

Penggunaan Algoritma *Fuzzy C-Means* Untuk Analisis Web Usage Mining (Studi Kasus : Aktifitas Internet Telkom University)

Stefan Bimarsana Syarief¹, Eko Darwiyanto², Veronikha Effendy²

^{1,2,3} *Teknik Informatika, Fakultas Teknik Informatika, Telkom University*

Jalan Telekomunikasi No.1, Dayeuh Kolot, Bandung 40257

stefansyarief@gmail.com¹, ekodarwiyanto@telkomuniversity.ac.id², veffendy@telkomuniversity.ac.id³

Abstrak

Saat ini penggunaan internet sebagai sumber informasi sedang berkembang dengan pesat. Dengan memanfaatkan media internet ini, pengguna dapat memperoleh informasi secara cepat, di mana saja dan kapan saja tanpa terhalang oleh ruang dan waktu. Informasi-informasi tersebut disimpan dalam sebuah wadah yang biasa disebut dengan website. Agar proses pencarian informasi berjalan lebih efektif maka perlu diperhatikan performansi dan kualitas website tersebut sehingga informasi yang didapat oleh pengguna merupakan informasi yang memang diperlukan. Salah satu tolak ukur yang bisa digunakan untuk meningkatkan performansi dan kualitas dari satu website adalah dengan melihat pola kecenderungan pengguna dalam mengakses website tersebut menggunakan ilmu Web Mining. Pada penelitian ini, ilmu web mining yang dipakai adalah clustering menggunakan algoritma fuzzy c-means. Data log yang digunakan berasal dari situs baa.itelkom.ac.id yang kemudian akan dilakukan preprocessing untuk mengambil bagian yang dibutuhkan seperti ip address, url, dan waktu akses user. selanjutnya cluster akan dibentuk dan dianalisis.

Kata Kunci: Fuzzy c-means, web usage mining, clustering, web log, transaksi, pola akses.

Abstract

Nowdays, the usage of internet as the source of information is growing rapidly. By using the internet media, user can get information from anywhere and anytime without the boundaries of time and space. Those information is kept in a place that are usually called websites. The quality and performance of the website should be a concern to make the searching process more effective so that the information that the user is getting is really the information that they need. One of the benchmark that can be used to improve the performance and quality of a certain website is to see user pattern that accessed the web using Web Mining method. In this research the web mining method that is used is clustering and using fuzzy c-means algorithm. The data log that are used originated from baa.itelkom.ac.id sites then the data is preprocessed to get the bits that are needed such as ip address, url, and user access time. The next step is to build the cluster and analyse it.

Keywords : Fuzzy c-Means, web usage mining, clustering, web log, transaction, access pattern

1. Pendahuluan

Saat ini penggunaan internet sebagai sumber informasi sedang berkembang dengan pesat. Informasi tentang berbagai macam aspek kehidupan dapat kita peroleh menggunakan media internet. Dengan memanfaatkan media internet ini, pengguna dapat memperoleh informasi secara cepat, di mana saja dan kapan saja tanpa terhalang oleh ruang dan waktu. Informasi-informasi tersebut disimpan dalam sebuah wadah yang biasa disebut dengan website. Sebuah website dapat menyimpan banyak informasi yang dibutuhkan oleh lebih dari satu pengguna. Agar proses pencarian informasi berjalan lebih efektif maka perlu diperhatikan performansi dan kualitas website tersebut sehingga informasi yang didapat oleh pengguna merupakan informasi yang memang diperlukan. Salah satu tolak ukur yang bisa digunakan untuk meningkatkan performansi dan kualitas dari satu website adalah dengan melihat pola kecenderungan pengguna dalam mengakses website tersebut menggunakan ilmu Web Mining.

