

**ANALISIS SENTIMEN TWITTER
MENGGUNAKAN METODE K-NEAREST NEIGHBOR DAN
PENDEKATAN LEXICON**

TUGAS AKHIR

Diajukan Sebagai Salah Satu Syarat
Untuk Memperoleh Gelar Sarjana Teknik
Pada JurusanTeknik Informatika

Oleh :

DEFRI ROSDIANSYAH
11051102066



**FAKULTAS SAINS DAN TEKNOLOGI
UNIVERSITAS ISLAM NEGERI SULTAN SYARIF KASIM RIAU
PEKANBARU
2014**

LEMBAR PENGESAHAN

ANALISIS SENTIMEN TWITTER MENGGUNAKAN METODE K-NEAREST NEIGHBOR DAN PENDEKATAN LEXICON

TUGAS AKHIR

Oleh :

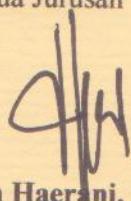
DEFRI ROSDIANSYAH
11051102066

Telah dipertahankan di depan sidang dewan penguji
Sebagai salah satu syarat untuk memperoleh gelar sarjana Teknik Informatika
Fakultas Sains dan Teknologi Universitas Islam Negeri Sultan Syarif Kasim Riau
Di Pekanbaru, pada tanggal 10 September 2014

Pekanbaru, 10 September 2014

Mengesahkan,

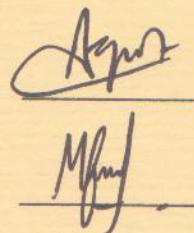
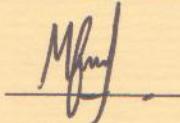
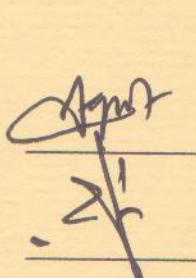
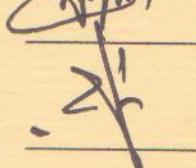
Ketua Jurusan


Elin Haerani, S.T, M.Kom
NIP. 19810523 200710 2 003



DEWAN PENGUJI

- Ketua : Surya Agustian, S.T, M.Kom
Sekretaris : Surya Agustian, S.T, M.Kom
Penguji I : M. Fikry, S.T, M.Sc
Penguji II : M. Safrizal, S.T, M.Cs

ANALISIS SENTIMEN TWITTER MENGGUNAKAN METODE K-NEAREST NEIGHBOR DAN PENDEKATAN LEXICON

DEFRI ROSDIANSYAH

11051102066

Tanggal Sidang: 10 September 2014

Periode Wisuda : November 2014

Jurusan Teknik Informatika

Fakultas Sains dan Teknologi

Universitas Islam Negeri Sultan Syarif Kasim Riau

ABSTRAK

Analisis sentimen dapat dikatakan sebagai *opinion mining*. Penelitian Analisis sentimen dapat dilakukan pada sosial media Twitter. Namun dari sejumlah penelitian analisis sentimen Twitter bahasa Indonesia tidak dilakukan evaluasi kalimat negasi. Pada penelitian ini akan dilakukan klasifikasi orientasi sentimen dalam 3 jenis yaitu positif, negatif dan netral menggunakan kombinasi *K-Nearest Neighbor* (*k*-NN) dan pendekatan *lexicon*. Topik penelitian ini adalah menentukan analisis sentimen dari kasus politik, khususnya sentimen publik terhadap tokoh politik pada kandidat calon presiden Republik Indonesia 2014. Algoritma *k*-NN pada penelitian ini akan bertindak sebagai algoritma yang akan melakukan klasifikasi orientasi sentimen sementara. Sedangkan pendekatan *lexicon* untuk dapat melakukan identifikasi kalimat netral pada tahapan *preprocessing query* juga melakukan evaluasi kalimat negasi. Pengujian sistem analisis sentimen Twitter ini dilakukan dengan model *confusion matrix* pada data uji *dan 10-fold Cross Validation* pada evaluasi data latih untuk mendapatkan data latih terbaik. Pengujian juga dilakukan pada penentuan nilai *k* pada algoritma *k*-NN. Kesimpulan dari pengujian kombinasi *k*-NN dan *lexicon* dapat meningkatkan akurasi orientasi sentimen. Nilai *k* terbaik pada algoritma *k*-NN adalah *k*=5 dan *k*=15. Sedangkan Metode yang terbaik adalah kombinasi B yang menggunakan pendekatan *lexicon* dan algoritma *k*-NN dengan akurasi 82 %.

