

KOMPARASI ALGORITMA FUZZY *C-MEANS* DAN ALGORITMA *K-NEAREST NEIGHBOR* DALAM PENGELOMPOKAN ROMBONGAN BELAJAR SISWA BARU (STUDI KASUS: SISWA BARU MADRASAH ALIYAH NEGERI 01 KOTA BENGKULU)

Ernawati¹, Budiman², Diyah Puspitaningrum³

Program Studi Teknik Informatika, Fakultas Teknik, Universitas Bengkulu.

Jl. Raya Kandang Limun, Bengkulu. Telp. (0736) 21170

¹w_ier_na@yahoo.com, ²budiman_st@yahoo.co.id, ³diyahpuspitaningrum@gmail.com

ABSTRAK

Setiap tahun ajaran baru Madrasah Aliyah 01 Model, Bengkulu akan mengadakan penerimaan siswa baru. Kegiatan ini menimbulkan masalah dalam kelompok studi mahasiswa baru ke dalam kelompok belajar yang sama. Proses pengelompokan ini dikenal sebagai *clustering*. Beberapa algoritma *clustering* yang digunakan adalah algoritma *fuzzy c-means* dan *k-nearest neighbor*. Untuk memenuhi kebutuhan akan aplikasi *clustering* yang dapat digunakan dalam memecahkan masalah tersebut, penelitian ini bertujuan membangun aplikasi pengelompokan kelas bagi siswa baru dan juga untuk membandingkan kinerja algoritma *fuzzy c-means* dan *k-nearest neighbor*. Penelitian ini menggunakan NetBeans IDE 6.0.1 sebagai perangkat lunak, metode pengembangan sekuensial linier, perangkat lunak analisis dan desain menggunakan *Unified Modeling Language* (UML). Hasil penelitian ini menunjukkan bahwa sistem ini mampu mengklasifikasikan kelompok belajar (RomBel) siswa secara sama, dengan sedikit perbedaan antara kelompok yang satu dengan yang lainnya dalam nilai akademis. Selain algoritma *fuzzy c-means* lebih baik daripada algoritma *k-nearest neighbor* dalam hal pengelompokan, tetapi algoritma *k-nearest neighbor* memiliki waktu CPU yang memproses jauh lebih cepat daripada algoritma *fuzzy c-means*.

Kata Kunci: *Fuzzy C-Means, K-nearest Neighbor, Clustering*

I. PENDAHULUAN

Pendidikan merupakan proses mengubah sikap, perilaku, pola pikir dan tingkat pengetahuan setiap individu atau kelompok. Dengan pendidikan diharapkan kualitas hidup seseorang dapat menjadi lebih baik. Pendidikan juga merupakan salah satu faktor penentu bagi kemajuan suatu bangsa dan negara.

Pendidikan dibedakan menjadi dua jenis, yaitu pendidikan formal dan pendidikan informal. Pendidikan formal adalah pendidikan yang teorganisasi oleh pemerintah dan berjenjang. Sedangkan pendidikan informal adalah pendidikan yang terdapat di dalam suatu keluarga atau masyarakat yang bentuknya tidak terorganisasi.

Pembagian atau pengelompokan kelas bagi siswa-siswi baru setiap tahun pada umumnya masih menggunakan cara konvensional, yaitu secara acak ataupun dengan pengurutan nilai ujian akhir dan belum melibatkan kemajuan teknologi. Pembagian siswa/i dengan menggunakan cara konvensional tersebut akan memunculkan suatu permasalahan kesenjangan siswa antar kelas dan kesenjangan siswa dalam satu kelas. Hal ini terjadi dikarenakan adanya perbedaan kecerdasan yang jauh berbeda antar siswa dalam satu kelas/antar kelas.

Mengingat pembagian kelas pada tahun ajaran baru merupakan masalah pengelompokan (*clustering*), maka metode yang seharusnya diterapkan dalam pembagian kelas bagi siswa baru ini adalah metode pengelompokan. Tidak hanya dalam bidang pendidikan, dalam berbagai bidang lain misalnya bidang bisnis metode pengelompokan juga sangat banyak digunakan. Untuk itu diperlukan pemilihan algoritma yang tepat untuk menyelesaikan permasalahan pembagian kelas siswa baru ini. Algoritma *fuzzy c-means* dan algoritma *k-nearest neighbor* merupakan dua algoritma yang dapat digunakan dalam menyelesaikan permasalahan pengelompokan (*Clustering*), maka pada penelitian ini akan mengkaji sejauh mana efektifitas kedua algoritma ini dalam menyelesaikan permasalahan pengelompokan kelas bagi siswa baru.

Berdasarkan latar belakang diatas maka penulis memutuskan untuk membuat suatu aplikasi komparasi pengelompokan siswa baru dengan menggunakan algoritma *fuzzy c-means* dan algoritma *k-nearest neighbor* di MAN 01 Kota Bengkulu. Hasil dari penelitian ini diharapkan dapat digunakan menyelesaikan pembagian kelas siswa baru secara objektif dengan memanfaatkan kemajuan dalam bidang teknologi informasi.

