

IMPLEMENTASI ALGORITMA FUZZY K-MEANS PADA DATA REKAM MEDIS BERDASARKAN KODE PENYAKIT INTERNASIONAL (ICD10)

Mulya Negara Bachtiar¹, Haryanto Hanny²

¹Mahasiswa, ²Dosen Pembimbing

Program Studi Teknik Informatika/Fakultas Ilmu Komputer

Jl. Nakula I No. 5-11, Semarang, 50131, 024 3517261

Email : tiarmulya@yahoo.co.id¹, hanny.haryanto@dsn.dinus.ac.id²

ABSTRAK

Penelitian ini dilakukan untuk mengolah data rekam medis pada RSUD Kardinah Tegal, dengan adanya data rekam medis yang semakin bertumpuk maka akan mempersulit pada pengambilan informasi, selain itu kebanyakan dokter tidak bisa membaca penyakit pasien secara langsung dengan hanya melihat kode penyakit pasien berdasarkan kode internasional(ICD10). Agar bisa membaca kode penyakit pasien maka akan dilakukan pengelompokan atau pengklasteran terlebih dahulu penyakit pasien berdasarkan kode penyakit internasional. Untuk mencapai tujuan penelitian, maka dipilih 1 atribut yang ada pada data rekam medis. Data rekam medis yang dimaksud terdiri dari atribut diagnosa penyakit berdasarkan International Classification of Diseases-10th (ICD-10). Penelitian ini menggunakan algoritma Fuzzy K-Means dikarenakan dalam data set tersebut mempunyai berbagai macam sub-sub kode penyakit berdasarkan kode penyakit internasional (ICD10). Dengan adanya masalah tersebut maka akan dilakukan penelitian untuk mengambil informasi pengelompokan dan pembentukan klaster mana saja yang masuk kedalam penyakit tertentu berdasarkan kode penyakit internasional (ICD10). Hasil dari pengklasteran tersebut dengan menggunakan Fuzzy K-Means selanjutnya akan dilakukan pengukuran tingkat kesalahan atau Distance Error dari analisis data pasien berdasarkan ICD10.

Kata Kunci : Rekam Medis, Kode Penyakit Internasional (ICD10), Klustering, Fuzzy C-Means

ABSTRACT

This research is to process medical record data on RSUD Kardinah Tegal, with medical record data that continuously stacking, make it harder to gather information, furthermore many of doctors can't read patient directly with only look at patient disease code based on International Code Disease (ICD10). In order to read the patient code disease, grouping or clustering the patient based on international code disease first. to meet the research goal. Then selected one attribute that available on medical record data. medical record data that selected consist of disease diagnose attribute based on International Classification Disease-10th (ICD10). This research using Fuzzy C-Means algorithm because the data set have many disease sub-codes based on International Disease Code(ICD10).Result from the clustering using Fuzzy C-Means, next we need to measure distance error from analysis patient data based on ICD10

Keywords: Medical record, International Code Disease(ICD10), Clustering, Fuzzy K-Means.

1. PENDAHULUAN

1.1 Latar Belakang

Pertumbuhan yang sangat pesat dari akumulasi data telah menciptakan kondisi

data tapi minim informasi. Data mining sendiri sering disebut sebagai *knowledge discovery in database (KDD)* adalah kegiatan yang meliputi pengumpulan,

pemakaian data historis untuk menemukan keteraturan, pola hubungan dalam set data berukuran besar. Salah satu algoritma yang dikenal dalam data mining yaitu Fuzzy K-Means. Pengertian Fuzzy K-Means (FCM), 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. pertama kali adalah menentukan pusat cluster yang akan menandai lokasi rata-rata untuk tiap cluster. 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 dapat dilihat bahwa pusat cluster akan menuju lokasi yang tepat. Perulangan ini didasarkan pada minimasi fungsi obyektif [1].

