

IMPLEMENTASI K-MEANS KLUSTERING UNTUK REKOMENDASI TEMA TUGAS AKHIR PADA STMIK ASIA MALANG

Lia Farokhah¹⁾, Rendy Aditya²⁾

^{1,2}Teknik Informatika, STMIK ASIA Malang
email: ¹farokhah@asia.ac.id,
²rendya@asia.ac.id

Abstract : *Each student who attended the lecture at a college will inevitably undergo thesis examination or final project to finish the S1 degree. However, students still often have difficulty in determining the title theme to be lifted. On the other hand, many students take random theme of the thesis title or final project, just following friends or colleagues or looking for a single reference from the library. Therefore, it takes a process of data mining to assist students in determining the proper final assignment theme. The process of data mining is done by using K-Means algorithm with input value students courses that have reached as the determining aspect, so that it found a pattern of students interest which is used to recommended them in determining the final project or thesis theme that suits their ability. In its application, classification as done against the final project themes were divided into 7 groups/clusters. Using students score as the input, the process is done using the K-Means algorithm by taking into account distance between data to the center cluster (centroid). So that there is no more data move to the other cluster. Based on the testing against the centroid for 15 times, obtained a result that second centroid has the highest truth compared to other centroid with the value was 90.24%. Therefore, the system will use the second centroid as a new reference in determining the theme of the final project for each student by viewing the closest data to the centroid point.*

Keywords : *K-Means, Recommendation, Themes, Final Project*

PENDAHULUAN

Setiap mahasiswa yang menempuh kuliah di suatu perguruan tinggi pasti akan menjalani ujian skripsi atau tugas akhir. Namun seringkali mahasiswa mengalami kesulitan dalam menentukan tema judul yang akan diangkat. Hal serupa juga kerap terjadi pada mahasiswa STMIK Asia Malang. Banyaknya mahasiswa yang kebingungan dalam menentukan tema judul pada akhirnya berdampak pada lamanya proses penyelesaian tugas akhir dan nilai yang tidak memuaskan. Banyak mahasiswa yang mengambil tema judul skripsi atau tugas akhir secara acak, hanya mengikuti teman sejawat atau mencari referensi tunggal dari perpustakaan. Selain itu, tidak sedikit juga mahasiswa yang mencari referensi judul dari internet yang seringkali tidak sesuai dengan kemampuan, yang berakibat pada penguasaan materi yang tidak maksimal.

Agar proses pengerjaan tugas akhir menjadi sebuah penelitian yang sesuai dengan kemampuan mahasiswa, diperlukan suatu proses penggalian data untuk membantu mahasiswa dalam menentukan tema judul tugas akhir yang tepat. Serangkaian proses untuk mendapatkan pengetahuan atau pola dari sekumpulan data disebut *data mining*. Aspek yang digunakan sebagai penentu tema tugas akhir meliputi nilai mata kuliah wajib dan pilihan yang telah ditempuh selama kuliah. Dengan menggunakan algoritma K-Means, dilakukan proses penggalian data sehingga ditemukan pola minat mahasiswa yang digunakan untuk membantu pengambilan keputusan dalam menentukan tema judul yang akan diangkat oleh mahasiswa sebagai tugas akhir.

Berdasarkan latar belakang tersebut maka dibuatlah sebuah penelitian yang berjudul "Implementasi Algoritma K-

Means untuk Rekomendasi Tema Tugas Akhir pada STMIK Asia Malang”.

METODE PENELITIAN

Kebutuhan Input

Data masukan yang digunakan terbagi dalam 12 inputan yang diambil dari data nilai mahasiswa per mata kuliah yaitu:

1. *Artificial Intelligence*
2. *Computer Network 2*
3. *Data Mining*
4. Keamanan Jaringan
5. Pemrograman Berorientasi Objek
6. Pemrograman Dasar
7. Pemrograman Terstruktur 2
8. Pemrograman Web 2
9. Pengolahan Citra dan Pola
10. Sistem Pakar
11. Sistem Pendukung Keputusan
12. Statistik dan Probabilitas

Dari sekian banyak mata kuliah yang ditawarkan mulai dari semester satu sampai tujuh, tidak semuanya digunakan sebagai parameter input. Mata kuliah diatas dipilih karena dianggap mendukung dalam penentuan tema tugas akhir.

