



**ANALISIS SENTIMEN TWITTER UNTUK MENENTUKAN  
KEPRIBADIAN SESEORANG BERDASARKAN MODEL *THE  
BIG FIVE PERSONALITY TRAITS* DENGAN  
MENGUNAKAN ALGORITMA *NAÏVE BAYES***

**LAPORAN SKRIPSI**

Oleh

**KHAIRUL SHOLECH**

**41519120068**

**PROGRAM STUDI TEKNIK INFORMATIKA  
FAKULTAS ILMU KOMPUTER  
UNIVERSITAS MERCU BUANA  
JAKARTA**

**2023**



**ANALISIS SENTIMEN TWITTER UNTUK MENENTUKAN  
KEPRIBADIAN SESEORANG BERDASARKAN MODEL *THE  
BIG FIVE PERSONALITY TRAITS* DENGAN  
MENGUNAKAN ALGORITMA *NAÏVE BAYES***

**LAPORAN SKRIPSI**

Oleh  
**KHAIRUL SHOLECH**  
41519120068

**Diajukan sebagai salah satu syarat untuk memperoleh gelar sarjana**

**PROGRAM STUDI TEKNIK INFORMATIKA  
FAKULTAS ILMU KOMPUTER  
UNIVERSITAS MERCU BUANA  
JAKARTA  
2023**

## HALAMAN PERNYATAAN KARYA SENDIRI

Saya yang bertanda tangan di bawah ini:

Nama : Khairul Sholech  
NIM : 41519120068  
Program Studi : Teknik Informatika  
Judul Laporan Skripsi : Analisis Sentimen Twitter untuk Menentukan Kepribadian Seseorang Berdasarkan Model *The Big Five Personality Traits* dengan Menggunakan Algoritma *Naïve Bayes*

Menyatakan bahwa Laporan Skripsi 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 Skripsi saya terdapat unsur plagiat, maka saya siap mendapatkan sanksi akademis yang berlaku di Universitas Mercu Buana.

Jakarta, 27 Desember 2023  
UNIVERSITAS  
MERCU BUANA



Khairul Sholech

## HALAMAN PENGESAHAN

Laporan Skripsi ini diajukan oleh:

Nama : Khairul Sholech  
NIM : 41519120068  
Program Studi : Teknik Informatika  
Judul Laporan Skripsi : Analisis Sentimen Twitter untuk Menentukan Kepribadian Seseorang Berdasarkan Model *The Big Five Personality Traits* dengan Menggunakan Algoritma *Naïve Bayes*

Telah berhasil dipertahankan pada sidang di hadapan Dewan Penguji dan diterima sebagai persyaratan yang diperlukan untuk memperoleh gelar Sarjana Sastra I pada Program Studi Teknik Informatika, Fakultas Ilmu Komputer, Universitas Mercu Buana.

Disahkan oleh:

Pembimbing : Saruni Dwiasnati, ST,MM., M.Kom (  )  
NIDN : 0325128802  
Ketua Penguji : Lukman Hakim, ST., M.Kom (  )  
NIDN : 0327107701  
Penguji 1 : Dr. Harwikarya, MT (  )  
NIDN : 0014075805  
Penguji 2 : Dr. Rahmat Budiarto, M.Eng (  )  
NIDN : 0316106106

Jakarta, 27 Desember 2023

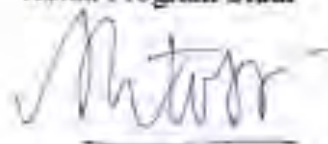
Mengetahui,

Dekan



Dr. Bambang Jekonowo, S.Si., M.T.I

Ketua Program Studi



Dr. Hadi Santoso, S.Kom., M.Kom

## KATA PENGANTAR

بِسْمِ اللَّهِ الرَّحْمَنِ الرَّحِيمِ

*Assalammu'alaikum Warrahmatullahi Wabarakatuh.*

Alhamdulillah rabbil'alamin Puji syukur penulis panjatkan kehadirat Allah SWT yang telah memberikan kesehatan, rahmat serta hidayah-Nya kepada penulis. Sholawat serta salam semoga senantiasa tercurahkan kepada junjungan kita Nabi Muhammad SAW.

