FK-NN

Seperti halnya pada teori fuzzy, sebuah data mempunyai nilai keanggotaan pada setiap kelas [2][3], yang artinya sebuah data bisa dimiliki oleh kelas yang berbeda dengan nilai derajat keanggotaan dalam interval [0,1]. Teori himpunan fuzzy men-generalisasi teori K-NN klasik dengan mendefinisikan nilai keanggotaan sebuah data pada masing-masing kelas. Formula yang digunakan [2][3]:

Dimana *u*(*x,yi*) adalah nilai keanggotaan data *x* ke kelas *ci*, *K* adalah jumlah tetangga terdekat yang digunakan, *u*(*xk,yi*) adalah nilai keanggotaan data tetangga dalam *K* tetangga pada kelas *yi*, nilainya 1 jika data latih *xk* milik kelas *ci* atau 0 jika bukan milik kelas *ci*, untuk *d*(*x, xk*) adalah jarak dari data *x* ke data *xk* dalam *K* tetangga terdekat, m adalah bobot pangkat (*weight exponent*) yang besarnya *m* > 1.

Nilai keanggotaan suatu data pada kelas sangat dipengaruhi oleh jarak data itu ke tetangga terdekatnya, semakin dekat ke tetangganya maka semakin besar nilai keanggotaan data tersebut pada kelas tetangganya, begitu pula sebaliknya. Jarak tersebut diukur dengan N dimensi (fitur) data.

Pengukuran jarak (ketidakmiripan) dua data yang digunakan dalam FK-NN digeneralisasi dengan [2][3]:

Dimana N adalah dimensi (jumlah fitur) data. Untuk p adalah penentu jarak yang digunakan, jika p=1 maka jarak yang digunakan adalah Manhattan, jika p=2 maka jarak yang digunakan adalah Euclidean, jika p=∞ maka jarak yang digunakan adalah Chebyshev.

Meskipun FK-NN menggunakan nilai keanggotaan untuk menyatakan keanggotaan data pada setiap kelas, tetapi untuk memberikan keluaran akhir, FK-NN tetap harus memberikan kelas akhir hasil prediksi, untuk keperluan ini, FKNN memilih kelas dengan nilai keanggotaan terbesar pada data tersebut.

Algoritma Prediksi dengan FK-NN

1. Normalisasikan data menggunakan nilai terbesar dan terkecil data pada setiap fitur.
2. Cari K tetangga terdekat untuk data uji x menggunakan persamaan (2).
3. Hitung nilai keanggotaan u(x, yi) menggunakan persamaan (1) untuk setiap i, dimana 1 ≤ i ≤ C.
4. Ambil nilai terbesar v = u(x, yi) untuk semua 1 ≤ i ≤ C, C adalah jumlah kelas.
5. Berikan label kelas v ke data uji x yaitu yi.

K-NN

Pada algoritma K-NN , data berdimensi N, dapat dihitung jara dari data tersebut ke data yang lain, nilai jarak ini yang digunakan sebagai nilai kedekatan/ketidakmiripan antara data uji dengan data latih. Nilai K pada K-NN berarti K-data terdekat dari data uji.

Pada algoritma K-NN, sebuah data uji z = (x’,y’), dimana x’ ada vektor/atribut data uji, sedangkan y’ adalah label kelas data uji yang belum diketahui, kemudian menghitung jarak (atau kemiripan) data uji ke setiap data latih d(x’,x), kemudian mengambil K tetangga terdekat pertama dalam Dz. Setelah itu dihitung jumlah data yang mengikuti kelas yang ada dari K tetangga tersebut. Kelas dengan data terbanyak yang mengikutinya menjadi kelas pemenang yang diberikan sebagai label kelas pada data uji y’.

Algoritma Prediksi dengan K-NN

1. z = (x’,y’), adalah data uji dengan vektor x’ dan label kelas y’ yang belum diketahui.
2. Hitung jarak d(x’,x), jarak diantara data uji z ke setiap vektor data latih, simpan dalam D
3. Pilih Dz D, yaitu K tetangga terdekat dari z

Salah satu masalah yang dihadapi K-NN adalah pemilihan nilai K yang sulit, cara voting mayoritas dari K-tetangga untuk nilai K yang besar bisa mengakibatkan distorsi data yang besar, jika K terlalu kecil bisa menyebabkan algoritma terlalu sensitif terhadap noise. Meskipun begitu, sebenarnya penentuan nilai K yang terbaik juga dapat dilakukan dengan menggunakan teknik crossvalidation (Tan, 2006).