Kebutuhan Proses

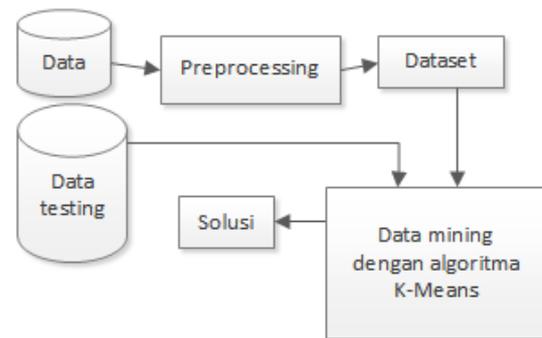
Proses awal yang dilakukan adalah dengan mengolah data nilai mahasiswa dari masing-masing mata kuliah hingga didapatkan *dataset* yang kemudian akan dilakukan proses *data mining* dengan menggunakan algoritma *K-Means*.

Kebutuhan Output

Keluaran yang diharapkan dari pengolahan data ini yaitu dapat memberikan rekomendasi tema tugas akhir yang akan diambil oleh mahasiswa sesuai dengan kemampuan berdasarkan nilai-nilai dari mata kuliah yang telah diambil.

Diagram Blok Sistem

Gambar 1 merupakan alur dari diagram blok sistem:



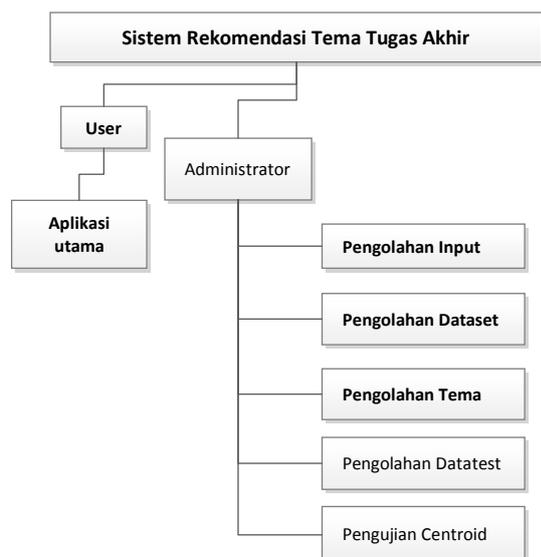
Gambar 1 Diagram Blok Sistem

1. *Data*
Pemilihan data (*data selection*) dari sekumpulan data operasional perlu dilakukan sebelum tahap penggalian informasi dimulai.
2. *Preprocessing*
Pemrosesan pendahuluan seperti sinkronisasi data, penambahan atribut, eliminasi atribut, transformasi data dan konversi data.
3. *Dataset*
Data hasil dari *preprocessing* yang nantinya akan digunakan dalam proses *mining*.
4. *Data testing*
Data yang digunakan untuk menguji sistem yang telah dibuat. Struktur *data testings* sama dengan *dataset*.
5. *Data mining* dengan algoritma *K-Means*
Proses mencari pola atau informasi yang menarik dalam data yang sebelumnya tidak diketahui dengan menggunakan metode *K-Means*.
6. *Solusi*
Penerjemahan pola-pola yang dihasilkan dari *data mining*, yang ditampilkan dalam bentuk yang mudah dimengerti oleh pengguna.

Perancangan Sistem

1. Struktur Sistem

Sistem dibagi menjadi 2 bagian utama, yaitu bagian *user* (pengguna) dan bagian *administrator*. Gambar 2 menunjukkan struktur sistem secara keseluruhan.

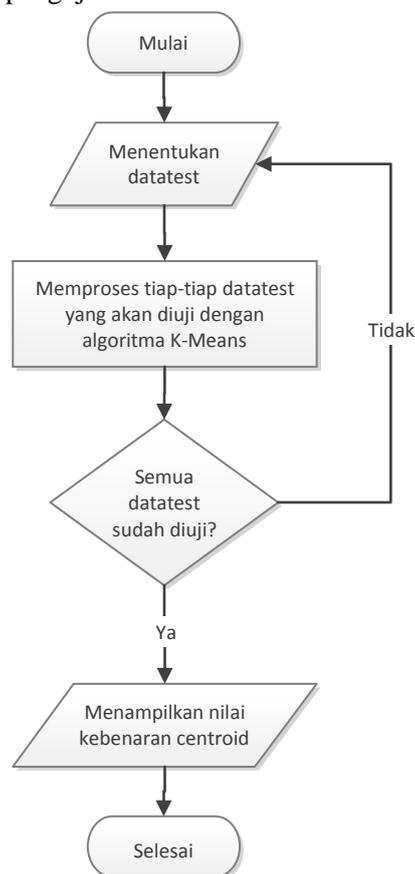


Gambar 2 Struktur Sistem

Penjelasan dari tiap-tiap bagian struktur sistem yaitu:

1. **Aplikasi utama**
Aplikasi utama merupakan aplikasi rekomendasi tema tugas akhir yang bisa diakses oleh pengguna (*user*). Aplikasi utama menampilkan sebuah *form* yang berisi inputan nilai-nilai mata kuliah yang telah ditempuh oleh mahasiswa dan sebuah tombol untuk melakukan proses serta menampilkan hasil rekomendasi tema tugas akhir.
2. **Pengolahan *input***
Pada bagian pengolahan *input*, admin dapat melakukan pengaturan terhadap mata kuliah yang akan dijadikan sebagai parameter dalam menentukan rekomendasi tema tugas akhir.
3. **Pengolahan *dataset***
Pada bagian pengolahan *dataset*, admin dapat melakukan pengolahan data-data yang dijadikan sebagai *dataset*.
4. **Pengolahan tema**
Pada bagian pengolahan tema, admin dapat melakukan pengolahan tema-tema tugas akhir yang dijadikan sebagai *cluster*.
5. **Pengolahan *datatest***
Pada bagian pengolahan *datatest*, admin dapat melakukan pengolahan data-data yang dijadikan sebagai data uji.
6. **Pengujian *centroid***

Pada bagian pengujian, admin dapat melakukan pengujian terhadap *centroid* dengan menggunakan *datatest* yang telah tersimpan pada *database* guna mendapatkan pusat *cluster* (*centroid*) dengan nilai kebenaran yang paling baik. Gambar 3 menunjukkan alur proses pengujian *centroid*.

Gambar 3 Alur Proses Pengujian *Centroid*

HASIL DAN PEMBAHASAN

Implementasi Perangkat Lunak

Sistem dibagi menjadi 2 bagian, yaitu bagian *user* (pengguna) dan bagian *administrator*.

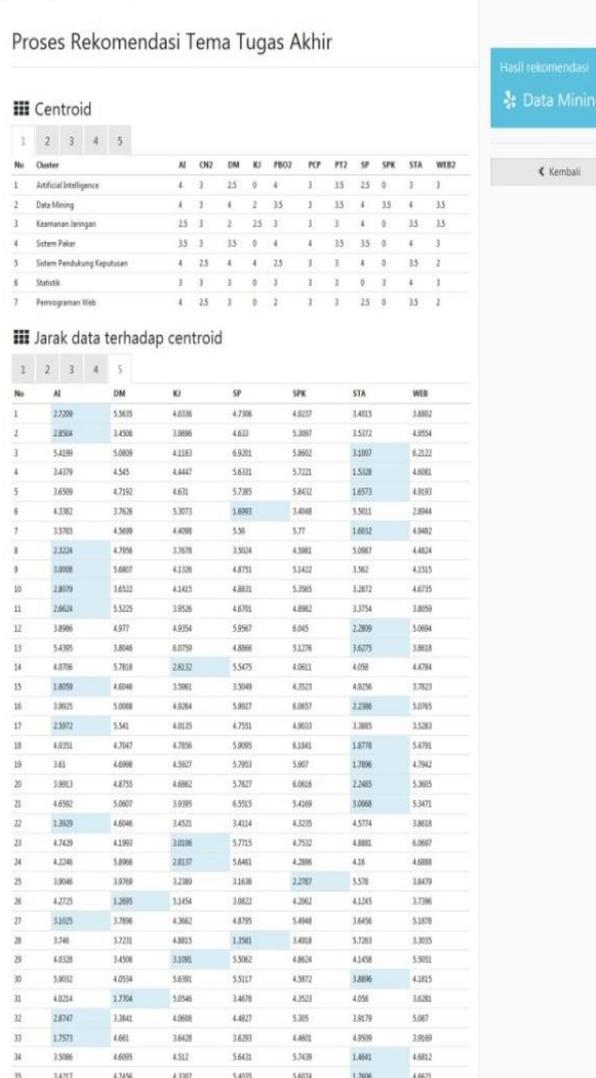
Bagian *User*

Pada dasarnya, *user* (pengguna) hanya dapat memberikan input berupa nilai-nilai mata kuliah yang telah ditempuh dan mendapatkan hasil berupa suatu rekomendasi tema tugas akhir tanpa perlu mengetahui proses, seperti ditunjukkan pada Gambar 4.



