

## **Analisa Web Usage Mining dengan Algoritma Particle Swarm Optimization (PSO) (Studi Kasus : Aktifitas Internet di Universitas Telkom)**

Arby Fahrizi Subiyanto<sup>1</sup>, Eko Darwiyanto<sup>2</sup>, Rita Rismala<sup>3</sup>

<sup>1,2,3</sup> *Teknik Informatika, Fakultas Informatika, Telkom University*

*Jalan Telekomunikasi No.1, Dayeuh Kolot, Bandung 40257*

[fahrizi@gmail.com](mailto:fahrizi@gmail.com)<sup>1</sup>, [ekodarwiyanto@telkomuniversity.ac.id](mailto:ekodarwiyanto@telkomuniversity.ac.id)<sup>2</sup>, [ritaris.staff@telkomuniversity.ac.id](mailto:ritaris.staff@telkomuniversity.ac.id)<sup>3</sup>

---

### **Abstrak**

Website BAA Universitas Telkom dipilih karena merupakan salah satu *website* di Universitas Telkom yang cukup sering diakses oleh mahasiswa untuk mencari informasi tentang panduan akademik dan berbagai kebutuhan informasi lainnya tentang perkuliahan dan wisuda. Pola pengunjung *web* dapat digunakan untuk mengetahui halaman apa saja yang telah dikunjungi oleh *user* dalam suatu *website*. Hal tersebut bertujuan sebagai acuan dalam perbaikan kualitas *website* dan menjamin kepuasan *user* dalam mengakses *website* tersebut dalam menemukan informasi didalamnya sesuai dengan kebutuhannya. Dalam tugas akhir ini, digunakan *web server log* dari BAA Universitas Telkom. *Web server log* dari BAA Universitas Telkom kemudian akan diproses dengan mengimplementasikan salah satu metode pada *web usage mining* yaitu *clustering*. *Data log* tersebut awalnya diolah dengan tahap *preprocessing*, kemudian dilakukan proses *clustering* dengan menggunakan algoritma *Particle Swarm Optimization* (PSO). Algoritma PSO digunakan karena memiliki kelebihan yaitu memiliki *swarm* yang dapat mencari solusi terbaik dalam penentuan hasil *cluster*. Penggunaan algoritma PSO pada penelitian ini ditemukan hasil *cluster* terbaik adalah delapan dan dari hasil *cluster* tersebut dihasilkan rekomendasi terhadap *website* BAA Universitas Telkom.

**Kata Kunci:** *web usage mining, web log, clustering, preprocessing, swarm, particle swarm optimization.*

---

### **Abstract**

BAA Telkom University website chosen because it is one of the website at the University of Telkom which quite often accessed by students to search for information about academic guidelines and various other information needs lectures and graduation. Web visitor pattern can be used to determine which pages have been visited by the user in a website. It is intended as a reference in the improvement of the quality of the website and ensure user satisfaction in accessing the website to find information in it according to his needs. In this thesis, use web server logs of BAA Telkom University. In addition the website is used because the author has conducted an analysis and found that the appropriate requirements to serve as a case study of this thesis. Web server logs of BAA Telkom University will then be processed by implementing one of the methods on web usage mining namely clustering. The log data is initially processed by preprocessing stage, then the clustering process is carried out using Particle Swarm Optimization algorithm (PSO). The PSO algorithm is used because it has the advantage of having *swarm* to find the best solution in determining the results of the cluster. In this case PSO algorithm was found the best result is eight cluster and that cluster results can generated recommendations for the BAA website Telkom University.

