

**KLASIFIKASI SENTIMEN ANALISIS TWITTER UNTUK
MENILAI OPINI PUBLIK TERHADAP PERTANDINGAN SEPAK
BOLA MENGGUNAKAN METODE NAÏVE BAYES**

SKRIPSI

Diajukan untuk memenuhi salah satu syarat mencapai derajat Sarjana
Program Studi Informatika



disusun oleh

HAFIIDH KHOIRU PRADANA

19.11.2645

Kepada

**FAKULTAS ILMU KOMPUTER
UNIVERSITAS AMIKOM YOGYAKARTA
YOGYAKARTA**

2023

**KLASIFIKASI SENTIMEN ANALISIS TWITTER UNTUK
MENILAI OPINI PUBLIK TERHADAP PERTANDINGAN SEPAK
BOLA MENGGUNAKAN METODE NAÏVE BAYES**

SKRIPSI

untuk memenuhi salah satu syarat mencapai derajat Sarjana
Program Studi Informatika



disusun oleh

HAFIIDH KHOIRU PRADANA

19.11.2645

Kepada

**FAKULTAS ILMU KOMPUTER
UNIVERSITAS AMIKOM YOGYAKARTA
YOGYAKARTA**

2023

HALAMAN PERSETUJUAN

SKRIPSI

**KLASIFIKASI SENTIMEN ANALISIS TWITTER UNTUK
MENILAI OPINI PUBLIK TERHADAP PERTANDINGAN SEPAK
BOLA MENGGUNAKAN METODE NAÏVE BAYES**

yang disusun dan diajukan oleh

Hafidh Khoiru Pradana

19.11.2645

telah disetujui oleh Dosen Pembimbing Skripsi
pada tanggal 30 Maret 2023

Dosen Pembimbing,



Yuli Astuti, M.Kom

NIK. 190302146

HALAMAN PENGESAHAN

SKRIPSI

**KLASIFIKASI SENTIMEN ANALISIS TWITTER UNTUK MENILAI OPINI
PUBLIK TERHADAP PERTANDINGAN SEPAK BOLA MENGGUNAKAN
METODE NAÏVE BAYES**

yang disusun dan diajukan oleh

Hafidh Khoiru Pradana

19.11.2645

Telah dipertahankan di depan Dewan Penguji
pada tanggal 30 Maret 2023

Susunan Dewan Penguji

Nama Penguji

Tanda Tangan

**Yoga Pristyanto, S.Kom, M.Eng.
NIK. 190302412**



**Rumini, M.Kom
NIK. 190302246**



**Yuli Astuti, M.Kom
NIK. 190302146**



Skripsi ini telah diterima sebagai salah satu persyaratan
untuk memperoleh gelar Sarjana Komputer
Tanggal 30 Maret 2023

DEKAN FAKULTAS ILMU KOMPUTER



**Hanif Al Fatta, S.Kom., M.Kom.
NIK. 190302096**

HALAMAN PERNYATAAN KEASLIAN SKRIPSI

Yang bertandatangan di bawah ini,

Nama mahasiswa : Hafidh Khoiru Pradana

NIM : 19.11.2645

Menyatakan bahwa Skripsi dengan judul berikut:

Klasifikasi Sentimen Analisis Twitter Untuk Menilai Opini Publik Terhadap Pertandingan Sepak Bola Menggunakan Metode Naïve Bayes

