

DETEKSI PENYAKIT DIABETES DENGAN METODE FUZZY C-MEANS CLUSTERING DAN K-MEANS CLUSTERING

Abdi Praja¹, Chairisni Lubis², Dyah Erny Herdiwindiati³

¹ Jalan Kramat I, Pondok Pinang, Keb. Lama: Jakarta Selatan, 12310, Indonesia
^{2,3} Universitas Tarumanagara, Jalan S Parman no 1: Jakarta Barat, 11440, Indonesia
Email: ¹abdipn@gmail.com.

Abstrak

Diabetes adalah penyakit yang terjadi ketika kandungan glukosa di dalam darah tinggi. Tes glukosa yang menghasilkan keakuratan tinggi harus dilakukan beberapa kali untuk mendeteksi diabetes di dalam tubuh. Beberapa indikator di dalam tubuh dapat menjadi titik awal untuk mendeteksi diabetes. Bagaimanapun juga, keterbatasan seorang tenaga medis dalam mendeteksi dalam jumlah data yang sangat besar dengan cara manual menjadi kendala. Salah satu solusi untuk gap tersebut adalah menggunakan komputer sebagai perhitungan matematika dalam metode pengelompokan K-Means dan Fuzzy C-Means. Pengelompokan terdiri dari kelompok diabetes dan non-diabetes. Pengujian untuk masing-masing metode dilakukan terhadap 9 data. Hasil pengujian terbaik metode K-Means adalah 73,438% dan untuk metode Fuzzy C-Means adalah 82,812%.

Kata kunci: Diabetes, Clustering, K-Means, Fuzzy C-Means

Abstract

Diabetes is a disease that occurs when the sugar in the blood is high. Glucose tests should be taken of several times to make sure diabetes disease detects in the body. Some indicators of the body conditions can be a starting point to detect diabetes. However, the limitations of medical specialist to detect on the large data manually may still become a constraint. One of the solution to cover up the gap between them is performing technology's in term of computation with the calculation of K-Means clustering and Fuzzy C-Means Clustering method. It clusters are composed of a group of diabetic and non-diabetic. Dataset used in the evaluation process is nine variance data. The highest accurate rate for method of K-Means is 73,438% and for the Fuzzy C-Means result is 82,812%.

Keywords: Diabetes, Clustering, K-Means, Fuzzy C-Means

1. PENDAHULUAN

1.1 Latar Belakang Pemasalahan

Menurut pengertian *National Institute of Diabetes and Digestive and Kidney Diseases Health Information Center*, diabetes adalah penyakit yang terjadi ketika kandungan gula di dalam darah terlalu tinggi. Seiring berjalannya waktu, kadar gula di dalam darah yang tinggi dapat menyebabkan masalah kesehatan. Diantaranya adalah sakit jantung, kerusakan saraf, penglihatan, dan penyakit ginjal. International Diabetes Federation tahun 2015 menunjukkan

data, sekitar 415 juta manusia di dunia mempunyai penyakit diabetes, dengan perbandingan satu diantara sebelas orang dewasa yang terjangkit, dan dari setiap enam detik satu orang meninggal karena penyakit ini.

Tes gula darah yang dilakukan di laboratorium dapat dilakukan untuk mengontrol kandungan gula darah dalam tubuh. Namun, tes darah tunggal tidak bisa langsung menghasilkan hasil yang akurat. Tes harus dilakukan beberapa kali lagi untuk memastikan kandungan gula di dalam tubuh. Beberapa indikator kesehatan tubuh dapat menjadi acuan awal deteksi terhadap penyakit diabetes. Namun, keterbatasan tenaga medis untuk mendeteksi secara manual pada jumlah data yang sangat besar masih menjadi kendala. Salah satu cara untuk menutupi gap diantara keduanya adalah memanfaatkan peranan teknologi dengan perhitungan matematika di bidang medis.