Kata Kunci: Analisis Sentimen, Kalimat Negasi, *K-Nearest Neighbor*, *Lexicon*, Algoritma, *Confusion Matrix*, *Cross Validation*

TWITTER SENTIMENT ANALYSIS USING K-NEAREST NEIGHBOR AND LEXICON APPROACH

DEFRI ROSDIANSYAH

11051102066

Final Exam Date: September 10th, 2014

Graduation Ceremony Period: November 2014

Informatics Engineering Department

Faculty of Science and Technology

Universitas Islam Negeri Sultan Syarif Kasim Riau

ABSTRACT

Sentiment analysis also called opinion mining. Study of sentiment analysis can be research on twitter . However, a number of studies sentiment analysis of Twitter Indonesian sentence negation not be evaluated. The study will be conducted orientation sentiment classification into 3 types: positive, negative and neutral using a combination of K-Nearest Neighbor (k-NN) approach and lexicon. The topic of this study is to determine the sentiment analysis of political cases, Especially public sentiment against political figures on the presidential candidate of the Republic of Indonesia 2014. The k-NN algorithm in this study will act as a classification algorithm that will perform temporary sentiment orientation. While the lexicon approach to identifying sentence neutral on preprocessing query stage also conduct evaluation of sentence negation. The testing of sentiment analysis of Twitter using confusion matrix models to test data and 10-fold Cross Validation to the evaluation of training data to get the best training data. tests were also done on the determination of the value of k in the k-NN algorithm. Conclusion of testing a combination of k-NN and lexicon can improve the accuracy of sentiment orientation. The best value of k in k-NN algorithm is k = 5 and k = 15. While the best method is a combination approach that uses a lexicon B and k-NN algorithm by about 82% in accuracy.

Key Words: Sentiment Analysis, Sentence Negation, K-Nearest Neighbor, Lexicon, Algorithm, Confusion Matrix, Cross Validation

KATA PENGANTAR



Alhamdulillaahi Robbil'alamin, penulis bersyukur ke-hadirat Allah SWT, karena atas segala limpahan rahmat dan karunia-Nya yang diberikan sehingga penulis dapat menyelesaikan penelitian dan penulisan laporan tugas akhir ini. *Allahumma sholli'ala Muhammad wa'ala ali sayyidina Muhammad*, yang tidak lupa penulis haturkan juga untuk Rosul Allah, Muhammad SAW.

Laporan tugas akhir ini merupakan salah satu prasyarat untuk memenuhi persyaratan akademis dalam rangka meraih gelar kesarjanaan di Jurusan Teknik Informatika, Fakultas Sains dan Teknologi, Universitas Islam Negeri Sultan Syarif Kasim Riau (UIN SUSKA Riau). Selama menyelesaikan tugas akhir ini, penulis telah banyak mendapatkan bantuan, bimbingan, dan petunjuk dari banyak pihak baik secara langsung maupun tidak langsung. Untuk itu dalam kesempatan ini penulis ingin mengucapkan terimakasih yang sebesar-besarnya kepada:

1. Prof. Dr. H. M. Nazir, selaku Rektor Universitas Islam Negeri Sultan Syarif Kasim Riau.
2. Dra. Yenita Morena, M.Si, selaku Dekan Fakultas Sains dan Teknologi Universitas Islam Negeri Sultan Syarif Kasim Riau.
3. Elin Haerani, ST, M.Kom, selaku Ketua Jurusan Teknik Informatika, Fakultas Sains dan Teknologi.
4. Surya Agustian, ST, M.Kom, selaku dosen pembimbing tugas akhir. Terimakasih pak untuk waktu yang selalu bapak luangkan untuk penulis, ilmu, semangat, dan motivasinya yang luar biasa.
5. M. Fikry, S.T, M.Sc, selaku dosen penguji 1 yang membantu dan memberi masukan penulis dalam penyempurnaan laporan tugas akhir ini.
6. M. Safrizal, S.T, M.Cs, selaku dosen penguji 2 yang membantu dan memberi masukan penulis dalam penyempurnaan laporan tugas akhir ini.
7. Yusra, ST, MT, yang membantu dan memberi masukan penulis dalam penyempurnaan laporan tugas akhir ini.

