

# PENERAPAN FUZZY C-MEANS PADA KLASTERISASI KARAKTERISTIK PENGUNJUNG WEBSITE PMB STT – PLN UNTUK MENINGKATKAN KEPADATAN KUNJUNGAN

Sely Karmila<sup>1</sup>, dan Rakhmat Arianto<sup>2</sup> Rahmat Rezki<sup>3</sup>

<sup>1,2,3</sup>Jurusan Teknik Informatika (Institut Teknologi PLN)

Jl. Lingkar Luar Barat Duri Kosambi Cengkareng Jakarta Barat, Indonesia

E-mail: <sup>1</sup>[sely\\_karmila@itpln.ac.id](mailto:sely_karmila@itpln.ac.id), <sup>2</sup>[arianto@sttpln.ac.id](mailto:arianto@sttpln.ac.id), <sup>3</sup>[rahmat1531037@sttpln.ac.id](mailto:rahmat1531037@sttpln.ac.id)

## ABSTRACT

*Bringing high website visitors is a difficult challenge to face because of the many interesting website contents popping up. This problem is also faced by the STT-PLN New Student Admission (PMB) website. One of the things that need to be considered in increasing website visitors is by knowing the demographic data of visitor characteristics. Changes in behaviour are also caused by the attractiveness of quality content that can attract visitor interest and can improve SEO (Search Engine Optimization) to be able to gain visitor density (traffic) so that it is easily indexed by Google's search engine. The data used in this study are location data, web browser data, parent attraction data and teenage attraction data of visitors to the PMB STT-PLN website. The method that can be done is to cluster the demographic characteristics of visitors using Fuzzy C-Means and Root Mean Square Error to test the accuracy of the cluster. Optimal accuracy results obtained for each data are when the browser data is grouped into 4 clusters with an error value of 0.000275229, location data into 2 clusters with an error value of 0.003197591, teenage attractiveness data into 4 clusters with an error value of 0.000322929, and data the appeal of parents into 2 clusters with an error value of 0.084997533.*

**Keywords:** Clustering, Fuzzy C-Means, Data Analytics, Web Usage Mining, Root Mean Square Error

## ABSTRAK

*Mendatangkan pengunjung website yang tinggi merupakan tantangan yang sulit untuk dihadapi karena banyaknya konten website yang menarik bermunculan. Permasalahan ini juga dihadapi oleh website Penerimaan Mahasiswa Baru (PMB) STT-PLN. Salah satu hal yang perlu diperhatikan dalam meningkatkan pengunjung website adalah dengan mengetahui data demografi karakteristik pengunjung. Perubahan perilaku juga disebabkan oleh daya tarik dari konten yang berkualitas yang dapat menarik daya tarik (interest) pengunjung serta dapat meningkatkan SEO (Search Engine Optimization) dengan tujuan dapat mendulang kepadatan pengunjung (traffic) sehingga mudah terindeks oleh mesin pencari Google. Data yang digunakan pada penelitian ini adalah data lokasi, data web browser, data daya tarik orang tua serta data daya tarik remaja pengunjung website PMB STT-PLN. Cara yang dapat dilakukan adalah dengan mengklasterisasi data demografi karakteristik pengunjung menggunakan Fuzzy C-Means dan Root Mean Square Error untuk menguji tingkat akurasi klaster. Hasil akurasi optimal yang diperoleh masing-masing data yaitu pada saat data browser dikelompokkan menjadi 4 klaster dengan nilai error sebesar 0.000275229, data lokasi menjadi 3 klaster dengan nilai error sebesar 0.003197591, data daya tarik remaja menjadi 4 klaster dengan nilai error sebesar 0.000322929, dan data daya tarik orang tua menjadi 2 klaster dengan nilai error sebesar 0.084997533.*

**Kata kunci:** Klasterisasi, Fuzzy C-Means, Data Analytics, Web Usage Mining, Root Mean Square Error

## 1. PENDAHULUAN

Mendatangkan pengunjung *website* yang tinggi merupakan tantangan yang sulit untuk dihadapi, baik karena semakin banyak *website* yang bermunculan dengan temanya sama maupun karena kesulitan membuat konten atau isi yang menarik dikarenakan ketidaktahuan instansi atau perusahaan terhadap karakteristik pengunjung *website* mereka. Untuk melihat karakteristik pengunjung *website*, terdapat beberapa alat bantu (*tools*) salah satunya adalah *Google Analytics*. *Google Analytics* merupakan perangkat lunak gratis yang disediakan oleh *Google* yang digunakan untuk melihat statistik *website*, berupa jumlah pengunjung, kota, sistem operasi, browser, penyedia layanan, resolusi layar, lalu lintas hingga alur perilaku serta menyajikan hasil informasi hasil web usage mining yang menggunakan teknik *page tags* [1].

