



UNIVERSITAS
MERCU BUANA

**IMPLEMENTASI NAIVE BAYES DENGAN PENDEKATAN
LEKSIKON DAN TF-IDF DALAM ANALISIS SENTIMEN**

LAPORAN TUGAS AKHIR

Huzaifah Qahar Djauhari 41820010070

Vincent Lius 41820010037

Ahmad Harits Ramadhani 41820010051

**PROGRAM STUDI SISTEM INFORMASI
FAKULTAS ILMU KOMPUTER
UNIVERSITAS MERCU BUANA
JAKARTA
2024**



UNIVERSITAS
MERCU BUANA

**IMPLEMENTASI NAIVE BAYES DENGAN PENDEKATAN
LEKSIKON DAN TF-IDF DALAM ANALISIS SENTIMEN**

LAPORAN TUGAS AKHIR

Huzaifah Qahar Djauhari 41820010070

Vincent Lius 41820010037

Ahmad Harits Ramadhani 41820010051

Diajukan sebagai salah satu syarat untuk memperoleh gelar sarjana

PROGRAM STUDI SISTEM INFORMASI

FAKULTAS ILMU KOMPUTER

UNIVERSITAS MERCU BUANA

JAKARTA

2024

HALAMAN PERNYATAAN KARYA SENDIRI

Saya yang bertanda tangan di bawah ini:

Nama Mahasiswa : Huzaiifah Qahar Djauhari
NIM : 41820010070
Program Studi : Sistem Informasi
Judul Laporan : IMPLEMENTASI NAIVE BAYES DENGAN
Skripsi : PENDEKATAN LEKSIKON DAN TF-IDF DALAM
ANALISIS SENTIMEN

Menyatakan bahwa Laporan Tugas Akhir ini adalah hasil karya saya sendiri dan bukan plagiat, serta semua sumber baik yang dikutip maupun dirujuk telah saya nyatakan dengan benar. Apabila ternyata ditemukan di dalam Laporan Tugas Akhir saya terdapat unsur plagiat, maka saya siap mendapatkan sanksi akademis yang berlaku di Universitas Mercu Buana.



Jakarta, 19 Juni 2024



Huzaiifah Qahar Djauhari

UNIVERSITAS
MERCU BUANA

LEMBAR PENGESAHAN

Nama Mahasiswa (1) : Huzaifah Qahar Djauhari
 NIM : 41820010070
 Nama Mahasiswa (2) : Vincent Lius
 NIM : 41820010037
 Nama Mahasiswa (3) : Ahmad Harits Ramadhani
 NIM : 41820010051
 Judul Tugas Akhir : IMPLEMENTASI NAIVE BAYES DENGAN
 PENDEKATAN LEKSIKON DAN TF-IDF
 DALAM ANALISIS SENTIMEN

Tugas Akhir ini telah diperiksa dan disidangkan sebagai salah satu persyaratan untuk memperoleh gelar Sarjana pada Program Studi Sistem Informasi, Fakultas Ilmu Komputer, Universitas Mercu Buana.

Jakarta, 19 Juni 2024

Menyetujui

Pembimbing : Andi Nugroho, ST, M.Kom
 NIDN : 0305098303
 Ketua Penguji : Fajar Masya, Ir. MMSI
 NIDN : 0313036701
 Penguji 1 : Kurnia Gusti Ayu, M.Kom
 NIDN : 0302088704
 Penguji 2 : Inna Sabily Karima, S.Kom, M.Kom
 NIDN : 0324018902





Mengetahui,



Dr. Bambang Jokonowo, S.Si., M.T.I.
 Dekan Fakultas Ilmu Komputer



Dr. Ruci Meiyanti, M.Kom
 Ka.Prodi Sistem Informasi

KATA PENGANTAR

Segala Puji dan Syukur saya panjatkan kehadiran Allah SWT yang telah memberikan Rahmat, Karunia, Dan Hidayah-nya sehingga penulis dalam menyelesaikan Tugas Akhir yang berjudul “IMPLEMENTASI NAIVE BAYES DENGAN PENDEKATAN LEKSIKON DAN TF-IDF DALAM ANALISIS SENTIMEN” sebagai salah satu syarat Program Sarjana (S1) pada Program Sarjana Fakultas Ilmu Komputer Jurusan Sistem Informasi Universitas Mercu Buana.