Dosen Pembimbing : Yuli Astuti, M.Kom

1. Karya tulis ini adalah benar-benar **ASLI** dan **BELUM PERNAH** diajukan untuk mendapatkan gelar akademik, baik di Universitas AMIKOM Yogyakarta maupun di Perguruan Tinggi lainnya.
2. Karya tulis ini merupakan gagasan, rumusan dan penelitian **SAYA** sendiri, tanpa bantuan pihak lain kecuali arahan dari Dosen Pembimbing.
3. Dalam karya tulis ini tidak terdapat karya atau pendapat orang lain, kecuali secara tertulis dengan jelas dicantumkan sebagai acuan dalam naskah dengan disebutkan nama pengarang dan disebutkan dalam Daftar Pustaka pada karya tulis ini.
4. Perangkat lunak yang digunakan dalam penelitian ini sepenuhnya menjadi tanggung jawab **SAYA**, bukan tanggung jawab Universitas AMIKOM Yogyakarta.
5. Pernyataan ini **SAYA** buat dengan sesungguhnya, apabila di kemudian hari terdapat penyimpangan dan ketidakbenaran dalam pernyataan ini, maka **SAYA** bersedia menerima **SANKSI AKADEMIK** dengan pencabutan gelar yang sudah diperoleh, serta sanksi lainnya sesuai dengan norma yang berlaku di Perguruan Tinggi.

Yogyakarta, 22 Maret 2023

Yang Menyatakan,



Hafidh Khoiru Pradana

HALAMAN PERSEMBAHAN

Alhamdulillahilahirabbil'alamin, segala puji bagi ALLAH SWT berkat rahmat serta hidayah-Nya sehingga penyusunan skripsi ini dapat terselesaikan dengan lancar. Skripsi ini saya persembahkan untuk:

1. Kedua orang tua saya yang telah selalu memberikan doa, restu serta dukungan motivasi kepada saya. Saya sangat bersyukur memiliki keluarga yang saling menyayangi.
2. Ibu Yuli Astuti, M.Kom selaku dosen pembimbing saya yang senantiasa membimbing saya dengan sabar dan ikhlas membantu saya ketika saya kesulitan dalam menyusun skripsi sehingga saya dapat mengerjakan skripsi dengan lancar. Terima Kasih untuk ilmu yang telah diberikan kepada saya.
3. Teman-teman saya di kelas 19-IF01 atas pertemanan dan kekeluargaan selama 4 tahun ini. Bantuan dan dukungan kalian sangat membantu saya selama perkuliahan.
4. Teman-Teman Takeshi Castle yang selalu mendukung dan membantu saya selama perkuliahan dan penyusunan skripsi. Terima kasih untuk kenangan, persahabatan, dan kekeluargaan selama ini semoga kita bisa meraih mimpi kita masing-masing.
5. Warmindo Gorongan terkhusus untuk teman saya Ari yang telah sedia menyediakan tempat untuk nongkrong selama kuliah dan tempat untuk mengerjakan skripsi.

KATA PENGANTAR

Alhamdulillahirobbil'alamin, segala puji dan syukur penulis panjatkan kepada Allah SWT yang telah memberikan nikmat, rahmat, hidayah dan inayah-Nya sehingga penulis dapat menyelesaikan skripsi ini. Penulis juga mengucapkan terimakasih kepada pihak-pihak yang telah membimbing dan membantu sehingga penulis dapat menyelesaikan skripsi yang berjudul "Klasifikasi Sentimen Analisis Twitter Untuk Menilai Opini Publik Terhadap Pertandingan Sepak Bola Menggunakan Metode Naïve Bayes" dengan sebaik-baiknya.

Selanjutnya penulis mengucapkan terimakasih atas segala dukungan dan bantuan yang diberikan selama mengerjakan penelitian ini kepada :

1. Bapak Prof. Dr. M. Suyanto, MM. selaku rektor Universitas Amikom Yogyakarta.
2. Ibu Yuli Astuti, M.Kom. selaku dosen pembimbing yang selalu membimbing penulis dalam melakukan penelitian ini.
3. Kedua orang tua saya yang telah selalu memberikan doa, restu serta dukungan motivasi kepada saya.
4. Bapak dan Ibu dosen yang telah memberikan ilmu dan pengetahuan yang sangat berguna bagi penulis selama masa perkuliahan.

Penulis menyadari bahwa makalah ini jauh dari kata sempurna dan masih terdapat beberapa kekurangan, oleh karena itu penulis sangat mengharapkan saran dan kritik yang membangun dari pembaca untuk penyempurnaan makalah ini.

