



SKRIPSI

**ANALISIS SENTIMEN BERBASIS ASPEK
PADA SISTEM LAYANAN PENGADUAN
MASYARAKAT DI KOTA SURABAYA
MENGUNAKAN METODE LATENT
DIRICHLET ALLOCATION DAN NAÏVE BAYES**

FERDY ATMAJA
NPM 20082010118

DOSEN PEMBIMBING
Dr. Eng. Agussalim, MT.
Eka Dyar Wahyuni, S.Kom., M.Kom.

**KEMENTERIAN PENDIDIKAN, KEBUDAYAAN, RISET, DAN TEKNOLOGI
UNIVERSITAS PEMBANGUNAN NASIONAL VETERAN JAWA TIMUR
FAKULTAS ILMU KOMPUTER
PROGRAM STUDI INFORMATIKA
SURABAYA
2024**

Halaman ini sengaja dikosongkan

LEMBAR PENGESAHAN

ANALISIS SENTIMEN BERBASIS ASPEK PADA SISTEM LAYANAN PENGADUAN MASYARAKAT DI KOTA SURABAYA MENGGUNAKAN METODE LATENT DIRICHLET ALLOCATION DAN NAÏVE BAYES


Oleh :

FERDY ATMAJA

NPM. 20082010118

Telah dipertahankan dihadapan dan diterima oleh Tim Penguji Skripsi Prodi Sistem Informasi Fakultas Ilmu Komputer Universitas Pembangunan Nasional Veteran Jawa Timur Pada tanggal 13 Desember 2024.

Dr. Eng. Agussalim, MT.
NIP. 19850811 2019031 005


..... (Pembimbing I)


Eka Dyar Wahyuni, S.Kom., M.Kom.
NIP. 19841201 2021212 005


..... (Pembimbing II)

Nur Cahyo Wibowo, S.Kom., M.Kom.
NIP. 19790317 2021211 002


..... (Ketua Penguji)


Eristya Maya Safitri, S.Kom., M.Kom.
NIP. 19930316 2019032 020


..... (Anggota Penguji II)

Reisa Permatasari, S.T., M.Kom.
NIP. 19920514 2022032 007


..... (Anggota Penguji III)

Mengetahui,
Dekan Fakultas Ilmu Komputer


Prof. Dr. Ir. Novirina Hendrasarie, MT
NIP. 19681126 199403 2 001

Halaman ini sengaja dikosongkan

LEMBAR PERSETUJUAN

**ANALISIS SENTIMEN BERBASIS ASPEK PADA SISTEM LAYANAN
PENGADUAN MASYARAKAT DI KOTA SURABAYA MENGGUNAKAN
METODE LATENT DIRICHLET ALLOCATION DAN NAÏVE BAYES**

Oleh:
FERDY ATMAJA
NPM. 20082010118

Telah disetujui untuk mengikuti ujian skripsi

Menyetujui,

Dosen Pembimbing I



Dr. Eng. Agussalim, MT.

NIP. 19850811 2019031 005

Dosen Pembimbing II



Eka Dyar Wahyuni, S.Kom., M.Kom.

NIP. 19841201 2021212 005

Koordinator Skripsi

Prodi Sistem Informasi Fakultas Ilmu Komputer



Eristya Maya Safitri, S.Kom., M.Kom.

NIP. 19930316 2019032 020

Halaman ini sengaja dikosongkan

SURAT PERNYATAAN ORISINALITAS

Yang bertandatangan di bawah ini:

Nama Mahasiswa : FERDY ATMAJA
Program Studi : Sistem Informasi
Dosen Pembimbing : 1. Dr. Eng. Agussalim, MT.
2. Eka Dyar Wahyuni, S.Kom., M.Kom.

dengan ini menyatakan bahwa isi sebagian maupun keseluruhan disertasi dengan judul:

ANALISIS SENTIMEN BERBASIS ASPEK PADA SISTEM LAYANAN PENGADUAN MASYARAKAT DI KOTA SURABAYA MENGGUNAKAN METODE LATENT DIRICHLET ALLOCATION DAN NAÏVE BAYES

adalah benar-benar hasil karya intelektual mandiri, diselesaikan tanpa menggunakan bahan-bahan yang tidak diizinkan dan bukan merupakan karya pihak lain yang saya akui sebagai karya sendiri. Semua referensi yang dikutip maupun dirujuk telah ditulis secara lengkap pada daftar pustaka. Apabila ternyata pernyataan ini tidak benar, saya bersedia menerima sanksi sesuai peraturan yang berlaku.

