
Segmentasi Citra CT Kanker Payudara Menggunakan Metode *K-Means*

Dian Rani Angraini¹, I Wayan Sudiarta¹ ✉, Rahadi Wirawan¹,

¹Program Studi Fisika, Fakultas Matematika dan Ilmu Pengetahuan Alam, Universitas Mataram, Jl. Majapahit No. 62, Mataram, Lombok-Nusa Tenggara Barat, Indonesia, 83125

Info Artikel

Sejarah Artikel:

Diterima:
DD-MM-YYYY

Disetujui:
DD-MM-YYYY

Dipublikasikan:
DD-MM-YYYY

Keywords:
Tumor, PSNR
gaussian filter,
median filter,
cluster optimal

ABSTRAK

Kanker payudara adalah salah satu jenis kanker penyebab utama kematian pada wanita. Upaya awal deteksi dini kanker dapat dilakukan melalui pemeriksaan radiologis *Computed Tomography Scanner* (CT-Scan). Namun citra CT masih memiliki *noise* pada citra CT yang berpotensi dapat menimbulkan kesalahan interpretasi diagnosa dokter. Penelitian ini bertujuan untuk mengimplementasikan metode segmentasi *clustering K-means* untuk mendeteksi dan memperoleh daerah terdeteksi objek tumor pada CT kanker payudara. Metode pengurangan *noise* citra yang digunakan yaitu *gaussian filter* dan *median filter*. Hasil filtering citra berdasarkan nilai PSNR menunjukkan bahwa pada citra CT *median filter* memiliki nilai PSNR yang lebih tinggi dibandingkan citra CT *gaussian filter*. Hasil segmentasi menunjukkan jumlah *cluster* optimal yang sama yaitu pada jumlah *cluster* $K=7$ untuk citra *gaussian filter* dan *median filter* yang tersegmentasi *K-means clustering*.

ABSTRACT

Breast cancer is one of the leading causes of death in women. Early detection of cancer can be done through Computed Tomography Scanner (CT-Scan) radiological examination. However, CT images still have noise in the CT image, potentially leading to misinterpretation of the doctor's diagnosis. This study aims to implement the K-Means Clustering segmentation method to detect and obtain detected areas of tumor objects on CT breast cancer. The image noise reduction methods used are gaussian filter and median filter. The segmentation results show the same optimal number of clusters at the number of $K = 7$ for CT gaussian filter image and CT median filter.

✉ Alamat korespondensi:
Jl. Sawo Manila, Pejaten, Ps. Minggu, Jakarta
E-mail: nilarasati@civitas.unas.ac.id

© 2023 Universitas Mataram

p-ISSN xxxx-xxxx
e-ISSN xxxx-xxxx

PENDAHULUAN

Kanker payudara adalah satu jenis kanker yang paling umum menjadi penyebab utama kematian Wanita. Data Global Cancer Observatory (GCO) tahun 2021 mengungkapkan kanker payudara telah menjadi kanker dengan insidensi terbanyak dan melebihi insidensi kanker paru. Oleh karena itu perlu adanya Tindakan deteksi secara dini dengan melakukan pemeriksaan radiologis guna memastikan keberadaan adanya kanker [1]. Pemeriksaan radiologis yang umum dilakukan adalah pemeriksaan *Computed Tomography Scanner* (CT-Scan) [2]. CT Scan merupakan suatu modalitas imaging diagnostic yang menggunakan gabungan sinar x dan komputer untuk mendapatkan citra atau gambar berupa variasi irisan tubuh manusia [3]. Namun citra CT yang dihasilkan masih memiliki banyak *noise* dan belum terekonstruksi dengan baik. Sehingga hal tersebut dapat berpotensi menimbulkan kesalahan interpretasi diagnosa dokter. Salah satu cara untuk meningkatkan kualitas citra adalah dengan menggunakan metode reduksi *noise filtering*. Setelah dilakukan proses penapisan derau, citra akan dilakukan proses segmentasi guna memberikan kepastian keberadaan objek kanker [4].