Web Mining dapat dibedakan menjadi tiga tipe berbeda yaitu web content Mining, web structure Mining, dan web usage mining. Web usage mining adalah teknik data mining untuk melakukan pencarian terhadap data akses dari data suatu web. Data yang bisa menjadi inputan Web usage mining antara lain access logs, browser logs, user profiles, registration data, user session, dan data lainnya yang merupakan hasil transaksi dengan web. Data kemudian diolah melalui beberapa tahap, Teknik yang digunakan dalam web usage mining merupakan implementasi dari teknik data mining. Teknik-teknik data mining yang sering digunakan antara lain classification, clustering, dan association. Clustering adalah teknik data mining yang secara umum bekerja dengan cara mengelompokkan data berdasarkan kesamaan yang dimiliki oleh kumpulan data yang ada.

Fuzzy c-means adalah algoritma clustering yang dikembangkan berdasarkan algoritma K-means. Secara umum cara kerja kedua algoritma ini adalah dengan menghitung jarak antara data terhadap titik pusat cluster kemudian data digolongkan ke dalam cluster dengan jarak terkecil dari data tersebut. Selanjutnya titik pusat cluster akan bergeser berdasarkan rata-rata jarak seluruh anggota cluster

dengan titik pusatnya. Perbedaannya adalah pada algoritma Fuzzy c-means penggolongan data ke dalam sebuah cluster menggunakan derajat keanggotaan data terhadap semua cluster. Kondisi ini memungkinkan data untuk dibandingkan ke semua cluster sehingga dapat mengatasi masalah data yang memiliki jarak yang sama terhadap dua pusat cluster atau lebih. Dengan demikian pengguna dapat retrieve dan menganalisa informasi lebih banyak dibandingkan dengan penggunaan algoritma hard clustering. Algoritma fuzzy c-means juga termasuk algoritma yang mudah untuk beradaptasi terhadap analisis yang kompleks. Oleh karena itu algoritma ini cocok digunakan untuk kasus web usage mining.

2. Landasan Teori

2.1 Web

World Wide Web atau WWW merupakan salah satu fenomena teknologi yang berkembang sangat pesat saat ini. WWW menyediakan berbagai layanan informasi mengenai berita, iklan, pendidikan, e-commerce dan sebagainya. Informasi yang tersedia dalam WWW tersebut memiliki ukuran yang sangat besar dan terdistribusi secara global di seluruh dunia. Web juga mengandung kekayaan informasi dilihat dari struktur dan penggunaannya (web usage). Web merupakan kumpulan data dan informasi yang sangat berpotensi untuk dilakukan penggalian (mining) agar menghasilkan pengetahuan (knowledge) yang dapat berguna bagi masyarakat maupun pihak-pihak tertentu.

2.2 Web Mining

Menurut Etzioni Oren, web mining diartikan sebagai suatu usaha mengaplikasikan teknik data mining untuk menggali dan mengekstrak informasi yang berguna dari dokumen-dokumen yang tersimpan dalam halaman web secara otomatis. Meskipun memiliki akar terminologi yang sama dengan data mining, namun web mining memiliki perbedaan dari data mining, diantaranya berhubungan dengan sifat datanya yang tidak terstruktur dan sumber datanya yang tidak disimpan di sebuah *data warehouse* namun tersebar di berbagai sumber. Web mining terbagi menjadi 3 (tiga) kategori yaitu *web content mining*, *web structure mining* dan *web usage mining* (Madria, et al., 1999). *Web content* mining berfokus pada usaha untuk menggali informasi dari isi atau *content* yang disajikan di web. *Teknik web content mining* lebih

banyak berhubungan dengan disiplin ilmu *information retrieval* (IR). Sedangkan *web structure mining*, membahas mengenai penggalian informasi web dilihat dari struktur halaman web itu sendiri. *Web structure mining* banyak digunakan untuk menggali keterkaitan antara suatu halaman web dengan halaman web lainnya. Sementara *web usage mining*, berusaha melihat pola atau pattern dari user dalam mengakses web.

