

PENGELOMPOKAN KECAMATAN BERDASARKAN HASIL PRODUKSI KOMODITAS UNGGULAN DI KABUPATEN JEMBER MENGGUNAKAN ALGORITMA *PARTITIONING AROUND MEDOIDS*

Alfian Hadiyatullah¹, Deni Arifianto², Habibatul Azizah Al Faruq³

Program Studi Teknik Informatika, Fakultas Teknik, Universitas Muhammadiyah Jember

alfianhadi36@gmail.com¹, deniarifianto@unmuhjember.ac.id², habibatulazizah@unmuhjember.ac.id³

ABSTRAK

Clustering adalah proses mengelompokkan sejumlah data ke dalam cluster, sehingga dalam setiap cluster akan berisi data yang semirip mungkin. Salah satu algoritma clustering adalah Partitioning Around Medoids (PAM). Algoritma PAM menggunakan objek pada sekumpulan objek untuk mewakili sebuah cluster. Pada penelitian ini membahas tentang penerapan algoritma PAM untuk mengelompokkan kecamatan di kabupaten Jember berdasarkan hasil produksi komoditas unggulan menurut Badan Pusat Statistik (BPS) tahun 2018. Data yang dijadikan sumber penelitian bersumber dari situs website resmi Badan Pusat Statistik Kabupaten Jember. Hasil dari penerapan algoritma PAM dalam mengelompokkan kecamatan di kabupaten Jember pada hasil produksi komoditas unggulan diperoleh cluster optimum pada skenario 2 cluster sampai dengan 10 cluster berdasarkan nilai Davies Bouldin Index terkecil yaitu 0,087 pada 2 cluster dengan anggota cluster 1 terdapat 30 kecamatan dan cluster 2 terdapat terdapat 1 kecamatan, tetapi pada cluster 2 yang mempunyai 1 kecamatan ini terindikasi outliers, jadi tidak bisa dilakukan proses perhitungan selanjutnya. Untuk itu dilakukan proses perhitungan pada cluster optimum yang kedua yaitu pada 4 cluster yang mempunyai nilai 0,612.

Kata kunci : Hasil Produksi, Clustering, PAM, Davies Bouldin Index.

ABSTRACT

Clustering is the process of grouping data into clusters, so that in each cluster will contain data that is as similar as possible. One of the clustering algorithms is Partitioning Around Medoids (PAM). The PAM algorithm uses objects in a group to represent a cluster. This study discusses the PAM algorithm application to classify sub-districts in Jember Regency based on superior production results according to the Central Statistics Agency (BPS) in 2018. The data sources used in this research are sourced from the official website of the Central Statistics Agency of Jember. The results of the PAM algorithm application in grouping districts in Jember Regency on the yield of superior commodity production obtained optimum clusters in scenario 2 clusters up to 10 clusters based on the smallest Davies Bouldin Index value that is 0.087 in 2 clusters with cluster 1 members there are 30 districts and cluster 2 there are contained 1 sub-district, but in cluster 2 which has 1 sub-district indicated outliers, so the next calculation process cannot be done. For this reason, the calculation process is performed on the second optimum cluster, which is 4 clusters which have 0.612 values.

Keywords: Production Results, Clustering, PAM, Davies Bouldin Index.

1. PENDAHULUAN

Pertanian adalah salah satu faktor utama dalam pembangunan ekonomi di Indonesia, karena Indonesia termasuk dalam daerah agraris. Peranan sektor pertanian sangat luas, diantaranya sebagai sektor penyerap tenaga kerja terbesar, sebagai penghasil makanan penduduk, dan sebagai penentu stabilitas harga. Pemerintah juga mengharapkan produksi pertanian tanaman pangan mengalami peningkatan tiap tahunnya (Rivani, 2010).