Segmentasi merupakan proses pengolahan citra yang bertujuan untuk memisahkan wilayah citra objek dan wilayah *background* pada citra dengan akurat [5]. Salah satu metode segmentasi yang sering digunakan adalah *clustering*. metode segmentasi *clustering* bekerja dengan data multidimensi untuk mengelompokkan piksel citra ke dalam kluster/kelompok berdasarkan kedekatan jarak antar piksel [6]. Metode *K-means clustering* menjadi metode yang paling sering digunakan karena mampu mengkluster data variabel yang besar dengan sangat cepat [7]. Penelitian ini bertujuan untuk mendapatkan *cluster* yang optimal dalam menampilkan daerah terdeteksi objek tumor menggunakan *K-means clustering* serta menunjukkan perbandingan kualitas dari penggunaan metode *noise filtering* dalam mengurangi derau pada citra secara kuantitatif menggunakan PSNR (*Peak Signal to Noise Ratio*).

METODE

Penelitian ini menggunakan data sampel citra sebanyak 20 orang pasien kanker payudara CT Simulator axial thoraks. Citra dalam bentuk file *.dcm atau DICOM (*Digital Imaging and Communication in Medicine*) dengan resolusi 512×512 piksel dengan *Bit depth* 16 Bit. Adapun alat-alat yang digunakan laptop, bahasa pemrograman *Python* ver 3.8 dan *Jupyter Notebook* sebagai aplikasi editing program. *Preprocessing* sampel citra diawali dengan *rescale windowing* nilai CT HU. Dilakukan *windowing window width* (WW): 600 HU dan *window Level* (WL) 40 HU citra CT asli dengan jenis *window* yang lebar (*wide*) menjadi WW: 350 HU; WL: 50 HU dengan jenis *window* sempit (*narrow*) untuk memfokuskan pada jaringan lunak dan tumor. Kemudian citra hasil *windowing* dilakukan penapisan derau dengan *gaussian filter* dan *median filter*. Adapun formula yang digunakan dari metode *gaussian filter* pada citra dua dimensi pada Persamaan (1) dan Persamaan (2) untuk formula metode *median filter*. Digunakan ukuran dimensi atau matriks kernel 3×3 untuk metode *gaussian filter* dan *median filter*. Selanjutnya seluruh citra hasil *filtering* dilakukan uji parameter nilai *Peak Signal to Noise Ratio* (PSNR) untuk mengetahui kualitas dari citra hasil *filtering* menggunakan Persamaan (3). Semakin besar parameter PSNR semakin mirip dengan citra asli. Kriteria kualitas citra dapat dilihat dari nilai PSNR, seperti pada Tabel 1 [8].

$$f(x, y) = e^{-\frac{x^2+y^2}{2\sigma^2}} \quad (1)$$

Dimana $f(x, y)$ adalah *gaussian filtering* pada dimensi dua, e adalah epsilon (2,71828), x jarak dari titik asal pada sumbu horizontal, dan y jarak dari titik asal pada sumbu vertikal.

Table 1: Nilai PSNR Kriteria kualitas citra

Nilai PSNR	Kualitas Citra
60 dB	<i>Excellent</i> , tanpa derau
50 dB	<i>Good</i> , terdapat sejumlah derau tapi kualitas citra masih bagus
40 dB	<i>Reasonable</i> , terdapat butiran halus atau seperti salju di dalam citra
30 dB	<i>Poor picture</i> , terdapat banyak derau
20 dB	<i>Unusable</i>

$$x = \frac{n+1}{2} \tag{2}$$

Dimana x adalah nilai baru median dan n adalah jumlah data

$$PSNR = 20 \cdot \text{Log}_{10} \left(\frac{255}{\sqrt{MSE}} \right) \tag{3}$$

$$MSE = \frac{1}{mn} \sum_{i=0}^{m-1} \sum_{j=0}^{n-1} |fa_{(i,j)} - fb_{(i,j)}|^2 \tag{4}$$

Dengan MSE sebagai nilai *Mean Square Error* dari citra; m sebagai panjang citra (dalam piksel); n sebagai lebar citra (dalam piksel), (i,j) merupakan koordinat masing-masing piksle; I sebagai nilai intensitas citra asli; K adalah nilai intensitas citra *filtering*.

