

Pengelompokan Dokumen Berita Menggunakan *Fuzzy C-means*

*Diajukan Untuk Menyusun Tugas Akhir
di Jurusan Teknik Informatika Fakultas Ilmu Komputer UNSRI*



Oleh :

**DWI ASTUTI
09101002015**

**JURUSAN TEKNIK INFORMATIKA
FAKULTAS ILMU KOMPUTER UNIVERSITAS SRIWIJAYA
2016**

LEMBAR PENGESAHAN TUGAS AKHIR

**PENGELOMPOKAN DOKUMEN BERITA
MENGUNAKAN FUZZY C- MEANS**

Oleh :

DWI ASTUTI

NIM : 09101002015

Palembang, 2016

Menyetujui

a/n Ketua Jurusan Teknik Informatika
sekretaris jurusan


Rifkie Primartha, M.T
NIP. 197706012009121004

Pembimbing I,


Rifkie Primartha, M.T
NIP. 197706012009121004

TANDA LULUS UJIAN SKRIPSI SIDANG TUGAS AKHIR

Pada hari jum'at 18 Maret 2016 telah dilaksanakan Ujian Skripsi Tugas Akhir oleh Jurusan Teknik Informatika Fakultas Ilmu Komputer Universitas Sriwijaya.

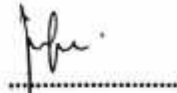
Nama : Dwi Astuti

NIM : 09101002015

Judul TA : Pengelompokan Dokumen Berita Menggunakan *Fuzzy C-Means*

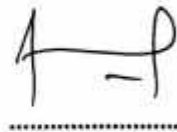
1. Ketua Penguji

Rifkie Primartha, M.T
NIP. 197706012009121004



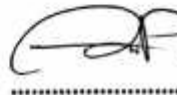
2. Penguji I

M. Fachrurrozi, M.T
NIP. 198005222008121002




3. Penguji II

Anggina Primanita, M.IT
NIP.198908062015042002



Menyetujui,
a/n Ketua Jurusan Teknik Informatika
sekretaris jurusan


Rifkie Primartha, M.T
NIP.197706012009121004

MOTTO DAN PERSEMBAHAN

“Disaat semangat juang mengundur, aku teringat lagi bahwa kalian yang ku perjuangkan (keluarga)” – Dwi Astuti

“Tidak ada sesuatu yang tidak mungkin bagi Allah, semua mungkin terjadi, asalkan kita dengan keras terus mengejarnya. Pekerjaan sesulit apapun apabila kita menyukainya maka akan tampak mudah dan menggembirakan ”

Mario Teguh

Kupersembahkan karya tulis ini kepada:

- Allah SWT
- Bapak, Mamak, Kakak dan Adik tercinta
- Keluarga Besarku
- Seluruh dosen Teknik Informatika Fasilkom Unsri
- Seluruh sahabat IF Reguler angkatan 2010
- Almamater
- Staff operasional Fasilkom Unsri

PENGELOMPOKAN DOKUMEN BERITA
MENGGUNAKAN *FUZZY C MEANS*

Oleh:

Dwi Astuti

Nim: 09101002015

ABSTRAK

Berita disajikan dalam jumlah dan jenis yang banyak akan sulit dikelompokkan secara manual karena berita harus dibaca dengan teliti satu persatu. Hal ini yang kemudian menyebabkan pencari berita kesulitan dalam mencari informasi yang sesuai. Penelitian ini mengimplementasikan *Fuzzy C-Means* untuk mengelompokkan dokumen berita berbahasa Indonesia dengan melihat kesamaan isi kata dari sebuah dokumen. Metode *Fuzzy C-Means* mengelompokkan data ke dalam *cluster* berdasarkan nilai derajat keanggotaan sehingga memungkinkan data dapat memasuki lebih dari satu *cluster*. Prinsip pengelompokan ini adalah meminimumkan nilai dari fungsi objektif terhadap penentuan nilai derajat keanggotaan awal. Semakin besar nilai derajat keanggotaan dalam data suatu *cluster* maka semakin besar pula data tersebut menjadi anggota *cluster* tersebut. Hasil ujicoba terhadap 200 dataset yang diambil dari media online menunjukkan tingkat akurasi *f-measure* tertinggi yaitu 0.90. nilai akurasi yang didapat pada saat ujicoba menunjukkan hasil yang didapat sangat tergantung pada jumlah dari variasi kata yang digunakan dalam pembobotan.

