

KINERJA METODE ALGORITMA FUZZY C-MEANS CLUSTERING UNTUK DETEKSI PENYAKIT LIVER

Nuril Amrullah
1410651068
Jurusan Teknik Informatika
Fakultas Teknik
Universitas Muhammadiyah Jember
Jl. Karimata No. 49 Jember, Jawa Timur, Indonesia
nurilamrullah@gmail.com

ABSTRAK

Penyakit hati atau penyakit liver adalah penyakit yang disebabkan oleh berbagai faktor yang merusak hati, seperti virus dan penggunaan alkohol. Obesitas juga berhubungan dengan kerusakan hati. Seiring waktu, kerusakan hati berdampak pada luka di jaringan (sirosis), yang dapat menyebabkan gagal hati, suatu kondisi yang dapat mengancam jiwa. Pada tahun 2011 dilakukan penelitian oleh Bendi Venkata Ramana, Prof. M.Surendra Prasad Babu dan Prof. N. B. Venkateswarlu yang bertujuan untuk meneliti tingkat akurasi metode data mining algoritma naive bayes, c4.5, back propagation k-nearest neighbor, support vector machines (SVM) menggunakan data yang bersumber dari University of California di Irvine (UCI) Machine Learning Repository tentang data penyakit liver. Berdasarkan penelitian sebelumnya yang belum ada penelitian yang mengolah data penyakit liver yang bersumber dari UCI Machine Learning Repository menggunakan metode clustering maka peneliti berniat untuk melakukan penelitian mengenai tingkat akurasi yang dapat dihasilkan menggunakan metode fuzzy c-means untuk mendeteksi penyakit liver. Setelah dilakukan perhitungan didapatkan hasil akurasi terbaik dari penggunaan metode fuzzy c-means untuk deteksi penyakit liver yaitu sebesar 59,86%.

Kata Kunci : Cluster, FCM, Liver, Data Mining

1. PENDAHULUAN

Kesehatan merupakan suatu keadaan baik secara fisik, mental dan psikologis. Menurut undang – undang republik Indonesia nomor 36 tahun 2009 tentang kesehatan menyatakan bahwa kesehatan merupakan keadaan sehat, baik secara fisik, mental, spiritual maupun sosial yang memungkinkan setiap orang untuk hidup produktif secara sosial dan ekonomis.

Penyakit hati atau penyakit liver adalah penyakit yang disebabkan oleh berbagai faktor yang merusak hati, seperti virus dan penggunaan alkohol. Obesitas juga berhubungan dengan kerusakan hati. Seiring waktu, kerusakan hati berdampak

pada luka di jaringan (sirosis), yang dapat menyebabkan gagal hati, suatu kondisi yang mengancam jiwa.

Pada tahun 2011 dilakukan penelitian oleh Bendi Venkata Ramana, Prof. M.Surendra Prasad Babu dan Prof. N. B. Venkateswarlu (2011) dengan judul penelitian “A Critical Study of Selected Classification Algorithms for Liver Disease Diagnosis“, dari hasil yang didapat dari penelitian dengan menggunakan algoritma naive bayes, c4.5, back propagation k-nearest neighbor, support vector machines (SVM) menggunakan data yang bersumber dari University of California di Irvine (UCI) Machine Learning Repository

tentang data penyakit liver. Hasil yang didapat dari penggunaan metode yang digunakan yaitu pada tabel 2.1 :

Classification Algorithms	Accuracy	Precision	Sensitivity	Specificity
NBC	56,52	48,91	77,93	41
C 4.5	68,69	65,81	53,1	80
Back Propagation	71,59	69,74	57,34	82
K-NN	62,89	55,78	56,51	67,5
SVM	58,26	1	68,9	1

Berdasarkan penelitian sebelumnya yang cenderung menggunakan metode klasifikasi dalam mengukur tingkat akurasi menggunakan dataset penyakit liver dan belum adanya penelitian yang mengolah data penyakit liver yang bersumber dari UCI Machine Learning Repository menggunakan metode clustering maka peneliti berniat untuk melakukan penelitian mengenai tingkat akurasi yang dapat dihasilkan menggunakan metode fuzzy c-means untuk mendeteksi penyakit liver.