2.3 Web Usage Mining

Pada web usage mining, untuk mengatasi masalah dokumen web yang terus berubah, data mentah yang digunakan umumnya akan dibekukan untuk sementara waktu. Kemudian perilaku pengunjung pun dianalisis dalam jangka waktu tertentu [5]. Dari data mentah tersebut akan dimodelkan pola perilaku dan profil dari pengunjung web. Dari pola-pola tersebut digali informasi yang dapat dimanfaatkan, beberapa contohnya selain yang dilakukan pada penelitian ini yaitu untuk kegiatan komersial, digunakan untuk menyimpulkan demographic facts pengguna, yang berguna untuk meningkatkan efektivitas iklan internet. Pada penelitian sebelumnya pun [2, 6] telah menyajikan strategi untuk membantu web servers mengurangi waktu loading web. Hal ini dilakukan dengan melakukan mining web logs untuk menemukan aturan bentuk “document 1 – document 2” sehingga web server dapat mempersiapkan “document 2” ketika “document 1” direquest. Disimpulkan usage pattern web dari user sebelumnya, dapat memainkan peran dalam membantu pengguna lain selanjutnya.

2.4 Fuzzy Clustering

Fuzzy C-means Clustering (FCM) atau dikenal juga sebagai *Fuzzy ISODATA* merupakan salah satu metode *clustering* yang merupakan bagian dari metode *Hard K-Means*. FCM menggunakan model pengelompokan *fuzzy* sehingga data dapat menjadi anggota dari semua kelas atau *cluster* terbentuk dengan derajat atau tingkat keanggotaan yang berbeda antara 0 hingga 1. Tingkat keberadaan data dalam suatu kelas atau *cluster* ditentukan oleh derajat keanggotaannya.

Konsep dasar FCM, pertama kali adalah menentukan pusat *cluster* yang akan menandai lokasi rata-rata untuk tiap-tiap *cluster*. Pada kondisi awal, pusat *cluster* ini masih belum akurat, tiap-tiap data memiliki derajat keanggotaan untuk tiap-tiap *cluster*. Dengan cara memperbaiki pusat *cluster* dan nilai keanggotaan tiap-tiap data secara berulang, maka dapat dilihat bahwa pusat *cluster* akan menuju lokasi yang tepat dan nilai keanggotaan akan menuju ke *cluster* yang tepat. Perulangan ini didasari pada minimasi fungsi objektif dan maksimum iterasi yang telah ditentukan oleh *user*. Langkah-langkah dalam algoritma *Fuzzy C-Means* adalah sebagai berikut :

- Input data yang akan di *cluster* X yang berukuran $i \times n$ (i = jumlah data, n = atribut setiap data)
- Tentukan parameter-parameter dalam *Fuzzy C-Means*. Parameter-parameter tersebut antara lain jumlah *cluster* (c), derajat keanggotaan (m), maksimum iterasi, dan *error* terkecil
- Bangkitkan bilangan random sebagai titik centroid awal
- Hitung derajat keanggotaan

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

- Hitung titik pusat *cluster* yang baru

$$c_j = \frac{\sum_{i=1}^N \delta_{ij}^m \cdot x_i}{\sum_{i=1}^N \delta_{ij}^m}$$

- Hitung fungsi objektif di tiap terasi

$$J = \sum_{i=1}^N \sum_{j=1}^C \delta_{ij} \|x_i - c_j\|^2$$

- Cek kondisi berhenti

$$\epsilon = \Delta_i^N \Delta_j^C \|\delta_{ij}^{k+1} - \delta_{ij}^k\|$$

Jika perubahan derajat keanggotaan < error terkecil atau t > maksimum iterasi maka berhenti, jika ulangi dari penghitungan derajat keanggotaan

2.5 Self Organizing Map

Cluster validity index digunakan untuk mengevaluasi kualitas dari partisi yang dihasilkan oleh algoritma clustering. Cluster validity index yang akan digunakan adalah Partition coefficient and exponential separation (PCAES) index. Cara kerjanya adalah sebagai berikut:

Pertama kita asumsikan $X = \{x_1, \dots, x_n\}$ merupakan data set dalam R^s dan

$\mu = \{ \mu_1, \dots, \mu_n \}$ merupakan fuzzy c-partition dari fuzzy clustering.