**Keywords :** *web usage mining, web log, clustering, preprocessing, swarm, particle swarm optimization.*

---

## 1. Pendahuluan

Saat ini internet menjadi suatu kebutuhan yang penting bagi seseorang, dengan kebutuhan tersebut internet juga semakin berkembang mengikuti perkembangan zaman. Hal ini membuat para pengembang web berusaha untuk meningkatkan performa dan kualitas *web* masing – masing sehingga pengembang perlu untuk menggali *knowledge* yang ada dari aktifitas user dalam mengakses website. Aplikasi web yang digunakan cenderung bertujuan untuk penyebaran dan penerimaan informasi sehingga memberikan kemudahan bagi pengguna (*user*) yang membutuhkannya. Oleh sebab itu muncul sebuah ilmu yang bernama *Web Mining*. *Web mining* adalah sebuah ilmu yang mempelajari penggalian data yang berhubungan dengan *web*. Pada *web mining* terdapat tiga kategori, yaitu *Web Content Mining*, *Web Structure Mining* dan *Web Usage Mining*. Pada *Web Content Mining* dan *Web Structured Mining* proses miningnya mencari informasi dari data yang tersedia secara *on-line*, sedangkan pada *Web Usage Mining* data yang akan diolah dapat diambil secara *off-line*. Data yang diambil tersebut berupa data log akses user yang tersimpan dalam *web log server*. Pengolahan data yang dilakukan secara *off-line* akan lebih mudah dikarenakan dalam pelaksanaan proses mining tidak membutuhkan koneksi internet. Dalam kasus ini pengelompokan berdasarkan pada web log dari *website* yang ada di Telkom University, yaitu BAA Telkom University. Website yang tersebut *www.baa.telkomuniversity* adalah salah satu web yang cukup sering diakses di Universitas Telkom, hal tersebut telah dibuktikan dengan melakukan survey ke SISFO Universitas Telkom. Dengan dasar tersebut perlu dilakukan analisa untuk perkembangan dan dapat dijadikan bahan evaluasi terhadap website tersebut. Metode yang diterapkan pada kasus ini adalah *clustering*, *clustering* digunakan untuk mengetahui pola ketertarikan pengunjung *web*. Untuk mendukung metode tersebut digunakan algoritma *Particle Swarm Optimization* (PSO). Algoritma ini bekerja dengan mempertimbangkan halaman *web* yang paling sering dikunjungi dan waktu kunjungannya. Algoritma PSO dipilih karena memiliki *swarm* yang dapat mencari solusi terbaik dengan menghitung nilai *fitness* sehingga nantinya dapat dihasilkan *cluster* yang baik dibandingkan dengan algoritma *clustering* yang lain.

## 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. Data dan informasi yang tersimpan di dalam *web* memiliki karakteristik yang berbeda dengan data yang tersimpan dalam penyimpanan konvensional seperti DBMS.

### 2.2 Web Mining

Menurut Oren [9], 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 [13] *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 Algoritma K – Means

Algoritma *K-Means* yang digunakan pada penelitian ini menggunakan perhitungan *Symmetric distance* untuk mengukur jarak antar data. *Symmetric* dipilih karena data sesi yang berbentuk binary.

1. Inisiasi *centroid* awal menggunakan inisiasi *K-Means*.
2. Tetapkan semua *data point* (yang merepresentasikan transaksi) ke *cluster* terdekat (diukur dari *centroid cluster*). Dengan cara menyajikan *data point x* dan hitung kemiripan(jarak) *d* dari input ke setiap pusat *cluster j* (*centroid*). Masukkan data ke dalam cluster yang pusat *cluster*-nya berjarak minimum ke *data point x*.  

$$\leftarrow \text{argmin}_{j \in \{1, 2, \dots, K\}} (d(x, c_j))$$
3. Hitung kembali pusat dari setiap *cluster* sebagai *centroid* dari semua *data point* di setiap *cluster*. *Centroid* baru dihitung dengan menggunakan pendekatan yang diproposes oleh Grana [10]:  
 Jika tiap bit vektor data transaksi adalah  $\{1, 2, 3, \dots, n\}$   
 Penentuan dilakukan untuk tiap bit transaksi, jika mayoritas bit  $v_i$  pada vektor didalam  $C_d$  adalah “1” maka bit  $v_i$  untuk  $c$  baru di-assign “1”, begitu juga sebaliknya untuk “0”.
4. Ulangi proses 2-3 sampai tidak ada *centroid* yang berubah.

