

## Algoritma *Partitioning Around Medoids (PAM) Clustering* untuk Melihat Gambaran Umum Skripsi Mahasiswa

Herny Februariyanti<sup>1</sup>, Dwi Budi Santoso<sup>2</sup>

<sup>1,2</sup>Program Studi Sistem Informasi, Fakultas Teknologi Informasi, Universitas Stikubank  
Email: <sup>1</sup>hernyfeb@gmail.com, <sup>2</sup>nayantaka@gmail.com

### Abstrak

Informatika dan Program Studi Sistem Informasi Fakultas Teknologi Informasi Universitas Stikubank Semarang. Fakultas Teknologi Informasi dipilih karena jumlah data penelitiannya cukup banyak. Setiap tahunnya Fakultas Teknologi Informasi meluluskan mahasiswa dengan penelitian skripsi yang beragam. Setiap tahunnya jumlah data skripsi selalu bertambah. Semakin bertambahnya penelitian skripsi dengan mata kuliah terbatas menyebabkan semakin banyak pula mahasiswa yang mengambil penelitian yang mirip tema, objek, atau metode penelitian dengan penelitian sebelumnya. Pengelompokan data penelitian yang umumnya berbentuk teks dapat dilakukan dengan text mining.

Algoritma yang digunakan adalah *Partitioning Around Medoids (PAM) Clustering* untuk melihat gambaran umum kemampuan akademik mahasiswa. Aplikasi dapat memberikan layanan pemilihan medoids sejumlah 3 dokumen sebagai objek medoid. Untuk menghitung similaritas antar dokumen yang diuji dengan medoids yang telah ditentukan menggunakan algoritma Jaccard Coefficient.

Aplikasi dapat melakukan proses menentukan sejumlah 3 (tiga) medoids oleh user dari dokumen abstrak skripsi mahasiswa.. Aplikasi juga dapat melakukan perhitungan jarak antara dokumen abstrak skripsi yang dipilih sebagai medoids dengan dokumen diluar medoid dan menampilkan hasil klastering dokumen abstrak skripsi menggunakan algoritma PAM

**Kata kunci** : *Algoritma Partitioing Around Meoids (PAM), Jaccard Coefficient*

### PENDAHULUAN

Data penelitian dapat dikelompokkan berdasarkan kemiripan tema, objek maupun metode penelitian. Hasil pengelompokan data penelitian dapat memperlihatkan bagaimana pola kemiripan penelitian dari waktu ke waktu. Hasil pengelompokan dapat menunjukkan kapan waktu penelitian mahasiswa banyak mengambil materi yang sama dan kapan waktu penelitian mahasiswa beragam. Hasil pengelompokan juga dapat memperlihatkan materi yang banyak diambil mahasiswa dan yang jarang diambil mahasiswa pada waktu tertentu. Informasi tersebut diharapkan dapat membantu dosen dalam mengevaluasi metode pembelajaran yang telah dilakukan pada materi yang banyak ataupun sedikit diambil sebagai bahan penelitian mahasiswa. Penelitian ini mengelompokkan dokumen skripsi Program Studi Teknik Informatika dan Program Studi Sistem Informasi Fakultas Teknologi Informasi Universitas Stikubank Semarang. Fakultas Teknologi Informasi dipilih karena jumlah data penelitiannya cukup banyak. Setiap tahunnya

Fakultas Teknologi Informasi meluluskan mahasiswa dengan penelitian skripsi yang beragam. Setiap tahunnya jumlah data skripsi selalu bertambah. Semakin bertambahnya penelitian skripsi dengan mata kuliah terbatas menyebabkan semakin banyak pula mahasiswa yang mengambil penelitian yang mirip tema, objek, atau metode penelitian dengan penelitian sebelumnya. Pengelompokan data penelitian yang umumnya berbentuk teks dapat dilakukan dengan text mining. Tujuan dari text mining adalah untuk mencari kata-kata yang dapat mewakili isi dari dokumen sehingga dapat dilakukan analisa keterhubungan antar dokumen (Han, J., Kamber, M, 2006).