Ada dua faktor yang dijadikan bahan pertimbangan untuk memvalidasi sebuah cluster. Pertama adalah partition coefficient dan yang kedua adalah exponential separation. Kedua faktor tersebut kemudian digabungkan untuk membuat indeks validitas baru yang disebut dengan partition coefficient and exponential separation (PCAES) index. Definisi indeks PCAES untuk cluster i adalah sebagai berikut

$$PCAES_i = \sum_{j=1}^n \mu_{ij}^2 / \mu_M - \exp \left(- \min_{k \neq i} \{ \|a_i - a_k\|^2 \} / \beta_T \right) \tag{12}$$

where

$$\mu_M = \min_{1 \leq i \leq c} \left\{ \sum_{j=1}^n \mu_{ij}^2 \right\} \quad \text{and} \quad \beta_T = \frac{\sum_{i=1}^c \|a_i - \bar{a}\|^2}{c} \tag{13}$$

Salah satu term dari normalized partition coefficient (NPC) digunakan untuk mengukur kepadatan dari cluster i .

$$\sum_{j=1}^n \mu_{ij}^2 / \mu_M$$

Sedangkan untuk perhitungan terpisahnya satu cluster dengan cluster yang lain dengan memanfaatkan fungsi eksponensial yang mengukur jarak antara cluster i dengan cluster terdekatnya.

$$\exp \left(- \min_{k \neq i} \{ \|a_i - a_k\|^2 \} / \beta_T \right)$$

Dengan demikian cluster dapat dilihat dari dua sudut pandang. Pertama adalah dari sudut pandang normalized validation coefficient dan yang kedua adalah sudut pandang exponential partition. Nilai PCAESi yang besar mengindikasikan bahwa cluster i padat didalamnya dan terpisah dari cluster yang lain. Nilai PCAESi yang kecil atau bahkan negatif mengindikasikan bahwa cluster yang terbentuk kurang baik. Rumus PCAES validity index adalah :

$$\begin{aligned} PCAES(c) &= \sum_{i=1}^c PCAES_i \\ &= \sum_{i=1}^c \sum_{j=1}^n \mu_{ij}^2 / \mu_M \\ &\quad - \sum_{i=1}^c \exp \left(- \min_{k \neq i} \{ \|a_i - a_k\|^2 \} / \beta_T \right) \end{aligned} \tag{19}$$

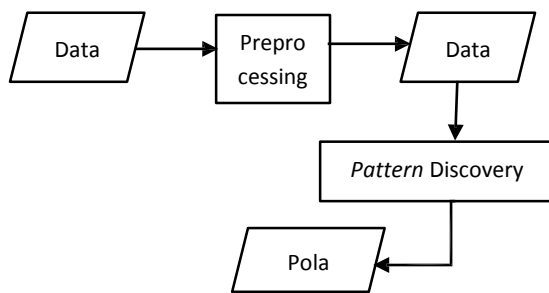
Obviously,

$$-c < PCAES(c) < c \tag{20}$$

3. Perancangan Sistem

Sistem yang akan dibangun pada tugas akhir ini menerapkan metode clustering, yaitu algoritma fuzzy c-means. Pada aplikasi ini akan dibangun sebuah sistem yang berfungsi untuk mengklusterisasi data web log untuk menemukan pola web usage. Dalam mengklusterisasi data tersebut, digunakan algoritma Fuzzy C-Means. Pertama data raw log dari server akan melalui tahapan pre processing

Setelah proses preprocessing selesai, data akan berbentuk sebagai urutan transaksi. Urutan transaksi tersebut kemudian akan di proses clustering menggunakan fuzzy c-means clustering sehingga menghasilkan grup-grup transaksi sesuai dengan karakteristiknya.



Gambar 3.1 Diagram Blok

3.1 Preprocessing

Proses data formatting and cleansing, bertujuan untuk membersihkan data dari bagian-bagian yang tidak diperlukan sehingga yang tersisa hanya bagian data yang diperlukan. Proses ini menggunakan fungsi substring yang ada dengan mengelompokkan baris-baris data menjadi beberapa bagian sebelumnya

```

    39.213.217.135 - - [01/Sep/2013:00:00:11
    +0700] "GET /wp-
    content/themes/reports_and_folders_ote115/
    images/item-right.png HTTP/1.1" 200 535
    
```

