

PENGELOMPOKAN POTENSI DESA DI KABUPATEN PROBOLINGGO DENGAN CLUSTERING MENGGUNAKAN ALGORITMA K-MEANS

Tri Gunantohadi

SMK Negeri 1 Pasuruan

Magister Informatika, Fakultas Sains dan Teknologi UIN Maulana Malik Ibrahim Malang

Email: 200605210005@student.uin-malang.ac.id

Abstrak

Dengan perbedaan kultur, geografis, dan SDM suatu daerah dapat menjelaskan keberagaman antar daerah, termasuk keragaman antar kabupaten. Potensi desa memberikan data tentang keberadaan, ketersediaan dan pengembangan potensi setiap wilayah administrasi. Untuk mengetahui tingkatan atau level suatu daerah yang memiliki karakteristik yang sama, dilakukan pengelompokan menggunakan analisis cluster. Analisis cluster adalah sebuah pengelompokan objek atau kasus ke dalam grup yang lebih kecil dimana setiap grup berisi objek itu mirip satu sama lain. Proses clustering dilakukan untuk 24 Kecamatan di Kabupaten Probolinggo. Dengan algoritma K-Means menggunakan aplikasi Rapidminer. Tiga cluster dipilih untuk menentukan Kecamatan mana saja yang memiliki potensi desa sesuai kriteria Desa Mandiri, Desa Berkembang dan Desa Tertinggal. Dari analisis menggunakan metode Elbow dan Silhouette, 1 Cluster diperoleh Kecamatan yang memiliki Desa-desanya dengan Potensi Maju atau Desa mandiri. Cluster ke-2 dengan Kecamatan yang memiliki Desa-desanya dengan Potensi Sedang atau Desa Berkembang. Cluster ke-3 dengan Kecamatan yang memiliki Desa-desanya dengan Potensi rendah atau Desa Tertinggal. Dari hasil analisis diharapkan dapat membantu pemerintah Kabupaten Probolinggo dalam mengambil kebijakan atau program lebih tepat sasaran.

Kata Kunci: Analisis Cluster, Kabupaten, Algoritma K-Means, Rapidminer, Potensi Desa

1. PENDAHULUAN

Kabupaten Probolinggo merupakan satu dari 38 kabupaten/kota yang menjadi bagian wilayah dari Provinsi Jawa Timur. Secara administratif, kabupaten ini terbagi dalam 24 kecamatan, 5 kelurahan dan 325 desa (Kabupaten Probolinggo, 2019). Potensi sumber daya alam Kabupaten Probolinggo tergolong melimpah sehingga memiliki banyak lokasi pariwisata yang ramai dikunjungi wisatawan. Selain itu, potensi perekonomian pun terus bertumbuh dengan berdirinya beberapa pabrik di Kabupaten Probolinggo. Potensi sumber daya alam, pariwisata dan potensi perekonomian ini tersebar di beberapa wilayah kecamatan di Kabupaten Probolinggo.

Suatu wilayah mempunyai karakter tertentu, yang meliputi aspek fisik dan aspek non fisik. Aspek fisik diantaranya sarana, prasarana dan akses. Aspek non fisik diantaranya sosial-budaya, ekonomi, dan penduduk (Sutami, 1977). PODES (Potensi Desa) menyediakan data tentang keberadaan, ketersediaan, dan perkembangan potensi yang dimiliki setiap wilayah administrasi JURNAL GAUSSIAN Vol. 5, No. 4, Tahun 2016 Halaman 802 pemerintahan. Potensi tersebut meliputi: sarana dan prasarana wilayah serta potensi ekonomi, sosial, budaya, dan aspek kehidupan masyarakat lainnya (BPS, 2019). Untuk mengetahui wilayah kecamatan yang memiliki karakteristik wilayah yang sama, dapat dilakukan pengelompokan dengan

mengelompokkan kecamatan-kecamatan berdasarkan sarana, prasarana, sosial, ekonomi, dan penduduk.