Tujuan yang ingin dicapai dalam penelitian ini adalah merancang dan melakukan implementasi algoritma *Partitioning Around Medoids (PAM) Clustering* untuk Melihat Gambaran Umum Kemampuan Akademik Mahasiswa.

Adapun manfaat yang diharapkan dari penelitian ini adalah dapat mengetahui trend judul skripsi dan mencegah plagiat dalam

pembuatan skripsi

## METODE PENELITIAN

### Teks Mining

Text mining, yang juga disebut sebagai Teks Data Mining (TDM) atau *Knowledge Discovery in Text* (KDT), secara umum engacu ada proses ekstraksi informasi dari dokumen-dokumen teks tak terstruktur (unstructured). Text mining dapat didefinisikan sebagai penemuan informasi baru dan tidak diketahui sebelumnya oleh komputer, yang secara otomatis mengekstrak informasi dari sumber - sumber teks tak terstruktur yang berbeda. Kunci dari proses ini adalah menggabungkan informasi yang berhasil diekstraksi dari berbagai sumber (Tan, 1999).

Tujuan utama text mining adalah mendukung proses *knowledge discovery* pada koleksi dokumen yang besar. Pada prinsipnya, text mining adalah bidang ilmu multidisipliner, melibatkan information retrieval (IR), text analysis, *information extraction* (IE), *clustering*, *categorization*, *visualization*, *database technology*, *natural language processing* (NLP), *machinelearning*, dan data mining. Dapat pula dikatakan bahwa text mining merupakan salah satu bentuk aplikasi kecerdasan buatan (*artificial intelligence / AI*)

Text mining mencoba memecahkan masalah information overload dengan menggunakan teknik- teknik dari bidang ilmu yang terkait. Text mining dapat dipandang sebagai suatu perluasan dari data mining atau *knowledge - discovery in database* (KDD), yang mencoba untuk menemukan pola-pola menarik dari basis data berskala besar. Namun text mining memiliki potensi komersial yang lebih tinggi dibandingkan dengan data mining, karena kebanyakan format alami dari penyimpanan informasi adalah berupa teks. Text mining menggunakan informasi teks tak terstruktur dan mengujinya dalam upaya mengungkap struktur dan arti yang tersembunyi di dalam teks. Perbedaan mendasar antara text mining dan data mining terletak pada sumber data yang digunakan. Pada data mining, pola-pola diekstrak dari basis data yang terstruktur, sedangkan di text mining, pola -pola diekstrak

dari data tekstual (*natural language*). Secara umum, basis data didesain untuk program dengan tujuan melakukan pemrosesan secara otomatis, sedangkan teks ditulis untuk dibaca langsung oleh manusia (Hearst, 2003).

### Term Weighing

Pembobotan kata (term weighting) adalah proses pembobotan pada kata. Pembobotan dasar dilakukan dengan menghitung frekuensi kemunculan term dalam dokumen. Frekuensi kemunculan (term frequency) merupakan petunjuk sejauh mana term tersebut mewakili isi dokumen. Semakin besar kemunculan suatu term dalam dokumen akan memberikan nilai kesesuaian yang semakin besar.

Faktor lain yang diperhatikan dalam pemberian bobot adalah kejarangmunculan kata (term scarcity) dalam koleksi. Kata yang muncul pada sedikit dokumen harus dipandang sebagai kata yang lebih penting (uncommon tems) daripada kata yang muncul pada banyak dokumen. Pembobotan akan memperhitungkan faktor kebalikan frekuensi dokumen yang mengandung suatu kata (inverse document frequency) (Mandala dan Setiawan, 2002). Faktor terakhirnya adalah faktor normalisasi terhadap panjang dokumen. Dokumen dalam suatu koleksi memiliki karakteristik panjang yang beragam.