Gambar 3.2 Baris log sebelum formatting

Gambar diatas merupakan contoh baris log sebelum dilakukan preprocessing. Sistem

kemudian akan memprosesnya menjadi seperti berikut.

```

    [39.213.217.135]--[01/Sep/2013:00:00:11]+0700] "[GET]
    [/wp-
    content/themes/reports_and_folders_ote115/images/item-
    right.png HTTP/1.1"] [200] [535]
    
```

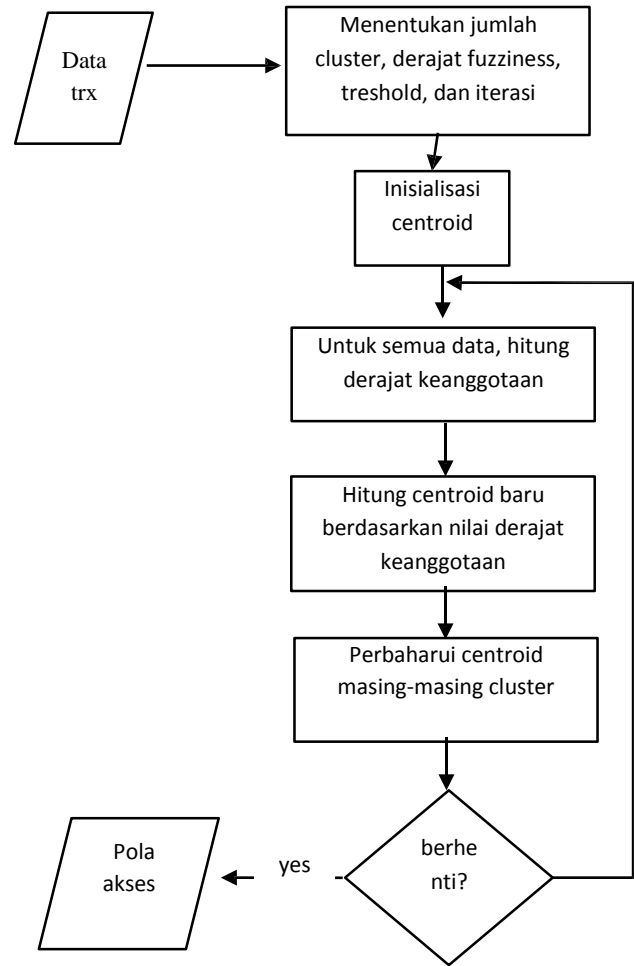
Setelah proses formatting dilakukan, langkah selanjutnya adalah membersihkan baris log dari bagian yang tidak perlu seperti data berekstensi .jpg, .gif, ukuran byte, dan status. Langkah selanjutnya adalah menghilangkan log yang tidak berhubungan dengan penelitian. Ada tiga kategori log yang akan dihapus. Pertama adalah log yang didalamnya terdapat IP address yang mengacu ke server. Kedua adalah log yang menghasilkan eror, biasanya log yang mengakses halaman kosong dan log yang memiliki respons kurang dari 200 dan lebih dari 299. Ketiga adalah log yang mengakses robot.txt. Setelah selesai menghilangkan log yang tidak berhubungan dengan penelitian, semua karakter akan diubah menjadi lowercase untuk kepentingan saat membandingkan data.

1. User Identification merupakan proses mengidentifikasi unique user. Pada penelitian ini IP address dipilih untuk menjadi identitas user.
2. User Session Identification adalah proses mengidentifikasi aktivitas transaksi pada selang waktu tertentu. Jangka waktu yang ditentukan adalah 30 menit. Jika selang waktu Siantar request melebihi jangka waktu tersebut maka akan dianggap sebagai sesion yang berbeda yang berarti transaksi selanjutnya.