Salah satu teknik algoritma yang digunakan adalah *clustering* atau pengelompokan. Metode *clustering* yang akan digunakan pada perancangan ini adalah *K-Means Clustering* dan *Fuzzy C-Means Clustering*. *K-Means Clustering* adalah metode yang mengelompokkan sejumlah objek ke dalam partisi-partisi berdasarkan kategori yang ada dengan melihat titik tengah yang diberikan. Cluster objek dilihat dari jarak objek dengan titik tengah yang terdekat. Setelah mengetahui titik tengah terdekat, objek tersebut akan diklasifikasikan sebagai anggota atau bukan dari kategori tersebut.

Sedangkan algoritma pada *Fuzzy C-Means Clustering* mengakomodasi adanya ketidakpastian dan ketidaktepatan data input. Dengan didasari pada teori himpunan, maka pada tahun 1965, Lotfi A. Zadeh memperkenalkan sebuah logika *fuzzy*. Dasar logika *fuzzy* adalah teori himpunan *fuzzy*. Pada teori himpunan *fuzzy*, peranan derajat keanggotaan sebagai penentu keberadaan elemen dalam suatu himpunan sangatlah penting. Nilai keanggotaan atau derajat keanggotaan atau *membership degree* menjadi ciri utama dalam penalaran dengan logika *fuzzy* tersebut (Sri Kusumadewi, 2003). Jika nilai keanggotaan pada logika Boolean bernilai 1 untuk benar dan 0 untuk salah, pada logika *fuzzy* nilai keanggotaan terletak pada rentan 0 sampai 1.

Logika *fuzzy* dapat diterapkan dalam data clustering atau pengelompokan data, dikenal dengan *Fuzzy Clustering*. *Fuzzy Clustering* adalah salah satu teknik untuk menentukan kelompok optimal dalam suatu ruang vector. Ada beberapa algoritma pengelompokan data pada *fuzzy*, salah satu diantaranya adalah *Fuzzy C-Means Clustering*, metode yang akan digunakan dalam penelitian ini. *Fuzzy C-Means Clustering* adalah suatu teknik pengelompokan yang mana keberadaannya tiap-tiap titik data dalam suatu *cluster* ditentukan oleh derajat keanggotaan. Keluaran dari *Fuzzy C-Means Clustering* ini merupakan deretan pusat cluster dan beberapa derajat keanggotaan untuk tiap-tiap titik data. Informasi ini dapat digunakan untuk membangun suatu *fuzzy inference system*.

Jadi metode *K-Means* akan mengelompokkan suatu data yang menjadi anggota salah satu cluster yang memenuhi syarat, dan data tersebut tidak menjadi anggota cluster yang lainnya. Namun, pada metode *Fuzzy C-Means Clustering* data ditentukan dengan derajat keanggotaan yang terbesar, dengan kata lain data dapat menjadi anggota lebih dari satu cluster. Berasal dari latar belakang tes gula darah yang harus dilakukan beberapa kali untuk mendapatkan hasil yang akurat dan keterbatasan tenaga medis untuk menentukan deteksi pada penyakit diabetes pada beberapa indikator dengan data yang banyak, maka penelitian ini akan merancang aplikasi deteksi penyakit diabetes dengan metode *Fuzzy C-Means Clustering* dan *K-Means Clustering*.

1.2 Diabetes

Menurut *Team Alodokter*, diabetes adalah penyakit yang terjadi ketika kandungan gula di dalam darah terlalu tinggi. Sel-sel dalam tubuh manusia membutuhkan energi dari gula

(glukosa) untuk bisa berfungsi dengan normal. Gula di dalam darah tersebut dikendalikan oleh hormon insulin. Jika tubuh kekurangan insulin atau muncul resistansi terhadap insulin pada sel-sel tubuh, kadar zat gula (glukosa) darah akan meningkat drastis. Inilah yang memicu dan menjadi penyebab penyakit diabetes (diabetes melitus).

Diabetes terdiri dari dua tipe, yaitu diabetes tipe 1 dan diabetes tipe 2. Diabetes tipe 1 adalah karakteristik dari kurangnya produksi hormon insulin. Sedangkan diabetes tipe 2 adalah ketidakefektifan dari penggunaan hormon insulin oleh tubuh. Sekitar 90% penderita diabetes di dunia adalah diabetes tipe 2 [12].

