Algoritma K-Nearest Neighbor adalah untuk memprediksi nilai kelas baru dari dataset yang telah mempunyai nilai kelas sebelumnya.

2

keterangan :

* d(x,y) adalah jarak antara data x dan y
* Xi adalah data testing ke-*i*
* Yi adalah data training ke-*i*

Langkah-langkah algoritma K-NN :

1. Tentukan nilai parameter *k*(Nilai *k* dipilih secara manual).
2. Hitung jarak antara data training dan data testing.
3. Urutkan data training berdasarkan jarak terkecil.
4. Menetapkan kelas, dimana kelas yang dipilih adalah kelas dengan jumlah nilai *k* terbanyak pada data testing.

Dataset dibagi menjadi 2 bagian yaitu data training dan data testing. Data training adalah data yang sudah mempunyai kelas. Data testing adalah data yang akan dicari nilai kelasnya dari data training yang sudah mempunyai nilai kelas. Data training akan membentuk suatu model/pola/pengetahuan yang akan menjadi acuan dalan data testing. Data testing digunakan untuk pengukuran evaluasi algoritma.

Data training

|  |  |  |
| --- | --- | --- |
| X | Y | Kelas |
| 1,2 | 2,3 | A |
| 2,5 | 4,6 | A |
| 4 | 1 | B |
| 5,6 | 1,2 | B |
| 6,0 | 3,5 | B |

Data Testing

|  |  |  |
| --- | --- | --- |
| X | Y | Kelas |
| 3 | 2 | ? |

1. Menentukan Nilai Parameter *k*

Nilai parameter *k* pada kasus ini adalah 3. Dapat disimpulkan data testing yang akan dicari kelasnya berdasarkan 3 data training terdekat dengan data testing

1. Hitung Jarak antara Data Training dan Data Testing
2. Urutkan Data testing berdasarkan jarak terkecil

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Data training |  |  | Data testing |  |  |
| X | Y | Kelas | X | Y | Testing |
| 4 | 1 | B | 3 | 2 | 1,414 |
| 1,2 | 2,3 | A | 3 | 2 | 1,825 |
| 2,5 | 4,6 | A | 3 | 2 | 2,648 |
| 5,6 | 1,2 | B | 3 | 2 | 2,721 |
| 6 | 3,5 | B | 3 | 2 | 3,354 |

1. Menetapkan kelas

Pilih 3 data terdekat (karena k =3) antara data training dan data testing. Tiga data dengan jarak terdekat antara data training dan data testing.

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Data training |  |  | Data testing |  |  |
| X | Y | Kelas | X | Y | Testing |
| 4 | 1 | B | 3 | 2 | 1,414 |
| 1,2 | 2,3 | A | 3 | 2 | 1,825 |
| 2,5 | 4,6 | A | 3 | 2 | 2,648 |

Dapat disimpulkan baha kelas A lebih banyak dari pada kelas B denga proporsi kelas A sebesar 67%, sedangkan kelas B sebesar 33%. Dapat disimpulkan bahwa data testing dengan nilai X = 3 dan Y = 2 memiliki kelas A.