

REKOMENDASI PEMINATAN STUDI MENGGUNAKAN METODE *K-NEAREST NEIGHBOR* DAN *FUZZY C-MEANS*

Taysa Natalia¹⁾ Tri Sutrisno²⁾

¹⁾ Teknik Informatika Universitas Tarumanagara Jakarta
Jl. S. Parman No 1, Jakarta 11440 Indonesia
email : Tasyanatalia95@gmail.com

²⁾ FMIPA Universitas Gajah Mada Yogyakarta
Jl. Bulaksumur, Yogyakarta 55281 Indonesia
email : Tris@fti.untar.ac.id

ABSTRAK

Information Technology Advancements has become inseparable things to human's life in this modern era, technological advancement has already give positive progress and additional value in educational segment in Indonesia.

Technology have role to help undergraduate and college student choose specialization in their field of study respectively, which utilize K-Nearest Neighbor and Fuzzy C-Means Clustering calculation method.

Clustering consist of 5 distinct specialization which are DA, IS, R, GD, ACS. Data used in evaluation process are 97 Undergraduate Student with 34 Subject which already decided and 25 Student's Data to be testing subject.

Kata Kunci

Clustering, Fuzzy C-Means, K-Nearest Neighbor

1. Pendahuluan

Indonesia adalah negara berkembang dimana infrastruktur pendidikan masih minim untuk dikembangkan sehingga pendidikan yang didapat masih terbatas di sebagian wilayah di Indonesia. Untuk kota-kota besar di Indonesia seperti Jakarta, Surabaya, Medan, dan kota-kota besar lainnya. Dengan berkembangnya teknologi saat ini membuat sebagian besar pelajar di Indonesia menyadari akan pentingnya kemajuan teknologi dan mendorong minat mereka untuk mempelajari lebih dalam dan rinci mengenai teknologi informasi, hal tersebut dapat dilihat dari perkembangan institusi-institusi Pendidikan yang memasukan program studi teknologi informatika sebagai program studi andalan dengan jumlah mahasiswa yang memilih program studi teknologi informatika cukup banyak. Algoritma yang digunakan untuk *clustering* dalam penelitian ini adalah *K-Nearest Neighbor* dan *Fuzzy C-Means*. Tujuan dari perancangan yang disusun ini adalah untuk mengelompokkan data mahasiswa sesuai dengan peminatan yang akan diambil dan sesuai dengan

kemampuan mahasiswa tersebut berdasarkan nilai yang sebelumnya sudah diperoleh.

2. *K-Nearest Neighbor* dan *Fuzzy C-Means*

2.1 Algoritma *K-Nearest Neighbor*

Algoritma *k-nearest neighbor* atau KNN adalah sebuah metode untuk melakukan klasifikasi terhadap objek berdasarkan data pembelajaran yang jaraknya paling dekat dengan objek tersebut.

Secara umum algoritma *K-Nearest Neighbor* adalah:

1. Tentukan parameter K = jumlah banyaknya tetangga terdekat.
2. Hitung jarak antara data baru dan semua data yang ada di data training.
3. Urutkan jarak tersebut dan tentukan tetangga mana yang terdekat.
4. Tentukan kategori dari tetangga terdekat.
5. Gunakan kategori mayoritas yang sederhana dari tetangga yang terdekat tersebut sebagai nilai prediksi dari data yang baru.

Rumus untuk menghitung jarak yang digunakan adalah *Euclidean Distance* [1] :

$$d(X_j, C_j) = \sqrt{\sum_{j=1}^n (X_j - C_j)^2} \quad (1)$$

Keterangan:

d = jarak;
j = banyaknya data;
c = centroid;
x = data.

2.2 Algoritma Fuzzy C-Means

Fuzzy C-means Clustering (FCM), atau dikenal juga sebagai Fuzzy ISODATA, merupakan salah satu metode clustering yang merupakan bagian dari metode Hard K-Means. FCM menggunakan model pengelompokan fuzzy sehingga data dapat menjadi anggota dari semua kelas atau cluster terbentuk dengan derajat atau tingkat keanggotaan yang berbeda antara 0 hingga 1. Tingkat keberadaan data dalam suatu kelas atau cluster ditentukan oleh derajat keanggotaannya. Teknik ini pertama kali diperkenalkan oleh Jim Bezdek pada tahun 1981. Pada kondisi awal, pusat *cluster* ini masih belum akurat. Tiap-tiap data memiliki derajat keanggotaan untuk tiap-tiap *cluster*. Dengan cara memperbaiki pusat *cluster* dan nilai keanggotaan tiap-tiap data secara berulang, maka akan dapat dilihat bahwa pusat *cluster* akan bergerak menuju lokasi yang tepat. Algoritma pada *Fuzzy C-Means Clustering* adalah sebagai berikut:

1. Tentukan:
 - a. Matriks X berukuran $n \times m$
 n = jumlah data yang akan di-*cluster*
 m = jumlah variable (atribut);
 - b. Jumlah cluster yang akan dibentuk = c
 c harus bernilai ≥ 2 ;
 - c. Pangkat (pembobot) = w
 w bernilai ≥ 2 ;
 - d. Maksimum iterasi = $MaxIter$;
 - e. Kondisi iterasi berhenti (error terkecil) = ξ ;
 - f. Iterasi awal, $t = 1$ dan $\Delta = 1$;
2. Bangkitkan bilangan random μ_{ik} , dimana $i = 1, 2, \dots, n$; $k = 1, 2, \dots, c$; sebagai elemen-elemen matriks partisi awal U.