Adapun rumusan masalah yang dibahas pada penelitian ini adalah :

1. Bagaimana membangun aplikasi pengelompokan rombongan belajar siswa baru dengan menggunakan algoritma *fuzzy c-means* dan algoritma *k-nearest neighbor*.
2. Bagaimana kinerja dari algoritma *fuzzy c-means* dan algoritma *k-nearest neighbor* pada kasus pengelompokan rombongan belajar siswa baru.

Adapun Tujuan yang ingin dicapai dari pelaksanaan dan penulisan tugas akhir ini adalah :

1. Untuk membangun aplikasi pengelompokan rombongan belajar siswa dan siswi baru dengan menggunakan algoritma *fuzzy c-means* dan algoritma *k-nearest neighbor*.
2. Untuk menganalisa perbandingan akurasi hasil algoritma *fuzzy c-means* dan algoritma *k-nearest neighbor* dalam kasus pengelompokan rombongan belajar siswa dan siswi baru.

Manfaat penelitian yang dapat diambil adalah:

1. Aplikasi yang dibangun dapat digunakan untuk pengelompokan rombongan belajar siswa dan siswi baru.
2. Penelitian ini dapat memberikan analisa perbandingan akurasi hasil algoritma *fuzzy c-means* dan algoritma *k-nearest neighbor* dalam kasus pengelompokan rombongan belajar siswa dan siswi baru.

II. LANDASAN TEORI

A. Klustering (Clustering)

Klaster atau *cluster* adalah kumpulan *record* yang memiliki kemiripan suatu dengan yang lainnya dan memiliki ketidakmiripan dengan *record* dalam klaster lain. Sedangkan Klustering atau *clustering* atau pengklasteran merupakan proses pengelompokan *record* untuk membentuk kelas objek-objek yang memiliki kemiripan.

Algoritma *clustering* mencoba untuk melakukan pembagian terhadap keseluruhan data menjadi kelompok-kelompok yang memiliki kemiripan (homogen) sehingga kemiripan dengan *record* dalam kelompok lain akan bernilai minimal (Wijaya,2011)[1]. Fuzzy C-Means selain termasuk dalam karakteristik Hierarchical clustering juga merupakan bagian dari metode dalam fuzzy Clustering. Dalam Fuzzy Clustering terdapat 2 metode dasar yaitu metode Fuzzy subtractive clustering (atau fuzzy Equivalence Relation) dan Fuzzy C-Means. Metode Fuzzy subtractive clustering adalah metode yang dalam proses pembentukan klaster dimana banyaknya cluster tidak ditentukan sebelumnya, sedangkan Fuzzy C-Means banyaknya cluster ditentukan sebelumnya.

B. Logika Fuzzy

Fuzzy secara bahasa dapat diartikan sebagai kabur atau samar-samar. Suatu nilai dapat bernilai benar atau salah secara bersamaan. Menurut Rinaldi Munir (2009)[2] Logika fuzzy pertama kali dikembangkan oleh Lotfi Asker Zadeh melalui tulisannya pada tahun 1965 tentang teori himpunan fuzzy. Lotfi Asker Zadeh adalah seorang ilmuwan Amerika Serikat berkebangsaan Iran dari Universitas California di Barkeley. Meskipun logika fuzzy dikembangkan di Amerika, namun lebih populer dan banyak diaplikasikan secara luas oleh praktisi Jepang dengan mengadaptasikannya mulai dari teori kendali hingga inteligensia buatan. Salah satu alasan mengapa logika fuzzy lebih berkembang di jepang dibanding amerika adalah kultur orang barat yang cenderung memandang suatu persoalan sebagai hitam-putih, ya-tidak, bersalah-tidak bersalah, sukses-gagal, atau yang setara dengan dunia logika biner Aristoteles, sedangkan kultur orang Timur lebih dapat menerima dunia “abu-abu” atau fuzzy. Logika fuzzy umumnya diterapkan pada masalah-masalah yang mengandung unsur ketidakpastian (uncertainty), ketidaktepatan (imprecise), noisy, dan sebagainya. Dalam fuzzy dikenal derajat keanggotaan yang memiliki rentang nilai 0 (nol) hingga 1(satu). Berbeda dengan himpunan tegas yang memiliki nilai 1 atau 0 (ya atau tidak).

Logika fuzzy adalah suatu cara yang tepat untuk memetakan suatu ruang input kedalam suatu ruang output, mempunyai nilai kontinyu. Fuzzy dinyatakan dalam derajat dari suatu keanggotaan dan derajat dari kebenaran. Oleh sebab itu sesuatu dapat dikatakan sebagian benar dan sebagian salah pada waktu yang sama (Kusumadewi. 2004)[3]. Kesamaran (fuzzyness) antara benar atau salah. Dalam teori logika fuzzy suatu nilai bisa bernilai benar atau salah secara bersamaan, Namun berapa besar suatu keberadaan dan kesalahan tergantung pada bobot keanggotaan yang dimilikinya. Logika fuzzy memiliki derajat keanggotaan dalam rentang 0 hingga 1. Berbeda dengan logika digital yang hanya memiliki dua nilai 1 atau 0. Logika fuzzy digunakan untuk menterjemahkan suatu besaran yang diekspresikan menggunakan bahasa (linguistic), misalkan besaran kecepatan laju kendaraan yang diekspresikan dengan pelan, agak cepat, cepat, dan sangat cepat. Dan logika fuzzy menunjukkan sejauh mana suatu nilai itu benar dan sejauh mana suatu nilai itu salah. Tidak seperti logika klasik (scrisp)/tegas, suatu nilai hanya mempunyai 2 kemungkinan yaitu merupakan suatu anggota himpunan atau tidak. Derajat keanggotaan 0 (nol) artinya nilai bukan merupakan anggota himpunan dan 1 (satu) berarti nilai tersebut adalah anggota himpunan.