Surabaya, 13 Desember 2024
Yang Membuat Pernyataan,



FERDY ATMAJA
NPM. 20082010118

Halaman ini sengaja dikosongkan

ABSTRAK

Nama Mahasiswa / NPM : Ferdy Atmaja / 20082010118
Judul Skripsi : Analisis Sentimen Berbasis Aspek Pada Sistem Layanan Pengaduan Masyarakat Di Kota Surabaya Menggunakan Metode Latent Dirichlet Allocation Dan Naïve Bayes
Dosen Pembimbing : 1. Dr. Eng. Agussalim, MT.
2. Eka Dyar Wahyuni, S.Kom., M.Kom.

Sistem layanan pengaduan masyarakat merupakan sarana penting bagi pemerintah kota untuk menerima, mengelola, dan menindaklanjuti keluhan masyarakat, namun peningkatan volume data pengaduan di Kota Surabaya menjadi tantangan dalam pengelolaan manual yang kurang efektif. Analisis mendalam terhadap data pengaduan diperlukan untuk memahami opini masyarakat terhadap kualitas layanan, sehingga penelitian ini bertujuan untuk menganalisis sentimen berbasis aspek pada data pengaduan masyarakat dengan menggunakan metode Latent Dirichlet Allocation (LDA) dan Naïve Bayes. Metode LDA digunakan untuk mengidentifikasi aspek-aspek utama yang sering muncul dalam pengaduan masyarakat, seperti pelayanan, infrastruktur, dan kebersihan, sedangkan metode Naïve Bayes diterapkan untuk mengklasifikasikan sentimen masyarakat ke dalam kategori positif, negatif, atau netral. Dataset yang digunakan terdiri dari 10.847 record data aduan yang dikumpulkan melalui aplikasi WargaKu dan Media Center Kota Surabaya sepanjang tahun 2023. Hasil penelitian menunjukkan bahwa metode LDA berhasil mengidentifikasi 17 topik utama, termasuk administrasi, infrastruktur, dan informasi publik, sementara model Naïve Bayes dengan resampling mencapai akurasi 80% dalam klasifikasi sentimen. Penelitian ini memberikan wawasan berbasis data yang relevan untuk membantu pemerintah kota memprioritaskan perbaikan pada isu utama serta menawarkan pendekatan yang dapat diadopsi oleh sistem pengaduan serupa..

Kata Kunci: Analisis sentimen, *Latent Dirichlet Allocation*, *Naïve Bayes*, Pengaduan masyarakat, Layanan publik

Halaman ini sengaja dikosongkan

ABSTRACT

Student Name / NPM : Ferdy Atmaja / 20082010118
Thesis Title : Aspect-Based Sentiment Analysis on Public
Complaint Service Systems in Surabaya City
Using Latent Dirichlet Allocation and Naïve
Bayes
Advisor : 1. Dr. Eng. Agussalim, MT.
2. Eka Dyar Wahyuni, S.Kom., M.Kom.

The public complaint service system is a crucial platform for city governments to receive, manage, and respond to citizens' grievances. However, the increasing volume of complaints in Surabaya poses challenges for manual management, which is often inefficient. In-depth analysis of complaint data is essential to understand public opinions regarding service quality. This study aims to perform aspect-based sentiment analysis on public complaint data using Latent Dirichlet Allocation (LDA) and Naïve Bayes methods. LDA is employed to identify key aspects frequently mentioned in complaints, such as services, infrastructure, and cleanliness, while Naïve Bayes is applied to classify public sentiment into positive, negative, or neutral categories. The dataset consists of 10,847 complaint records collected through the WargaKu application and the Media Center of Surabaya City throughout 2023. The findings reveal that LDA successfully identified 17 key topics, including administration, infrastructure, and public information, while the Naïve Bayes model, enhanced with resampling techniques, achieved an accuracy of 80% in sentiment classification. This study provides data-driven insights to help city governments prioritize improvements on critical issues and offers an approach that can be adopted by similar complaint management systems.