#### 4.5 Algoritma Particle Swarm Optimization (PSO)

Particle Swarm Optimization adalah salah satu metode optimasi yang terinspirasi dari perilaku gerakan kawanan hewan seperti ikan (*school of fish*), hewan herbivor (*herd*), dan burung (*flock*) yang kemudian tiap objek hewan disederhanakan menjadi sebuah partikel. Suatu partikel dalam ruang memiliki posisi yang dikodekan sebagai vektor koordinat. Vektor posisi ini dianggap sebagai keadaan yang sedang ditempati oleh suatu partikel di ruang pencarian. Setiap posisi dalam ruang pencarian merupakan alternatif solusi yang dapat dievaluasi menggunakan fungsi objektif. Setiap partikel bergerak dengan kecepatan  $v$ .

Particle Swarm Optimization atau yang kita dikenal dengan PSO menerapkan sifat masing-masing individu dalam satu kelompok besar. Kemudian menggabungkan sifat-sifat tersebut untuk menyelesaikan permasalahan. Particle Swarm Optimization pertama kali dimunculkan pada tahun 1995, sejak saat itulah para peneliti banyak menurunkan dan mengembangkan metode PSO.

Ciri khas dari PSO adalah pengaturan kecepatan partikel secara heuristik dan probabilistik. Jika suatu partikel memiliki kecepatan yang konstan maka jika jejak posisi suatu partikel divisualisasikan akan membentuk garis lurus. Dengan adanya faktor eksternal yang membelokkan garis tersebut yang kemudian menggerakkan partikel dalam ruang pencarian maka diharapkan partikel dapat mengarah, mendekati, dan pada akhirnya mencapai titik optimal. Faktor eksternal yang dimaksud antara lain posisi terbaik yang pernah dikunjungi suatu partikel, posisi terbaik seluruh partikel (diasumsikan setiap partikel mengetahui posisi terbaik setiap partikel lainnya), serta faktor kreativitas untuk melakukan eksplorasi.

Particle Swarm Optimization memiliki kesamaan sifat dengan teknik komputasi seperti Algoritma Genetika (*Genetic Algorithm*). Sistem PSO diinisialisasi oleh sebuah populasi solusi secara acak dan selanjutnya mencari titik optimum dengan cara meng-update tiap hasil pembangkitan. Metode optimasi yang didasarkan pada swarm intelligence ini disebut algoritma behaviorally inspired sebagai alternatif dari algoritma genetika, yang sering disebut evolution-based procedures. Dalam konteks optimasi multivariabel, kawanan diasumsikan mempunyai ukuran tertentu atau tetap dengan setiap partikel posisi awalnya terletak di suatu lokasi yang acak dalam ruang multidimensi. Setiap partikel diasumsikan memiliki dua karakteristik: posisi dan kecepatan. Setiap partikel bergerak dalam ruang/space tertentu dan mengingat posisi terbaik yang pernah dilalui atau ditemukan terhadap sumber

makanan atau nilai fungsi objektif. Setiap partikel menyampaikan informasi atau posisi bagusnya kepada partikel yang lain dan menyesuaikan posisi dan kecepatan masing-masing berdasarkan informasi yang diterima mengenai posisi yang bagus tersebut. Sebagai contoh, misalnya perilaku burung-burung dalam kawanan burung. Meskipun setiap burung mempunyai keterbatasan dalam hal kecerdasan, biasanya ia akan mengikuti kebiasaan (rule) seperti berikut :

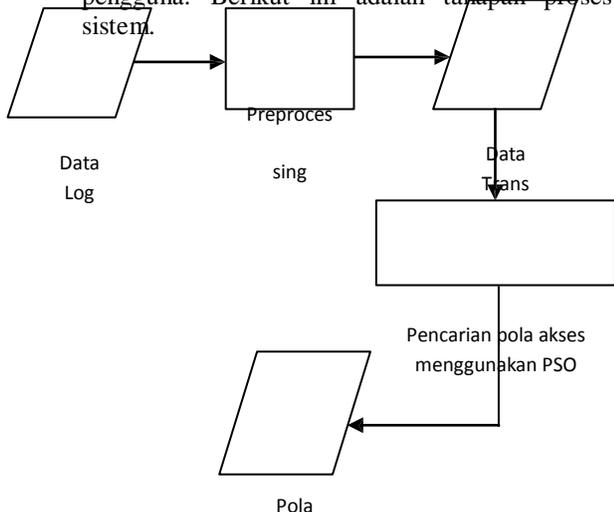
1. Seekor burung tidak berada terlalu dekat dengan burung yang lain
2. Burung tersebut akan mengarahkan terbangnya ke arah rata-rata keseluruhan burung
3. Akan memposisikan diri dengan rata-rata posisi burung yang lain dengan menjaga sehingga jarak antar burung dalam kawanan itu tidak terlalu jauh

