

Nama : Putra Satria Nagara Mahmud
NIM : 2230511130
Kelas : Teknik Informatika (5C)
Mata Kuliah : Kecerdasan Buatan

UJIAN AKHIR SEMESTER

Soal :

1. Jelaskan dengan bahasa Anda sendiri bagaimana cara kerja algoritma K-Means

Clustering dan sebutkan 3 komponen utama yang harus ada dalam perhitungan K

Means! (Minimal sitasi dari 1 sumber)

2. Berdasarkan studi kasus pada jurnal, mengapa algoritma K-Means cocok digunakan

untuk pengelompokan kelas unggulan? Jelaskan minimal 2 alasan!

3. Perhatikan data nilai 4 siswa berikut:

a. Siswa A: Matematika=80, Bahasa Indonesia=85, Bahasa Inggris=78

b. Siswa B: Matematika=85, Bahasa Indonesia=82, Bahasa Inggris=80

c. Siswa C: Matematika=75, Bahasa Indonesia=70, Bahasa Inggris=72

d. Siswa D: Matematika=78, Bahasa Indonesia=75, Bahasa Inggris=73

Jika ditentukan centroid awal:

C1 = (85,82,80) dari nilai tertinggi

C2 = (75,70,72) dari nilai terendah

Jawablah pertanyaan berikut:

a. Hitunglah jarak Euclidean Distance antara Siswa A dengan C1

b. Tentukan Siswa A masuk ke cluster mana dengan membandingkan jaraknya ke C1 dan C2

Jawaban :

1. Ada Beberapa cara kerja dalam algoritma k-means clustering, salah satu nya adalah :

● *Inisialisasi*

Inisialisasi dapat menentukan beberapa jumlah cluster “k” dan menginisialisasi centroid awal, baik secara acak atau berdasarkan nilai

maksimum atau minimum.

● Pengelompokan Data

Untuk mengelompokan sebuah data, hitung jarak dalam setiap data terhadap centroid menggunakan euclidean distance :

$$D = \sqrt{\sum (x_i - y_i)^2}$$

Diatas ini merupakan rumus euclidean distance yang akan dipakai ketika mengelompokan sebuah data terhadap jarak. Data yang dikelompokan ke cluster dengan jarak terpendek dari centroid. Dan untuk penjelasan dari rumus yang dipakai adalah :

1. “D” : Merupakan Jarak antara Data dan Centroid
2. x_j : Merupakan Koordinat (X) data dari ke- j
3. y_j : Merupakan koordinat (Y) centroid ke- j
4. n : Jumlah Dimensi dari banyak atribut atau data

● Pembaruan Centroid

Pada memperbarui centroid, centroid di hitung ulang dan sebagai posisi data dalam setiap cluster

$$v_{ij} = \frac{1}{N_i} \sum_{k=1}^{N_i} x_{kj}$$

Dan Untuk penjelasan rumusnya adalah :

1. v_{ij} : Nilai dari centroid baru untuk sebuah cluster i dan atribut j .
2. N_i : Merupakan Jumlah Data dalam Cluster i
3. x_{kj} : Nilai dari suatu data ke- k untuk atribut j

● Fungsi Objektif

Pada Fungsi ini bertujuan untuk meminimalkan beberapa fungsi objektif pada k-means, yang dimana total kuadrat jarak antara setiap sebuah data dengan beberapa centroid-nya :

$$J = \sum_{i=1}^k \sum_{j=1}^{N_i} ||x_j - v_i||^2$$

1. J : Total sebuah Fungsi Objektif
2. k : Jumlah dari cluster
3. N_i : Jumlah Data dalam Cluster I
4. $||x_j - v_i||^2$: Jarak Kuadrat antara x_j dengan centroid v_i .

● Iterasi

Iterasi merupakan sebuah proses pengulangan dalam mengelompokkan dan sekaligus pembaruan centroid sehingga tidak ada perubahan secara signifikan dalam posisi centroid.

