

**KLASIFIKASI KEPRIBADIAN BERDASARKAN METODE THE FOUR
TEMPERAMENTS PADA CUITAN TWITTER MENGGUNAKAN
ALGORITMA NAÏVE BAYES CLASSIFIER**

SKRIPSI



disusun oleh

Rizky Nur Hidayatullah

17.11.1247

**PROGRAM SARJANA
PROGRAM STUDI INFORMATIKA
FAKULTAS ILMU KOMPUTER
UNIVERSITAS AMIKOM YOGYAKARTA
YOGYAKARTA
2021**

**KLASIFIKASI KEPERIBADIAN BERDASARKAN METODE THE FOUR
TEMPERAMENTS PADA CUITAN TWITTER MENGGUNAKAN
ALGORITMA NAÏVE BAYES CLASSIFIER**

SKRIPSI

untuk memenuhi sebagian persyaratan
mencapai gelar Sarjana
pada Program Studi Informatika



disusun oleh

Rizky Nur Hidayatullah

17.11.1247

**PROGRAM SARJANA
PROGRAM STUDI INFORMATIKA
FAKULTAS ILMU KOMPUTER
UNIVERSITAS AMIKOM YOGYAKARTA
YOGYAKARTA
2021**

PERSETUJUAN

SKRIPSI

**KLASIFIKASI KEPERIBADIAN BERDASARKAN METODE THE FOUR
TEMPERAMENTS PADA CUITAN TWITTER MENGGUNAKAN
ALGORITMA NAÏVE BAYES CLASSIFIER**

yang dipersiapkan dan disusun oleh

Rizky Nur Hidayatullah

17.11.1247

telah disetujui oleh Dosen Pembimbing Skripsi
pada tanggal 4 April 2020

Dosen Pembimbing,

Anggit Dwi Hartanto, M.Kom.
NIK. 190302163

PENGESAHAN

SKRIPSI

KLASIFIKASI KEPERIBADIAN BERDASARKAN METODE THE FOUR TEMPERAMENTS PADA CUITAN TWITTER MENGGUNAKAN ALGORITMA NAÏVE BAYES CLASSIFIER

yang dipersiapkan dan disusun oleh

Rizky Nur Hidayatullah

17.11.1247

telah dipertahankan di depan Dewan Penguji
pada tanggal 15 Januari 2021

Susunan Dewan Penguji

Nama Penguji

Tanda Tangan

Arif Dwi Laksito, M.Kom
NIK. 190302150

Anggit Dwi Hartanto, M.Kom
NIK. 190302163

Ika Nur Fajri, M.Kom
NIK. 190302268

Skripsi ini telah diterima sebagai salah satu persyaratan
untuk memperoleh gelar Sarjana Komputer
Tanggal 30 Januari 2021.

DEKAN FAKULTAS ILMU KOMPUTER

Krisnawati, S.Si, M.T.
NIK. 190302038

PERNYATAAN

Saya yang bertandatangan dibawah ini menyatakan bahwa, skripsi ini merupakan karya saya sendiri (ASLI), dan isi dalam skripsi ini tidak terdapat karya yang pernah diajukan oleh orang lain untuk memperoleh gelar akademis di suatu institusi pendidikan tinggi manapun, dan sepanjang pengetahuan saya juga tidak terdapat karya atau pendapat yang pernah ditulis dan/atau diterbitkan oleh orang lain, kecuali yang secara tertulis diacu dalam naskah ini dan disebutkan dalam daftar pustaka.

Segala sesuatu yang terkait dengan naskah dan karya yang telah dibuat adalah menjadi tanggungjawab saya pribadi.

Yogyakarta, 29 Januari 2021



Rizky Nur Hidayatullah

NIM. 17.11.1247

MOTTO

“Banyak orang gagal karena mereka tidak memahami usaha yang diperlukan untuk menjadi sukses”

“Karena sesungguhnya sesudah kesulitan itu ada kemudahan.
(QS. Al-Insyirah: 5)”

“Jika kau tidak berusaha semampumu, maka kau akan menyesal nantinya”



PERSEMBAHAN

Dengan penuh rasa syukur saya persembahkan skripsi ini kepada semua pihak yang terlibat secara langsung maupun tidak langsung dalam proses pembuatan skripsi.

