

PENGELOMPOKAN JENIS TUMOR MENGGUNAKAN METODE FUZZY C-MEANS

Lasthania¹, Suyanto^{2, 3}

¹Teknik Informatika, Fakultas Teknik Informatika, Universitas Telkom

Abstrak

Dalam teori fuzzy dikenal adanya istilah clustering atau pengelompokan. Clustering merupakan suatu proses pengelompokan data ke dalam kelas-kelas atau cluster-cluster sehingga data dalam suatu kelas tersebut memiliki tingkat persamaan yang tinggi antara satu dengan yang lainnya akan tetapi berbeda dengan data dalam kelas yang lainnya. Fuzzy clustering adalah salah satu teknik untuk menentukan cluster optimal dalam suatu ruang vektor yang didasarkan pada bentuk normal Euclidian untuk jarak antar vektor. Salah satu kasus yang berkaitan dengan pengelompokan data yaitu pengelompokan suatu tumor yang diderita pasien ke dalam dua kelas yaitu tumor jinak (benign) atau tumor ganas (malignant). Pertumbuhan tumor tersebut dapat mengarah ke pertumbuhan jinak (benign) atau ke pertumbuhan ganas (malignant). Tumor pada payudara yang mengalami pertumbuhan ganas itu lah yang merupakan kanker payudara. Algoritma fuzzy c-means merupakan salah satu algoritma fuzzy clustering yang paling banyak digunakan karena berusaha mengelompokkan setiap elemen-elemen yang merupakan anggota dari suatu cluster ke dalam suatu cluster. Algoritma fuzzy c-means dapat digunakan untuk mengelompokkan data yang sudah diketahui jumlah cluster-nya. Pertama-tama proses awal yang dilakukan adalah preprocessing data (dilakukan di luar sistem) dengan menggunakan tools weka yaitu proses attribute selection. Proses ini dilakukan untuk mendapatkan atribut-atribut pilihan yang akan digunakan sebagai input pada sistem. Selanjutnya data akan diproses menggunakan algoritma fuzzy c-means. Proses terakhir yang dilakukan yaitu melakukan penghitungan akurasi sistem. Hasil pengujian menunjukkan bahwa metode fuzzy c-means mampu melakukan identifikasi resiko terkena kanker payudara dengan nilai akurasi yang baik.

Kata Kunci : kanker payudara, fuzzy clustering, fuzzy c-means

Abstract

In fuzzy theory, there's something called clustering or grouping. Clustering is data processing to classes or clusters so that the data in a class has high similarity with the others which happen to be in the same class, but it is different with the data in the other class. Fuzzy clustering is a technique to determine the optimal cluster in a vector that is based on the normal Euclidian for the distance between vectors. One of the cases that related to the data grouping is a patient's tumor grouping into two classes: benign and malignant. The growth of the tumor can lead to a benign tumor or the malignant one. The tumor in breast tumor that is a malignant growth can be called breast cancer. Fuzzy C-Means algorithm is one of the fuzzy clustering algorithms that was used the most because it tries to cluster every elements that is the member of a cluster to a cluster. Fuzzy C-Means algorithm can be used to cluster data that was known the number of the cluster. First of all, the initial process is to do preprocessing data (outside the system) using weka tools, that is the attribute selection process. This process is done to get the selected attribute that will be used as inputs for the system. Then data will be processed with fuzzy c-means algorithm. The last proses to be done is calculating the system's accuracy. The testing result shows that the fuzzy c-means method is capable to identificate the breast cancer risk with the good accuracy.

Keywords : breast cancer, fuzzy clustering, fuzzy c-means.

1. Pendahuluan

1.1 Latar Belakang

Dalam teori *fuzzy* dikenal adanya istilah *clustering* atau pengelompokan. [9] *Clustering* merupakan suatu proses pengelompokan data ke dalam kelas-kelas atau *cluster-cluster* sehingga data dalam suatu kelas tersebut memiliki tingkat persamaan yang tinggi antara satu dengan yang lainnya akan tetapi berbeda dengan data dalam kelas yang lainnya. [7] *Fuzzy clustering* adalah salah satu teknik untuk menentukan *cluster* optimal dalam suatu ruang vektor yang didasarkan pada bentuk normal *Euclidian* untuk jarak antar vektor.