Pertimbangan yang juga perlu diperhatikan, bahwa perubahan perilaku disebabkan oleh daya tarik dari konten atau artikel yang berkualitas yang dapat menarik daya tarik (*interest*) pengunjung serta dapat meningkatkan SEO (*Search Engine Optimization*) dengan tujuan dapat mendulang *traffic* sehingga mudah terindeks oleh mesin pencari *Google*.

Adapun permasalahan saat ini adalah untuk mengetahui informasi tentang karakteristik pengunjung *website* PMB STT-PLN dengan tujuan untuk meningkatkan kepadatan kunjungan, Bagian Pemasaran dan Admisi (BPA) STT-PLN selaku pengelola *website* merasa kesulitan dalam melihat serta mendapatkan informasi terkait karakteristik perilaku pengunjung *website*, sehingga Bagian Pemasaran dan Admisi (BPA) STT-PLN membutuhkan bantuan *Google Analytics* dalam melihat serta mendapatkan data terkait karakteristik pengunjung *website* tersebut. Hasil yang diperoleh dari *Google Analytics* berupa data karakteristik pengunjung yang diambil dari data lokasi, data browser serta data daya tarik atau minat pengunjung kemudian dianalisa menggunakan bantuan *Microsoft Excel* dengan fitur *pivoting*. Namun hasil yang diperoleh dari proses *pivoting* tersebut memiliki kekurangan berupa tidak diketahuinya kelompok terhadap hasil analisis yang dapat membantu Bagian Pemasaran dan Admisi (BPA) STT-PLN dalam menentukan tingkatan karakteristik pengunjung yang dapat dijadikan acuan dalam meningkatkan kepadatan kunjungan *website* serta upaya dalam menentukan strategi *marketing* yang tepat untuk kedepannya.

## 2. LANDASAN TEORI

### a. Web Mining

*Web mining* atau yang lebih sering dikenal dengan istilah data mining web merupakan suatu proses dalam menemukan hubungan intrinsik (misalkan, informasi yang menarik dan bermanfaat) dari data web, yang disajikan dalam bentuk teks, *link*, atau informasi penggunaan web itu sendiri. *Web mining* merupakan ekstraksi pola-pola penting dan bermanfaat namun tersimpan secara implisit pada kumpulan data yang relatif besar pada layanan *world wide web* [2].

*Web mining* terbagi menjadi 3 (tiga) kategori yaitu *web content mining*, *web usage mining*, dan *web structure mining*.

### b. Clustering

*Clustering* merupakan proses pembagian obyek ke dalam kelompok atau "*cluster*" sehingga obyek dalam suatu kelompok cenderung lebih mirip satu sama lain dibandingkan dengan obyek milik kelompok yang berbeda. Berdasarkan pendekatan *partitional* dalam *clustering* terdapat istilah *Hard Clustering* dan *Fuzzy Clustering*. Pada dasarnya klaster dapat dilihat sebagai himpunan bagian dari himpunan data. Himpunan dapat berupa himpunan *fuzzy* atau *crisp* (hard). Metode *hard clustering*

berdasarkan teori himpunan klasik membagi cluster secara tegas yang mengharuskan objek dapat menjadi anggota atau bukan anggota dari suatu klaster [3].

**c. Fuzzy C-Means**

*Fuzzy C-Means* (FCM) adalah salah satu teknik pengklasteran data *unsupervised learning* berupa keberadaan tiap-tiap titik data dalam suatu *cluster* ditentukan oleh derajat keanggotaan. Teknik ini pertama kali diperkenalkan oleh James C Bezdek pada tahun 1981. Algoritma *Fuzzy C-Means* merupakan algoritma klastering yang paling banyak dikenal dan memiliki ketepatan pada penempatan pusat klaster, namun memiliki kelemahan pada penentuan jumlah klaster optimal [4].