Dan Berikut ini ada 3 Komponen utama dalam K-Means Clustering :

1. Centroid yang merupakan sebuah titik tengah (center) pada setiap cluster.
2. Data Point, yang dimana data akan dikelompokkan berdasarkan kemiripan dari sebuah data clustering
3. Metode Penghitung Jarak, pada komponen ini biasanya menggunakan euclidean distance untuk menentukan sebuah kedekatan dari satu data dengan centeroid (Titik Tengah pada Cluster)

2. Ada dua alasan mengenai pengelompokan K-Means dari kelas unggulan yang ada pada Jurnal yang diberikan :

a. Segi **Kesederhanaan dan Efisiensi** : K-Means sangat cepat dalam memproses sebuah data yang berukuran besar dan sehingga efisien untuk mengelompokkan beberapa data, pada kasus dalam jurnal, ada sebuah kasus pengelompokan 192 Siswa yang menjadi dua kelompok Unggulan dan tidak unggulan, yang dimana algoritma ini mampu mengelompokkan data siswa secara cepat dan efektif.

b. **Flexibilitas, Penentuan dan Kemampuan Identifikasi Pola dari Cluster** : Pada Algoritma K-Means ini memungkinkan pengguna untuk menemukan beberapa jumlah cluster dari satu data yang diinginkan, termasuk pada kasus ini yaitu menganalisis data siswa, dan dari segi kemampuan identifikasi terhadap pola, K-Means dapat

mengidentifikasi Beberapa pola nilai Siswa berdasarkan beberapa atribut seperti Mata Pelajaran Matematika, Ilmu Pengetahuan Alam, Bahasa Inggris yang dimana sangat relevan untuk menentukan suatu kelas unggulan.

3. Ada 3 data nilai pada siswa, yaitu :

Matematika = 80, Bahasa Indonesia = 85 dan Bahasa Inggris = 78.
dan untuk Centroid Awal terbagi menjadi 2 yaitu C1 yang mengelompokkan Nilai tertinggi seperti (85,82,80) dan C2 yang mengelompokkan Nilai Terendah seperti (75,70,72).
Dan Pertanyaannya adalah :

- A. Hitung Jarak Euclidean Distance antara Siswa dengan C1.
- B. Tentukan Cluster Siswa A Dengan Membandingkan Jarak C1 dan C2.

- Untuk Menghitung Euclidean Distance, yaitu menggunakan rumus ini :

$$D = \sqrt{\sum_{i=1}^n (x_i - y_i)^2}$$

Yang dimana, x_i merupakan koordinat Data Siswa, untuk x_1 adalah 80, x_2 adalah 85 dan untuk x_3 adalah 78.

Selanjutnya adalah y_i yang merupakan sebuah koordinat centroid ($y_1 = 85$, $y_2 = 82$, $y_3 = 80$)

Hasil dari Substitusi nilai adalah :

$$\begin{aligned} D(A, C1) &= \sqrt{(80 - 85)^2 + (85 - 82)^2 + (78 - 80)^2} \\ &= \sqrt{(-5)^2 + 3^2 + (-2)^2} \\ &= \sqrt{25 + 9 + 4} = \sqrt{38} \approx 6.16 \end{aligned}$$

Disini ditentukan bahwa Jarak Siswa ke C1 adalah 6.16

Selanjutnya adalah menghitung substitusi nilai dari C2 yang dimana isi dari data C2 adalah $y_1 = 75$, $y_2 = 70$, $y_3 = 72$

$$\begin{aligned}
 D(A, C2) &= \sqrt{(80 - 75)^2 + (85 - 70)^2 + (78 - 72)^2} \\
 &= \sqrt{(5)^2 + (15)^2 + (6)^2} \\
 &= \sqrt{25 + 225 + 36} = \sqrt{286} \approx 16.92
 \end{aligned}$$

Hasil dari substitusi nilai dari C2 adalah 16.92

- Untuk menentukan Cluster pada Suatu data yaitu mencari jarak yang lebih kecil, dan disini jarak yang lebih kecil adalah C1 dengan hasil 6.16, maka Siswa akan masuk ke Kelas Cluster C1