Analisis kluster (Cluster Analysis) merupakan salah satu metode statistika yang dapat digunakan untuk melakukan proses pengelompokan. Dalam pengelompokannya digunakan suatu ukuran yang dapat menerangkan kedekatan antar data untuk menerangkan struktur grup sederhana dari data yang kompleks, yaitu ukuran jarak. Ukuran jarak yang sering digunakan adalah ukuran jarak Euclidean (Johnson dan Wichern, 1982). Berdasarkan uraian diatas, penulis akan melakukan proses pengelompokan kecamatan di Kabupaten Probolinggo berdasarkan pada data Potensi Desa Kabupaten Probolinggo 2019 menggunakan analisis kluster non-hierarki. Untuk mencari jarak Euclidean dan clustering data penulis menggunakan aplikasi Rapidminer 5.3.015.

2. LANDASAN TEORI

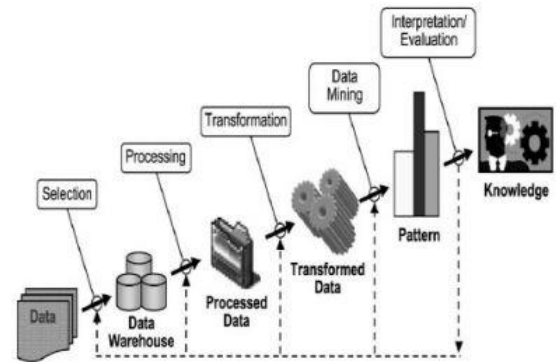
2.1 Data Mining

Data mining merupakan proses menemukan korelasi baru yang bermanfaat, pola dan trend dengan menambang sejumlah repository data dalam jumlah besar, menggunakan teknologi pengenalan pola seperti statistik dan teknik matematika. Data mining disebut juga dengan knowledge discovery in database (KDD) ataupun pattern recognition. Data Mining dapat dibagi menjadi empat kelompok, yaitu model prediksi (prediction modelling), analisis kelompok (Cluster analysis), analisis asosiasi (association analysis) dan deteksi anomaly (anomaly detection).

2.2 Tahapan Data Mining

Tahapan yang dilakukan pada proses data mining diawali dari seleksi data dari data sumber ke data target, tahap pre-processing untuk memperbaiki kualitas data, transformasi, data mining serta tahap interpretasi dan evaluasi yang menghasilkan output berupa pengetahuan baru yang diharapkan memberikan kontribusi yang

lebih baik. Secara detail dijelaskan sebagai berikut:



Gambar 1. Tahapan Data Mining

Berikut ini adalah penjelasan tahapan data mining berdasarkan gambar 1:

1. Data selection

Pemilihan (seleksi) data dari sekumpulan data operasional perlu dilakukan sebelum tahap penggalian informasi dalam KDD dimulai. Data hasil seleksi yang digunakan untuk proses data mining, disimpan dalam suatu berkas, terpisah dari basis data operasional.

2. Pre-processing /cleaning

Sebelum proses data mining dapat dilaksanakan, perlu dilakukan proses cleaning pada data yang menjadi fokus KDD. Proses cleaning mencakup antara lain membuang duplikasi data, memeriksa data yang inkonsisten, dan memperbaiki kesalahan pada data.

3. Transformation

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.

4. Data mining

Data mining adalah proses mencari pola atau informasi menarik dalam data terpilih dengan menggunakan teknik atau metode tertentu. Teknik, metode, atau algoritma dalam data mining sangat bervariasi. Pemilihan metode atau algoritma yang tepat sangat bergantung

pada tujuan dan proses KDD secara keseluruhan.

5. Interpretation/evaluation

Pola informasi yang dihasilkan dari proses data mining perlu ditampilkan dalam bentuk yang mudah dimengerti oleh pihak yang berkepentingan. Tahap ini merupakan bagian dari proses KDD yang disebut interpretation. Tahap ini mencakup pemeriksaan apakah pola atau informasi yang ditemukan bertentangan dengan fakta atau hipotesis yang ada sebelumnya.

2.3 Clustering

Clustering atau klasterisasi adalah metode pengelompokan data. Menurut Tan, 2006 clustering adalah sebuah proses untuk mengelompokkan data ke dalam beberapa cluster atau kelompok sehingga data dalam satu cluster memiliki tingkat kemiripan yang maksimum dan data antar cluster memiliki kemiripan yang minimum.