Penelitian ini menerapkan ilmu data mining yaitu pada permasalahan kode penyakit internasional yang terlalu banyak sehingga sulit untuk dikelompokkan sesuai dengan kategori penyakit yang semestinya dan kode penyakit berdasarkan kode internasional yang semakin banyak maka akan mengakibatkan penumpukan data

yang akan semakin rumit untuk diolah dan sulit menghasilkan suatu informasi.

Dari penelitian ini data yang belum diolah berada pada sub level terendah sehingga kita belum tahu kode penyakit tersebut masuk dalam kelas tertentu yang ada pada level tertinggi atau super kelasnya. Hasil dari penelitian ini dengan menggunakan algoritma Fuzzy K-Means yaitu kita dapat mengetahui kode penyakit pasien yang diproses akan masuk dalam klaster tertentu yang membentuk suatu informasi bahwa data tersebut akan menghasilkan suatu kesatuan klaster penyakit pasien pada level tertinggi dan juga dapat mengetahui *distance error* algoritma Fuzzy K-Mean pada pengelompokan data penyakit pasien.

1.2 Perumusan Masalah

Berdasarkan dari latar belakang diatas dapat dirumuskan, bagaimana cara mengimplementasikan algoritma Fuzzy K-Means untuk pengelompokan penyakit pasien berdasarkan kode internasional atau yang biasa disebut dalam dunia medis yaitu International Classification of Diseases-10th (ICD-10) dan bagaimana cara untuk mengetahui tingkat *distance error* algoritma Fuzzy K-Means pada data penyakit pasien.

1.3 Batasan Masalah

Adapun batasan-batasan masalah dalam penelitian ini adalah:

1. Kriteria yang digunakan sebagai dasar, diperoleh dari RSUD Kardinah Tegal Jawa tengah. Yaitu berupa data set rekam medis pasien.
2. Data mining hanya sebagai alat bantu untuk menentukan pengelompokan kode penyakit pasien berdasarkan kode internasional.
3. Metode yang digunakan dalam proses ini adalah Klustering dan menggunakan algoritma Fuzzy K-Means yang nantinya akan membagi data kedalam kluster tertentu dan akan dilakukan penghitungan tingkat akurati.
4. Software yang digunakan adalah Matlab R2012a.

1.4 Tujuan Penelitian

Tujuan dari penelitian berdasarkan rumusan masalah yang telah dijelaskan diatas yaitu apakah metode Klustering dengan menggunakan algoritman Fuzzy K-Means dapat dipakai dalam penerapan untuk membagi pengelompokan kluster mana saja yang masuk dalam kode penyakit internasional (ICD10) dan untuk mengetahui *distance error* pada pengimplementasian Fuzzy K-Means pada data penyakit.

1.5 Manfaat Penelitian

Manfaat yang akan diperoleh dari penelitian ini adalah untuk mengetahui

seberapa besarkah tingkat *distance error* yang dihasilkan dari penerapan Fuzzy K-Means pada data penyakit pasien, apakah mendapatkan tingkat *distance error* yang rendah atau malah sebaliknya. Lalu mempermudah pihak RSUD Kardinah dalam proses pengelompokan penyakit pasien, khususnya pada bagian rekam medis. Bagian rekam medis akan dimudahkan untuk mengolah data yang tadinya data tersebut diolah secara manual, tidak tertata, dan hanya menghasilkan informasi terbatas, maka akan di permudah untuk mengelompokan penyakit pasien berdasarkan kode internasional yang akan dibuat dengan menggunakan algoritma Fuzzy K-Means dan metode Klustering.

2. METODE

2.1 Data Mining

Data mining adalah proses yang menggunakan teknik statistic, matematika, kecerdasan buatan, dan machine learning untuk mengekstraksi dan mengidentifikasi informasi yang bermanfaat dan pengetahuan yang terkait dari berbagai database besar [2], berikut adalah langkah-langkah dalam data mining :

1. pemilihan data (data selection), pemilihan data relevan yang didapat dari basis data.
2. pembersihan data (data cleaning), proses menghilangkan noise dan data yang tidak konsisten atau data tidak relevan.