Algoritma dari *Fuzzy C-Means* adalah sebagai berikut :

1. Masukkan data yang akan di cluster  $x$ , berupa matriks berukuran  $n \times m$  ( $n$ = jumlah sampel data,  $m$ =atribut setiap data)  $X_{ij}$  data ke- $i$  ( $i=1,2,\dots,n$ ) , atribut ke- $j$  ( $j=1,2,\dots,m$ )

2. Inisialisasi = tentukan jumlah cluster ( $k \geq 2$ ), tetukan bobot pangkat ( $w > 1$ ), tentukan jumlah maksimal iterasi, tentukan ambang batas perubahan nilai fungsi objektif (jika perlu juga perubahan nilai centroid)

Ket :

- a. Jumlah cluster ( $c$ )
- b. Pangkat ( $w$ )
- c. Maksimum iterasi
- d. Error terkecil yang diharapkan
- e. Fungsi objektif awal ( $P_0=0$ )
- f. Iterasi awal ( $t=1$ )

3. Berikan nilai awal pada matriks fuzzy pseudo-partition, dengan syarat

$$\sum_{j=1}^k u_{ij} < n$$

4. Lakukan langkah

$$\begin{matrix} u_{11}(x_1) & u_{12}(x_1) & u_{13}(x_1) \\ u_{21}(x_2) & u_{22}(x_2) & u_{23}(x_2) \end{matrix} \text{ dan langkah } C_{ij} = \frac{\sum_{i=1}^n (U_{ij})^w X_{ij}}{\sum_{i=1}^n (U_i)^w}$$

- a. Apabila perubahan pada nilai fungsi objektif masih di atas nilai ambang batas yang ditentukan atau
- b. Perubahan pada nilai centroid masih diatas nilai ambang batas yang ditentukan atau
- c. Iterasi maksimal belum tercapai

5. Hitung nilai centroid dari masing – masing cluster menggunakan persamaan

$$C_{ij} = \frac{\sum_{i=1}^n (U_{ij})^w X_{ij}}{\sum_{i=1}^n (U_i)^w}$$

6. Hitung kembali matrik fuzzy pseudo – partition menggunakan persamaan

$$U_{ij} = \frac{\sum_{i=1}^n D(X_i, C_i)^{\frac{-2}{w-1}}}{\sum_{i=1}^k D(X_i, C_i)^{\frac{-2}{w-1}}}$$

7. Cek kondisi berhenti :

- a. Jika  $t > \max$  iter maka berhenti
- b. Jika tidak,  $t=t+1$ , ulangi langkah ke-6

**d. Root Mean Square Error**

RMSE merupakan metode alternative yang digunakan untuk mengevaluasi teknik peramalan yang digunakan melalui tingkat akurasi hasil prakiraan pada suatu model. RMSE dapat diambil dari nilai rata-rata jumlah kuadrat kesalahan, serta dapat menyatakan ukuran besarnya kesalahan yang dihasilkan oleh suatu model prakiraan. RMSE merupakan parameter yang digunakan untuk mengevaluasi nilai hasil dari pengukuran terhadap nilai sebenarnya atau nilai dianggap benar [5].

Persamaan untuk menghitung nilai RMSE adalah sebagai berikut :

$$RMSE = \sqrt{\sum_{i=1}^n \frac{(X_i' - X_i)^2}{n}}$$

**Dimana :**

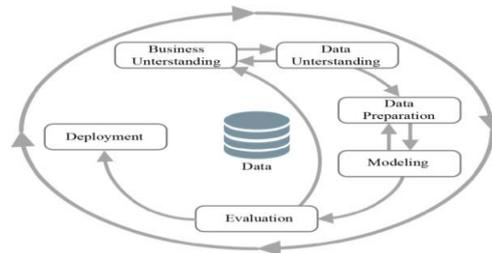
- $X_i'$  = nilai data training / perhitungan manual data
- $X_i$  = nilai exact / nilai aplikasi / nilai estimasi
- $N$  = jumlah data

**3. METODE/PERANCANGAN PENELITIAN**

Pada bagian perencanaan penelitian ini, penulis akan menjelaskan mengenai tahapan perancangan yang digunakan.

**2.1. Perancangan CRISP-DM**

Pada penelitian ini penulis menggunakan metode pengembangan CRISP-DM dalam. Berikut adalah 6 fase siklus CRISP-DM :



**Gambar 1** Alur Perancangan CRISP-DM

**1. Business Understanding Phase (Fase Pemahaman Bisnis)**

Pada tahap pemahaman bisnis ini terfokus terhadap pemahaman kebutuhan berdasarkan penilaian pebisnis. Tujuan bisnis melakukan penelitian untuk melihat karakteristik pengunjung website PMB STT-PLN.