1.2.1 Ciri-ciri Penderita Diabetes

Ciri-ciri penderita diabetes dapat diketahui secara fisik dan hasil tes pada Glucose Tolerance Test di laboratorium. Ciri fisik dan gejala diabetes antara lain:

- Peningkatan Jumlah Urin
- Nafas Berbau
- Badan Mudah Lelah
- Penglihatan Kabur
- Tekanan Darah Tinggi

Selain beberapa ciri-ciri fisik, penyakit diabetes juga dapat diindikasikan dengan menggunakan tes toleransi kandungan glukosa dalam darah di laboratorium atau Glucose Tolerance Test, indikatornya dapat dilihat pada tabel dibawah:

Table 1. Hasil Glucose Tolerance Test pada indikasi penyakit diabetes [10]

Darah diambil Ketika	Prediabetes	Diabetes
Puasa	100-125 mg/dL	> 126 mg/dL
Setelah 2 jam	140-199 mg/dL	> 200 mg/dL

1.3 Tujuan dan Kegunaan Rancangan

Tujuan dari perancangan yang disusun ini adalah untuk mengelompokkan data pasien penderita diabetes dan non diabetes menggunakan metode *K-Means Clustering* dan *Fuzzy C-Means Clustering* dengan dataset *National Institute of Diabetes and Digestive and Kidney Diseases*.

Unjuk kerja rancangan ini dievaluasi dengan perbandingan data hasil pengelempokan masing-masing metode terhadap data sebenarnya. Pembagian kelompok pada dataset adalah diabetes dengan nilai 1 dan non diabetes dengan nilai 0. Hasil pengelempokan dari masing-masing metode juga menghasilkan kelas dengan nilai 1 pada penderita diabetes dan kelas 0 pada penderita non diabetes. Kegunaan dari sistem ini adalah mengelompokkan data penderita diabetes dan non diabetes dari dataset. Sehingga pada hasil prosesnya didapat dua nilai centroid. Nilai centroid tersebut adalah nilai titik tengah data diabetes dan non diabetes yang menjadi acuan pada data pengelempokan. Dataset terdiri dari 8 indikator tubuh, yaitu:

- A. Number of Pregnancy (jumlah kehamilan)
- B. Plasma Glucose Concentration a 2 hours in an Oral Glucose Tolerance Test (konsentrasi kandungan glukosa pada Glucose Tolerance Test)
- C. Diastolic Blood Pressure (tekanan diastol)
- D. Triceps Skin Fold Thickness (Ketebalan Lipatan Kulit Trisep)

- E. 2-Hour Serum Insulin (2 jam Serum Insulin)
- F. Body Mass Index (masa tubuh)
- G. Diabetes Pedigree Function (Silsilah keturunan diabetes)
- H. Age (umur)

Sepuluh baris pertama dataset penderita diabetes dan non-diabetes yang digunakan adalah:

Table 2. Sepuluh baris dataset

A	B	C	D	E	F	G	H	Result	Detail
8	133	72	0	0	32.9	0.270	39	1	Diabetes
3	158	76	36	245	31.6	0.851	28	1	Diabetes
11	103	68	40	0	46.2	0.126	42	0	Non Diabetes
5	99	54	28	83	34.0	0.499	30	0	Non Diabetes
5	139	80	35	160	31.6	0.361	25	1	Diabetes
1	89	66	23	94	28.1	0.167	21	0	Non Diabetes
2	155	52	27	540	38.7	0.240	25	1	Diabetes
0	125	96	0	0	22.5	0.262	21	0	Non Diabetes
6	104	74	18	156	29.9	0.722	41	1	Diabetes
6	96	0	0	0	23.7	0.190	28	0	Non Diabetes

1.4 Batasan Rancangan

Batasan rancangan yang terdapat pada sistem ini adalah sebagai berikut:

- Dataset terdiri dari wanita keturunan Pima Indian dan berusia minimal 21 tahun yang mempunyai penyakit diabetes dan tidak.
- *Clustering* terdiri dari 2, yaitu penderita diabetes dan non diabetes.