8. Muhammad Affandes, MT, sebagai koordinator tugas akhir yang telah memberi masukan arahan pada penulis untuk penyelesaian tugas akhir ini. Trimakasih pak waktu luang dan arahannya.
9. Orang Tua penulis serta kakak abang dan keluarga besar yang selalu memberikan doa, dorongan dan semangat untuk penulis.
10. Seluruh staff dan karyawan Fakultas Sains & Teknologi Khususnya Jurusan Teknik Informatika.
11. Teman-teman karib satu angkatan 2010 khususnya teman-teman kelas D, terima kasih atas kebersamaannya.
12. Keluarga besar PAS Media yang telah memberikan semangat dan dukungan.
13. Teman-teman karib satu angkatan 2010 Teknik Informatika.
14. Seluruh pihak yang turut membantu penulis dalam bentuk apapun.

Akhirnya, penulis menyadari dalam penulisan laporan ini masih terdapat kekurangan. Oleh karena itu, saran dan kritik sangat penulis harapkan untuk kemajuan penulis secara pribadi. Terimakasih.

Pekanbaru, 10 September 2014

Defri Rosdiansyah

DAFTAR ISI

	Halaman
HALAMAN JUDUL LAPORAN	i
LEMBAR PERSETUJUAN	ii
LEMBAR PENGESAHAN	iii
LEMBAR HAK ATAS KEKAYAAN INTELEKTUAL.....	iv
LEMBAR PERNYATAAN.....	v
LEMBAR PERSEMBAHAN	vi
ABSTRAK	vii
<i>ABSTRACT</i>	viii
KATA PENGANTAR	ix
DAFTAR ISI	xi
DAFTAR GAMBAR	xiv
DAFTAR TABEL	xvi
DAFTAR RUMUS	xviii
DAFTAR SIMBOL	xix
BAB I PENDAHULUAN	
1.1 Latar Belakang	I-1
1.2 Rumusan Masalah.....	I-4
1.3 Batasan Masalah	I-4
1.4 Tujuan	I-4
1.5 Sistematika Penulisan	I-5
BAB II LANDASAN TEORI	
2.1 Sentimen Analisis	II-1
2.2.1 Sentimen Analisis pada Twitter	II-2
2.2 Struktur Data Twitter	II-3
2.3 Metode Klasifikasi Analisis Sentimen.....	II-4
2.2.1 <i>Supervised Learning</i>	II-4

2.2.2 <i>Lexicon (Unsupervised Aproche)</i>	II-5
2.3 <i>Text Mining</i>	II-6
2.3.1 Pembangunan Index	II-8
2.3.2 Algoritma Nazief Adriani	II-10
2.3.3 Pembobotan Kata	II-13
2.4 Algoritma <i>k-Nearest Neighbor (k-NN)</i>	II-15
BAB III METODOLOGI PENELITIAN	
3.1 Identifikasi Masalah	III-1
3.2 Merumuskan Masalah	III-2
3.3 Study Literatur	III-2
3.4 Analisa Aplikasi	III-2
3.5 Perancangan Aplikasi.....	III-5
3.6 Implementasi	III-6
3.7 Pengujian.....	III-6
3.8 Kesimpulan dan Saran	III-7
BAB IV ANALISA DAN PERANCANGAN	
4.1 Analisa Orientasi Sentimen Twitter.....	IV-1
4.1.1 Analisa Penentuan <i>Query</i>	IV-2
4.1.2 Analisa Menggunakan Algoritma <i>k-NN</i>	IV-3
4.1.3 Analisa Menggunakan Pendekatan <i>Lexicon</i>	IV-21
4.1.4 Analisa Kombinasi Algortima k-NN dan <i>Lexicon</i>	IV-25
4.1 Perancangan Aplikasi	IV-27
4.1.1 Perancangan File Teks (<i>Flat File</i>)	IV-28
4.1.2 Perancangan Database Relasional.....	IV-29
4.1.3 Perancangan Struktur Menu	IV-30
4.1.4 Perancangan <i>Interface</i>	IV-30
BAB V IMPLEMENTASI dan PENGUJIAN	
5.1 Implementasi	V-1
5.1.1 Batasan Implementasi	V-1
5.1.2 Lingkungan Implementasi	V-1
5.1.3 Implementasi Antarmuka Aplikasi	V-2