3. pengintegrasian data (data integration), penggabungan data dari berbagai basisdata ke dalam satu basisdata baru
4. transformasi data, data diubah atau digabung ke dalam format yang sesuai untuk diproses dalam data mining.
5. data mining, suatu proses di mana metoda diterapkan untuk menemukan pengetahuan berharga dan tersembunyi dari data;
6. evaluasi pola (pattern evaluation), untuk mengidentifikasi pola-pola menarik untuk di representasikan kedalam knowledge based.
7. representasi pengetahuan (knowledge presentation), visualisasi dan penyajian pengetahuan mengenai teknik yang digunakan untuk memperoleh pengetahuan yang diperoleh pengguna.

2.2 Fuzzy K-Means

Fuzzy C-Means adalah suatu teknik klusterisasi yang mana keberadaannya tiap-tiap titik data dalam suatu klaster ditentukan oleh derajat keanggotaan [3]. Konsep dari *Fuzzy C-Means* pertama kali adalah menentukan pusat klaster, yang akan menandai lokasi rata-rata untuk tiap-tiap klaster. Pada kondisi awal, pusat klaster ini masih belum akurat. Tiap-tiap titik data memiliki derajat keanggotaan

untuk tiap-tiap klaster. Dengan cara memperbaiki pusat klaster dan derajat keanggotaan tiap-tiap titik data secara berulang, maka akan dapat dilihat bahwa pusat klaster akan bergerak menuju lokasi yang tepat. Perulangan ini didasarkan pada minimasi fungsi obyektif yang menggambarkan jarak dari titik data yang diberikan kepusat *cluster* yang terbobot oleh derajat keanggotaan titik data tersebut. Langkah-langkah algoritma FCM adalah sebagai berikut [3] :

1. Tentukan :
 - a. Matriks X berukuran $n \times m$, dengan n = jumlah data yang akan dicluster; dan m = jumlah variabel (kriteria).
 - b. Jumlah variabel klaster (variabel C) yang akan dibentuk, dimana untuk awal mula klaster di-setting bernilai lebih besar sama dengan 2 ($C \geq 2$)
 - c. Tentukan besar variabel pembobot (variabel w), pada fase inialisasi nilai variabel w diberikan lebih dari 1.
 - d. Jumlah maksimum iterasi.
 - e. Kriteria penghentian yang diberikan pada variabel ϵ (ϵ = nilai positif yang sangat kecil).
2. Bentuk matriks partisi awal U (derajat keanggota-an dalam klaster) matriks partisi awal biasanya dibuat secara acak.

3. Hitung pusat cluster V untuk setiap cluster dengan menggunakan rumus dibawah ini :

$$V_{ij} = \frac{\sum_{k=1}^n (\mu_{ik})^w X_{kj}}{\sum_{k=1}^n (\mu_{ik})^w}$$

4. Setelah mendapatkan Centroid pada tiap clusternya langkah selanjutnya adalah menghitung jarak data ke Centroid, dengan menggunakan rumus dibawah ini :

$$D(x_1, c_1) = \sqrt{\sum_{j=1}^p (x_1 - c_1)^2}$$

5. Menghitung nilai keanggotan data untuk membentuk matrix baru, dengan menggunakan rumus dibawah ini :

$$U_{11} = \frac{D(x_1, c_1)^{\frac{-2}{w-1}}}{\sum_{i=1}^k D(x_1, c_i)^{\frac{-2}{w-1}}}$$

$$= \frac{D(x_1, c_1)^{\frac{-2}{w-1}}}{D(x_1, c_1)^{\frac{-2}{w-1}} + D(x_1, c_2)^{\frac{-2}{w-1}} + D(x_1, c_3)^{\frac{-2}{w-1}}}$$