Dengan demikian perilaku kawanan burung akan didasarkan pada kombinasi dari 3 faktor simpel berikut:

- 5 1. Kohesi - terbang bersama
- 6 2. Separasi - jangan terlalu dekat
- 7 3. Penyesuaian(alignment) - mengikuti arah bersama.

**8 Perancangan Sistem**

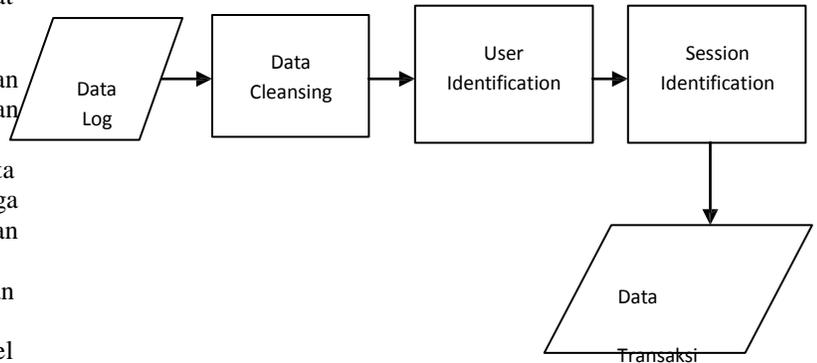
Dalam Tugas akhir ini terdapat dua proses yang dijalankan. Proses yang pertama adalah preprocessing dan yang kedua adalah Pattern Discovery. Jadi data Raw Log pertama akan masuk ke dalam proses preprocessing. Proses tersebut akan menghasilkan data transaksi untuk kemudian dijadikan input untuk proses selanjutnya, yaitu Pattern discovery. Pada proses ini data akan diolah dan dikelompokkan kedalam cluster sesuai dengan kemiripan datanya. Untuk kemudian hasil dari proses ini adalah akses pengguna. Berikut ini adalah tahapan proses sistem.



Gambar 3.1 Diagram Balok

**3.1 Preprocessing**

Preprocessing adalah sebuah proses untuk mempersiapkan data agar data bisa diolah lebih lanjut oleh proses selanjutnya. Berikut ini adalah langkah dalam preprocessing.



Gambar 3.2 Proses Preprocessing

1. Proses data cleansing, bertujuan untuk membersihkan data dari bagian – bagian yang tidak diperlukan sehingga yang tersisa hanya bagian data yang diperlukan. Dalam proses ini digunakan fungsi substring yang ada dengan mengelompokkan baris-baris data menjadi beberapa bagian sebelumnya.

```
[39.213.217.138 -- [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 3.3 Baris Log Sebelum Preprocessing

Pada gambar 3.3 merupakan contoh baris log sebelum dilakukan preprocessing, untuk kemudian sistem akan memprosesnya menjadi seperti data berikut.

```
[39.213.217.138][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 3.4 Contoh baris log setelah proses formatting

Setelah melalui proses formatting, proses selanjutnya adalah membersihkan baris log dari bagian yang tidak perlu seperti data berekstensi .jpg, .gif, ukuran

