

**PERBANDINGAN NAÏVE BAYES, SVM DAN DECISION  
TREE PADA ANALISIS SENTIMEN TERHADAP  
PENYELENGGARAAN SEA GAMES 2023 DI TWITTER**

**SKRIPSI**

Diajukan untuk memenuhi salah satu syarat mencapai derajat Sarjana  
Program Studi Sistem Informasi



disusun oleh

**Muhammad Fathy Islam**

**18.12.0976**

Kepada

**FAKULTAS ILMU KOMPUTER  
UNIVERSITAS AMIKOM YOGYAKARTA  
YOGYAKARTA**

**2023**

**PERBANDINGAN NAÏVE BAYES, SVM DAN DECISION  
TREE PADA ANALISIS SENTIMEN TERHADAP  
PENYELENGGARAAN SEA GAMES 2023 DI TWITTER  
SKRIPSI**

Diajukan untuk memenuhi salah satu syarat mencapai derajat Sarjana  
Program Studi Sistem Informasi



disusun oleh

**Muhammad Fathy Islam**

**18.12.0976**

Kepada

**FAKULTAS ILMU KOMPUTER  
UNIVERSITAS AMIKOM YOGYAKARTA  
YOGYAKARTA  
2023**

**HALAMAN PERSETUJUAN**

**SKRIPSI**

**PERBANDINGAN NAÏVE BAYES, SVM DAN DECISION  
TREE PADA ANALISIS SENTIMEN TERHADAP  
PENYELENGGARAAN SEA GAMES 2023 DI TWITTER**

yang disusun dan diajukan oleh

**Muhammad Fathy Islam**

**18.12.0976**

telah disetujui oleh Dosen Pembimbing Skripsi  
pada tanggal 7 Agustus 2023

Dosen Pembimbing,



**Norhikmah, M.Kom**  
**NIK. 190302245**

HALAMAN PENGESAHAN

SKRIPSI

**Perbandingan Naïve Bayes, SVM dan Decision Tree Pada Analisis Sentimen Terhadap Penyelenggaraan SEA Games 2023 di Twitter**

yang disusun dan diajukan oleh

**Muhammad Fathy Islam**

18.12.0976

Telah dipertahankan di depan Dewan Penguji  
pada tanggal 7 Agustus 2023

**Susunan Dewan Penguji**

**Nama Penguji**

**Tanda Tangan**

Ferian Fauzi Abdullah, M.Kom

NIK. 190302276

Yoga Pristvanto, S.Kom, M.Eng

NIK. 190302412

Norhikmah, M.Kom

NIK. 190302245

Skripsi ini telah diterima sebagai salah satu persyaratan  
untuk memperoleh gelar Sarjana Komputer  
Tanggal 7 Agustus 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 : **Muhammad Fathy Islam**  
NIM : **18.12.0976**

Menyatakan bahwa Skripsi dengan judul berikut:

**Perbandingan Naïve Bayes, SVM dan Decision Tree Pada Analisis Sentimen Terhadap Penyelenggaraan SEA Games 2023 di Twitter**

Dosen Pembimbing : **Norhikmah, 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, 7 Agustus 2023

Yang Menyatakan,



Muhammad Fathy Islam

## KATA PENGANTAR

Segala puji bagi Allah SWT atas nikmat dan rahmat-Nya sehingga penulis dapat menyelesaikan penyusunan skripsi ini. Penulisan skripsi ini merupakan langkah penting dalam pencapaian gelar Sarjana Komputer Program Studi Informatika Fakultas Ilmu Komputer Universitas Amikom Yogyakarta. Dalam proses ini, banyak pihak yang memberikan bantuan, dukungan, serta saran yang sangat berarti. Oleh karena itu, penulis ingin mengucapkan rasa terima kasih kepada:

1. Allah SWT, yang telah memberikan petunjuk, pertolongan, dan kekuatan dalam proses penulisan skripsi ini.
2. Kedua orangtua penulis, Bapak Agung Hamid dan Ibu Panca Budi Yanuakhiriyah, yang selalu memberikan doa, dukungan, dan motivasi dalam menyelesaikan skripsi ini.
3. Ibu Norhimah, M.Kom, selaku dosen pembimbing skripsi, atas waktu, bimbingan, arahan, dan masukan berharga yang diberikan hingga penyelesaian skripsi.
4. Seluruh dosen Fakultas Ilmu Komputer Universitas Amikom Yogyakarta yang telah berkontribusi dalam pembentukan ilmu pengetahuan penulis
5. Teman-teman dan semua pihak yang telah berbagi pengalaman dan pengetahuan selama masa studi.

Skripsi ini mungkin masih memiliki kekurangan, oleh karena itu, kritik, saran, dan masukan dari pembaca sangat diharapkan. Semoga skripsi ini dapat bermanfaat, memberikan wawasan, dan pengetahuan bagi para pembaca.

Yogyakarta, 7 Agustus 2023

Penulis

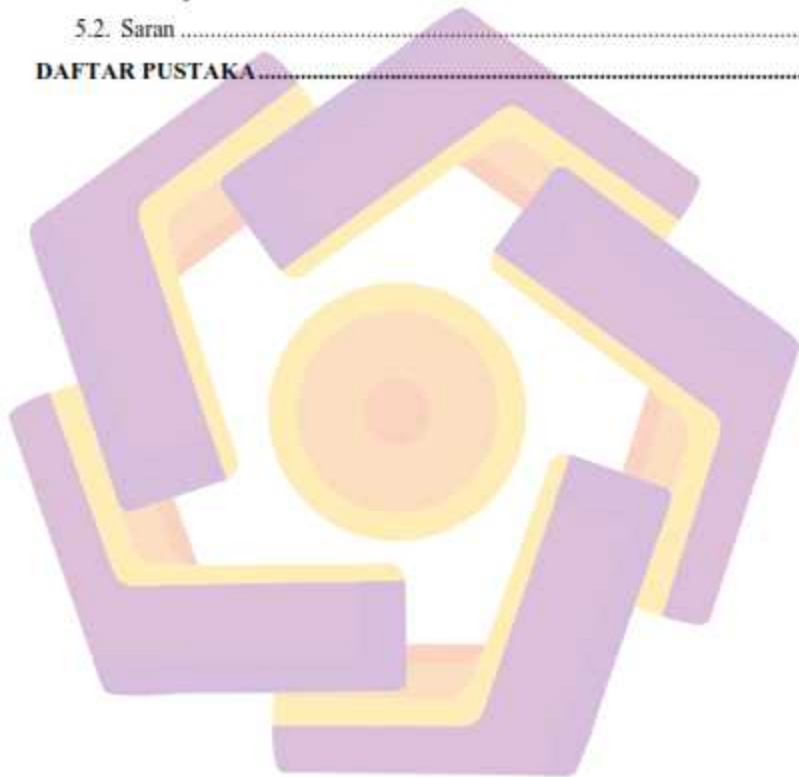
## DAFTAR ISI

JUDUL .....	i
HALAMAN PERSETUJUAN .....	ii
HALAMAN PENGESAHAN .....	iii
HALAMAN PERNYATAAN KEASLIAN SKRIPSI.....	iv
KATA PENGANTAR.....	v
DAFTAR ISI.....	vi
DAFTAR GAMBAR.....	ix
DAFTAR TABEL .....	x
INTISARI .....	xi
ABSTRACT .....	xii
<b>BAB I PENDAHULUAN.....</b>	<b>1</b>
1.1 Latar Belakang .....	1
1.2 Rumusan Masalah.....	4
1.3 Batasan Masalah .....	4
1.3.1. <i>Batasan Temporal</i> .....	4
1.3.2. <i>Batasan Data</i> .....	4
1.3.3. <i>Batasan Klasifikasi Sentimen</i> .....	5
1.3.4. <i>Batasan Ruang Lingkup Penelitian</i> .....	5
1.4 Tujuan Penelitian .....	5
1.5 Manfaat Penelitian .....	6
1.6 Sistematika Penulisan .....	6
<b>BAB II TINJAUAN PUSTAKA .....</b>	<b>8</b>
2.1 Studi Literatur .....	8