Citra hasil dari penepisan derau *gaussian filter* dan *median filter* akan dilakukan proses segmentasi *K-means clustering*. Proses tahapan segmentasi dengan metode *K-means clustering* disusun dalam algoritma *Python*. Banyak kelompok yang digunakan sebanyak tujuh kelompok/*cluster* dengan nilai $k = 4, 5, 6, \dots, 10$. Pusat *cluster* akan dipilih yang mewakili “C”, kemudian dihitung jarak *Euclidean*. Setiap piksel diambil dan ditetapkan kedalam *cluster* yang sesuai jika jarak *Euclidean* minimum antar *cluster* dan piksel. Setelah itu dihitung ulang pusaat *cluster* baru menggunakan Persamaan (6). Setelah kedua data citra *filtering* disegmentasi, dilakukan penentuan nilai k terbaik atau *cluster* yang optimal menggunakan metode *Elbow*. Penerapan metode ini memiliki fokus pada perbandingan nilai dari perhitungan SSE (*Sum of Square distances Euclidean*) disetiap *cluster*-nya dengan menggunakan Persamaan (7). *Cluster* didefinisikan dengan $K_{init}=4$ dan $K_{end}=10$ yang menunjukkan jumlah *cluster* sebanyak 7 kelompok, kemudian dihitung nilai jarak kedua objek terdekat dengan menggunakan Persamaan (8). Setelah diperoleh hasil nilai optimal *cluster* dari kedua data citra segmentasi, dilakukan ekstraksi objek tumor payudara dari setiap citra dengan nilai *cluster* yang optimal.

$$J(v) = \sum_{i=1}^C \sum_{j=1}^{C_i} (\|x_i - v_j\|)^2 \tag{5}$$

$$V_i = \left(\frac{1}{C_i} \right) \sum_{j=1}^{C_i} x_i \tag{6}$$

Dengan x_i adalah piksel, v_j adalah pusat *cluster*, $\|x_i - v_j\|$ adalah jarak *Euclidean* antara x_i dan v_j , C_i adlaah banyaknya titik data untuk *cluster* ke I dan C adalah jumlah pusat *cluster*.

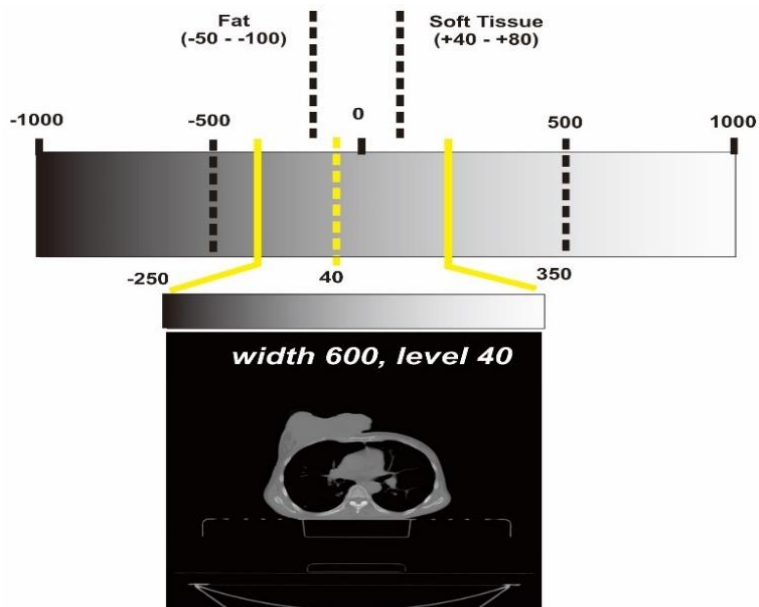
$$SSE = \sum_{j=1}^k \left(\sum_{p \in C_i} \|p_{ij} - v_i\|^2 \right) \tag{7}$$

$$d_{(x,y)} = \|x - y\|^2 = \sqrt{\sum_{i=1}^n (x_i - y_i)^2} \tag{8}$$

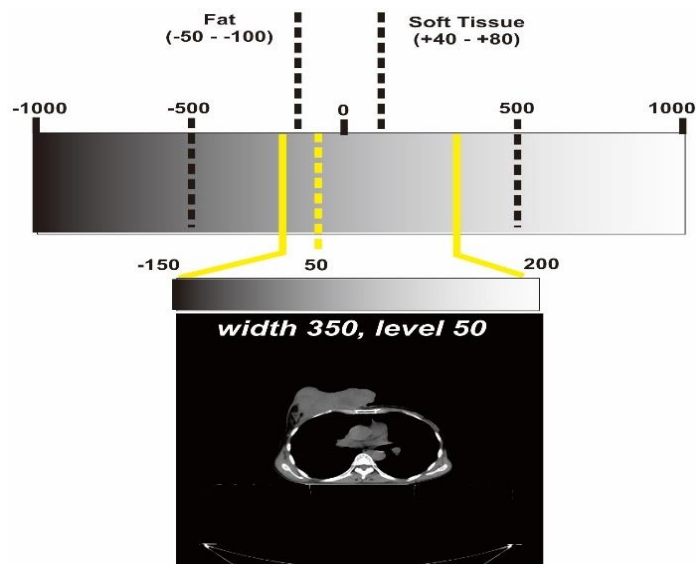
Dimana p_{ij} sebagai titik sampel j dalam *cluster* i , v_i sebagai pusaat *cluster* i , dan k sebagai jumlah pusat *cluster*.

