

**ANALISIS SENTIMEN PARIWISATA DI KOTA MALANG
MENGGUNAKAN METODE NAIVE BAYES DAN
SELEKSI FITUR QUERY EXPANSION RANKING**

SKRIPSI

Untuk memenuhi sebagian persyaratan
memperoleh gelar Sarjana Komputer

Disusun oleh:
Shima Fanissa
NIM: 135150200111053



PROGRAM STUDI TEKNIK INFORMATIKA
JURUSAN TEKNIK INFORMATIKA
FAKULTAS ILMU KOMPUTER
UNIVERSITAS BRAWIJAYA
MALANG
2017

PENGESAHAN

Analisis Sentimen Pariwisata di Kota Malang menggunakan Metode *Naive Bayes*
dan Seleksi Fitur *Query Expansion Ranking*

Skripsi

Diajukan untuk memenuhi sebagian persyaratan
memperoleh gelar Sarjana Komputer

Disusun Oleh :
Shima Fanissa
NIM: 135150200111053

Skripsi ini telah diuji dan dinyatakan lulus pada
15 Desember 2017
Telah diperiksa dan disetujui oleh:

Dosen Pembimbing I

Dosen Pembimbing II

M. Ali Fauzi, S.Kom, M.Kom
NIK: 201502 890101 1 001

Sigit Adinugroho, S.Kom., M.Sc
NIK: 201607 880701 1 000

Mengetahui
Ketua Jurusan Teknik Informatika

Tri Astoto Kurniawan, S.T, M.T, Ph.D
NIP: 19710518 200312 1 001

PERNYATAAN ORISINALITAS

Saya menyatakan dengan sebenar-benarnya bahwa sepanjang pengetahuan saya, di dalam naskah skripsi ini tidak terdapat karya ilmiah yang pernah diajukan oleh orang lain untuk memperoleh gelar akademik di suatu perguruan tinggi, dan tidak terdapat karya atau pendapat yang pernah ditulis atau diterbitkan oleh orang lain, kecuali yang secara tertulis disitasi dalam naskah ini dan disebutkan dalam daftar pustaka.

Apabila ternyata didalam naskah skripsi ini dapat dibuktikan terdapat unsur-unsur plagiasi, saya bersedia skripsi ini digugurkan dan gelar akademik yang telah saya peroleh (sarjana) dibatalkan, serta diproses sesuai dengan peraturan perundang-undangan yang berlaku (UU No. 20 Tahun 2003, Pasal 25 ayat 2 dan Pasal 70).

Malang, November 2017

Shima Fanissa

NIM: 135150200111053

KATA PENGANTAR

Dengan menyebut nama Allah SWT Yang Maha Pengasih dan Maha Penyayang. Puji syukur kehadirat Allah SWT karena limpahan rahmat dan hidayah-Nya sehingga penulis dapat menyelesaikan skripsi dengan judul “Analisis Sentimen Pariwisata di Kota Malang menggunakan Metode *Naive Bayes* dan Seleksi Fitur *Query Expansion Ranking*”. Skripsi ini disusun untuk memenuhi salah satu syarat memperoleh gelar Sarjana Komputer di Program Studi Teknik Informatika Jurusan Teknik Informatika Fakultas Ilmu Komputer Universitas Brawijaya Malang (FILKOM UB).

Pada kesempatan ini penulis juga ingin menyampaikan terima kasih yang sebesar-besarnya kepada semua pihak yang telah memberikan bantuan dan dukungan dalam menyelesaikan skripsi ini, antara lain :

1. Tuhan yang Maha Esa, Allah SWT.
2. Ibuk Ismuwati, Bapak Paryono, Pak Jadi, Mbak Ika, Mas Nofel, Mas Nas, Mbak Devi, Chila, Taqiy, keluarga besar Mukaeni dan Mbah Kromo karena kalian adalah keluarga terbaik.
3. M. Ali Fauzi, S.Kom., M.Kom., selaku Dosen Pembimbing I yang telah membimbing, mengarahkan dan memberikan banyak saran kepada penulis sehingga dapat menyelesaikan skripsi ini.
4. Sigit Adinugroho, S.Kom., M.Sc selaku Dosen Pembimbing II yang telah membimbing, mengarahkan dan memberikan banyak saran kepada penulis sehingga dapat menyelesaikan skripsi ini.
5. Teman-teman saya yang menemani saya selama kuliah yaitu Sambalado, MVC, Arek RWT, BIOS Tari Karawitan. Teman-teman SD, SMP, SMA, dll
6. Dosen dan Karyawan FILKOM UB.
7. Lalisa.id, Black Pink, Sechskies, Bigbang, YG Family.