**a. Menentukan Tujuan Bisnis (Determine Business Objectives)**

Tujuan bisnis melakukan penelitian untuk melihat karakteristik pengunjung website PMB STT-PLN.

**b. Menilai Situasi (Asses the Situation)**

Data *browser*, data daya tarik remaja, data daya tarik orang tua, dan data lokasi tempat mengakses web merupakan data yang akan digunakan.

**c. Mentukan Tujuan Data Mining (Determine Data Mining Goals)**

Tujuan data mining atau tujuan dari penelitian ini adalah mengklusterisasi karakteristik pengunjung web PMB STT-PLN dengan menggunakan metode *data mining*.

**2. Data Understanding Phase (Fase Pemahaman Data)**

Pada penelitian ini data yang akan digunakan adalah berupa data lokasi akses, data *interest* sektor yang menitik beratkan kepada kategori remaja usia 18-24 tahun serta kategori orang tua dengan usia 45-54 tahun, dan data *browser* yang dipakai pada saat mengakses web PMB STT-PLN.

**3. Data Preparation Phase (Fase Persiapan Data)**

Pada tahap ini adalah melakukan beberapa proses penyederhanaan data antara lain :

**a. Mendeskripsikan Data Set ( Data Set Description )**

Pada tahap ini merupakan pengumpulan data-data yang diperlukan untuk proses pengklasteran karakteristik pengunjung *website*.

**b. Memilih Data ( Data Selection )**

Pemilihan atribut – atribut data yang digunakan mau atribut data yang tidak digunakan.

**Tabel 1.** Atribut yang tidak digunakan

Nama Atribut	Keterangan Tidak Digunakan
Admisi STT-PLN (Goal 1 Conversion Rate)	Ketiga atribut ini merupakan atribut hanya sebagai identitas yang tidak dibutuhkan dalam proses klasterisasi karakteristik pengunjung <i>website</i> PMB
Admisi STT-PLN (Goal 1 Completions)	
Admisi STT-PLN (Goal 1 Value)	

**Tabel 2.** Atribut yang Digunakan

Nama Atribut	Keterangan Tidak Digunakan
<i>Users</i>	Atribut ini digunakan untuk melihat jumlah pengunjung keseluruhan yang mengunjungi <i>website</i> PMB STT-PLN.
<i>New Users</i>	Atribut ini digunakan untuk melihat jumlah pengunjung yang baru pertama kali yang mengunjungi <i>website</i> PMB STT-PLN.
<i>Sessions</i>	Atribut ini digunakan untuk melihat jumlah kunjungan mengunjungi <i>website</i> PMB STT-PLN.
<i>Bounce Rate</i>	Atribut ini digunakan untuk melihat persentase kunjungan dari user yang hanya membuka 1 halaman saja pada <i>website</i> PMB STT-PLN.
<i>Pages/Session</i>	Atribut ini digunakan untuk melihat jumlah halaman yang diakses atau dibuka oleh pengunjung dalam setiap kunjungan ( <i>sessions</i> )
<i>Avg Session Duration</i>	Atribut ini digunakan untuk menggambarkan waktu pengunjung terhadap sebuah <i>website</i> dengan kata lain dapat menunjukkan lamanya waktu yang digunakan pengunjung dalam setiap kunjungannya terhadap <i>website</i>

c. Membersihkan Data ( *Data Cleaning* )

Dari hasil atribut– atribut yang tidak digunakan, akan dilakukan pembersihan pada data yang terdapat pada atribut ini. Tujuan menghapus atribut yang tidak digunakan adalah agar proses klusterisasi dengan metode *Fuzzy C-Means* hasil yang dikeluarkan lebih tepat dan hasil klaster tidak terdapat nilai *none* (tidak diketahui).

1. Data Browser

Dari total sebanyak 18 data yang didapat pertama kali, terdapat 1 noise data yang mengganggu proses yang nilai datanya berupa *not set* (tidak ditentukan).sehingga hasilnya menjadi 17 data.

2. Data Lokasi

Dari total sebanyak 176 data lokasi yang berupa data kota-kota yang ada di Indonesia, maka data akan dibersihkan dan dan diubah menjadi 33 data.