Kabupaten Jember adalah sebuah wilayah kabupaten yang merupakan bagian dari wilayah Provinsi Jawa Timur dengan 31 kecamatan didalamnya. Kabupaten Jember memiliki beberapa jenis komoditas yang berkembang di wilayah ini, diantaranya komoditas tanaman pangan, hortikultura, perkebunan, perikanan, peternakan, dan pertanian. Kabupaten Jember juga merupakan salah satu kabupaten yang mempunyai sektor pertanian yang bagus dan berkembang baik di wilayah karesidenan Besuki. Sektor pertanian adalah sektor unggulan di kabupaten Jember dengan banyaknya kecamatan yang menjadi sektor basis pada lapangan usaha tersebut. Sub sektor ini yang mencakup komoditi pangan seperti padi, jagung, ketela pohon, ketela rambat, kacang tanah, kacang hijau, kedelai, umbi-umbian, sayur-sayuran, buah – buahan dan kentang (Mytriani, 2007).

Pada tahun 2018, pertumbuhan sektor pertanian kabupaten Jember sebesar 0,08%. Selama lima tahun, pertumbuhan sektor pertanian terus menurun, hal ini berdampak pada PDRB (Produk Domestik Regional Bruto) Kabupaten Jember (Hakim, 2019). Menurut Jumantoro (2019), Ketua Forum Komunikasi Petani Jember ini menilai program mengenai penataan dan peningkatan SDM di sektor pertanian di kabupaten Jember terbilang gagal. Hal ini mengakibatkan banyak potensi pertanian yang ditinggalkan. Seperti sektor irigasi, budidaya dan pemasaran ditinggalkan yang mengakibatkan proses pelaksanaan kebijakan di sektor pertanian tidak tepat

sehingga kondisi pertanian semakin terpinggirkan atau menurun.

Guna menunjang program pemerintah dalam meningkatkan hasil produksi pertanian di kabupaten Jember, maka dilakukan pengelompokan kecamatan di kabupaten Jember berdasarkan hasil produksi komoditas unggulan menurut BPS (Badan Pusat Statistik) tahun 2019. Pengelompokan kecamatan di kabupaten Jember perlu dilakukan sebagai saran dan perencanaan untuk program – program pemerintah dalam bidang pertanian agar bantuan yang diberikan pemerintah lebih tepat sasaran.

Pada penelitian ini dibahas tentang pemanfaatan algoritma *Partitioning Around Medoids* (PAM) untuk mengelompokkan kecamatan di kabupaten Jember berdasarkan kemiripan karakteristik kecamatan yang ditinjau dari lima komponen hasil produksi komoditas unggulan di setiap kecamatan pada kabupaten Jember. Komponen tersebut adalah hasil produksi komoditas unggulan pada tahun 2018 yang dilaporkan dalam data BPS (Badan Pusat Statistik) kabupaten Jember, BPS menyebutnya data laporan kecamatan dalam angka 2019 di kabupaten Jember yang meliputi padi, jagung, pisang, pepaya dan cabe rawit. Berdasarkan latar belakang tersebut, maka penelitian ini berjudul “Pengelompokan kecamatan berdasarkan hasil produksi komoditas unggulan di kabupaten Jember menggunakan algoritma *Partitioning Around Medoids*”. Tujuan penelitian ini adalah untuk mengelompokkan kecamatan di kabupaten Jember berdasarkan lima komponen tersebut.

2. TINJAUAN PUSTAKA

2.1 Data Mining

Data mining merupakan proses yang menggunakan teknik statistik, matematika, kecerdasan buatan, dan *machine learning* untuk mengekstraksi dan mengidentifikasi informasi yang bermanfaat dan pengetahuan yang terkait dari berbagai *database* besar (Kusrini dan Luthfi, 2009). *Data mining* adalah analisis dari peninjauan

kumpulan data untuk menemukan hubungan tidak diduga dan meringkas data dengan cara yang berbeda dengan sebelumnya, yang dapat dipahami dengan bermanfaat bagi pemilik data (Larose, 2005).