2. BAHAN DAN METODE

Clustering

Clustering merupakan proses partisi satu set objek data ke dalam himpunan bagian yang disebut dengan cluster. Objek yang di dalam cluster memiliki kemiripan karakteristik antar satu sama lainnya dan berbeda dengan cluster yang lain.

Logika Fuzzy

logika fuzzy merupakan salah satu komponen pembentuk Soft Computing. 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

function menjadi ciri utama dari penalaran dengan logika fuzzy tersebut.

Fuzzy C-Means

Fuzzy c – means merupakan suatu teknik pengclusteran yang keberadaan tiap – tiap data dalam suatu cluster ditentukan oleh derajat keanggotaan. Konsep fuzzy c-means tahap yang pertama adalah menentukan pusat cluster, yang akan menandai lokasi rata-rata untuk tiap-tiap cluster. Pada iterasi awal, pusat cluster ini masih belum akurat. Tiap-tiap titik data memiliki derajat keanggotaan untuk tiap-tiap cluster. Langkah berikutnya adalah Dengan memperbaiki pusat cluster dan derajat keanggotaan tiap-tiap titik data secara berulang, maka akan terlihat bahwa pusat cluster akan bergerak menuju lokasi yang tepat. Perulangan ini didasarkan pada minimasi fungsi obyektif yang menggambarkan jarak dari titik data yang diberikan ke pusat cluster yang terbobot oleh derajat keanggotaan titik data tersebut.

Langkah-langkah algoritma *Fuzzy C-Means* adalah sebagai berikut :

1. Menentukan data yang akan di *cluster X*, berupa matriks berukuran $n \times m$ (n = jumlah sampel data, m = atribut setiap data). X_{ij} = data sampel ke- i ($i = 1,2,\dots,n$), atribut ke- j ($j = 1,2,\dots,m$).
2. Tentukan jumlah *cluster* (c), pangkat (w), maksimum iterasi (MaxIter), *error* terkecil yang diharapkan (ϵ), fungsi obyektif awal ($P_0 = 0$), iterasi awal ($t = 1$).
3. Membangkitkan bilangan *random* μ_{ik} , $i = 1, 2, \dots, n$; $k = 1, 2, \dots, c$; μ_{ik} adalah derajat keanggotaan yang merujuk pada seberapa besar kemungkinan suatu data bisa menjadi anggota ke dalam suatu *cluster*. Posisi dan nilai matriks dibangun secara *random*. Dimana nilai keanggotaan terletak pada interval 0 sampai dengan 1. Pada posisi awal matriks partisi masih belum akurat begitu juga pusat clusternya. Sehingga kecenderungan data untuk masuk suatu *cluster* juga belum akurat.

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

dengan,

c = jumlah *cluster*

k = indeks *cluster*

μ_{ik} = nilai keanggotaan

4. Hitung pusat *cluster* ke- $k = V_{kj}$, dengan $k = 1, 2, \dots, c$; dan $j = 1, 2, \dots, m$

$$V_{kj} = \frac{\sum_{i=1}^n ((\mu_{ik})^w * X_{ij})}{\sum_{i=1}^n (\mu_{ik})^w} \quad (2)$$

dengan,

V_{kj} = hasil pusat *cluster* yang dihasilkan

n = jumlah data

μ_{ik} = nilai keanggotaan

w = pangkat

X_{ij} = data ke- i dan parameter ke- j

5. Hitung fungsi obyektif pada iterasi ke- t , P_t = Fungsi obyektif digunakan sebagai syarat perulangan untuk mendapatkan pusat *cluster* yang tepat. Sehingga diperoleh kecenderungan data untuk masuk ke *cluster* mana pada step akhir.