akses

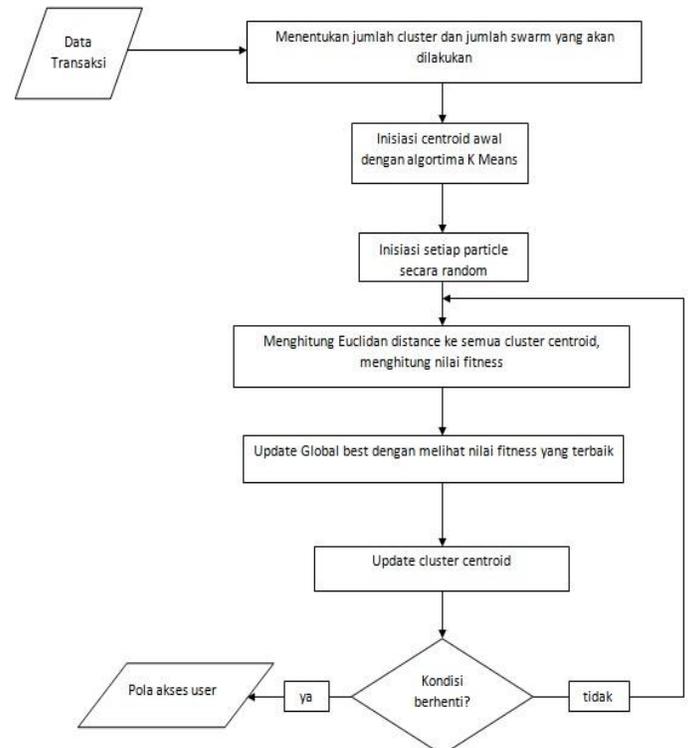
4

byte dan status. Langkah selanjutnya adalah menghilangkan log yang tidak berhubungan dengan penelitian. Ada tiga kategori log yang akan dihapus berdasarkan (Chitraa, 2010). Pertama adalah log yang didalamnya terdapat IP address yang mengacu ke server. Kedua adalah log yang menghasilkan error, 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.

2. User Identification merupakan proses mengidentifikasi unique user, pada penelitian ini IP address digunakan untuk menjadi identitas user.
3. User Session Identification adalah proses mengidentifikasi aktivitas transaksi pada selang waktu tertentu. Jangka waktu yang ditentukan adalah 30 menit. Jika selang waktu request melebihi jangka waktu tersebut maka akan dianggap sebagai session yang berbeda atau disebut masuk ke 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 proses preprocessing. Proses pattern discovery menggunakan algoritma K-Means untuk menentukan posisi awal inisialisasi swarm untuk kemudian dijalankan dengan menggunakan algoritma PSO. Algoritma PSO digunakan untuk mengoptimasi pusat cluster sehingga didapatkan hasil terbaik dari cluster tersebut dengan penghitungan nilai fitness. Nilai fitness terbaik akan menjadi solusi untuk menentukan cluster terbaik.



Gambar 3.5 Pencarian Pola Akses

Pertama, sistem harus diberi input berupa jumlah cluster dan jumlah swarm. Selanjutnya inisialisasi centroid awal dilakukan dengan mengambil sampel data secara acak sebanyak jumlah cluster yang ditentukan. Kemudian untuk mendapatkan cluster awal digunakan algoritma K - Means. Setelah cluster awal didapatkan dengan menggunakan algoritma K - Means, kemudian akan dimasukkan inisiasi awal swarm yang kemudian swarm akan mencari solusi terbaik dan cluster terbaik dengan menghitung euclidan distance dan menghitung nilai fitness terbaik. Setelah didapatkan nilai fitness terbaik global best akan diupdate dan jika fitness terbaik yang didapatkan tidak lebih baik dari nilai fitness terbaik nilai global best dan cluster centroid tidak akan diupdate. Kemudian cek kondisi berhenti, apabila belum memenuhi kondisi berhenti maka kembali lakukan perhitungan nilai fitness. Ada dua kriteria yang memenuhi kondisi berhenti. Pertama adalah jika sudah tidak ada perubahan pada nilai fitness terbaik atau tidak ditemukan lagi nilai fitness yang lebih baik dari sebelumnya. Kedua adalah jika maksimum iterasi telah tercapai. Setelah kondisi berhenti tercapai maka didapatkan lah pola akses user.

