

**OTOMATISASI PENDETEKSI KATA BAKU DAN TIDAK
BAKU PADA DATA TWITTER BERBASIS KBBI
MENGUNAKAN METODE *LEXICON-BASED***



Skripsi

Untuk memenuhi syarat memperoleh Derajat
Sarjana Komputer (S.T.)

Oleh:

M Irfan Raif

2001020056

**JURUSAN TEKNIK INFORMATIKA
FAKULTAS TEKNIK DAN TEKNOLOGI KEMARITIMAN
UNIVERSITAS MARITIM RAJA ALI HAJI
TANJUNGPINANG
2024**

**OTOMATISASI PENDETEKSI KATA BAKU DAN TIDAK
BAKU PADA DATA TWITTER BERBASIS KBBI
MENGUNAKAN METODE *LEXICON-BASED***



Pembimbing I

Tekad Matulatan, S.Sos., S.Kom., M.Inf.Tech
NIP. 197308282021211006

Pembimbing II

Nuraisa Novia Hidayati S.Kom., M.Kom
NIP. 1989112320140220001

HALAMAN PENGESAHAN

Judul : Otomatisasi Pendeteksi Kata Baku dan Tidak Baku pada Data Twitter Berbasis KBBI Menggunakan Metode *Lexicon-Based*
Nama Mahasiswa : M Irfan Raif
NIM : 2001020056
Jurusan : Teknik Informatika

telah dipertahankan di depan Dewan Penguji dan dinyatakan lulus pada tanggal 15 Juli 2024

Susunan Tim Pembimbing dan Penguji

Jabatan	Nama Dosen	Tanda Tangan	Tanggal
Pembimbing I	: Tekad Matulatan, S.Sos., S.Kom., M.Inf.Tech. 24 Juli 2024
Pembimbing II	: Nuraisa Novia Hidayati S.Kom., M.Kom 24/7/2024
Ketua Penguji	: Nurfalinda, S.T., M.Cs 29/09-2024
Anggota Penguji 1	: Muhamad Radzi Rathomi, S.Kom., M.Cs 24/7/24
Anggota Penguji 2	: Nerfita Nikentari, S.T., M.Cs 24/7/2024

Tanjungpinang, 29 Juli 2024
Universitas Maritim Raja Ali Haji
Dekan Fakultas Teknik dan Teknologi Kemaritiman



Ir. Sapta Nugraha, S.T., M.Eng
NIP. 198904132015041005

PERNYATAAN ORISINALITAS

Dengan ini saya menyatakan bahwa skripsi saya yang berjudul Otomatisasi Pendeteksi Kata Baku dan Tidak Baku pada Data Twitter Berbasis KBBI Menggunakan Metode *Lexicon-Based* adalah benar karya saya dengan arahan dari komisi pembimbing dan belum diajukan dalam bentuk apa pun kepada perguruan tinggi mana pun. Sumber informasi yang berasal atau dikutip dari karya yang diterbitkan maupun tidak diterbitkan dari penulis lain telah disebutkan dalam teks dan dicantumkan dalam Daftar Pustaka di bagian akhir skripsi ini.

Jika kemudian hari ternyata terbukti pernyataan saya ini tidak benar dan melanggar peraturan yang sah dalam karya tulis dan hak intelektual maka saya bersedia ijazah yang telah saya terima untuk ditarik kembali oleh Universitas Maritim Raja Ali Haji.

Tanjungpinang, 28 Juni 2024

Yang menyatakan



(M Irfan Raif)

HALAMAN PERSEMBAHAN

Dengan penuh rasa syukur dan terima kasih, karya ini saya persembahkan kepada:

