

KLASTERISASI DATA JENIS-JENIS PEKERJAAN PENDUDUK DENGAN MENGGUNAKAN ALGORITMA K-MEANS

Wili Aprianto¹, Rudi Kurniawan², Yudhistira Arie Wijaya³

¹ Teknik Informatika, STMIK IKMI Cirebon, Cirebon

² Rekayasa Perangkat Lunak, STMIK IKMI, Cirebon

³ Sistem Informasi, STMIK IKMI, Cirebon

Jalan Perjuangan No. 10B, Majasem, Cirebon

wiliaprianto39@gmail.com

ABSTRAK

Pekerjaan di desa memiliki peran penting dalam mengembangkan ekonomi lokal dan meningkatkan kesejahteraan masyarakat. Namun, belum ada studi yang mendalam mengenai pola dan karakteristik pekerjaan di Desa Babakanmulya sehingga kurangnya informasi atas pekerjaan penduduk Desa Babakanmulya Kecamatan Jalaksana Kabupaten Kuningan dengan maksud ingin mengetahui informasi karakteristik dan pola pekerjaan dengan mendapatkan parameter apa yang terbaik dari *EuclideanDistance*, *CamberraDistance*, *CorrelationSimilarity*. Dengan data pekerjaan yang dikumpulkan Metode yang digunakan dalam penelitian ini adalah algoritma *K-Means*, yang merupakan salah satu metode pengelompokan data dalam analisis kluster. Dengan menggunakan algoritma *K-Means*, diharapkan dapat ditemukan kelompok pekerjaan yang memiliki kesamaan karakteristik dan memudahkan dalam pengambilan keputusan terkait pengembangan ekonomi di Desa Babakan Mulya dan dapat memberikan kontribusi dalam pengembangan strategi pembangunan ekonomi lokal dan memberikan rekomendasi kebijakan yang relevan untuk meningkatkan kualitas pekerjaan di desa. Hasil dari penelitian ini diharapkan dapat memberikan pemahaman yang lebih baik tentang pola pekerjaan di desa, memetakan kelompok pekerjaan yang serupa, serta mengidentifikasi faktor-faktor yang memengaruhi jenis pekerjaan yang dipilih oleh penduduk desa. Hasil Analisis diperoleh parameter terbaik yaitu *Numerical Measure Types EuclideanDistance*.

Kata kunci : Data Pekerjaan, Desa Babakan Mulya, Algoritma K-Means, Analisis Data, Cluster, Algoritma

1. PENDAHULUAN

Dalam era perkembangan pesat di bidang informatika, pengolahan dan analisis data menjadi komponen penting dalam berbagai aspek kehidupan, termasuk dalam memahami pola dan karakteristik pekerjaan penduduk di suatu wilayah. Seseorang atau kelompok orang yang tidak memiliki pekerjaan atau penghasilan tetap dianggap sebagai pengangguran [1] Angkatan kerja yang sedang berusaha mencari pekerjaan dan mempersiapkan usaha juga termasuk dalam kategori pengangguran. Pengangguran terjadi karena jumlah tenaga kerja tidak sebanding dengan pertumbuhan kesempatan kerja, akibat minimnya lapangan pekerjaan di perusahaan maupun tempat usaha di suatu wilayah. Metode yang tepat untuk analisis data pekerjaan penduduk adalah dengan melakukan pengelompokan data berdasarkan jenis pekerjaan, tingkat pendapatan, dan tingkat pendidikan.

Pekerjaan dapat mencakup berbagai jenis kegiatan, baik yang bersifat fisik maupun mental, dan melibatkan berbagai tingkat keterampilan, pekerjaan ini juga dipengaruhi oleh yaitu lapangan kerja dan padatnya penduduk [3]. Desa Babakan Mulya, Kecamatan Jalaksana, Kabupaten Kuningan merupakan salah satu daerah yang memiliki potensi sumber daya manusia yang perlu dianalisis lebih lanjut terkait dengan jenis pekerjaan yang ada di desa tersebut [4].

Penelitian ini bertujuan untuk melakukan analisis statistik dan data mining untuk mengklasifikasikan hubungan antar penduduk [5]. Penelitian ini bertujuan untuk memberikan informasi yang dapat membantu Desa Babakanmulya dalam mengoptimalkan pekerjaan masyarakat. Dengan demikian, tujuan penelitian ini adalah untuk memberikan pemahaman yang lebih baik tentang karakteristik jenis-jenis pekerjaan di wilayah tersebut sehingga Desa dapat membuat kebijakan yang lebih efektif dalam menangani masalah pekerjaan [6]

Dalam penelitian ini, metode analisis data yang digunakan adalah metode Knowledge Discovery in Database (KDD). KDD adalah proses yang bertujuan untuk mengidentifikasi keabsahan data, potensi, manfaat, dan akhirnya menghasilkan pola data yang dapat dimengerti [5]. KDD terkait erat dengan teknik integrasi dan penemuan ilmiah, interpretasi, serta visualisasi pola-pola yang terdapat dalam berbagai set data. Metode ini merujuk pada strategi-strategi yang digunakan untuk memproses dan menafsirkan data yang diperoleh dari suatu penelitian.