#### 4. Pengujian dan Analisis Sistem

##### 4.1 Mencari Jumlah Cluster Terbaik dan Jumlah Swarm Terbaik Sistem

Untuk mencari jumlah *cluster* terbaik dan jumlah *swarm* terbaik sistem, dilakukan perulangan dengan merubah jumlah *cluster* dan jumlah *swarm* kemudian dilihat nilai *fitness* yang didapatkan. Pengujian dilakukan secara berurutan dimulai dari dua *cluster* hingga 20 *cluster* dan jumlah *swarm* yang Dimulai dari dua *swarm* hingga sembilan *swarm*. Semua ajan diuji satu per satu secara berurutan dimulai dari dua *cluster* dengan dua *swarm*, kemudian dua *cluster* dengan tiga *swarm* dan seterusnya hingga 20 *cluster* dengan sembilan *swarm*. Hasil pengujian tersebut dapat dilihat di tabel dibawah ini :

Tabel 4.1 Hasil Pengujian Cluster dan Swarm

Swarm \ Cluster	2	3	4	5	6	7	8	9
2	24	14	17	22	15	15	13	12
3	44	68	44	26	33	19	17	15
4	48	37	34	23	39	19	21	17
5	78	76	77	60	59	51	33	32
6	220	213	160	110	98	101	105	92
7	222	224	174	193	163	141	112	94
8	199	155	151	152	119	99	107	90
9	377	306	225	265	186	196	173	119
10	414	386	352	361	347	289	265	231
11	433	408	382	368	316	303	292	287
12	505	476	452	468	431	419	404	392
13	566	514	487	461	458	447	432	429
14	608	594	572	583	569	541	519	483
15	739	723	706	718	704	682	663	640
16	800	788	767	754	766	742	729	713
17	1099	1087	1069	1058	1048	1051	1036	1001
18	1132	1116	1095	1098	1074	1062	1048	1037
19	1509	1464	1429	1396	1387	1329	1255	1207
20	1819	1702	1622	1544	1468	1329	1287	1216

Pada tabel 4.1 menunjukkan hasil pengujian jumlah *cluster* dan jumlah *swarm*. Dari tabel tersebut dapat dilihat bahwa semakin banyak jumlah *cluster* maka nilai *fitness* terbaik yang dihasilkan menjadi semakin besar pula, sedangkan pada jumlah *swarm* yang semakin banyak maka nilai *fitness* yang dihasilkan menjadi semakin kecil. Nilai *fitness* terbaik pada PSO adalah nilai *fitness* yang memiliki nilai kecil. Analisis pertama dilakukan pada jumlah *cluster*. Pada *cluster* ke delapan (berwarna merah) dapat dilihat nilai *fitness* terbaik yang dihasilkan memiliki nilai kecil (*fitness* terbaik) dibandingkan dengan jumlah *cluster* lain yang semakin besar dan diikuti oleh nilai *fitness* terbaiknya yang besar pula. Sehingga dapat diketahui jika *cluster* optimum yang dihasilkan adalah delapan. Hasil *cluster* dua, tiga, empat dan lima tidak dipilih meskipun memiliki nilai *fitness* yang lebih baik karena dengan jumlah *cluster* tersebut memiliki jumlah yang lebih sedikit dibandingkan *tab* utamanya yaitu sepuluh, maka *url* yang ada didalam *cluster* tersebut nantinya masih terlalu banyak dan sulit untuk dianalisis. Oleh sebab itu *cluster* delapan dipilih karena memiliki jumlah *cluster* yang banyak dan memiliki pendekatan yang lebih baik dengan jumlah *tab* utama. *Tab* utama pada *web* BAA Universitas Telkom berjumlah sepuluh. Sehingga dengan delapan *cluster* dapat mencakup jumlah *tab* tersebut untuk kemudian dikelompokkan tiap *url* yang sering dikunjungi oleh user. Selain itu *cluster* delapan dipilih karena memiliki nilai *fitness* best yang baik dibandingkan hasil *cluster* besar yang lain seperti tujuh, sembilan dan seterusnya hingga 20 *cluster*. Kemudian analisis yang kedua adalah jumlah *swarm*, dari tabel tersebut dapat diketahui hasil dari pengujian dengan merubah parameter jumlah *swarm*. Dengan menggunakan jumlah *swarm* yang lebih banyak yaitu sembilan dapat menghasilkan nilai *fitness* terbaik yang semakin kecil. Sehingga dengan jumlah *swarm* berjumlah sembilan memberikan nilai *fitness* terbaik sistem dan merupakan jumlah *swarm* optimum untuk sistem. Jumlah *swarm* berpengaruh terhadap hasil nilai *fitness* yang didapatkan, karena semakin banyak *swarm* peluang untuk mendapatkan solusi maksimal akan semakin besar, karena pada PSO *swarm* adalah sekumpulan nilai yang nantinya akan menjadi sebuah solusi. Hasil nilai *fitness* terbaik masing – masing *cluster* didapatkan dengan jumlah *cluster* maksimum, yaitu sembilan.

