

**ANALISIS SENTIMEN PADA DATASET SENTIPOL  
MENGUNAKAN METODE NEIGHBOR WEIGHTED K-NEAREST  
NEIGHBOR (NWKNN)**

Skripsi



oleh  
**HENDRAWAN**  
**71130080**

PROGRAM STUDI TEKNIK INFORMATIKA FAKULTAS TEKNOLOGI  
INFORMASI UNIVERSITAS KRISTEN DUTA WACANA  
2017

**ANALISIS SENTIMEN PADA DATASET SENTIPOL  
MENGUNAKAN METODE NEIGHBOR WEIGHTED K-NEAREST  
NEIGHBOR (NWKNN)**

Skripsi



Diajukan kepada Program Studi Teknik Informatika Fakultas Teknologi Informasi  
Universitas Kristen Duta Wacana  
Sebagai Salah Satu Syarat dalam Memperoleh Gelar  
Sarjana Komputer

Disusun oleh

**HENDRAWAN**  
**71130080**

PROGRAM STUDI TEKNIK INFORMATIKA FAKULTAS TEKNOLOGI  
INFORMASI UNIVERSITAS KRISTEN DUTA WACANA 2017

## PERNYATAAN KEASLIAN SKRIPSI

Saya menyatakan dengan sesungguhnya bahwa skripsi dengan judul:

### **ANALISIS SENTIMEN PADA DATASET SENTIPOL MENGGUNAKAN METODE NEIGHBOR WEIGHTED K-NEAREST NEIGHBOR (NWKNN)**

yang saya kerjakan untuk melengkapi sebagian persyaratan menjadi Sarjana Komputer pada pendidikan Sarjana Program Studi Teknik Informatika Fakultas Teknologi Informasi Universitas Kristen Duta Wacana, bukan merupakan tiruan atau duplikasi dari skripsi kesarjanaan di lingkungan Universitas Kristen Duta Wacana maupun di Perguruan Tinggi atau instansi manapun, kecuali bagian yang sumber informasinya dicantumkan sebagaimana mestinya.

Jika dikemudian hari didapati bahwa hasil skripsi ini adalah hasil plagiasi atau tiruan dari skripsi lain, saya bersedia dikenai sanksi yakni pencabutan gelar kesarjanaan saya.

Yogyakarta, 2 Agustus 2017



HENDRAWAN  
71130080

## HALAMAN PERSETUJUAN

Judul : ANALISIS SENTIMEN PADA KOMENTAR PEMILU  
PRESIDEN INDONESIA TAHUN 2014  
MENGUNAKAN METODE NEIGHBOR WEIGHTED  
K-NEAREST NEIGHBOR (NWKNN)

Nama : Hendrawan

NIM : 71130080

Mata Kuliah : Tugas Akhir

Kode : TIW276

Semester : Genap

Tahun Akademik : 2016/2017

Telah diperiksa dan disetujui

Di Yogyakarta,


Pada Tanggal 16 Juli 2017

Dosen Pembimbing I



Antonius Rachmat C., S.Kom., M.Cs.

Dosen Pembimbing II



Yuan Lukito, S.Kom., M.Cs.

**HALAMAN PENGESAHAN**

**ANALISIS SENTIMEN PADA DATASET SENTIPOL MENGGUNAKAN  
METODE NEIGHBOR WEIGHTED K-NEAREST NEIGHBOR (NWKNN)**

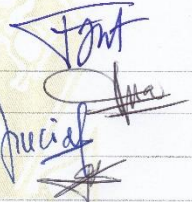
Oleh: HENDRAWAN / 71130080

Dipertahankan di depan Dewan Penguji Skripsi  
Program Studi Teknik Informatika Fakultas Teknologi Informasi  
Universitas Kristen Duta Wacana - Yogyakarta  
Dan dinyatakan diterima untuk memenuhi salah satu syarat memperoleh gelar  
Sarjana Komputer  
pada tanggal 26 Juli 2017.

Yogyakarta, 2 Agustus 2017  
Mengesahkan,

Dewan Penguji:

1. Antonius Rachmat C., S.Kom.,M.Cs.
2. Yuan Lukito, S.Kom., M.Cs.
3. Lucia Dwi Krishawati, Dr.
4. Willy Sudiarto Raharjo, S.Kom.,M.Cs.