Salah satu kasus yang berkaitan dengan pengelompokan data yaitu pengelompokan suatu tumor yang diderita pasien ke dalam dua kelas yaitu tumor jinak (*benign*) atau tumor ganas (*malignant*). Identifikasi seseorang terkena kanker payudara bisa diketahui dari jenis tumor payudara yang dideritanya. [2] Tumor merupakan kumpulan dari sel abnormal yang terbentuk dari sel-sel yang tumbuh secara terus menerus secara tidak terbatas. Pertumbuhan tumor tersebut dapat mengarah ke pertumbuhan jinak (*benign*) atau ke pertumbuhan ganas (*malignant*). Tumor pada payudara yang mengalami pertumbuhan ganas itu lah yang merupakan kanker payudara. Pada penelitian ini digunakan beberapa baris data yang terdiri dari beberapa atribut yang salah satu atributnya merupakan suatu kelas yang terdiri dari dua jenis kelas yaitu *benign* dan *malignant*. Atribut-atribut tersebut terdiri dari masukan *real-valued features* yang dihitung dari setiap sel nukleus dari setiap pasien yang menderita tumor pada payudaranya. Berdasarkan masukan *real-valued features* tersebut dapat diprediksi apakah tumor yang diderita pasien tersebut tergolong jinak (*benign*) atau ganas (*malignant* atau kanker).

Terdapat beberapa algoritma *clustering* data salah satunya yaitu *fuzzy c-means* (FCM). Pendekatan *clustering* menghasilkan partisi dimana dalam partisi tersebut setiap anggota hanya dimiliki oleh satu *cluster*. *Fuzzy clustering* mempunyai sebuah skema *soft clustering* yang mana setiap pola berasosiasi dengan seluruh *cluster* dengan menggunakan fungsi keanggotaan. Nilai keanggotaan yang besar mempunyai kemungkinan yang besar pula suatu pola menjadi bagian dari suatu *cluster*. [6] Algoritma *fuzzy c-means* merupakan salah satu algoritma *fuzzy clustering* yang paling banyak digunakan karena berusaha mengelompokkan setiap elemen-elemen yang merupakan anggota dari suatu *cluster* ke dalam suatu *cluster*. Algoritma *fuzzy c-means* dapat digunakan untuk mengelompokkan data yang sudah diketahui jumlah *cluster*-nya. Algoritma *fuzzy c-means* merupakan algoritma yang

sederhana, mudah diimplementasikan, dan memiliki kemampuan untuk mengelompokkan data dalam jumlah besar.

1.2 Perumusan Masalah

Rumusan masalah yang terdapat pada penelitian ini adalah sebagai berikut:

1. Bagaimana mengimplementasikan metode *fuzzy c-means* untuk memprediksi resiko terkena kanker payudara berdasarkan pengelompokan jenis tumor yang dihasilkan?
2. Bagaimana proses yang dilakukan untuk mendapatkan keluaran berupa status resiko terkena kanker payudara dengan menggunakan metode *fuzzy c-means*?
3. Bagaimana menganalisis tingkat akurasi yang dihasilkan oleh metode *fuzzy c-means* dalam memprediksi seseorang terkena kanker payudara atau tidak berdasarkan beberapa *real-valued features* yang diberikan?

Adapun batasan masalah dalam tugas akhir ini yaitu :

1. Menggunakan data *breast cancer wisconsin (diagnostic) data set* yang diperoleh dari *UCI repository*.
2. *Preprocessing* data dilakukan di luar sistem program.

1.3 Tujuan

Berdasarkan rumusan masalah di atas, maka tujuan yang ingin dicapai dari tugas akhir ini adalah:

1. Mengimplementasikan metode *fuzzy c-means* untuk memprediksi resiko terkena kanker payudara berdasarkan karakteristik tumor yang dihasilkan.
2. Melakukan proses untuk mendapatkan keluaran berupa status resiko terkena kanker payudara dengan menggunakan metode *fuzzy c-means*.
3. Menganalisis tingkat akurasi yang dihasilkan oleh metode *fuzzy c-means* dalam memprediksi seseorang terkena kanker payudara atau tidak berdasarkan beberapa *real-valued features* yang diberikan.

1.4 Metodologi Penyelesaian Masalah

Metode yang digunakan dalam penyelesaian tugas akhir ini adalah :

1. Melakukan studi Literatur
 - a. mencari referensi dan sumber-sumber lain yang berhubungan dengan tugas akhir ini seperti *fuzzy logic*, *fuzzy clustering*, dan *fuzzy c-means*.
 - b. mempelajari dan memahami materi yang berhubungan dengan tugas akhir ini.
2. Mempelajari algoritma *fuzzy c-means* yang akan digunakan dalam penelitian ini.