- Allah SWT yang senantiasa memberikan nikmat dan kasih sayang kepada semua umat-Nya.
- Kedua orang tua dan keluarga saya yang selalu memberikan do'a, semangat, dan dukungan tak henti – hentinya kepada saya.
- Bapak Anggit Dwi Hartanto, M.Kom yang telah membimbing saya dari awal hingga akhir pembuatan skripsi ini.
- Dosen – dosen Universitas AMIKOM Yogyakarta yang telah memberikan banyak ilmu pengetahuan selama perkuliahan.
- Kakak – kakak Informatika Agung Krisna Wijaya, Andia Enggar Mayasari, Zauvik Rizaldi Ma'ruf yang memberikan saran terkait pembuatan skripsi.
- Teman – teman kelas 17 IF 05 yang selalu menemani perkuliahan, mendukung dan memberikan semangat sampai saat ini.
- Teman – teman LIDE Dev yang sudah membantu saya dalam membangun sistem dalam skripsi ini.
- Teman skripsi saya Rizki Khairunnisa yang sudah berpartisipasi dalam mengumpulkan data pengguna Twitter.
- Nanda Fauzia Fitriyani yang selalu menemani, memberikan semangat, dan berbagi keluh kesah kepada saya.

KATA PENGANTAR

Puji dan syukur kehadiran Allah SWT yang telah melimpahkan rahmat, taufik, dan hidayah-Nya kepada kita semua serta shalawat dan salam penulis curahkan kepada junjungan kita Nabi Muhammad SAW sehingga penulis dapat menyelesaikan skripsi yang berjudul “KLASIFIKASI KEPERIBADIAN BERDASARKAN METODE THE FOUR TEMPERAMENTS PADA CUITAN TWITTER MENGGUNAKAN ALGORITMA NAÏVE BAYES CLASSIFIER”.

Skripsi ini disusun guna menyelesaikan studi jenjang Strata Satu (S1) pada program studi Informatika fakultas Ilmu Komputer Universitas AMIKOM Yogyakarta. Selain itu juga sebagai bukti bahwa mahasiswa telah memenuhi salah satu syarat untuk memperoleh gelar Sarjana Komputer.

Dalam penyusunan skripsi ini, penulis menyadari masih banyak kekurangan dan jauh dari kata sempurna. Selain itu dengan terselesaikannya skripsi ini, penulis ingin menyampaikan rasa terima kasih kepada:

1. Allah SWT atas segala berkah dan ridho-Nya penulis dapat menyelesaikan skripsi ini dengan baik.
2. Nabi Muhammad SAW yang selalu menjadi panutan dan suri tauladan.
3. Kedua orang tua dan keluarga penulis sebagai sosok luar biasa yang selalu memberikan dukungannya dalam bentuk apapun.
4. Bapak Prof. Dr. M. Suyanto, M.M., selaku Rektor Universitas AMIKOM Yogyakarta.
5. Bapak Anggit Dwi Hartanto, M.Kom, selaku dosen pembimbing yang telah memberikan bimbingan, arahan, dan masukan kepada penulis.

6. Seluruh staff pengajar dan karyawan Universitas AMIKOM Yogyakarta yang telah memberikan ilmu yang bermanfaat.
7. Teman – teman kelas 17 IF 05 yang telah memberikan semangat dalam menyelesaikan skripsi ini.
8. Teman – teman LIDE Dev yang sudah membantu dalam membangun sistem pada skripsi ini.
9. Dia yang selalu menemani, mendo'akan, dan memberi semangat kepada penulis.
10. Semua pihak yang telah membantu penulis baik secara langsung maupun tidak langsung.

Semoga Allah SWT memberikan balasan lebih kepada semua yang telah ikut membantu penulis hingga terselesainya skripsi ini. Semoga skripsi ini dapat bermanfaat bagi penulis maupun pembaca.