Kata Kunci : pengelompokan berita, *fuzzy c-means*

CLUSTERING OF DOCUMENTS NEWS USING *FUZZY C MEANS*

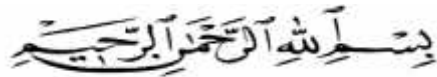
By:
Dwi Astuti
Nim: 09101002015

ABSTRACT

News is presented in the amount and type of which many will be difficult grouped manually for news should be read carefully one by one. This then led to a search of news difficulty in finding the appropriate information. This study implements Fuzzy C-Means to classify documents in Indonesian language news by viewing common content words in a document. Fuzzy C-Means method of classifying the data into clusters based on the degree of membership that allows data to be able to enter more than one cluster. The principle of this grouping is to minimize the value of the objective function to determine the value of the degree of initial membership. The greater the degree of membership in a cluster of data, the greater the data into the cluster members. Results of tests on 200 datasets extracted from online media shows the level of accuracy highest f-measure 0.98. accuracy values obtained during the experiment showed results obtained is highly dependent on the number of variations of the words used in weighting.

Keywords: news clustering, fuzzy c-means

KATA PENGANTAR



Puji dan syukur kehadiran Allah SWT atas segala nikmat, rahmat, dan karunia-Nya penulis dapat menyelesaikan tugas akhir ini. Tugas akhir yang berjudul **“Pengelompokan Dokumen Berita Menggunakan Fuzzy C-Means”** disusun untuk memenuhi salah satu persyaratan kelulusan tingkat sarjana (S1) pada Jurusan Teknik Informatika Fakultas Ilmu Komputer Universitas Sriwijaya.

Pada kesempatan ini, penulis ingin mengucapkan terima kasih kepada pihak-pihak yang telah memberikan dukungan, bimbingan, kerjasama, dan motivasi kepada penulis untuk menyelesaikan tugas akhir ini, antara lain:

1. Bapak, Mamak, Kakak Aris dan adik Naniky yang telah senantiasa memberikan doa restu dan dukungan bagi penulis selama kegiatan akademik maupun dalam proses penyusunan tugas akhir;
2. Bapak Jaidan Jauhari, M.T. selaku Dekan Fakultas Ilmu Komputer Universitas Sriwijaya;
3. Bapak Julian Supardi, M.T. selaku Ketua Jurusan Teknik Informatika Universitas Sriwijaya;
4. Bapak Rifkie Primartha, M.T. selaku dosen pembimbing I dan Bapak Julian Supardi, M.T. selaku dosen pembimbing yang telah memberikan bimbingan, ilmu pengetahuan, dukungan, kritik dan saran yang membangun kepada penulis dalam proses penyusunan tugas akhir ini;
5. Bapak M. Fachrurrozi, M.T. selaku dosen penguji I dan Ibu Anggina Primanita, M.IT. selaku dosen penguji II yang telah memberikan kritik dan saran kepada penulis dalam penyempurnaan tugas akhir ini;
6. Bapak Abdiansah, M.Cs. dan Bapak Danny Matthew Saputra, M.Sc. selaku dosen pembimbing akademik yang telah memberikan bimbingan dan saran kepada penulis selama kegiatan akademik;