Metode TF/IDF merupakan suatu cara untuk memberikan bobot hubungan suatu kata (term) terhadap dokumen. Metode ini menggabungkan dua konsep untuk perhitungan bobot yaitu, frekuensi kemunculan sebuah kata di dalam sebuah dokumen tertentu yang disebut Term Frequency (TF) dan inverse frekuensi dokumen yang mengandung kata yang disebut Inverse Document Frequency (IDF). Frekuensi kemunculan kata di dalam dokumen yang diberikan menunjukkan seberapa penting kata tersebut di dalam dokumen. Sehingga bobot hubungan antara sebuah kata dan sebuah dokumen akan tinggi apabila frekuensi kata tinggi di dalam dokumen dan frekuensi keseluruhan dokumen yang mengandung kata tersebut akan rendah pada kumpulan dokumen. Rumus umum untuk Tf -Idf :

$$W_{dt} = TF_{dt} * IDF_d$$

Dimana:

d = dokumen ke-d

t = kata ke-t dari kata kunci

W = bobot dokumen ke-d terhadap kata ke-t

tf = banyaknya kata yang dicari pada sebuah dokumen

IDF = Inverse Dokument Frequency

D = total dokumen

df = banyak dokumen yang mengandung kata yang dicari

### Jaccard Coefficient

Untuk menentukan similaritas antar dokumen yang satu dengan yang lain dilakukan proses hitung jarak antar dokumen. Pada penelitian ini akan dihitung jarak antar dokumen untuk menentukan similaritas antar dokumen sedangkan metode yang digunakan adalah metode *Jaccard Coefficient*. *Jaccard Coefficient* merupakan metode yang digunakan untuk menghitung tingkat kesamaan (*similarity*) antar dua buah objek. Untuk tujuan klastering dokumen, fungsi yang baik adalah fungsi Jaccard Coefficient (Salton, 1989). Untuk notasi himpunan dapat digunakan rumus

$$Similarity(X, Y) = \frac{|X \cap Y|}{|X| + |Y| - |X \cap Y|}$$

dari notasi himpunan di atas dapat dibuat persamaan matematika sebagai berikut :

$$Similarity(X, Y) = \frac{\sum_{i=1}^t x_i y_i}{\sum_{i=1}^t x_i^2 + \sum_{i=1}^t y_i^2 - \sum_{i=1}^t x_i y_i}$$

Untuk merealisasi konsep ini, setiap dokumen harus dikorelasikan dengan subyek dengan relasi *many to many*, artinya satu subyek bisa memiliki beberapa dokumen, sebaliknya satu dokumen bisa juga memiliki beberapa subyek. Untuk dapat melakukan pengelompokan dokumen terhadap subyek dapat dilakukan dengan 2 cara, yaitu:

1. Memasukkan setiap dokumen secara langsung

kedalam subyek

2. Memasukkan dokumen secara tidak langsung kedalam suatu subyek dengan menggunakan bantuan term.

### Klastering

*Clustering* (klastering) didefinisikan sebagai upaya mengelompokkan data ke dalam klaster sedemikian sehingga data-data di dalam klaster yang sama memiliki lebih kesamaan dibandingkan dengan data-data pada klaster yang berbeda. Bisa juga diartikan sebagai proses untuk mendefinisikan pemetaan / *mapping*  $f : D \rightarrow C$  dari beberapa data  $D = \{t, t\}$  meminimalkan kesalahan, tapi PAM bekerja dengan Medoids, yang merupakan entitas dari dataset yang mewakili kelompok di mana ia dimasukkan, dan K-means bekerja dengan Sentroid, yang artifisial diciptakan entitas yang mewakili cluster.

Algoritma PAM sering dikenal dengan algoritma k-Medoids adalah sebuah algoritma yang merepresentasikan cluster yang dibentuk menggunakan medoids. Proses pembentukan cluster dimulai dengan menentukan k objek dari dataset secara acak sebagai medoid, selanjutnya hitung cost setiap objek non-medoid dengan k objek,  $cost = \{c^1, c^2, \dots, c^k\}$  terkecil setiap objek non-medoid terhadap medoid akan masuk dalam cluster dimana medoid tersebut  $n = 1, 2, \dots, n$  berdasarkan kesamaan antar  $t_j$ . Sebuah klaster adalah berada. Secara iteratif ditentukan kembali objek baru sekumpulan obyek yang digabung bersama karena persamaan atau kedekatannya.