Laporan Tugas Akhir ini dengan judul **“Analisis Sentimen Twitter untuk Menentukan Kepribadian Seseorang Berdasarkan Model *The Big Five Personality Traits* dengan Menggunakan Algoritma *Naïve Bayes*”**, dimaksudkan dalam rangka memenuhi salah satu persyaratan kelulusan Program Studi Strata Satu (S1) pada jurusan Teknik Informatika, Universitas Mercu Buana.

Dalam penulisan Laporan Tugas Akhir ini, penulis banyak menghadapi kesulitan-kesulitan karena keterbatasan penulis dalam penguasaan materi baik secara teori maupun praktek serta juga keterbatasan waktu. Namun, berkat rahmat Allah SWT, doa orang tua, bimbingan dan pengarahan dari dosen pembimbing, serta bantuan semua pihak, akhirnya penulis dapat menyelesaikan penulisan Laporan Tugas Akhir ini. Maka dengan segala kerendahan hati penulis menyampaikan rasa terimakasih yang setulus-tulusnya kepada :

1. Bapak Prof. Dr. Ir. Andi Adriansyah, M.Eng selaku Rektor Universitas Mercu Buana.
2. Bapak Dr. Bambang Jokonowo, S.Si., M.T.I selaku Dekan Fakultas Ilmu Komputer.
3. Bapak Dr. Hadi Santoso, S.Kom., M.Kom selaku Ketua Program Studi Teknik Informatika.
4. Ibu Saruni Dwiasnati, ST.MM., M.Kom selaku Dosen Pembimbing Tugas Akhir yang telah menyediakan waktu, tenaga, dan pikiran untuk mengarahkan penulis dalam penyusunan laporan tugas akhir ini.

5. Ibu Dr. Afiyati., S.Si., MT selaku Dosen Pembimbing Akademik yang telah memberikan saran, dorongan serta motivasi kepada penulis sehingga dapat menyelesaikan laporan tugas akhir ini dengan baik.
6. Bapak Lukman Hakim, ST., M.Kom selaku Dosen Ketua Penguji Tugas Akhir, Bapak Dr. Harwikarya, MT dan Bapak Dr. Rahmat Budiarto, M.Eng selaku Dosen Penguji 1 dan 2 yang telah meluangkan waktunya untuk memberikan koreksi dan arahan serta masuknya dalam penulisan dan menguji laporan tugas akhir dari penulis.
7. Dosen-dosen pengajar Program S1 serta seluruh staff Tata Usaha Teknik Informatika Fakultas Ilmu Komputer Universitas Mercu Buana yang telah membantu selama masa perkuliahan sampai penyusunan laporan tugas akhir.
8. Rasa terimakasih yang terdalam juga penulis ucapkan secara khusus kepada kedua orang tua penulis, Bapak Sukarno dan Ibu Sumiati. Terimakasih atas segala kasih sayang yang diberikan dalam mendidik dan membimbing penulis selama ini sehingga penulis dapat terus berjuang dalam meraih mimpi dan cita-citanya.
9. Eva wulandari dan Evi Sri Gumilang selaku kakak-kakak yang penulis cintai dan hormati yang telah banyak memberikan doa dan dukungan baik moril maupun materil serta saran maupun motivasi untuk melewati suka duka kehidupan penulis selama ini.
10. Direktur PT. Indojoya Sukses Makmur, Bapak Surian Sutrisno S.H yang telah memberikan beasiswa pendidikan kepada penulis, sehingga penulis bisa berkuliah di Universitas Mercu Buana.
11. Bapak Edi Gunawan S. Kom M. M dan Bapak Ir. Sugeng Subroto selaku pimpinan di PT. Indojoya Sukses Makmur yang telah memberikan dukungan, arahan, motivasi dan bantuan lainnya selama masa perkuliahan.
12. Tatiek Hurun Raehan Maemanah selaku *support system* bagi penulis yang selalu memberikan saran dan dukungan penuh kepada penulis demi terselesainya laporan tugas akhir ini.