2. TINJAUAN PUSTAKA

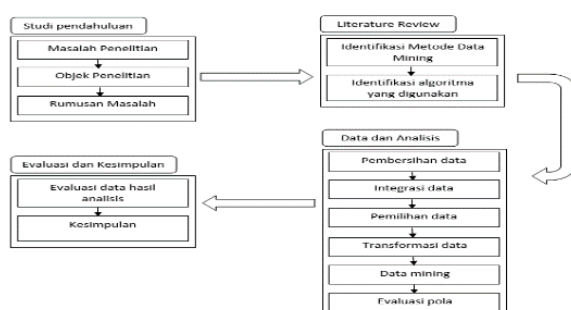
Hasil literature review yang telah dilakukan pada jurnal-jurnal penelitian terkait topik algoritma data mining *K-means Clustering* dapat dijabarkan sebagai berikut :

Paper ke-1 [10] Membahas tentang persentase penduduk miskin usia 15 tahun ke atas yang telah menyelesaikan pendidikan, status pekerjaan, pengeluaran per kapita untuk makanan, dan status kemiskinan. Hasil analisis menunjukkan bahwa terdapat dua klaster yang terbentuk, di mana klaster 1 terdiri dari Kota Tangerang, Kota Cilegon, dan Kota Tangerang Selatan, sementara klaster 2 terdiri dari Kabupaten Pandeglang, Kabupaten Lebak, Kabupaten Tangerang, Kabupaten Serang, dan Kota Serang. Evaluasi menggunakan Silhouette Index menunjukkan bahwa pembentukan dua klaster memiliki nilai keakuratan terbaik dibandingkan dengan pembentukan tiga klaster dan empat klaster.

Paper ke-2 [9] Membahas tentang menerapkan algoritma K-Means clustering dalam mengelompokkan distribusi sosial ekonomi masyarakat perumahan berdasarkan demografi kependudukan di Kota Pekanbaru. Dalam penelitian ini, dilakukan analisis statistik dan machine learning untuk mengklasifikasikan hubungan sosial ekonomi dengan demografi penduduk. Hasil dari penelitian ini diharapkan dapat memberikan rekomendasi kepada pemerintah dalam pengambilan kebijakan terhadap masyarakat dalam bidang sosial ekonomi di wilayah Kota Pekanbaru.

Paper ke-3 (Ananda et al., 2019) yang berjudul “Sistem Rekomendasi Pemilihan Peminatan Menggunakan Density Canopy K-Means” dalam penelitian ini membahas tentang membangun sistem rekomendasi pemilihan peminatan bagi mahasiswa di sebuah perguruan tinggi berdasarkan data mutu akademik semester 1 dan 2. Penelitian ini mengimplementasikan algoritme Density Canopy K-Means (DCKM) untuk membangun sistem rekomendasi. DCKM merupakan algoritme clustering yang terdiri dari dua proses utama, yaitu penentuan centroid awal dan lokasinya dengan algoritme Density Canopy serta proses clustering dengan algoritme K-Means. Penelitian ini juga mengusulkan penambahan kriteria (alpha) pada algoritme Density Canopy untuk meningkatkan kinerja algoritme tersebut.

3. METODE PENELITIAN



Gambar 1. Tahapam Metode Penelitian

Jalan Penelitian bertujuan untuk membuat target yang hendak dicapai dalam penelitian secara keseluruhan berjalan dengan baik sesuai apa yang

direncanakan atau dikehendaki sehingga proses dan tujuan dari penelitian tersebut dalam memecahkan masalah bisa berjalan dengan baik pula jelas dan terstruktur berikut tahapan yang diambil dalam penelitian ini.

4. HASIL DAN PEMBAHASAN

4.1. Hasil Penelitian

Implementasi clustering dengan algoritma K-means Clustering dirancang menggunakan perangkat RapidMiner versi 10.1. Pemilihan perangkat ini disebabkan oleh kelengkapan fitur serta kemudahan penggunaannya dalam menerapkan algoritma clustering. Dalam penelitian ini, proses Knowledge Discovery in Database melibatkan beberapa langkah, seperti pemilihan data, pra-pemrosesan, transformasi, data mining, dan evaluasi. Langkah pemilihan data, pra-pemrosesan, dan transformasi dilakukan untuk memastikan validitas dan kesiapan data yang akan dianalisis. Langkah data mining bertujuan untuk menemukan pola dan relasi dalam dataset, dengan menggunakan algoritma K-means clustering untuk mengelompokkan data ke dalam beberapa klaster. Evaluasi dilakukan pada tahap akhir untuk menilai hasil dari proses data mining tersebut

