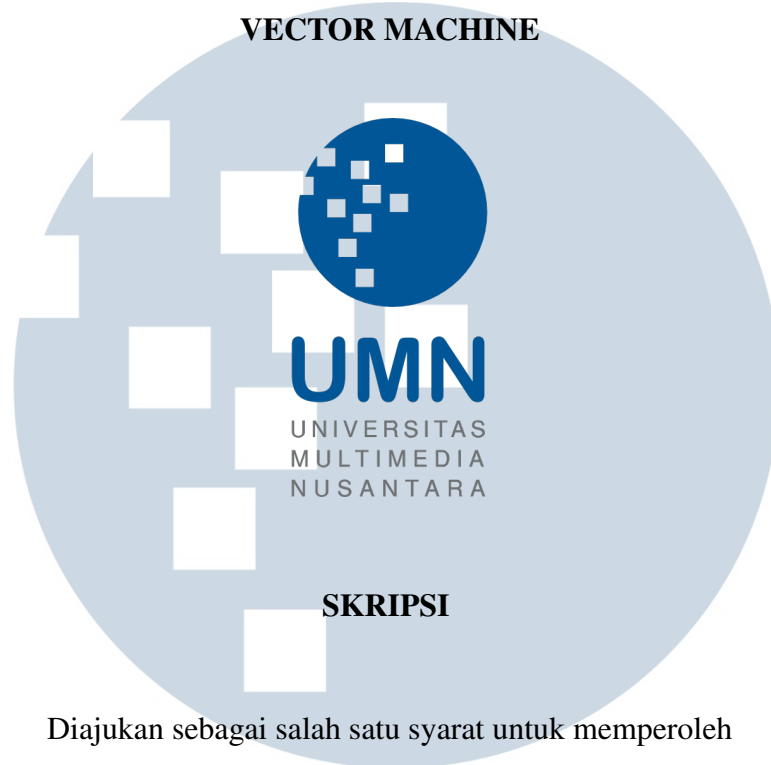


**ANALISIS SENTIMEN TWITTER TERKAIT PEMBATALAN PIALA
DUNIA U-20 DI INDONESIA MENGGUNAKAN METODE SUPPORT
VECTOR MACHINE**



SKRIPSI

Diajukan sebagai salah satu syarat untuk memperoleh
Gelar Sarjana Komputer (S.Kom.)