6. Menghitung nilai fungsional objektif, dengan menggunakan rumus dibawah ini:

$$J = \sum_{i=0}^n \sum_{i=0}^n (u_{ij})^w D(x_i, c_l)^2$$

2.3 Klastering

Pada teknik klastering akan dikelompokkan menjadi klaster-klaster berdasarkan kemiripan satu data dengan yang lain. Prinsip dari klastering adalah memaksimalkan kesamaan antar anggota satu klaster dan meminimumkan kesamaan

antar anggota klaster yang berbeda. Kategori algoritma klastering yang banyak dikenal adalah Hirarki Klastering. Hirarki Klastering adalah salah satu algoritma klastering yang dapat digunakan untuk meng-klaster dokumen (document klastering). Dari teknik hirarki klastering, dapat dihasilkan suatu kumpulan partisi yang berurutan, dimana dalam kumpulan tersebut terdapat [4]:

- a. Klaster – klaster yang mempunyai poin – poin individu. Klaster – klaster ini berada di level yang paling bawah.
- b. Sebuah klaster yang didalamnya terdapat poin – poin yang dipunyai semua klaster didalamnya. Single klaster ini berada di level yang paling atas.

2.4 ICD10

ICD-10 adalah klasifikasi statistik, yang berarti bahwa ICD-10 berisi nomor-nomor terbatas dari kategori kode eksklusif yang menggambarkan seluruh konsep penyakit. Struktur hirarki dengan subdivisi-subdivisi untuk mengidentifikasi kelompok besar dan sesuatu yang spesifik. Tujuan ICD-10 diantaranya adalah untuk mendapatkan rekaman sistematis, melakukan analisis, interpretasi dan Menyediakan informasi diagnosis dan tindakan bagi *riset*, edukasi dan kajian *assesment* kualitas keluaran [5].

2.5 Metode Pengumpulan Data

Sesuai dengan sumber data dan tujuan penyusunan penelitian ini serta untuk mendapatkan data yang benar-benar akurat dan relevan, maka dalam pengumpulan data penulis menggunakan beberapa teknik yaitu teknik Observasi. Penelitian ini penulis menggunakan observasi pada database pasien RSUD Kardinah Kota Tegal. Observasi langsung digunakan untuk mengamati atau mendapatkan data-data yang telah ditentukan aspek yang akan diobservasi dari data pasien pada bagian rekam medis di RSUD Kardinah. Observasi langsung juga memungkinkan bagi seorang observer untuk mendapatkan data secara detail atau terperinci sesuai apa yang dibutuhkan. Parameter yang dijadikan acuan yaitu Diagnosa penyakit berdasarkan International Classification of Diseases - 10th (ICD-10) (Diagnosis Utama), berikut gambar field data yang akan diolah.

| TTD | Diagnosa | |
|-----|----------|------------|
| | Utama | Sekunder |
| 10 | 11 | 12 |
| ada | N189 | |
| ada | O020 | |
| ada | O064 | |
| ada | D638 | I120 |
| ada | E145 | I619; G819 |
| ada | O820 | O755; Z301 |
| ada | O809 | |
| ada | P211 | P399 |
| ada | O700 | O721 |
| ada | O200 | |
| ada | H160 | |
| ada | K746 | I10; R162 |
| ada | C229 | |
| ada | D638 | I120 |
| ada | P071 | P229; P599 |
| ada | D638 | I120 |
| ada | O809 | O141 |
| ada | N40 | |
| ada | C921 | O630 |
| ada | D332 | |
| ada | D319 | |