4.2. Data

Penelitian ini memanfaatkan data sekunder, yang diperoleh dari lembaga pemerintah desa. Data sekunder yang digunakan berasal dari sumber resmi pemerintah desa, Kumpulan data ini dikumpulkan dari 88 sampel yang mencakup representasi pekerjaan penduduk Desa Babakanmulya. Dataset yang dipresentasikan dalam penelitian ini dengan jumlah record sebanyak 88 record, terdiri dari 37 atribut. Rinciannya dapat ditemukan dalam yang terlampir pada Tabel 1 berikut ini

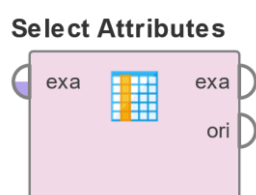
Tabel 1. Data Pekerjaan Penduduk

No	Atribut	Type
1.	Pekerjaan	Nominal
2.	Kode	Numeric
3.	Rt 01 L	Numeric
4.	Rt 01 P	Numeric
5.	Rt 02 L	Numeric
6.	Rt 02 P	Numeric
7.	Rt 03 L	Numeric
8.	Rt 03 P	Numeric
9.	Rt 04 L	Numeric
10.	Rt 04 P	Numeric
11.	Rt 05 L	Numeric
12.	Rt 05 P	Numeric
13.	Rt 06 L	Numeric
14.	Rt 06 P	Numeric
15.	Rt 07 L	Numeric
16.	Rt 07 P	Numeric
17.	Rt 08 L	Numeric
18.	Rt 08 P	Numeric
19.	Rt 09 L	Numeric
20.	Rt 09 P	Numeric
21.	Rt 10 L	Numeric

No	Atribut	Type
22.	Rt 10 P	Numeric
23.	Rt 11 L	Numeric
24.	Rt 11 P	Numeric
25.	Rt 12 L	Numeric
26.	Rt 12 P	Numeric
27.	Rt 13 L	Numeric
28.	Rt 13 P	Numeric
29.	Rt 14 L	Numeric
30.	Rt 14 P	Numeric
31.	Rt 15 L	Numeric
32.	Rt 15 P	Numeric
33.	Rt 16 L	Numeric
34.	Rt 16 P	Numeric
35.	L	Numeric
36.	P	Numeric
37.	Jumlah	Numeric

4.3. Data Selection

Dalam penelitian ini, data yang diambil adalah data mengenai individu yang mengalami masalah data pekerjaan penduduk Desa Babakanmulya. Total data yang digunakan mencapai 88 catatan dan bersifat publik. Meskipun dataset awal memiliki 37 atribut. Untuk menyaring atribut data yang akan digunakan dalam proses dan analisis selanjutnya, digunakan operator Select Attributes pada aplikasi RapidMiner. Dataset mengenai pekerjaan penduduk sebenarnya memiliki 37 atribut, tetapi melalui operator Select Attributes, data difilter sehingga hanya terdapat 36 atribut yang terpilih. Atribut yang terpilih antara lain: Pekerjaan, Rt 01 L, Rt 01 P, Rt 02 L, Rt 02 P, Rt 03 L, Rt 03 P, Rt 04 L, Rt 04 P, Rt 05 L, Rt 05 P, Rt 06 L, Rt 06 P, Rt 07 L, Rt 07 P, Rt 08 L, Rt 08 P, Rt 09 L, Rt 09 P, Rt 10 L, Rt 10 P, Rt 11 L, Rt 11 P, Rt 12 L, Rt 12 P, Rt 13 L, Rt 13 P, Rt 14 L, Rt 14 P, Rt 15 L, Rt 15 P, Rt 16 L, Rt 16 P, P, L, Jumlah. Tampilan operator Select Attributes dapat dilihat pada Gambar 2 berikut ini



Gambar 2. Operator Select Attributes.

Pada operator *Select Attributes* terdapat parameter yang perlu disesuaikan. Parameter yang digunakan dalam operator ini dapat dilihat pada Tabel 2 sebagai berikut:

Tabel 2. Parameter Select Attributes.

Parameter	Isi
Type	Include Attributes
Attribute filter type	A Subset
Select Attributes	Pekerjaan, Rt 01 L, Rt 01 P, Rt 02 L, Rt 02 P, Rt 03 L, Rt 03 P, Rt 04 L, Rt 04 P, Rt 05 L, Rt 05 P, Rt 06 L, Rt 06 P, Rt 07 L, Rt 07 P, Rt 08 L, Rt 08 P, Rt 09 L, Rt 09 P, Rt 10 L, Rt 10 P, Rt 11 L, Rt 11 P, Rt 12 L, Rt 12 P, Rt 13 L, Rt 13 P, Rt 14 L, Rt 14 P, Rt 15 L, Rt 15 P, Rt 16 L, Rt 16 P, P, L, Jumlah