3. Data Daya Tarik Orang Tua

Data daya tarik orang tua diperoleh dari data daya tarik (*interest*). Data awal daya tarik keseluruhan yang terdapat sejumlah 577 data dan dipilih hanya daya tarik yang targetnya berupa orang tua dengan rentang nilai 48-52 tahun sehingga menghasilkan data sebanyak 50 data.

4. Data Daya Tarik Remaja

Data daya tarik orang tua diperoleh dari data daya tarik (*interest*). Data awal daya tarik keseluruhan yang terdapat sejumlah 577 data. Setelah dilakukan pembersihan data terhadap target berupa remaja dengan rentang nilai 18-24 tahun, maka data yang diperoleh sebanyak 229 data.

**4. Modelling Phase (Fase Pemodelan)**

Dalam fase ini penulis melakukan percobaan klaster dengan 2,3 dan 4 klaster menggunakan metode *Fuzzy C-Means* sebagai berikut :

**2 Klaster**

- a. Langkah pertama yang harus dilakukan adalah menginisialisasi matriks partisi awal secara acak, berupa matriks berukuran n x m (n= jumlah sampel data, m=atribut setiap data)

**Tabel 3.** Matriks Partisi Awal

Data ke i	C1	C2
1	0.197828015	0.393869781
2	0.440373354	0.319901777

Data ke i	C1	C2
3	0.718157037	0.510791219
4	0.267551396	0.063730385
5	0.618225606	0.582126118
6	0.215504573	0.869907214
7	0.623281521	0.242027618
8	0.067153888	0.306135801
9	0.034671372	0.468398926
10	0.875264577	0.217311284
11	0.618818303	0.154455479
12	0.7708221	0.553153874
13	0.612318103	0.997599182
14	0.15694481	0.099332973
15	0.733513654	0.096185383
16	0.856853372	0.55853262
17	0.006959905	0.190967276

- b. Setelah matrix partisi awal telah di tentukan, langkah selanjutnya adalah menghitung nilai pusat kluster dengan menentukan terlebih dahulu bobot pangkat yang digunakan ( $k > 1$ ) serta jumlah kluster yang diinginkan, dan juga *epsilon* (error terkecil yang diharapkan) dari proses iterasi nantinya. Setelah itu lakukan perhitungan pusat kluster dengan menggunakan rumus :

$$C_{ij} = \frac{\sum_{i=1}^n (U_{ij})^w X_{ij}}{\sum_{i=1}^n (U_i)^w}$$

Sehingga di dapatkan pusat cluster berikut :

**Tabel 4 Centroid Kluster**

Pusat Cluster (V)	1040.04	1036.301	1945.957	0.789845	1.2953	86.777
	2228.287	2217.364	4195.426	0.829549	1.273804	83.74545

- c. Setelah pusat cluster di dapat, langkah selanjutnya adalah menghitung nilai fungsi objektif yang nilainya digunakan sebagai kunci iterasi berhenti.

**Tabel 5 Fungsi Objektif**

Data ke i	C1^2	C2^2	L1	L2	L1+L2+L3
1	0.304046267	0.000478422	362342570.5	617221.7115	362959792.2
2	0.008838704	0.043573171	99968.53521	531447.1116	631415.6468
3	0.081573772	0.018654075	637023.0575	100159.0506	737182.1081

Data ke i	C1^2	C2^2	L1	L2	L1+L2+L3
4	0.000766149	0.677480743	6088.589891	3252207.519	3258296.108
5	0.136550074	0.101108011	1041440.404	270535.226	1311975.63
6	0.000967087	0.005891982	7526.972755	8798.241377	16325.21413
7	0.737775519	0.444627643	5863088.355	610121.0704	6473209.425
8	0.264694833	0.73579607	2727630.89	198292.9822	2925923.872
9	0.835263682	0.55674969	8670393.599	167984.3614	8838377.96
10	0.007168675	0.401284503	78823.34261	146916.9126	225740.2552
11	0.032490147	0.076094279	358003.1726	28341.68247	386344.855
12	0.059177679	0.991702983	655289.5103	376851.2081	1032140.718
13	0.26513888	0.086631886	2939071.275	32807.62245	2971878.898
14	0.201421114	0.807359898	2236395.658	308109.9082	2544505.566
15	0.400334968	0.335860199	4470737.183	134032.1627	4604769.346
16	0.624066301	0.258336686	6963682.798	101359.7896	7065042.587
17	0.27535826	0.04560851	3074155.372	17947.80931	3092103.181
<b>Fungsi Obejktif</b>					409075023.6

d. Setelah itu maka proses perhitungan akan dilanjutkan dengan menghitung matriks *fuzzy pseudo-partition* (matriks U) untuk melakukan proses iterasi selanjutnya.