13. Terimakasih kepada semua pihak yang terlibat yang tidak bisa disebutkan satu per-satu oleh penulis yang telah banyak membantu dalam penyusunan laporan tugas akhir ini.

Akhir kata, penulis berharap Allah SWT berkenan membalas segala kebaikan semua pihak yang telah membantu. Penulis menyadari bahwa laporan ini jauh dari kata sempurna, namun penulis berharap agar Laporan Tugas Akhir ini membawa manfaat bagi pengembangan ilmu. Penulis juga mengharapkan adanya masukan berupa kritik atau saran yang bersifat membangun untuk perbaikan Laporan Tugas Akhir ini menjadi lebih baik lagi.

*Wassalammu'alaikum Warohmatullahi Wabarokatuh.*

Jakarta, 27 Desember 2023

Khairul Sholeh





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

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

Nama : Khairul Sholech  
NIM : 41519120068  
Program Studi : Teknik Informatika  
Judul Laporan Skripsi : Analisis Sentimen Twitter untuk Menentukan Kepribadian Seseorang Berdasarkan Model *The Big Five Personality Traits* dengan Menggunakan Algoritma *Naïve Bayes*

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 diatas 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, 27 Desember 2023

Yang menyatakan,



Khairul Sholech



## ABSTRAK

Nama : Khairul Sholeh  
NIM : 41519120068  
Program Studi : Teknik Informatika  
Judul Laporan Skripsi : Analisis Sentimen Twitter untuk Menentukan Kepribadian Seseorang Berdasarkan Model *The Big Five Personality Traits* dengan Menggunakan Algoritma *Naïve Bayes*  
Dosen Pembimbing : Saruni Dwiasnati, ST.MM., M.Kom

Penelitian ini mengimplementasikan algoritma *Naïve Bayes* untuk menganalisis kepribadian seseorang berdasarkan data Twitter sesuai dengan *The Big Five Personality Traits* dengan pengguna Twitter yang berpengaruh di Indonesia dengan masing-masing 500 tweet per-pengguna. Melalui proses Crawling data menghasilkan 50.024 tweets dan setelah melalui Pre-processing data menjadi 46.028 tweet. Langkah selanjutnya dengan menerapkan teknik pembobotan Term Frequency-Inverse Document Frequency (TF-IDF) dan melakukan analisis NRC-Lexicon, penelitian ini berhasil menggali informasi sentimen dan emosi yang terkandung dalam setiap tweet. Kombinasi kedua metode ini menjadi dasar penentuan label pada dimensi kepribadian yang mengacu pada model OCEAN (Openness, Conscientiousness, Extraversion, Agreeableness, Neuroticism). Algoritma *Naïve Bayes* digunakan untuk klasifikasi, mencapai akurasi tertinggi 75.56% pada dataset uji dengan perbandingan 90:10. Evaluasi model menunjukkan peningkatan performa seiring dengan pertumbuhan ukuran dataset uji, menekankan pentingnya dataset yang lebih besar untuk meningkatkan ketepatan klasifikasi. Penelitian ini memberikan wawasan mendalam untuk riset lebih lanjut dan aplikasi praktis dalam memahami dinamika kepribadian yang diperoleh dari media sosial.