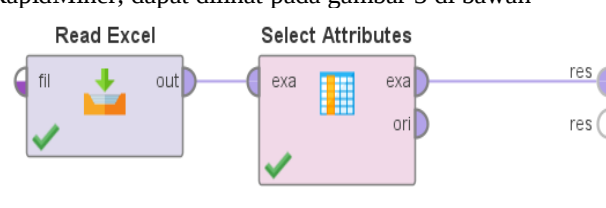
Parameter	Isi
	09 L, Rt 09 P, Rt 10 L, Rt 10 P, Rt 11 L, Rt 11 P, Rt 12 L, Rt 12 P, Rt 13 L, Rt 13 P, Rt 14 L, Rt 14 P, Rt 15 L, Rt 15 P, Rt 16 L, Rt 16 P, P, L, Jumlah

Dari hasil pembacaan operator *Select Attributes* yang telah dilakukan didapat informasi pada Tabel 3 sebagai berikut :

Tabel 3. Statistik Select Attributes.

No	Uraian	Keterangan
1.	Record	88
2.	Special Attributes	0
3.	Regular Attributes	36
4.	Attributes:	
5.	Pekerjaan	Integer, missing 0
6.	Rt 01 L	Nominal, missing 0
7.	Rt 01 P	Nominal, missing 0
8.	Rt 02 L	Integer, missing 0
9.	Rt 02 P	Integer, missing 0
10.	Rt 03 L	Integer, missing 0
11.	Rt 03 P	Integer, missing 0
12.	Rt 04 L	Integer, missing 0
13.	Rt 04 P	Integer, missing 0
14.	Rt 05 L	Integer, missing 0
15.	Rt 05 P	Integer, missing 0
16.	Rt 06 L	Integer, missing 0
17.	Rt 06 P	Integer, missing 0
18.	Rt 07 L	Integer, missing 0
19.	Rt 07 P	Integer, missing 0
20.	Rt 08 L	Integer, missing 0
21.	Rt 08 P	Integer, missing 0
22.	Rt 09 L	Integer, missing 0
23.	Rt 09 P	Integer, missing 0
24.	Rt 10 L	Integer, missing 0
25.	Rt 10 P	Integer, missing 0
26.	Rt 11 L	Integer, missing 0
27.	Rt 11 P	Integer, missing 0
28.	Rt 12 L	Integer, missing 0
29.	Rt 12 P	Integer, missing 0
30.	Rt 13 L	Integer, missing 0
31.	Rt 13 P	Integer, missing 0
32.	Rt 14 L	Integer, missing 0
33.	Rt 14 P	Integer, missing 0
34.	Rt 15 L	Integer, missing 0
35.	Rt 15 P	Integer, missing 0
36.	Rt 16 L	Integer, missing 0
37.	Rt 16 P	Integer, missing 0
38.	L	Integer, missing 0
39.	P	Integer, missing 0
40.	Jumlah	Integer, missing 0

Proses pemodelan pada tahap *selection* di RapidMiner, dapat dilihat pada gambar 3 di bawah



Gambar 3. Proses Pemodelan Tahap Selection.

4.4. Preprocessing

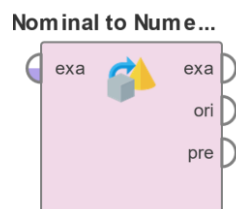
Tahap *preprocessing* dalam *Knowledge Discovery in Database (KDD)* bertujuan untuk menghilangkan *Missing Value* pada data yang akan digunakan agar proses selanjutnya tidak mengalami kesalahan. *Missing Value* perlu ditangani karena dapat mengganggu proses *modelling data mining* selanjutnya. Pada dataset pekerjaan penduduk Desa Babakanmulya, tidak ditemukan *Missing Value* atau data yang tidak memiliki nilai. Oleh karena itu, tahap *processing data* tidak dilakukan. Tampilannya dapat dilihat pada Gambar 4 sebagai berikut :

Name	Type	Missing	Statistics	Filter (36 / 36 attributes)	Search for A
Id Pekerjaan	Real	0	Min 0	Max 88	
RT 01 L	Integer	0	Min 0	Max 171	
RT 01 P	Integer	0	Min 0	Max 158	
RT 02 L	Integer	0	Min 0	Max 149	
RT 02 P	Integer	0	Min 0	Max 151	
RT 03 L	Integer	0	Min 0	Max 98	
RT 03 P	Integer	0	Min 0	Max 100	
RT 04 L	Integer	0	Min 0	Max 97	

Gambar 4. Tahap *Preprocessing*.