Klastering biasa digunakan pada banyak bidang, seperti : *data mining*, *pattern recognition* (pengenalan pola), *image classification* (pengklasifikasian gambar), ilmu biologi, pemasaran, perencanaan kota, pencarian dokumen, dan lain sebagainya.

Tujuan dari klastering adalah untuk menentukan pengelompokan dari suatu set data. Akan tetapi tidak ada "ukuran terbaik" untuk pengelompokan data. Untuk pengelompokan data tergantung tujuan akhir dari klastering, maka diperlukan suatu kriteria sehingga hasil

klastering seperti yang diinginkan.

Penelitian tentang *clustering document* (klastering dokumen) telah banyak dilakukan. Secara umum klastering dokumen adalah proses mengelompokkan dokumen berdasarkan kemiripan antara satu dengan yang lain dalam satu klaster (Gordon, 1991; Ellis, 1996).

Tujuan klastering dokumen adalah untuk memisahkan dokumen yang relevan dari dokumen yang tidak *relevan* (Zhang, dkk., 2001). Atau dengan kata lain, dokumen-dokumen yang *relevan* dengan suatu *query* cenderung memiliki kemiripan satu sama lain dari pada dokumen yang tidak relevan, sehingga dapat dikelompokkan ke dalam suatu klaster.

Klastering dokumen dapat dilakukan sebelum atau sesudah proses temu kembali (Zhang, dkk., 2001). Pada klastering dokumen yang dilakukan sebelum proses temu kembali informasi, koleksi dokumen dikelompokkan ke dalam klaster berdasarkan kemiripan (*similarity*) antar dokumen. Selanjutnya dalam proses temu kembali informasi, apabila suatu dokumen ditemukan maka seluruh dokumen yang berada dalam klaster yang sama dengan dokumen tersebut juga dapat ditemukan.

### Algoritma PAM (Partitioning Around Medoids)

Algoritma PAM dikembangkan oleh Leonard Kaufman dan Peter J. Rousseeuw, dan algoritma ini sangat mirip dengan K-means, terutama karena keduanya algoritma partitional, dengan kata lain, keduanya memecah dataset menjadi kelompok-kelompok, dan keduanya bekerja berusaha untuk secara acak, lalu proses perhitungan cost dilakukan kembali. Apabila total cost yang dihasilkan lebih kecil dari cost setiap objek dengan medoid lama, maka objek baru tersebut dapat menjadi medoid baru. Iterasi berakhir hingga tidak ada perubahan terhadap cost yang dihasilkan oleh medoid baru.

Algoritma *PAM (Partitioning Around Medoids)* menggunakan metode partisi *clustering* untuk mengelompokkan sekumpulan  $n$  objek menjadi sejumlah  $k$  *cluster*. Algoritma ini menggunakan objek pada kumpulan objek untuk mewakili sebuah *cluster*. Objek yang terpilih untuk

mewakili sebuah *cluster* disebut *medoid*. *Cluster* dibangun dengan menghitung kedekatan yang dimiliki antara *medoid* dengan objek *non-medoid*.

Menurut Han dan Kamber (2006, hal 406) algoritma K-medoids adalah sebagai berikut.

1. Secara acak pilih  $k$  objek pada sekumpulan  $n$  objek sebagai *medoid*.
2. Ulangi:
3. Tempatkan objek *non-medoid* ke dalam *cluster* yang paling dekat dengan *medoid*.
4. Secara acak pilih oacak: sebuah objek *non-medoid*.
5. Hitung total biaya,  $S$ , dari pertukaran medoid oj dengan orandom.
6. Jika  $S < 0$  maka tukar oj dengan oacak untuk membentuk sekumpulan  $k$  objek baru sebagai *medoid*.
7. Hingga tidak ada perubahan.