**Kata Kunci** : Algoritma *Naïve Bayes*, *The Big Five Personality Traits*, NRC-Lexicon, Analisis Sentimen, TF-IDF.

## ABSTRACT

Name : Khairul Sholech  
NIM : 41519120068  
Study Program : Informatics Engineering  
Title Thesis : Twitter Sentiment Analysis to Determine Personality Based on The Big Five Personality Traits Model Using Naïve Bayes algorithm  
Counsellor : Saruni Dwiasnati, ST.MM., M.Kom

This research implements the Naïve Bayes algorithm to analyze a person's personality based on Twitter data according to The Big Five Personality Traits with influential Twitter users in Indonesia with a total data of 500 tweets per user. Crawling data produces 50.024 tweets and after data becomes 46.028 tweets. The next step by applying the Term Frequency-Inverse Document Frequency (TF-IDF) weighting technique and conducting NRC-Lexicon analysis, this research succeeded in extracting sentiment and emotion information contained in each tweet. The combination of the two methods is the basis for determining the label on the personality dimension that refers to the OCEAN (Openness, Conscientiousness, Extraversion, Agreeableness, Neuroticism). The Naïve Bayes algorithm was used for classification, achieving the highest accuracy of 75.56% on a test dataset with a 90:10 ratio. Model evaluation showed improved performance as the size of the test dataset increased, emphasizing the importance of larger dataset to improve classification accuracy. This research provides insights for further research and practical applications in understanding personality dynamics derived from social media.

**Keywords** : Naïve Bayes algorithm, The Big Five Personality Traits, NRC-Lexicon, Sentiment Analysis, TF-IDF.

## DAFTAR ISI

<b>HALAMAN JUDUL .....</b>	<b>i</b>
<b>HALAMAN PERNYATAAN KARYA SENDIRI .....</b>	<b>ii</b>
<b>HALAMAN PENGESAHAN .....</b>	<b>iii</b>
<b>KATA PENGANTAR.....</b>	<b>iv</b>
<b>HALAMAN PERNYATAAN PERSETUJUAN PUBLIKASI TUGAS AKHIR UNTUK KEPENTINGAN AKADEMIS .....</b>	<b>vii</b>
<b>ABSTRAK .....</b>	<b>viii</b>
<b>ABSTRACT .....</b>	<b>ix</b>
<b>DAFTAR ISI.....</b>	<b>x</b>
<b>DAFTAR TABEL .....</b>	<b>xii</b>
<b>DAFTAR GAMBAR.....</b>	<b>xiii</b>
<b>DAFTAR LAMPIRAN .....</b>	<b>xiv</b>
<b>BAB I PENDAHULUAN .....</b>	<b>1</b>
1.1 Latar Belakang .....	1
1.2 Rumusan Masalah .....	3
1.3 Batasan Masalah.....	3
1.4 Tujuan dan Manfaat Penelitian.....	4
1.5 Sistematika Penulisan.....	4
<b>BAB II TINJAUAN PUSTAKA.....</b>	<b>6</b>
2.1 Teori Penelitian Terkait.....	6
2.1.1 <i>Natural Language Processing</i> .....	6
2.1.2 Analisis Sentimen .....	7
2.1.3 Model <i>Big Five Personality Traits</i> .....	8
2.1.4 Python .....	10
2.1.5 Algoritma <i>Naïve Bayes</i> .....	10
2.2 Penelitian Terkait .....	11
2.3 <i>Critical Review</i> .....	17
<b>BAB III METODE PENELITIAN .....</b>	<b>29</b>
3.1 Jenis Penelitian .....	29
3.2 Sarana Pendukung .....	30
3.2.1 Perangkat Keras ( <i>Hardware</i> ).....	30
3.2.2 Perangkat Lunak ( <i>Software</i> ) .....	30