7. Seluruh Bapak dan Ibu Dosen Jurusan Teknik Informatika Universitas Sriwijaya yang telah memberikan ilmu pengetahuan kepada penulis selama kegiatan akademik berlangsung;
8. Seluruh Karyawan/i Fasilkom Universitas Sriwijaya yang telah membantu penulis dalam urusan administrasi selama kegiatan akademik;
9. Tak lupa kepada sahabat-sahabatku, Penta Agustina, S.kom, Media Intan permata, Iora Ershita, Leni Safitri, S.kom, Erni Sulpah dan seluruh teman-teman rusunawa yang turut menyumbangkan ide, kerjasama, maupun solusi masalah yang dihadapi penulis selama penyusunan tugas akhir;
10. Seluruh teman seperjuangan dari IF Reguler 2010 A dan 2010 B yang telah memberikan motivasi, kerjasama, dan bantuan selama proses kegiatan akademik berlangsung;
11. Seluruh pihak yang telah membantu dalam penyusunan dan penyempurnaan tugas akhir ini yang tidak dapat penulis sebutkan satu-persatu.

Penulis menyadari bahwa tugas akhir ini jauh dari kata sempurna. Untuk itu, penulis mengharapkan kritik dan saran yang membangun untuk kesempurnaan tugas akhir ini. Semoga tugas akhir ini dapat memberikan manfaat kepada pihak-pihak yang membutuhkan.

Palembang, April 2016

Penulis

DAFTAR ISI

	Halaman
HALAMAN JUDUL	i
HALAMAN PENGESAHAN.....	ii
TANDA LULUS UJIAN SIDANG TUGAS AKHIR	iii
MOTTO DAN PERSEMBAHAN	iv
ABSTRACT	v
ABSTRAK	vi
KATA PENGANTAR.....	viii
DAFTAR ISI	ix
DAFTAR GAMBAR	xiii
DAFTAR TABEL	xv
BAB I PENDAHULUAN	
1.1 Latar Belakang	I-1
1.2 Rumusan Masalah	I-3
1.3 Tujuan Penelitian	I-4
1.4 Manfaat Penelitian	I-4
1.5 Batasan Masalah	I-4
1.6 Metodologi Penelitian	
1.6.1 Tahapan-tahapan penelitian	1-5
1.6.2 Metode pengumpulan data	
1.6.2.1 Jenis Data	1-5
1.6.2.2 Sumber Data	1-6
1.6.3 Metode pengembangan perangkat lunak.....	

1.6.3.1 Fase Insepsi	1-6
1.6.3.2 Fase Elaborasi	1-7
1.6.3.3 Fase Konstruksi.....	1-7
1.6.3.4 Fase Transisi	1-8
1.7 Sistematika penulisan	1-8

BAB II TINJAUAN PUSTAKA

2.1 Penelitian Terkait.....	II-1
2.2 Berita.....	II-2
2.2.1 Definisi Berita	II-2
2.2.2 Struktur Berita	II-2
2.3 Pengelompokan dokumen berita	II-4
2.4 <i>Text mining</i>	II-4
2.5 <i>Text Preprocessing</i>	II-5
2.6 Pembobotan Kata.....	II-7
2.7 Teknik pengelompokan.....	II-8
2.8 Logika Fuzzy.....	II-8
2.8.1 Pengertian Logika Fuzzy	II-9
2.8.2 Fuzzy Clustering	II-10
2.8.3 Fuzzy C-Means (FCM)	II-11
2.9 Analisis Deskripsi Persentase Perangkat Lunak	II-14
3.0 Metode RUP.....	II-17