2.4 K-Means

K-means merupakan salah satu teknik pengelompokan yang bekerja berdasarkan partitioned clustering. Prinsip kerja dari pengelompokan hierarchical clustering dilakukan secara bertahap. Dan disetiap iterasi dari pengelompokan hierarchical clustering hanya ada satu pemilihan penggabungan suatu item terhadap item lainnya. Langkah-langkah melakukan Clustering dengan metode K-Means adalah sebagai berikut:

1. Pilih jumlah cluster k.
2. Inisialisasi k pusat cluster ini bisa dilakukan dengan berbagai cara. Namun yang paling sering dilakukan adalah dengan cara random. Pusat-pusat cluster diberi nilai awal dengan angka-angka random.
3. Alokasikan semua data/ objek ke cluster terdekat. Kedekatan dua objek ditentukan berdasarkan jarak kedua objek tersebut. Untuk menghitung jarak semua data ke setiap titik pusat cluster dapat menggunakan teori jarak

Euclidean yang dirumuskan sebagai berikut:

$$D(i,j) = \sqrt{(X_{1i} - X_{1j})^2 + (X_{2i} - X_{2j})^2 + \dots + (X_{ki} - X_{kj})^2}$$

dimana:

$D(i,j)$ = Jarak data ke i ke pusat cluster j

X_{ki} = Data ke i atribut data ke k

X_{kj} = Titik pusat ke j pada atribut ke k

4. Hitung kembali pusat cluster dengan keanggotaan cluster yang sekarang. Pusat cluster adalah rata-rata dari semua data/objek dalam cluster tertentu. Jika dikehendaki bisa juga menggunakan median dari cluster tersebut. Jadi rata-rata (mean) bukan satu-satunya ukuran yang bisa dipakai.

$$R_k = \frac{1}{N_k} (X_{1k} + X_{2k} + \dots + X_{nk})$$

dimana :

RR_k = Rata-rata baru.

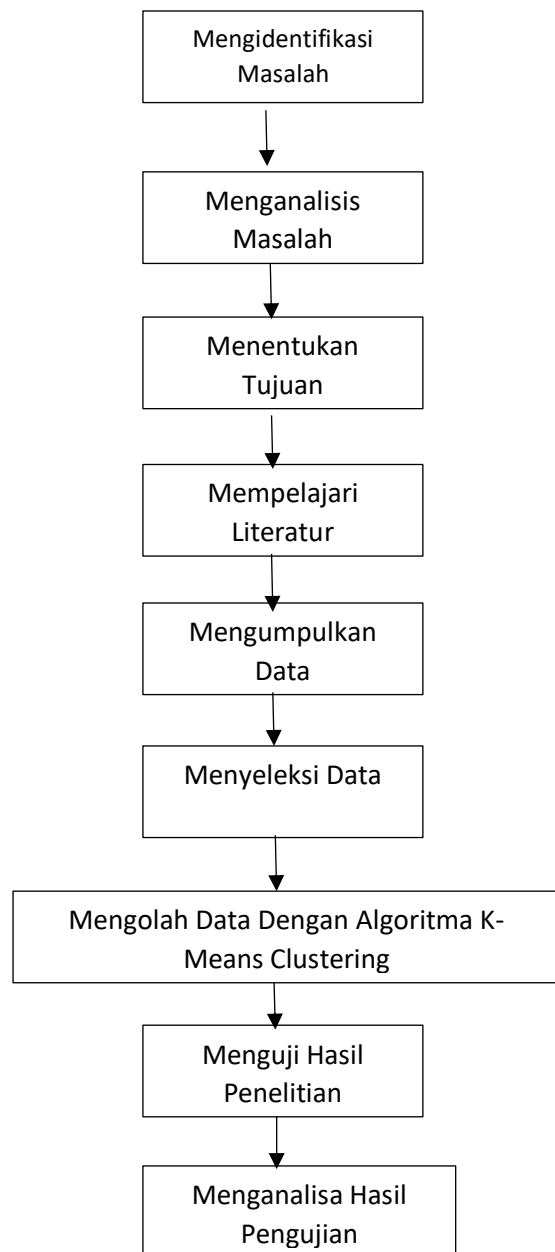
N_k = Jumlah training pattern pada cluster (k).