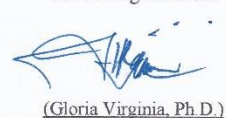


Dekan



(Budi Susanto, S.Kom., M.T.)

Ketua Program Studi



(Gloria Virginia, Ph.D.)

## UCAPAN TERIMA KASIH

Dalam menyelesaikan penelitian tugas akhir ini, penulis telah banyak mendapatkan bimbingan, saran serta dukungan dari berbagai pihak. Oleh karena itu penulis mengucapkan terima kasih kepada :

1. Bapak Budi Susanto, S.Kom., M.T. selaku Dekan Fakultas Teknologi Informasi Universitas Kristen Duta Wacana.
2. Ibu Gloria Virginia, S.Kom., MAI. selaku Ketua Program Studi Teknik Informatika Universitas Kristen Duta Wacana.
3. Bapak Antonius Rachmat, S.Kom., M.Cs. selaku dosen pembimbing I yang telah membantu memberikan saran dan bimbingan dalam penyusunan tugas akhir ini.
4. Bapak Yuan Lukito, S.Kom., M.Cs. selaku dosen pembimbing II yang telah membantu memberikan saran dan bimbingan dalam penyusunan tugas akhir ini.
5. Orangtua, saudara-saudara dan teman-teman terdekat yang memberikan dukungan baik serta motivasi untuk menyelesaikan tugas akhir ini.
6. Semua pihak yang tidak dapat disebutkan satu per satu yang ikut memberikan arahan, saran maupun nasihat dalam penelitian tugas akhir ini.

Penulis menyadari dalam penelitian ini terdapat banyak kekurangan, baik dalam penelitian maupun penulisan laporan akhir ini. Akhir kata penulis ingin berterima kasih pada pihak yang telah mendukung penulis, dan semoga tugas akhir ini dapat bermanfaat bagi semua pihak terutama Universitas Kristen Duta Wacana.

Penulis

## KATA PENGANTAR

Puji syukur dan terima kasih kepada Tuhan Yang Maha Esa karena berkat dan bimbingan-Nya, penulis dapat menyelesaikan penelitian tugas akhir dengan judul “ANALISIS SENTIMEN PADA DATASET SENTIPOL MENGGUNAKAN METODE NEIGHBOR WEIGHTED K-NEAREST NEIGHBOR (NWKNN)”.

Penulisan laporan tugas akhir ini diajukan guna untuk melengkapi sebagai syarat dalam mencapai gelar sarjana strata satu (S1) di Fakultas Teknologi Informasi Program Studi Teknik Informatik Universitas Kristen Duta Wacana Yogyakarta. Dalam penulisan laporan tugas akhir ini penulis menyadari banyak kekurangan dalam melakukan penelitian serta penulisan, namun penulis sudah melakukan dan berusaha semaksimal mungkin untuk mencapai hasil yang terbaik. Penulis mengharapkan kritik dan saran yang membangun untuk menyempurnakan tugas akhir ini.

Pada pengerjaan tugas akhir ini banyak sekali kendala dan tantangan, namun berkatnya Tuhan Yang Maha Esa, laporan serta penelitian ini dapat selesai dengan lancar dan baik. Penulis juga banyak mengucapkan terimakasih kepada Bapak Antonius Rachmat dan Bapak Yuan Lukito yang telah membimbing penulis dari awal hingga akhir ini. Terima kasih untuk waktu serta ide-ide yang telah diungkapkan untuk membantu penulis dalam melakukan penulisan serta melakukan penelitian.

Penulis

# INTISARI

## ANALISIS SENTIMEN PADA DATASET SENTIPOL MENGUNAKAN METODE NEIGHBOR WEIGHTED K- NEAREST NEIGHBOR (NWKNN)

Pada umumnya pengklasifikasian teks mengasumsikan bahwa data latih yang digunakan terdistribusikan secara sama rata pada setiap kategorinya. Padahal kenyataannya sering sekali kumpulan data yang digunakan untuk latih tidak seimbang sehingga memunculkan data yang masuk ke dalam kategori mayoritas dan minoritas. Seperti pada pengujian ini menggunakan dataset sentipol pemilihan presiden Indonesia pada tahun 2014 memiliki dataset yang tidak seimbang. Dengan tidak seimbang data latih ini menyebabkan klasifikasi pada KNN menjadi tidak akurat dan hasil analisisnya selalu cenderung ke kategori yang mayoritas. Dengan keadaan inilah NWKNN muncul untuk mengurangi kesalahan dalam data yang tidak seimbang.