Penulis menyadari bahwa penulisan ini tidak dapat terselesaikan tanpa bantuan dari berbagai pihak. Dan dengan ini penulis ingin mengucapkan terimakasih sebesar besar kepada semua pihak yang terlibat dalam penyusunan skripsi ini terutama:

1. Prof. Dr. Ir. Andi Adriansyah, M.Eng selaku Rektor Universitas Mercu Buana
2. Dr. Bambang Jokonowo, S.Si., M.T.I. selaku Dekan Fakultas Ilmu Komputer.
3. Dr. Ruci Meiyanti, M.Kom, selaku Kepala Program Studi Sistem Informasi Fakultas Ilmu Komputer.
4. Dosen pembimbing kami, Andi Nugroho, ST, M.Kom yang senantiasa memberikan arahan. Dan tanpa henti memberikan dukungan dan dengan senang hati selalu memotivasi kami untuk menyelesaikan Tugas Akhir ini dengan baik.
5. Kami ingin menyampaikan rasa terima kasih yang sebesar-besarnya kepada Kurnia Gusti Ayu, M.Kom dan Inna Sabily Karima, S.Kom, M.Kom, dosen penguji kami, yang telah memberikan arahan, dukungan, dan masukan berharga selama proses sidang Tugas Akhir. Kehadiran dan bimbingan beliau sangat berarti bagi kami, dan kami sangat menghargai semua waktu serta usaha yang telah beliau curahkan. Terima kasih atas dedikasi dan komitmen yang luar biasa dalam membantu kami menyelesaikan Tugas Akhir ini dengan baik.
6. Bapak/Ibu Dosen dan Staff Universitas Mercu Buana, khususnya Fakultas Ilmu Komputer Program Studi Sistem Informasi yang telah memberikan ilmu dan pengalaman yang berharga di dalam maupun di luar kelas.
7. Kedua orang tua, serta keluarga yang selalu mendoakan, memberikan kasih sayang, memberi dukungan, serta membiayai dalam segala hal.
8. Seluruh pihak yang tidak dapat penulis sebutkan satu per satu, yang telah membantu penulis dalam menyelesaikan Tugas Akhir ini.

Lalu ucapan terima kasih oleh Huzaifah kepada :

1. Nur Fadilah, selaku kekasih tersayang yang telah membantu dan memotivasi saya selama mengerjakan penelitian dan penulisan Tugas Akhir ini. Terima kasih telah hadir dan selalu ada untuk saya dalam keadaan apapun, semoga segala cita-cita dan tujuan kita tercapai dan terpenuhi.

Besar Harapan kami, laporan tugas akhir ini dapat bermanfaat bagi kebaruan ilmu pengetahuan, terkhusus bidang studi ilmu komputer. Semoga dapat bermanfaat pula bagi para pembaca untuk dijadikan referensi dalam melakukan penelitian yang lebih baik selanjutnya. Kritik dan saran dengan senang hati penulis terima untuk hal yang lebih baik kedepannya.

Akhir kata, penulis ucapkan terima kasih.

Jakarta, 19 Juni 2024



Huzaifah Qahar Djauhari



UNIVERSITAS
MERCU BUANA

**HALAMAN PERNYATAAN PERSETUJUAN PUBLIKASI TUGAS AKHIR
UNTUK KEPENTINGAN AKADEMIS**

Sebagai sivitas akademik Universitas Mercu Buana, saya yang bertanda tangan di bawah ini:

Nama Mahasiswa : Huzaifah Qahar Djauhari
NIM : 41820010070
Program Studi : Sistem Informasi
Judul Laporan : IMPLEMENTASI NAIVE BAYES DENGAN
Skripsi : PENDEKATAN LEKSIKON DAN TF-IDF DALAM
ANALISIS SENTIMEN

Demi pengembangan ilmu pengetahuan, dengan ini memberikan izin dan menyetujui untuk memberikan kepada Universitas Mercu Buana Hak Bebas Royalti Non-Eksklusif (Non-exclusive Royalty-Free Right) atas karya ilmiah saya yang berjudul di atas beserta perangkat yang ada (jika diperlukan).

Dengan Hak Bebas Royalti Non-Eksklusif ini Universitas Mercu Buana berhak menyimpan, mengalihmedia/format-kan, mengelola dalam bentuk pangkalan data (database), merawat, dan mempublikasikan Laporan Magang/Skripsi/Tesis/Disertasi saya selama tetap mencantumkan nama saya sebagai penulis/pencipta dan sebagai pemilik Hak Cipta.

Demikian pernyataan ini saya buat dengan sebenarnya.