C. Fuzzy C-Means

Metode clustering yang akan digunakan dalam penelitian ini adalah fuzzy c-means. Metode ini pertama kali dikenalkan oleh Jim Bezdek pada tahun 1973 (Angelina Widya, 2011)[4]. Untuk selanjutnya metode Fuzzy C-Means akan penulis singkat tulisannya menjadi FCM. Berikut langkah-langkah dari metode FCM dalam pengelompok data siswa :

1. Susunlah matrix X berukuran n x m sebagai tempat dari data yang akan di-cluster. Di mana n merupakan banyaknya data dan m merupakan banyaknya atribut dari data tersebut. X_{ij} = data pada sampel ke-i dan atribut ke-j.
2. Inisiasikan nilai pada komponen-komponen berikut ini:
 - a. Banyaknya cluster yang diinginkan $\rightarrow k$
 - b. Pangkat (pembobot) $\rightarrow w$
 - c. Maksimum Iterasi $\rightarrow \maxIter$
 - d. Error terkecil $\rightarrow e$
 - e. Fungsi Objektif awal $\rightarrow P_0 = 0$
 - f. Iterasi awal $\rightarrow iter = 1$.
 - g. Jumlah Rombel $\rightarrow R$
3. Bangkitkan matriks bilangan acak U_{ik} , dimana banyaknya i sejumlah data dan k sejumlah cluster.
4. Hitunglah pusat cluster dengan rumus:

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

5. Hitung fungsi objektif pada iterasi ke-iter dengan rumus:

$$P_{iter} = \sum_{i=1}^n \sum_{k=1}^c \left[\left[\sum_{j=1}^m (X_{ij} - V_{kj})^2 \right] (U_{ik})^w \right] \dots(2)$$

6. Perbaharui matrix partisi U dengan menggunakan rumus:

$$U_{ik} = \frac{\left[\sum_{j=1}^m (X_{ij} - V_{kj})^2 \right]}{\sum_{k=1}^c \left[\sum_{j=1}^m (X_{ij} - V_{kj})^2 \right]} \quad \dots(3)$$

7. Cek kondisi berhenti, jika $P_{iter} - P_{iter-1} < e$ atau $iter \geq \maxIter$ maka proses perhitungan berhenti. Namun jika keduanya tidak memenuhi, maka $iter = iter+1$ dan kembali ke langkah-4.
8. Urutkan data berdasarkan nilai cluster.
9. Pemberian nilai R (RomBel) pada setiap data secara berurutan.
10. Urutkan data berdasarkan nilai R.

D. K-Nearest Neighbor

Algoritma *k-nearest neighbor* (k-NN atau KNN) merupakan metode untuk melakukan *clustering* terhadap objek berdasarkan data pembelajaran yang jaraknya paling dekat dengan objek tersebut. Data pembelajaran diproyeksikan ke ruang berdimensi banyak, dimana masing-masing dimensi merepresentasikan fitur dari data. Ruang ini dibagi menjadi bagian-bagian berdasarkan klasifikasi data pembelajaran. Algoritma *k-nearest neighbor* biasanya dihitung berdasarkan *euclidean distance space* atau biasa disebut dengan *euclidean* saja yang dipresentasikan sebagai berikut :

$$d = \sqrt{(X_1 - X_2)^2 + (Y_1 - Y_2)^2 + \dots + (N_1 - N_2)^2}$$

Keterangan :

- d = Jarak Skalar antara kedua atribut
- X_1 = Atribut pertama dari record pertama
- X_2 = Atribut pertama dari record kedua
- Y_1 = Atribut kedua dari record pertama
- Y_2 = Atribut kedua dari record kedua
- N_1 = Atribut ke-n dari record pertama
- N_2 = Atribut ke-n dari record kedua

Dalam *Euclidean* perhitungan yang dipilih merupakan jarak terpendek antara dua titik. Jika ada n titik pengamatan dengan p variabel, maka sebelum dilakukan pengelompokan data atau objek, terlebih dahulu menentukan ukuran kedekatan sifat antar data.

Berikut langkah-langkah algoritma *k-nearest neighbor* dalam melakukan pengelompokan data siswa :

1. Inputkan jumlah Tetangga (T), Jumlah Cluster (k) dan Jumlah RomBel (R).
2. Tentukan *interval* batas-batas cluster dengan langkah berikut ini :
 - a. Hitung nilai batas maksimum *euclidean cluster* dengan menggunakan rumus :

$$\text{Max. } k = \sqrt{(X_{\max} - X_{\min})^2 + (Y_{\max} - Y_{\min})^2 + \dots + (N_{\max} - N_{\min})^2}$$

Keterangan :

- Max. k = Batas maksimum *euclidean cluster*
- X_{\max} = Nilai maximum atribut pertama
- X_{\min} = Nilai minimum atribut pertama
- Y_{\max} = Nilai maximum atribut kedua
- Y_{\min} = Nilai minimum atribut kedua
- N_{\max} = Nilai maximum atribut ke-n
- N_{\min} = Nilai minimum atribut ke-n

- b. Tentukan *interval* masing-masing cluster sebanyak *kcluster* dengan rumus
3. Hitung jarak *euclidean* antar data dengan menggunakan persamaan 4 dan pilih nilai *euclidean* terkecil.