5.2 Hipotesa Pengujian	V-6
5.3 Pengujian	V-7
5.3.1 Rencana Pengujian	V-7
5.3.2 Implementasi Pengujian	V-8
5.3.3 Kesimpulan Pengujian	V-25

BAB VI PENUTUP

6.1 Kesimpulan	VI-1
6.2 Saran	VI-1

DAFTAR PUSTAKA

DAFTAR RIWAYAT HIDUP

DAFTAR GAMBAR

Gambar	Halaman
2.1 Sistem Arsitektur <i>Text Mining</i>	II-7
3.1 Tahapan Penelitian.....	III-1
4.1 <i>Flowchart</i> Penentuan <i>Query</i>	IV-3
4.2 <i>Flowchart Diagram</i> Analisa Algoritma <i>k-NN</i>	IV-3
4.3 Proses Pembuatan <i>Inverted Index</i>	IV-4
4.4 Contoh Proses <i>Case Folding</i>	IV-4
4.5 Contoh Proses Menghapus Karakter Selain UTF-8 dan <i>link url</i>	IV-5
4.6 Contoh Proses <i>Tokenization</i>	IV-5
4.7 <i>Flowchart</i> Algoritma Nazief dan Adriani.....	IV-6
4.8 Contoh Proses <i>Linguistic Preprocessing: Stopword dan Stemming</i>	IV-6
4.9 <i>Flowchart</i> Normalisasi.....	IV-8
4.10 Contoh Proses Normalisasi	IV-8
4.11 Proses Pembuatan <i>Inverted Index Query</i>	IV-15
4.12 <i>Flowchart</i> Proses Klasifikasi <i>Query</i>	IV-17
4.13 <i>Flowchart Diagram</i> Analisa Pendekatan <i>Lexicon</i>	IV-21
4.14 Proses Orientasi Sentimen pada Pendekatan <i>Lexicon</i>	IV-23
4.15 <i>Flowchart Diagram</i> Analisa Kombinasi A	IV-26
4.16 <i>Flowchart Diagram</i> Analisa Kombinasi B	IV-27
4.17 Rancangan Struktur Menu	IV-30
4.18 Rancangan <i>Interface</i> Halaman Beranda.....	IV-31
4.19 Rancangan <i>Interface</i> Halaman Data Latih	IV-32
4.20 Rancangan <i>Interface</i> Halaman Tambah Data Latih	IV-32
4.21 Rancangan <i>Interface</i> Halaman Pengujian	IV-33
4.22 Rancangan <i>Interface</i> Halaman Lihat Akurasi.....	IV-33
4.23 Rancangan <i>Interface</i> Halaman Hasil Pengujian	IV-34
4.24 Rancangan <i>Interface</i> Halaman Sub Menu Ambil Data.....	IV-34

4.25	Rancangan <i>Interface</i> Halaman Sub Menu Ambil Data.....	IV-35
4.26	Rancangan <i>Interface</i> Halaman Sub Menu Lihat Hasil	IV-35
4.27	Rancangan <i>Interface</i> Halaman Pengaturan	IV-36
5.1	Tampilan Menu Beranda.....	V-2
5.2	Tampilan Menu Data Latih	V-2
5.3	Tampilan Tambah Data Latih	V-2
5.4	Tampilan Menu Pengujian.....	V-4
5.5	Tampilan Lihat Akurasi	V-4
5.6	Tampilan Menu Hasil Pengujian	V-5
5.7	Tampilan Sub Menu Ambil Data	V-5
5.8	Tampilan Menu Pengaturan	V-6

