

TUGAS AKHIR

**KLUSTERISASI JUDUL BERITA PADA WEBSITE DETIK.COM
MENGUNAKAN ALGORITMA KMEANS**



**ADHIF ADHITAMA T A
1810651116**

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

2024

TUGAS AKHIR

**KLUSTERISASI JUDUL BERITA PADA WEBSITE DETIK.COM
MENGUNAKAN ALGORITMA KMEANS**

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



ADHIF ADHITAMATA

1810651116

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

2024

HALAMAN PERSETUJUAN SIDANG TUGAS AKHIR

**KLUSTERISASI JUDUL BERITA PADA WEBSITE DETIK.COM
MENGUNAKAN ALGORITMA KMEANS**

Oleh :

ADHIF ADHITAMA T A

1810651116

Telah disetujui bahwa Laporan Tugas Akhir ini untuk diajukan pada sidang Tugas Akhir sebagai salah satu syarat kelulusan dan mendapatkan gelar Sarjana Komputer (S.Kom)

di
Universitas Muhammadiyah Jember

Disetujui oleh,

Pembimbing I

Pembimbing II



Svarif Hidayatullaah, S.Kom., M.Kom

Miftahur Rahman, S.Kom., M.Kom

NIDN. 0723049203

NIDN. 0724039201

HALAMAN PENGESAHAN

KLUSTERISASI JUDUL BERITA PADA WEBSITE DETIK.COM
MENGUNAKAN ALGORITMA KMEANS

Oleh :

ADHIF ADHITAMA T A

1810651116

Telah mempertanggung jawabkan Laporan Tugas Akhir pada sidang Tugas Akhir
tanggal 2024 sebagai salah satu syarat kelulusan dan mendapatkan gelar Sarjana
Komputer (S.Kom)

di

Universitas Muhammadiyah Jember

Disetujui oleh,

Penguji I

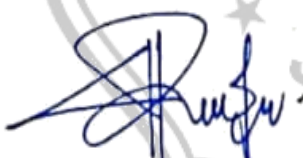
Pembimbing I


Ari Eko Wardono, ST.,M.Kom
NIDN. 0014027501


Syarif Hidayatullah, S.Kom., M.Kom
NIDN. 0723049203

Penguji II

Pembimbing II


Wiwik Suharso, ST.,M. Kom
NIDN. 0006097601


Miftahur Rahman, S.Kom., M.Kom
NIDN. 0724039201

Dekan Fakultas Teknik

Ketua Program Studi Teknik Informatika


Prof. Dr. Ir. Nanang Saiful Rizal, MT., IPM.
NIDN. 0705047806

ii


Rosita Yandarti, S.Kom., M.CS
NIDN. 0629018601

PERNYATAAN

Yang bertanda tangan di bawah ini :

Nama : ADHIF ADHITAMA T A
NIM : 1810651116
Program Studi : Teknik Informatika
Perguruan Tinggi : Universitas Muhammadiyah Jember

Menyatakan bahwa Tugas Akhir yang berjudul “**KLUSTERISASI JUDUL BERITA PADA WEBSITE DETIK.COM MENGGUNAKAN ALGORITMA KMEANS**” Tugas akhir ini tidak boleh merupakan hasil karya orang lain, baik sebagian maupun seluruhnya, kecuali jika dikutip secara benar.

Oleh karena itu, surat pernyataan ini dibuat dengan sejujurnya dan apabila dirasa tidak benar, penulis siap menghadapi konsekuensi akademis.

Jember, 8 July 2024



ADHIF ADHITAMA T A

KATA PENGANTAR

Puji syukur kami panjatkan kehadirat Allah SWT, Pencipta Alam Semesta, yang telah melimpahkan rahmat dan rezekinya kepada penulis sehingga penulis dapat menyelesaikan skripsi ini sesuai dengan persyaratan gelar sarjana. Komputer (S1) Fakultas Teknik Informatika Universitas Muhammadiyah Jember.

Penulisan skripsi ini tentunya bukan tanpa cela, karena penulis sangat menyadari keterbatasan pengetahuan dan kemampuannya. Penulis dapat menyelesaikan skripsi ini berkat doa, bantuan, dan dukungan dari banyak pihak, baik moril maupun materii.

Penulis mengucapkan terima kasih kepada semua pihak atas dukungan dan bantuannya dalam penyusunan skripsi ini. Penulis bertujuan agar temuan penelitian ini dapat bermanfaat bagi pembaca dan khususnya Universitas Muhammadiyah Jember. Pastikan untuk menunjukkan penghargaan kepada penulis :

1. Bapak Prof. Dr. Ir. Nanang Saiful Rizal, MT., IPM. Selaku Dekan Fakultas Teknik Universitas Muhammadiyah Jember.
2. Ibu Rosita Yanuarti, S.Kom., M.CS selaku Ketua Program Studi Teknik Informatika
3. Bapak Syarif Hidayatullaah, S.Kom., M.Kom Selaku dosen pembimbing 1
4. Bapak Miftahur Rahman, S.Kom., M.Kom Selaku Dosen Pembimbing 2
5. Bapak Ari Eko Wardoyo, ST.,M.Kom Selaku Dosen Penguji 1
6. Bapak Wiwik Suharso, ST.,M.Kom Selaku Dosen Penguji 2
7. Penulis mengucapkan terima kasih kepada para dosen dan pegawai Universitas Muhammadiyah Jember atas bimbingan, arahan, dan dorongannya selama proses penyusunan skripsi.

Akhir kata penulis ingin mengucapkan terima kasih kepada semua pihak yang telah berkontribusi dalam penyelesaian skripsi ini, termasuk pihak-pihak yang tidak dapat disebutkan namanya satu per satu. Saya berharap penelitian ini bermanfaat bagi semua pembaca Amin.