Sehingga pengertian dari *data mining* adalah proses menganalisa data dengan menggunakan pencarian pola data yang tidak diketahui atau tidak dapat diperkirakan sebelumnya.

2.2 Clustering

Clustering (pengelompokan data) mempertimbangkan sebuah pendekatan penting untuk mencari kesamaan dalam data dan menempatkan data yang sama ke dalam kelompok-kelompok. *Clustering* membagi kumpulan data ke dalam beberapa kelompok dimana kesamaan dalam sebuah kelompok adalah lebih besar dari pada kelompok-kelompok yang lain (RuiXu dan Donald, 2009). Tujuan dari *clustering* yaitu untuk meminimalkan jarak di dalam *cluster* dan memaksimalkan jarak antar *cluster*.

Clustering merupakan suatu proses pengelompokan *record*, observasi, atau mengelompokkan kelas yang memiliki kesamaan objek (Larose, 2006). Sehingga definisi dari *clustering* yaitu suatu kumpulan dari keseluruhan pengelompokan. Perbedaan *clustering* dengan klasifikasi yaitu tidak adanya variabel target dalam melakukan pengelompokan data pada proses *clustering*.

2.3 Algoritma Partitioning Around Medoids

Algoritma *Partitioning Around Medoids* menggunakan metode partisi *clustering* untuk mengelompokkan sekumpulan n objek menjadi sejumlah k *cluster*. Algoritma *Partitioning Around Medoids* (PAM) atau disebut dengan *K-Medoids* merupakan algoritma pengelompokan yang berkaitan dengan algoritma *K-Means* dan algoritma *medoidshift*. Algoritma *Partitioning Around Medoids* (PAM) dikembangkan oleh Leonard Kaufman dan Peter J. Rousseeuw. Algoritma ini sangat mirip dengan algoritma

K-Means. Perbedaan dari kedua algoritma ini adalah algoritma PAM menggunakan objek sebagai perwakilan (*medoid*) dalam pusat *cluster* untuk setiap *cluster*, sedangkan *K-Means* menggunakan nilai rata-rata (*mean*) sebagai pusat *cluster* (Kaur dkk, 2014).

Algoritma *Partitioning Around Medoids* memiliki kelebihan untuk mengatasi kelemahan dari algoritma *K-Means* yang sensitif terhadap *noise* atau *outlier*, dan objek dengan nilai yang besar yang memungkinkan menyimpang dari distribusi data, sedangkan kelebihan lainnya yaitu hasil proses *clustering* tidak bergantung pada urutan masuk dataset (Furqon dkk, 2015). Langkah-langkah dari algoritma PAM (Han and Kamber, 2006), yaitu:

- 1) Secara acak pilih k objek pada sekumpulan n objek sebagai *medoids*.
- 2) Ulangi langkah 3 hingga langkah 6
- 3) Tempatkan objek *non-medoids* ke dalam *cluster* yang paling dekat dengan *medoids*.
- 4) Secara acak pilih o_{random} : sebuah objek *non-medoids*.
- 5) Hitung total *cost*, S (selisih), dari pertukaran *medoids* o_j dengan o_{random} . Nilai total biaya atau *cost* dinyatakan sebagai berikut:

$$Total\ Cost = \sum \sqrt{\sum_{k=1}^n (x_k - y_k)^2} \quad (1)$$
 Dengan:
 n = jumlah suatu data
 k = indeks data
 x_k = nilai atribut ke- k dari x
 y_k = nilai atribut ke- k dari y
 Nilai S dinyatakan dalam persamaan:
 $S = Total\ Cost\ baru - Total\ Cost\ Lama \quad (2)$
 Dengan:
 S = Selisih
 $Total\ Cost\ Baru$ = jumlah *cost non-medoids*
 $Total\ Cost\ Lama$ = jumlah *cost medoids*
- 6) Jika $S < 0$ maka tukar o_j dengan o_{random} untuk membentuk sekumpulan k objek baru sebagai *medoids*.
- 7) Hingga tidak ada perubahan.