4.5. Transformation

Proses transformasi data diperlukan dalam data mining untuk mengubah tipe data atribut agar sesuai dengan analisis yang dibutuhkan. Dalam penelitian ini, transformasi data dilakukan karena terdapat beberapa atribut dengan tipe data nonnumerik yang perlu diubah menjadi bentuk numerik agar dapat diproses lebih lanjut. Atribut non numerik yang perlu diubah adalah Pekerjaan. Atribut ini kemudian ditransformasikan ke dalam bentuk numerik menggunakan operator Integer to Numerical. Tampilan operator Integer to Numerical dapat dilihat pada Gambar 5 berikut ini.



Gambar 5. Operator *Nominal To Numerical*.

Pada operator *Nominal to Numerical* terdapat parameter yang perlu disesuaikan. Parameter yang digunakan dalam operator ini tampak pada Tabel 4 sebagai berikut:

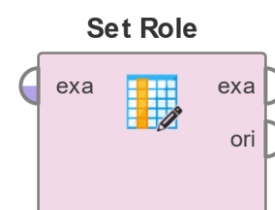
Parameter	Isi
Attribute filter type	All
Coding type	Unique Integers

Dari hasil pembacaan operator *Nominal to Numerical* yang telah dilakukan didapat informasi pada Tabel 5 sebagai berikut:

Tabel 5. Statistik *Nominal to Numerical*.

No	Uraian	Keterangan
1.	Record	88
2.	Special Attributes	0
3.	Regular Attributes	36
4.	Attributes:	
5.	Pekerjaan	Integer, missing 0
6.	Rt 01 L	Nominal, missing 0
7.	Rt 01 P	Nominal, missing 0
8.	Rt 02 L	Integer, missing 0
9.	Rt 02 P	Integer, missing 0
10.	Rt 03 L	Integer, missing 0
11.	Rt 03 P	Integer, missing 0
12.	Rt 04 L	Integer, missing 0
13.	Rt 04 P	Integer, missing 0
14.	Rt 05 L	Integer, missing 0
15.	Rt 05 P	Integer, missing 0
16.	Rt 06 L	Integer, missing 0
17.	Rt 06 P	Integer, missing 0
18.	Rt 07 L	Integer, missing 0
19.	Rt 07 P	Integer, missing 0
20.	Rt 08 L	Integer, missing 0
21.	Rt 08 P	Integer, missing 0
22.	Rt 09 L	Integer, missing 0
23.	Rt 09 P	Integer, missing 0
24.	Rt 10 L	Integer, missing 0
25.	Rt 10 P	Integer, missing 0
26.	Rt 11 L	Integer, missing 0
27.	Rt 11 P	Integer, missing 0
28.	Rt 12 L	Integer, missing 0
29.	Rt 12 P	Integer, missing 0
30.	Rt 13 L	Integer, missing 0
31.	Rt 13 P	Integer, missing 0
32.	Rt 14 L	Integer, missing 0
33.	Rt 14 P	Integer, missing 0
34.	Rt 15 L	Integer, missing 0
35.	Rt 15 P	Integer, missing 0
36.	Rt 16 L	Integer, missing 0
37.	Rt 16 P	Integer, missing 0
38.	L	Integer, missing 0
39.	P	Integer, missing 0
40.	Jumlah	Integer, missing 0

Nama atribut koordinat dan prediksi posisi dapat dibedakan dengan mengelompokkan keduanya ke dalam kategori "ID". Hal ini dilakukan agar data "ID" tidak ikut terhitung pada atribut No dan mengubah hasil pengkategorian. Tampilan operator *Set Role* dapat dilihat pada Gambar 6 berikut ini



Gambar 6. Operator *Set Role*.

Pada operator *Set Role* terdapat Atribut yang perlu disesuaikan. Atribut yang digunakan dalam operator ini adalah Nama Kecamatan untuk dijadikan sebagai id seperti pada Tabel 6 berikut ini.

Tabel 6. Parameter *Set Role*

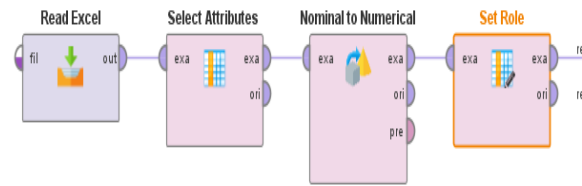
Parameter	Isi
Attribute Name	Register
Target Role	Id

Dari hasil pembacaan operator *Set Role* yang telah dilakukan didapat informasi pada Tabel 7 sebagai berikut:

Tabel 7. Statistik Set role

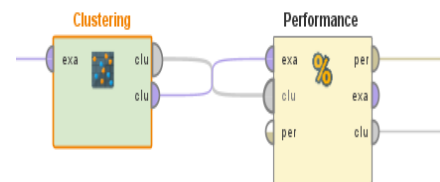
No	Uraian	Keterangan
1.	Record	88
2.	Special Attributes	0
3.	Regular Attributes	36
4.	Attributes:	
5.	Pekerjaan	Integer, missing 0
6.	Rt 01 L	Nominal, missing 0
7.	Rt 01 P	Nominal, missing 0
8.	Rt 02 L	Integer, missing 0
9.	Rt 02 P	Integer, missing 0
10.	Rt 03 L	Integer, missing 0
11.	Rt 03 P	Integer, missing 0
12.	Rt 04 L	Integer, missing 0
13.	Rt 04 P	Integer, missing 0
14.	Rt 05 L	Integer, missing 0
15.	Rt 05 P	Integer, missing 0
16.	Rt 06 L	Integer, missing 0
17.	Rt 06 P	Integer, missing 0
18.	Rt 07 L	Integer, missing 0
19.	Rt 07 P	Integer, missing 0
20.	Rt 08 L	Integer, missing 0
21.	Rt 08 P	Integer, missing 0
22.	Rt 09 L	Integer, missing 0
23.	Rt 09 P	Integer, missing 0
24.	Rt 10 L	Integer, missing 0
25.	Rt 10 P	Integer, missing 0
26.	Rt 11 L	Integer, missing 0
27.	Rt 11 P	Integer, missing 0
28.	Rt 12 L	Integer, missing 0
29.	Rt 12 P	Integer, missing 0
30.	Rt 13 L	Integer, missing 0
31.	Rt 13 P	Integer, missing 0
32.	Rt 14 L	Integer, missing 0
33.	Rt 14 P	Integer, missing 0
34.	Rt 15 L	Integer, missing 0
35.	Rt 15 P	Integer, missing 0
36.	Rt 16 L	Integer, missing 0
37.	Rt 16 P	Integer, missing 0
38.	L	Integer, missing 0
39.	P	Integer, missing 0
40.	Jumlah	Integer, missing 0

Proses pemodelan pada tahap *Transformation* di RapidMiner dapat dilihat pada Gambar 7 di bawah ini.

Gambar 7. Pemodelan *Transformation*

4.6. Data Mining

Pada tahap data mining ini, teknik yang digunakan adalah pengklasteran dengan mengimplementasikan algoritma *K-Means*. Operator ini merupakan operator utama pemodelan ini agar dapat diketahui hasil pengklasteran dataset ini. Tampilan operator clustering menggunakan *K-Means* dapat dilihat pada Gambar 8 berikut ini.

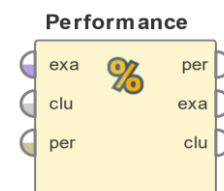
Gambar 8. Operator *K-Means*.

Pada operator *Clustering* menggunakan *K-Means* terdapat parameter yang perlu disesuaikan. Parameter yang digunakan dalam operator ini tampak pada Tabel 8 sebagai berikut:

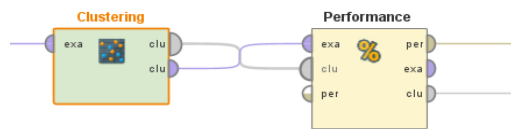
Tabel 8. Operator *K-means*

Parameter	Isi
K	2-5
Measure Types	EuclideanDistance, CamberraDistance, CorrelationSimilarity
Max Optimization	10

Peneliti menggunakan operator *Cluster Distance Performance* untuk mengukur kinerja jarak antarkelompok dengan metode DBI. Operator ini bertujuan untuk mengetahui nilai DBI dari proses klasterisasi yang dilakukan. Tampilan operator *Cluster Distance Performance* dapat dilihat pada Gambar 9 sebagai berikut.

Gambar 9. Operator *Cluster Distance Performance*.

Proses pemodelan pada tahap *Data Mining* di RapidMiner, dapat dilihat pada Gambar 10 di bawah ini.



Gambar 10. Pemodelan Data Mining.

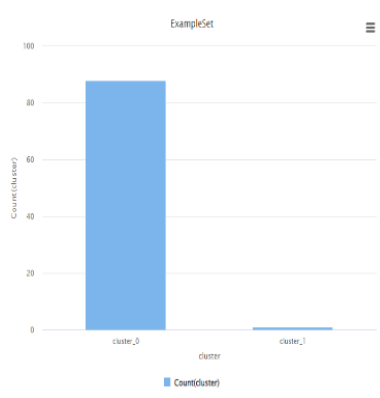
Hasil pembacaan operator *K-Means clustering* dengan parameter jenis *EuclideanDistance*, *CamberraDistance*, *CorrelationSimilarity*, serta pembacaan operator *Cluster Distance Performance*, dengan *Main Criterion* DBI, dapat dilihat pada Tabel 9 di bawah ini:

Tabel 9. Hasil *K-means*.