Yogyakarta, November 2022

Penulis

DAFTAR ISI

HALAMAN JUDUL	i
HALAMAN PERSETUJUAN	ii
HALAMAN PENGESAHAN	iii
HALAMAN PERNYATAAN KEASLIAN SKRIPSI	iv
HALAMAN PERSEMBAHAN	v
KATA PENGANTAR	vi
DAFTAR ISI	vii
DAFTAR TABEL	xi
DAFTAR GAMBAR	xiii
DAFTAR LAMBANG DAN SINGKATAN	xv
DAFTAR ISTILAH	xvi
INTISARI	xvii
ABSTRACT	xviii
BAB I PENDAHULUAN	1
1.1 Latar Belakang	1
1.2 Rumusan Masalah	2
1.3 Batasan Masalah	2
1.4 Tujuan Penelitian	3
1.5 Manfaat Penelitian	3
1.6 Sistematika Penulisan	3
BAB II TINJAUAN PUSTAKA	5
2.1 Studi Literatur	5
2.2 Dasar Teori	14
2.2.1 Twitter	14
2.2.2 Klasifikasi	14
2.2.3 Analisa Sentimen	14

2.2.4	Sepak Bola	15
2.2.5	Naïve Bayes	15
2.2.6	Support Vector Machine	16
2.2.7	Decision Tree	16
2.2.8	Python	17
2.2.9	Google Colab	17
2.2.10	Confusion Matrix	17
2.2.11	TF-IDF (Term Frequency-Inverse Document Frequency)	18
2.2.12	VADER (Valence Aware Dictionary and Sentiment Reasoner)	18
BAB III METODE PENELITIAN		19
3.1	Objek Penelitian	19
3.2	Alur Penelitian	19
3.2.1	Pengumpulan data	20
3.2.2	Preprocessing data	21
3.2.3	Pemodelan	21
3.2.4	Evaluasi dan pengujian	21
3.2.5	Dokumentasi	22
3.3	Alat dan Bahan	22
3.3.1	Bahan Penelitian	22
3.3.2	Alat	22
3.4	Metodologi Penelitian	23
3.4.1	Pengumpulan Data	23
3.4.2	Text Preprocessing	23
3.4.3	Pelabelan	30
3.4.4	Pelabelan Otomatis	30
3.4.5	Pembagian Data	31

3.4.6	Pembobotan TF-IDF	31
3.4.7	Implementasi Metode Naïve Bayes	36
3.4.8	Pengujian Label Manual	41
3.4.9	Pengujian Label Otomatis	51
BAB IV HASIL DAN PEMBAHASAN		60
4.1	Pembahasan	60
4.1.1	Crawling data	60
4.1.2	Case Folding	61
4.1.3	Menghapus Data Kosong	65
4.1.4	Normalisasi	65
4.1.5	Translator	68
4.1.6	VADER (Valence Aware Dictionary and Sentiment Reasoner)	69
4.1.7	Stopword	70
4.1.8	Tokenizing	71
4.1.9	Stemming	71
4.1.10	Pembobotan TF-IDF	72
4.1.11	Split Data	73
4.1.12	Implementasi Metode	74
4.2	Pengujian dan Hasil	75
4.2.1	Library dan Module	75
4.2.2	Pengujian Confusion Matrix Label Manual	75
4.2.3	Pengujian Confusion Matrix Label Otomatis	77
4.2	Hasil	78
BAB V PENUTUP		80
5.1	Kesimpulan	80
5.2	Saran	80

