

PENERAPAN METODE FUZZY C-MEANS PADA PENGELOMPOKAN PASIEN KANKER PAYUDARA PASCA OPERASI MENGGUNAKAN HABERMAN'S SURVIVAL DATASET

Hetty Rohayani. AH¹⁾, Afrizal. J²⁾

Jurusan Teknologi Informasi Fakultas Teknik dan Ilmu Komputer Universitas Adiwangsa Jambi¹⁾

hettyrohayani@gmail.com

urusan Sistem Komputer Sekolah Tinggi Manajemen Informatika dan Komputer Nurdin Hamzah Jambi²⁾

Afrizal_79@yahoo.com

Abstrak

Kanker payudara adalah penyakit yang merupakan urutan kedua sebagai penyebab kematian di dunia. Dimana penanganan kanker payudara diantaranya dilakukan dengan operasi, tetapi penanganan kanker payudara dengan jalan operasi tidaklah merupakan suatu jalan mengatasi masalah, karena kemudian masalah yang timbul adalah keberuntungan hidup pasien pasca dilakukannya operasi. Dalam penelitian ini dengan menggunakan metode fuzzy clustering means mencoba mengelompokkan pasien kanker payudara pasca dilakukan operasi, dimana pengelompokan dijadikan dua kelompok yang pertama adalah keberuntungan hidup pasien kanker payudara pasca operasi lebih dari lima tahun dan keberuntungan hidup pasien pasca operasi kanker payudara yang kurang dari lima tahun. Dihasilkan pengelompokan berdasarkan atribut yang meliputi usia saat pasien terkena kanker, tahun dilakukannya operasi dan identitas secara medis. Yang mana berhasil mengelompokkan data yang berasal dari data set haberman survival menjadi dua kelompok.

Key word : FCM, Kanker payudara, Data set, Haberman Survival.

Pendahuluan

Kanker payudara adalah sekelompok penyakit sebagai akibat dari pertumbuhan sel-sel jaringan tubuh pada payudara dan tumbuh diluar kendali, yang bila tidak cepat ditangani dan diobati akan menyebabkan kematian (*American Cancer Society*, 2013). Kanker merupakan penyebab kematian utama kedua yang memberikan kontribusi 13% kematian dari 22% kematian akibat penyakit tidak menular utama di dunia (Oemiati et al., 2011). Selain itu kecenderungan peningkatan prevalensinya tidak dapat dihindari. Ditambah lagi kematian karena kanker payudara masih tinggi, terutama pada negara-negara sedang berkembang, karena keterlambatan diagnosis, yang berarti juga keterlambatan pengobatan (Bustan, 2007).

Kanker payudara menyerang lebih banyak wanita Eropa dan Amerika dengan prevalensi tiga kali lipat lebih tinggi dibandingkan dengan wanita di Asia Timur. Akan tetapi, peningkatan jumlah kasus kanker payudara juga terjadi di sejumlah negara maju di Asia seperti Jepang, Singapura dan Hongkong. Resiko sepanjang hidup untuk mengalami kanker payudara di Hongkong

diperkirakan mencapai 5,3% (*Leung et al.*, 2008).

Di Indonesia ada lima besar provinsi mempunyai prevalensi tumor di atas angka nasional (>5,03%), yang pertama DIY menduduki urutan prevalensi tertinggi di Indonesia yaitu sebesar 9,66%, disusul Jawa Tengah 8,06%, DKI Jakarta 7,44%, Banten 6,35%, selanjutnya Sulawesi Utara 5,76%, kanker payudara mempunyai prevalensi kedua tertinggi setelah kanker ovarium dan cervix uteri di Indonesia pada tahun 2011 (Oemiati et al., 2011).