BAB III ANALISA DAN PERANCANGAN

3.1 Analisis Masalah	III-1
----------------------------	-------

3.1.1 Analisis Data	III-
3.1.2 Analisis <i>Preprocessing</i>	III-2
3.1.3 Analisis Pembobotan Kata	III-5
3.1.4 Analisis Metode <i>Fuzzy C-Means</i>	III-7
3.2 Analisis Perangkat Lunak	
3.2.1 Deskripsi Umum Perangkat lunak	III-15
3.2.2 Analisis Kebutuhan perangkat lunak	III-17
3.2.3 Model <i>Use Case</i>	III-17
3.2.3.1 Diagram <i>Use Case</i>	III-17
3.2.3.2 Tabel Definisi Aktor	III-18
3.2.3.3 Tabel Definisi <i>Use Case</i>	III-19
3.2.3.4 Skenario <i>Use Case</i>	III-19
3.2.3.5 Kelas Analisis	III-22
3.2.3.6 <i>Sequence Diagram</i>	III-24
3.2.3.7 Kelas Diagram.....	III-27
3.3 Perancangan Perangkat Lunak	III-30
3.4 Perancangan Antarmuka	III-30
BAB IV IMPLEMENTASI DAN PENGUJIAN	
4.1 Implementasi perangkat lunak	IV-1
4.1.1 Lingkungan Implementasi.....	IV-1
4.1.2 Implementasi Kelas	IV-2
4.1.3 Implementasi Antarmuka	IV-4
4.2 Pengujian Perangkat lunak.....	IV-4
4.2.1 Lingkungan Pengujian	IV-4

4.2.2 Rencana Pengujian	IV-5
4.3 Kasus Uji.....	IV-7
4.4 Hasil Pengujian Perangkat Lunak	IV-10
4.4.1 Hasil pengujian <i>Use Case</i> melakukan browser	
memilih dokumen	IV-10
4.4.2 Hasil pengujian <i>Use Case</i> melakukan <i>pra-processing</i>	IV-10
4.4.3 Hasil pengujian <i>Use Case</i> melakukan <i>Clustering</i>	IV-11
4.4.4 Hasil Pengujian pengelompokan berita.....	IV-12
BAB V KESIMPULAN DAN SARAN	
5.1 Kesimpulan	V-1
5.2 Saran	V-1
DAFTAR PUSTAKA	xvi
Lampiran 1. KODING PROGRAM.....	xviii
Lampiran 2. HASIL PENGUJIAN.....	xxxix

DAFTAR GAMBAR

	Halaman
Gambar II-1. Struktur Berita	II-3
Gambar II-2. Contoh tahap <i>case folding</i>	II-5
Gambar II-3. Contoh tahap <i>tokenizing</i>	II-6
Gambar II-4. Contoh tahap <i>filtering</i>	II-6
Gambar II-5. Struktur RUP (Kruchen, 2003)	II-17
Gambar III-1 Skema Diagram Proses Kerja Perangkat Lunak	III-15
Gambar III-2 Diagram <i>Use Case</i>	III-18
Gambar III-3 Kelas Analisis <i>Browser</i> Memilihi Dokumen	III-22
Gambar III-4 Kelas Analisis Melakukan <i>praprocessing</i>	III-23
Gambar III-5 Kelas Analisis Melakukan <i>Clustering</i>	III-23
Gambar III-6 <i>Sequence Diagram Browser</i> Memilihi Dokumen	III-24
Gambar III-7 <i>Sequence Diagram</i> Melakukan <i>praprocessing</i>	III-25
Gambar III-8 <i>Sequence Diagram</i> Melakukan <i>Custering</i>	III-26

Gambar III-9 Kelas Diagram <i>UseCase</i> keseluruhan.....	III-27
Gambar III-10 Kelas Diagram <i>UseCase browser</i> kumpulan dokumen	III-28
Gambar III-11 Kelas Diagram <i>UseCase</i> melakukan <i>preprocessing</i>	III-28
Gambar III-12 Kelas Diagram <i>Use Case</i> melakukan <i>Clustering</i>	III-29
Gambar III-13 Rancangan Antarmuka Form Utama	III-30
Gambar IV-1 Antarmuka perangkat lunak.....	IV-4
Gambar IV-2 Hasil Pengujian <i>Use Case</i> browser	
memilih dokumen U-1-101	IV-10
Gambar IV-3 Hasil Pengujian <i>Use Case</i> Melakukan	
<i>pra-processing</i> U-1-101	IV-11
Gambar IV-4 Hasil Pengujian <i>Use Case</i> Melakukan	
Clustering U-1-101	IV-11
Gambar IV.5 Grafik Hasil Uji Tingkat Akurasi Pengelompokan Data	IV-13