Pada penelitian ini akan dilakukan proses *preprocessing* pada biasanya seperti *case folding*, *cleansing*, *tokenizing*, sinonim kamus, *stop word removal* dan *stemming*. Lalu akan dilakukan pembobotan menggunakan TF-IDF dan *cosine similarity* untuk mencari kemiripan dokumen. Kemudian akan dilakukan klasifikasi menggunakan NWKNN. Dari hasil tersebut akan dibandingkan hasil dari KNN biasa dengan NWKNN.

Penelitian ini membuktikan bahwa NWKNN memang cocok digunakan dalam pembelajaran data latih yang tidak seimbang. Penelitian ini menghasilkan kenaikan akurasi sebesar 6.25% dari KNN biasa dengan nilai akurasi dari 58.75% menjadi 65%. Penelitian lainnya dengan menambahkan kamus sinonim pada proses analisis NWKNN sehingga dapat meningkatkan akurasi sebesar 6.25% dengan nilai akurasi dari 50% menjadi 56.25%. Dalam penelitian ini juga penggunaan kamus sinonim yang dibuat penulis dapat meningkatkan akurasi dari KNN tanpa kamus dengan NWKNN menggunakan kamus sebesar 7.5% dari 66.25% menjadi 73.75%.

**Kata Kunci :** [*KNN, neighbor weighted K-nearest neighbor (NWKNN), text mining, sentiment analysis*]



## DAFTAR ISI

PERNYATAAN KEASLIAN SKRIPSI .....	iii
HALAMAN PERSETUJUAN.....	iv
HALAMAN PENGESAHAN .....	v
UCAPAN TERIMA KASIH.....	vi
KATA PENGANTAR .....	vii
INTISARI .....	viii
DAFTAR TABEL.....	xi
DAFTAR GAMBAR .....	xii
DAFTAR LAMPIRAN.....	xiv
BAB I.....	1
1.1. Latar Belakang.....	1
1.2. Rumusan Masalah.....	2
1.3. Batasan Masalah .....	3
1.4. Tujuan Penelitian .....	3
1.5. Manfaat Penelitian .....	3
1.6. Metode Penelitian .....	4
1.7. Sistematika Penulisan .....	4
BAB II.....	6
2.1. Tinjauan Pustaka.....	6
2.2. Landasan Teori .....	7
BAB III .....	17
3.1. Spesifikasi Sistem.....	17
3.2. Perancangan Proses Sistem.....	22
3.3. Perancangan Antar Muka Sistem.....	28
3.4. Perancangan Pengujian dan Evaluasi Sistem.....	32
BAB IV .....	35
4.1. Implementasi Sistem.....	35
4.2. Pengujian dan Analisis Sistem.....	46

BAB V .....	62
5.1. Kesimpulan .....	62
5.2. Saran .....	62
DAFTAR PUSTAKA .....	64
LAMPIRAN.....	66