Gambar 2.1 data yang akan diolah

Data diatas akan di pindahkan ke excel dengan ekstensi .xlsx seperti pada

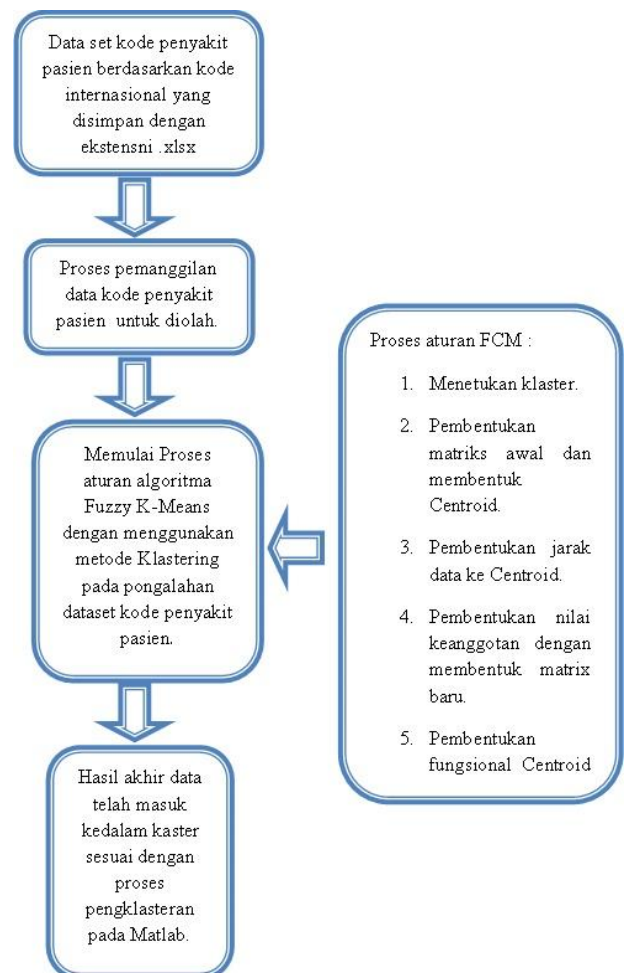
gambar dibawah ini berupa contoh data sample yang diolah :

| | A |
|----|----------------------------|
| 1 | Data kode penyakit (ICD10) |
| 2 | 1,150 |
| 3 | 7,109 |
| 4 | 4,145 |
| 5 | 1,309 |
| 6 | 7,113 |
| 7 | 19,062 |
| 8 | 10,069 |
| 9 | 1,209 |
| 10 | 15,820 |
| 11 | 14,359 |
| 12 | 1,162 |
| 13 | 7,602 |
| 14 | 11,746 |
| 15 | 2,638 |
| 16 | 2,598 |
| 17 | 18,421 |
| 18 | 11,628 |
| 19 | 19,031 |
| 20 | 18,509 |

Gambar 2.2 Data mentah

2.6 Metodeologi Penelitian

Langkah dan aturan penelitian ini ditunjukkan oleh gambar dibawah ini:



Gambar 2.3 analisis perancangan

2.7 Implementasi Penelitian

Disini akan dibahas langkah-langkah implementasi dan analisis hasil penelitian data penyakit pasien berdasarkan kode internasional untuk membentuk sebuah kluster penyakit yang lebih tinggi levelnya dengan menggunakan algoritma Fuzzy K-Means.

Data yang digunakan pada penelitian ini adalah dataset kode penyakit berdasarkan kode internasional pada RSUD Kardinah Kota Tegal. Data yang digunakan dengan jumlah 234 dataset penyakit pasien berdasarkan kode penyakit internasional. Untuk menganalisis kode penyakit yang masuk dalam Kluster tertentu maka akan digunakan Software Matlab untuk menghasilkan pembentukan Kluster pada data set tersebut.