2.4 Davies Bouldin Indeks

Davies Bouldin Index (DBI) merupakan cara validasi *cluster* yang ditemukan oleh D.L.Davies. DBI adalah fungsi rasio dari jumlah distribusi didalam *cluster* untuk pemisahan antar *cluster*. Dalam penelitian ini, DBI digunakan untuk melakukan validasi data pada setiap *cluster*. Adapun langkah-langkah untuk menghitung Davies Bouldin Index (Sujacka, 2019), yaitu:

1. Menghitung *Sum of Square Within cluster* (SSW).

Sum of Square Within cluster (SSW) merupakan persamaan yang digunakan untuk mengetahui matriks kohesi dalam sebuah *cluster* ke-*i* yang dirumuskan sebagai berikut:

$$SSW_i = \frac{1}{m_i} \sum_{j=i}^{m_i} d(x_j C_i) \quad (3)$$

Dengan:

m_i = jumlah data dalam *cluster* ke-*i*

C_i = *centroid cluster* ke-*i*

$d(x_j C_j)$ = merupakan jarak setiap data terhadap *centroid*

2. Menghitung *Sum of Square Between cluster* (SSB).

Sum of Square Between cluster (SSB) merupakan persamaan yang digunakan untuk mengetahui separasi antar *cluster* yang dihitung menggunakan persamaan sebagai berikut:

$$SSB_{i,j} = \|C_1 - C_2\|^2 \quad (4)$$

3. Menghitung jumlah *ratio* (rasio).

Setelah nilai SSW dan SSB diperoleh, kemudian melakukan pengukuran rasio (R_{ij}) untuk mengetahui nilai perbandingan antara *cluster* ke-*i* dan *cluster* ke-*j*. nilai rasio dihitung menggunakan persamaan sebagai berikut:

$$R_{ij} = \frac{SSW_i + SSW_j}{SSB_{ij}} \quad (5)$$

4. Menghitung nilai *Davies Bouldin Index* (DBI).

Nilai rasio yang diperoleh tersebut digunakan untuk mencari nilai *Davies Bouldin Indeks* (DBI) dengan menggunakan persamaan berikut:

$$DBI = \frac{1}{k} \sum_{i=1}^k \max_{i \neq j} (R_{ij}) \quad (6)$$

Dengan:

k = jumlah *cluster* yang ditentukan

R_{ij} = rasio *cluster* ke-*i* dan *cluster* ke-*j*

Menurut Salazar, dkk (dalam Hermadi, dkk.) Skema clustering yang optimal menurut index pengukuran Davies Bouldin adalah *cluster* yang memiliki nilai indeks terkecil atau minimal.

2.5 Metode Elbow

Metode *elbow* merupakan suatu metode penentuan jumlah *cluster* optimum untuk menghasilkan suatu informasi dengan cara melihat persentase hasil perbandingan antara jumlah *cluster* yang akan membentuk siku pada suatu titik.

Dalam penelitian ini, metode *elbow* juga digunakan untuk melakukan validasi data pada setiap *cluster*. Untuk mendapatkan perbandingannya adalah dengan menghitung *Sum of Square Error* (SSE) dari masing – masing nilai *cluster*. Karena semakin besar jumlah nilai *cluster* K , maka nilai SSE akan semakin kecil. Rumus SSE sesuai dengan persamaan berikut:

$$SSE = \sum_{k=1}^k \sum_{x_i \in S_k} |X_i - C_k|_2^2 \quad (7)$$

Dengan:

X_i = Fitur atau atribut data ke – *i*

C_k = Fitur atau atribut titik pusat *Cluster* ke – *i*

2.6 Rapidminer

Menurut Dennis, dkk (2013) dalam Asistiyasari dan Baidawi (2017) *rapidminer* merupakan perangkat lunak yang bersifat terbuka (*open source*). *Rapidminer* adalah sebuah solusi untuk melakukan analisis terhadap data mining, text mining dan analisis prediksi. *Rapidminer* menggunakan berbagai teknik deskriptif dan prediksi dalam memberikan wawasan kepada pengguna sehingga dapat membuat keputusan yang paling baik. *Rapidminer* memiliki beberapa sifat sebagai berikut.