Jakarta, 19 Juni 2024

Yang Menyatakan,



Huzaifah Qahar Djauhari

ABSTRAK

Nama Mahasiswa (1) : Huzaifah Qahar Djauhari
NIM (1) : 41820010070
Nama Mahasiswa (2) : Vincent Lius
NIM (2) : 41820010037
Nama Mahasiswa (3) : Ahmad Harits Ramadhani
NIM (3) : 41820010051
Dosen TA : Andi Nugroho, ST, M.Kom
Judul Tugas Akhir : Implementasi Naive Bayes Dengan Pendekatan Leksikon
Dan TF-IDF Dalam Analisis Sentimen

Di era digital dengan peningkatan penggunaan media sosial sebagai platform ekspresi publik, analisis sentimen yang akurat menjadi semakin penting, terutama untuk evaluasi ulasan aplikasi. Penelitian ini menekankan pentingnya klasifikasi sentimen yang akurat, terutama karena banyaknya ulasan aplikasi yang negatif. Kami mengusulkan pendekatan kombinasi yang mengintegrasikan algoritma Naive Bayes dengan pelabelan sentimen berbasis leksikon dan pembobotan TF-IDF untuk pelatihan model. Melalui penggunaan dataset yang terdiri dari 5000 ulasan diambil dari Kaggle mengenai ulasan aplikasi di Indonesia, penelitian ini mengeksplorasi leksikon bahasa Indonesia, khususnya InSet dan SentiStrengthID. Evaluasi efektivitas penggabungan Naive Bayes dengan TF-IDF dan metode berbasis leksikon menghasilkan kontribusi signifikan dalam pemahaman yang lebih mendalam tentang analisis sentimen. Dari evaluasi yang kami lakukan, kami mencapai kesimpulan bahwa leksikon SentiStrength menunjukkan kinerja yang lebih baik di semua parameter. Model yang menggunakan leksikon SentiStrengthID mencapai akurasi sebesar 86%, dengan nilai presisi 85%, recall 86%, dan F1-score 85%, yang lebih baik dibandingkan dengan penggunaan leksikon InSet, yang memiliki akurasi 70%, presisi 80%, recall 70%, dan F1-score 71%.

Kata kunci:

Analisis Sentimen, Leksikon InSet, Leksikon SentiStrengthID, Naive bayes, Ulasan Aplikasi.

ABSTRACT

Name (1) : Huzaifah Qahar Djauhari
NIM (1) : 41820010070
Name (2) : Vincent Lius
NIM (2) : 41820010037
Name (3) : Ahmad Harits Ramadhani
NIM (3) : 41820010051
Counsellor : Andi Nugroho, ST, M.Kom
Title : Implementation Naive Bayes Algorithm using Lexicon
Approach and TF-IDF for Sentiment Analysis of Reviews

In the digital era marked by the increased use of social media as a platform for public expression, accurate sentiment analysis has become crucial, particularly in evaluating app reviews. This research highlights the need for precise sentiment classification, driven by the prevalence of negative feedback in several app reviews. To address this challenge, we propose a combined approach that integrates the Naive Bayes algorithm with lexicon-based sentiment labeling and TF-IDF weighting for model training. Utilizing a dataset of 5000 reviews from Kaggle regarding app reviews in Indonesia, this study explores Indonesian lexicons, specifically InSet and SentiStrengthID, for sentiment labeling. Evaluating the effectiveness of combining Naive Bayes with TF-IDF and lexicon-based methods provides significant insights into sentiment analysis in the context of app reviews. Based on our evaluation, we concluded that the SentiStrength lexicon performs better on all parameters. The model using the SentiStrengthID lexicon achieved an accuracy of 86%, with a precision of 85%, recall of 86%, and F1-score of 85%, which is better compared to the InSet lexicon, which had an accuracy of 70%, precision of 80%, recall of 70%, and F1-score of 71%.

Keywords:

Sentiment Analysis, Lexicon InSet, Lexicon SentiStrengthID, Naive Bayes, App Review.

DAFTAR ISI

HALAMAN JUDUL	i
HALAMAN PERNYATAAN KARYA SENDIRI	ii
HALAMAN PENGESAHAN	iii
KATA PENGANTAR	iv
HALAMAN PERNYATAAN PERSETUJUAN PUBLIKASI TUGAS AKHIR UNTUK KEPENTINGAN AKADEMIS	vi
ABSTRAK	vii
ABSTRACT	viii
DAFTAR ISI.....	ix
DAFTAR TABEL	xii
DAFTAR GAMBAR.....	xiii
DAFTAR LAMPIRAN	xiv
BAB I PENDAHULUAN.....	1
1.1 Latar Belakang.....	1
1.2 Rumusan Masalah	3
1.3 Tujuan Penelitian	3
1.4 Batasan Masalah.....	3
1.4.1 Data	3
1.4.2 Metode.....	3
1.4.3 Hasil.....	4
1.5 Manfaat Penelitian	4
1.6 Sistematika Penulisan.....	5
BAB II TINJAUAN PUSTAKA.....	6
2.1 Teori/Konsep Terkait	6
2.1.1 Analisis Sentimen.....	6
2.1.6 Pendekatan Leksikon.....	9
2.1.6.1 Leksikon InSet.....	9
2.1.6.2 Leksikon SentiStrengthID	9
2.1.7 TF-IDF.....	10
2.1.8 Naive Bayes.....	11
2.1.9 <i>Confusion Matrix</i>	11
2.1.10 <i>Preprocessing Data</i>	13