HASIL DAN PEMBAHASAN

Hasil pengolahan citra CT ditunjukkan pada Gambar 1. untuk citra CT tanpa proses *windowing* dan Gambar 2. citra CT dengan proses *windowing*. Terlihat jenis *window width* (WW) yang digunakan pada citra CT asli yaitu *wide window* atau jendela yang melebar dengan rentang 400-2000 HU. Akibatnya sebaran intensitas warna gelap ke warna yang lebih terang terdistribusi ke seluruh skala intensitas. Sehingga daerah paru-paru atau jaringan kortikal pada citra CT terlihat dengan jelas, dimana terdapat adanya udara dan pembuluh darah yang berdampingan. Sedangkan pada Gambar 2. merupakan citra CT setelah dilakukan *rescale* menggunakan jenis *window width* (WW) yang sempit atau *narrow window* yang terdefinisikan 50-350 HU. Terlihat citra hasil *rescale windowing* menampilkan jaringan lunak dengan sangat baik yang sebelumnya terlihat kabur dengan kontras warna yang tinggi.

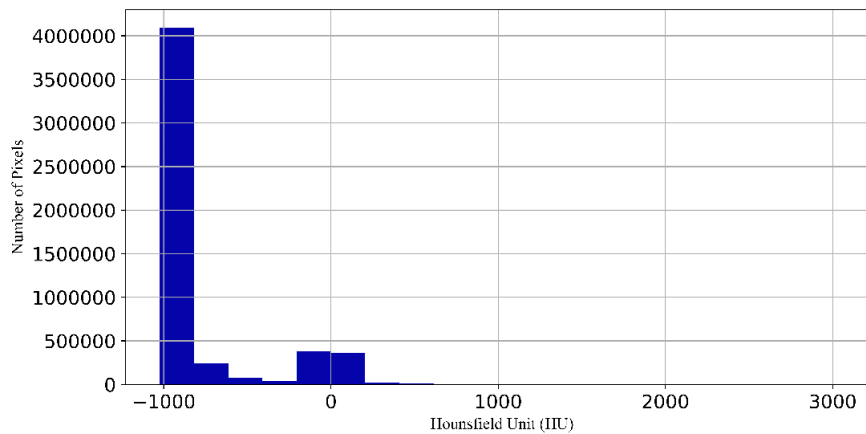


Gambar 1. Sebaran Warna Citra CT Asli berdasarkan Nilai CT HU

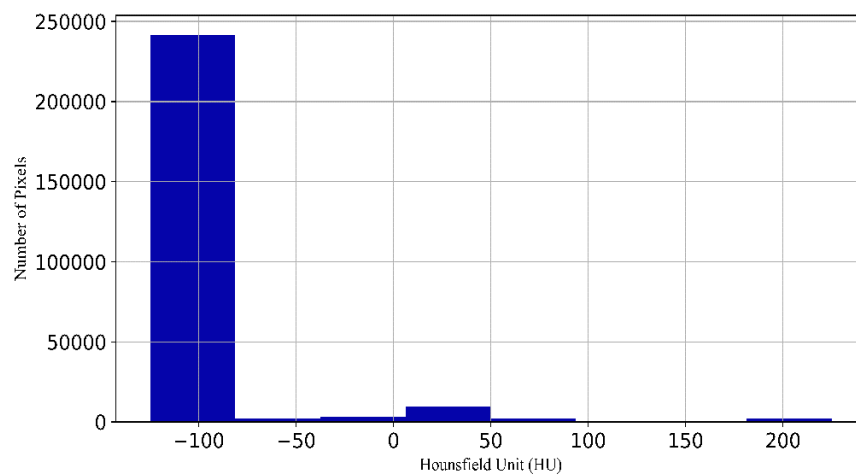


Gambar 2. Sebaran Warna Citra CT Rescale Windowing berdasarkan Nilai CT HU

Setelah dilakukan proses *rescale windowing* ditampilkan perbandingan histogram sebaran nilai piksel HU citra CT asli pada Gambar 3. dan Gambar 4. untuk citra CT *rescale windowing*. Pada Gambar 3. terlihat histogram batang histogram tertinggi dengan nilai CT sebesar -1000 HU, dimana nilai tersebut merepresentasikan elemen udara pada daerah paru-paru. Maka dari itu citra CT asli tidak dapat menampilkan jaringan lunak lain dengan warna kontras keabuan yang tinggi