3.3	Teknik Pengumpulan Data .....	31
3.4	Diagram Alir Penelitian.....	32
3.4.1	<i>Crawling Data</i> .....	33
3.4.2	<i>Preprocessing</i> .....	33
3.4.3	Pembobotan TF-IDF .....	34
3.4.4	NRC-Lexicon.....	35
3.4.5	Label <i>The Big Five Personality Traits</i> .....	35
3.4.6	Pembagian Data .....	36
3.4.7	Klasifikasi <i>Naïve Bayes</i> .....	36
3.4.8	Model Prediksi.....	36
3.4.9	Hasil Evaluasi .....	36
<b>BAB IV</b>	<b>HASIL DAN PEMBAHASAN.....</b>	<b>37</b>
4.1	Dataset .....	37
4.2	Pre-processing .....	37
4.3	<i>Feature Extraction</i> .....	39
4.4	Pembuatan Model.....	42
4.5	Visualisasi Data .....	42
4.6	Pengujian .....	43
4.7	Analisis Hasil .....	44
4.8	Perhitungan Manual .....	45
4.8.1	Perhitungan Probabilitas Kelas Pribadi ( <i>Prior</i> ).....	45
4.8.2	Perhitungan Probabilitas Kelas Fitur Diberikan Kelas ( <i>Likelihood</i> ) .	48
4.8.3	Perhitungan Probabilitas <i>Posterior</i> .....	56
4.8.4	Prediksi .....	57
<b>BAB V</b>	<b>KESIMPULAN DAN SARAN .....</b>	<b>58</b>
5.1	Kesimpulan.....	58
5.2	Saran.....	58
	<b>DAFTAR PUSTAKA .....</b>	<b>60</b>
	<b>LAMPIRAN.....</b>	<b>63</b>

## DAFTAR TABEL

tabel 2. 1 Penelitian Terkait .....	11
tabel 2. 2 <i>Critical Review</i> .....	17



## DAFTAR GAMBAR

gambar 3. 1 Diagram Alir Penelitian .....	32
gambar 4. 1 Crawling Dataset.....	37
gambar 4. 2 Proses dari Program Pre-processing Data .....	38
gambar 4. 3 Visualisasi Jumlah Tweets Setelah Pre-processing Data .....	39
gambar 4. 4 Hasil Pre-processing Data Per-pengguna.....	39
gambar 4. 5 Hasil Word Cloud dari Salah Satu Pengguna .....	40
gambar 4. 6 Hasil dari NRC-Lexicon .....	41
gambar 4. 7 Final Dataset .....	42
gambar 4. 8 Visualisasi Hasil Akurasi dan Ukuran Data Uji.....	43
gambar 4. 9 <i>Confusion Matrix</i> Ukuran Dataset Uji 0.9 .....	43
gambar 4. 10 <i>Classification Report</i> Ukuran Dataset Uji 0.9 .....	44
gambar 4. 11 Tabel Performa Model dengan Berbagai Ukuran Uji .....	44
gambar 4. 12 Tabel <i>Instance</i> Kelas <i>Openness</i> .....	46
gambar 4. 13 <i>Tabel Instance</i> Kelas <i>Conscientiosness</i> .....	46
gambar 4. 14 <i>Tabel Instance</i> Kelas <i>Extraversion</i> .....	47
gambar 4. 15 <i>Tabel Instance</i> Kelas <i>Agreeableness</i> .....	47
gambar 4. 16 <i>Tabel Instance</i> Kelas <i>Neuroticism</i> .....	47
gambar 4. 17 Hitungan <i>Mean</i> untuk Semua Kelas.....	50
gambar 4. 18 Hitungan <i>Variance</i> untuk Semua Kelas.....	53
gambar 4. 19 Hitungan <i>Gaussian Distribution</i> untuk Semua Kelas .....	56
gambar 4. 20 Contoh Data Uji untuk Diprediksi .....	57



## DAFTAR LAMPIRAN

Lampiran 1 Asistensi Bimbingan.....	64
Lampiran 2 Halaman Persetujuan .....	65
Lampiran 3 Bukti Submit / Published Artikel Ilmiah / HKI.....	66
Lampiran 4 Naskah Artikel Jurnal .....	67
Lampiran 5 Curriculum Vitae .....	84
Lampiran 6 Plagiarisme Check .....	85
Lampiran 7 Surat Pernyataan Hak Cipta.....	87
Lampiran 8 Surat Pengalihan Hak Cipta.....	88
Lampiran 9 Sertifikat BNSP/Bukti Daftarnya .....	89



UNIVERSITAS  
MERCU BUANA