Penyembuhan kanker payudara bisa dilakukan dengan jalan operasi, tapi pada kenyataan setelah menjalani operasi kanker payudara ada pasien yang bertahan hidup tapi ada pula yang tidak. Dalam penelitian ini dengan menggunakan data Haberman's Survival yang diambil dari UCI Machine Learning Repository (

<https://archive.ics.uci.edu/ml/datasets/Haberman%27s+Survival>) akan dilakukan pengelompokan keberuntungan hidup pasien pasca dilakukannya operasi kanker payudara. Dalam penelitian akan diterapkan pengelompokan menggunakan Fuzzy C-Means dalam Haberman's Survival dataset. Dengan

PENERAPAN METODE FUZZY C-MEANS PADA PENGELOMPOKAN PASIEN KANKER PAYUDARA PASCA OPERASI MENGGUNAKAN HABERMAN'S SURVIVAL DATASET

penelitian ini diharapkan dapat dilihat pengelompokan pasien kanker payudara yang dapat bertahan hidup dalam waktu lima tahun atau lebih pasca operasi dan pasien yang meninggal dalam waktu lima tahun pasca dilakukannya operasi.

1. Landasan Teori

Fuzzy C-Means (FCM) adalah suatu teknik pengelompokan data yang mana keberadaan tiap-tiap titik data dalam suatu kelompok (*cluster*) ditentukan oleh derajat keanggotaan. Kelebihan dari Fuzzy C-Means adalah dapat melakukan clustering lebih dari satu variabel secara sekaligus. Teknik ini pertama kali diperkenalkan oleh Jim Bezdek pada tahun 1981. (Kusumadewi, 2010).

Konsep dasar FCM, pertama kali adalah menentukan pusat cluster, yang akan menandai lokasi rata-rata untuk tiap-tiap kelompok (Cluster). Pada kondisi awal, pusat cluster ini masih belum akurat. Tiap-tiap titik data memiliki derajat keanggotaan untuk tiap-tiap kelompok (cluster). Dengan memperbaiki pusat cluster dan derajat keanggotaan tiap-tiap data secara berulang, maka akan dapat dilihat bahwa pusat cluster akan bergerak menuju lokasi yang tepat. Perulangan ini didasarkan pada minimalisasi fungsi obyektif yang menggambarkan jarak dari titik data yang diberikan ke pusat cluster yang terbobot oleh derajat keanggotaan titik data tersebut.

FCM didasarkan pada minimisasi fungsi obyektif berikut dirumuskan sebagai berikut:

$$J_m = \sum_{i=1}^D \sum_{j=1}^N \mu_{ij}^m \|x_i - c_j\|^2,$$

Data Set

Data set yang digunakan dalam penelitian ini diambil dari UCI Machine Learning Repository yaitu Haberman's Survival Data Set (<https://archive.ics.uci.edu/ml/datasets/Haberman%27s+Survival>) dimana data ini berisi kasus dari penelitian yang dilakukan antara tahun 1958 dan 1970 di University of Chicago's Billings Hospital mengenai kelangsungan hidup pasien yang telah menjalani operasi kanker payudara. Informasi atribut dari data yang digunakan adalah usia pasien pada saat

dilakukan operasi, tahun melaksanakan operasi (tahun 1900) serta jumlah nodus aksila positif yang terdeteksi, yang mana dalam pengelompokan ini diharapkan akan dapat terbentuk cluster pasien yang bertahan hidup 5 tahun atau lebih dan pasien yang meninggal pasca operasi dalam waktu 5 tahun.

Haberman's Survival Data Set

Download: [Data Folder](#), [Data Set Description](#)

Abstract: Dataset contains cases from study conducted on the survival of patients who had undergone surgery for breast cancer

Data Set Characteristics:	Multivariate	Number of Instances:	306	Area:	Life
Attribute Characteristics:	Integer	Number of Attributes:	3	Date Donated:	1999-03-04
Associated Tasks:	Classification	Missing Values?	No	Number of Web Hits:	118521

Gambar 1. Tampilan UCI Machine learning Repository Haberman's Survival Data Set

(Sumber :

<https://archive.ics.uci.edu/ml/datasets/Haberman%27s+Survival>)

Informasi Atribut dari Dataset yang digunakan:

1. Usia pasien pada saat operasi berupa data numerik
2. Tahun operasi pasien diambil pada tahun 1900 dengan data berupa data numerik
3. Jumlah nodus aksila positif yang terdeteksi berupa data numerik

Atribut pengelompokan yang diharapkan dalam analisa ini adalah :

-1 untuk data pasien yang bertahan lima tahun atau lebih pasca operasi.

-2 untuk data pasien yang meninggal dalam waktu lima tahun.