2.2 Dasar Teori .....	15
2.2.1 <i>Sentimen Analisis</i> .....	15
2.2.2 <i>Text Mining</i> .....	15
2.2.3 <i>Text Preprocessing</i> .....	15
2.2.4 <i>Term Frequency/Inverse Document Frequency (TF/IDF)</i> .....	16
2.2.5 <i>Naive Bayes</i> .....	16
2.2.6 <i>Support Vector Machine (SVM)</i> .....	17

2.2.7	<i>Decision Tree</i>	17
2.2.8	<i>Confusion Matrix</i>	18
2.2.9	<i>Python</i>	19
2.2.10	<i>Twitter</i>	20
2.2.11	<i>SEA Games</i>	20
<b>BAB III METODE PENELITIAN</b>		<b>21</b>
3.1	Alat dan Bahan Penelitian	21
3.1.1	<i>Perangkat Keras/Hardware</i>	21
3.1.2	<i>Perangkat Lunak/Software</i>	21
3.2	Alur Penelitian	22
3.2.1	<i>Scraping data</i>	22
3.2.2	<i>Labeling</i>	22
3.2.3	<i>Text Preprocessing</i>	23
3.2.4	<i>Splitting data</i>	25
3.2.5	<i>Pembobotan Kata</i>	25
3.2.6	<i>Klasifikasi Data</i>	26
3.2.7	<i>Pengujian dan evaluasi</i>	27
<b>BAB IV HASIL DAN PEMBAHASAN</b>		<b>29</b>
4.1	Pengumpulan Data	29
4.2	Labeling	30
4.3	Preprocessing Data	30
4.3.1	<i>Case Folding</i>	31
4.3.2	<i>Cleaning</i>	32
4.3.3	<i>Tokenizing</i>	35
4.3.4	<i>Normalisasi</i>	36
4.3.5	<i>Stopword Removal</i>	38
4.3.6	<i>Stemming</i>	39
4.4	Splitting	43
4.5	Pembobotan Kata (TF-IDF)	44
4.6	Klasifikasi Data	46
4.6.1	<i>Naive Bayes</i>	46
4.6.2	<i>Support Vector Machine</i>	47
4.6.3	<i>Decision Tree</i>	48



4.7. Evaluasi dan pengujian tingkat akurasi dengan confusion matrix.....	50
4.7.1. <i>Gaussian Naive Bayes</i> .....	53
4.7.2. <i>Support Vector Machine Linear Kernel</i> .....	55
4.7.3. <i>Decision Tree Information Gain</i> .....	57
<b>BAB V PENUTUP</b> .....	<b>60</b>
5.1. Kesimpulan.....	60
5.2. Saran.....	60
<b>DAFTAR PUSTAKA</b> .....	<b>62</b>