Pesatnya kemajuan teknologi di dunia informasi dalam berbagai bidang telah membawa dampak besar bagi umat manusia. Pada bidang pendidikan contohnya, teknologi telah membantu kemudahan proses belajar mengajar bagi pengajar atau pelajar dengan membuat proses kerja menjadi lebih efisien, contohnya pelajar bisa melihat nilai atau mengambil mata kuliah dimana saja secara online atau pengajar bisa mengambil dan membawakan materi dengan mudah melalui perangkat keras laptop atau komputer.

Contoh aplikasi akademis yang akan membantu dalam bidang pendidikan adalah aplikasi untuk melakukan prediksi status kelulusan seorang mahasiswa di awal masa perkuliahan, sehingga mahasiswa bisa menyadari hal yang diperlukan untuk membantu dirinya di semester lanjut dan membantu pekerjaan pihak yang berwenang dalam menangani mahasiswa yang dianggap akan mengalami kesulitan untuk lulus tepat waktu.

Dari hasil Tanya jawab diketahui bahwa untuk sekarang, belum ada sistem yang mengurus tentang prediksi kelulusan mahasiswa. Di fakultas teknologi informasi universitas tarumanagara, khususnya Penilaian mahasiswa yang kira-kira perlu mendapat bimbingan dilakukan dengan cara melihat jumlah IPK dan SKS yang dapat diambil mahasiswa di semester empat. Jika IPK dinilai terlalu rendah (dua ke bawah) dan SKS yang dapat diambil 10 atau

kurang, maka mahasiswa akan dipanggil untuk menanyakan penyebab dan kesulitannya berkuliah. Setelah itu mahasiswa akan diberi solusi atau sugesti yang dapat membantunya.

Untuk itu dibuatlah sebuah aplikasi prediksi kelulusan mahasiswa yang dibuat untuk Dosen Ketua Jurusan Universitas Tarumanagara yang akan berfungsi untuk memprediksi status kelulusan mahasiswa (lulus tepat waktu atau lulus tidak tepat waktu) berdasarkan nilai mahasiswa dalam mata kuliah yang dianggap sulit dan/atau penting di semester 1 sampai 4 sehingga dosen ketua jurusan bisa mengambil keputusan dengan lebih cepat dan akurat. Aplikasi prediksi kelulusan ini akan menggunakan metode *Linear Discriminant Analysis* yaitu sebuah metode yang menggunakan nilai dari beberapa variable bebas untuk menentukan kecenderungan masuknya suatu objek ke dalam satu kelompok (variable terikat) yang telah ditentukan yaitu lulus tepat waktu dan tidak tepat waktu. Nama mahasiswa kemudian ditampilkan dengan prediksi dan skor determinan yang didapat dari proses perhitungan fungsi diskriminan. Hal ini diharapkan bisa mempermudah proses prediksi kelulusan untuk mahasiswa yang mengalami masalah dalam mengikuti beberapa pembelajaran, sehingga pihak fakultas bisa membantu dengan memanggil mahasiswa bersangkutan untuk keperluan konseling atau pencarian solusi.

2. PENELITIAN TERKAIT

Penelitian terkait yang sudah dibuat sebelumnya adalah:

- Prediksi diabetes dengan menggunakan dataset yang sama pada penelitian ini oleh *Igor Shvartser* pada tahun 2014. Algoritma klasifikasi yang digunakan adalah Naïve Bayes Classification. Tingkat keakuratan yang dihasilkan adalah 75,41%.
- Program Aplikasi Verifikasi Simbol pada Peta dengan Menggunakan Metode Fuzzy C-Means Clustering. Proses pertama yang dilakukan adalah mencari ekstraksi ciri dengan region pada piksel image dan menggunakan ekstraksi ciri untuk dikelompokkan dengan metode Fuzzy C-Means Clustering sebagai nilai pada verifikasi input. Hasil yang didapatkan adalah 64,1%.

3. METODE PENELITIAN

3.1 *K-Means Clustering*

Metode *K-Means Clustering* pertama kali diperkenalkan oleh MacQueen JB pada tahun 1976. Metode ini adalah salah satu metode yang paling umum digunakan [5]. Metode ini termasuk dalam teknik *partition-based clustering* yang membagi atau memisahkan objek ke – k dalam daerah bagian yang terpisah. Pada metode *K-Means Clustering*, setiap objek akan masuk ke kelompok tertentu dalam suatu proses tertentu.