Keywords: *Sentiment Analysis, Latent Dirichlet Allocation, Naïve Bayes, Public complaints, Public services*

Halaman ini sengaja dikosongkan

KATA PENGANTAR

Dengan memanjatkan puji syukur kehadirat Allah SWT atas limpahan rahmat, karunia, dan kasih sayang-Nya, sehingga penulis dapat menyelesaikan tugas akhir skripsi yang berjudul “**Analisis Sentimen Berbasis Aspek Pada Sistem Layanan Pengaduan Masyarakat Di Kota Surabaya Menggunakan Metode *Latent Dirichlet Allocation* dan *Naïve Bayes***”. Tugas akhir skripsi ini diajukan sebagai salah satu syarat kelulusan program Sarjana pada program studi Strata Satu Sistem Informasi, Fakultas Ilmu Komputer, Universitas Pembangunan Nasional “Veteran” Jawa Timur.

Perjalanan penyusunan skripsi ini menjadi pengalaman yang penuh pembelajaran, tantangan, dan momen berharga yang tidak mungkin terlewat tanpa dukungan berbagai pihak. Oleh karena itu, penulis dengan tulus menyampaikan rasa terima kasih yang sebesar-besarnya kepada:

1. Kedua orang tua penulis, terima kasih sebesar-besarnya atas segala bentuk kasih sayang, dukungan, kepercayaan, dan doa yang selalu dipanjatkan, sehingga penulis dapat menyelesaikan skripsi ini dengan baik.
2. Keluarga besar penulis yang selalu memberikan doa, semangat, dan dukungan moral tanpa henti
3. Bapak Dr. Eng. Agussalim, M.T. selaku dosen pembimbing pertama yang telah memberikan waktu, dukungan, arahan, dan pengetahuan selama pengerjaan skripsi, sehingga penulis dapat menyelesaikan skripsi ini dengan baik.
4. Ibu Eka Dyar Wahyuni, S.Kom, M.Kom. sebagai pembimbing kedua yang telah memberikan dukungan untuk segera menyelesaikan skripsi, memberikan saran, dan memberikan solusi terhadap kesalahan yang muncul dalam penyusunan skripsi ini.
5. Bapak Tri Lathif Mardi Suryanto, S.Kom, M.T. selaku dosen wali yang senantiasa memberikan bimbingan dan arahan selama masa perkuliahan.
6. Bapak Agung Brastama Putra, S.Kom., M.Kom., selaku Koordinator Program Studi Sistem Informasi yang selalu memberikan kemudahan bagi mahasiswanya.

7. Seluruh Bapak dan Ibu dosen Program Studi Sistem Informasi UPN “Veteran” Jawa Timur yang telah memberikan ilmu pengetahuan yang bermanfaat selama di jenjang perkuliahan.
8. Pihak Dinas Komunikasi dan Informatika Kota Surabaya, yang telah memberikan izin dan dukungan penuh sehingga penelitian ini dapat berjalan dengan lancar.
9. Seluruh teman-teman Sistem Informasi angkatan 2020 yang telah bersama-sama melewati perjuangan dari awal hingga akhirnya dapat mencapai tahap ini.
10. Semua pihak yang tidak dapat penulis sebutkan satu persatu yang telah membantu dan memberikan dukungan dalam penyelesaian skripsi ini.
11. Terima kasih kepada diri saya sendiri yang sudah berjuang dan tidak putus asa menghadapi berbagai rintangan, sehingga pada akhirnya skripsi ini dapat terselesaikan.

Akhir kata, penulis menyadari bahwa skripsi ini masih jauh dari sempurna. Oleh karena itu, masukan, kritik, dan saran yang membangun sangat diharapkan. Semoga skripsi ini dapat memberikan manfaat nyata bagi perkembangan ilmu pengetahuan, khususnya di bidang sistem informasi.

Surabaya, Desember 2024

Penulis

DAFTAR ISI

LEMBAR JUDUL SKRIPSI.....	i
LEMBAR PENGESAHAN	iii
LEMBAR PERSETUJUAN.....	v
SURAT PERNYATAAN ORISINALITAS	vii
ABSTRAK	ix
KATA PENGANTAR	xiii
DAFTAR ISI.....	xv
DAFTAR GAMBAR	xix
DAFTAR TABEL.....	xxiii
DAFTAR NOTASI	xxv
BAB I PENDAHULUAN	1
1.1. Latar Belakang	1
1.2. Rumusan Masalah	6
1.3. Batasan Masalah.....	6
1.4. Tujuan Penelitian.....	6
1.5. Manfaat Penelitian.....	7
1.6. Sistematika Penulisan.....	7
BAB II TINJAUAN PUSTAKA.....	9
2.1. Penelitian Terdahulu.....	9
2.2. Dasar Teori	12
2.1.1. Media Center.....	12
2.1.2. Aplikasi WargaKu.....	13
2.1.3. Python	14
2.1.4. Text Mining.....	14