3. Metodologi Algoritma Fuzzy C-Means

Algoritma Fuzzy C-Means (FCM) adalah sebagai berikut (Zimmerman, 1996);(Yan,1994)

1. Input data yang akan dicluster X, berupa matriks n x m (n=jumlah sampel data, m =atribut setiap data), X_{ij} =data sampel ke-l ($i=1,2,\dots,m$).

PENERAPAN METODE FUZZY C-MEANS PADA PENGELOMPOKAN PASIEN KANKER PAYUDARA PASCA OPERASI MENGGUNAKAN HABERMAN'S SURVIVAL DATASET

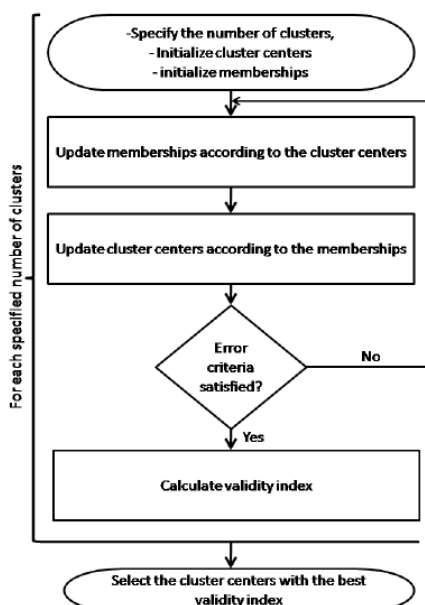
2. Tentukan
 - a. jumlah cluster
= c
 - b. pangkat(bobot)
= w
 - c. maksimum iterasi
= maxIter
 - d. error terkecil yang dihadapi = ξ
 - e. fungsi obyektif awal
= $P_0=0$
 - f. iterasi awal
= $t=1$
3. Mengkalkulasikan pusat cluster dengan rumus sebagai berikut :

$$c_j = \frac{\sum_{i=1}^D \mu_{ij}^m x_i}{\sum_{i=1}^D \mu_{ij}^m}$$

4. Pengupdatean nilai μ_{ij} dilakukan dengan formula sebagai berikut:

$$\mu_{ij} = \frac{1}{\sum_{k=1}^N \left(\frac{\|x_i - c_j\|}{\|x_i - c_k\|} \right)^{\frac{2}{m-1}}}$$

5. Bangkitkan bilangan random μ_{ij} , $i=1,2,\dots,n; k=1,2,\dots,c$ sebagai elemen-elemen matriks, partisi awal U.



Gambar 2. Flowchart for the fuzzy c-means (FCM) algorithm (https://www.researchgate.net/publication/224039986_Dynamic_Fuzzy_C-Means_dFCM_Clustering_and_its_Application_to_Calorimetric_Data_Reconstruction_in_High_Energy_Physics/figures?lo=1)

4. Analisa Fuzzy C-Means dengan dataset Haberman's Survival menggunakan Matlab

Menampilkan data set
Berikut ini adalah listing untung memanggil dataset

```
% memanggil haberman dataset
```

```
load haberman.dat
```

Pengaturan Parameter

Dalam dataset haberman's survival ini diharapkan akan terbentuk 2 cluster yang mana cluster pertama adalah kebertahanan hidup pasien setelah operasi kanker payudara dalam waktu lima tahun atau lebih, dan cluster kedua adalah kebertahanan hidup pasien pasca operasi kanker payudara kurang dari lima tahun.

```
[centers,U] = fcm(haberman,2);
```

Dimana jumlah maksimum iterasi dalam proses penggunaan haberman's dataset ditampilkan sebagai berikut :