Untuk iterasi awal nilai $t = 1$.

$$P_t = \sum_{i=1}^n \sum_{k=1}^c \left(\left[\sum_{j=1}^m (X_{ik} - V_{kj})^2 \right] (\mu_{ik})^w \right) \quad (3)$$

dengan,

P_t = fungsi obyektif

n = Jumlah data

C = jumlah *cluster*

m = jumlah parameter

μ_{ik} = nilai keanggotaan

X_{ik} = data ke- i dan parameter ke- $j = V_{kj}$

= hasil pusat *cluster* yang dihasilkan

6. Hitung perubahan matriks partisi:

$$\mu_{ik} = \frac{[\sum_{j=1}^m (X_{ik} - V_{kj})^2]^{-\frac{1}{w-1}}}{\sum_{k=1}^c [\sum_{j=1}^m (X_{ik} - V_{kj})^2]^{-\frac{1}{w-1}}} \quad (4)$$

7. Cek kondisi berhenti:

1). $|P_t - P_{t-1}| < \zeta$ atau $(t > \text{MaxIter})$ maka berhenti;

2). Jika tidak, iterasi dinaikkan $t = t+1$, ulangi langkah ke-4.

3. HASIL DAN PEMBAHASAN

Setelah program selesai dibuat, dilakukan tahap pengujian sistem. Data yang digunakan adalah data sekunder berjumlah 583 data yang diperoleh dari media internet pada website *UCI Machine Learning Repository* data set tentang *liver* sebagai kasus yang dibahas dalam tugas akhir ini.

Untuk melakukan uji coba pada aplikasi *fuzzy c-means* maka dilakukan skenario uji coba sebagai berikut :

4.2.1 Pengujian Mencari Nilai Matrix Awal

Percobaan dilakukan sampai 10 kali yang bertujuan untuk mencari nilai akurasi tertinggi dengan jumlah iterasi yang sedikit, setelah ditemukan nilai akurasi tertingginya maka akan diambil nilai matriks randomnya untuk penentuan matriks awal.

Skema uji dilakukan dengan memasukkan nilai iterasi maksimum 100, jumlah *cluster* 2 dan nilai *error* terkecil 10^{-5} . Hasil uji coba ditampilkan dalam tabel di bawah ini :

Tabel 4.1 : Hasil Uji Coba

Pengujian Ke	Jumlah Iterasi	Nilai Akurasi
1	59	59,86%
2	53	59,86%
3	57	40,13%
4	56	59,86%
5	53	40,13%
6	54	59,86%
7	54	59,86%
8	55	59,86%
9	54	59,86%
10	54	59,86%

Dari hasil percobaan di atas dapat diambil kesimpulan bahwa pengujian ke-2 dapat digunakan karena memiliki jumlah iterasi yang sedikit dengan nilai akurasi tertinggi sebesar 59,86%. setelah proses selesai kemudian diambil matriks awal untuk digunakan dalam proses perhitungan *fuzzy c-means*. Kemudian nilai matriks yang didapat akan disimpan ke dalam *data*

base untuk selanjutnya akan dilakukan proses perhitungan oleh sistem. Nilai matriks awal dapat dilihat di lampiran ke-1.

4.2.2 Pengujian dengan mengubah nilai error

Percobaan ini dilakukan dengan memasukkan nilai *error* mulai dari 10^4 sampai 10^{-4} untuk di cari nilai akurasi tertinggi dengan proses perhitungan tercepat. Berikut adalah tabel hasil percobaan yang dilakukan dengan mengubah nilai eror :

Tabel 4.2 : Hasil Percobaan Nilai Error

nilai eror	jumlah iterasi	nilai akurasi	waktu	fungsi objectif
10000	11	59,34%	7,18	7073,05
1000	14	59,51%	8,20	533,67
100	17	59,86%	7,34	62
10	21	59,86%	8,05	6,29
0,01	35	59,86%	10,1	0,006
0,001	39	59,86%	12,1	0,0008
0,0001	44	59,86%	12	(44847858.770648 - 44847858.770648) = 7.7612698078156E-5
0,00001	49	59,86%	11,5	(44847858.770535 - 44847858.770535) = 6.8098306655884E-6

Dari percobaan di atas dapat disimpulkan bahwa nilai *error* terbaik adalah nilai *error* yang memiliki jumlah iterasi yang sedikit dan nilai akurasi yang besar serta proses perhitungan yang cepat. Berdasarkan tabel 4.2 di atas pada pengujian ke-3 dengan nilai *error* 100 memiliki jumlah iterasi yang sedikit dibandingkan dengan percobaan yang lain, nilai akurasi yang di hasilkan sebesar 59,86% dengan lama proses terendah dari uji coba lainnya yaitu sebesar 7,34 detik.