Gambar 4 Halaman User

Namun pada sistem, ditambahkan juga pilihan untuk melihat proses penghitungan dengan menggunakan algoritma *K-Means* hingga didapatkan hasil rekomendasi tema tugas akhir. Gambar 5 menunjukkan proses penghitungan *K-Means*.



Gambar 5 Proses Penghitungan Algoritma *K-Means*

Secara umum, kedua halaman yang ditunjukkan pada Gambar 4 dan Gambar 5 mempunyai proses pengerjaan yang sama. Berikut potongan kode program yang menunjukkan proses penghitungan dengan menggunakan algoritma *K-Means*.

```
// Proses penghitungan algoritma K-
Means
// Iterasi terus berjalan hingga
data jarak tidak mengalami
perubahan
$loop = -1;
do {
    $loop++;

    // Menentukan centroid
    if ($loop == 0) {
        $centroid[$loop] = $this-
        >getCentroid();
    } else {
        $centroid[$loop] = $this-
        >getCentroid($member[$loop-1]);
    }

    // Menghitung jarak data terhadap
    centroid
    $distance[$loop] = $this-
    >calcDistance($centroid[$loop]);

    // Mengelompokkan tiap-tiap data
    kedalam cluster
    $temp = $this-
    >clusteringData($distance[$loop])
    ;
    $member[$loop] = $temp[0];
    $group[$loop]
    = $temp[1];
} while ( $loop == 0 || ($loop > 0
&& ($group[$loop] != $group[$loop-
1])) );
```

Bagian Administrator

Seorang admin dapat melakukan pengelolaan lebih lanjut terkait data-data pada *database*, yaitu data *input* mata kuliah, *dataset*, tema, *datatest* (data uji), serta pengujian, seperti ditunjukkan pada Gambar 6.

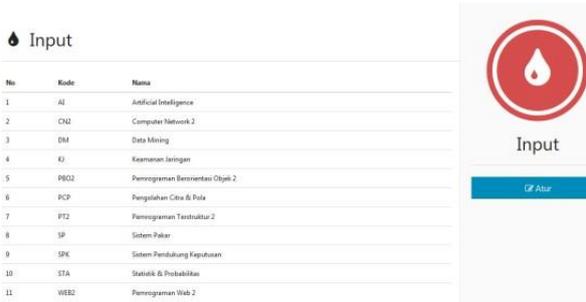


Gambar 6 Halaman Administrator

Terdapat beberapa fitur yang dapat diakses pada bagian *administrator*, yaitu:

A. Halaman *Input*

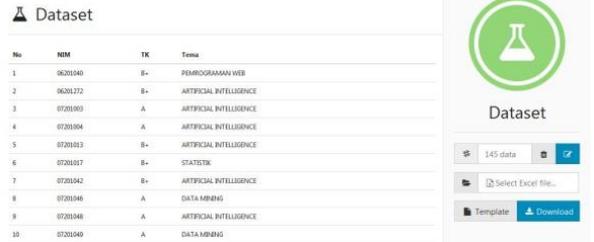
Pada halaman *input*, admin dapat melakukan pengaturan terhadap mata kuliah yang akan dijadikan sebagai parameter dalam menentukan rekomendasi tema tugas akhir. Gambar 7 merupakan tampilan halaman *input*.



Gambar 7 Halaman *Input*

B. Halaman *Dataset*

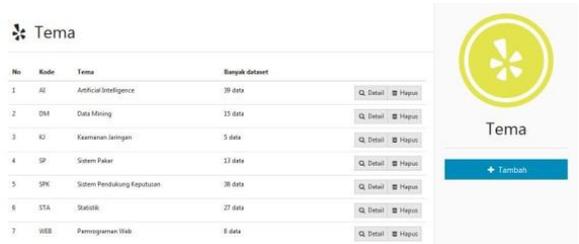
Pada halaman *dataset*, admin dapat melakukan pengaturan terhadap data-data yang digunakan untuk menentukan rekomendasi. Gambar 8 merupakan tampilan halaman *dataset*.



