

**KATEGORISASI DOKUMEN TEXT MENGGUNAKAN METODE
K-NEAREST NEIGHBOR PADA DOKUMEN TUGAS AKHIR
UNIVERSITAS MUHAMMADIYAH JEMBER**

TUGAS AKHIR

Diajukan Untuk Memenuhi Persyaratan Guna Meraih Gelar Sarjana Komputer
Teknik Informatika Universitas Muhammadiyah Jember



JA'FAR SHODIQ

1110651085

**PROGRAM STUDI TEKNIK INFORMATIKA
FAKULTAS TEKNIK
UNIVERSITAS MUHAMMADIYAH JEMBER**

2017

HALAMAN PENGESAHAN

**KATEGORISASI DOKUMEN TEXT MENGGUNAKAN METODE
K-NEAREST NEIGHBOR PADA DOKUMEN TUGAS AKHIR
UNIVERSITAS MUHAMMADIYAH JEMBER**

Ja'far Shodiq

1110651085

Telah mempertanggung jawabkan Laporan Tugas Akhirnya pada Sidang Tugas Akhir Tanggal 12 Juni 2017 sebagai salah satu syarat kelulusan dan mendapatkan gelar Sarjana Komputer (S.Kom) di Universitas Muhammadiyah Jember

Disetujui Oleh:

Dosen Pembimbing:

Lutfi Ali Muharom, S.Si, M.Si

NPK. 10 09 550

Dosen Penguji I:

Dosen Penguji II:

Ulya Anisatur R., S.Kom, M.Kom

Yeni Dwi Rahayu, S.ST, M.Kom

NPK. 12 03 705

NPK. 11 03 590

Mengesahkan,

Mengetahui,

Dekan Fakultas Teknik

Ketua Program Studi Teknik Informatika

Ir. Suhartinah, M.T

Yeni Dwi Rahayu, S.ST, M.Kom

NPK. 95 05 246

NPK. 11 03 590

KATA PENGANTAR

Puji syukur kehadirat Allah SWT atas limpahan rahmat dan hidayah sehingga penelitian dengan judul “Kategorisasi Dokumen Text Menggunakan Metode K-Nearest Neighbor Pada Tugas Akhir Universitas Muhammadiyah Jember” dapat terselesaikan. Penulis menyadari bahwa dalam penelitian maupun penulisan laporan ini banyak pihak yang telah membantu menyelesaikannya. Maka dari itu, saya ucapkan terima kasih kepada:

1. Ibu Ir. Suhartinah, ST., MT., selaku Dekan Fakultas Teknik Universitas Muhammadiyah Jember
2. Ibu Yeni Dwi Rahayu, S.ST., M.Kom., selaku Ketua Jurusan Teknik Informatika Fakultas Teknik Universitas Muhammadiyah Jember
3. Bapak Lutfi Ali Muharrom, S. Si, M.Si., selaku dosen pembimbing yang telah banyak memberi petunjuk dan arahan
4. Ibu Ulya Anisatur Rosyidah, S.Kom., M.Kom dan Ibu Yeni Dwi Rahayu, S.ST., M.Kom., selaku dosen penguji

Jember,..

Penulis

DAFTAR ISI

JUDUL.....	i
HALAMAN PENGESAHAN.....	ii
HALAMAN PERNYATAAN.....	iii
MOTTO.....	iv
HALAMAN PERSEMBAHAN.....	v
KATA PENGANTAR.....	vi
ABSTRAK.....	vii
ABSTRACT.....	viii
DAFTAR ISI.....	ix
DAFTAR TABEL.....	xi
DAFTAR GAMBAR.....	xii
DAFTAR RUMUS.....	xiii
LAMPIRAN.....	xiv
BAB I LATAR BELAKANG.....	1
1.1 Latar Belakang Masalah	1
1.2 Rumusan Masalah	2
1.3 Batasan Masalah	2
1.4 Tujuan Penelitian	3
1.5 Manfaat Penelitian	3
BAB II TINJUAN PUSTAKA.....	4
2.1. Dokumen Teks.....	4
2.2. Preprosesing Dokumen	4