User identification dan user session identification membentuk sebuah transaksi atau disebut juga dengan proses transaction identification. Transaksi didefinisikan sebagai urutan web pages yang diakses seorang user pada sebuah session tertentu

3.2 Pattern Discovery

Setelah data melalui tahap preprocessing, data kemudian akan diolah untuk mencari pola penggunaan user. Pada tahap ini dilakukan pembentukan grup cluster terhadap list transaksi yang dihasilkan dari proses transaction identification pada preprocessing. Proses pattern discovery menggunakan algoritma fuzzy c-means. Algoritma fuzzy c-means bekerja dengan cara mengelompokkan transaksi berdasarkan perhitungan derajat keanggotaan masing-masing transaksi terhadap semua pusat cluster. Transaksi yang memiliki derajat keanggotaan terbesar terhadap sebuah cluster, maka transaksi itu akan masuk sebagai anggota cluster tersebut. Perhitungan derajat keanggotaan berdasarkan pada jarak transaksi tersebut dan diolah menggunakan satu formula perhitungan tertentu.



4. Pengujian dan Analisis Sistem

4.1 Mencari Jumlah Cluster Terbaik Sistem

Untuk mencari jumlah cluster terbaik sistem, dilakukan perulangan percobaan untuk mencari nilai PCAES yang terbaik. Pengujian dilakukan secara berurutan dimulai dari cluster dua pengujian pertama hingga tiga belas pengujian ke empat. Tiga belas dijadikan jumlah cluster maksimum karena adanya keterbatasan kemampuan hardware untuk memproses data jika cluster yang digunakan lebih dari tiga belas. Karena sistem menggunakan fungsi random untuk memulai proses, hasil tiap terasi memberikan output yang berbeda-beda.

| Cluster | Percobaan ke- | FCM eror | PCAES |
|---------|---------------|----------|----------|
| 2 | 1 | 12162.64 | 1.910318 |
| 2 | 2 | 11958.23 | 1.797971 |
| 2 | 3 | 12150.78 | 1.90994 |

| | | | |
|-----|-----|----------|----------|
| 3 | 1 | 12216.77 | 1.037803 |
| 3 | 2 | 12217.65 | 1.128195 |
| 3 | 3 | 12217.91 | 1.038972 |
| ... | ... | ... | ... |
| 9 | 1 | 2410.524 | 2.251975 |
| 9 | 2 | 2165.263 | 2.086605 |
| 9 | 3 | 2818.229 | 1.934038 |
| ... | ... | ... | ... |

Dikarenakan hasil yang selalu berubah-ubah tetapi cenderung memiliki nilai yang berdekatan maka diambil keputusan bahwa percobaan ke empat akan dijadikan sumber data yang dipakai untuk menunjukkan hasil dengan asumsi data tersebut dapat mewakili data yang lainnya.

| Cluster | percobaan ke- | Koefisien fuzzy | Error terkecil | max iterasi | FCM Error | PCAES |
|---------|---------------|-----------------|----------------|-------------|-----------|----------|
| 2 | 4 | 1,5 | 0.001 | 50 | 12162.22 | 1.793651 |
| 3 | 4 | 1,5 | 0.001 | 50 | 12217.18 | 1.128204 |
| 4 | 4 | 1,5 | 0.001 | 50 | 5999.585 | 1.223252 |
| 5 | 4 | 1,5 | 0.001 | 50 | 3989.592 | 1.438399 |
| 6 | 4 | 1,5 | 0.001 | 50 | 4118.352 | 1.785297 |
| 7 | 4 | 1,5 | 0.001 | 50 | 3520.164 | 2.35254 |
| 8 | 4 | 1,5 | 0.001 | 50 | 3111.219 | 2.446648 |
| 9 | 4 | 1,5 | 0.001 | 50 | 3049.432 | 0.861592 |
| 10 | 4 | 1,5 | 0.001 | 50 | 2111.286 | 2.356366 |
| 11 | 4 | 1,5 | 0.001 | 50 | 1849.4 | 2.017029 |
| 12 | 4 | 1,5 | 0.001 | 50 | 2359.453 | 0.30804 |
| 13 | 4 | 1,5 | 0.001 | 50 | 2250.733 | -0.02085 |

Dari tabel diatas dapat dilihat bahwa nilai PCAES tertinggi ada ketika jumlah cluster yang dipakai adalah delapan cluster. Hal ini menunjukkan bahwa ketika cluster berjumlah delapan cluster yang terbentuk paling solid dan data didalamnya paling terpisah. Oleh karena itu jumlah cluster delapan lah yang paling optimal untuk kasus penelitian ini.