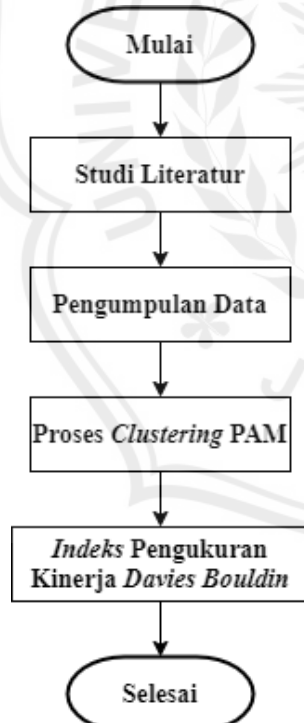
1. Ditulis dengan bahasa pemrograman *java* sehingga dapat dijalankan di berbagai sistem operasi.
2. Proses penemuan pengetahuan dimodelkan sebagai operator *trees*.

3. Representasi XML internal untuk memastikan format standar pertukaran data.
4. Bahasa *scripting* memungkinkan untuk eksperimen skala besar dan otomatisasi eksperimen.
5. Konsep *multilayer* untuk menjamin tampilan data yang efisien dan menjamin penanganan data.
6. Memiliki GUI, *command line mode*, dan *java API* yang dapat dipanggil dari program lain.

3. METODOLOGI

3.1 Tahapan Penelitian

Adapun metode penelitian yang dilakukan dalam pemanfaatan algoritma Partitioning Around Medoids (PAM) untuk mengelompokkan kecamatan di kabupaten Jember berdasarkan index kinerja Davies Bouldin pada kasus hasil produksi komoditas unggulan, memiliki tahapan sebagai berikut:



Gambar 1. Tahapan Penelitian

3.2 Studi Literatur

Tahapan awal dari penelitian ini adalah dengan mencari dan mempelajari masalah yang akan diteliti, kemudian menentukan ruang lingkup masalah, latar belakang, dan mempelajari beberapa literatur yang berkaitan dengan permasalahan dan bagaimana mencari solusi dari masalah tersebut. Untuk mencapai tujuan yang ditentukan, maka penulis perlu mempelajari beberapa literatur yang digunakan, kemudian literatur tersebut diseleksi untuk ditentukan sebagai literatur yang akan digunakan dalam penelitian.

3.3 Pengumpulan Data

Penelitian ini menggunakan data BPS (Badan Pusat Statistik) kabupaten Jember, BPS menyebutnya data laporan kecamatan dalam angka 2019 di kabupaten Jember yang berupa data sekunder hasil produksi komoditas unggulan pada tahun 2018. Variable yang digunakan dalam pengelompokan kecamatan di kabupaten Jember berdasarkan hasil produksi komoditas unggulan meliputi padi, jagung, pisang, pepaya dan cabe rawit pada tahun 2018. Data tersebut dapat diunduh dari situs <https://jemberkab.bps.go.id/>.

3.4 Proses Clustering

Pada penelitian ini, metode analisa data yang digunakan adalah *Partitioning Around Medoids*. *Partitioning Around Medoids* merupakan teknik *cluter* atau mengelompokkan sekumpulan objek sebagai perwakilan (*medoid*) dalam pusat *cluster* untuk setiap *cluster*. Kelebihan dari PAM yaitu untuk mengatasi kelemahan dari algoritma *K-Means* yang sensitif terhadap *noise* atau *outlier*, dan objek dengan nilai yang besar yang memungkinkan menyimpang dari distribusi data.