Muhammad Armanda

0000041497

UMN

UNIVERSITAS

MULTIMEDIA

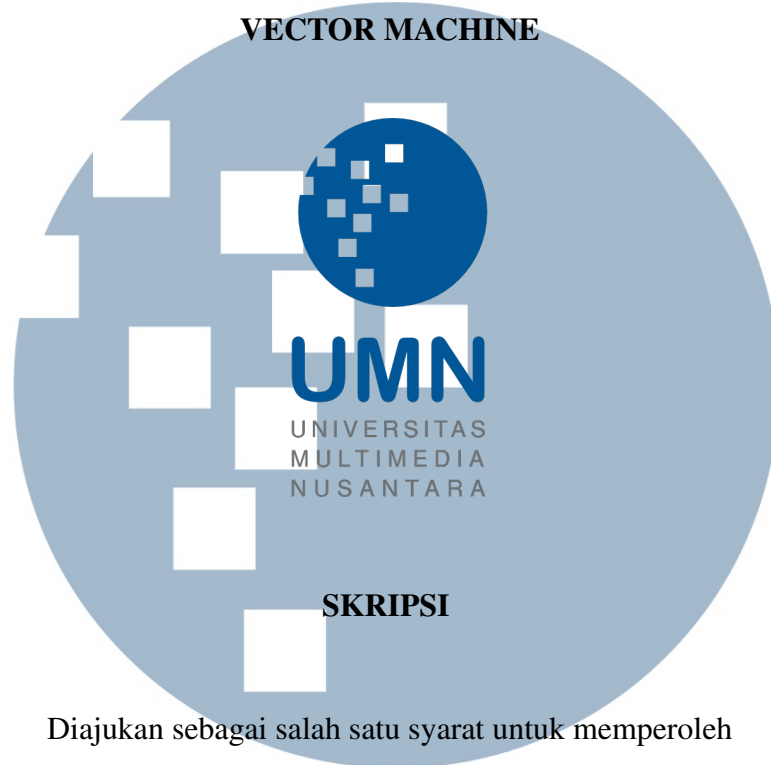
NUSANTARA

**PROGRAM STUDI INFORMATIKA
FAKULTAS TEKNIK DAN INFORMATIKA
UNIVERSITAS MULTIMEDIA NUSANTARA**

TANGERANG

2023

**ANALISIS SENTIMEN TWITTER TERKAIT PEMBATALAN PIALA
DUNIA U-20 DI INDONESIA MENGGUNAKAN METODE SUPPORT
VECTOR MACHINE**



SKRIPSI

Diajukan sebagai salah satu syarat untuk memperoleh
Gelar Sarjana Komputer (S.Kom.)

Muhammad Armanda

0000041497

UMN

UNIVERSITAS

MULTIMEDIA

NUSANTARA

**PROGRAM STUDI INFORMATIKA
FAKULTAS TEKNIK DAN INFORMATIKA
UNIVERSITAS MULTIMEDIA NUSANTARA**

TANGERANG

2023

HALAMAN PERNYATAAN TIDAK PLAGIAT

Dengan ini saya,

Nama : Muhammad Armanda
Nomor Induk Mahasiswa : 00000041497
Program Studi : Informatika

Skripsi dengan judul:

Analisis Sentimen Twitter Terkait Pembatalan Piala Dunia u-20 di Indonesia Menggunakan Metode Support Vector Machine

merupakan hasil karya saya sendiri bukan plagiat dari karya ilmiah yang ditulis oleh orang lain, dan semua sumber baik yang dikutip maupun dirujuk telah saya nyatakan dengan benar serta dicantumkan di Daftar Pustaka.

Jika di kemudian hari terbukti ditemukan kecurangan/ penyimpangan, baik dalam pelaksanaan Skripsi maupun dalam penulisan laporan Skripsi, saya bersedia menerima konsekuensi dinyatakan TIDAK LULUS untuk Tugas akhir yang telah saya tempuh.

Tangerang, 07 Juni 2023



(Muhammad Armanda)

UMM
UNIVERSITAS
MULTIMEDIA
NUSANTARA

HALAMAN PERSETUJUAN

Skripsi dengan judul

ANALISIS SENTIMEN TWITTER TERKAIT PEMBATALAN PIALA DUNIA U-20 DI INDONESIA MENGGUNAKAN METODE SUPPORT VECTOR MACHINE

oleh

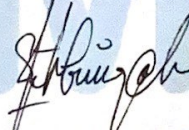
Nama : Muhammad Armanda
NIM : 00000041497
Program Studi : Informatika
Fakultas : Fakultas Teknik dan Informatika

Telah disetujui untuk diajukan pada

Sidang Ujian Skripsi Universitas Multimedia Nusantara

Tangerang, 07 Juni 2023

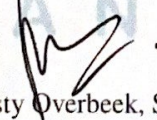
Pembimbing



(Fenina Adline Twince Tobing, S.Kom., M.Kom)

NIDN: 0406058802

Ketua Program Studi Informatika,



(Marlinda Vasty Overbeek, S.Kom., M.Kom.)

NIDN: 0818038501

HALAMAN PENGESAHAN

Skripsi dengan judul

**ANALISIS SENTIMEN TWITTER TERKAIT PEMBATALAN PIALA
DUNIA U-20 DI INDONESIA MENGGUNAKAN METODE SUPPORT
VECTOR MACHINE**

oleh

Nama : Muhammad Armanda
NIM : 00000041497
Program Studi : Informatika
Fakultas : Fakultas Teknik dan Informatika