dengan jenis jendela lebar yang hanya difokuskan pada daerah paru-paru saja. Setelah dilakukan *rescale windowing* histogram sebaran nilai piksel HU citra CT pada Gambar 4. memperlihatkan Batang histogram tertinggi mewakili sebaran dari nilai *window width* (WW) dengan -150 HU yang didefinisikan sebagai nilai CT dari lemak (*fat*). Lalu batang histogram yang terletak pada posisi tengah merupakan *window center* dengan nilai 50 HU. Batang histogram terakhir yang memiliki nilai 200 HU merupakan bagian dari dari nilai *window width* (WW) yang mendefinisikan warna terang untuk jaringan lunak (*soft tissue*). Sehingga dengan dilakukannya proses *rescale* jaringan-jaringan lunak serta tumor pada daerah *thorax* dapat terlihat lebih jelas.

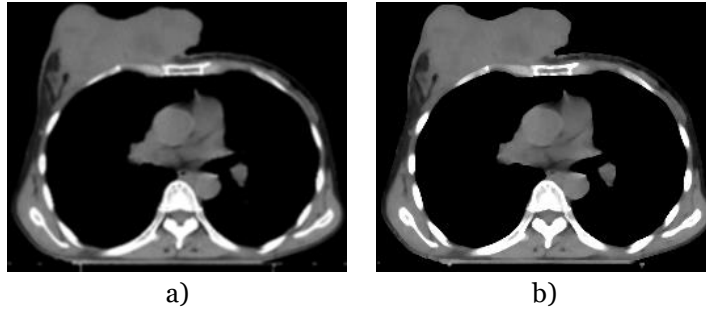


Gambar 3. Histogram Sebaran Nilai Pixel HU Citra CT Asli



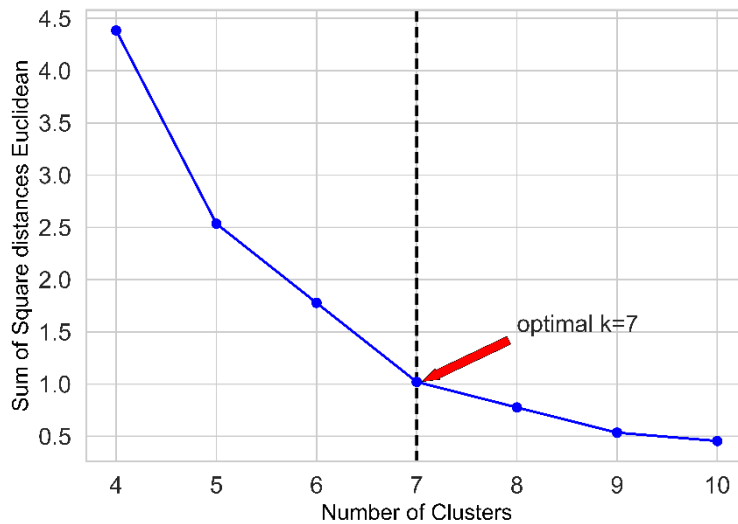
Gambar 4. Histogram Sebaran Nilai Pixel HU Citra CT Rescale Windowing

Hasil *filtering* citra CT menggunakan metode *gaussian filter* dan *median filter* ditunjukkan pada Gambar 5. Berdasarkan hasil perhitungan uji parameter nilai PSNR terhadap citra hasil *filtering* menunjukkan bahwa citra CT dengan metode *median filter* memperoleh nilai PSNR >40 dB dengan kriteria *Reasonable* yang artinya tidak adanya derau atau butiran halus seperti salju. Sedangkan hasil uji nilai PSNR pada citra CT dengan metode *gaussian filter* memperlihatkan berada pada kriteria *Poor* dengan nilai 30 dB yang berarti masih terdapat banyak derau pada citra CT hasil *filtering* menggunakan metode *gaussian filter*.