2.3. K-Nearest Neighbor.....	6
2.4. Penelitian Sebelumnya.....	8
BAB III METODOLOGI.....	9
3.1. Metodologi Penelitian	9
3.2. Deskripsi Data.....	9
3.3. Arsitektur Sistem	9
3.4. Tahapan Uji Coba.....	11
BAB IV HASIL DAN PEMBAHASAN.....	28
4.1 Uji Coba.....	28
4.2 Hasil Pengujian	28
4.3 Analisa Hasil Penelitian.....	31
4.4 Pembahasan.....	32
BAB V KESIMPULAN DAN SARAN.....	33
3.1. Kesimpulan.....	33
3.2. Saran.....	33
DAFTAR PUSTAKA.....	34

DAFTAR TABEL

Tabel 3.1 Contoh Data Latih.....	12
Tabel 3.2 Contoh Data Uji.....	16
Tabel 3.3 Hasil tf danIdf Dokumen.....	18
Tabel 3.4 Hasil Bobot Setiap Term.....	19
Tabel 3.5 Hasil Jarak Dokumen.....	21
Tabel 3.6 Hasil Kategorisasi dengan k=7.....	25
Tabel 3.7 Hasil Kategorisasi dengan k=3.....	26
Tabel 3.8 Hasil Kategorisasi dengan k=5.....	27
Tabel 4.1 Deskripsi Sekenario Data.....	28
Tabel 4.7 Hasil Kategorisasi k=3.....	29
Tabel 4.7 Hasil Pengujian Skenario 1 dengan k=3.....	29
Tabel 4.7 Hasil Kategorisasi k=5.....	30
Tabel 4.8 Hasil Pengujian Skenario 2 dengan k=5.....	30
Tabel 4.7 Hasil Kategorisasi k=7.....	31
Tabel 4.9 Hasil Pengujian Skenario 3 dengan k=7.....	31
Tabel 5.0 Akurasi.....	32

DAFTAR GAMBAR

Gambar 3.1 Arsitektur Sistem.....	10
-----------------------------------	----

DAFTAR RUMUS

Rumus 2.1 Perhitungan Bobot Dokumen.....	5
Rumus 2.2 Perhitungan K-Nearest Neighbor.....	5
Rumus 2.3 Perhitungan Akurasi.....	6

LAMPIRAN

Preprosesing Data Lanjutan..... 35

DAFTAR PUSTAKA

- G. Toker and Ö. Kırmemi , “TEXT CATEGORIZATION USING K-NEARESTNEIGHBOR CLASSIFICATION.”
- H. Februariyanti and E. Zuliarso, “Klasifikasi Dokumen Berita Teks Bahasa Indonesia menggunakan Ontologi,” J. Teknol. Inf. Din., vol. 17, no. 1, pp. 14–23, 2012.
- H. Muhammad Isa , “Klasifikasi Dokumen Teks Menggunakan Metode Support Vector Mechine Dengan Pemilihan fitur Chi-Square” , Jember, 2016.
- J. Samodra, S. Sumpeno, and M. Hariadi, “Klasifikasi Dokumen Teks Berbahasa Indonesia dengan Menggunakan Naïve Bayes,” Semin. Nas. Electr. Informatic, IT’s Educ., pp. 1–4, 2009
- Kusrini, Emha Taufiq Luthfi, “Algoritma Data Mining”, Yogyakarta, Andi, 2009.
- R. Rakhmat Sani , J. Zeniarja , A. Luthfiarta “Penerapan Algoritma K-Nearest Neighbor pada Information Retrieval dalam Penentuan Topik Referensi Tugas Akhir” Journal of Applied Intelligent System, Vol. 1, No. 2, Juni 2016: 123 – 133
- Zhang, X.F, Huang, H.Y, Zhang K.L. 2009. KNN Text Categorization Algorithm Based on Semantic Centre. 2009 International Conference on Information Technology and Computer Science.