Algoritma metode *K-Means* merupakan model *centroid*. Model *centroid* adalah model yang menggunakan titik tengah suatu *cluster* yang berupa nilai. *Centroid* digunakan untuk menghitung jarak suatu objek terhadap titik tengah. Suatu objek data termasuk dalam suatu cluster jika memiliki jarak terpendek terhadap *centroid cluster* tersebut.

Secara umum algoritma K-Means adalah:

1. Menentukan banyaknya *cluster* (k).
2. Menentukan *centroid*.
3. Hitung jarak objek ke *centroid*.
4. Data dikelompokkan dengan jarak terdekat.
5. Cek apakah ada objek yang berpindah:
Jika ya, kembali ke langkah 2.

Jika tidak, selesai.

Rumus untuk menghitung jarak yang digunakan adalah *Euclidean Distance*. Rumus tersebut adalah (Mathworld Wolfram, 2016):

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

Keterangan:

d = jarak;

j = banyaknya data;

c = centroid;

x = data.

3.2 Fuzzy Logic

Ide himpunan *fuzzy* (*fuzzy set*) diawali dari matematika dan teori sistem Lotfi. A Zadeh pada tahun 1965. Jika diterjemahkan, “*fuzzy*” artinya tidak jelas/buram, tidak pasti. Himpunan *fuzzy* adalah cabang dari matematika yang mempelajari proses bilangan random; teori probabilitas, statistik matematik, teori informasi, dan lainnya. Penyelesaian masalah dengan himpunan *fuzzy* lebih mudah dari pada dengan menggunakan teori probabilitas (konsep pengukuran).

Fuzzy Logic merupakan perkembangan dari logika *Boolean* yang berhubungan dengan konsep kebenaran parsial. Dapat dikatakan kebenaran nilai logikanya diekspresikan secara tidak mutlak. *Fuzzy Logic* menggunakan nilai antara 0 dan 1 yang menggambarkan derajat kepercayaan / kebenaran suatu logika, atau disebut juga membership degree. Semakin besar nilai pada membership degree atau semakin mendekati 1, maka nilai kepercayaan atau kebenarannya semakin besar.

Pada teori himpunan klasik atau *crisp*, keberadaan suatu elemen pada suatu himpunan A, hanya akan memiliki 2 kemungkinan keanggotaan, yaitu menjadi anggota A atau tidak menjadi anggota A (Chak, 1998). Suatu nilai yang menunjukkan seberapa besar suatu tingkat keanggotaan suatu elemen (x) dalam suatu himpunan (A), sering dikenal dengan nama nilai keanggotaan atau derajat keanggotaan, dinotasikan dengan $\mu_A(X)$. Pada himpunan klasik ini hanya ada 2 nilai keanggotaan, yaitu $\mu_A(X) = 1$ untuk x menjadi anggota A; dan $\mu_A(X) = 0$ untuk x bukan anggota A.

Sedangkan pada himpunan *fuzzy*, jika X adalah koleksi obyek-obyek yang dinotasikan secara generik oleh x, maka suatu himpunan fuzzy A, dalam X adalah himpunan pasangan yang berurutan:

$$A = \{(X, \mu_A(X)) \mid x \in X\} \quad (2)$$

Dengan $\mu_A(X)$ adalah derajat keanggotaan x di A yang memetakan X ke ruang keanggotaan M yang terletak pada rentang [0,1].

3.3 Fuzzy C-Means Clustering

Fuzzy C-Means Clustering adalah teknik *clustering* atau pengelompokan data yang keberadaan tiap titik data cluster ditentukan oleh derajat keanggotaan fuzzy. Teknik *Fuzzy* ini pertama kali diperkenalkan oleh Jim Bezdek tahun 1981 [7]. Konsep dasar algoritma ini adalah menentukan pusat cluster yang akan menandai lokasi rata-rata untuk setiap clusternya.