Cluster	Numerical Measure Types	Davies Bouldin Index
2	EuclideanDistance	-0.001
2	CamberraDistance	-0.001
2	CorrelationSimilarity	-0.001
3	EuclideanDistance	-0.009
3	CorrelationSimilarity	-0.009
3	CamberraDistance	-0.009
4	CorrelationSimilarity	-0.009
4	EuclideanDistance	-0.009
4	CamberraDistance	-0.009
5	EuclideanDistance	-0.008
5	CamberraDistance	-0.008
5	CorrelationSimilarity	-0.008

4.7. Evaluation

Hasil *clustering* dengan menggunakan algoritma *K-means* menghasilkan 2 kelompok terbaik, dengan nilai DBI 0.21. Jumlah anggota *cluster* 0 adalah 88 item, *cluster* 1 adalah 1 item. Grafik bar *cluster* terlampir pada Gambar 11 berikut ini.



Gambar 11. Jumlah Anggota Cluster.

4.8. Pembahasan

Pada tahap ini, setelah melalui proses analisis metode data mining menggunakan Knowledge Discovery in Database dengan melakukan pengujian menggunakan algoritma *K-means clustering*, dengan tujuan menjawab rumusan masalah yang telah ditetapkan pada awal penelitian.

Hasil dari nilai *K* optimal, yang didasarkan pada Davies Bouldin Index (DBI), diperoleh melalui

eksperimen dengan rentang *K*=2 hingga *K*=10. Nilai DBI yang paling rendah, mendekati 0, menunjukkan kualitas cluster yang lebih baik dibandingkan dengan cluster yang memiliki nilai DBI yang lebih tinggi. Oleh karena itu, hasil penelitian menunjukkan bahwa nilai *K*=2 menghasilkan DBI sebesar 0,021, dengan parameter terbaik adalah *EuclideanDistance*.

Tabel 10. Nilai *K* Optimal

Cluster	Iteration	Numerical Measure Types	Davies Bouldin Index
2	1	EuclideanDistance	0.021
2	1	CorrelationSimilarity	0.021
2	1	CamberraDistance	0.021
3	1	EuclideanDistance	0.301
3	1	CorrelationSimilarity	0.301
3	1	CamberraDistance	0.301
4	1	EuclideanDistance	0.315
4	1	CorrelationSimilarity	0.315
4	1	CamberraDistance	0.315
5	1	EuclideanDistance	0.277
5	1	CorrelationSimilarity	0.277
5	1	CamberraDistance	0.277
6	1	EuclideanDistance	0.304
6	1	CorrelationSimilarity	0.392
6	1	CamberraDistance	0.392
7	1	EuclideanDistance	0.253
7	1	CorrelationSimilarity	0.253
7	1	CamberraDistance	0.315
8	1	EuclideanDistance	0.194
8	1	CorrelationSimilarity	0.194
8	1	CamberraDistance	0.194
9	1	EuclideanDistance	0.066
9	1	CorrelationSimilarity	0.066
9	1	CamberraDistance	0.066
10	1	EuclideanDistance	0.134
10	1	CorrelationSimilarity	0.134
10	1	CamberraDistance	0.134

Hasil pengujian 10 kali percobaan untuk mencari parameter yang paling berpengaruh terhadap nilai *K* optimal dari pengelompokan data dengan jumlah cluster dari 2 hingga 10, masing-masing dengan nilai iterasi dari 1 hingga 10, menggunakan menggunakan tiga jenis *Numerical measure type*, yaitu *Euclidean Distance*, *Camberra Distance*, *Correlation Similarity*, type, Hasil pengujian menunjukkan bahwa parameter terbaik adalah mixed measure dengan nilai *K* optimal sebesar 0,021. Tabel hasil uji eksperimen dapat dilihat pada Tabel 11 sebagai berikut.

Tabel 11. Parameter Terbaik

Cluster	Iteration	Numerical Measure Types	Davies Bouldin Index
2	1	EuclideanDistance	0.021
2	1	CorrelationSimilarity	0.021
2	1	CamberraDistance	0.021

5. KESIMPULAN DAN SARAN

Setelah menjalani penelitian data mining menggunakan metode *K-Means Clustering* terkait dengan pekerjaan penduduk Desa Babakanmulya

Kecamatan Jalaksana Kabupaten Kuningan, kesimpulan yang dapat ditarik adalah sebagai berikut:

Setelah menganalisis data dengan menggunakan Davies Bouldin Index (DBI), nilai K Optimal yang ditemukan adalah 2, dengan nilai DBI sebesar 0,021.

Setelah melakukan penelitian ini, penulis memberikan beberapa saran yaitu studi yang mendalam mengenai pola dan karakteristik pekerjaan di Desa Babakanmulya sehingga kurang nya informasi atas pekerjaan penduduk Desa Babakanmulya Kecamatan Jalaksana Kabupaten Kuningan.