Yogyakarta, 8 Januari 2021

Rizky Nur Hidayatullah

DAFTAR ISI

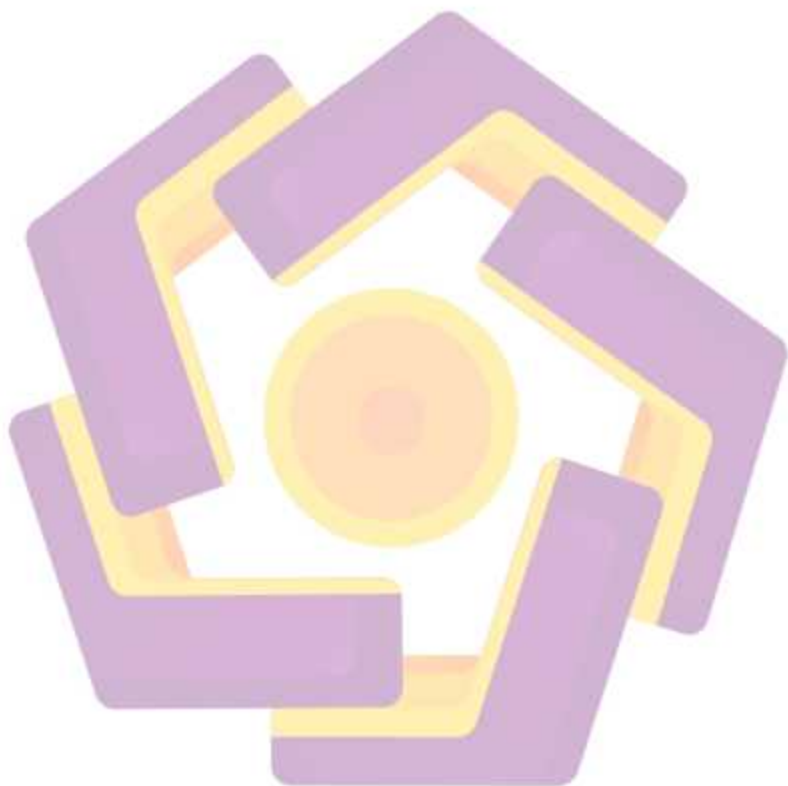
HALAMAN JUDUL.....	i
PERSETUJUAN.....	iii
PENGESAHAN.....	iv
PERNYATAAN.....	iv
MOTTO.....	v
PERSEMBAHAN.....	vi
KATA PENGANTAR.....	vii
DAFTAR ISI.....	ix
DAFTAR TABEL.....	xi
DAFTAR GAMBAR.....	xiii
INTISARI.....	xv
<i>ABSTRACT</i>	xvi
BAB I PENDAHULUAN	1
1.1 Latar Belakang.....	1
1.2 Rumusan Masalah.....	3
1.3 Batasan Masalah.....	4
1.4 Maksud Dan Tujuan Penelitian.....	4
1.5 Manfaat Penelitian.....	5
1.6 Metode Penelitian.....	5
1.7 Sistematika Penulisan.....	7
BAB II LANDASAN TEORI	9
2.1 Kajian Pustaka.....	9
2.2 The Four Temperaments.....	12
2.3 Naïve Bayes Classifier.....	15
2.4 Data Mining.....	17
2.5 Python.....	18
2.6 K-Fold Cross Validation.....	20
BAB III ANALISIS DAN PERANCANGAN	25
3.1 Analisis Kebutuhan.....	25

3.2	Gambaran Umum Sistem.....	27
3.3	Analisis Algoritma.....	29
3.4	Analisis Kebutuhan Data.....	57
3.5	Perancangan Antarmuka Pengguna.....	62
BAB IV IMPLEMENTASI		67
4.1	Deskripsi Implementasi.....	67
4.2	Pengumpulan Data.....	67
4.3	Preprocessing.....	74
4.4	Pelatihan.....	79
4.5	Pengujian.....	83
4.6	Antarmuka.....	85
BAB V PENUTUP.....		88
5.1	Kesimpulan.....	88
5.2	Saran.....	88
DAFTAR PUSTAKA		90
LAMPIRAN.....		92