Perhitungan akurasi yang digunakan dalam percobaan di atas dilakukan menggunakan rumus sebagai berikut :

$$Akurasi = \frac{TP + TN}{TP + FN + FP + TN} \times 100\%$$

Keterangan :

TP = *True Positif* merupakan nilai benar dengan nilai aktual 1 prediksi 1.

TN = *True Negatif* merupakan nilai benar dengan nilai aktual 2 prediksi 2.

FN = *False Negatif* merupakan nilai salah dengan nilai aktual 1 prediksi 2

FP = *False Positif* merupakan nilai salah dengan nilai aktual 2 prediksi 1

Sehingga perhitungan akurasi pada perbandingan metode *fuzzy c-means* dapat digambarkan dengan perhitungan berikut:

Diketahui :

Nilai TP = 345 data

Nilai TN = 4 data

Total seluruh data = 583 data

$$Akurasi = \frac{349}{583} \times 100\%$$

$$Akurasi = 59,86\%$$

Setelah dilakukan perhitungan didapatkan hasil akurasi dari penggunaan metode *fuzzy c-means* untuk deteksi penyakit *liver* yaitu sebesar 59,86%.

KESIMPULAN

Beberapa kesimpulan yang dapat diambil dari hasil pengujian yang telah dilakukan sebagai berikut :

1. Dalam mencari nilai matriks awal diperlukan beberapa kali uji coba, dari hasil uji coba yang sudah dilakukan diambil nilai matriks dengan jumlah iterasi terendah dengan tingkat akurasi yang tinggi sehingga matriks awal dapat digunakan untuk proses selanjutnya.
2. Berdasarkan hasil uji coba yang dilakukan dengan menggunakan matriks awal yang sudah ditemukan sebelumnya, waktu terbaik yang dihasilkan dari penggunaan matriks tersebut yaitu sebesar 7,14 detik dengan tingkat akurasi sebesar 59,86%. Perhitungan waktu ini menggunakan fungsi *microtime* di pemrograman PHP.

3. Dari hasil uji coba pada skenario kedua yaitu mengubah nilai error ditemukan bahwa nilai error di bawah 10^{-3} nilai fungsi objektif tidak dapat menunjukkan nilai yang sesuai, sehingga memakan waktu yang lama dalam memproses data.

Saran

Beberapa saran yang dapat dilakukan untuk pengembangan penelitian ini selanjutnya adalah :

1. Penentuan matriks awal sebaiknya menggunakan kombinasi dengan algoritma Fuzzy Subtractive Clustering agar mendapatkan partisi pengelompokan yang lebih baik dari pada algoritma standar.
2. Penelitian ini dapat dilanjutkan dengan menggunakan metode-metode clustering data mining yang lain seperti K-Means.