$$Q_i = \sum_{k=1}^c \mu_{ik} \quad (2)$$

Dengan $j=1, 2, \dots, n$

3. Hitung pusat *cluster* ke- k : V_{kj} , untuk setiap *cluster*:

$$V_{kj} = \frac{\sum_{i=1}^n ((\mu_{ik})^w * X_{ij})}{\sum_{i=1}^n (\mu_{ik})^w} \quad (3)$$

Keterangan:

V = Pusat *cluster*

$k = 1, 2, \dots, c$

$j = 1, 2, \dots, m$

$i = 1, 2, \dots, n$

4. Hitung fungsi objektif pada iterasi ke- t , P_t :
5. Hitung perubahan derajat keanggotaan yang baru pada matriks partisi:

Keterangan:

$i = 1, 2, \dots, n$

$k = 1, 2, \dots, c$

6. Cek kondisi:

- a. Jika: $(|P_t - P_{t-1}| < \xi)$ atau $(t > MaxIter)$, maka berhenti;
- b. Jika tidak, $t = t + 1$, ulangi langkah ke-3.

Jika salah satu kondisi 6a terpehuni, maka proses komputasi berhenti. Proses yang dihasilkan sebelumnya adalah pusat *cluster* pada algoritma 3 dan derajat keanggotaan yang baru. Penentuan data untuk dimasukkan ke dalam sebuah anggota *cluster* adalah berdasarkan besarnya derajat keanggotaan pada *cluster* tersebut. Nilai pada derajat keanggotaan dibandingkan. Nilai yang terbesar akan masuk di dalam anggota *cluster* tersebut. Matriks pusat *cluster* (V) yang didapat merepresentasikan nilai *centroid* pada *cluster*.

3. Hasil Percobaan

Data yang digunakan pada proses evaluasi adalah 97 data mahasiswa dengan 30 mata

- a. Tabel 2 Hasil metode *Fuzzy C-Means* Terdapat 9 Data pada tabel Pengujian *K-Nearest Neighbor* yang memiliki Hasil sesuai dengan peminatan awal dengan persentase **36%**.
- b. Terdapat 10 Data pada tabel Pengujian *Fuzzy C-Means* dengan 7 Mata Kuliah yang berkorelasi yang memiliki Hasil sesuai dengan peminatan awal dengan persentase **40%**.
- c. Terdapat 5 Data pada tabel Pengujian *Fuzzy C-Means* dengan 12 Mata Kuliah yang berkorelasi yang memiliki Hasil sesuai dengan peminatan awal dengan persentase **20%**.
- d. Terdapat 7 Data pada tabel Pengujian *Fuzzy C-Means* dengan 26 Mata Kuliah yang berkorelasi yang memiliki Hasil sesuai dengan peminatan awal dengan persentase **28%**.

4. Kesimpulan

Kesimpulan hasil penelitian “Rekomendasi Peminatan Studi Menggunakan Metode *K-Nearest Neighbor* dan *Fuzzy C-Means*” adalah sebagai berikut:

1. Hasil *clustering* pada kedua metode menunjukkan hasil yang relatif berbeda, dikarenakan adanya perbedaan pendekatan pada kedua metode, pada metode *K-Nearest Neighbor* cenderung merujuk pada peminatan yang telah diambil oleh mahasiswa sebelumnya, sedangkan pada metode *Fuzzy C-Means* peminatan dipengaruhi oleh mata kuliah yang sudah di *clusterkan* berdasarkan korelasinya terhadap peminatan tertentu.

2. Pada metode *Fuzzy C-Means* hasil clustering akan ditemukan lebih cepat jika semua data mahasiswa memiliki kelengkapan transkrip nilai yang baik.
3. Metode *K-Nearest Neighbor* dapat menentukan peminata mahasiswa uji walaupun transkrip nilai mahasiswa terbatas tanpa mempengaruhi kecepatan pemrosesan peminatan dengan konsekuensi hasil *Euclidean distance* cenderung tinggi.
4. Berdasarkan hasil pengujian pada metode *Fuzzy C-Means* bisa didapat kesimpulan bahwa korelasi antara peminatan dengan 7 mata kuliah memiliki hasil pengujian paling besar, dikarenakan mata kuliah yang saling beririsan pada setiap peminatan tidak lebih dari 1 mata kuliah saja.
5. Pengujian metode *Fuzzy C-Means* untuk 12 dan 26 mata kuliah yang berkorelasi memiliki tingkat ketepatan yang masih rendah karena banyaknya mata kuliah yang beririsan pada setiap peminatan.

REFERENSI

- [1] Weisstein, Eric W, Distance, <http://mathworld.wolfram.com/Distance.html>, 24 September 2016. Li, Shujun., Zheng, Xuan., Mei 2002, "Cryptanalysis of a Chaotic Image Encryption Method", Proceeding IEEE – ISACS, Vol.2, Scottsdale -Arizona.
- Sri Kusumadewi dan Sri Hartanti, *Neuro-Fuzzy: Integrasi Sistem Fuzzy & Jaringan Syaraf*, (Jogyakarta: Graha Ilmu, 2006), h.28.

Taysa Natalia merupakan mahasiswa tingkat akhir Program Studi Teknik Informatika Fakultas Teknologi Informasi Universitas Tarumanagara, Jakarta.

Tri Sutrisno, memperoleh gelar S.Si dari Universitas Diponegoro kemudian memperoleh gelar M.Sc dari Universitas Gajah Mada, Saat ini masih aktif sebagai dosen tetap Program Studi Teknik Informatika Fakultas Teknologi Informasi Universitas Tarumanagara, Jakarta