DAFTAR TABEL

	Halaman
Tabel III-1. <i>Tokenizing</i>	III-3
Tabel III-2. <i>Filtering</i>	III-4
Tabel III-3 Perhitungan TF IDF	III-6
Tabel III-4. Data-data yang akan dicluster	III-8
Tabel III-5. Matriks Keanggotaan	III-8
Tabel III-6 Cluster Pusat	III-9
Tabel III-7. Hasil Cluster Pusat	III-11
Tabel III-8. Fungsi Objektif	III-
11Tabel III-9 Perubahan Matriks U	III-13
Tabel III-10. Hasil Pengelompokan	III-14
Tabel III-11. Kebutuhan Fungsional	III-17
Tabel III-12. Kebutuhan Non Fungsional	III-17
Tabel III-13. Definisi Aktor	III-18
Tabel III-14. Definisi <i>Use Case</i>	III-18

Tabel III-15. Skenario <i>Use Case</i> browser Kumpulan Dokumen	III-19
Tabel III-16 Skenario <i>Use Case</i> Melakukan <i>preprocessing</i>	III-20
Tabel III-17 Skenario <i>Use Case</i> Melakukan <i>Clustering</i>	III-21
Tabel IV-1 Daftar Implementasi Kelas	IV-2
Tabel IV-2 Rencana Pengujian <i>Use Case</i> Melakukan browser	
memilih dokumen	IV-5
Tabel IV-3 Rencana Pengujian <i>Use Case</i> Melakukan <i>pra-processing</i>	IV-5
Tabel IV-4 menunjukkan rencana pengujian <i>use case clustering</i>	IV-6
Tabel IV-5 Pengujian <i>Use Case</i> Melakukan <i>browser</i> memilih dokumen	IV-7
Tabel IV-6 Pengujian <i>Use Case</i> Melakukan <i>pra-processing</i>	IV-8
Tabel IV-7 Pengujian <i>Use Case</i> Melakukan <i>clustering</i>	IV-9

BAB I

PENDAHULUAN

1.1 Latar Belakang

Berita sudah merupakan kebutuhan bagi setiap orang. Hal ini dikarenakan manusia membutuhkan informasi. Menurut Budiman. K (2005) berita adalah laporan mengenai suatu peristiwa atau kejadian yang terbaru, laporan mengenai fakta-fakta yang aktual, menarik perhatian, dinilai penting dan luar biasa. Seseorang dapat mengakses sebuah berita melalui berbagai media, antara lain: media cetak, elektronik, dan internet.

Penyebaran berita sekarang ini telah berkembang sangat pesat, perkembangan sarana penyebaran berita ini dipengaruhi oleh faktor bertambahnya jumlah manusia yang haus akan informasi baru. Salah satu media penyebaran berita yang banyak diminati adalah media *online*. Berita penyebaran menggunakan media *online* menjadi sangat populer saat ini, sehingga penyebarannya begitu pesat. Terbukti bertambah banyaknya situs perusahaan maupun situs personal yang tersedia, sehingga semakin meningkatnya jumlah informasi dan data.

Pekembangan informasi begitu cepat semestinya dapat memberikan manfaat bagi pengguna. Keterbatasan waktu dan perangkat lunak menyebabkan pengguna tidak mudah untuk mengakses setiap informasi yang berkaitan. Dari permasalahan di atas diperlukan suatu

teknik pengelompokan dokumen berita. tersebut, sehingga dapat membantu pengguna dalam memahami informasi yang ada didalamnya.

Teknik pengelompokan adalah suatu proses metode pembelajaran untuk mengkategorikan kumpulan dari dokumen kedalam kelasnya berdasarkan kontennya (Isa, Kallimaniand Lee, 2009). Pengelompokan dokumen telah banyak dilakukan penelitian dengan berbagi metode, diantaranya *k-means* (Congnan, Li and Chung, 2009), *K-NearestNeighbor* (KNN) (Buana, Janet and Outra, 2012), *naïve Bayes* (Mc Callum and Nigam, 2003) dan *Fuzzy C-means* (Yudha & Sandhi, 2012).