DAFTAR TABEL

Tabel 2. 1 Perbandingan Penelitian Sebelumnya.....	10
Tabel 3. 1 <i>Tweet</i> hasil unduhan.....	33
Tabel 3. 2 <i>Tweet</i> sebelum proses <i>cleaning</i>	35
Tabel 3. 3 <i>Tweet</i> setelah proses <i>cleaning</i>	37
Tabel 3. 4 <i>Tweet</i> setelah proses tokenisasi.....	38
Tabel 3. 5 <i>Tweet</i> setelah proses <i>slang replacement</i>	39
Tabel 3. 6 <i>Tweet</i> setelah proses <i>stemming</i>	40
Tabel 3. 7 <i>Tweet</i> setelah proses <i>stopword removal</i>	41
Tabel 3. 8 <i>Tweet</i> setelah proses <i>POS Tagging</i>	42
Tabel 3. 9 <i>Tweet</i> setelah proses <i>POS Filtering</i>	44
Tabel 3. 10 <i>Tweet</i> setelah proses <i>Common Word Removal</i>	45
Tabel 3. 11 <i>Tweet</i> setelah proses penggabungan.....	46
Tabel 3. 12 <i>Term frequency</i> kepribadian Sanguinis.....	47
Tabel 3. 13 <i>Term frequency</i> kepribadian Koleris.....	48
Tabel 3. 14 <i>Term frequency</i> kepribadian Melankolis.....	48
Tabel 3. 15 <i>Term frequency</i> kepribadian Plegmatis.....	49
Tabel 3. 16 <i>Tweet</i> terangkaning.....	49
Tabel 3. 17 Daftar <i>Feature</i>	50
Tabel 3. 18 Tiga kata teratas.....	51
Tabel 3. 19 Proses pembuatan array biner.....	52
Tabel 3. 20 Daftar label/ <i>target</i>	53
Tabel 3. 21 Model <i>data training</i>	53
Tabel 3. 22 Penghitungan probabilitas $P(H)$	54
Tabel 3. 23 Penghitungan probabilitas $P(X H)$	55
Tabel 3. 24 Penghitungan probabilitas $P(X H)*P(H)$	55
Tabel 3. 25 Penghitungan probabilitas dengan <i>Laplace Correction</i>	56
Tabel 3. 26 Struktur tabel users.....	57
Tabel 3. 27 Struktur tabel personality.....	58
Tabel 3. 28 Struktur tabel filter_users.....	58

Tabel 3. 29 Struktur tabel tweets.....	59
Tabel 3. 30 Struktur tabel prediction_results.....	59
Tabel 3. 31 Struktur tabel prediction_type.....	59



DAFTAR GAMBAR

Gambar 3. 1 Alur sistem	28
Gambar 3. 2 Alur <i>preprocessing</i>	34
Gambar 3. 3 Daftar <i>feature</i> untuk dataset	61
Gambar 3. 4 Dataset untuk <i>training</i> dan <i>testing</i>	61
Gambar 3. 5 Rancangan halaman beranda	62
Gambar 3. 6 Rancangan halaman tentang kami	63
Gambar 3. 7 Rancangan halaman <i>login</i>	64
Gambar 3. 8 Rancangan halaman metode prediksi	64
Gambar 3. 9 Rancangan halaman kuesioner	65
Gambar 3. 10 Rancangan halaman hasil prediksi	66
Gambar 4. 1 Implementasi halaman beranda	68
Gambar 4. 2 Implementasi halaman tentang kami	68
Gambar 4. 3 Implementasi halaman login	69
Gambar 4. 4 Implementasi halaman kuesioner	69
Gambar 4. 5 Tabel <i>users</i>	70
Gambar 4. 6 Tabel <i>personality</i>	70
Gambar 4. 7 Tabel <i>personality</i> yang telah divalidasi	71
Gambar 4. 8 Script untuk <i>tweet scraping</i>	72
Gambar 4. 9 Tabel <i>tweets</i>	72
Gambar 4. 10 Tabel <i>filter_users</i>	73
Gambar 4. 11 Script untuk <i>tweet cleaning</i>	74
Gambar 4. 12 Script untuk <i>tokenization</i>	75
Gambar 4. 13 Kamus bahasa <i>slang</i>	75
Gambar 4. 14 Script untuk <i>slang replacement</i>	76
Gambar 4. 15 Script untuk <i>tweet stemming</i>	76
Gambar 4. 16 Script untuk <i>stopword removal</i>	76
Gambar 4. 17 Script untuk <i>POS Tagging</i>	77
Gambar 4. 18 Script untuk <i>POS Filtering</i>	77
Gambar 4. 19 Kamus <i>common word</i>	78