©UKDW

## DAFTAR TABEL

Tabel 2.1 Contoh soal NWKNN .....	13
Tabel 2.2 Contoh hasil dari tahapan <i>preprocessing</i> .....	13
Tabel 2.3. Contoh hasil perhitungan bobot TF-IDF .....	13
Tabel 2.4. Contoh perhitungan total kuadrat bobot data latih dan data uji, serta akarnya .....	14
Tabel 2.5. Contoh hasil <i>cosine similarity</i> .....	14
Tabel 2.6. Contoh hasil <i>sorting</i> .....	14
Tabel 2.7. Contoh penentuan <i>score</i> KNN .....	14
Tabel 2.8. Contoh hasil <i>Weight</i> ke-I .....	15
Tabel 3.1 Test Case Skenario Pengujian .....	33
Tabel 4.1 Skenario Pengujian .....	47
Tabel 4.2 Tabel pengujian k optimal pada skenario 1 .....	47
Tabel 4.3 Tabel pengujian k optimal pada skenario 1(lanjutan).....	48
Tabel 4.4 Hasil pengujian <i>confusion matrix</i> semua skenario tanpa kamus .....	50
Tabel 4.5 Hasil pengujian <i>precision, recall, f-measure</i> dan akurasi semua skenario tanpa kamus .....	51
Tabel 4.6 Contoh kamus singkatan .....	53
Tabel 4.7 Hasil pengujian <i>confusion matrix</i> semua skenario menggunakan kamus .....	53
Tabel 4.8 Hasil pengujian <i>precision, recall, f-measure</i> dan akurasi semua skenario menggunakan kamus .....	53
Tabel 4.9 Selisih akurasi pada setiap skenario KNN dengan NWKNN tanpa kamus .....	58

## DAFTAR GAMBAR

Gambar 2.1. <i>Confusion Matrix</i> .....	16
Gambar 3.1. <i>Use case</i> diagram sistem klasifikasi komentar .....	18
Gambar 3.2. Flowchart sistem klasifikasi secara umum .....	23
Gambar 3.3. Flowchart sistem dalam <i>pre-processing</i> .....	24
Gambar 3.4. Flowchart pembobotan .....	25
Gambar 3.5. Flowchart sistem dalam melakukan klasifikasi .....	26
Gambar 3.6. Skema Basis Data .....	27
Gambar 3.7. Rancangan Antar Muka Analisis Komentar ( <i>User</i> ) .....	28
Gambar 3.8. Rancangan Antar Muka Hasil Analisis Komentar ( <i>User</i> ) .....	29
Gambar 3.9. Rancangan Antar Muka History Analisis ( <i>User</i> ) .....	30
Gambar 3.10. Rancangan Antar Muka Analisis Komentar ( <i>User</i> ) .....	30
Gambar 3.11. Rancangan Antar Muka Hasil Analisis Komentar ( <i>Admin</i> ) .....	31
Gambar 3.12. Rancangan Antar Muka History Analisis ( <i>Admin</i> ) .....	31
Gambar 3.13. Rancangan Antar Muka Data latih ( <i>Admin</i> ) .....	32
Gambar 4.1 Halaman utama .....	36
Gambar 4.2 Halaman <i>history analysis</i> .....	37
Gambar 4.3 Halaman Data Latih .....	38
Gambar 4.4 Halaman <i>Stop list</i> .....	38
Gambar 4.5 Halaman Kamus Singkatan .....	39
Gambar 4.6 Hasil evaluasi dari skenario I tanpa kamus .....	50
Gambar 4.7 Hasil evaluasi dari skenario II tanpa kamus .....	51
Gambar 4.8 Hasil evaluasi dari skenario III tanpa kamus .....	52
Gambar 4.9 Hasil evaluasi dari skenario I menggunakan kamus .....	54
Gambar 4.10 Hasil evaluasi dari skenario II menggunakan kamus .....	54
Gambar 4.11 Hasil evaluasi dari skenario III menggunakan kamus .....	55
Gambar 4.12 Hasil evaluasi NWKNN dari skenario I perbandingan kamus dan tidak kamus .....	56
Gambar 4.13 Hasil evaluasi NWKNN dari skenario II perbandingan kamus dan tidak kamus .....	57

Gambar 4.14 Hasil <i>confusion matrix</i> NWKNN dari skenario II perbandingan kamus dan tidak kamus .....	57
Gambar 4.15 Hasil <i>confusion matrix</i> dari skenario I perbandingan KNN tanpa kamus, NWKNN tanpa kamus dan NWKNN kamus .....	59
Gambar 4.16 Hasil <i>confusion matrix</i> dari skenario II perbandingan KNN tanpa kamus, NWKNN tanpa kamus dan NWKNN kamus .....	60
Gambar 4.16 Hasil <i>confusion matrix</i> dari skenario III perbandingan KNN tanpa kamus, NWKNN tanpa kamus dan NWKNN kamus .....	60