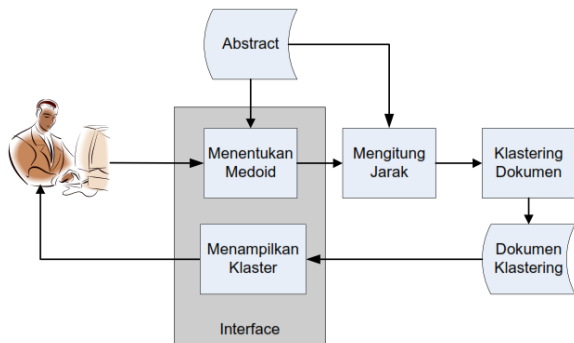
Bagian ini menunjukkan K-medoids pengelompokkan dengan fungsi pam () dan pamk (). K-medoids sangat mirip dengan K-means dan perbedaan utama diantara mereka adalah jika cluster diwakili dengan pusat dari cluster pada K-means, sedangkan pada K-medoids, cluster diwakili oleh objek terdekat dari pusat cluster.

## HASIL DAN PEMBAHASAN

### Arsitektur Sistem

Penelitian ini dibangun sistem klastering dokumen teks skripsi mahasiswa Program Studi Sistem Informasi Universitas Stikubank. Dokumen skripsi diambil langsung secara online yang ada di alamat [eprints.unisbank.ac.id](http://eprints.unisbank.ac.id). Dokumen skripsi tersimpan dalam database yang merupakan hasil penelitian terdahulu Ekstraksi Dokumen Skripsi Mahasiswa Fakultas Teknologi Informasi (Februariyanti H, Santoso D.W., 2016). Selanjutnya dari hasil proses ekstraksi dokumen skripsi telah dihasilkan database abstrak skripsi mahasiswa, pada penelitian ini akan dibangun klastering dokumen skripsi mahasiswa Program Studi Sistem Informasi Universitas Stikubank

Semarang dengan menggunakan algoritma PAM (Partitioning Arouds Medoids). Arsitektur dari sistem yang akan dikembangkan dapat dilihat pada gambar 1 di bawah ini:



Gambar 1. Arsitektur Sistem

**Data Percobaan**

Penelitian ini menggunakan dokumen text, yaitu dokumen skripsi mahasiswa Program Studi Sistem Informasi Fakultas Teknologi Informasi Universitas Semarang. Dokumen skripsi tersebut ada di alamat eprints.unisbank.ac.id sejumlah 50 dokumen. Dari data tersebut akan dilakukan proses kalkulasi untuk menentukan cluster yang akan menggambarkan topik- topik skripsi yang banyak diambil oleh mahasiswa Program Studi Sistem Informasi dalam mengambil matakuliah Skripsi. Dengan demikian kita dapat membuat analisis dan memetakan matakuliah yang akan dirumuskan dalam Program Studi Sistem Informasi kedepannya dalam peninjauan Kurikulum. Dokumen abstrak skripsi yang digunakan dalam penelitian merupakan dokumen hasil riset sebelumnya yang tersimpan dalam database terdiri dari dua tabel yaitu tabel abstract dan tabel indeks. Rancangan tabel hasil ekstraksi dokumen skripsi dapat dilihat pada tabel seperti terlihat pada gambar 2 di bawah ini:



Gambar 2 Database Abstrak Skripsi

Di dalam database extract terdapat dua buah tabel yang berfungsi untuk menyimpan dan

mengolah dokumen abstrak skripsi.

1. Tabel abstract

Tabel ini menyimpan data judul dan data deskripsi abstrak skripsi mahasiswa program studi informasi. Tabel filtered merupakan penanda apakah sebuah dokumen abstrak sudah diproses atau belum.

2. Tabel indeks

Tabel menampung data abstrak yang sudah melalui tahapan pre-processing. Di dalam tabel indeks terdapat id\_abstract dan token yang merupakan kata yang sudah mengalami pre-processing.