DAFTAR PUSTAKA	82
LAMPIRAN.....	85



DAFTAR TABEL

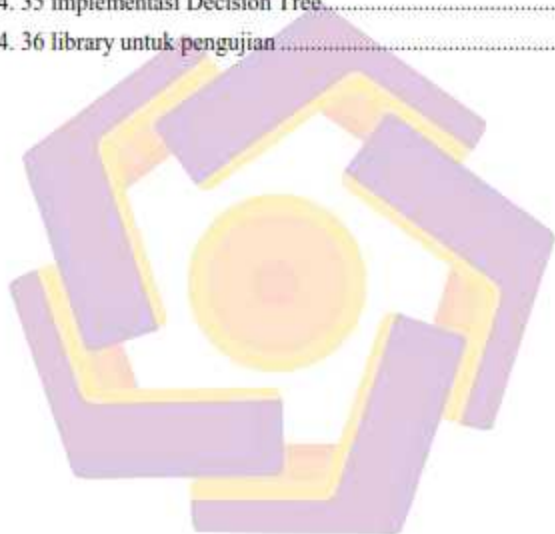
Tabel 2.1 Keaslian Penelitian	14
Tabel 3. 1 Hasil Case Folding.....	24
Tabel 3. 2 Hasil Normalisasi.....	26
Tabel 3. 3 Hasil Stopword	27
Tabel 3. 4 Hasil Stemming.....	28
Tabel 3. 5 Hasil Tokenize	29
Tabel 3. 6 Komentar Twitter.....	31
Tabel 3. 7 Hasil <i>Text Preprocessing</i>	32
Tabel 3. 8 Term Frequency (TF) dan Document Frequency (DF).....	32
Tabel 3. 9 Perhitungan Inverse Document Frequency (IDF).....	33
Tabel 3. 10 Perhitungan bobot kata	34
Tabel 3. 11 Tabel Data Latih	36
Tabel 3. 12 Tabel Data Uji.....	37
Tabel 3. 13 Hasil Perhitungan Term Pada Kelas	40
Tabel 3. 14 Hasil Pengujian NB pada perbandingan 70:30	42
Tabel 3. 15 Hasil akurasi pada perbandingan 70:30	42
Tabel 3. 16 Hasil Pengujian NB pada perbandingan 80:20	43
Tabel 3. 17 Hasil akurasi pada perbandingan 80:20	43
Tabel 3. 18 Hasil Pengujian NB pada perbandingan 90:10	44
Tabel 3. 19 Hasil akurasi pada perbandingan 90:10	44
Tabel 3. 20 Hasil Pengujian SVM pada perbandingan 70:30	45
Tabel 3. 21 Hasil akurasi pada perbandingan 70:30	45
Tabel 3. 22 Hasil Pengujian SVM pada perbandingan 80:20	46
Tabel 3. 23 Hasil akurasi pada perbandingan 80:20	46
Tabel 3. 24 Hasil Pengujian SVM pada perbandingan 90:10	47
Tabel 3. 25 Hasil pada perbandingan 90:10.....	47
Tabel 3. 26 Hasil Pengujian Decision Tree pada perbandingan 70:30	48
Tabel 3. 27 Hasil akurasi pada perbandingan 70:30	48
Tabel 3. 28 Hasil Pengujian Decision Tree pada perbandingan 80:20	49
Tabel 3. 29 Hasil akurasi pada perbandingan 80:20	49

Tabel 3. 30 Hasil Pengujian Decision Tree pada perbandingan 90:10	50
Tabel 3. 31 Hasil pada perbandingan 90:10.....	50
Tabel 3. 32 Pengujian Label Otomatis NB 70:30	51
Tabel 3. 33 Hasil perhitungan 70:30 label Otomatis	51
Tabel 3. 34 Pengujian Label Otomatis NB 80:20	52
Tabel 3. 35 Hasil perhitungan 80:20 label Otomatis	52
Tabel 3. 36 Pengujian Label Otomatis NB 90:10	53
Tabel 3. 37 Hasil perhitungan 90:10 label Otomatis	53
Tabel 3. 38 Pengujian Label Otomatis SVM 70:30	54
Tabel 3. 39 Hasil perhitungan SVM 70:30 label Otomatis.....	54
Tabel 3. 40 Pengujian Label Otomatis SVM 80:20	54
Tabel 3. 41 Hasil perhitungan SVM 80:20 label Otomatis.....	55
Tabel 3. 42 Pengujian Label Otomatis SVM 90:10	55
Tabel 3. 43 Hasil perhitungan SVM 90:10 label Otomatis.....	56
Tabel 3. 44 Pengujian Label Otomatis Decision Tree 70:30	56
Tabel 3. 45 Hasil perhitungan Decision Tree 70:30 label Otomatis.....	57
Tabel 3. 46 Pengujian Label Otomatis Decision Tree 80:20	57
Tabel 3. 47 Hasil perhitungan Decision Tree 80:20 label Otomatis.....	57
Tabel 3. 48 Pengujian Label Otomatis Decision Tree 90:10.....	58
Tabel 3. 49 Hasil perhitungan Decision Tree 90:10 label Otomatis.....	58
Tabel 4. 1 Hasil akurasi NB pada dataset label manual.....	76
Tabel 4. 2 Hasil akurasi SVM pada dataset label manual.....	76
Tabel 4. 3 Hasil akurasi Decision Tree pada dataset label manual	77
Tabel 4. 4 Hasil akurasi NB pada dataset label otomatis.....	77
Tabel 4. 5 Hasil akurasi SVM pada dataset label manual.....	78
Tabel 4. 6 Hasil akurasi Decision Tree pada dataset label manual	78