2.1.5.	Analisis Sentimen	15
2.1.6.	Analisis Sentimen Berbasis Aspek.....	16
2.1.7.	Dataset.....	16
2.1.8.	Preprocessing Text.....	17
2.1.9.	Exploratory data analysis (EDA)	19
2.1.10.	Algoritma <i>Naïve Bayes</i>	20
2.1.11.	<i>Topic Modelling</i>	22
2.1.12.	<i>Latent Dirichlet Allocation</i> (LDA).....	22
2.1.13.	<i>Term Frequency-Inverse Document Frequency</i> (TF-IDF)	23
BAB III METODOLOGI PENELITIAN.....		25
3.1.	Studi Literatur.....	26
3.2.	Pengumpulan Data	27
3.3.	<i>Exploratory Data Analysis</i> (EDA)	27
3.4.	Pelabelan Sentimen	27
3.5.	<i>Text Preprocessing</i>	29
3.6.	Tahapan Analisa Sentimen Berbasis Aspek.....	33
3.6.1.	Pemodelan Topik	34
3.6.2.	Klasifikasi Sentimen Menggunakan <i>Naïve Bayes</i>	35
3.6.3.	Analisis Sentimen Berbasis Aspek.....	37
3.7.	<i>Deployment</i>	37
3.8.	Kesimpulan.....	38
BAB IV PENGUJIAN DAN ANALISA		39
4.1.	Studi Literatur.....	39
4.2.	Pengumpulan Data	39
4.3.	<i>Exploratory Data Analysis</i> (EDA)	41
4.4.	Pelabelan Sentimen	43

4.4.1.	Labelling Otomatis.....	44
4.4.2.	Labelling Manual	52
4.5.	Text Preprocessing	54
4.5.1.	Case Folding	54
4.5.2.	Data Cleansing	55
4.5.3.	Normalisasi	63
4.5.4.	<i>Stopword Removal</i>	66
4.5.5.	<i>Stemming</i>	68
4.5.6.	Tokenizing.....	70
4.5.7.	<i>WordCloud</i>	71
4.6.	Tahapan Analisis Sentimen Berbasis Aspek	72
4.6.1.	Pemodelan Topik	72
4.6.2.	Klasifikasi Sentimen Menggunakan Naïve Bayes	83
4.6.3.	Analisis Sentimen Berbasis Aspek.....	100
4.7.	Deployment	102
4.7.1.	Beranda	103
4.7.2.	Dashboard	104
4.7.3.	Wordcloud.....	107
4.7.4.	Analisis.....	108
4.7.5.	Dokumentasi	112
BAB V KESIMPULAN DAN SARAN.....		113
5.1.	Kesimpulan.....	113
5.2.	Saran	114
DAFTAR PUSTAKA		115
LAMPIRAN.....		121

Halaman ini sengaja dikosongkan

DAFTAR GAMBAR

Gambar 2. 1 Halaman Website Media Center Surabaya.....	12
Gambar 2. 2 Halaman Aplikasi Wargaku	13
Gambar 3. 1 Alur Penelitian.....	26
Gambar 3. 2 Alur proses pelabelan sentimen.....	29
Gambar 3. 3 Alur proses preprocessing	29
Gambar 3. 4 Alur proses case folding.....	30
Gambar 3. 5 Alur proses cleansing	30
Gambar 3. 6 Alur proses normalisasi.....	31
Gambar 3. 7 Alur Proses Stopword Removal	32
Gambar 3. 8 Alur proses stemming.....	33
Gambar 3. 9 Alur proses tokenizing	33
Gambar 3. 10 Alur pemodelan topik.....	34
Gambar 3. 11 Alur Klasifikasi Sentimen menggunakan Naïve Bayes	36
Gambar 4. 1 Data aduan masyarakat dari PPID Kota Surabaya.....	40
Gambar 4. 2 Distribusi jumlah keluhan masyarakat per bulan tahun 2023	42
Gambar 4. 3 Word Cloud awal data aduan	43
Gambar 4. 4 Mentranslate teks Bahasa Indonesia kedalam Bahasa Inggris	44
Gambar 4. 5 Case folding.....	54
Gambar 4. 6 Remove HTML	56
Gambar 4. 7 Remove emoji	58
Gambar 4. 8 Remove Whitespace.....	59
Gambar 4. 9 Complete cleansing	61
Gambar 4. 10 Inisialisasi path untuk memanggil kamus normalisasi.....	64
Gambar 4. 11 Membaca file CSV	64
Gambar 4. 12 Membuat kamus normalisasi.....	64
Gambar 4. 13 Fungsi untuk melakukan proses normalisasi pada teks.....	65
Gambar 4. 14 Memasukkan teks hasil normalisasi ke dalam dataframe	65
Gambar 4. 15 Menghapus titik koma pada data.....	65
Gambar 4. 16 Mengambil Daftar Stopwords	66
Gambar 4. 17 Fungsi untuk menghapus stopwords	67