**Klustering Dokumen Abstrak Skripsi**

Algoritma PAM sering dikenal dengan algoritma-Medoids adalah sebuah algoritma yang merepresentasikan cluster yang dibentuk menggunakan medoids. Proses pembentukan cluster dimulai dengan menentukan k objek dari dataset secara acak sebagai medoid, selanjutnya hitung cost setiap objek non-medoid dengan k objek, cost terkecil setiap objek non-medoid terhadap medoid akan masuk dalam cluster dimana medoid tersebut berada. Secara iteratif ditentukan kembali objek baru secara acak, lalu proses perhitungan cost dilakukan kembali. Apabila total cost yang dihasilkan lebih kecil dari cost setiap objek dengan medoid lama, maka objek baru tersebut dapat menjadi medoid baru. Iterasi berakhir hingga tidak ada perubahan terhadap cost yang dihasilkan oleh medoid baru.

Pada penelitian ini user akan menentukan 3 (tiga) medoid yang dijadikan objek klaster. Interface untuk menentukan medoid dapat dilihat pada gambar 5.6 dibawah ini :

**PAM Clustering Dokumen Skripsi Prodi Sistem Informasi**

No	Judul Abstrak	Pilih Medoids
1	PERANCANGAN SISTEM INFORMASI BERBASIS E-COMMERCE PADA MYKAELA STORE DENGAN PHP DAN JQUERY	<input type="checkbox"/>
2	SISTEM INFORMASI E-COMMERCE ALAT TELEKOMUNIKASI UNTUK CEK SINYAL TELEFON SELLER PADA PT. NEXWAVE	<input type="checkbox"/>
3	RANCANG BANGUN APLIKASI ECOMMERCE MENGGUNAKAN FRAMEWORK CODEIGNITER (STUDI KASUS ACE BATAK & BANGYACRAFT)	<input type="checkbox"/>
4	RANCANG BANGUN SISTEM INFORMASI PEMESANAN ALAT PRAKTEK KEBOHONAN BERBASIS WEB PADA PROGRAM STUDI DIKEBOHONAN UNIVERSITAS MUHAMMADIYAH SENARANG	<input type="checkbox"/>
5	RANCANG BANGUN SISTEM INFORMASI PENYEWAAN LAPWIGAN METRO BERBASIS WEB	<input type="checkbox"/>
6	SISTEM INFORMASI CASH COLLECTION DI BPR WELERI JAWAPERSADA	<input type="checkbox"/>

Gambar 3 Interface Klastering PAM

Pada gambar 3 diperlihatkan tampilan interface sistem klastering dokumen skripsi menggunakan algoritma PAM. Pada tampilan diperlihatkan title kolom terdiri dari no, judul abstrak dan pilih medoids. No merupakan id\_indeks dari abstrak skripsi, judul abstrak berisi judul dari abstrak sekripsi mahasiswa dan user dapat menentukan medoids untuk objek klaster dengan mencenthang 3 (tiga) point di kotak kecil.

Setelah dipilih 3 (tiga) medoids sebagai objek klaster, maka akan dilanjutkan dengan proses menghitung jarak antar dokumen skripsi. Setelah proses menghitung jarak dilakukan maka sistem akan menampilkan interface perhitungan jarak antar dokumen seperti terlihat pada gambar 5.7 dibawah. Kolom tampilan terdiri dari no, id\_nm, medoids ke\_1, medoids ke-2, medoids ke\_3, term\_nm, m1->nm, m2-> nm, m3-> nm dan klaster. No menunjukan no urut proses perhitungan jarak antar dokumen, id\_nm adalah no id dari abstrak skripsi yng tersimpan dalam database, m2-> nm adalah perhitungan jarak antar dokumen dari dokumen dengan no\_id 1 dengan dokumen medoids 1, begitu juga m2 dan m3, sedangkan klaster akan berisi no\_id dari klaster yang terpilih dari objek yang diuji.