DAFTAR GAMBAR

Gambar 3. 1 Alur Penelitian	20
Gambar 3. 2 Hasil Pengumpulan Data.....	23
Gambar 3. 3 Kamus Alay.....	26
Gambar 3. 4 Tampilan <i>Dataset</i> Pada Microsoft Excel	30
Gambar 4. 1 Instal <i>library snsrape</i>	60
Gambar 4. 2 source code crawling data	60
Gambar 4. 3 hasil crawling data	61
Gambar 4. 4 source code <i>lowercase</i>	61
Gambar 4. 5 <i>source code</i>	62
Gambar 4. 6 source code import module Regular expression	62
Gambar 4. 7 source code menerapkan fungsi pada dataframe.....	62
Gambar 4. 8 <i>source code</i> menghapus angka	63
Gambar 4. 9 <i>source code</i> menghapus tanda baca	63
Gambar 4. 10 contoh penggunaan fungsi <i>strip()</i>	64
Gambar 4. 11 contoh penggunaan fungsi <i>strip()</i>	64
Gambar 4. 12 penggunaan <i>module Regular expression</i>	64
Gambar 4. 13 penggunaan <i>module Regular expression</i>	65
Gambar 4. 14 menghapus data kosong dan data duplicate	65
Gambar 4. 15 upload file.....	66
Gambar 4. 16 membaca file	66
Gambar 4. 17 kamus alay.....	67
Gambar 4. 18 <i>dictionary</i> kamus alay	67
Gambar 4. 19 membuat fungsi normalisasi	68
Gambar 4. 20 menerapkan fungsi normalisasi	68
Gambar 4. 21 import Translator.....	68
Gambar 4. 22 fungsi terjemah.....	69
Gambar 4. 23 nltk Vader lexicon	69
Gambar 4. 24 menghitung <i>score sentiment</i> dengan Vader	69
Gambar 4. 25 pemberian label pada dataset	70
Gambar 4. 26 menerapkan fungsi stopword	71

Gambar 4. 27 menerapkan fungsi <i>word_tokenize()</i>	71
Gambar 4. 28 menerapkan library Sastrawi pada stemming	72
Gambar 4. 29 menerapkan pembobotan TF-IDF	72
Gambar 4. 30 pembagian data latih dan data uji 1	73
Gambar 4. 31 pembagian data latih dan data uji 2.....	73
Gambar 4. 32 pembagian data latih dan data uji 3.....	73
Gambar 4. 33 implementasi <i>Naïve Bayes</i>	74
Gambar 4. 34 implementasi Support Vector Machine.....	74
Gambar 4. 35 implementasi Decision Tree.....	74
Gambar 4. 36 library untuk pengujian	75