X_{nk} = Pola ke (n) yang menjadi bagian cluster (k).

5. Tugaskan lagi tiap objek memakai pusat cluster yang baru. Jika pusat cluster tidak berubah lagi maka proses clustering selesai. Atau, kembali ke langkah nomor 3

3. METODE PENELITIAN

Dalam metodologi penelitian terdapat urutan kerangka kerja yang harus diikuti. Kerangka kerja ini merupakan langkah langkah yang dilakukan dalam penelitian. Adapun kerangka kerja yang digunakan dalam penelitian ini adalah seperti terlihat pada gambar dibawah ini.



Gambar 2 Kerangka Kerja Penelitian

4. HASIL DAN PEMBAHASAN

4.1 Sumber Data

Data yang digunakan dalam pengelompokan adalah data-data kecamatan di Kabupaten Probolinggo 2019. Data yang digunakan merupakan data sekunder yang didapat dari Potensi Desa (PODES) Kabupaten Probolinggo 2019 yang diperoleh dari Badan Pusat Statistik Kabupaten Probolinggo. Berikut tabel data potensi desa dari kecamatan yang ada di Kabupaten Probolinggo:

No	NamaKec	jmlDesa	jmlSD_MI	jmlSMP_MTs	jmlSMA_SM	Puskesmas	jmlPoliklinik	jmlPasar
1	Bantaran	10	41	18	6	4	6	1
2	Banyuwari	14	50	24	9	6	8	2
3	Besuk	17	39	9	5	6	14	3
4	Dringu	14	29	5	3	6	8	3
5	Gading	19	60	16	5	5	14	2
6	Gending	13	21	5	6	5	8	3
7	Kotanyar	13	29	6	2	4	1	3
8	Kraksaan	18	52	33	23	4	14	3
9	Krejengan	17	47	22	15	5	12	1
10	Krucil	14	64	16	10	5	9	2
11	Kahunpan	7	34	6	2	4	4	2
12	Leces	10	53	16	7	4	6	1
13	Lumbang	10	27	10	2	3	7	2
14	Maron	18	56	24	15	6	12	1
15	Paiton	20	55	24	18	5	14	4
16	Pajarakan	12	35	9	10	4	8	3
17	Pakuniran	17	54	22	6	5	3	2
18	Sukapura	12	21	10	2	3	9	2
19	Sumber	9	24	5	2	5	4	1
20	Sumberasih	13	38	16	10	5	7	6
21	Tegalsiwala	12	37	8	5	6	6	2
22	Tiris	16	92	30	11	6	10	2
23	Tongas	14	44	11	1	6	7	5
24	Wonomerto	11	31	11	3	4	7	1

Sumber: BPS Kabupaten Probolinggo
<https://probolinggakab.bps.go.id/>

4.2 Cendroid

Data Dalam penerapan algoritma K-means dihasilkan nilai titik tengah atau centroid dari data. Proses pencarian nilai titik tengah dilakukan dengan mengambil nilai terbesar (maksimum) untuk cluster tingkat tinggi (C1), nilai sedang untuk cluster tingkat sedang (C2) dan nilai terkecil (minimum) untuk cluster tingkat rendah (C3). Berikut ini nilai centroid awal dapat diketahui pada tabel berikut:

Attribute	cluster_0	cluster_1	cluster_2
jmlSD_MI	54.556	92	32.143
jmlSMP_MTs	21.889	30	9.214
jmlSMA_SMI	12	11	4.214
Puskesmas	5	6	4.643
jmlPoliklinik	10.222	10	6.857
jmlPasar	2	2	2.643

Tabel Centroid Data Awal

4.3 Clustering Data

Proses clustering dengan menggunakan nilai centroid awal yang terdapat pada tabel, akan memperoleh hasil pengelompokan yang dapat dilihat pada tabel dibawah ini:

A	B	K	L	M
No	NamaKec	C1	C2	C3
1	Bantaran			1
2	Banyuanyar	1		
3	Besuk			1
4	Dringu			1
5	Gading	1		
6	Gending			1
7	Kotaanyar			1
8	Kraksaan	1		
9	Krejengan	1		
10	Krucil	1		
11	Kahuripan			1
12	Leces	1		
13	Lumbang			1
14	Maron	1		
15	Paiton	1		
16	Pajarakan			1
17	Pakuniran	1		
18	Sukapura			1
19	Sumber			1
20	Sumberasih			1
21	Tegalsiwalan			1
22	Tiris		1	
23	Tongas			1
24	Wonomerto			1

Tabel Pengelompokan Data Awal

diperoleh pengelompokan: Cluster tingkat tinggi (C1) terdiri dari Kecamatan Banyuanyar, Gading, Kraksaan, Krejengan, Krucil, Leces, Maron, Paiton, Pakuniran, Cluster tingkat sedang (C2) yaitu Kecamatan Tiris, dan Cluster tingkat rendah (C3) terdiri dari Kecamatan Bantaran, Besuk, Dringu, Gending, Kotaanyar, Kahuripan, Lumbang, Pajarakan, Sukapura, Sumber, Sumberasih, Tegalsiwalan, Tongas dan Wonomerto. Proses iterasi akan terus berlangsung sampai hasil iterasi terakhir sama dengan hasil iterasi sebelumnya. Proses nilai titik tengah atau centroid akan menyesuaikan sesuai dengan iterasi yang ada. Dalam hal ini proses akan dilanjutkan dengan menggunakan aplikasi RapidMiner.

4.4 Penerapan aplikasi RapidMiner

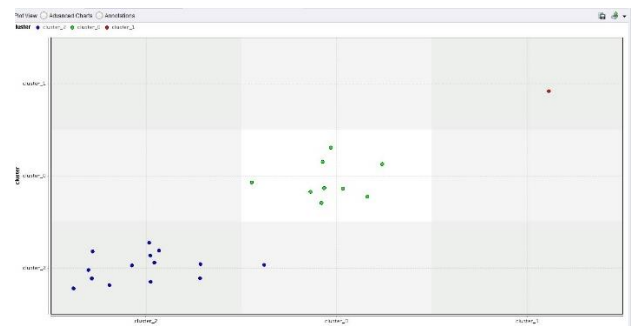
Hasil pengelompokan data Potensi Desa berdasarkan Kecamatan dengan Rapid Miner :

1. Data view: merupakan sheet di Result Perspective untuk menampilkan data yang telah diolah secara keseluruhan lengkap dengan klasternya dari example set (read csv). Data view dapat lihat pada gambar.

Row No.	No	cluster	jmiSD_MI	jmiSMP_MTs	jmiSMA_SM	Puskesmas	jmiPoliklinik	jmiPasar
1	1	cluster_2	41	18	6	4	6	1
2	2	cluster_0	50	24	9	6	8	2
3	3	cluster_2	39	9	5	6	14	3
4	4	cluster_2	29	5	3	6	8	3
5	5	cluster_0	60	16	5	5	14	2
6	6	cluster_2	21	5	6	5	8	3
7	7	cluster_2	29	6	2	4	1	3
8	8	cluster_0	52	33	23	4	14	3
9	9	cluster_0	47	22	15	5	12	1
10	10	cluster_0	64	16	10	5	9	2
11	11	cluster_2	34	6	2	4	4	2
12	12	cluster_0	53	16	7	4	6	1
13	13	cluster_2	27	10	2	3	7	2
14	14	cluster_0	56	24	15	6	12	1
15	15	cluster_0	55	24	18	5	14	4
16	16	cluster_2	35	9	10	4	8	3
17	17	cluster_0	54	22	6	5	3	2
18	18	cluster_2	21	10	2	3	9	2
19	19	cluster_2	24	5	2	5	4	1
20	20	cluster_2	38	16	10	5	7	6
21	21	cluster_2	37	8	5	6	6	2
22	22	cluster_1	92	30	11	6	10	2
23	23	cluster_2	44	11	1	6	7	5
24	24	cluster_2	31	11	3	4	7	1

