKI Jakarta yang telah dilakukan menggunakan Algoritma *K-Means Clustering*, maka mendapatkan hasil dan kesimpulan. Pertama *Implementasi* mengelompokkan kepadatan penduduk di Provinsi DKI Jakarta berdasarkan kepadatan perkeluharan melibatkan beberapa tahapan yaitu, *data selection, preprocessing, data transformation*, *Data mining* dan *evaluation*. Kedua Dari Nilai BDI (*Davies Bouldin Index*) yang dilakukan pada 4 *Cluster* didapatkan hasil yang mendekati 0 yaitu pada *Cluster_2* nilai DBI yang optimal untuk pengelompokan kepadatan penduduk adalah 0,521. Dan yang terakhir Pengelompokan kepadatan penduduk di Provinsi DKI Jakarta dapat menghasilkan 2 *Cluster* yaitu tertinggi dan terendah. Kepadatan penduduk tertinggi terdapat pada kelurahan Kali Anyar pada tahun 2019 yang berjumlah 95676.10063(jiwa/km), dan daerah kelurahan terendah terdapat pada kelurahan P.Harapan pada tahun 2018 dengan jumlah 1045.276234(jiwa/km). Saran pada skripsi ini Peneliti selanjutnya diharapkan dapat menambahkan atribut pendukung dalam mengetahui kepadatan.

DAFTAR PUSTAKA

[1] F. Sembiring, O. Octaviana, and S. Saepudin, "Implementasi Metode K-Means Dalam Pengklasteran Daerah Pungutan Liar Di Kabupaten Sukabumi (Studi Kasus: Dinas Kependudukan Dan Pencatatan Sipil)," *J. Tekno Insentif*, 2020, [Online]. Available: http://jurnal.lldikti4.id/index.php/jurnaltekno/article/view/165

[2] T. Hidayat *et al.*, "*Clustering* daerah rawan stunting di Jawa Barat menggunakan algoritma

k-means *Clustering* stunting-prone areas in West Java using k-means algorithm," vol. 4, pp. 137–146, 2023, [Online]. Available: http://jurnal.sttmcileungsi.ac.id/index.php/infotech

[3] Y. R. Sari, A. Sudewa, D. A. Lestari, and ..., "Penerapan Algoritma K-Means Untuk *Clustering* Data Kemiskinan Provinsi Banten Menggunakan Rapidminer," *CESS (Journal …*, 2020, [Online]. Available: https://jurnal.unimed.ac.id/2012/index.php/cess/article/view/18519

[4] J. Ilmu, S. Informasi, P. Marpaung, I. Pebrian, and W. Putri, "Penerapan *Data mining* Untuk Pengelompokan Kepadatan Penduduk Kabupaten Deli Serdang Menggunakan Algoritma K-Means," vol. 6, no. September, pp. 64–70, 2023.

[5] I. Nasution, A. P. Windarto, and M. Fauzan, "Penerapan Algoritma K-Means Dalam Pengelompokan Data Penduduk Miskin Menurut Provinsi," *Build. Informatics, Technol. Sci.*, vol. 2, no. 2, pp. 76–83, 2020, doi: 10.47065/bits.v2i2.492.

[6] A. Bidarti, *Teori kependudukan*. books.google.com, 2020. [Online]. Available: https://books.google.com/books?hl=en%5C&lr=%5C&id=YM35DwAAQBAJ%5C&oi=fnd%5C&pg=PP1%5C&dq=teori+penduduk%5C&ots=hO84UDxgeV%5C&sig=D5_RPyAgzTORIeykihR-3-w4qTE

[7] A. Neva, "Penerapan Metode K-Means *Clustering* dalam Mengelompokan Jumlah Wisatawan Asing di Jawa Barat," *Algor*, vol. 2, 2023, [Online]. Available: https://jurnal.buddhidharma.ac.id/index.php/algor/article/view/1872%0Ahttps://jurnal.buddhidharma.ac.id/index.php/algor/article/download/1872/1495

[8] N. Wakhidah, "*CLUSTERING* MENGGUNAKAN K-MEANS ALGORITHM (K-MEANS ALGORITHM *CLUSTERING*)".

[9] A. A. Az-zahra, A. F. Marsaoly, I. P. Lestyani, R. Salsabila, and W. O. Z. Madjida, "Penerapan Algoritma K-Modes *Clustering* Dengan Validasi Davies Bouldin Index Pada Pengelompokkan Tingkat Minat Belanja Online Di Provinsi Daerah Istimewa Yogyakarta," *J. MSA (Mat. dan Stat. serta Apl.)*, vol. 9, no. 1, p. 24, 2021, doi: 10.24252/msa.v9i1.18555.

[10] M. A. Fajar, "Klasterisasi Proses Seleksi Pemain Menggunakan Algoritma K-Means," *J. Tek. Inform.*, pp. 1–5, 2015, [Online]. Available: http://eprints.dinus.ac.id/16498/1/jurnal_15441.pdf (05 Juni 2020)

[11] F. M. Basysyar, Y. Arie Wijaya, I. Ali, and S. Anwar, "*Clustering* Data Disabilitas menggunakan Algoritma K-Means di Kabupaten Cirebon," *JURSIMA (Jurnal Sist. Inf. dan Manajemen)*, vol. 9, no. 3, pp. 247–255, 2021.