3. PEMBAHASAN PENELITIAN

3.1 Hasil dan Analisis

Hasil yang didapatkan keseluruhan sebanyak 44 iterasi dari 234 data penyakit berdasarkan kode penyakit internasional yang dikelompokkan kedalam 17 kluster dengan hasil perubahan fungsional objektif sebesar 13637927,740369 pada iterasi ke 44 dan data sudah terkluster kedalam kluster yang terbentuk.

```

Terbentuk Iterasi ke = 14, Nilai Fungsional Objektif = 13638206.942091
Terbentuk Iterasi ke = 15, Nilai Fungsional Objektif = 13638088.578882
Terbentuk Iterasi ke = 16, Nilai Fungsional Objektif = 13638020.228692
Terbentuk Iterasi ke = 17, Nilai Fungsional Objektif = 13637980.784702
Terbentuk Iterasi ke = 18, Nilai Fungsional Objektif = 13637958.092888
Terbentuk Iterasi ke = 19, Nilai Fungsional Objektif = 13637945.075371
Terbentuk Iterasi ke = 20, Nilai Fungsional Objektif = 13637937.625966
Terbentuk Iterasi ke = 21, Nilai Fungsional Objektif = 13637933.371299
Terbentuk Iterasi ke = 22, Nilai Fungsional Objektif = 13637930.944968
Terbentuk Iterasi ke = 23, Nilai Fungsional Objektif = 13637929.562903
Terbentuk Iterasi ke = 24, Nilai Fungsional Objektif = 13637928.776360
Terbentuk Iterasi ke = 25, Nilai Fungsional Objektif = 13637928.329034
Terbentuk Iterasi ke = 26, Nilai Fungsional Objektif = 13637928.074758
Terbentuk Iterasi ke = 27, Nilai Fungsional Objektif = 13637927.930273
Terbentuk Iterasi ke = 28, Nilai Fungsional Objektif = 13637927.848198
Terbentuk Iterasi ke = 29, Nilai Fungsional Objektif = 13637927.801586
Terbentuk Iterasi ke = 30, Nilai Fungsional Objektif = 13637927.775117
Terbentuk Iterasi ke = 31, Nilai Fungsional Objektif = 13637927.760089
Terbentuk Iterasi ke = 32, Nilai Fungsional Objektif = 13637927.751558
Terbentuk Iterasi ke = 33, Nilai Fungsional Objektif = 13637927.746715
Terbentuk Iterasi ke = 34, Nilai Fungsional Objektif = 13637927.743966
Terbentuk Iterasi ke = 35, Nilai Fungsional Objektif = 13637927.742405
Terbentuk Iterasi ke = 36, Nilai Fungsional Objektif = 13637927.741520
Terbentuk Iterasi ke = 37, Nilai Fungsional Objektif = 13637927.741017
Terbentuk Iterasi ke = 38, Nilai Fungsional Objektif = 13637927.740731
Terbentuk Iterasi ke = 39, Nilai Fungsional Objektif = 13637927.740570
Terbentuk Iterasi ke = 40, Nilai Fungsional Objektif = 13637927.740478
Terbentuk Iterasi ke = 41, Nilai Fungsional Objektif = 13637927.740425
Terbentuk Iterasi ke = 42, Nilai Fungsional Objektif = 13637927.740396
Terbentuk Iterasi ke = 43, Nilai Fungsional Objektif = 13637927.740379
Terbentuk Iterasi ke = 44, Nilai Fungsional Objektif = 13637927.740369
  
```

Gambar 3.1 hasil fungsional objektif

Hasil data yang mentah telah diolah sehingga terbentuk kluster dan dikelompokkan sesuai kluster yang terbentuk, terlihat seperti pada gambar dibawah ini :

| No | Data Penyakit (ICD10) | Kluster |
|----|-----------------------|---------|
| 1 | 1,150 | 5 |
| 2 | 7,109 | 13 |
| 3 | 4,145 | 10 |
| 4 | 1,309 | 5 |
| 5 | 7,113 | 13 |
| 6 | 19,062 | 12 |
| 7 | 10,069 | 8 |
| 8 | 1,209 | 5 |
| 9 | 15,820 | 14 |
| 10 | 14,359 | 15 |
| 11 | 1,162 | 5 |
| 12 | 7,602 | 13 |
| 13 | 11,746 | 9 |
| 14 | 2,638 | 5 |
| 15 | 2,598 | 5 |
| 16 | 18,421 | 4 |
| 17 | 11,628 | 9 |
| 18 | 19,031 | 12 |
| 19 | 18,509 | 4 |

Gambar 3.2 hasil kluster terbentuk

3.2 Analisis Pengujian Distance Error

Keseluruhan data yang telah diolah dengan dibentuk 17 kluster dari 234 dataset kode penyakit internasional maka dihasilkan data kode penyakit internasional yang sudah terkelompok atau sudah terkluster seperti pada gambar 2.5. Selanjutnya akan dilakukan penghitungan

distance error pada setiap data yang dibentuk oleh setiap klasternya. Pengujian dilakukan dengan menghitung jarak data pada setiap klaster menggunakan perhitungan jarak minkowski dengan rumus seperti dibawah ini :

$$RSS_K = \sum X \in \omega k | \vec{x} - \mu(\omega k)$$

Keterangan :

RSS_K = Perhitungan Jarak.