Gambar 4. 18 Menghapus stopwords dari dataframe.....	67
Gambar 4. 19 Daftar kata spesifik yang akan dihapus.....	67
Gambar 4. 20 Menghapus kata spesifik yang terdapat pada teks	67
Gambar 4. 21 Inisialisasi stemmer	69
Gambar 4. 22 Fungsi untuk melakukan stemming pada data teks.....	69
Gambar 4. 23 Mengisi dataframe dengan teks yang telah dilakukan stemming... 69	
Gambar 4. 24 Proses tokenizing.....	71
Gambar 4. 25 Word Cloud setelah dilakukan preprocessing pada data.....	72
Gambar 4. 26 Import library untuk pemodelan topik	73
Gambar 4. 27 Mengkonversi kolom ke dalam bentuk list	74
Gambar 4. 28 Membangun model bigram trigram	74
Gambar 4. 29 Membuat bigram dan trigram.....	75
Gambar 4. 30 Proses vektorisasi menggunakan TF IDF.....	75
Gambar 4. 31 Membuat dictionary dan corpus.....	76
Gambar 4. 32 Fungsi untuk menghitung coherence score	76
Gambar 4. 33 Grafik hasil perhitungan coherence score	77
Gambar 4. 34 Visualisasi LDA dengan 19 topik	79
Gambar 4. 35 Import library untuk pemodelan klasifikasi	84
Gambar 4. 36 Import Library untuk resampling	85
Gambar 4. 37 Menggunakan data hasil stopword removal.....	86
Gambar 4. 38 Menggunakan data hasil stemming	86
Gambar 4. 39 Splitting data menggunakan metode holdout	87
Gambar 4. 40 Splitting data dalam cross-validation	87
Gambar 4. 41 Inisialisasi TF-IDF untuk vektorisasi.....	88
Gambar 4. 42 Vektorisasi pada metode Holdout	88
Gambar 4. 43 Vektorisasi pada metode cross-validation.....	89
Gambar 4. 44 Proses resampling pada skenario holdout RUS.....	89
Gambar 4. 45 Proses resampling pada skenario holdout SMOTE.....	90
Gambar 4. 46 Proses resampling pada skenario Cross-Validation RUS	90
Gambar 4. 47 Proses resampling pada skenario Cross-Validation SMOTE.....	91
Gambar 4. 48 Train Model Naïve Bayes	91
Gambar 4. 49 Train model Naïve Bayes dengan Cross-Validation.....	92