Gambar 8 Halaman *Dataset*

C. Halaman *Tema*

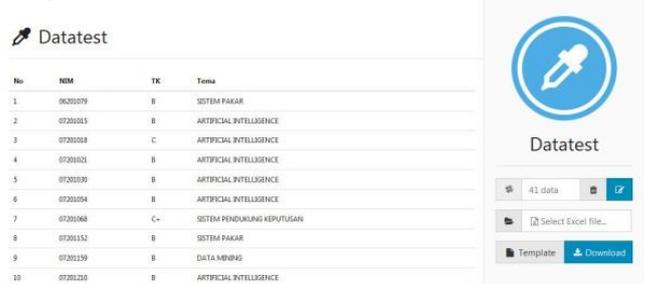
Pada halaman *tema*, admin dapat melakukan pengaturan terhadap tema-tema tugas akhir. Gambar 9 merupakan tampilan halaman *tema*.



Gambar 9 Halaman *Tema*

D. Halaman *Datatest*

Pada halaman *datatest*, admin dapat melakukan pengaturan terhadap data-data yang digunakan sebagai data uji dalam proses pengujian. Gambar 10 merupakan tampilan halaman *datatest*.



Gambar 10 Halaman *Datatest*

E. Halaman *Pengujian*

Pada halaman *pengujian*, admin dapat melakukan pengujian dengan menggunakan *datatest* yang telah tersimpan pada *database* guna mendapatkan pusat *cluster (centroid)* dengan nilai kebenaran yang paling baik. Gambar 11 merupakan tampilan halaman *pengujian*.

No	Cluster	AI	CK2	DM	KI	PRO2	PCP	PT2	SP	SPK	STA	WER2
1	Artificial Intelligence	4	3	2,5	0	4	3	3,5	2,5	0	3	3
2	Data Mining	4	3	4	2	3,5	3	3,5	4	3,5	4	3,5
3	Kemampuan Jaringan	2,5	3	2	2,5	3	3	3	4	0	3,5	3,5
4	Sistem Pakar	3,5	3	3,5	0	4	4	3,5	3,5	0	4	3
5	Sistem Pendukung Keputusan	4	2,5	4	4	2,5	3	3	4	0	3,5	2
6	Statistika	3	3	3	0	3	3	3	0	3	4	3
7	Pemrograman Web	4	2,5	3	0	2	3	3	2,5	0	3,5	2

No	NIM	TK	Tema
1	06200279	B	SISTEM PAKAR
2	07200263	B	ARTIFICIAL INTELLIGENCE
3	07200268	C	ARTIFICIAL INTELLIGENCE
4	07200263	B	ARTIFICIAL INTELLIGENCE
5	07200260	B	ARTIFICIAL INTELLIGENCE
6	07200264	B	ARTIFICIAL INTELLIGENCE

Gambar 11 Halaman Penguujian

Penguujian Perangkat Lunak

Dengan dilakukan penguujian, akan diketahui tingkat kelayakan dari sistem yang dibuat. Proses penguujian melibatkan *datatest* yang telah tersimpan pada *database* dan bertujuan untuk mendapatkan pusat *cluster* (*centroid*) yang memiliki nilai kebenaran paling baik. Pusat *cluster* (*centroid*) hasil penguujian akan dijadikan sebagai acuan dalam menentukan rekomendasi tema tugas akhir.

Proses Penguujian Centroid

Proses penguujian dilakukan sebanyak 15 kali. Parameter yang digunakan untuk menilai tingkat kebenaran yaitu nilai tugas khusus yang telah diambil oleh mahasiswa.

Untuk tiap-tiap *datatest* yang diuji, penguujian terhadap *centroid* dinyatakan benar apabila memenuhi salah satu persyaratan berikut:

1. Rekomendasi sistem sama dengan tema tugas khusus yang telah diambil mahasiswa dan nilai tugas khusus yang diperoleh A atau B+
2. Rekomendasi sistem tidak sama dengan tema tugas khusus yang telah diambil mahasiswa namun nilai tugas khusus yang diperoleh dibawah B+

Hasil Penguujian

Selain melakukan penguujian terhadap 15 buah *centroid* yang didapatkan secara acak, sistem juga melakukan

penguujian pada *centroid* yang tersimpan di *database* guna membandingkan tingkat akurasi terhadap kelima belas *centroid* lain. Hasil dari penguujian *centroid* dapat dilihat pada Gambar 12.