4. Urutkan data berdasarkan *cluster*.
5. Pemberian nilai R (RomBel) pada setiap data secara berurutan.
6. Urutkan data berdasarkan nilai R.

E. Interval Nilai Rata-Rata Siswa

Pada penelitian ini penulis akan mengelompokkan data siswa berdasarkan nilai rata-rata siswa ke dalam 3 kelompok utama. Pengelompokan nilai rata-rata ini nantinya akan digunakan sebagai acuan dalam menganalisis hasil yang telah diproses oleh perangkat lunak yang telah dibangun. Dengan standarisasi 3 kelompok ini penulis akan menganalisis komposisi setiap kelompok belajar yang tercipta berdasarkan algoritma fuzzy c-means dan algoritma k-nearest neighbor. Berdasarkan data yang penulis gunakan dalam penelitian ini yaitu data nilai siswa baru MAN 01 Kota Bengkulu sebanyak 194 record diketahui bahwa nilai rata-rata siswa terendah adalah 6.04 atas nama Khairurroziqin (NIS 6886), sedangkan nilai rata-rata siswa tertinggi adalah 8.70 diraih oleh Ikhsan Bintang (NIS 6736). Sehingga interval nilai rata-rata siswa adalah sebagai berikut:

$$\begin{aligned} \text{Panjang kelas} &= \frac{\text{Nilai max} - \text{Nilai min}}{\text{Banyak kelas}} \\ \text{interval} &= \frac{8.70 - 6.04}{3} \quad (6) \\ &= 0.88 \end{aligned}$$

Interval untuk setiap kategori data menjadi sebagai berikut:

$$\begin{aligned} \text{Rata-rata rendah} &: 6.04 \leq x \leq 6.92 \\ \text{Rata-rata menengah} &: 6.93 \leq x \leq 7.81 \\ \text{Rata-rata tinggi} &: 7.82 \leq x \leq 8.7 \end{aligned}$$

Interval nilai rata-rata siswa yang didapat pada setiap kelompok ini juga memberikan informasi mengenai batasan-batasan nilai pada kategori kecerdasan siswa yaitu kategori kecerdasan rendah, menengah dan tinggi.

Adapun prosedur pengelompokan rombongan belajar (RomBel) pada penelitian ini adalah sebagai berikut :

1. Pertama semua data 194 siswa yang telah didapat dari pengumpulan data diinputkan ke dalam database dengan menggunakan aplikasi yang telah dibangun. Untuk proses input data kedalam database dapat dilakukan dengan 2 cara, yaitu pertama dengan menginputkan satu persatu data siswa melalui form tambah data siswa yang ada pada aplikasi yang sudah dibangun. Sedangkan cara kedua yaitu dengan menggunakan fitur import data yang tersedia pada tools database XAMPP, semua data siswa yang awalnya disajikan dalam bentuk file microsoft excel dirubah formatnya menjadi

*.csv. file *.csv inilah yang akan di upload/import ke database yang telah disediakan.

2. Selanjutnya data yang telah disimpan ke dalam database akan diolah dengan menggunakan algoritma fuzzy c-means dan algoritma k-nearest neighbor. Proses pengolahan data ini disebut dengan clustering. Proses ini akan memberikan nilai cluster pada setiap record data dan data akan diurutkan berdasarkan nilai cluster seperti C1,C2,C3,...,Cn.
3. Setelah dilakukan proses clustering langkah selanjutnya adalah membagi 194 data siswa ke dalam 9 RomBel. Sehingga akan dihasilkan jumlah siswa yang merata. Proses yang dilakukan pada pembagian siswa ini adalah hasil pada langkah ke 2 setiap record data yang telah diurutkan berdasarkan nilai cluster diberi nilai R1 s.d R9 secara berurutan hingga data yang terakhir (data ke 194). Pada langkah ini akan dihasilkan jumlah siswa R1 s.d R5 berjumlah 22 orang siswa dan R6 s.d R9 berjumlah 21 orang siswa.
4. Selanjutnya dilakukan pengurutan data (sorting) berdasarkan nilai R (RomBel) yang telah dilakukan dalam hal ini R1 s.d R9 kemudian ditampilkan sebagai hasil akhir dari proses pengelompokan rombongan belajar (RomBel).

Setelah dilakukan prosedur pengelompokan, selanjutnya akan dilakukan analisis perbandingan algoritma. Perbandingan algoritma untuk proses pengelompokan RomBel siswa baru akan dibagi menjadi 2 bagian, yaitu :

1. Perbandingan algoritma berdasarkan waktu CPU proses.
2. Perbandingan algoritma berdasarkan distribusi nilai rata-rata RomBel pada hasil aplikasi terhadap data asli.