©UKDW

## DAFTAR LAMPIRAN

A.1 Listing Program 1 – <i>Preprocessing</i> .....	Lampiran A - 1
A.2 Listing Program 2 – <i>Preprocessing 2</i> .....	Lampiran A - 1
A.3 Listing Program 3 – TF-IDF .....	Lampiran A - 2
A.4 Listing Program 4 – Pembobotan .....	Lampiran A - 3
A.5 Listing Program 5 – <i>Cosine Similarity</i> .....	Lampiran A - 4
A.6 Listing Program 6 – <i>Score KNN</i> .....	Lampiran A - 4
A.7 Listing Program 7 – <i>NWKNN</i> .....	Lampiran A - 5
A.8 Listing Program 8 – Simpan hasil analisis <i>NWKNN</i> ke <i>database</i> .....	Lampiran A - 6
.....	Lampiran A - 6
A.9 Listing Program 9 – <i>Confusion matrix 1</i> .....	Lampiran A - 6
A.10 Listing Program 10 – <i>Confusion matrix 2</i> .....	Lampiran A - 6
A.11 Listing Program 11 – Simpan hasil <i>confusion matrix 2</i> .....	Lampiran A - 7
B.1 Kamus Singkatan .....	Lampiran B - 1
C.1 Contoh <i>Dataset</i> Sentipol .....	Lampiran C - 1
D.1 Scan Kartu Konsultasi A.....	Lampiran D – 1
D.2 Scan Kartu Konsultasi B.....	Lampiran D – 2
E.1 Scan Form Revisi .....	Lampiran E – 1

© UKDW

© UKDW



# BAB I

## PENDAHULUAN

### 1.1 Latar Belakang Masalah

Media sosial saat ini telah menguasai kehidupan pada pengguna *Internet*. Dengan media sosial kita menjadi mudah berkomunikasi. Pada umumnya, layanan yang ada pada media sosial adalah *chatting*, *e-mail*, berbagi pesan berbagi *video* ataupun foto. Tetapi dalam implementasinya, media sosial kini lebih banyak digunakan untuk kegiatan jual beli, menyampaikan informasi, bahkan sebagai media untuk mengekspresikan diri. Salah satu media sosial yang sangat dikenal yaitu Facebook. Facebook merupakan sebuah media sosial yang sudah sangat mendunia. Hampir setiap orang yang bisa mengakses *internet* pasti memiliki akun Facebook. Menurut KOMINFO (2013) dalam websitenya, pengguna *internet* di Indonesia saat ini mencapai 63 juta orang. Dari angka tersebut, 95 persennya menggunakan internet untuk mengakses media sosial. Dari data tersebut sangat jelas bahwa seberapa penting media sosial bagi masyarakat Indonesia.

Selain itu, media sosial juga dapat dimanfaatkan untuk *me-review* sebuah produk atau melakukan *review* terhadap suatu masalah baik politik maupun masalah sosial terkini yang sedang menjadi *trending* topik. Salah satu yang terjadi pada pemilu presiden Indonesia pada tahun 2014. Banyak terjadi sebuah kontradiksi antara 2 kubu, Kubu Jokowi dan kubu Prabowo. Dalam hal ini, dapat dilakukan sebuah analisis sentimen yang digunakan untuk mendapatkan arti-arti atau sebuah makna dari komentar-komentar yang ada.

Menurut Pang, Lee & Vaithyanathan (2002), analisis sentimen merupakan proses memahami, mengekstrak, dan mengolah data tekstual secara otomatis untuk mendapatkan informasi. Dari sini dapat disimpulkan bahwa analisis sentimen dapat digunakan untuk mengambil informasi yang tersurat maupun tersirat di dalam pesan. Analisis sentimen dapat dilakukan dengan membagi informasi kedalam

kelas-kelas. Kelas yang dimaksud dalam hal tersebut adalah kelas positif dan kelas negatif. Informasi yang mengandung informasi bernilai baik atau setuju akan dimasukkan ke dalam kelas positif. Sebaliknya, jika informasi yang mengandung arti tidak baik atau tidak setuju akan dimasukkan ke dalam kelas negatif.