Dan untuk pembaca skripsi ini, saya ucapan terima kasih banyak. Saya berpesan jangan takut dengan skripsi karena skripsi akan berlalu, jika takut maka tidak akan pernah selesai. Semoga Tuhan melimpahkan rahmatNya untuk kita.

Dengan kerendahan hati, penulis menyadari bahwa skripsi ini memiliki banyak kekurangan. Oleh karena itu kritik dan saran yang bersifat membangun sangat dibutuhkan untuk penelitian selanjutnya. Penulis berharap semoga skripsi ini dapat bermanfaat bagi diri sendiri maupun bagi semua pihak.

Malang, Oktober 2017

Shima Fanissa

Shimafanissa@gmail.com

ABSTRAK

Pariwisata merupakan salah satu sektor usaha untuk memajukan suatu daerah. Kota Malang pada saat ini memiliki sebuah branding city yang bernama “*Beautiful Malang*”. Masyarakat Indonesia memilih pariwisata Kota Malang sebagai destinasi dan mengulasnya pada *website* salah satunya adalah TripAdvisor. Dengan demikian penelitian ini mencoba menganalisis ulasan dari masyarakat tentang pariwisata Kota Malang melalui analisis sentimen dan diklasifikasikan menjadi dua kelas yaitu positif dan negatif. Pada penelitian ini metode yang digunakan adalah *Naive Bayes* dan seleksi fitur *Query Expansion Ranking* untuk mengurangi jumlah fitur pada proses klasifikasi. Proses dari analisis sentimen terdiri dari preprocessing, seleksi fitur dengan metode *Query Expansion Ranking*, dan klasifikasi dengan *Naive Bayes*. Pengujian pada penelitian ini adalah uji akurasi dengan menggunakan variasi rasio seleksi fitur, hasilnya seleksi fitur 75% memiliki akurasi terbaik sebesar 86.6%.

Kata kunci: Analisis Sentimen, *Query Expansion Ranking*, *Naïve Bayes*, *Multinomial Naïve Bayes*

ABSTRACT

Tourism is one of effort to promote a city. Malang currently has a branding city called "Beautiful Malang". Indonesian choose Malang tourism as a destination and review it on the website, one of them is TripAdvisor. Thus this research tried to analyze the reviews from the public about the tourism of Malang City through sentiment analysis and classified into two classes, that is positive and negative. In this research the method used is Naive Bayes with Query Expansion Ranking feature selection to reduce the number of features in the classification process. The process of sentiment analysis consists of preprocessing, feature selection with Query Expansion Ranking method, and classification with Naive Bayes. This research is testing the accuracy by using the variation of feature selection ratio, the result of 75% feature selection has the best accuracy of 86.6%.

Keywords: Sentiment Analysis, Query Expansion Ranking, Naïve Bayes, Multinomial Naïve Bayes

DAFTAR ISI

PENGESAHAN	ii
PERNYATAAN ORISINALITAS	iii
KATA PENGANTAR.....	iv
ABSTRAK.....	v
ABSTRACT.....	vi
DAFTAR ISI	vii
DAFTAR TABEL.....	x
DAFTAR GAMBAR.....	xi
DAFTAR LAMPIRAN	xii
BAB 1 PENDAHULUAN.....	1
1.1 Latar belakang.....	1
1.2 Rumusan masalah.....	2
1.3 Tujuan	3
1.4 Manfaat.....	3
1.4.1 Bagi Peneliti.....	3
1.4.2 Bagi Pemerintah Kota Malang	3
1.4.3 Bagi Masyarakat.....	3
1.5 Batasan masalah	3
1.6 Sistematika pembahasan.....	4
BAB 2 LANDASAN KEPUSTAKAAN	5
2.1 Tinjauan Penelitian	5
2.2 Pariwisata.....	5
2.3 TripAdvisor®	5
2.4 Pemrosesan Teks	6
2.5 <i>Preprocessing</i>	6
2.5.1 Tokenisasi.....	6
2.5.2 <i>Filtering</i>	7
2.5.3 <i>Stemming</i>	8
2.6 Analisis Sentimen.....	10
2.7 <i>Naive Bayes Classifier</i>	10