Gambar 4. 20 Script untuk mencari <i>common word</i>	78
Gambar 4. 21 Script untuk <i>common word removal</i>	78
Gambar 4. 22 Script untuk penggabungan <i>tweet</i>	79
Gambar 4. 23 Script untuk <i>term frequency</i>	79
Gambar 4. 24 Script untuk <i>term frequency</i> kepribadian	80
Gambar 4. 25 Script untuk konversi biner	81
Gambar 4. 26 Script untuk <i>data training</i>	82
Gambar 4. 27 Script untuk <i>data testing</i>	83
Gambar 4. 28 Pengujian menggunakan data uji (dataset tidak <i>balance</i>)	83
Gambar 4. 29 Pengujian menggunakan data uji (dataset <i>balance</i>)	83
Gambar 4. 30 Script untuk <i>K-Fold Cross Validation</i>	84
Gambar 4. 31 Pengujian <i>K-Fold Cross Validation</i> (dataset tidak <i>balance</i>).....	85
Gambar 4. 32 Pengujian <i>K-Fold Cross Validation</i> (dataset <i>balance</i>).....	85
Gambar 4. 33 Implementasi halaman pilih metode prediksi.....	86
Gambar 4. 34 Implementasi antarmuka hasil kepribadian.....	87



INTISARI

Informasi di era digital kini berkembang semakin pesat. Sosial media menjadi tempat penyampaian informasi yang mudah digunakan. Pada media sosial Twitter, informasi tersebut disebarakan melalui cuitan – cuitan yang dapat dilihat oleh banyak orang. Cuitan ini berupa teks, gambar, suara, dan video yang dapat dianalisis menggunakan algoritma Naïve Bayes Classifier (NBC).

Naïve Bayes Classifier adalah suatu algoritma yang dapat mengklasifikasikan data ke dalam kelompok – kelompok yang memiliki kesamaan tertentu. Algoritma ini digunakan untuk mengelompokkan pengguna Twitter ke dalam teori *The Four Temperaments* dimana kepribadian seseorang dibagi menjadi empat tipe, yaitu: sanguinis, koleris, melankolis, dan plegmatis.

Cuitan seseorang di Twitter diunduh kemudian diolah melalui tahap *preprocessing*, *term frequency*, dan dikonversi ke dalam bentuk yang dapat diklasifikasi oleh algoritma NBC. Hasil dari klasifikasi kepribadian ini nantinya dapat digunakan sebagai alat bantu dalam berkomunikasi dengan orang lain, mengenali diri sendiri lebih baik, mencari rekan yang tepat, dan dapat digunakan oleh perusahaan atau organisasi dalam perekrutan anggotanya.

Kata Kunci: Klasifikasi kepribadian, The Four Temperaments, Naïve Bayes Classifier

ABSTRACT

The information in the digital age is now growing more rapidly. Social media is an easy-to-use place for delivering information. In Twitter, this information is spread through tweets that can be seen by many people. These tweets are in the form of text, images, voice, and video that can be analyzed using the Naïve Bayes Classifier (NBC) algorithm.

Naïve Bayes Classifier is an algorithm that can classify data into groups that have certain similarities. This algorithm is used to classify Twitter users into The Four Temperaments theory in which a person's personality is divided into four types, that is: sanguine, choleric, melancholy, and phlegmatic.

Someone's tweets on Twitter are downloaded and then processed through preprocessing steps, term frequency, and converted into a form that can be classified by the NBC algorithm. The results of this personality classification can later be used as a tool in communicating with others, knowing yourself better, looking for the right partner, and can be used by companies or organizations in recruiting their members.

Keyword: *Personality classification, The Four Temperaments, Naïve Bayes Classifier*