Analisis sentimen sendiri merupakan cabang dari topik *Text Mining*. Kegiatan penting dari *Text Mining* adalah klasifikasi dan kategorisasi. Pada pemrosesan *Text Mining* akan selalu melibatkan proses pre-processing sebelum informasi *text* diolah. Contoh *pre-processing* yang digunakan seperti *casefolding*, *stopword removal*, dan *tokenisasi*. Metode yang digunakan dalam penelitian ini adalah *Neighbor Weighted K-Nearest Neighbor*. *Neighbor Weighted K-Nearest Neighbor* (NWKNN) merupakan modifikasi dari algoritma *K-Nearest Neighbor* (KNN).

Pada dasarnya sistem akan tetap menjalankan algoritma KNN seperti biasanya, akan tetapi sesaat sebelum masuk ke tahapan KNN akan ada pembobotan berdasarkan perhitungan yang ada. Pada bagian pembobotan ini lah yang membedakan KNN dengan NWKNN. Kemudian dari hasil pembobotan akan digunakan kedalam perhitungan di tahapan KNN.

Penelitian ini akan digunakan untuk pemrosesan data komentar pemilu presiden di Indonesia pada tahun 2014. Diharapkan dengan adanya sistem ini maka komputer dengan mudah dapat mengelompokan / mengklasifikasi komentar-komentar kedalam kelas positif dan negatif.

## 1.2 Rumusan Masalah

Rumusan masalah dalam penelitian ini adalah :

1. Bagaimana cara mengimplementasikan metode *Neighbor Weighted K-Nearest Neighbor* untuk digunakan analisis sentimen komentar pemilu presiden Indonesia pada tahun 2014?
2. Seberapa akuratkah hasil klasifikasi ke dalam kelas-kelas dengan menggunakan metode *Neighbor Weighted K-Nearest Neighbor* menggunakan *precision*, *recall* dan *F-measure*?

3. Bagaimana meningkatkan akurasi dengan menggunakan metode *Neighbor Weighted K-Nearest Neighbor*

### 1.3 Batasan Masalah

Batasan masalah dalam penelitian ini adalah:

1. Analisis sentimen menggunakan bahasa Indonesia.
2. Klasifikasi komentar akan menghasilkan dua kelompok komentar, yaitu komentar positif dan komentar negatif.
3. *Dataset* sentipol (sentiment politik) yang digunakan mengacu pada Rachmat C & Lukito (2016) yang diambil dari *account facebook page* yang membahas tentang pemilu presiden Indonesia pada tahun 2014.
4. *Dataset* yang tergolong data netral akan digabung dalam data negatif.

### 1.4 Tujuan Penelitian

Tujuan dari penelitian ini adalah untuk mengimplementasikan metode *Neighbor Weighted K-Nearest Neighbor* ke dalam analisis komentar pemilu presiden Indonesia pada tahun 2014 dan menganalisis seberapa akurat hasil yang didapatkan dari analisis komentar pemilu presiden Indonesia pada tahun 2014 dengan menggunakan metode tersebut.

### 1.5 Manfaat Penelitian

Manfaat dalam penelitian ini adalah:

1. Dari hasil *output* yang dihasilkan dapat digunakan untuk klasifikasi secara *general*.
2. Menghasilkan kamus yang spesifik pada *dataset* sentipol.
3. Meningkatkan hasil akurasi dari *Neighbor Weighted K-Nearest Neighbor*.

## 1.6 Metode Penelitian

Metode yang akan digunakan sebagai langkah-langkah untuk pemecahan masalah dalam penelitian adalah sebagai berikut:

1. Studi Literatur

Studi literatur dilakukan dengan cara mempelajari teori-teori yang diperlukan dalam penelitian melalui jurnal-jurnal yang berkaitan dengan algoritma *Neighbor Weighted K-Nearest Neighbor* untuk melakukan *Sentiment analysis*.

2. Merancang Desain Sistem

Desain yang akan dirancang adalah desain aplikasi *web base* yang akan digunakan untuk mengklasifikasikan sebuah *string* yang dimasukan sehingga menghasilkan *output* hasil dari *Sentiment analysis*.