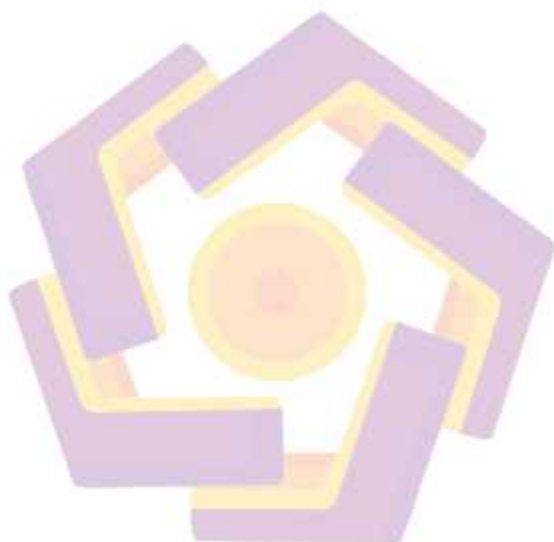
DAFTAR LAMBANG DAN SINGKATAN

nltk	Natural Language Toolkit
regex	Regular Expression
sklearn	Scikit-learn
SVM	Support Vector Machine



DAFTAR ISTILAH

Trending Topic	Topik atau berita yang paling sering dibahas
Tweet	Komentar pada media sosial <i>twitter</i>
Library	Kumpulan kode yang sebelumnya sudah dikompilasi.



INTISARI

Pada tanggal 1 Oktober 2022 terjadi sebuah tragedi di Stadion Kanjuruhan Malang yang terjadi pada pendukung dari tim Arema FC. Tragedi Kanjuruhan menuai banyak komentar di berbagai media sosial, hal tersebut membuat peneliti tertarik untuk melakukan penelitian dengan judul “Klasifikasi Sentimen Analisis Twitter Untuk Menilai Opini Publik Terhadap Pertandingan Sepak Bola Menggunakan Metode Naïve Bayes”. Penelitian ini dilakukan sebagai bentuk informasi perbandingan jumlah komentar positif dan negative di media social Twitter dan untuk mengetahui akurasi dari metode klasifikasi yang digunakan.

Penelitian ini menggunakan metode klasifikasi *Naïve Bayes*. Data yang digunakan dalam penelitian ini sebanyak 2052 data yang terdiri dari 1015 data positif dan 1037 data negatif. Data yang digunakan ini nanti akan dibagi menjadi data latih dan data uji. Untuk mengetahui pengaruh jumlah data latih pada akurasi yang dihasilkan, maka pengujian akan dilakukan sebanyak tiga kali dengan kombinasi data latih dan data uji yang berbeda yaitu 70:30, 80:20, dan 90:10. Hasil dari penelitian ini diperoleh nilai akurasi tertinggi sebesar 79%, *precision* 79%, dan *recall* 79%. Program ini dapat dikembangkan untuk platform media sosial lainnya seperti facebook, instagram, dan lainnya.

Kata kunci: *Naïve Bayes*, Komentar ,Sentimen Analisis, Sepak Bola, Kanjuruhan.

ABSTRACT

On October 1, 2022, a tragedy occurred at the Malang Kanjuruhan Stadium, which occurred to supporters of the Arema FC team. The Kanjuruhan tragedy reaped a lot of comments on various social media, this made researchers interested in conducting research with the title "Classification of Twitter Analysis Sentiments to Assess Public Opinion of Football Matches Using the Naïve Bayes Method". This research was conducted as a form of comparative information on the number of positive and negative comments on social media Twitter and to determine the accuracy of the classification method used.

This study uses the Naïve Bayes classification method. The data used in this study were 2052 data consisting of 1015 positive data and 1037 negative data. The data used will be divided into training data and test data. To find out the effect of the amount of training data on the resulting accuracy, the test will be carried out three times with different combinations of training data and test data, namely 70:30, 80:20, and 90:10. The results of this study obtained the highest accuracy values of 79%, 79% precision, and 79% recall. This program can be developed for other social media platforms such as Facebook, Instagram, and others.

Keyword: *Naïve Bayes, Commentary, Sentiment Analysis, Football, Kanjuruhan.*