

“ANALISIS SENTIMEN DARI MEDIA SOSIAL MENGUNAKAN METODE *INFORMATION GAIN* DAN *NAÏVE BAYES*”

Muhamad Fahmi Ansori

(14.04.1.1.1.00102)

Pembimbing I : Mula'ab, S.Si., M.Kom.

Pembimbing II : Eka Mala Sari R, S.Kom., M.Kom.

ABSTRAK

Pada media sosial, pengguna dapat dengan bebas mengekspresikan pendapat atau opini. Media sosial dapat menjadi alternatif solusi untuk melakukan analisis sentimen disamping survei konvensional yang memerlukan biaya mahal dan waktu yang cukup lama. Salah satu metode yang banyak digunakan untuk analisis sentimen adalah *naïve bayes*. Metode ini mempunyai akurasi yang baik untuk klasifikasi sentimen dan efisien dalam hal komputasi. Proses analisis akan dimulai dengan pengumpulan data, kemudian melakukan preprocessing, seleksi fitur dan pembuatan model klasifikasi. Seleksi fitur dilakukan menggunakan metode *information gain*, dengan menghitung keterkaitan kata terhadap kelas dan klasifikasi akan dilakukan dengan *naïve bayes*, dengan menghitung probabilitas setiap kata terhadap kelas. Penelitian ini bertujuan untuk mengetahui bagaimana performa metode *information gain* dan *naïve bayes* dalam analisis sentimen. Hasil pengukuran dari performa metode antara lain nilai akurasi rata-rata 56,6%, *precision* 0,33, dan *recall* sebesar 0,39.

Kata kunci: analisis sentimen, klasifikasi, seleksi fitur, *naïve bayes*, *information gain*.

DAFTAR ISI

HALAMAN PENGESAHAN.....	iii
HALAMAN PERNYATAAN ORISINALITAS	v
HALAMAN PERSEMBAHAN.....	vii
ABSTRAK.....	ix
KATA PENGANTAR	xi
DAFTAR ISI.....	xiii
DAFTAR GAMBAR	xv
DAFTAR TABEL.....	xvii
BAB I PENDAHULUAN	1
1.1. Latar Belakang	1
1.2. Rumusan Masalah	2
1.3. Tujuan dan Manfaat Penelitian.....	2
1.4. Batasan Masalah.....	2
1.5. Metode Penelitian.....	3
1.6. Sistematika Penulisan.....	3
BAB II STUDI PUSTAKA.....	5
2.1. Penelitian Sebelumnya	5
2.2. Analisis Sentimen.....	7
2.3. Media Sosial.....	8
2.4. <i>Text Mining</i>	8
2.5. <i>Preprocessing</i>	10
2.6. Seleksi Fitur	11
2.7. Klasifikasi	13
2.8. <i>Cross Validation</i>	14
2.9. Contoh Perhitungan.....	16
BAB III PERANCANGAN SISTEM	23

3.1. Input data	23
3.2. Preprocessing	24
3.3. Vector Space Model.....	24
3.4. Seleksi fitur	24
3.5. Training model klasifikasi	25
3.6. Pengujian dan evaluasi.....	25
BAB IV IMPLEMENTASI SISTEM	27
4.1. Lingkungan pengembangan	27
4.2. Pembuatan program	27
4.2.1. Pembuatan GUI	27
4.2.2. Memuat dataset	30
4.2.3. <i>Preprocessing</i> dan pembuatan VSM	30
4.2.4. Training model klasifikasi.....	32
4.2.5. Evaluasi	33
4.3. Evaluasi dan hasil analisa sistem	34
BAB V PENUTUP	51
5.1. Kesimpulan	51
5.2. Saran	51
DAFTAR PUSTAKA	53

DAFTAR GAMBAR

Gambar 2-1. Proses pada <i>text mining</i>	9
Gambar 2-2. Confusion matrix.	15
Gambar 3-1. Rancangan sistem.....	23
Gambar 4-1. Rancangan GUI.....	27
Gambar 4-2. Tampilan tab setup.	28
Gambar 4-3. Tampilan tab <i>fix word</i>	29
Gambar 4-4. Tampilan tab <i>evaluation</i>	29
Gambar 4-5. Uji coba memuat data.	34
Gambar 4-6. Uji coba memperbaiki kata.	35
Gambar 4-7. Uji coba proses evaluasi.....	35
Gambar 4-8. Hasil akurasi pada setiap skenario.	37
Gambar 4-9. Hasil <i>precision</i> pada setiap skenario.	37
Gambar 4-10. Hasil <i>recall</i> pada setiap skenario.....	38
Gambar 4-11. Hasil akurasi setelah pengaturan threshold seleksi fitur.	39
Gambar 4-12. Hasil <i>precision</i> setelah pengaturan threshold seleksi fitur.	39
Gambar 4-13. Hasil <i>recall</i> setelah pengaturan threshold seleksi fitur... ..	40
Gambar 4-14. Hasil akurasi pada skenario tambahan.	42
Gambar 4-15. Hasil <i>precision</i> pada skenario tambahan.	43
Gambar 4-16. Hasil <i>recall</i> pada skenario tambahan.	43
Gambar 4-17. Hasil akurasi setelah pengaturan threshold seleksi fitur pada skenario 6.	44
Gambar 4-18. Hasil <i>precision</i> setelah pengaturan threshold seleksi fitur pada skenario 6.	44
Gambar 4-19. Hasil <i>recall</i> setelah pengaturan threshold seleksi fitur pada skenario 6.	45
Gambar 4-20. Perbandingan akurasi.	46
Gambar 4-21. Perbandingan nilai <i>precision</i>	46
Gambar 4-22. Perbandingan nilai <i>recall</i>	47
Gambar 4-23. Hasil prediksi kalimat positif.	48
Gambar 4-24. Hasil prediksi kalimat negatif.	48
Gambar 4-25. Hasil prediksi kalimat netral.	48
Gambar 4-26. Hasil klasifikasi kalimat negasi.....	48

DAFTAR TABEL

Tabel 2-1. Penelitian sebelumnya.	6
Tabel 4-1. Lingkungan pengembangan.	27
Tabel 4-2. Hasil dari skenario uji coba.	36
Tabel 4-3. Hasil dari pengaturan threshold pada seleksi fitur untuk skenario 3.	36
Tabel 4-4. Hasil dari skenario uji coba tambahan.	41
Tabel 4-5. Hasil dari pengaturan threshold pada seleksi fitur untuk skenario 6.	41