Jember, 8 July 2024



ADHIF ADHITAMA T A



DAFTAR ISI

ABSTRAK	iv
ABSTRACT	v
KATA PENGANTAR	vi
DAFTAR ISI	viii
DAFTAR GAMBAR	x
DAFTAR TABEL	xi
BAB I PENDAHULUAN	1
1.1 Latar Belakang	1
1.2 Rumusan Masalah	2
1.3 Tujuan Masalah	2
1.4 Manfaat Penelitian	2
1.5 Batasan Penelitian	3
BAB II TINJAUAN PUSTAKA	4
2.1 Teori Tentang Topik Skripsi	4
2.1.1. <i>Text Mining</i>	4
2.1.2. <i>Clustering</i>	9
2.1.3. TF-IDF	13
2.2 Teori Tentang Tools / Software Yang Digunakan	14
2.2.1 <i>Google Colabs</i>	14
2.2.2 <i>Python</i>	16
2.3 Penelitian Terdahulu	16
BAB III METODOLOGI PENELITIAN	18
3.1. Tahapan penelitian	18
3.2. Pengumpulan Data dan Analisis Kebutuhan	19
3.3. Preprocessing Dokumen	19
3.3.1 <i>Case Folding</i>	19
3.3.2 <i>Stopword Removal</i>	20
3.3.3 <i>Lemmatization</i>	21
3.3.4 <i>Stemming</i>	21

3.3.5	<i>Tokenizing</i>	22
3.4.	<i>Term Weighting</i>	22
3.5.	<i>Proses Clustering</i>	24
3.6.	<i>Evaluasi</i>	25
BAB IV HASIL DAN PEMBAHASAN		28
4.1.	<i>Pembahasan</i>	28
4.1.1.	<i>Text Preprocessing</i>	28
4.1.2	<i>Case Folding</i>	29
4.1.3	<i>Stopword Removal</i>	31
4.1.4	<i>Lemmatization</i>	33
4.1.4	<i>Stemming</i>	34
4.1.5	<i>Tokenizing</i>	36
4.1.6	<i>Term-weighting</i>	38
BAB V KESIMPULAN DAN SARAN		51
5.1	<i>Kesimpulan</i>	51
5.2	<i>Saran</i>	51
DAFTAR PUSTAKA		52



DAFTAR GAMBAR

Gambar 2. 1 Tahapan <i>Text Mining</i>	4
Gambar 2. 2 <i>Source Code Kmeans</i>	13
Gambar 2. 3 Menjalankan Kode di <i>Google Colab</i>	15
Gambar 3. 1 Contoh Pengelompokan Berdasarkan Bentuk Dengan <i>Kmeans</i>	18
Gambar 3. 2 Diagram Alur <i>Preprocessing</i> Dokumen.....	19
Gambar 3. 3 Ilustrasi Pembentukan Kluster dengan Metode <i>Kmeans</i>	25
Gambar 4. 1 <i>Import Library</i>	29
Gambar 4. 2 <i>Source Code Casefolding</i>	30
Gambar 4. 3 <i>Source Code Stopwords</i>	32
Gambar 4. 4 <i>Source Code Lemmatizer</i>	33
Gambar 4. 5 <i>Source Code Stemming</i>	35
Gambar 4. 6 <i>Source Code Tokenizing</i>	37
Gambar 4. 7 <i>Source Code Kmeans</i>	45
Gambar 4. 8 Menampilkan Hasil <i>Cluster</i>	46
Gambar 4. 9 Menampilkan <i>WordCloud</i>	46
Gambar 4. 10 <i>Cluster 1</i>	47
Gambar 4. 11 <i>Cluster 2</i>	48
Gambar 4. 12 <i>Cluster 3</i>	48
Gambar 4. 13 <i>Cluster 4</i>	49
Gambar 4. 14 <i>Cluster 5</i>	50
Gambar 4. 15 Grafik Rata-Rata Jumlah yang sering keluar dalam cluster judul berita Masing-masing cluster memiliki judul dengan jumlah kata yang paling sering muncul.....	50

DAFTAR TABEL

Tabel 2. 1 Contoh <i>Tokenizing</i>	6
Tabel 2. 2 Contoh <i>Stopwords</i>	6
Tabel 2. 3 Contoh <i>lemmatization</i>	7
Tabel 2. 4 Penelitian Terdahulu.....	16
Tabel 3. 1 Hasil <i>Case Folding</i>	20
Tabel 3. 2 Hasil <i>Stopword</i>	20
Tabel 3. 3 Hasil <i>Lemmatization</i>	21
Tabel 3. 4 Hasil <i>Stemming</i>	21
Tabel 4. 1 Dataset	29
Tabel 4. 2 Hasil <i>Case Folding</i>	30
Tabel 4. 3 Hasil <i>Stopword</i>	32
Tabel 4. 4 Hasil <i>Lemmatization</i>	34
Tabel 4. 5 Hasil <i>Stemming</i>	35
Tabel 4. 6 Hasil <i>Tokenizing</i>	37
Tabel 4. 7 Hasil <i>Term Dari Preprocessing</i>	38
Tabel 4. 8 Hasil Perhitungan TF.....	39
Tabel 4. 9 Hasil Perhitungan IDF.....	41
Tabel 4. 10 Hasil Perhitungan TF*IDF	43
Tabel 4. 11 Hasil <i>Cluster 1</i>	46
Tabel 4. 12 Hasil <i>Cluster 2</i>	47
Tabel 4. 13 Hasil <i>Cluster 3</i>	48
Tabel 4. 14 Hasil <i>Cluster 4</i>	49
Tabel 4. 15 Hasil <i>Cluster 5</i>	49