3. Implementasi Sistem

Sistem akan dilakukan dengan menginputkan *dataset* kemudian masuk pada *pre-processing* dengan menggunakan beberapa metode. Seperti, *Case Folding*, *Tokenizing*, *stop word removal* dan *stemming*. Kemudian dari hasil *pre-processing* tersebut akan dijadikan pola latih dalam sistem ini. Langkah selanjutnya proses klasifikasi / penentuan komentar dalam kelas positif atau negatif dengan menggunakan *Neighbor Weighted K-Nearest Neighbor*. Hasil yang didapatkan bisa berupa klasifikasi komentar positif atau komentar negatif.

4. Pengujian dan Evaluasi

Setelah dilakukan pengujian, tahap selanjutnya adalah menganalisis keakuratan hasil dengan menggunakan *Precision*, *Recall* dan *F-Measure*.

## 1.7 Sistematika Penulisan

Sistematika penulisan Laporan Tugas Akhir ini dikelompokan 5 bab, yaitu:

Bab I, Pendahuluan. Bab ini berisi gambaran umum mengenai penelitian yang akan dilakukan yang mencakup antara lain latar belakang, perumusan masalah, batasan masalah, tujuan penelitian, manfaat penelitian, metode penelitian, sistematika penelitian,

Bab II, Landasan Teori, Bab ini berisi mengenai landasan teori yang digunakan dalam penelitian ini. Pada landasan teori memuat penjelasan tentang konsep dan prinsip utama yang digunakan dalam pembangunan sistem penelitian ini.

Bab III, Perancangan sistem. Bab ini berisi tentang analisis teori yang digunakan dan bagaimana teori tersebut dapat diimplementasikan kedalam sistem yang akan digunakan. Bab ini juga menjelaskan proses-proses perancangan sistem dengan teori yang telah dipilih, berisi perancangan basis pengetahuan, alir kerja sistem, antarmuka pemakai dan perancangan penelitian.

Bab IV, Implementasi dan analisis Sistem. Bab ini berisi tentang implementasi sistem yang telah dirancang sebagaimana sudah dirancang pada bab III dan juga dilakukan pembahasan analisis sistem yang juga disertai dengan hasil capture sistem yang telah dibangun.

Bab V, Kesimpulan dan Saran. Bab ini memuat kesimpulan dari penelitian yang dilakukan dan saran yang dapat dikembangkan untuk penelitian selanjutnya.

## BAB V

### KESIMPULAN DAN SARAN

#### 5.1. Kesimpulan

1. Penggunaan metode NWKNN sukses menyeimbangkan data latih yang tidak seimbang karena adanya penambahan bobot pada kategori minoritas sehingga kesalahan klasifikasi data pada kategori minoritas dapat dikurangi. Tingkat akurasi tertinggi pada KNN dan NWKNN tanpa menggunakan kamus adalah 68.75% dan akurasi KNN dan NWKNN menggunakan kamus adalah 75%.
2. NWKNN tanpa kamus dengan data seimbang tidak terjadi perubahan akurasi maupun *f-measure*, kemudian pada skenario II dengan 1260 data latih positif dan 420 data latih negatif didapatkan kenaikan akurasi dari KNN dan NWKNN sebesar 6.25% dengan kenaikan dari 58.75% menjadi 65%. Lalu pada skenario III dengan 2900 data latih positif dan 420 data latih negatif menghasilkan peningkatan akurasi sebesar 1.25% dengan kenaikan dari 51.25% menjadi 52.50%.
3. NWKNN menggunakan kamus dengan data seimbang tidak terjadi perubahan akurasi maupun *f-measure*, kemudian pada skenario II dengan 1260 data latih positif dan 420 data latih negatif didapatkan kenaikan akurasi dari KNN dan NWKNN sebesar 5% dengan kenaikan dari 60% menjadi 65%. Lalu pada skenario III dengan 2900 data latih positif dan 420 data latih negatif menghasilkan peningkatan akurasi sebesar 6.25% dengan kenaikan dari 50% menjadi 56.25%.
4. Kamus yang dibuat oleh peneliti berhasil meningkatkan akurasi NWKNN sebesar 7.5% dari 66.25% menjadi 73.75%.

#### 5.2. Saran

1. Pengujian *dataset* sentipol perlu menggunakan *K-fold cross validation* sehingga semua data dapat menjadi data uji sebanyak satu kali dan akan menjadi data latih sebanyak k-1. Dari pengujian ini akan diambil rata-

rata sebanyak iterasi yang digunakan untuk mendapatkan hasil dari pengujian.