Algoritma *k-means* berdasarkan pada ide bahwa *centroid* dapat merepresentasikan suatu kelompok. Proses optimasi pengelompokan dokumen ke k cluster sehingga fungsi kriteria tertentu diminimalkan atau dimaksimalkan. Jika dua dokumen cukup mirip, mereka dianggap sebagai tetangga satu sama lain. Dan hubungan antar dua dokumen menunjukkan jumlah kesamaan tetangga mereka (Congnan, Li and Chung, 2009).

Algoritma *K-NearestNeighbor* (KNN) merupakan teknik pelatihan dan pengklasifikasian dengan mencari grup objek k yang dekat dengan objek t es pada saat pelatihan. KNN memiliki perhitungan yang kompleks untuk mengetahui tetangga terdekat dari objek k. Jika menggunakan data latih yang sedikit maka waktu pengklasifikasian tidak signifikan, namun jika data latih yang berjumlah besar maka KNN akan membutuhkan waktu yang lama untuk menghitung kesamaan (Buana, Jannet and Putra, 2012).

Algoritma *naïve Bayes* sangat sering digunakan karena kesederhanaannya. *naïve Bayes* merupakan klasifikasi probabilistik sederhana dengan asumsi independensi yang kuat. Berdasarkan asumsi independensi ini, parameter setiap atribut dapat belajar terpisah, hal ini sangat menyederhanakan pembelajaran terutama ketika jumlah atributnya banyak (McCallum and Nigam, 2003).

FCM merupakan suatu teknik pengclusteran yang mana keberadaan tiap titik data dalam suatu *cluster* ditentukan oleh derajat keanggotaan. Dalam pengelompokan dokumen dengan menggunakan metode ini dibutuhkan proses *lowercasing*, penghilangan tanda baca dan angka, tokenisasi, stopwords, dan pembobotan term. FCM menggunakan model pengelompokan fuzzy sehingga data dapat menjadi anggota dari semua kelas atau tingkat keanggotaan yang berbeda antara 0 hingga 1 (Yudha & Sandhi, 2012).

FuzzyC-means sangat sering digunakan dalam pengelompokan dalam bidang citra dan masih jarang diterapkan dalam bidang pemrosesan text, sebab itu penulis akan melakukan penelitian tentang pengelompokan dokumen berita berbahasa Indonesia menggunakan *FuzzyC-means*.

1.2 Rumusan Masalah

Pengelompokan dokumen berita secara manual merupakan hal yang sulit dikarenakan jumlah kata yang banyak sehingga untuk mengelompokkannya harus dibaca dengan teliti atau persatu. Karena kondisi tersebut menyebabkan waktu yang diperlukan menjadi cukup lama. Sehingga masalah yang akan diangkat pada penelitian ini adalah bagaimana menerapkan metode *Fuzzy c-means* dalam.