2.2 Literature Review	15
2.3 Analisis Literature Review	28
BAB III METODE PENELITIAN	29
3.1 Deskripsi Sumber Data	29
3.2 Teknik Pengumpulan Data	31
3.3 Diagram Alir Penelitian	32
3.4 Pemrograman.....	33
BAB IV HASIL DAN PEMBAHASAN	45
4.1 Analisa Deskriptif.....	45
4.2 <i>Preprocessing Data</i>	45
4.2.1 <i>Cleansing Data</i>	45
4.2.2 <i>Case Folding</i>	46
4.2.3 <i>Tokenization</i>	46
4.2.4 <i>Stopwords Removal</i>	47
4.2.5 <i>Stemming</i>	48
4.2.6 Normalization.....	48
4.3 <i>Processing Data</i>	49
4.3.1 Pelabelan Data.....	49
4.3.2 Ekstraksi Fitur TF-IDF.....	52
4.3.3 Klasifikasi Split Data.....	53
4.3.4 Klasifikasi Algoritma Naive Bayes.....	54
4.4 Evaluasi Model.....	56
4.4.1 Lexicon InSet dengan Skenario (80:20).....	56
4.4.2 Lexicon InSet dengan Skenario (70:30).....	57
4.4.3 Lexicon InSet dengan Skenario (60:40).....	58
4.4.4 Lexicon SentiStrengthID dengan Skenario (80:20)	60
4.4.5 Lexicon SentiStrengthID dengan Skenario (70:30)	61
4.4.6 Lexicon SentiStrengthID dengan Skenario (60:40)	62

BAB V PENUTUP	64
5.1 Kesimpulan.....	64
5.2 Saran.....	64
DAFTAR PUSTAKA	65
LAMPIRAN	68



DAFTAR TABEL

Tabel 2.1 <i>Confusion Matrix</i>	12
Tabel 2.2 <i>Literature Review</i>	16
Tabel 4.1 Contoh Beberapa Dataset Ulasan Gojek.....	30
Tabel 4.2 Perhitungan Lexicon	50
Tabel 4.3 Jumlah Sentimen Lexicon InSet dan SentiStrengthID	51
Tabel 4.4 Hasil TF-IDF Datasheet Ulasan.....	52
Tabel 4.5 Rasio Pembagian Data	53
Tabel 4.6 Rangkuman Evaluasi MNB (InSet)	59
Tabel 4.7 Rangkuman Evaluasi MNB (SentiStrengthID).....	63
Tabel 5.1 Hasil Rata-Rata Evaluasi	64



DAFTAR GAMBAR

Gambar 3.1 Diagram Alir Penelitian.....	32
Gambar 4.1 Hasil Proses <i>Cleansing</i>	46
Gambar 4.2 Hasil Proses <i>Case Folding</i>	46
Gambar 4.3 Hasil Proses <i>Tokenization</i>	47
Gambar 4.4 Hasil Proses <i>Stopwords Removal</i>	47
Gambar 4.5 Hasil Proses <i>Stemming</i>	48
Gambar 4.6 Hasil Proses <i>Normalization</i>	48
Gambar 4.7 Visualisasi Pelabelan Lexicon InSet	51
Gambar 4.8 Visualisasi Pelabelan Lexicon SentiStrengthID.....	52
Gambar 4.9 Visualisasi <i>Confusion Matrix</i> (Lexicon InSet 80:20)	56
Gambar 4.10 Hasil Evaluasi MNB (Inset 80:20).....	57
Gambar 4.11 Visualisasi <i>Confusion Matrix</i> (InSet 70:30).....	57
Gambar 4.12 Hasil Evaluasi MNB (InSet 70:30)	58
Gambar 4.13 Visualisasi <i>Confusion Matrix</i> (InSet 60:40).....	58
Gambar 4.14 Hasil Evaluasi MNB (InSet 60:40).....	59
Gambar 4.15 Visualisasi <i>Confusion Matrix</i> (SentiStrengthID 80:20)	60
Gambar 4.16 Hasil Evaluasi MNB (SentiStrengthID 80:20).....	60
Gambar 4.17 Visualisasi <i>Confusion Matrix</i> (SentiStrengthID 70:30)	61
Gambar 4.18 Hasil Evaluasi MNB (SentiStrengthID 70:30).....	61
Gambar 4.19 Visualisasi <i>Confusion Matrix</i> (SentiStrengthID 60:40)	62
Gambar 4.20 Hasil Evaluasi MNB (SentiStrengthID 60:40).....	62

DAFTAR LAMPIRAN

Lampiran 1.1 Laporan Bimbingan	68
Lampiran 1.2 Submitted Dan Accepted Jurnal.....	73
Lampiran 1.3 CV	74
Lampiran 1.4 BNSP.....	77