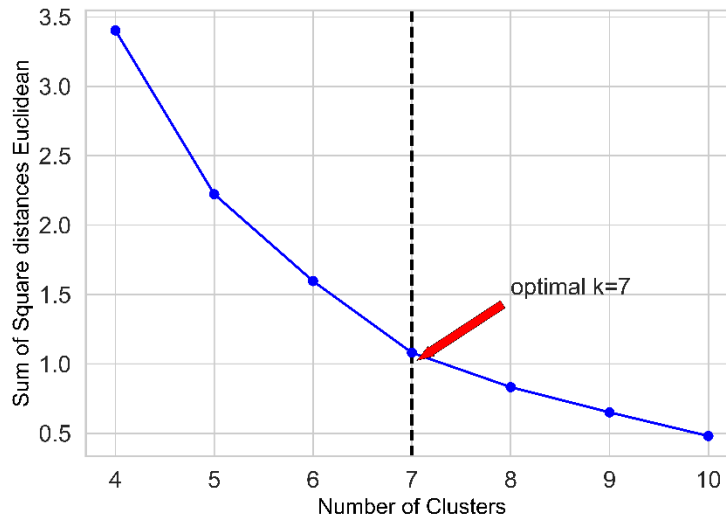


Gambar 5. A) Citra CT *Gaussian Filter* b) Citra CT *Median Filter*

Setelah dilakukan segmentasi *K-means clustering* pada citra CT *filtering*, dilakukan uji optimal *cluster* menggunakan metode *elbow*. Hasil uji metode *elbow* untuk memperoleh nilai *cluster* yang optimal ditunjukkan pada Gambar 6. dan Gambar 7. Visualisasi grafik *elbow* memperlihatkan penurunan seperti siku pada nilai *cluster* yang sama untuk citra CT *gaussian filter* dan *median filter* yakni pada nilai $K=7$ dengan masing-masing nilai *SSE/inertia* 1,020 dan 1,080.



Gambar 6. Grafik *Elbow* Sampel Citra CT *Gaussian Filter* Segmentasi *K-means Clustering*



Gambar 7. Grafik *Elbow* Citra CT *Median Filter* Segmentasi *K-means Clustering*

SIMPULAN

Penerapan metode *median filter* menunjukkan performa yang lebih baik dalam mengurangi derau pada citra CT dibandingkan metode *gaussian filter* berdasarkan parameter nilai PSNR. Segmentasi metode *K-means clustering* citra CT mampu menampilkan daerah terdeteksi objek tumor dengan nilai *cluster* yang optimal pada nilai K ke tujuh.

REFERENSI

- Dhanachandra, N. M. (2015). Image Segmentation Using K-Means Clustering ALgorithm and Subtractive Clsutering Algorithm. *Procedia Comput Sci*, Vol.54,764-771.
- Gurusamy, R. d. (2017). A Machine Learning Approach for MRI Brain Tumor Classifictaion . *Computers, Material and Continua*, 53.(2),91-109.
- Hidayat, E. Y. (2013). Analisis Steganografi Metode Least Significant Bit (LSB) dengan Penyisipan Sekuensial dan Acak Secara Kuantitatif dan Visual. *Techno.COM*, Vol.12.No.3.
- Lampignano, J. d. (2010). *Bontrager's Handbook Of Radiographic Positioning And Techniques, 9th Edition*. United States of America: Elsevier.
- (n.d.). Retrieved from IARC: <https://geo.iarc.fr>
- Rakesh, M. R. (2013). Hybrid Median FIlter for Impluse Noise Removal of an Image in Image Restoration . *International Journal of Advanced Research in Electrical Electronics and Instrumentation Engineering*, 10pp.5117-5124.
- Utami, T. A. (2017). Perbandingan K-Means dan Fuzzy C-Means untuk Pengelompokkan Data User Knowledge Modeling. *Seminar Nasional Teknologi Informasi, Komunikasi dan Industri (SNTIKI) 9:Fakultas Sains dan Tkenologi, UIN Sultan Syarif Kasim Riau*, 219-226.