2. Pada *dataset* sentipol perlu melakukan pengujian terhadap penggunaan nilai *exponent* pada rumus NWKNN, karena nilai *exponent* ini juga berpengaruh terhadap besar kecilnya pembobotan setiap kategori sehingga hasilnya dapat mempengaruhi hasil dari pengujian yang dilakukan.
3. Pengembangan kamus bisa ditambahkan dengan *spelling correction/levenshtein distance*.

©UKDW

## Daftar Pustaka

- Feldman, R., & Sanger, J. (2007). *The Text Mining Handbook*. New York: Cambridge University Press. Retrieved 05 December 2016 from [https://wtlab.um.ac.ir/images/e-library/text\\_mining/The%20Text%20Mining%20HandBook.pdf](https://wtlab.um.ac.ir/images/e-library/text_mining/The%20Text%20Mining%20HandBook.pdf)
- KOMINFO, P. (2013). *Kominfo : Pengguna Internet di Indonesia 63 Juta Orang*. Website Resmi Kementerian Komunikasi dan Informatika RI. Retrieved 15 September 2016, from [https://www.kominfo.go.id/content/detail/3415/kominfo-pengguna-internet-di-indonesia-63-juta-orang/0/berita\\_satker](https://www.kominfo.go.id/content/detail/3415/kominfo-pengguna-internet-di-indonesia-63-juta-orang/0/berita_satker)
- Larose, Daniel T. (2005). *Discovering Knowledge in Data: An Introduction to Data Mining*. Retrieved 05 December 2016, from [http://secs.ac.in/wp-content/CSEPORTAL/DataMining\\_Daniel.pdf](http://secs.ac.in/wp-content/CSEPORTAL/DataMining_Daniel.pdf)
- Liu, B. (2012). *Sentiment Analysis and Opinion Mining*. Morgan & Claypool Publishers. Retrieved 05 December 2016, from <https://www.cs.uic.edu/~liub/FBS/SentimentAnalysis-and-OpinionMining.pdf>
- Maatwebsite/Laravel-Excel. (n.d.). Retrieved from [www.github.com](http://www.github.com): <https://github.com/Maatwebsite/Laravel-Excel>
- Pang, B., Lee, L., & Vaithyanathan, S. (2002). *Thumbs up? Sentiment classification using machine learning techniques*, Proceedings of the Conference on Empirical Methods in Natural Language Processing



- Rachmat C, A. & Lukito, Y. (2016). Sentipol: Dataset Sentimen Komentar pada Kampanye Pemilu Presiden Indonesia 2014 dari Facebook Page. *Konferensi Nasional Teknologi Informasi dan Komunikasi*. Retrieved from <https://knastik.ukdw.ac.id/2016/makalah/artikel/e7-j1.pdf>
- Ridok, A. & Indrianti. (2016). Sentiment Analysis For Review Mobile Applications Using Neighbor Method Weighted K-Nearest Neighbor (NWKNN). *Journal Of Environmental Engineering & Sustainable Technology*, 3(1), 23-32. Retrieved from <http://jeest.ub.ac.id/index.php/jeest/article/view/45>
- Tan, S. (2005). Neighbor-Weighted K-nearest Neighbour For Unbalanced Text Corpus. *Expert System with Applications*, 28, 667-671. Retrieved from <http://www.sciencedirect.com/science/article/pii/S0957417404001708>
- Vijayarani, Dr. S, Ms. J Ilamanthi, and Ms Nithya. (2011). Preprocessing Techniques For Text Mining - An Overview. *International Journal of Computer Science & Communication Networks* 5.1 7-16. Web. 10 Oct. 2016. Retrieved from <http://www.ijcsn.com/Documents/Volumes/vol5issue1/ijcsn2015050102>
- Van Rijsbergen, C.J. *Information Retrieval*. 2nd ed. Skotlandia: Butterworth-Heinemann, 1976. Web. 5 Dec. 2016.
- Witten, Ian H and Eibe Frank. *Data Mining Practical Machine Learning Tools And Technique*. 2nd ed. San Francisco, Calif.: Morgan Kaufmann, 2005. Print.