DAFTAR PUSTAKA

- [1] Haris Kurniawan, Sarjon Defit, and Sumijan, "Data Mining Menggunakan Metode K-Means Clustering Untuk Menentukan Besaran Uang Kuliah Tunggal," *J. Appl. Comput. Sci. Technol.*, vol. 1, no. 2, pp. 80–89, 2020, doi: 10.52158/jacost.v1i2.102.
- [2] R. Maliki, K. Falgenti, S. Priani, F. Fithri, M. Suherman, and D. S. Nugraha, "Perbandingan Tingkat Pengangguran Terbuka Provinsi di Indonesia Berbasis Metode K-Means Clustering," *Comput. Sci.*, vol. 2, no. 2, pp. 109–116, 2022.
- [3] S. Suhartini and R. Yuliani, "Penerapan Data Mining untuk Mengcluster Data Penduduk Miskin Menggunakan Algoritma K-Means di Dusun Bagik Endep Sukamulia Timur," *Infotek J. Inform. dan Teknol.*, vol. 4, no. 1, pp. 39–50, 2021, doi: 10.29408/jit.v4i1.2986.
- [4] T. Kartika, L. Afriza, and K. Fajri, "Pemberdayaan Masyarakat Di Desa Wisata Cibuntu Kabupaten Kuningan Provinsi Jawa Barat," *J. Indones. Tour. Hosp. Recreat.*, vol. 2, no. 1, pp. 11–24, 2019, doi: 10.17509/jithor.v2i1.16427.
- [5] E. Dianawati, P. P. Yanti, and Y. Suryandari, "Klustering Jumlah Penumpang pada Halte Bus Rapid Transit Kota Tangerang," *J. Sist. Cerdas*, vol. 2, no. 3, pp. 163–172, 2019, doi: 10.37396/jsc.v2i3.34.
- [6] Y. Diansyah, U. Juhardi, U. Muhammadiyah Bengkulu, and I. Korespondensi, "Penerapan Metode Clustering K-Means Dalam Pendataan Kependudukan Di Desa Argamulya Kabupaten Kaur," *J. Innov. Informatics*, vol. 2, pp. 223–238, 2023.
- [7] N. P. Dharshinni, G. Singh, A. Simamora, J. A. T. Naibaho, and R. D. L. Tobing, "Penerapan Algoritma K-Means Pada Data Pengangguran Di Jawa Barat," *Data Sci. Indones.*, vol. 3, no. 1, pp. 23–34, 2023, doi: 10.47709/dsi.v3i1.2767.
- [8] T. Susilowati, D. Sugiarto, and I. Mardianto, "Uji Validasi Algoritme Self-Organizing Map (SOM) dan K-Means untuk Pengelompokan Pegawai," *Tek. Inform. Fak. Teknol. Ind.*, vol. 4, no. 6, pp. 1171–1178, 2021.
- [9] D. Syaputri, P. H. Noprita, and S. Romelah, "Implementasi Algoritma K-Means untuk Pengelompokan Distribusi Sosial Ekonomi Masyarakat Berdasarkan Demografi Kependudukan," *MALCOM Indones. J. Mach. Learn. Comput. Sci.*, vol. 1, no. 1, pp. 1–6, 2021, doi: 10.57152/malcom.v1i1.5.
- [10] M. Tadi and B. A. Ningsi, "Analisis Klaster Kemiskinan Kabupaten Kota Di Provinsi Banten Menggunakan Metode K-Means," *J. Lebesgue J. Ilm. Pendidik. Mat. Mat. dan Stat.*, vol. 4, no. 1, pp. 374–385, 2023, doi: 10.46306/lb.v4i1.224.
- [11] R. Novita, A. N. Khomarudin, R. Aulia, J. Jamaluddin, A. Yuditihwa, and A. Ayuri, "Penerapan Algoritma K-Means dan Analisisnya untuk Menentukan Kebijakan Strategis Penyelesaian Studi Mahasiswa," *J. SAINTIKOM (Jurnal Sains Manaj. Inform. dan Komputer)*, vol. 22, no. 2, p. 401, 2023, doi: 10.53513/jis.v22i2.8461.
- [12] R. J. Kasim, S. Bahri, and S. Amir, "Implementasi Metode K-Means Untuk Clustering Data Penduduk Miskin Dengan Systematic Random Sampling," *Pros. SISFOTEK*, vol. 5, no. 1, pp. 95–101, 2021.
- [13] R. Ananda, M. Zidny Naf'an, A. B. Arifa, and A. Burhanuddin, "Terakreditasi SINTA Peringkat 2 Sistem Rekomendasi Pemilihan Peminatan Menggunakan Density Canopy K-Means," *Masa Berlaku Mulai*, vol. 1, no. 3, pp. 172–179, 2017.
- [14] R. Nurfalah, Dwiza Riana, and Anton, "Identifikasi Citra Beras Menggunakan Algoritma Multi-SVM Dan Neural Network Pada Segmentasi K-Means," *J. RESTI (Rekayasa Sist. dan Teknol. Informasi)*, vol. 5, no. 1, pp. 55–62, 2021, doi: 10.29207/resti.v5i1.2721.