DAFTAR TABEL

Tabel	Halaman
4.1 Dokumen Data Latih.....	IV-9
4.2 Hasil <i>Case Folding</i> Data Latih	IV -9
4.3 Hasil Penghapusan Karakter selain UTF-8 dan <i>link URL</i>	IV -10
4.4 Hasil Proses Tokenisasi Setiap Dokumen.....	IV -11
4.5 Hasil Proses <i>Linguistic Preprocessing Stopword</i> dan <i>Steamming</i>	IV -11
4.6 Hasil Proses Normalisasi	IV -12
4.7 Hasil Proses Indexing (Pembuatan <i>Inverted Index</i>).....	IV -12
4.8 Hasil Pembobotan Kata.....	IV -13
4.9 Hasil Bobot Kata terhadap Kata Hasil <i>Indexing</i>	IV -17
4.10 Hasil Identifikasi Kata Sentimen	IV -24
4.11 Hasil Identifikasi Kata Negasi	IV -24
4.12 Rancangan File Teks (<i>Flat File</i>)	IV -28
4.13 Konseptual Data Model Tabel Hak Akses	IV -29
4.14 Konseptual Data Model Tabel Hasil Pengujian	IV -29
4.15 Konseptual Data Model Tabel Konfigurasi	IV -30
5.1 Hasil <i>10-Fold Cross Validation</i> Datalatih	V-8
5.2 Hasil Pengujian nilai <i>k</i> pada Algoritma <i>k</i> -NN	V -9
5.3 Hasil Pengujian Dokumen 1 dengan Sentimen Netral.....	V -10
5.4 Hasil Pengujian Dokumen 2 dengan Sentimen Netral.....	V -11
5.5 Hasil Pengujian Dokumen 3 dengan Sentimen Netral.....	V -11
5.6 Hasil Pengujian Dokumen 4 dengan Sentimen Positif	V -12
5.7 Hasil Pengujian Dokumen 5 dengan Sentimen Positif	V -12
5.8 Hasil Pengujian Dokumen 6 dengan Sentimen Positif	V -13
5.9 Hasil Pengujian Dokumen 7 dengan Kalimat Negatif.....	V -13
5.10 Hasil Pengujian Dokumen 8 dengan Kalimat Negatif.....	V -14
5.11 Hasil Pengujian Dokumen 9 dengan Kalimat Negatif.....	V -15
5.12 Hasil Pengujian Dokumen 10 dengan Kalimat Negasi.....	V -15

5.13	Hasil Pengujian Dokumen 11 dengan Kalimat Negasi.....	V -16
5.14	Hasil Pengujian Dokumen 12 dengan Kalimat Negasi.....	V -17
5.15	Hasil Pengujian Dokumen 13 dengan Klausula Tapi-	V -17
5.16	Hasil Pengujian Dokumen 14 dengan Klausula Tapi-	V -18
5.17	Hasil Pengujian Dokumen 15 dengan Klausula Tapi-	V -18
5.18	Hasil Pengujian Dokumen 16 dengan dengan Sarkasme	V -19
5.19	Hasil Pengujian Dokumen 17 dengan dengan Sarkasme	V -20
5.20	Hasil Pengujian Dokumen 18 dengan dengan Sarkasme	V -20
5.21	Hasil Pengujian Akurasi <i>k</i> -NN.....	V -21
5.22	Hasil Pengujian Akurasi <i>Lexicon</i>	V -22
5.23	Hasil Pengujian Akurasi Kombinasi A	V -22
5.24	Hasil Pengujian Akurasi Kombinasi B	V -23
5.25	Hasil Pengujian Akurasi Keseluruhan	V -24
5.26	Hasil Pengujian <i>Real Time</i>	V -24

DAFTAR RUMUS

Rumus	Halaman
2.1 <i>Agregat Lexicon</i>	II-6
2.2 <i>Logarithm term frequency</i>	II-13
2.3 <i>Augmented Term Frequency</i>	II-14
2.4 <i>Inverse Document Frequency</i>	II-14
2.5 <i>Cosine Similarity</i>	II-15
2.6 Hitung banyak ketetanggan terdekat algoritma <i>k</i> -NN.....	II-15

DAFTAR SIMBOL

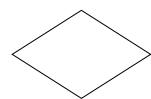
Flowchart



Terminator : Simbol *terminator* (Mulai/Selesai) merupakan tanda bahwa sistem akan dijalankan atau berakhir.



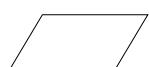
Proses : Simbol yang digunakan untuk melakukan pemrosesan data baik oleh *user* maupun komputer (sistem).



Verifikasi : Simbol yang digunakan untuk memutuskan apakah valid atau tidak validnya suatu kejadian.



Data Store : Simbol yang digunakan untuk mewakili suatu penyimpanan data (*database*).



Data : Simbol yang digunakan untuk mendeskripsikan data yang digunakan



Laporan : Simbol yang digunakan untuk menggambarkan laporan.