## 5. Kesimpulan dan Saran

### 5.1 Kesimpulan

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

1. Algoritma *Particle Swarm Optimization* dapat diimplementasikan untuk membentuk *cluster* berdasarkan data transaksi hasil *preprocessing* untuk studi kasus penelitian ini.
2. Dari hasil pengujian yang dilakukan yaitu perubahan jumlah *cluster* dan melihat nilai *fitness* terbaiknya, jumlah *cluster* optimal yang dihasilkan adalah delapan *cluster*.
3. Parameter lain dari pengujian algoritma *Particle Swarm Optimization* ini adalah jumlah *swarm*, dimana diperoleh jumlah *swarm* terbaik pada pengujian yang telah dilakukan adalah sembilan *swarm*.

### 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. Dari hasil penelitian ini, hasil rekomendasi dapat digunakan oleh SISFO Universitas Telkom sebagai bahan pertimbangan untuk pengembangan *web* baa.telkomuniversity.
2. Untuk penelitian selanjutnya dapat digunakan data *log* dari *website* lain yang ada di Universitas Telkom atau *website* selain yang ada di Universitas Telkom yang dibutuhkan untuk dapat dianalisa dan dijadikan bahan pertimbangan untuk pengembangan *web* tersebut.

## Daftar Pustaka

- [1] Aini, Fithratul. (2011). *Web Usage Mining Menggunakan Algoritma Adaptive Web Access Pattern Tree (AWAPT)*.
- [2] Alam, Shafiq, Dobbie, Gillian, Riddle Patricia. (2008). *Particle Swarm Optimization Based Clustering Of Web Usage Data*.
- [3] Arsih, Kansul. (2012). *Perancangan dan Implementasi Aplikasi Analisis Log Menggunakan Metode JST Adaptive Resonance Theory 2 Dalam Memprediksi Tingkah Laku Pengguna Internet (Studi Kasus : Aktifitas Internet di Institut Teknologi Telkom)*.
- [4] Bimarsana, Stefan. (2014). *Analisis dan Implementasi Algoritma Fuzzy C-Means Untuk Web Usage Mining (Studi Kasus : Aktifitas Internet Telkom University)*.
- [5] 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.
- [6] 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.
- [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] D. W. Van der Merwe, and A. P. Engelbrecht (2003). *Data Clustering using Particle Swarm Optimization*.
- [9] Etzioni, Oren. *The World Wide Web: Quagmire or gold mine?*. Comm. Of ACM. 1996.
- [10] Grana, Constantino, Daniele B, Rita C, *A Fast Approach For Integrating ORB Descriptor in the Bag Of Words Model*, Dipartimento di Ingegneria "Enzo Ferrari"- Universita degli Studi di Modena e Reggio Emilia.
- [11] 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.
- [12] J. Kennedy, and R. C. Eberhart. (1995). *Particle Swarm Optimization*.

- [13] Madria, Sanjay Kumar, Sourav S Bhowmick, W-K Ng, Ee-Peng Lim. *Data Warehousing and Knowledge Discovery*. 1999. Springer Berlin Hiedelberg.
- [14] Maryowati, Damnaseka. (2011). *Analisis dan Implementasi Web Usage Mining dengan menggunakan Algoritma C5.0 (Study Kasus pada Web Student IT Telkom)*.
- [15] Prasetyo, Eko.2012.*Data Mining Konsep Dan Aplikasi menggunakan Matlab*. Penerbit Andi.Yogyakarta.
- [16] S.C. M. Cohen, and L. N. de Castro. (2006). *Data Clustering with Particle Swarms*.
- [17] Sudhamathy, G., C. Jothi V. (2011). *Web Log Clustering Approaches – A Survey*.