Untuk melakukan 2 perbandingan tersebut perlu dilakukan pengujian pada masing-masing algoritma, dimana pengujian dilakukan dengan melakukan kombinasi pada beberapa variabel algoritma. Adapun variabel yang akan dibandingkan yaitu :

1. Membandingkan variabel jumlah cluster (k) pada kedua algoritma.
2. Membandingkan variabel Maximum Iterasi (MaxIter) pada algoritma fuzzy c-means terhadap variabel Tetangga (T) pada algoritma k-nearest neighbor.

Total pengujian yang dilakukan adalah 30 kali yaitu 15 kali pengujian pada algoritma fuzzy c-means dan 15 kali pengujian pada algoritma *k-nearest neighbor*.

III. METODE PENELITIAN

A. Jenis dan Tempat Penelitian

Jenis penelitian yang akan dilakukan pada tugas akhir ini adalah penelitian terapan. Penelitian lebih berfokus kepada penerapan penelitian tersebut dalam kehidupan sehari-hari. Hal ini sesuai dengan tujuan penelitian penulis, yaitu menghasilkan aplikasi yang dapat langsung diterapkan untuk membagi kelompok kelas siswa baru di Sekolah Menengah Atas (SMA) atau sederajat.

Penelitian ini dilaksanakan di Sekolah Madrasah Aliyah Negeri 01 Kota Bengkulu. Waktu penelitian mulai dihitung bulan September 2013-Desember 2013.

Teknik pengumpulan data yang digunakan dalam penelitian terapan ini terdiri dari:

1. Dokumentasi

Pada teknik dokumentasi data-data yang berkenaan dengan penelitian didapat dengan cara mengumpulkan data-data yang sudah ada di Sekolah Madrasah Aliyah Negeri 01 Kota Bengkulu. Data-data tersebut berupa data siswa baru kelas 1 tahun 2013. Data yang didapat berupa data nilai ujian nasional dan nilai ujian sekolah siswa baru sebanyak 194 orang siswa dan siswi.

2. Studi pustaka

Studi pustaka merupakan salah satu teknik pengumpulan data yang bersumber dari buku-buku dan dokumen yang berhubungan dengan sistem yang dibangun pada penelitian. Studi kepustakaan yang digunakan untuk memperoleh data serta informasi dalam penelitian ini, meliputi:

- a) Buku referensi yang berhubungan dengan algoritma fuzzy C-means, k-Nearest Neighbor dan clustering, serta referensi lain yang berhubungan dengan penelitian.
- b) Jurnal dan skripsi yang berhubungan dengan penelitian.

B. Metode Pengembangan Sistem

Metode pengembangan sistem yang digunakan pada penelitian ini adalah model pengembangan sistem sekuensial linier. Menurut Pressman (2002) Model sekuensial linier merupakan salah satu dari metode yang digunakan untuk pengembangan sistem[5]. Sekuensial linier sering disebut juga dengan “siklus kehidupan klasik” atau “model air terjun”. Model sekuensial linier mengusulkan sebuah pendekatan kepada perkembangan perangkat lunak sistematis dan sekuensial yang mulai pada tingkat dan kemajuan sistem pada seluruh analisis, desain, kode, pengujian dan pemeliharaan. Berikut penjelasan tahap-tahap model sekuensial linier dalam penelitian ini:

1. Analisis kebutuhan.

Analisis kebutuhan merupakan tahap pertama yang menjadi dasar proses pembuatan sistem selanjutnya. Pada tahapan ini yang pertama kali dilakukan adalah pengumpulan data berupa data dan informasi siswa-siswi, nilai ujian nasional dan ujian sekolah siswa-siswi kelas 1 MAN 01 Kota Bengkulu serta spesifikasi aplikasi yang akan dirancang, seperti software, hardware dan database.

2. Desain sistem

Tahapan desain sistem ini dikerjakan setelah tahap analisis kebutuhan selesai dikerjakan. Pada tahapan ini hasil analisis yang telah dilakukan pada tahapan sebelumnya diterjemahkan ke dalam bentuk rancangan antarmuka (interface), rancangan basis data, rancangan struktur data dan rancangan prosedur algoritma. Metode desain sistem yang digunakan adalah *Unified Modeling Language* (UML). Menurut Chonocles:

‘Sebagai bahasa, berarti UML memiliki sintaks dan semantik. Ketika kita membuat model menggunakan konsep UML ada aturan-aturan yang harus diikuti. Bagaimana elemen pada model-model yang kita buat berhubungan dengan lainnya harus mengikuti standar yang ada. UML bukan hanya sekedar diagram, tetapi juga menceritakan konteksnya. Ketika pelanggan memesan sesuatu dari sistem, transaksi, penanganan error yang terjadi, keamanan sistem dapat dijawab dengan UML’ (Widodo dan Herlawati, 2011)[6].

3. Penulisan Kode Program

Setelah tahapan desain sistem selesai hasil rancangan akan diubah menjadi bentuk yang dimengerti oleh mesin yaitu ke dalam bahasa pemrograman yang telah ditentukan melalui proses penulisan program. Dalam penelitian ini, software yang digunakan adalah Java Netbean IDE 6.0.1 dan XAMPP.