1. Allah SWT, Tuhan Yang Maha Esa, atas segala rahmat, hidayah, dan kekuatan yang diberikan sehingga saya dapat menyelesaikan skripsi ini. Semoga segala usaha dan doa yang dipanjatkan membawa berkah dan manfaat.
2. Orang tua saya tercinta, bapak M Arifin dan Ibu Rasidah, yang selalu menjadi sumber inspirasi dan motivasi saya. Terima kasih atas doa, cinta, dukungan, serta segala pengorbanan yang telah diberikan. Tanpa kalian, perjalanan ini tidak akan mungkin tercapai.
3. Dosen pembimbing, yang telah dengan sabar membimbing dan memberikan ilmu serta arahan yang sangat berharga selama proses penulisan skripsi ini.
4. Sahabat-sahabat terbaik, yang selalu ada di saat suka dan duka. Terima kasih atas kebersamaan, tawa, dan dukungan yang tak ternilai selama masa kuliah. Kalian adalah keluarga kedua bagi saya.
5. Rekan-rekan satu jurusan, yang telah menjadi bagian dari perjalanan akademik ini. Terima kasih atas kerja sama, bantuan, dan semangat yang telah kita bagi bersama.
6. Riset - Pemrosesan Bahasa Alami – Serpong, atas dukungan dan fasilitas yang diberikan selama penelitian ini. Bantuan yang diberikan sangat berarti dalam menyelesaikan skripsi ini, terutama dalam aspek pengumpulan data dan analisis. Penulis sangat menghargai kesempatan untuk bekerja dengan tim yang berdedikasi dan berkompeten di bidang pemrosesan bahasa alami, yang telah memberikan banyak wawasan berharga dan membantu penulis dalam mengembangkan penelitian ini

Semoga skripsi ini dapat memberikan kontribusi yang berarti dan menjadi langkah awal untuk mencapai kesuksesan di masa depan. Harapan saya, semoga hasil karya ini bermanfaat bagi pengembangan ilmu pengetahuan dan dapat memberikan inspirasi bagi pembaca.

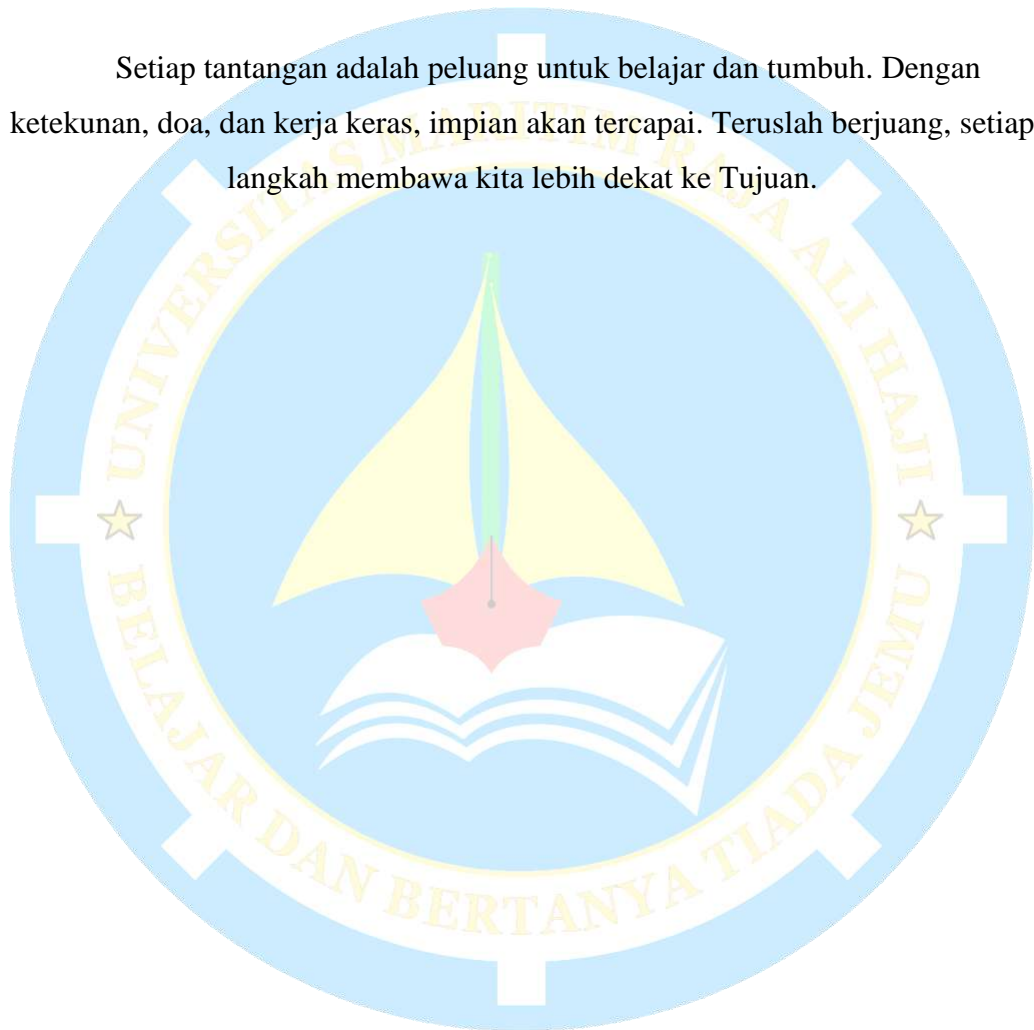
HALAMAN MOTO

“Success is not final, failure is not fatal, It is the courage to continue that counts”