4.2 Pengaruh parameter terhadap hasil performansi

Untuk menguji pengaruh parameter terhadap hasil, digunakan jumlah cluster terbaik hasil pengujian sebelumnya sebagai tolak ukur hasil percobaan. Semua pengujian dilakukan empat kali dengan pengujian satu sampai tiga sebagai acuan nilai, dan pengujian keempat dianggap pengujian yang mewakili pengujian-pengujian sebelumnya

Koefisien Fuzzy

Pada pengujian ini kita akan melihat pengaruh perbedaan koefisien fuzzy dengan mengubah-ubah nilainya. Ada tiga nilai yang akan diujikan, pertama jika derajat fuzzy sama dengan dua, dua koma lima, dan empat.

| Cluster | derajat fuzziness | Error terkecil | max iterasi | fcm error | PCAES |
|---------|-------------------|----------------|-------------|-----------|----------|
| 8 | 1.5 | 0.001 | 50 | 3111.219 | 2.446648 |
| 8 | 2 | 0.001 | 50 | 8139.568 | 1.465172 |
| 8 | 2.5 | 0.001 | 50 | 30600.29 | -6.99998 |
| 8 | 4 | 0.001 | 50 | 122401.2 | -7 |

Tabel 4.3 pengaruh koefisien fuzzy terhadap performansi

Error Terkecil

Error terkecil merupakan nilai perubahan minimum yang boleh terjadi sebelum algoritma berhenti. Diasumsikan bahwa jika perubahan yang terjadi saat proses perulangan lebih kecil dari error terkecil maka perubahan tersebut dianggap tidak ada dan algoritma akan berhenti

| Cluster | derajat fuzziness | Error terkecil | max iterasi | fcm error | PCAES |
|---------|-------------------|----------------|-------------|-----------|----------|
| 8 | 1.5 | 0.001 | 50 | 3111.219 | 2.446648 |
| 8 | 1.5 | 0.0001 | 50 | 2436.399 | 2.240231 |
| 8 | 1.5 | 0.01 | 50 | 2305.871 | 1.906085 |
| 8 | 1.5 | 0.1 | 50 | 7158.608 | 0.19462 |

Maksimum Iterasi

| Cluster | derajat fuzziness | Error terkecil | max iterasi | fcm error | PCAES |
|---------|-------------------|----------------|-------------|-----------|----------|
| 8 | 1.5 | 0.001 | 50 | 3111.219 | 2.446648 |
| 8 | 1.5 | 0.001 | 5 | 3334.191 | 0.233813 |
| 8 | 1.5 | 0.001 | 10 | 2724.66 | 1.937483 |
| 8 | 1.5 | 0.001 | 15 | 2838.826 | 0.341579 |

Maksimum iterasi merupakan variabel yang berfungsi untuk mencegah terjadinya looping forever pada system. Dengan melihat hasil dapat dilihat bahwa dengan membatasi proses looping sistem maka performansi sistem juga berubah. Hal ini dikarenakan sistem belum selesai memproses semua data tetapi sudah masuk pada proses keluar dari sistem

5 Kesimpulan dan Saran

5.1 Kesimpulan

Berdasarkan hasil dari pengujian yang telah dilakukan maka dapat disimpulkan hal-hal sebagai berikut:

1. Algoritma Fuzzy c-Means dapat diimplementasikan untuk membentuk cluster berdasarkan data transaksi hasil preprocessing dimana jumlah cluster optimal yang dihasilkan adalah delapan cluster dengan nilai validitas PCAES 2.446648 dan

nilai objektif FCM sebesar 3111.219. saat percobaan nilai parameter fuzzy lainnya saat percobaan adalah koefisien fuzzy sebesar 1.5, error terkecil 0.001, dan maksimum iterasi 50. Nilai PCAES 2.446648 adalah nilai paling besar diantara percobaan-percobaan yang lain mengindikasikan bahwa cluster

yang dibentuk memiliki kepadatan yang baik dan jarak antara satu cluster dengan cluster lainnya tidak berdekatan.