4. Pengujian sistem

Pada tahap ini akan dilakukan pengujian untuk melihat apakah sistem yang dibuat sesuai dengan rancangan. Setelah diketahui hasil perhitungan aplikasi yang sama dengan perhitungan manual maka tahap selanjutnya yang dilakukan adalah menganalisa perbandingan algoritma fuzzy c-means dengan algoritma k-nearest neighbor. Pada pengujian program akan dilakukan pengujian *black box*, *white box* dan *akurasi hasil*. Pengujian *black box* digunakan untuk menguji fungsional aplikasi seperti kesalahan interface, kesalahan kinerja dan kesalahan akses database. Jika pada pengujian *black box* lebih ditunjukkan pada interface aplikasi maka pada *white box* lebih ditunjukkan pada unit aplikasi. Pada pengujian akurasi hasil akan dilakukan pengujian dengan menggunakan sejumlah kecil sampel data, pengujian akurasi hasil ini untuk menguji apakah hasil perhitungan aplikasi

sudah tepat/sama dengan hasil pada perhitungan manual.

Selain ketiga pengujian diatas juga akan dilakukan 30x pengujian pada aplikasi dengan menggunakan variasi nilai variabel masing-masing algoritma. Variabel yang digunakan pada pengujian ini adalah variabel jumlah klaster (k), variabel maximum iterasi (MaxIter) dan variabel jumlah tetangga (T). Hasil yang didapat pada 30 kali pengujian ini akan dianalisis dan dibandingkan dengan kondisi data real di lapangan. Analisis hasil ini berupa perbandingan waktu CPU proses kedua algoritma dan perbandingan persebaran anggota antar rombongan belajar pada setiap pengujian. Persebaran anggota rombongan belajar akan dianalisis berdasar selisih nilai rata-rata terbesar antar rombongan belajar pada tiap mata pelajaran.

5. Pemeliharaan

Tahap pemeliharaan merupakan tahap akhir pengembangan dan implementasi sistem. Agar sistem dapat digunakan dalam jangka panjang perlu dilakukan pemeliharaan sistem. Pemeliharaan sistem dilakukan bukan hanya sekedar proses memperbaiki kesalahan program tetapi proses yang memiliki karakteristik memperbaiki kesalahan yang tidak ditemukan pada tahapan sebelumnya atau menambahkan fungsi baru yang belum ada pada program.

IV. HASIL DAN PEMBAHASAN

Total pengujian yang dilakukan adalah 30 kali pengujian yaitu 15 kali pengujian pada algoritma fuzzy c-means dan 15 kali pengujian pada algoritma k-nearest neighbor. Hasil setiap pengujian disimpan dalam bentuk file microsoft excel berupa tabel pengelompokan data siswa beserta atributnya. Setelah dilakukan 30x pengujian pengelompokan rombongan belajar siswa baru pada masing-masing algoritma maka waktu CPU proses pada setiap pengujian dapat dilihat pada Tabel-1 berikut ini.

Tabel-1. Perbandingan waktu proses

No	Variabel	Waktu Proses (mili second)	
		Fuzzy C-Means	K-Nearest Neighbor
1	k=3, Maxiter/T=5	4.359	500
2	k=3, Maxiter/T=10	8.828	1.250
3	k=3, Maxiter/T=20	18.157	2.094
4	k=6, Maxiter/T=5	9.593	500
5	k=6, Maxiter/T=10	19.985	1.312
6	k=6, Maxiter/T=20	25.797	1.797
7	k=9, Maxiter/T=5	14.000	453
8	k=9, Maxiter/T=10	27.484	1.344
9	k=9, Maxiter/T=20	52.891	1.938
10	k=12, Maxiter/T=5	16.203	516
11	k=12, Maxiter/T=10	38.110	1.219
12	k=12, Maxiter/T=20	42.938	1.687
13	k=15, Maxiter/T=5	20.750	515
14	k=15, Maxiter/T=10	45.390	1.297
15	k=15, Maxiter/T=20	89.047	1.782

Berdasarkan Tabel-1 terlihat jelas bahwa algoritma *fuzzy c-means* membutuhkan waktu proses yang lebih lama dibandingkan algoritma *k-nearest neighbor* dalam melakukan proses pengelompokan rombongan belajar siswa baru. Selain itu tabel 1 tersebut juga menunjukkan pengaruh nilai *k* (*cluster*) dan maximum *iterasi*(MaxIter) terhadap waktu proses algoritma *fuzzy c-means* dimana semakin tinggi nilai *k* (jumlah *cluster*) maka akan semakin lama waktu CPU proses yang dibutuhkan. Sedangkan waktu proses algoritma *k-nearest neighbor* berbanding lurus dengan nilai variabel Tetangga (T).