DAFTAR PUSTAKA

- Astuti, Nur., Oni Soesanto, dan Dwi Kartini. 2017. *Algoritma Fuzzy C-Means (FCM) Untuk Penentuan Nilai Center Radial Basis Function (RBF) Pada Klasifikasi Data Penyakit Karies Gigi*. Universitas Lambung Mangkurat, Kalimantan Selatan.
- Arief, M.Rudianto. 2011. *Pemrograman Web Dinamis Menggunakan Php dan Mysql*. Yogyakarta : ANDI.
- Cahyaningsih, Dwi. 2017. *Optimasi Aturan Fuzzy Dalam Sistem Fuzzy Sugeno Orde Nol Dengan Fuzzy C-Means Clustering Pada Diagnosis Kanker Otak*. Skripsi. Program Studi Matematika Fakultas Matematika Dan Ilmu Pengetahuan Alam Universitas Negeri Yogyakarta.
- Efiyah, Uum. 2014. *Penerapan algoritma fuzzy C-Means untuk pengelompokan harga gabah di tingkat penggilingan berdasarkan kualitas gabah*. Skripsi. Universitas Islam Negeri Maulana Malik Ibrahim..
- Irwansyah, Edy. 2017. *Clustering*. Diakses 18, April, 2018 dari <https://socs.binus.ac.id/2017/03/09/clustering/>.
- Junqueira LC, Carneriro J. 2007. *Histologi dasar teks dan atlas*. Jakarta : EGC
- Kusumadewi, sri dan Hari Purnomo. 2010. *Aplikasi Logika Fuzzy untuk Pendukung Keputusan Edisi 2*. Yogyakarta : Graha Ilmu.
- Kusumadewi, Sri dan Purnomo H., 2004, *Aplikasi Logika Fuzzy untuk Pendukung Keputusan*. Yogyakarta : Penerbit Graha Ilmu.
- Kusrini, Luthfi, Taufiq Emha, 2009. *Algoritma Data Mining*, Penerbit Andi, Yogyakarta.
- M, Nimas Mita Etika. 2018. *Apa Artinya Jika Kadar Alkali Fosfatase Tinggi Setelah Tes Darah?*. Diakses 20, April, 2018 dari <https://helohehat.com/hidup-sehat/tips-sehat/alkali-fosfatase-tinggi/>.
- Menteri Kesehatan. 2009. *Undang-Undang Republik Indonesia Nomor 36 Tahun 2009 Tentang Kesehatan*. Jakarta.
- Muhlisin, Ahmad. 2018. *SGOT dan SGPT : Nilai Normal, Tinggi, Rendah & Maknanya*. Diakses 20, April 2018 dari <https://mediskus.com/sgot-sgpt>.
- Nugroho, Adi. 2006. *E-commerce*. Bandung : Informatika Bandung.
- Ramana, Bendi Venkata, M.Surendra Prasad Babu dan N. B. Venkateswarlu. 2011. *A Critical Study of Selected Classification Algorithms for Liver Disease Diagnosis*. India

- Ritonga, Pahmi 2015. *Pengertian mysql menurut para pakar*. Diakses 18, April 2018 dari <http://www.bangpahmi.com/2015/03/pengertian-mysql-menurut-para-fakar.html>
- Riyanto. 2010. *Sistem Informasi Penjualan Dengan PHP Dan MySQL*. Yogyakarta : Gava Media.
- Samiadi, Lika Aprilia. 2018. *Apa itu penyakit hati (penyakit liver)?*. Diakses 20, April 2018 dari <https://hellosehat.com/penyakit/penyakit-hati-penyakit-liver/>.
- Samiadi, Lika Aprilia. 2017. *Cara Kerja Tes Fungsi Hati untuk Mendeteksi Risiko Kerusakan Hati*. Diakses 20, April 2018 dari <https://hellosehat.com/pusat-kesehatan/hepatitis/pemeriksaan-tes-fungsi-hati/>.
- Satriyanto, Edi. 2009. *Clustering*. Diakses 20, April 2018 dari <http://kangedi.lecturer.pens.ac.id/materi%20kuliah/matakuliah%20statistika/>.
- Tim Alodokter. 2016. *Mengenal Bilirubin Dan Penyebab Jumlah Bilirubin Meningkat*. Diakses 20, April 2018 dari <https://www.alodokter.com/mengenal-bilirubin-dan-penyebab-jumlah-bilirubin-meningkat>.
- Tim Alodokter.com. 2016. *Fungsi Hati Begitu Penting, Maka Sayangi Dia*. Diakses 5, Juni 2018 dari <https://www.alodokter.com/fungsi-hati-begitu-penting-maka-sayangi-dia>,
- Turban, E., dkk. 2005. *Decision Support System and Intelligent System*. Yogyakarta: Andi Offset
- Zadeh. 1965. *Fuzzy Sets*. L. A : Information and Control