Telah diujikan pada hari Senin, 12 Juni 2022

Pukul 10.00 s/d 12.00 dan dinyatakan


LULUS

Dengan susunan penguji sebagai berikut

Ketua Sidang

Penguji



(Marlinda Vasty Overbeek, S.Kom.,
M.Kom.)
NIDN: 0818038501


(Dennis Gunawan, S.Kom., M.Sc.)
NIDN: 0320059001

Pembimbing


(Fenina Adline Twince Tobing, S.Kom., M.Kom.)
NIDN: 0406058802

Ketua Program Studi Informatika,


(Marlinda Vasty Overbeek, S.Kom., M.Kom.)
NIDN: 0818038501

**HALAMAN PERSETUJUAN PUBLIKASI KARYA ILMIAH UNTUK
KEPENTINGAN AKADEMIS**

Sebagai sivitas akademik Universitas Multimedia Nusantara, saya yang bertanda tangan di bawah ini:

Nama : Muhammad Armanda
NIM : 00000041497
Program Studi : Informatika
Fakultas : Teknik dan Informatika
Jenis Karya : Skripsi

Demi pengembangan ilmu pengetahuan, menyetujui untuk memberikan kepada **Universitas Multimedia Nusantara** hak Bebas Royalti Non-eksklusif (*Non-exclusive Royalty-Free Right*) atas karya ilmiah saya yang berjudul:

**ANALISIS SENTIMEN TWITTER TERKAIT PEMBATALAN PIALA
DUNIA U-20 DI INDONESIA MENGGUNAKAN METODE SUPPORT
VECTOR MACHINE**

Beserta perangkat yang ada (jika diperlukan). Dengan Hak Bebas Royalti Non eksklusif ini Universitas Multimedia Nusantara berhak menyimpan, mengalih media / format-kan, mengelola dalam bentuk pangkalan data (*database*), merawat, dan mempublikasikan tugas akhir saya selama tetap mencantumkan nama saya sebagai penulis / pencipta dan sebagai pemilik Hak Cipta. Demikian pernyataan ini saya buat dengan sebenarnya.

Tangerang, 07 Juni 2023

Yang menyatakan



Muhammad Armanda

U N I V E R S I T A S
M U L T I M E D I A
N U S A N T A R A

Halaman Persembahan / Motto

”Berbuat kebaikan seperti memakai baju dalam, itu penting namun tidak untuk ditunjukkan”

Husein Ja’far Al Hadar

”Bismillah lulus”

Muhammad Armanda

UMMN

UNIVERSITAS
MULTIMEDIA
NUSANTARA

KATA PENGANTAR

Puji Syukur atas berkat dan rahmat kepada Tuhan Yang Maha Esa, atas selesainya penulisan laporan Skripsi ini dengan judul: Analisis Sentimen Twitter Terkait Pembatalan Piala Dunia u-20 di Indonesia Menggunakan Metode Support Vector Machine dilakukan untuk memenuhi salah satu syarat untuk mencapai gelar Sarjana Komputer Jurusan Informatika Pada Fakultas Teknik dan Informatika Universitas Multimedia Nusantara. Saya menyadari bahwa, tanpa bantuan dan bimbingan dari berbagai pihak, dari masa perkuliahan sampai pada penyusunan skripsi ini, sangatlah sulit bagi saya untuk menyelesaikan skripsi ini. Oleh karena itu, saya mengucapkan terima kasih kepada:

1. Bapak Dr. Ninok Leksono, selaku Rektor Universitas Multimedia Nusantara.
2. Dr. Eng. Niki Prastomo, S.T., M.Sc., selaku Dekan Fakultas Teknik dan Informatika Universitas Multimedia Nusantara.
3. Ibu Marlinda Vasty Overbeek, S.Kom., M.Kom., selaku Ketua Program Studi Informatika Universitas Multimedia Nusantara.
4. Ibu Fenina Adline Twince Tobing, S.Kom., M.Kom, sebagai Pembimbing pertama yang telah banyak meluangkan waktu untuk memberikan bimbingan, arahan dan motivasi atas terselesainya tesis ini.
5. Orang Tua dan keluarga saya yang telah memberikan bantuan dukungan material dan moral, sehingga penulis dapat menyelesaikan tesis ini.
6. Kepada diri saya sendiri karena telah mampu dan tidak menyerah sampai menyelesaikan tesis ini.

Semoga skripsi ini bermanfaat, baik sebagai sumber informasi maupun sumber inspirasi, bagi para pembaca.