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 [7]

1. Tentukan:
 - Matriks X berukuran n x m
n = jumlah data yang akan di-cluster
m = jumlah variable (atribut);
 - Jumlah cluster yang akan dibentuk = c
c harus bernilai ≥ 2 ;
 - Pangkat (pembobot) = w
w bernilai ≥ 2 ;
 - Maksimum iterasi = MaxIter;
 - Kondisi iterasi berhenti (error terkecil) = ξ ;
 - Iterasi awal, t = 1 dan $\Delta = 1$;
2. Bangkitkan bilangan random Q_i dimana $i = 1,2,3,\dots,n$; $k = 1,2,3,\dots,c$; sebagai elemen-elemen matriks partisi awal U.

$$Q_i = \sum_{k=1}^c Q_{ik} \quad (3)$$

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

Hitung:

$$\mu_{ik} = \frac{Q_{ik}}{Q_i} \quad (4)$$

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

$$V_{kj} = \frac{\sum_{i=1}^n ((\mu_{ik})^w * X_{ij})}{\sum_{i=1}^n (\mu_{ik})^w} \quad (5)$$

Keterangan:

V = Pusat *cluster*

k = 1,2,...c

j = 1,2,...m

i = 1,2,...n

4. Hitung fungsi objektif pada iterasi ke-t, P_t :

$$P_t = \sum_{i=1}^n \sum_{k=1}^c ([\sum_{j=1}^m (X_{ij} - V_{kj})^2] (\mu_{ik})^w) \quad (6)$$

5. Hitung perubahan derajat keanggotaan yang baru pada matriks partisi:

$$\mu_{ik} = \frac{[\sum_{j=1}^m (X_{ij} - V_{kj})^2]^{-\frac{1}{w-1}}}{\sum_{k=1}^c [\sum_{j=1}^m (X_{ij} - V_{kj})^2]^{-\frac{1}{w-1}}} \quad (7)$$

Keterangan:

i = 1,2,...n

k = 1,2,...c

6. Cek kondisi:
 - a. Jika: $(|P_t - P_{t-1}| < \xi)$ atau $(t > \text{MaxIter})$, 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. PEMBAHASAN

Hasil eksperimen dengan metode *K-Means Clustering*, dihasilkan nilai presentase tertinggi yaitu eksperimen menggunakan Data Acak 4. Banyaknya iterasi yang dihasilkan adalah 7 kali dan tingkat kebenarannya adalah 73,438%. Sedangkan hasil terendah yang diperoleh adalah eksperimen dengan menggunakan Data Nol, dengan banyak iterasi 2 kali dan tingkat kebenarannya 61,979%. Eksperimen dengan metode *K-Means Clustering* pada semua data ditunjukkan pada tabel 4 dibawah ini.

Table 4. Hasil Eksperimen Metode *K-Means Clustering*

No	Nama Data	Banyaknya Iterasi	Tingkat Kebenaran
1.	Data Keseluruhan	4	65,625%
2.	Data Nol	2	61,979%
3.	Data Non Nol	11	67,263%
4.	Data Ambigu	5	71,812%
5.	Data Non Ambigu	8	72,340%
6.	Data Acak 1	7	64,062%
7.	Data Acak 2	9	68,056%
8.	Data Acak 3	10	70,667%
9.	Data Acak 4	7	73,438%

Eksperimen dengan metode *Fuzzy C-Means* dilakukan dengan menetapkan nilai *Minimum Iteration*: 500, *Maximum Error*: 0,000005. Data eksperimen Acak 1 sampai 4 menggunakan nilai *range* eksponen 2 sampai 6. Setelah dilakukan eksperimen, dihasilkan nilai presentase tingkat kebenaran tertinggi yaitu eksperimen menggunakan Data Acak 4 dengan nilai eksponen 6. Banyaknya iterasi yang dihasilkan adalah 38 kali dan tingkat kebenarannya

mencapai 82,812%. Sedangkan hasil terendah yang diperoleh adalah eksperimen dengan menggunakan Data Keseluruhan, dengan hasil iterasi 43 kali dan tingkat kebenarannya 65,885%. Hasil eksperimen dengan metode *Fuzzy C-Means Clustering* pada semua data eksperimen, ditunjukkan pada table 5 di bawah ini.