2.7.1 <i>Multinomial Naive Bayes</i>	11
2.8 Seleksi Fitur.....	12
2.8.1 <i>Query Expansion Ranking</i>	12
2.9 Evaluasi	13
BAB 3 METODOLOGI	15
3.1 Studi literatur.....	15
3.2 Analisis Kebutuhan	16
3.2.1 Kebutuhan Fungsional.....	16
3.2.2 Kebutuhan Non Fungsional.....	16
3.3 Pengumpulan Data	16
3.4 Perancangan Sistem.....	17
3.5 Implementasi	18
3.6 Pengujian dan Analisis	18
3.7 Kesimpulan.....	18
BAB 4 PERANCANGAN.....	19
4.1 Deskripsi Sistem	19
4.2 Preprocessing.....	19
4.2.1 Tokenisasi.....	20
4.2.2 <i>Filtering</i>	21
4.2.3 <i>Stemming</i>	22
4.3 Seleksi Fitur <i>Query Expansion Ranking</i>	25
4.4 Klasifikasi Sentimen menggunakan <i>Naïve Bayes</i>	27
4.5 Manualisasi	28
4.5.1 Data	28
4.5.2 <i>Preprocessing</i>	29
4.5.3 Seleksi Fitur <i>Query Expansion Ranking</i>	31
4.5.4 Klasifikasi Sentimen menggunakan <i>Naïve Bayes</i>	32
4.6 Perancangan Pengujian	35
BAB 5 IMPLEMENTASI	37
5.1 Preprocessing.....	37
5.2 Seleksi Fitur <i>Query Expansion Ranking</i>	38
5.3 Klasifikasi Sentimen dengan <i>Naïve Bayes</i>	39

BAB 6 PENGUJIAN DAN ANALISIS.....	42
6.1 Pengujian Variasi Rasio Seleksi Fitur.....	42
6.1.1 Skenario Pengujian Variasi Rasio Seleksi Fitur.....	42
6.1.2 Analisis Pengujian Variasi Rasio Seleksi Fitur.....	42
BAB 7 PENUTUP	44
7.1 Kesimpulan.....	44
7.2 Saran	44
DAFTAR PUSTAKA.....	45
LAMPIRAN A DATASET LATIH ULASAN PARIWISATA DI KOTA MALANG	47
LAMPIRAN B DATASET UJI ULASAN PARIWISATA DI KOTA MALANG	83
LAMPIRAN C DAFTAR FITUR DATA LATIH	87
LAMPIRAN D SURAT KESEDIAAN PAKAR.....	92

DAFTAR TABEL

Tabel 2.1 <i>Confusion Matrix</i>	14
Tabel 4.1 Data latih manualisasi	28
Tabel 4.2 Hasil manualisasi tokenisasi	29
Tabel 4.3 Hasil manualisasi <i>filtering</i>	30
Tabel 4.4 Hasil manualisasi <i>stemming</i>	30
Tabel 4.5 Manualisasi <i>query expansion ranking</i>	31
Tabel 4.6 Seleksi fitur	32
Tabel 4.7 Data uji	32
Tabel 4.8 Tokenisasi data uji	33
Tabel 4.9 <i>Filtering</i> data uji.....	33
Tabel 4.10 <i>Stemming</i> data uji	33
Tabel 4.11 Hasil pencocokan data uji dan seleksi fitur	33
Tabel 4.12 Hasil Peluang setiap Fitur	34
Tabel 4.13 Nilai <i>Naive Bayes</i>	35
Tabel 4.14 Perancangan pengujian.....	35
Tabel 6.1 Hasil pengujian seleksi fitur.....	42

DAFTAR GAMBAR

Gambar 2.1 Proses tokenisasi	7
Gambar 2.2 Proses <i>filtering</i>	8
Gambar 2.3 Proses <i>stemming</i>	10
Gambar 3.1 Diagram blok metode penelitian	15
Gambar 3.2 Arsitektur sistem	17
Gambar 4.1 Diagram alir sistem.....	19
Gambar 4.2 Diagram alir <i>preprocessing</i>	20
Gambar 4.3 Diagram alir tokenisasi.....	21
Gambar 4.4 Diagram alir <i>filtering</i>	21
Gambar 4.5 Diagram alir <i>stemming</i>	22
Gambar 4.6 Diagram alir <i>stemming</i> kata jamak pada <i>library</i> sastrawi.....	23
Gambar 4.7 Diagram alir <i>stemming</i> kata tunggal	24
Gambar 4.8 Diagram alir seleksi fitur <i>query expansion ranking</i>	25
Gambar 4.9 Hitung nilai pf	26
Gambar 4.10 Hitung nilai qf	26
Gambar 4.11 Diagram alir klasifikasi sentimen dengan <i>naive bayes</i>	28
Gambar 6.1 Pengujian seleksi fitur	43

DAFTAR LAMPIRAN

LAMPIRAN A DATASET LATIH ULASAN PARIWISATA DI KOTA MALANG	47
LAMPIRAN B DATASET UJI ULASAN PARIWISATA DI KOTA MALANG	83
LAMPIRAN C DAFTAR FITUR DATA LATIH	87
LAMPIRAN D SURAT KESEDIAAN PAKAR.....	92