No	Centroid	Nilai Kebenaran
1	Centroid 1	75,61%
2	Centroid 2	90,24%
3	Centroid 3	75,61%
4	Centroid 4	36,36%
5	Centroid 5	75,61%
6	Centroid 6	80,49%
7	Centroid 7	76,03%
8	Centroid 8	80,49%
9	Centroid 9	85,37%
10	Centroid 10	75,61%
11	Centroid 11	85,37%
12	Centroid 12	80,49%
13	Centroid 13	86,20%
14	Centroid 14	73,17%
15	Centroid 15	73,17%
16	Centroid database	87,8%

Gambar 12 Hasil Penguujian Centroid

Pusat *cluster* (*centroid*) terbaik akan dijadikan sebagai acuan dalam menentukan rekomendasi tema tugas akhir. Berdasarkan Gambar 12, hasil penguujian *centroid* dengan nilai kebenaran terbaik didapatkan oleh *centroid* kedua dengan nilai 90.24%. Dengan demikian, sistem akan menggunakan *centroid* kedua sebagai acuan baru dalam menentukan tema tugas akhir menggantikan *centroid* lama yang tersimpan pada *database*.

4. KESIMPULAN DAN SARAN

4.1 Kesimpulan

Adapun kesimpulan dari penelitian adalah sebagai berikut:

1. Berdasarkan penguujian, diperoleh hasil bahwa *centroid* kedua mempunyai nilai kebenaran yang paling tinggi dibandingkan dengan *centroid* lain dengan nilai 90.24%.
2. Nilai kebenaran sistem didapatkan dari perbandingan antara tema judul yang telah diambil oleh mahasiswa dengan tema judul dari sistem serta nilai dari tugas khusus mahasiswa.

4.2 SARAN

1. Penelitian ini harus menambah *sample* mata kuliah yang lebih kompleks yang ada di Kartu Hasil Studi (KHS) supaya validasi keputusannya semakin valid.

2. Pada aplikasi ini belum ada perhitungan bobot keterkaitan antara mata kuliah yang sejajar yang berkelanjutan misalkan kelompok mata kuliah rumpun programming atau rumpun jaringan.

5. REFERENSI

- Agusta, Y. K-Means Penerapan, Permasalahan, dan Metode Terkait. *Jurnal Sistem dan Informatika*. 2007
- Fathansyah, Ir. *Basis Data*. Bandung. Informatika. 1999
- Han, Jiawei, and Kamber, Micheline. *Data Mining: Concepts and Techniques*. San Fransisco. Morgan Kaufman Publishers. 2000
- Kurgan, Lukasz, and Musilek, Petr. *A Survey of Knowledge Discovery and Data Mining Process Models*. New York. Cambridge University Press. 2006
- Kusumadewi, Sri. *Artificial Intelligence Teknik dan Aplikasinya*. Yogyakarta. Graha Ilmu. 2003
- Larose, Daniel T. *Discovering Knowledge in Data An Introduction to Data Mining*. New Jersey. John Wiley's and Son. 2005
- Peranginangin, Kasiman. *Aplikasi Web dengan PHP dan MySQL*. Yogyakarta. Andi. 2006
- Pramudiono, Iko. *Pengantar Data Mining: Menambang Permata Pengetahuan di Gunung Data*. Tokyo. The University of Tokyo. 2003
- Prasetyo, Eko. *Data Mining Konsep dan Aplikasi Menggunakan Matlab*. Yogyakarta. Andi. 2012
- Rich, Elaine, and Knight, Kevin. *Artificial Intelligence*. New York. McGraw-Hill Inc. 1991
- Santosa, Budi. *Data Mining Teknik Pemanfaatan Data untuk Keperluan Bisnis*. Yogyakarta. Graha Ilmu. 2007
- Widyawati, N. *Perbandingan Clustering Based on Frequent Word Sequence (CFWS) dan K-Means untuk Pengelompokan Dokumen Berbahasa Indonesia*. Bandung. Fakultas Pendidikan Matematika dan Ilmu Pengetahuan Alam Universitas Pendidikan Indonesia. 2010