DAFTAR PUSTAKA

- Arikunto, Suharsimi, 2010. *Prosedur Penelitian Suatu Pendekatan Praktek*, Jakarta: Penerbit Rineka Cipta.
- Bezdek, James C, 1981. *Pattern Recognition with Fuzzy Objective Function Algorithms*. ISBN 0-306-40671-3
- Budiman, K. 2005. *Dasar-Dasar Jurnalistik. Pelatihan Jurnalistik-info jawa* 12-15 desember 2005. In [http://pelitaku.sabda.org/dasar dasar jurnalistik](http://pelitaku.sabda.org/dasar_dasar_jurnalistik). Accessed in april, 6 2015
- Buana, P.W., Jannet, S., I.K.G.D. 2012. Combination of K-Nearest Neighbor and K-Means based on Term Re-weighting for Classify Indonesian News. *International Journal of Computer Applications* (0975-8887) Volume 50 – No. 11, July 2012
- Februariyanti, H., & Zuliarso, E. (2012). *Klasifikasi Dokumen Berita Teks Bahasa Indonesia menggunakan Ontologi*. *Dinamik-Jurnal Teknologi Informasi*. Universitas Teknologi Surabaya
- Februariyanti, H., Zuliarso, E., Anwaristyati, R., Hendriatmono, G., & Jupriyanto, J. (2011). *Klastering Dokumen Berita dari Web menggunakan Algoritma Single Pass Clustering*. Surabaya : Universitas Teknologi Surabaya
- Isa, Kallimani and Lee, 2009. Using the self organizing map for clustering of text documents. *An International Journal* Volume 36 Issue 5 July 2009
- Joviani, Amelia dan Syamsuryadi 2014. *Pengelompokan Dokumen Berita Berbahasa Indonesia Menggunakan Self Organizing Maps*. Laporan Tugas Akhir. FAKULTAS ILMU KOMPUTER UNSRI, Universitas Sriwijaya.
- Juwito, 2008. *Menulis Berita Dan Feature's*. Penerbit : Unesa University Press
- Kusumadewi, Sri dan Purnomo, Hari. (2004). *Aplikasi Logika Fuzzy untuk Pendukung Keputusan* (edisi pertama). Yogyakarta: Graha Ilmu. Zadeh 1965
- Kruchten, P. 2003. *The Rational Unified Process: An Introduction, Third Edition*. Addison-Wesley

- Lin, Ching-Teng; Lee, George. 1996. *Neural Fuzzy Systems : A Neuro-Fuzzy Synergism to Intelligent Systems*. United States of America. Prentice Hall International Inc.
- McCallum and Nigam, 2003. *A Comparison of Event Models for Naive Bayes Text Classification*. School of Computer Science Carnegie Mellon University Pittsburgh, PA 15213
- Mustopa, A.. 2009. *Klasifikasi Otomatis Dokumen Berita Kejadian Berbahasa Indonesia*. Skripsi. FAKULTAS SAINS dan TEKNOLOGI universitas Islam Negeri (UIN) Maulana Malik Ibrahim Malang
- Mooney, J. 2006. *CS 391L : Machine Learning Text Categorization*. Austin : University of Texas.
- Pang, B., & Lee, L. (2008). *Opinion mining and sentiment analysis*.
- Ron Papka. "On-Line New Event Detection, Clustering, and Tracking". Ph. D dissertation on University of Massachusetts, 1999
- Ross, T.J., 2010, *Fuzzy Logic With Engineering Application Thrid Edition*, Desktop Web, Graha Media, Yogyakarta
- Sree, K.P.N.V. Satya and Murthy, J.V.R. 2012. Clustering Based On Cosine Similarity Measure. *Internasional Journal of Engineering Science Advanced Technology* Volue-2, Issue-3, 508-512
- Trisedya, B. D. and Hardinal, J. 2009. *Klasifikasi Dokumen Menggunakan Naïve Bayes Classifier*. Jurusan Teknik Informatika FAKULTAS ILMU KOMPUTER, Universitas Indonesia.
- Yang, Y., Carbonell, J. G., Brown, R. D., Pierce, T., Archibald, B. T., & Liu, X. (1999). Learning approaches to topic detection and tracking.
- Yanjun Li, Congnan Luo, and S. M. Chung, "A Parallel Text Document Clustering Algorithm Based on Neighbors," *Cluster Computing*, April 2015, DOI 10.1007/s10586-015-0450-z, Springer US.
- Yudha C, Sandhi, 2012. *Clustering Dokumen Bahasa Indonesia dengan Menggunakan Fuzzy C-Means*. Universitas Telkom. Bandung
- Zadeh, L. (1965). Fuzzy sets. *Information and Control*, 8, 338 – 353.

TaryanaSuryana, 2007. *Metode RUP (Rational Unfied Process)* .STMIK LIKMI. Bandung

Zhang,W., Yoshida, T. and Tang, X., 2008. TF-IDF, LSI and Multi-word in Information Retrieval and Text Categorization. *IEEE International Conference on System, Mon and Cybernectics*.