PENERAPAN METODE FUZZY C-MEANS PADA PENGELOMPOKAN PASIEN KANKER PAYUDARA PASCA OPERASI MENGGUNAKAN HABERMAN'S SURVIVAL DATASET

```

Iteration count = 1, obj. fcn = 33602.617112
Iteration count = 2, obj. fcn = 27232.833720
Iteration count = 3, obj. fcn = 26863.397553
Iteration count = 4, obj. fcn = 25441.411251
Iteration count = 5, obj. fcn = 22985.708904
Iteration count = 6, obj. fcn = 21810.331261
Iteration count = 7, obj. fcn = 21640.389038
Iteration count = 8, obj. fcn = 21626.587942
Iteration count = 9, obj. fcn = 21625.366524
Iteration count = 10, obj. fcn = 21625.168793
Iteration count = 11, obj. fcn = 21625.116373
Iteration count = 12, obj. fcn = 21625.100183
Iteration count = 13, obj. fcn = 21625.095028
Iteration count = 14, obj. fcn = 21625.093378
Iteration count = 15, obj. fcn = 21625.092849
Iteration count = 16, obj. fcn = 21625.092680
Iteration count = 17, obj. fcn = 21625.092625
Iteration count = 18, obj. fcn = 21625.092608
Iteration count = 19, obj. fcn = 21625.092602
    
```

Klasifikasikan setiap titik data ke dalam cluster dengan nilai keanggotaan terbesar, diselesaikan dengan intruksi :

```

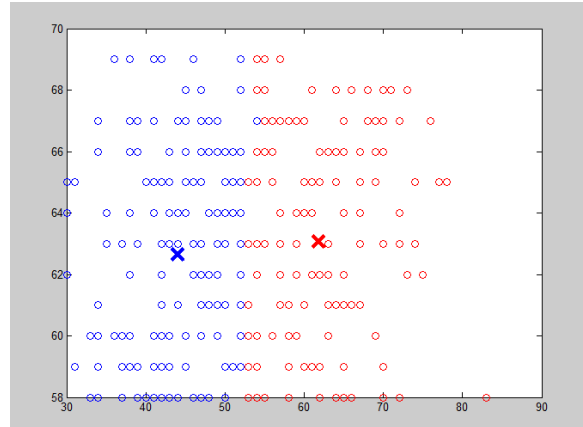
maxU = max(U);
index1 = find(U(1,:) == maxU);
index2 = find(U(2,:) == maxU);
    
```

Untuk memplot data cluster diselesaikan dengan intruksi:

```

plot(haberman(index1,1),haberman(index1,2), 'ob')
hold on
plot(haberman(index2,1),haberman(index2,2), 'or')
plot(centers(1,1),centers(1,2),'xb','MarkerSize',15,'LineWidth',3)
plot(centers(2,1),centers(2,2),'xr','MarkerSize',15,'LineWidth',3)
hold off
    
```

Tampilan data hasil pengelompokan menggunakan Fuzzy C-Means dengan aplikasi matlab ditunjukkan pada gambar berikut:



Gambar 3. Hasil Pengelompokan Dataset Haberman's Survival dengan FCM menggunakan Matlab

Pada tampilan gambar 3. Terlihat hasil plot dataset haberman's survival, dimana angka-angka yang tertera merupakan hasil dari pengklusteran.

5. Kesimpulan

- a. Dalam penelitian ini dihasilkan pengelompokan dataset haberman's survival dengan menggunakan Fuzzy C-Means.
- b. Dari iterasi terakhir yang dilakukan dalam setiap analisa ini terlihat hasil pengklusteran yang terbentuk menjadi 2 cluster.

Daftar Pustaka

1. Zimmerman, H.-J, Fuzzy set theory- and its application, 3rd ed, Kluwer, Boston, 1996.
2. SOCIETY, American Cancer. Cancer facts and figures 2013. 2013.
3. OEMIATI, Ratih; RAHAJENG, Ekowati; KRISTANTO, Antonius Yudi. Prevalensi tumor dan beberapa faktor yang mempengaruhinya di Indonesia. *Buletin Penelitian Kesehatan*, 2011, 39.4 Des: 190-204.
4. YAN, Jun. Using fuzzy logic: towards intelligent systems. 1994.
5. KUSUMADEWI, Sri; HARI, P. Logika Fuzzy. 2010.
6. <https://archive.ics.uci.edu/ml/datasets/Haberman%27s+Survival>
7. https://www.researchgate.net/publication/224039986_Dynamic_Fuzzy_c-means_dFCM_Clustering_and_its_Appli

[cation to Calorimetric Data Reconstruction in High Energy Physics/figures?l
o=1](#)

8. [Rohayani, Hetty; Sistem Pakar Pendeteksi Penyakit Kanker Ganas Yang Menyerang Wanita; Jurnal Media Processor, 2017.](#)