3. Memahami dan mempelajari data *breast cancer wisconsin (diagnostic) data set* yang diperoleh dari *UCI repository*.
4. Melakukan perancangan dan pembuatan sistem menggunakan metode *fuzzy c-means* untuk memperoleh keluaran berupa prediksi seseorang terkena kanker payudara atau tidak.
5. Melakukan pengujian akurasi terhadap metoda *fuzzy c-means* sesuai dengan kasus yang sedang dikerjakan dalam tugas akhir ini.
6. Pengambilan kesimpulan dan penyusunan laporan tugas akhir.



5. Kesimpulan dan Saran

5.1 Kesimpulan

1. Implementasi sistem dalam melakukan pendeteksian resiko terkena kanker payudara berdasarkan karakteristik tumor yang dihasilkan dengan menggunakan metode *fuzzy c-means* mampu melakukan pendeteksian dengan akurasi sistem sebesar 91,6084% dengan cara mengeluarkan keluaran berupa status resiko terkena kanker payudara atau tidak.
2. Diperlukan penentuan nilai parameter yang tepat agar dihasilkan nilai titik-titik pusat *cluster* yang tepat yang dapat berpengaruh pada hasil akurasi sistem. Pada tugas akhir ini nilai parameter yang digunakan adalah nilai maksimum iterasi (*maxIter*) sebesar 100, nilai error (*e*) sebesar 10^{-5} , dan nilai pangkat (*w*) sebesar 2.

5.2 Saran

1. Melakukan proses *clustering* data dengan menggunakan algoritma *clustering* selain *fuzzy c-means*.
2. Membandingkan hasil *clustering* data menggunakan algoritma *fuzzy c-means* dengan hasil *clustering* data menggunakan algoritma *clustering* yang lain.

Daftar Pustaka

- [1] A.Fernandez, S. Garcia, M.J. del Jesus, and F.Herrera. "A study of the behaviour of linguistic fuzzy rule based classification system in the framework of imbalanced data-sets". *Fuzzy Sets and systems*, 159 (18):2378-2398,2008.
- [2] Chrestella, Jessy. 2009. *Neoplasma*. repository.usu.ac.id/bitstream/123456789/2053/1/10E00541.pdf.
- [3] Guyon, Isabelle dan Andre Elisseeff. *An Introduction to Variable and Feature Selection*. Max Planck Institute. Germany. 2003.
- [4] <http://archive.ics.uci.edu/ml/>, diakses tanggal 10 Desember 2011.
- [5] http://repository.upi.edu/operator/upload/ta_mtk_0607196_chapter3.pdf, diakses tanggal 14 Desember 2011.
- [6] J.C, Obi dan A.A imianvan. 2011. *Fuzzy Cluster Means Expert System for the Diagnosis of Tuberculosis*. *Global Journal of Computer Science & Technology* Volume 11 Issue Version 1.0 April 2011.
- [7] Kusumadewi, Sri dan Purnomo, Hari. 2010. *Aplikasi Logika Fuzzy untuk Pendukung Keputusan*. Yogyakarta: Graha Ilmu.
- [8] Kusumadewi, Sri. 2002. *Tool Box Matlab*. Yogyakarta: Graha Ilmu.
- [9] Luthfi, Emha Taufik. 2007. *Fuzzy C-Means untuk Clustering Data (Studi Kasus: Data Performance Mengajar Dosen)*. Seminar Nasional Teknologi 2007 (SNT 2007).
- [10] Narayana, M. Kulkarni, S. 2012. *Comparison between Euclidean Distance Metric and SVM for CBIR using Level Set Features*. *International Journal of Engineering Science and Technology (IJEST)* Vol. 4 No.01 January 2012.
- [11] Szil'agyi, M. Szil'agyi, Beny'o.2008. *Analytical and Numerical Evaluation of the Suppressed Fuzzy C-Means Algorithm*. MDAI 2008, LNAI 5285, pp. 146–157, 2008._c Springer-Verlag Berlin Heidelberg.
- [12] Zou, K. Wang, Z. Ming Hu. 2008. *An New Initialization Method for Fuzzy C-Means Algorithm*. Springer Science+Business Media, LLC 2008.

Telkom
University