\vec{x} = Data Asli.

$\mu(\omega k)$ = Nilai dari Centeroid klaster.

Dengan memanfaatkan perhitungan jarak minkowski untuk melakukan evaluasi pada hasil pengklasteran Fuzzy K-Means dihasilkan *Distance Error* sebesar 2,553 dari pengklasteran yang telah diolah pada 234 database kode penyakit (ICD10) yang dibentuk 17 klaster. Dengan hasil *distance error* yang didapat maka dikatakan bagus dan baik dalam implementasi FCM pada data pengelompokan kode penyakit pasien, dikarenakan jarak *distance error* yang dihasilkan semakin mendekati angka 0.

4. EVALUASI KESELURUHAN

4.1 Kesimpulan dan Saran

Didapat kesimpulan untuk mereview hasil dari keseluruhan penelitian ini adalah sebagai berikut:

1. Teknik data mining dengan algoritma Fuzzy K-Means dapat diimplementasikan pada sistem pengelompokan kode penyakit

berdasarkan kode internasional (ICD10).

2. Didapatkan pengelompokan kode penyakit berdasarkan kode penyakit internasional menggunakan Fuzzy K-Means, dengan susunan pengelompokan kode penyakit dari 234 dataset sampel yang terbagi menjadi 17 klaster jenis penyakit yang tertera seperti pada Gambar 3.2.
3. Didapatkan hasil tingkat *distance error* sebesar 2,553. Berarti hasil implementasi algoritma FCM pada data kode penyakit berdasarkan kode internasional dikatakan memuaskan karena *distance error* hampir mendekati 0, jika tingkat *distance error* semakin mendekati angka 0 maka *error* pada pengklasteran data kode penyakit pasien semakin sedikit dan hasilnya baik

Tidak lupa pula didapatkan saran untuk mereview hasil dari keseluruhan penelitian ini adalah sebagai berikut:

1. Dapat dilakukan uji coba dengan melakukan modifikasi algoritma Fuzzy K-Means untuk mendapatkan hasil yang lebih maksimal agar tingkat *distance error* semakin rendah dan mendekati angka 0.
2. Penelitian dapat dilanjutkan dengan menggunakan data lain

yang berhubungan dengan pengklasteran atau pengelompokan data.

DAFTAR PUSTAKA

- [1] Idni Irsalina, Endang Supriyati, Tutik Khotimah, " Clustering Gender Berdasarkan Nilai Maksimum Minimum Amplitudo Suara Berbasis Fuzzy C-Means (FCM)," Prosiding SNATIF, 2014.
- [2] Dwi Prabantini, Sistem Pendukung dan Sistem Cerdas. Yogyakarta, 2005.
- [3] Berget Ingunn, Mevik Bjrnhelge, Ns Tormod. New Modifications and Applications of Fuzzy C-means Methodology, Computational Statistics and Data Analysis, (52) 5, 2008, pp. 2403-2418.
- [4] Yunus Abdul Halim. (2012, Desember) Zero - Center. [Online]. http://zero-fisip.web.unair.ac.id/artikel_detail-69819-Sistem%20Data%20%28Database%29-Hierarchical%20Clustering.html diakses pada tanggal 28 November 2014
- [5] Novita Yuliani, "Analisis keakurata Kode diagnosis penyakit Commotion Cerebri pasien rawat inap berdasarkan ICD-10 rekam medis rumah sakit islam Klaten," *INFOKES*, vol. 1, Februari 2010.
- [6] Talitah Titi Mei, " Pembuatan system cerdas untuk deteksi dini penyakit jantung dengan menggunakan algortma Decision Tree dan fuzzy Clustering," 2010.