-Winston Churchill

Bermimpilah setinggi langit, dan berusaha sekeras bumi.

Setiap tantangan adalah peluang untuk belajar dan tumbuh. Dengan ketekunan, doa, dan kerja keras, impian akan tercapai. Teruslah berjuang, setiap langkah membawa kita lebih dekat ke Tujuan.



KATA PENGANTAR

Puji dan syukur penulis panjatkan ke hadirat Allah SWT atas segala rahmat dan karunia-Nya sehingga penulis dapat menyelesaikan skripsi ini dengan judul "Otomatisasi Pendeteksi Kata Baku dan Tidak Baku pada Data Twitter Berbasis KBBI Menggunakan Metode *Lexicon-Based*". Skripsi ini disusun sebagai salah satu syarat untuk memperoleh gelar Sarjana di Fakultas Teknik dan Teknologi Kemaritiman, Universitas Maritim Raja Ali Haji.

Tujuan utama dari penelitian ini adalah membangun otomatisasi untuk mendeteksi kata baku dan tidak baku pada data Twitter berbasis KBBI guna meningkatkan keakuratan dan konsistensi dalam analisis teks di lingkungan media sosial, menilai dampak pendeteksian kata tidak baku terhadap peningkatan akurasi sentimen analisis, menggunakan metode *InSet* sebagai alat otomatis untuk sentimen analisis, serta menganalisis dan mendokumentasikan bagaimana pendeteksian kata tidak baku dapat membantu peneliti dalam mengidentifikasi kata-kata tidak baku di platform Twitter dan X. Selama pelaksanaan penelitian dan penulisan laporan ini, penulis menghadapi berbagai kesulitan non ilmiah, namun berkat dukungan dan bantuan dari berbagai pihak, penulis berhasil mengatasi kesulitan tersebut.

Penulis menyadari bahwa skripsi ini masih jauh dari sempurna, oleh karena itu, penulis mengharapkan kritik dan saran yang membangun dari semua pihak untuk perbaikan di masa yang akan datang. Semoga penelitian ini dapat bermanfaat bagi semua pihak yang berkepentingan. Akhir kata, penulis berharap agar skripsi ini dapat menjadi referensi yang berguna dan memberikan kontribusi positif dalam bidang *machine learning*.

Tanjungpinang, 28 Juni 2024



(M Irfan Raif)

DAFTAR ISI

HALAMAN JUDUL	i
HALAMAN PERSETUJUAN	ii
HALAMAN PENGESAHAN	iii
PERNYATAAN ORISINALITAS	iv
HALAMAN PERSEMBAHAN	v
HALAMAN MOTO	vi
KATA PENGANTAR	vii
DAFTAR ISI	viii
DAFTAR TABEL	ix
DAFTAR GAMBAR	x
GLOSARIUM	xi
ABSTRAK	xii
ABSTRACT	xiii
BAB I PENDAHULUAN	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	4
1.6 Sistematika Penelitian	5
BAB II LANDASAN TEORI	7
2.1 Tinjauan Pustaka	7
2.2 Landasan Teori	11
BAB III METODE PENELITIAN	18
3.1 Waktu dan Tempat Penelitian	18
3.2 Jenis Penelitian	18
3.3 Alur Penelitian	18
3.4 Studi Pustaka	19
3.5 Alat Penelitian	19
3.6 Pengumpulan Data	21
3.7 <i>Preprocessing</i> Data	21
3.8 Penentuan Polaritas	25
3.9 Analisis dan Perancangan	26
BAB IV HASIL DAN PEMBAHASAN	48
4.1 Komposisi Data	48
4.2 Hasil <i>Preprocessing</i>	48
4.3 Hasil Klasifikasi	74
4.4 Hasil Evaluasi	89
4.5 Implementasi Sistem	91
BAB V PENUTUP	93
5.1 Kesimpulan	93
5.2 Saran	94
DAFTAR PUSTAKA	95
LAMPIRAN	99

