NPM: 5210411144

Nama: Muhammad Izza Iqbal

1. Anda diberikan sebuah **data training** untuk mengklasifikasikan jumlah takaran dan jumlah sajian yang menghasilkan jumlah total energi yang tergolong jelek atau bagus. Berikut ini deskripsi atribut dari dataset yang dipakai yang terdiri dari nama atribut, dan tipe datanya:

a. **X1** | Takaran Saji (gram) | tipe: integer

b. **X2** | Jumlah Saji (persajian) | tipe: integer

c. **X3** | Total Energi (kalori) | tipe: integer

d. **Y** | Klasifikasi (Class) | tipe: diskrit (jelek/bagus:0/1)

| **X1** | **X2** | **X3** | **Y** |
| --- | --- | --- | --- |
| 40 | 5 | 60 | Jelek |
| 50 | 8 | 40 | Bagus |
| 50 | 7 | 30 | Jelek |
| 70 | 4 | 60 | Bagus |
| 80 | 4 | 80 | Bagus |
| 60 | 6 | 60 | Bagus |

Berbekal dataset (Data Training) tersebut, tuliskanlah langkah-langkah Anda untuk mengolah data pada tahapan **preproccessing**., dan jika kemudian diberikan sebuah data baru sebagai berikut:

| **X1** | **X2** | **X3** | **Y** |
| --- | --- | --- | --- |
| 50 | 3 | 40 | ? |

Maka, apakah prediksi nilai pada kolom ‘**Y**’ untuk tiap data menggunakan algoritma KNN (K=4 dan K=3)? Lakukan **perhitungan** **manual** dengan menguraikan jawaban Anda langkah demi langkah hingga mendapatkan hasil akhir yang diminta!

1. Sebuah Project menerapkan algoritma ML (*Machine Learning*) untuk memprediksi apakah mahasiswa berpotensi DO atau tidak. Setelah algoritma dijalankan hasilnya adalah:

| **NIM (NPM)** | **STATUS (DO) SEBENARNYA** | **HASIL**  **PREDIKSI** |
| --- | --- | --- |
| 001 | TIDAK | TIDAK |
| 002 | TIDAK | TIDAK |
| 003 | TIDAK | TIDAK |
| 004 | TIDAK | YA |
| 005 | TIDAK | YA |
| 006 | YA | TIDAK |
| 007 | YA | YA |
| 008 | YA | YA |
| 009 | YA | YA |
| 010 | YA | YA |

Berdasarkan data pada table hasil menjalankan algoritma diatas, maka bagaimana mengukur Kinerja Algoritma Klasifikasi tersebut dengan Confusion Matrix terkait hasil **perhitungan manual** dari:

* **Precision** yaitu tingkat ketepatan antara informasi yang diminta oleh pengguna dengan jawaban yang diberikan oleh sistem.
* **Recall** yaitu tingkat keberhasilan sistem dalam menemukan kembali sebuah informasi.
* **Accuracy** didefinisikan sebagai tingkat kedekatan antara nilai prediksi dengan nilai actual (sebenarnya).

Jawab

1. Perhitungan manual dengan k=4 menggunakan metode Euclidean:

Hitung jarak antara data uji dan data latih:

1. Jarak dengan data ke-1: √(40-50)² + (5-3)² + (60-40)² = 22.45
2. Jarak dengan data ke-2: √(50-50)² + (8-3)² + (40-40)² = 5
3. Jarak dengan data ke-3: √(50-50)² + (7-3)² + (30-40)² = 10.77
4. Jarak dengan data ke-4: √(70-50)² + (4-3)² + (60-40)² = 28.3
5. Jarak dengan data ke-5: √(80-50)² + (4-3)² + (80-40)² = 50.01
6. Jarak dengan data ke-5: √(60-50)² + (6-3)² + (60-40)² = 22.56

Untuk knn=4, data terdekat adalah:

1. Data ke-2 dengan jarak 5
2. Data ke-3 dengan jarak 10.77
3. Data ke-1 dengan jarak 22.45
4. Data ke-6 dengan jarak 22.56

Perhitungan manual dengan k=3 menggunakan metode Euclidean:

* 1. Jarak dengan data ke-1: √(40-50)² + (5-3)² + (60-40)² = 22.45
  2. Jarak dengan data ke-2: √(50-50)² + (8-3)² + (40-40)² = 5
  3. Jarak dengan data ke-3: √(50-50)² + (7-3)² + (30-40)² = 10.77
  4. Jarak dengan data ke-4: √(70-50)² + (4-3)² + (60-40)² = 28.3
  5. Jarak dengan data ke-5: √(80-50)² + (4-3)² + (80-40)² = 50.01
  6. Jarak dengan data ke-5: √(60-50)² + (6-3)² + (60-40)² = 22.56

Untuk knn=3, data terdekat adalah:

1. Data ke-2 dengan jarak 5
2. Data ke-3 dengan jarak 10.77
3. Data ke-1 dengan jarak 22.45

Berdasarkan hasil perhitungan knn, dengan knn=4 dan knn=3, dapat dilihat bahwa data baru 50 3 40 termasuk ke dalam kategori Jelek. Hal ini karena dari 4 atau 3 data terdekat, sebagian besar atau semua data tersebut termasuk dalam kategori Jelek. Oleh karena itu, dengan menggunakan algoritma knn, dapat disimpulkan bahwa data baru tersebut termasuk ke dalam kategori Jelek.

1. Tabel confusion matrix

|  | Sebenarnya Ya | Sebenarnya Tidak |
| --- | --- | --- |
| Prediksi Ya | 4 | 1 |
| Prediksi Tidak | 2 | 3 |

Dari tabel tersebut, kita dapat menghitung Precision, Recall, dan Accuracy dengan rumus-rumus berikut:

TP = True Positive

TN = True Negative

FP = False Positive

FN = False Negative

1. Precision = TP / (TP + FP)  
   Precision mengukur seberapa banyak prediksi "Ya" yang benar-benar benar dari semua prediksi "Ya" yang dilakukan oleh sistem.  
   Precision = 4 / (4 + 1) = 0.8
2. Recall = TP / (TP + FN)  
   Recall mengukur seberapa banyak nilai "Ya" yang berhasil ditemukan oleh sistem dari semua nilai "Ya" yang sebenarnya ada.  
   Recall = 4 / (4 + 2) = 0.67
3. Accuracy = (TP + TN) / (TP + FP + TN + FN)  
   Accuracy mengukur seberapa banyak prediksi yang benar dari semua prediksi yang dilakukan oleh sistem.  
   Accuracy = ( 4 + 3) / (4 + 1 + 3 + 2) = 0.7

Dari hasil perhitungan tersebut, dapat disimpulkan bahwa algoritma klasifikasi yang digunakan memiliki Precision sebesar 0.8, Recall sebesar 0.67, dan Accuracy sebesar 0.7.