Tangerang, 07 Juni 2023



Muhammad Armanda

ANALISIS SENTIMEN TWITTER TERKAIT PEMBATALAN PIALA DUNIA U-20 DI INDONESIA MENGGUNAKAN METODE SUPPORT VECTOR MACHINE

Muhammad Armanda

ABSTRAK

Pembatalan piala duni U-20 di Indonesia tahun 2023 menjadi perdebatan yang hangat dibicarakan dikalangan masyarakat Indonesia dikarenakan alasan dari pembatalan tersebut masih belum jelas. Banyaknya opinin pro dan kontra yang di unggah oleh masyarakat Indonesia di media sosial twitter menjadikan opini tersebut dapat menjadi data untuk menilai pendapat yang dibagi menjadi tiga kategori yaitu positif, negatif dan netral. Setelah dibagi menjadi tiga kategori selanjutnya akan dilakukan analisis sentimen menggunakan metode *support vector machine* dan membandingkan *kernel linear*, *polynomial* dan *rbf* untuk mendapatkan performa terbaik dari kernel yang ada dalam metode *support vector machine*. Dengan menggunakan confusion matrix untuk mengukur performa dari klasifikasi maka dapat dinilai accuracy, precision, recall dan f1-score. Didapatkan dengan perbandingan rasio data 80:20 memiliki akurasi tertinggi dari *kernel linear*, *polynomial*, *rbf* dan *kernel rbf* mendapatkan hasil yang lebih baik dibanding *kernel linear* dan *polynomial* yaitu Accuracy 78.15%, F1-Score, 76.30%, Precision 77.37% dan Recall 75.58%. Selain itu dari data yang didapatkan juga berhasil menganalisis teks berbahasa indonesia yang di input secara eksternal dan di kategorikan menjadi positif, netral dan negatif. Dari hasil yang telah didapatkan bahwa metode *support vector machine* berhasil di implementasi dalam analisis sentimen Pembatalan piala duni U-20 di Indonesia tahun 2023 pada media sosial twitter.

Kata kunci: Analisis Sentimen, confusion matrix, Piala Dunia U-20, Support Vector Machine, Twitter

U N I V E R S I T A S
M U L T I M E D I A
N U S A N T A R A

**TWITTER SENTIMENT ANALYSIS RELATED TO THE
CANCELLATION OF THE U-20 WORLD CUP IN INDONESIA USING
THE SUPPORT VECTOR MACHINE METHOD**

Muhammad Armanda

ABSTRACT

The cancellation of the U-20 world cup in Indonesia in 2023 has become a hot debate among Indonesians because the reasons for the cancellation are still unclear. The number of pro and con opinions uploaded by the Indonesian people on twitter social media makes these opinions can be used as data to assess opinions which are divided into three categories, namely positive, negative and neutral. After being divided into three categories, sentiment analysis will then be carried out using the *support vector machine* method and comparing *linear kernel*, *polynomial* and *rbf* to get the best performance of the kernel in the *support vector machine* method. By using confusion matrix to measure the performance of classification, accuracy, precision, recall and f1-score can be assessed. Obtained with a data ratio of 80:20 has the highest accuracy of *linear kernel*, *polynomial*, *rbf* and *rbf kernel* get better results than *linear kernel* and *polynomial* namely *Accuracy* 78.15%, *F1-Score*, 76.30%, *Precision* 77.37% and *Recall* 75.58%. In addition, the data obtained also successfully analyzed Indonesian texts that were input externally and categorized into positive, neutral and negative. From the results that have been obtained that the *support vector machine* method is successfully implemented in sentiment analysis of the U-20 world cup cancellation in Indonesia in 2023 on twitter social media.