Penulis selanjutnya menganalisa hasil sebaran data dari aplikasi pada pengelompokan rombongan belajar siswa baru berdasarkan masing-masing algoritma serta akan membandingkannya dengan data real di lapangan. Melalui *output* yang dihasilkan masing-masing algoritma akan dianalisis sebaran data siswa ke sejumlah rombongan belajar yang telah ditentukan, dalam pengujian ini jumlah rombongan belajar siswa ditentukan sebanyak 9 rombongan belajar. Dari sebaran data siswa ini akan dilihat sebaran siswa berdasarkan kategori nilai rata-rata setiap siswa, seperti diawal sudah disebutkan bahwa siswa dikategorikan menjadi 3 kelompok, yaitu siswa rata-rata rendah ($6.04 \leq x \leq 6.92$), sedang ($6.93 \leq x \leq 7.81$) dan tinggi ($7.82 \leq x \leq 8.7$). Selain itu juga akan dibandingkan nilai rata-rata setiap rombel pada setiap mata pelajaran untuk setiap pengujian. Contoh hasil pengujian dapat dilihat pada Tabel-2 dan Tabel-3 berikut ini:

Tabel-2. Hasil Pengujian Fuzzy C-Means K=6, MaxIter=5

No	Room	Cluster	NIS	Nama	...	Math	B.Indo	B.Ing	PPKN	Agama	Seni	IPA	IPS	Penjas
1	R1	C2	6721	Novita Sari	...	5.50	7.90	5.80	7.90	7.77	8.00	5.60	8.00	8.50
2	R1	C2	6746	Qharin Nurhasanah	...	4.50	7.40	6.20	7.80	8.22	8.80	6.25	8.60	8.50
...
34	R2	C2	7046	Raudhatul Jannah	...	6.75	8.20	6.40	9.50	9.00	8.20	3.00	8.50	8.00
35	R2	C5	6753	M.Rifqi Aziz	...	2.50	6.60	3.60	9.50	9.60	9.26	5.00	9.30	9.58
...
193	R9	C5	7030	Vini Duwi Putri Utami	...	2.75	8.00	2.60	8.50	7.20	9.00	4.75	9.00	8.20
194	R9	C5	7042	Monita Andriyani	...	2.75	8.20	2.80	7.40	7.24	9.30	5.50	7.00	7.30

Tabel-3. Hasil Pengujian K-Nearest Neighbor k=6, T=10

No	Room	Cluster	NIS	Nama	...	Math	B.Indo	B.Ing	PPKN	Agama	Seni	IPA	IPS	Penjas
1	R1	C1	7051	Nesi Fadillah Haki	...	3.25	8.20	4.20	7.70	6.50	8.00	4.50	7.80	7.50
2	R1	C2	6757	Muhammad Harits Abdillah	...	5.00	6.40	4.60	8.00	7.60	7.56	3.50	7.60	9.40
...
39	R2	C1	7052	M.Aditya Pratama	...	5.80	8.30	4.50	7.70	7.50	9.00	6.90	8.20	7.80
40	R2	C2	6758	Annesia Haripraptika	...	4.25	8.40	7.20	7.80	7.20	8.80	4.75	8.80	7.80
...
193	R9	C2	6921	Meysi	...	3.50	7.00	4.60	8.20	8.00	7.65	6.75	8.60	8.50
194	R9	C2	7041	M. Farhan Khotami	...	5.80	6.30	5.50	9.20	8.50	7.75	5.80	8.00	8.10

Setelah dilakukan 30x kasus uji berdasarkan kategori nilai rata-rata siswa terlihat jelas bahwa kedua algoritma tidak *fair* dalam melakukan pembagian komposisi rombongan belajar siswa baru. Hal ini dibuktikan dengan adanya perbedaan mencolok jumlah siswa pada setiap kategori antar rombongan belajar. Seperti pada kasus uji *fuzzy c-means k=3 MaxIter=5*, jumlah siswa kategori tinggi pada R1 berjumlah 7 orang sedangkan R6 berjumlah 0 orang. Hal yang sama juga terjadi pada pengujian KNN k=9 MaxIter=5 terlihat selisih jumlah yang mencolok, pada kategori nilai rata-rata tinggi pada R3 berjumlah 4 orang sedangkan pada R4 berjumlah 0 orang. Hal ini terjadi dikarenakan kedua algoritma memang dirancang untuk mengelompokkan siswa berdasarkan kemiripan 9 nilai mata pelajaran, tidak hanya melihat satu indikator berupa nilai rata-rata siswa. Akan tetapi jika kedua algoritma dibandingkan dengan data real maka terlihat kedua algoritma cenderung lebih baik dibandingkan data real, hal ini dibuktikan dengan selisih jumlah pada setiap kategori lebih rendah dibandingkan data real.

Dalam melakukan distribusi siswa ke dalam rombongan belajar, kedua algoritma mampu mendistribusikan secara merata jumlah siswa pada masing-masing rombongan belajar, yaitu R1 sampai dengan R4 berjumlah masing-masing 22 orang siswa dan R5 sampai dengan R9 berjumlah masing-masing 21 orang siswa.

Selanjutnya penulis akan menganalisis sebaran data siswa pada setiap rombel untuk setiap pengujian

dengan melihat nilai rata-rata setiap mata pelajaran pada setiap rombongan belajar. Kemudian akan dihitung selisih terbesar nilai rata-rata antar rombel. Semakin kecil nilai selisih yang didapat maka semakin kecil pula tingkat ketimpangan yang terjadi antar RomBel.