Gambar 3 Data View

2. Flot view : merupakan sheet di Result Perspective untuk menampilkan data yang telah diolah secara keseluruhan lengkap dengan klasternya dari example set (read excel) dalam bentuk diagram Scatter. Tampilannya dapat dilihat pada gambar berikut:



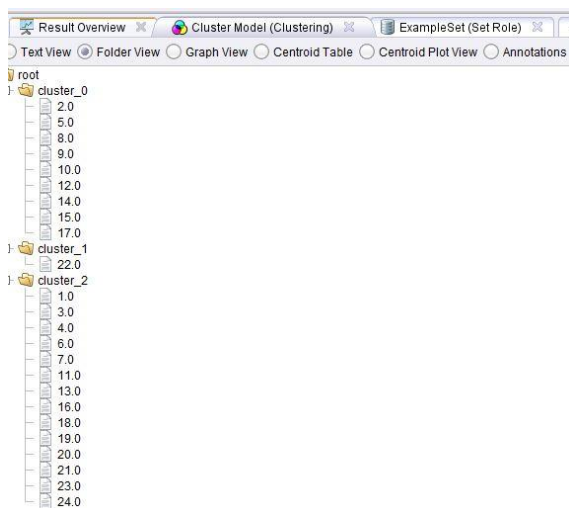
Gambar 4 Flot View

3. Text view : merupakan sheet untuk menampilkan database yang telah diolah secara keseluruhan lengkap dengan clusternya. Tampilan dari cluster model (clustering) dapat lihat pada gambar berikut.



Gambar 5 Text View

4. Folder view : merupakan sheet untuk menampilkan database yang telah diolah secara keseluruhan lengkap dengan clusternya. Tampilan folder view dapat dilihat pada gambar berikut.



Gambar 6 Folder View

5. KESIMPULAN

Dari hasil penelitian yang telah dilakukan, maka penulis dapat menarik beberapa kesimpulan, diantaranya sebagai berikut : 1. Pengujian yang dilakukan dalam penelitian ini, iterasi clustering pada data potensi desa pada kecamatan di Kabupaten Probolinggo terjadi sebanyak 6 kali iterasi. 2. Data tersebut diolah menggunakan Rapidminer untuk ditentukan nilai centroid dalam 3 cluster yaitu cluster tingkat tinggi (C1), cluster tingkat sedang (C2) dan cluster tingkat rendah (C3). 3. Dari 24 data Potensi Desa berdasarkan Kecamatan dapat diketahui, 9 Kecamatan Cluster tingkat tinggi yang memiliki potensi desa besar yakni Banyuwangi, Gading, Kraksaan, Krejengan, Krucil, Leces, Maron, Paiton,

Pakuniran, 1 Kecamatan cluster tingkat sedang dan 14 Kecamatan lainnya termasuk cluster tingkat rendah.

DAFTAR PUSTAKA

- Badan Pusat Statistika Kabupaten Probolinggo Data Tahun 2019
- Johnson, R. A. dan Wincern, D. W. 1982. Applied Multivariate Statistical Analysis. New Jersey : Prentice Hal, Inc.
- Johnson, R. A. dan Wincern, D. W. 2007. Applied Multivariate Statistical Analysis. Sixth Edition. New Jersey: Pearson International Edition.
- Laraswati, T. F. 2014. Perbandingan Kinerja Metode Complete Linkage, Metode Average Linkage, dan Metode K-Means Dalam Menentukan Hasil Analisis Cluster. Skripsi. Universitas Negeri Yogyakarta
- C. J. M. S. Fina Nasari, "Penerapan Algoritma K-Means Clustering Untuk Pengelompokan Penyebaran Diare Di Kabupaten Langkat," pp. 108–119.
- A. P. Windarto, "Implementation of Data Mining on Rice Imports by Major Country of Origin Using Algorithm Using K-Means Clustering Method," Int. J. Artif. Intell. Res., vol. 1, no. 2, pp. 26–33, 2017.
- R. Hidayat, R. Wasono, and M. Y. Darsyah, "Pengelompokan Kabupaten / Kota Di Jawa Tengah," pp. 240–250, 2017.