Table 5. Hasil Eksperimen Metode *Fuzzy C-Means Clustering*

No	Nama Data	Exponen (w)	Banyaknya Iterasi	Tingkat Kebenaran
1.	Data Keseluruhan	2	43	65,885%
2.	Data Nol	2	31	69,531%
3.	Data Non Nol	2	56	68,490%
4.	Data Ambigu	2	51	69,128%
5.	Data Non Ambigu	2	40	72,766%
		2	59	65,625%
		3	54	66,146%
6.	Data Acak 1	4	45	66,146%
		5	40	66,667%
		6	39	67,177%
		2	35	72,917%
		3	36	72,917%
7.	Data Acak 2	4	40	72,917%
		5	40	73,958%
		6	53	73,438%
		2	36	70,157%
		3	37	71,728%
8.	Data Acak 3	4	36	73,298%
		5	37	74,869%
		6	36	75,916%
		2	56	71,875%
		3	48	75,521%
9.	Data Acak 4	4	48	79,167%
		5	45	81,771%
		6	38	82,812%

4. KESIMPULAN

Kesimpulan dari penelitian deteksi diabetes dengan menggunakan metode *K-Means Clustering* dan *Fuzzy C-Means Clustering* adalah:

1. Hasil *clustering* pada kedua metode menunjukkan hasil yang lebih baik pada Data Acak 4.
2. Nilai Eksponen (w) pada *Fuzzy C-Means* Clustering semakin besar, jumlah looping semakin kecil, dan tingkat keakuratan semakin besar.
3. *Dataset* terbatas pada populasi Pima Indian, hal ini juga dapat menjadi awal yang baik untuk mendeteksi populasi yang lain. Keterbatasan *dataset* yang diambil pada 1960-1980 menjadi pertimbangan pada hasil *clustering*, saat ini tes diabetes dilengkapi dengan test urin dan hemoglobin A1c test untuk hasil yang lebih akurat.

DAFTAR PUSTAKA

- [1] American Heart Association, Understanding Blood Pressure Readings, http://www.heart.org/HEARTORG/Conditions/HighBloodPressure/AboutHighBloodPressure/Understanding-Blood-Pressure-Readings_UCM_301764_Article.jsp#.V82X75h97Dc, 5 September 2016.
- [2] David S Freedman et al, Relation of circumferences and skinfold thicknesses to lipid and insulin concentrations in children and adolescents: The Bogalusa Heart Study 1,2,3, <http://ajcn.nutrition.org/content/69/2/308.full>, 5 September 2016.
- [3] Erika Gebel, The Science of Diabetes and Pregnancy, <http://www.diabetesforecast.org/2010/aug/the-science-of-diabetes-and-pregnancy.html>, 5 September 2016.
- [4] John D. Cook, what is clustering, <http://bigdata-madesimple.com/what-is-clustering-in-data-mining/>, 31 Agustus 2016.
- [5] Noor Fitriana, Perbandingan Kinerja Metode Linkage, Metode Average Linkage, dan Metode K-Means Dalam menentukan hasil analisis cluster, (Jogyakarta: UNY, 2014), h.22.
- [6] Sudrajat, Dasar-dasar Fuzzy Logic, https://pustaka.unpad.ac.id/wp-content/uploads/2010/07/dasar_dasar_fuzzy_logic.pdf, 30 Agustus 2016.
- [7] Sri Kusumadewi dan Sri Hartanti, Neuro-Fuzzy: Integrasi Sistem Fuzzy & Jaringan Syaraf, (Jogyakarta: Graha Ilmu, 2006), h.282.
- [8] Team Alodokter, Pengertian Diabetes Tipe 2, <http://www.alodokter.com/diabetes-tipe-2/>, 30 Agustus 2016.
- [9] The National Institute of Diabetes and Digestive and Kidney Diseases Health Information Center, Diabetes, <https://www.niddk.nih.gov/health-information/diabetes>, 29 Agustus 2016.
- [10] Tricia Kinman, Glucose Tolerance Test, <http://www.healthline.com/health/glucose-tolerance-test#Overview1>, 5 September 2016.
- [11] Weisstein, Eric W, Distance, <http://mathworld.wolfram.com/Distance.html>, 24 September 2016.
- [12] WHO, Diabetes, (World Health Organization: NMH Fact Sheet February, 2010).