Keywords: *confusion matrix, World Cup U-20, Sentiment Analysis Support Vector Machine, Twitter*

U N I V E R S I T A S
M U L T I M E D I A
N U S A N T A R A

DAFTAR ISI

HALAMAN JUDUL	i
PERNYATAAN TIDAK MELAKUKAN PLAGIAT	ii
HALAMAN PERSETUJUAN	iii
HALAMAN PENGESAHAN	iv
HALAMAN PERSETUJUAN PUBLIKASI ILMIAH	v
HALAMAN PERSEMBAHAN/MOTO	vi
KATA PENGANTAR	vii
ABSTRAK	viii
ABSTRACT	ix
DAFTAR ISI	x
DAFTAR GAMBAR	xii
DAFTAR TABEL	xiii
DAFTAR KODE	xiv
DAFTAR LAMPIRAN	xv
BAB 1 PENDAHULUAN	1
1.1 Latar Belakang Masalah	1
1.2 Rumusan Masalah	3
1.3 Batasan Permasalahan	3
1.4 Tujuan Penelitian	3
1.5 Manfaat Penelitian	4
1.6 Sistematika Penulisan	4
BAB 2 LANDASAN TEORI	6
2.1 Piala Dunia U-20	6
2.2 Analisa Sentimen	6
2.3 TF-IDF	6
2.4 <i>Support Vector Machine</i>	7
2.4.1 Kernel Linear	8
2.4.2 Kernel Polynomial	9
2.4.3 Kernel RBF	10
2.5 <i>Confusion Matrix</i>	12
2.5.1 <i>Accuracy</i>	13
2.5.2 <i>Precision</i>	13
2.5.3 <i>Recall</i>	13
2.5.4 <i>F1-score</i>	13
BAB 3 METODOLOGI PENELITIAN	15
3.1 Metodologi Penelitian	15
3.2 Perancangan Sistem	16
3.2.1 Gambaran Umum Perancangan	16
3.2.2 <i>Crawling Data</i>	17
3.2.3 <i>Labeling Data</i>	17
3.2.4 <i>Text Preprocessing</i>	19
3.2.5 <i>Train-Test Split Data</i>	20
3.2.6 <i>Apply TF-IDF</i>	20
3.2.7 <i>Apply SVM</i>	21
3.2.8 <i>Evaluation</i>	22
BAB 4 HASIL DAN DISKUSI	23
4.1 Spesifikasi Sistem	23
4.2 Implementasi Sistem	23

4.2.1	Crawling Data	23
4.2.2	Labelling data	25
4.2.3	Text Pre-processing	27
4.2.4	Training dan Testing Split Data	32
4.2.5	Apply TF-IDF	33
4.2.6	Apply SVM	33
4.2.7	Evaluasi Model	33
4.3	Uji Coba	35
4.3.1	Pengujian dengan Pembagian Data Training dan Data Testing	35
4.3.2	Pengujian dengan Input teks Berbahasa Indonesia	41
4.3.3	Pengujian dengan Input File Csv Berbahasa Indonesia	42
BAB 5	SIMPULAN DAN SARAN	43
5.1	Simpulan	43
5.2	Saran	43
	DAFTAR PUSTAKA	44



DAFTAR GAMBAR

Gambar 2.1	Kernel Linear	9
Gambar 2.2	Kernel Polynomial	10
Gambar 2.3	Kernel RBF	10
Gambar 2.4	<i>Confusion Matrix</i>	12
Gambar 3.1	Flowchart Gambaran Umum Perancangan	17
Gambar 3.2	Flowchart Labeling Data	18
Gambar 3.3	Flowchart Text Preprocessing	19
Gambar 3.4	Flowchart Apply TF-IDF	21
Gambar 3.5	Flowchart Apply SVM	22
Gambar 4.1	Hasil Crawling Data	24
Gambar 4.2	Hasil Pelabelan Vader	26
Gambar 4.3	Pie Chart Hasil Vader	27
Gambar 4.4	Confusion Matrix Hasil Pemodelan	34
Gambar 4.5	Confusion Matrix linear 60:40	36
Gambar 4.6	Confusion Matrix polynomial 60:40	36
Gambar 4.7	Confusion Matrix rbf 60:40	37
Gambar 4.8	Confusion Matrix linear 70:30	38
Gambar 4.9	Confusion Matrix polynomial 70:30	38
Gambar 4.10	Confusion Matrix rbf 70:30	39
Gambar 4.11	Confusion Matrix linear 80:30	40
Gambar 4.12	Confusion Matrix polynomial 80:20	40
Gambar 4.13	Confusion Matrix rbf 80:20	40
Gambar 4.14	Hasil Input File Csv	41
Gambar 4.15	Hasil Input File Csv	42