2. Setiap parameter dari algoritma fuzzy c-means memiliki pengaruh masing-masing terhadap performansi algoritma. Koefisien fuzzy yang paling optimal untuk studi kasus ini adalah 1.5 dimana semakin besar koefisien fuzzy nya semakin rendah performansi algoritmanya. Error terkecil dan maksimum iterasi mempengaruhi performansi dikarenakan kedua faktor itu membatasi looping yang dibutuhkan algoritma untuk mencapai hasil yang optimal.

5.2 Saran

Berdasarkan apa yang sudah dilakukan pada Tugas akhir ini, mulai dari pemilihan data hingga analisis hasil pengujian, saran yang bisa disampaikan adalah

1. Untuk pembuatan algoritma fuzzy kedepannya dapat menggunakan metoda SSE untuk inisiasi cluster agar titik centroid awal yang dibentuk mempunyai distance yang saling berjauhan satu sama lain.
2. Untuk penelitian selanjutnya sebaiknya menggunakan data yang lebih variatif

Daftar Pustaka

- [1] A. Luotonen, The Common Logfile Format, 1995, <http://www.w3.org/pub/WWW/Daemon/User/Config/Logging.html>. Diakses : Desember 2013.
- [2] Beringer, Jürgen, *Adaptive Optimization of the Number of Clusters in Fuzzy Clustering*. Otto-von-Guericke University Faculty of Informatics:Germany
- [3] Bezdek, James C., Robert Ehrlich, William Full.1984. *FCM: The Fuzzy c-means Clustering Algorithm*, Pergamon Press Ltd: U.S.A.
- [4] Chen, Chi-Farn, Jyh-Ming Lee. *The Validity Measurement of Fuzzy c-Means Classifier for Remotely Sensed Images*.National Central University Centre for space and remote sensing Research:Taiwan.
- [5] Chitraa, V, Dr. Anthony Selvadoss Davamani. *An Efficient Path Completion Technique for Web Log Mining*. IEEE International Conference on Computational Intelligence and Computing Research. 2010.
- [6] Chitraa, V, Dr.Antony Selvadoss Davamani. *A Survey on Preprocessing Methods for Web Usage Data*. International Journal of Computer Science and information Security, Vol. 7, No. 3, 2010.
- [7] Cooley, Robert, B. Mobasher, J. Srivastava, *Data preparation for mining World Wide Web browsing patterns*, Journal of Knowledge and Information Systems 1 (1) (1999) 5 – 32.
- [8] Etzioni, Orem. *The Wold Wide Web: Quagmire or gold mine?*. Comm. Of ACM. 1996
- [9] Han, Seungyeop, Yong-Yeol Ahn, Sue Moon, and Hawoong Jeong. *Collaborative blog spam ltering using adaptive percolation search*. In *Workshop on the Weblogging Ecosystem*, 2006.
- [10] Lakheyana, Chhama, Usvir Kaur. *A survey on Web Usage Mining with Fuzzy c-Means Clustering Algorithm*. IJCSMC, vol 2, issue. 4, April 2013, pg. 160-163.
- [11] Lingras, Pawan, Rui Yan, Chad West.2003. *Fuzzy c-Means Clustering of Web Users for Educational Sites*.Springer-Verlag:Berlin Hiedelberg
- [12] Madria, Sanjay Kumar, Sourav S Bhowmick, W-K Ng, Ee-Peng Lim. *Data Warehousing and Knowledge Discovery*. 1999. Springer Berlin Hiedelberg.
- [13] Prasetyo, Eko.2012. *Data Mining Konsep Dan Aplikasi menggunakan Matlab*. Penerbit Andi.Yogyakarta.
- [14] Wu, Kuo-Lung, Miin-Shen Yang.2004. *A Cluster Validity Index for Fuzzy Clustering*.Elsevier B.V:Taiwan
- [15] Wu, Kuo-Lung.2010. *Analysis of Parameter Selection for Fuzzy c-Means*. Kun Shan University, Department of Information Management:Taiwan.
- [16] Yaikhom, Gagarine.2010. *Implementing the Fuzzy c-Means Algorithm*, Stamford, Faculty of Computer Science.