## DAFTAR GAMBAR

Gambar 3. 1	Alur Penelitian.....	22
Gambar 4. 1	Alur Preprocessing .....	31
Gambar 4. 2	Kode proses case folding.....	31
Gambar 4. 3	Kode proses cleaning.....	33
Gambar 4. 4	Kode Proses Tokenisasi.....	35
Gambar 4. 5	Kode Proses Normalisasi.....	36
Gambar 4. 6	Kode proses stopwords removal .....	38
Gambar 4. 7	Kode proses stemming .....	40
Gambar 4. 8	Kode proses word distribution .....	42
Gambar 4. 9	Grafik frekuensi kemunculan kata .....	43
Gambar 4. 10	Kode proses splitting data .....	44
Gambar 4. 11	Kode proses TF-IDF.....	45
Gambar 4. 12	Kode Gaussian Naive Bayes .....	46
Gambar 4. 13	Kode SVM linear kernel.....	47
Gambar 4. 14	Kode decision tree information gain .....	49
Gambar 4. 15	Kode diagram batang perbandingan.....	50
Gambar 4. 16	Diagram batang perbandingan model.....	51
Gambar 4. 17	Kode confusion matrix.....	52
Gambar 4. 18	Kode dan output confusion matrix Naive Bayes.....	53
Gambar 4. 19	Kode dan output confusion matrix SVM .....	55
Gambar 4. 20	Kode dan output confusion matrix decision tree.....	57

## DAFTAR TABEL

Tabel 2. 1	Keaslian Penelitian .....	10
Tabel 2. 2	Confusion Matrix .....	19
Tabel 4. 1	Hasil dari scraping Tweet Harvest .....	29
Tabel 4. 2	Hasil labeling.....	30
Tabel 4. 3	Output case folding .....	32
Tabel 4. 4	Output Cleaning .....	34
Tabel 4. 5	Output Tokenizing.....	35
Tabel 4. 6	Kamus slang word.....	36
Tabel 4. 7	Output normalisasi .....	37
Tabel 4. 8	Output stopword removal.....	39
Tabel 4. 9	Output stemming.....	41
Tabel 4. 10	Confusion matrix Naive Bayes.....	54
Tabel 4. 11	Confusion matrix SVM.....	56
Tabel 4. 12	Confusion matrix Decision Tree .....	58
Tabel 4. 13	Tabel Evaluasi dan Perbandingan confusion matrix .....	59

## INTISARI

Penelitian ini bertujuan untuk menganalisis sentimen pengguna Twitter terhadap SEA Games 2023 menggunakan metode Naïve Bayes, SVM, dan Decision Tree. Respons publik terhadap peristiwa olahraga internasional melalui media sosial dapat memiliki dampak besar pada citra penyelenggaraan acara tersebut. Data tweet dari Twitter dikumpulkan, diproses, dan dianalisis menggunakan algoritma-algoritma tersebut untuk mengklasifikasikan sentimen menjadi positif, negatif, atau netral.

Hasil penelitian menunjukkan bahwa algoritma Gaussian Naive Bayes mencapai akurasi tertinggi sebesar 92%, diikuti oleh SVM (91%) dan Decision Tree (83%). Kesimpulan ini mengindikasikan bahwa algoritma Gaussian Naive Bayes adalah yang terbaik dalam menganalisis sentimen pengguna Twitter terhadap SEA Games 2023.

Penelitian ini memberikan pemahaman lebih dalam tentang cara publik merespons peristiwa olahraga internasional melalui media sosial dan dapat membantu penyelenggara acara serupa untuk meningkatkan citra mereka.

**Kata kunci:** Klasifikasi, Sentimen, SEA Games, Naïve Bayes, SVM, Decision Tree.

## ABSTRACT

*This research aims to analyze Twitter users' sentiment towards SEA Games 2023 using the Naïve Bayes, SVM, and Decision Tree methods. Public responses to international sporting events on social media can significantly impact the event's image. Data from Twitter tweets were collected, preprocessed, and analyzed using these algorithms to classify sentiment as positive, negative, or neutral.*

*The results indicate that the Gaussian Naïve Bayes algorithm achieved the highest accuracy at 92%, followed by SVM (91%), and Decision Tree (83%). This suggests that the Gaussian Naïve Bayes algorithm is the most effective in analyzing Twitter users' sentiment towards SEA Games 2023.*

*This research contributes to a deeper understanding of how the public responds to international sporting events through social media and can provide valuable insights for organizers of similar events to enhance their image.*

**Keyword: Classification, Sentiments, SEA Games, Naïve Bayes, SVM, Decision Tree.**