Pada pengujian fuzzy c-means didapatkan selisih nilai rata-rata antar RomBel selalu lebih kecil dibandingkan pada data real (0.5), hal ini berarti bahwa algoritma fuzzy c-means lebih baik dalam mengelompokkan siswa dibandingkan dengan cara konvensional. Selain itu dibandingkan dengan cara konvensional algoritma k-nearest neighbor cenderung lebih baik, hal ini dibuktikan dengan nilai perbedaan pada kolom average cenderung lebih kecil. Dari pengujian yang telah dilakukan algoritma fuzzy c-means menghasilkan nilai terbaik saat variabel k=6 dan MaxIter =5 dengan nilai rata-rata perbedaan antar RomBel sebesar 0.15. Sedangkan algoritma k-nearest neighbor menghasilkan nilai terbaik saat variabel variabel k=3 dan T =5 dengan nilai rata-rata perbedaan antar RomBel sebesar 0.19.

V. KESIMPULAN DAN SARAN

Berdasarkan pengujian, analisis hasil dan pembahasan mengenai aplikasi komparasi algoritma fuzzy c-means dan algoritma k-nearest neighbor dalam pengelompokan siswa baru yang sudah dibangun, maka dapat disimpulkan bahwa:

DAFTAR PUSTAKA

1. Aplikasi komparasi algoritma fuzzy c-means dan algoritma k-nearest neighbor dalam pengelompokan rombongan belajar siswa baru telah berhasil dibangun dengan menggunakan analisis dan desain perangkat lunak Unified Modeling Language (UML), metode pengembangan sistem sekuensial linier dan software Netbeans IDE 6.0.1.
2. Algoritma k-nearest neighbor membutuhkan waktu CPU proses lebih cepat dibandingkan algoritma fuzzy c-means.
3. Waktu CPU proses algoritma fuzzy c-means dipengaruhi oleh nilai variabel Maximum Iterasi (MaxIter) sedangkan algoritma k-nearest neighbor dipengaruhi oleh variabel Tetangga (T).
4. Aplikasi yang telah dibangun mampu mendistribusikan semua data siswa baru ke sejumlah rombongan belajar dengan baik, yaitu tidak adanya perbedaan yang mencolok antar RomBel pada komposisi jumlah siswa.
5. Dalam melakukan pengelompokan rombongan belajar algoritma fuzzy c-means lebih baik dibandingkan algoritma k-nearest neighbor. Hal ini dibuktikan dengan lebih kecilnya perbedaan/selisih nilai rata-rata antar RomBel pada sebagian besar pengujian yang telah dilakukan.
6. Dalam melakukan pengelompokan rombongan belajar siswa baru algoritma fuzzy c-means dan k-nearest neighbor lebih baik dibandingkan dengan cara konvensional (data real). Hal ini dibuktikan dengan rendahnya perbedaan nilai rata-rata antar RomBel pada sejumlah besar pengujian yang telah dilakukan.
7. Dari pengujian yang telah dilakukan algoritma fuzzy c-means menghasilkan nilai terbaik saat variabel $k=6$ dan $MaxIter =5$ dengan nilai rata-rata perbedaan antar RomBel sebesar 0.15. Sedangkan algoritma k-nearest neighbor menghasilkan nilai terbaik saat variabel variabel $k=3$ dan $T =5$ dengan nilai rata-rata perbedaan antar RomBel sebesar 0.19.

Berdasarkan pengujian, analisis hasil dan pembahasan maka penulis menyarankan untuk penelitian selanjutnya pada kasus pengelompokan siswa baru sebaiknya dilakukan penyempurnaan pada algoritma *fuzzy c-means*, dikarenakan algoritma *fuzzy c-means* sangat bergantung kepada kualitas matriks bilangan acak (matriks *fuzzy*). Sehingga dibutuhkan suatu metode yang bisa memberikan nilai matriks bilangan acak yang optimal.

- [1] Kusumadewi, S dan Purnomo, H. 2004. *Logika Fuzzy Untuk Pendukung Keputusan*. Yogyakarta : Graha Ilmu.
- [2] Munir, Rinaldi. 2009. *Pengantar Logika Fuzzy* . Jurusan Teknik Informatika Institut Teknologi Bandung.
- [3] Pressman, Roger S. 2002. *Rekayasa Perangkat Lunak Pendekatan Praktisi (Buku Satu)*. Yogyakarta: Andi.
- [4] Widodo, Prabowo Pudjo dan Herlawati. 2011. *Menggunakan UML*. Bandung: Informatika Bandung.
- [5] Widya, Angelina. 2011. Mengenal Fuzzy C-Means. Diakses Melalui URL <http://www.megenep.com/artikel/category/0?sub=0> pada tanggal 11 Juli 2013.
- [6] Wijaya, Arim. 2011. Universitas Komputer Indonesia Bandung. Analisis Algoritma K-Means Untuk Membantu Penjurusan Siswa Di MAN Binong subang. [Online]. Tersedia: http://dir.unikom.ac.id/s1-final-project/fakultas-teknik-dan-ilmu-komputer/teknikinformatika/2011/jbptunikompp-gdl-arimwijaya-24414/4-unikom-a-a.pdf/pdf/4-uni_kom-a-a.pdf [25 September 2013].