Gambar 4. 50 Confusion Matrix Naïve Bayes	93
Gambar 4. 51 Classification Report Naïve Bayes.....	93
Gambar 4. 52 Konfigurasi parameter: Holdout tanpa resampling	94
Gambar 4. 53 Konfigurasi parameter: Holdout SMOTE.....	94
Gambar 4. 54 Konfigurasi parameter: Holdout RUS.....	94
Gambar 4. 55 Konfigurasi parameter: Cross-Validation tanpa resampling.....	95
Gambar 4. 56 Konfigurasi parameter: Cross-Validation SMOTE.....	95
Gambar 4. 57 Konfigurasi parameter: Cross-Validation RUS	95
Gambar 4. 58 Data hasil pemodelan topik dan klasifikasi sentimen	100
Gambar 4. 59 Perbandingan proporsi sentimen berdasarkan aspek keluhan	102
Gambar 4. 60 Tampilan menu beranda (overview)	103
Gambar 4. 61 Tampilan menu beranda (fitur kunci dan faq).....	103
Gambar 4. 62 Visualisasi data jumlah keluhan berdasarkan kategori	104
Gambar 4. 63 Diagram proporsi sentimen	104
Gambar 4. 64 Diagram rata-rata sentimen berdasarkan jumlah keluhan	105
Gambar 4. 65 Distribusi sentimen per aspek	105
Gambar 4. 66 Grafik tren keluhan bulanan.....	106
Gambar 4. 67 Ringkasan Analisis Sentimen.....	107
Gambar 4. 68 Tampilan Menu Wordcloud	108
Gambar 4. 69 Tampilan Menu Analisis dengan Input Teks	108
Gambar 4. 70 Tampilan menu analisis dengan unggahan file	109
Gambar 4. 71 Tampilan hasil analisis dengan input teks.....	110
Gambar 4. 72 Tampilan hasil analisis dengan unggahan file.....	110
Gambar 4. 73 Opsi untuk ekspor data hasil analisis	111
Gambar 4. 74 Tampilan dokumentasi penggunaan website.....	112

Halaman ini sengaja dikosongkan

DAFTAR TABEL

Tabel 4. 1 Data keluhan warga yang akan digunakan.....	41
Tabel 4. 2 Hasil pelabelan otomatis skenario pertama.....	45
Tabel 4. 3 Distribusi jumlah data hasil pelabelan otomatis skenario pertama	47
Tabel 4. 4 Hasil pelabelan otomatis dengan label yang tidak akurat	47
Tabel 4. 5 Hasil pelabelan otomatis skenario kedua	49
Tabel 4. 6 Distribusi jumlah data hasil pelabelan otomatis skenario kedua	50
Tabel 4. 7 Hasil pelabelan otomatis dengan label yang tidak akurat	51
Tabel 4. 8 Hasil Pelabelan Manual Sentimen	52
Tabel 4. 9 Hasil pengurangan data setelah pelabelan	54
Tabel 4. 10 Data hasil case folding	55
Tabel 4. 11 Data hasil remove HTML	57
Tabel 4. 12 Data hasil remove emoji	58
Tabel 4. 13 Data hasil remove whitespace.....	60
Tabel 4. 14 Data hasil complete cleansing	62
Tabel 4. 15 Kamus normalisasi	63
Tabel 4. 16 Data hasil normalisasi	65
Tabel 4. 17 Data hasil stopword removal.....	68
Tabel 4. 18 Data hasil stemming.....	70
Tabel 4. 19 Data hasil tokenizing.....	71
Tabel 4. 20 Tabel hasil perhitungan coherence score per Num Topic.....	78
Tabel 4. 21 Interpretasi aspek	80
Tabel 4. 22 Data keluhan yang sudah terlabeli aspek	83
Tabel 4. 23 Tabel analisis pengujian skenario	96
Tabel 4. 24 Hasil Evaluasi Model Naive Bayes pada Skenario Terbaik	99
Tabel 4. 25 Distribusi sentimen berdasarkan aspek keluhan	100

Halaman ini sengaja dikosongkan

DAFTAR NOTASI

$P(X Y)$: Probabilitas kejadian X dengan bukti Y.
$P(\text{Sentence} \text{Sentiment})$: Probabilitas kalimat muncul diberikan sentimen tertentu.
$P(V_j)$: Probabilitas kategori V_j dalam data.
$fd(V_j)$: Frekuensi dokumen dalam kategori V_j
$ D $: Total dokumen dalam data.
$P(W_k V_j)$: Probabilitas kata W_k muncul dalam kategori V_j .
$f(W_k V_j)$: Frekuensi kata W_k dalam kategori V_j
$ W $: Total kata unik dalam data pelatihan.
V_{map}	: Kategori dengan probabilitas maksimum.
$arg\ max$: Memilih kategori yang memiliki nilai probabilitas terbesar.
\prod_i	: Perkalian probabilitas untuk setiap kata i .
W	: Bobot suatu kata dalam model vector.
W_{dt}	: bobot dokumen ke-d terhadap kata ke-t.
TF_{dt}	: Frekuensi kemunculan kata tertentu dalam dokumen ke-d.
IDF_{ft}	: inverse document frequency
N	: Total dokumen.
df	: Jumlah dokumen yang mengandung kata yang dicari.

Halaman ini sengaja dikosongkan