DAFTAR TABEL

Tabel 4.1	Table Translated Data	25
Tabel 4.2	<i>Table Hasil Case Folding</i>	28
Tabel 4.3	Table Hasil <i>Cleaning Data</i>	29
Tabel 4.4	Table Hasil <i>Tokenizing</i>	29
Tabel 4.5	Table Hasil <i>Normalisasi</i>	30
Tabel 4.6	Table Hasil <i>Stopword</i>	31
Tabel 4.7	Table Hasil <i>Tokenizing</i>	32
Tabel 4.8	Tabel Perbandingan Dataset 60:40	37
Tabel 4.9	Tabel Perbandingan Dataset 70:30	39
Tabel 4.10	Tabel Perbandingan Dataset 80:20	41



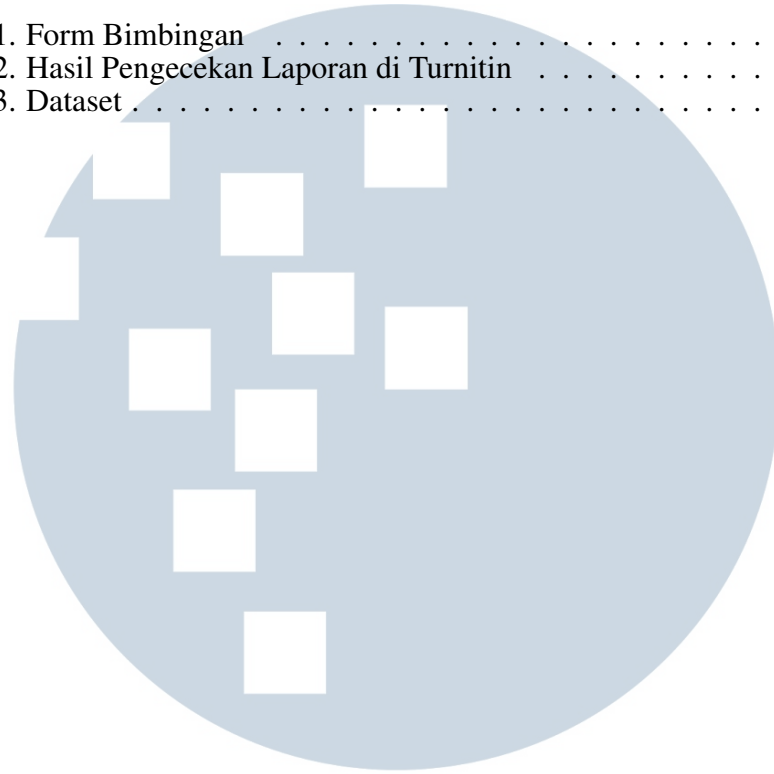
DAFTAR KODE

4.1	Potongan Kode Input File Csv	24
4.2	Potongan Kode Translated Data	25
4.3	Potongan Kode Labeling VADER	26
4.4	Potongan Kode <i>Case Folding</i>	27
4.5	Potongan Kode <i>Cleaning Data</i>	28
4.6	Potongan Kode <i>Tokenizing</i>	29
4.7	Potongan Kode <i>Normalisasi</i>	30
4.8	Potongan Kode <i>Stopword</i>	31
4.9	Potongan Kode Stemming	32
4.10	Potongan Kode Input Teks	32
4.11	Potongan Kode TF-IDF	33
4.12	Potongan Kode Apply SVM	33
4.13	Potongan Kode Ecaluasi Model	33
4.14	Potongan Kode Input Teks	41
4.15	Potongan Kode Input File Csv	42

UMMN
UNIVERSITAS
MULTIMEDIA
NUSANTARA

DAFTAR LAMPIRAN

Lampiran 1. Form Bimbingan	45
Lampiran 2. Hasil Pengecekan Laporan di Turnitin	47
Lampiran 3. Dataset	56



UMMN
UNIVERSITAS
MULTIMEDIA
NUSANTARA