**Tabel 6.** Matriks Partisi U

Data ke i	L1	L2	LT = L1+L2	U1=L1/LT	U2=L2/LT
1	7.74587E-10	6.6066E-10	1.43525E-09	0.54	0.46
2	1.02184E-05	9.44899E-08	1.03129E-05	0.99	0.01
3	7.71687E-07	2.17525E-07	9.89212E-07	0.78	0.22
4	6.81088E-07	3.1476E-07	9.95848E-07	0.68	0.32
5	2.88374E-07	8.0018E-07	1.08855E-06	0.26	0.74
6	1.81109E-07	2.53987E-06	2.72097E-06	0.07	0.93
7	1.65813E-07	3.54442E-06	3.71023E-06	0.04	0.96
8	7.23429E-08	1.7954E-06	1.86774E-06	0.04	0.96
9	7.16918E-08	1.71125E-06	1.78295E-06	0.04	0.96
10	6.45893E-08	1.08535E-06	1.14994E-06	0.06	0.94
11	6.43715E-08	1.07026E-06	1.13463E-06	0.06	0.94
12	6.38133E-08	1.0334E-06	1.09721E-06	0.06	0.94
13	6.36412E-08	1.02054E-06	1.08418E-06	0.06	0.94
14	6.34437E-08	1.00552E-06	1.06896E-06	0.06	0.94
15	6.29511E-08	9.78828E-07	1.04178E-06	0.06	0.94
16	6.29217E-08	9.73233E-07	1.03615E-06	0.06	0.94
17	6.28698E-08	9.70087E-07	1.03296E-06	0.06	0.94

- e. Setelah hasil perhitungan matriks partisi U didapatkan, periksa fungsi objektif apakah sudah dibawah nilai *epsilon* (error terkecil perhitungan) dengan mengurangkan fungsi objektif saat ini dengan fungsi objektif sebelumnya. Apabila fungsi objektif sudah dibawah *epsilon* maka proses iterasi berhenti dan hasil kluster diambil dari nilai maksimal dari matriks partisi U. Namun apabila nilai fungsi objektif masih diatas atau sama dengan nilai *epsilon*, maka proses iterasi dilanjutkan dengan mengulangi langkah – langkah sebelumnya dari langkah poin (b).

**5. Evaluation Phase (Fase Evaluasi)**

Evaluasi pada fase ini lebih memfokuskan pada model atau pola pola yang dihasilkan oleh metode *Fuzzy C-Means*. Model yang di dapatkan dari proses perhitungan atau analisa metode dapat dijadikan sebagai bahan acuan dalam menganalisa serta melihat karakteristik pengunjung yang lebih dominan.

**6. Deployment Phase (Fase Penyebaran)**

Pada fase deployment berupa laporan atau hasil report proses yang telah dilakukan dengan memvisualisasikan hasil pengetahuan (*knowledge*) dari proses perhitungan yang didapat.

**4. HASIL DAN PEMBAHASAN**

**3.1. Hasil Penentuan Kluster**

Berdasarkan pengujian akurasi yang dilakukan oleh penulis menggunakan analisis akurasi *Root Mean Square Error*, maka hasil akurasi tiap kluster yang didapatkan dari proses pengujian akurasi menggunakan RMSE ini adalah sebagai berikut :

**Tabel 7.** Hasil Pengujian Menggunakan RMSE

Jenis Data	2 Kluster	3 Kluster	4 Kluster
Data Browser	0.000336446	0.000286406	0.000275229
Data Lokasi	0.475740888	0.003197591	0.163787342
Data Daya Tarik Remaja	0.433091749	0.442163373	0.000322929
Data Daya Tarik Orang Tua	0.084997533	0.227404433	0.226435105

**3.2. Pembahasan**

Data yang dipakai dalam kasus ini adalah data yang diambil dari data analitik website PMB STT-PLN berupa data *browser*, data daya tarik remaja rentang usia 18-24 tahun, data daya tarik orang tua rentang usia 45-54 tahun, serta data lokasi pengunjung yang akan dikelompokkan berdasarkan kategori tinggi, sedang, rendah , atau tidak sama sekali. Untuk mengelompokkan data tersebut penulis menggunakan metode *Fuzzy C-Means*. Metode *Fuzzy C-Means* dipilih karena mudah diimplementasikan serta memiliki kemampuan untuk mengelompokkan jumlah data yang besar yang merupakan ciri-ciri dari jumlah data analitik.