**Menghitung Jarak Dokumen dengan Medoids**

Pada penelitian ini akan dihitung jarak antar dokumen untuk menentukan simmilaritas antar dokumen sedangkan metode yang digunakan adalah metode *Jaccard Coefficient*. *Jaccard Coefficient* merupakan metode yang digunakan untuk menghitung tingkat kesamaan (*similarity*) antar dua buah objek. Untuk tujuan

klastering dokumen, fungsi yang baik adalah fungsi *Jaccard Coefficient* (Salton, 1989). Untuk notasi himpunan dapat digunakan rumus

$$Similarity(X,Y) = \frac{|X \cap Y|}{|X \cup Y|} = \frac{|X \cap Y|}{|X| + |Y| - |X \cap Y|}$$

Penelitian ini akan menghitung jarak antar dokumen dengan menghitung jarak dokumen abstrak skripsi yang akan diuji dengan dokumen abstrak skripsi yang dipilih sebagai medoids. Seperti terlihat pada gambar 4 adalah hasil dari proses perhitungan jarak antara dokumen yang akan diuji (pada kolom ID NM sebagai variabel X dan medoids ke-1, medoid ke-2, medoid ke-3 sebagai variabel Y. Sehingga untuk perhitungan jarak untuk tiap dokumen akan dilakukan sejumlah medoids yang kita tentukan.

No	ID NM ke-1	Medoids ke-2	Medoids ke-3	Term NM	M1 -> NM	M2 -> NM	M3 -> NM	Klaster
1	1	20	25	28	1 20 / 145 = 0.137931034482759	10 / 134 = 0.13432835820896	15 / 137 = 0.13138686131307	1
1	2	20	25	28	2 29 / 221 = 0.13122171945701	21 / 210 = 0.1	19 / 215 = 0.088372093023256	1
1	3	20	25	28	3 34 / 291 = 0.11683848797251	25 / 255 = 0.088667071530249	24 / 295 = 0.081423740315789	1
1	4	20	25	28	4 39 / 369 = 0.10569105691057	25 / 264 = 0.088681318881819	27 / 305 = 0.07972982739729	1
1	5	20	25	28	5 45 / 450 = 0.09999109991099	26 / 450 = 0.057777777777778	33 / 431 = 0.073170731707317	1
1	6	20	25	28	6 45 / 547 = 0.082266919430475	26 / 547 = 0.047531992667396	36 / 540 = 0.066666666666667	1
1	7	20	25	28	7 47 / 624 = 0.075320513620513	28 / 623 = 0.044943818619583	40 / 615 = 0.065040650406504	1

Gambar 4 Interface Hasil Perhitungan Jarak

Setelah perhitungan jarak antara dokumen dengan medoids dihasilkan pengelompokkan dokumen abstrak skripsi menggunakan algoritma Partitioning Around Medoids, dihasilkan tampilan program seperti terlihat pada gambar 5 di bawah ini :