DAFTAR TABEL

Tabel 2. 1 Perbandingan Metode.....	15
Tabel 2. 2 <i>Confusion Matrix</i> Menggunakan 3x3.....	16
Tabel 3. 1 Proses Data <i>Cleanning</i>	27
Tabel 3. 2 Kata Tidak Baku.....	32
Tabel 3. 3 Proses Normalisasi	33
Tabel 3. 4 Penentuan Polaritas	34
Tabel 3. 5 Contoh Analisis Data <i>Confusion Matrix</i>	36
Tabel 4. 1 Proses Data <i>Cleaning</i> dari Twitter	50
Tabel 4. 2 Proses Data <i>Cleaning</i> dari X	54
Tabel 4. 3 Kata Baku dan Tidak Baku dari Data Twitter.....	60
Tabel 4. 4 Kata Baku dan Tidak Baku dari Data X.....	61
Tabel 4. 5 Kata yang Sering Muncul dari Data Twitter	64
Tabel 4. 6 Kata yang Sering Muncul dari Data X.....	64
Tabel 4. 7 Hasil Penghapusan Kata yang Berimbuhan dari Data Twitter.....	66
Tabel 4. 8 Hasil Penghapusan Kata yang Berimbuhan dari Data X.....	67
Tabel 4. 9 Kata yang Harusnya Dikecualikan dari Data Twitter	69
Tabel 4. 10 Kata yang Harusnya Dikecualikan dari Data X	70
Tabel 4. 11 Proses Normalisasi Data Twitter.....	71
Tabel 4. 12 Proses Normalisasi Data X.....	72
Tabel 4. 13 Penentuan Polaritas dari Data X Menggunakan <i>Stopword</i>	74
Tabel 4. 14 Penentuan Polaritas dari Data X Tidak Menggunakan <i>Stopword</i>	75
Tabel 4. 15 Hasil <i>Polarity</i> dari <i>InSet</i> Menggunakan Data X.....	78
Tabel 4. 16 Hasil Penentuan <i>Polarity</i> dari <i>InSet</i>	79
Tabel 4. 17 Hasil <i>Polarity</i> dari <i>InSet</i> Menggunakan Data Twitter.....	81
Tabel 4. 18 Penentuan Polaritas Data Twitter Menggunakan <i>Stopword</i>	83
Tabel 4. 19 Penentuan Polaritas Data Twitter Tidak Menggunakan <i>Stopword</i> ...	84
Tabel 4. 20 Hasil Evaluasi <i>Confusion Matrix</i>	88
Tabel 4. 21 Hasil Evaluasi.....	89

DAFTAR GAMBAR

Gambar 3. 1 Metodologi Penelitian.....	19
Gambar 3. 2 <i>Use Case</i> Diagram Pendeteksi Kata Tidak Baku	39
Gambar 3. 3 <i>Use Case</i> Diagram Analisis Sentimen <i>InSet</i>	40
Gambar 3. 4 <i>Activity</i> Diagram Menu Pendeteksi Kata Tidak Baku	41
Gambar 3. 5 <i>Activity</i> Diagram Menu Analisis Sentimen <i>InSet</i>	42
Gambar 3. 6 <i>Sequence</i> Diagram Menu Pendeteksi Kata Tidak Baku	44
Gambar 3. 7 <i>Sequence</i> Diagram Menu Analisis Sentimen <i>InSet</i>	45
Gambar 3. 8 Perancangan Aplikasi Pendeteksi Kata Tidak Baku.....	46
Gambar 3. 9 Perancangan Aplikasi Analisis Sentimen	47
Gambar 4. 1 Label 12.520 Data Twitter	48
Gambar 4. 2 Hasil Sentimen <i>Polarity</i> dari Data X.....	78
Gambar 4. 3 Hasil Sentimen <i>Polarity</i> dari Data Twitter	82
Gambar 4. 4 Implementasi Sistem Pendeteksi Kata Tidak Baku	91
Gambar 4. 5 Implementasi Sistem Analisis Sentimen <i>InSet</i>	92