Dari proses analisa data menggunakan metode *Fuzzy C-Means* yang diperoleh, penulis melakukan uji coba terhadap setiap data yang dipakai menggunakan model pengelompokan 2 kluster, 3 kluster dan 4 kluster. Untuk kategori karakteristik yang masuk ke dalam 2 kluster akan dikategorikan dengan tinggi dan rendah, untuk kategori 3 kluster dikategorikan dengan tinggi, sedang, dan rendah, sedangkan untuk kategori 4 kluster akan dikategorikan dengan tinggi, sedang, rendah, dan sangat rendah.

## 5. KESIMPULAN DAN SARAN

### 4.1. Kesimpulan

Berdasarkan hasil penelitian yang sudah dilakukan yang dimulai dari tahap awal sampai dengan proses pengujian aplikasi Klasterisasi Karakteristik Pengunjung Website PMB STT-PLN Untuk Meningkatkan Kepadatan Kunjungan Menggunakan *Fuzzy C-Means* dapat disimpulkan bahwa :

1. Untuk menentukan karakteristik pengunjung website menggunakan *Fuzzy C-Means* terdapat beberapa ketentuan yang digunakan sebagai atribut atau parameter, diantaranya adalah *users*, *new users*, *session*, *bounce rate*, *pages/session*, dan *average session duration* yang didapat dari data-data yang digunakan seperti data lokasi, data browser, data daya tarik orang tua, dan data daya tarik remaja.
2. Dari hasil uji coba yang di lakukan, diperoleh bahwa hasil klaster optimal yang diperoleh masing-masing data yaitu pada saat data browser dikelompokkan menjadi 4 klaster dengan nilai *error* sebesar 0.000275229, data lokasi menjadi 3 klaster dengan nilai *error* sebesar 0.003197591, data daya tarik remaja menjadi 4 klaster dengan nilai *error* sebesar 0.000322929, dan data daya tarik orang tua menjadi 2 klaster dengan nilai *error* sebesar 0.084997533.

### 4.2. Saran

Berdasarkan hasil penelitian yang telah dilakukan, adapun saran yang dapat diberikan oleh penulis demi kesempurnaan aplikasi ini adalah sebagai berikut:

1. Pada penelitian selanjutnya diharapkan dapat menambahkan metode klaster lainnya yang bertujuan untuk mengoptimalkan perbandingan antara hasil pengelompokan karakteristik pengunjung *website* yang dihasilkan.
2. Pada penelitian berikutnya diharapkan dapat menggabungkan hasil klaster antara data *analytic website* STT-PLN dengan data *insight* sosial media STT-PLN yang dikelola oleh Bagian Pemasaran dan Admisi STT-PLN untuk mengetahui lebih lanjut pola perilaku dan karakteristik pengunjung *website* dengan sosial media.

## DAFTAR PUSTAKA

- [1] A. A. Hendriadi, "Analisis web menggunakan google analytic untuk menyajikan laporan data situs dalam proses optimasi," vol. 1, no. 4, pp. 20–26, 2015.
- [2] T. Suratno, T. Prahasto, and A. Fatchur, "Web Usage Mining , Pattern Discovery dan Log File," vol. 02, pp. 94–99, 2011.
- [3] E. B. Susanto, "Evaluasi Hasil Klaster Pada Dataset Iris , Soybean-small , Wine Menggunakan Algoritma Fuzzy C-Means dan K-," vol. 2, no. 1, pp. 6–13, 2016.
- [4] S. Mashfuufah and D. Istiawan, "Penerapan Partition Entropy Index , Partition Coefficient Index dan Xie BeniIndex untuk Penentuan Jumlah Klaster Optimal pada Algoritma Fuzzy C-Means dalam Pemetaan Tingkat Kesejahteraan Penduduk Jawa Tengah," pp. 51–60, 2018.
- [5] F. Febrianti and A. H. Asyhar, "PERBANDINGAN PENGKLUSTERAN DATA IRIS MENGGUNAKAN METODE K-MEANS DAN FUZZY C-," vol. 02, no. 01, 2016.