Hasil Klaster Dokumen Abstrak Skripsi

Klaster 1	Klaster 2	Klaster 3
1 RANCANG BANGUN APLIKASI ECOMMERCE MENGGUNAKAN FRAMEWORK CODEIGNITER (STUDI KASUS ACE BATAK & BANGYACRAFT) 2 SISTEM INFORMASI E-COMMERCE ALAT TELEKOMUNIKASI UNTUK CEK SINYAL TELEFON SELLER PADA PT. NEXWAVE 3 RANCANG BANGUN APLIKASI ECOMMERCE MENGGUNAKAN FRAMEWORK CODEIGNITER (STUDI KASUS ACE BATAK & BANGYACRAFT) 4 PERANCANGAN SISTEM INFORMASI BERBASIS E-COMMERCE PADA MYKAELA STORE DENGAN PHP DAN JQUERY 5 RANCANG BANGUN APLIKASI ECOMMERCE MENGGUNAKAN FRAMEWORK CODEIGNITER (STUDI KASUS ACE BATAK & BANGYACRAFT)	1 RANCANG BANGUN APLIKASI ECOMMERCE MENGGUNAKAN FRAMEWORK CODEIGNITER (STUDI KASUS ACE BATAK & BANGYACRAFT) 2 PERANCANGAN SISTEM INFORMASI BERBASIS E-COMMERCE PADA MYKAELA STORE DENGAN PHP DAN JQUERY 3 RANCANG BANGUN APLIKASI ECOMMERCE MENGGUNAKAN FRAMEWORK CODEIGNITER (STUDI KASUS ACE BATAK & BANGYACRAFT)	1 RANCANG BANGUN APLIKASI ECOMMERCE MENGGUNAKAN FRAMEWORK CODEIGNITER (STUDI KASUS ACE BATAK & BANGYACRAFT) 2 PERANCANGAN SISTEM INFORMASI BERBASIS E-COMMERCE PADA MYKAELA STORE DENGAN PHP DAN JQUERY 3 RANCANG BANGUN APLIKASI ECOMMERCE MENGGUNAKAN FRAMEWORK CODEIGNITER (STUDI KASUS ACE BATAK & BANGYACRAFT) 4 RANCANG BANGUN APLIKASI ECOMMERCE MENGGUNAKAN FRAMEWORK CODEIGNITER (STUDI KASUS ACE BATAK & BANGYACRAFT)

Gambar 5 Tampilan Hasil Klastering PAM Abstrak Skripsi

**SIMPULAN DAN PENELITIAN LANJUTAN**

**Simpulan**

Dari penelitian yang telah dilakukan dihasilkan kesimpulan sebagai berikut :

1. Aplikasi dapat melakukan proses menentukan sejumlah 3 (tiga) medoids oleh user dari dokumen abstrak skripsi mahasiswa.
2. Aplikasi dapat melakukan perhitungan jarak antara dokumen abstrak skripsi yang dipilih sebagai medoids dengan dokumen diluar medoid.
3. Aplikasi dapat menampilkan hasil klastering dokumen abstrak skripsi menggunakan algoritma PAM.

### Penelitian Lanjutan

Untuk penelitian selanjutnya dapat dilakukan penelitian:

1. Klastering dengan algoritma yang berbeda untuk menganalisis dokumen abstrak skripsi apakah nantinya hasilnya sesuai atau tidak sebagai pengujian algoritma yang berbeda.
2. Penelitian Klasifikasi
3. Penelitian Pemaknaan dokumen

### DAFTAR PUSTAKA

- Even-Zohar Y, 2002. Introduction to Text Mining, Supercomputing
- hrisnanto, Yulison Herry , 2013, Algoritma Partitioning Around Medoids (PAM) Sebagai Teknik Clustering Pada Data Mining, Majalah Ilmiah Aristoteles FMIPA Unjani, Volume 11 Nomor 2, Oktober 2013
- Feldman, R., and J. Sanger. 2006. *The Text Mining Handbook, Advances Approaches in Analyzing Unstructured Data*. Cambridge: Cambridge University Press
- Han, J., Kamber, M., Data Mining Concept and Technique, 2<sup>nd</sup> Ed, Elsevier, 2006
- Hearst M, 2003, What is Text Mining ? <http://www.sims.berkeley.edu/~hearst/text-mining.html>
- Salton, G., 1989, *Automatic Text Processing, The Transformation, Analysis, and Retrieval of Information by Computer*, Addison – Wesley Publishing Company, Inc. All rights reserved.
- Salton, G. and Buckley, 1988, *Term Weigting*

*Approaches in Automatic Text Retrieval*, Department of Computer Science, Cornell University, Ithaca, NY 14853, USA

Tala, F.Z., 2003, *A Study of Stemming Effects on Information Retrieval in Bahasa Indonesia*. Institute for Logic, Language and Computation Universiteit van Amsterdam The Netherlands

Triawati Candra, 2009, *Metode Pembobotan Statistical Concept Based Untuk Klastering dan Kategorisasi Dokumen Berbahasa Indonesia*, Institut Teknologi Telkom Bandung