

**ANALISIS SENTIMEN KOMENTAR YOUTUBE TENTANG  
PERFORMA MANCHESTER UNITED DI PARUH AWAL LIGA  
INGGRIS MUSIM 2022-2023 MENGGUNAKAN METODE  
N-GRAM DAN SUPPORT VECTOR MACHINE**

**SKRIPSI**

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



disusun oleh

**ARIALZA TERA AHMAD**

**19.11.3215**

Kepada

**FAKULTAS ILMU KOMPUTER  
UNIVERSITAS AMIKOM YOGYAKARTA  
YOGYAKARTA**

**2023**

**ANALISIS SENTIMEN KOMENTAR YOUTUBE TENTANG  
PERFORMA MANCHESTER UNITED DI PARUH AWAL LIGA  
INGGRIS MUSIM 2022-2023 MENGGUNAKAN METODE  
N-GRAM DAN SUPPORT VECTOR MACHINE**

**SKRIPSI**

untuk memenuhi salah satu syarat mencapai derajat Sarjana  
Program Studi Informatika



disusun oleh

**ARIALZA TERA AHMAD**

**19.11.3215**

Kepada

**FAKULTAS ILMU KOMPUTER**

**UNIVERSITAS AMIKOM YOGYAKARTA**

**YOGYAKARTA**

**2023**

**HALAMAN PERSETUJUAN**

**SKRIPSI**

**ANALISIS SENTIMEN KOMENTAR YOUTUBE TENTANG PERFORMA  
MANCHESTER UNITED DI PARUH AWAL LIGA INGGRIS MUSIM  
2022-2023 MENGGUNAKAN METODE N-GRAM DAN SUPPORT  
VECTOR MACHINE**

yang disusun dan diajukan oleh

**Arialza Tera Ahmad**

**19.11.3215**

telah disetujui oleh Dosen Pembimbing Skripsi  
pada tanggal 18 Agustus 2023

**Dosen Pembimbing,**



**Theopilus Bayu Sasongko, S.Kom, M.Eng**

**NIK. 190302375**

**HALAMAN PENGESAHAN**

**SKRIPSI**

**ANALISIS SENTIMEN KOMENTAR YOUTUBE TENTANG PERFORMA  
MANCHESTER UNITED DI PARUH AWAL LIGA INGGRIS MUSIM  
2022-2023 MENGGUNAKAN METODE N-GRAM DAN SUPPORT  
VECTOR MACHINE**

yang disusun dan diajukan oleh

**Arialza Tera Ahmad**

**19.11.3215**

Telah dipertahankan di depan Dewan Penguji  
pada tanggal 18 Agustus 2023

**Susunan Dewan Penguji**

**Nama Penguji**

**Anna Baita, M.Kom**

**NIK. 190302290**

**Mulia Sulistiyono, M.Kom**

**NIK. 190302248**

**Theopilus Bayu Sasongko, S.Kom, M.Eng**

**NIK. 190302375**

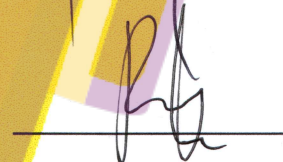
**Tanda Tangan**



---



---



---

Skripsi ini telah diterima sebagai salah satu persyaratan  
untuk memperoleh gelar Sarjana Komputer  
Tanggal 18 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 : Arialza Tera Ahmad**  
**NIM : 19.11.3215**

Menyatakan bahwa Skripsi dengan judul berikut:

**Analisis Sentimen Komentar Youtube tentang Performa Manchester United di Paruh Awal Liga Inggris Musim 2022-2023 menggunakan Metode N-Gram dan Support Vector Machine**

Dosen Pembimbing : Theopilus Bayu Sasongko, S.Kom, M.Eng

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, 18 Agustus 2023

Yang Menyatakan,



Arialza Tera Ahmad



## HALAMAN PERSEMBAHAN

Alhamdulillah, skripsi ini telah saya selesaikan dengan baik dan maksimal. Hal ini tentunya tak lepas dari karunia, rahmat serta hidayah yang telah diberikan oleh Allah swt. Sehingga saya mendapat kemudahan dan kelancaran dalam mengerjakan skripsi ini. Selain itu ada orang-orang hebat yang selalu memberikan dukungan, motivasi, dan bantuan baik secara langsung maupun tidak langsung, antar lain.

1. Kedua orang tua saya, Bapak Ali Sulthoni dan Ibu Azizah yang senantiasa mendoakan setiap langkah serta berbagai macam dukungan untuk mengantarkan saya menuju kesuksesan
2. Seluruh keluarga besar yang senantiasa mendoakan, mendukung, dan menantikan kelulusan saya.
3. Dosen pembimbing saya, Bapak Theopilus Bayu Sasongko, S.Kom, M.Eng yang telah membimbing saya dalam proses pengerjaan skripsi ini.
4. Teman-Teman mahasiswa Informatika yang tidak bisa saya sebutkan namanya satu-persatu yang telah membantu dan men-support baik secara langsung maupun tidak langsung

## KATA PENGANTAR

Puji dan syukur kami panjatkan kepada Tuhan Yang Maha Esa, karena atas berkat dan rahmat-Nya, penulis dapat menyelesaikan skripsi yang berjudul "Analisis Sentimen Komentar Youtube tentang Performa Manchester United di Paruh Awal Liga Inggris Musim 2022-2023 menggunakan Metode N-Gram dan Support Vector Machine" ini. Skripsi ini disusun dalam rangka memenuhi salah satu syarat untuk menyelesaikan program Sarjana (S1) Ilmu Komputer di Universitas Amikom Yogyakarta.

Selama penelitian ini, penulis mengucapkan syukur atas bantuan dari banyak pihak sehingga skripsi ini dapat terselesaikan dengan baik. Oleh karena itu, penulis mengucapkan terimakasih kepada :

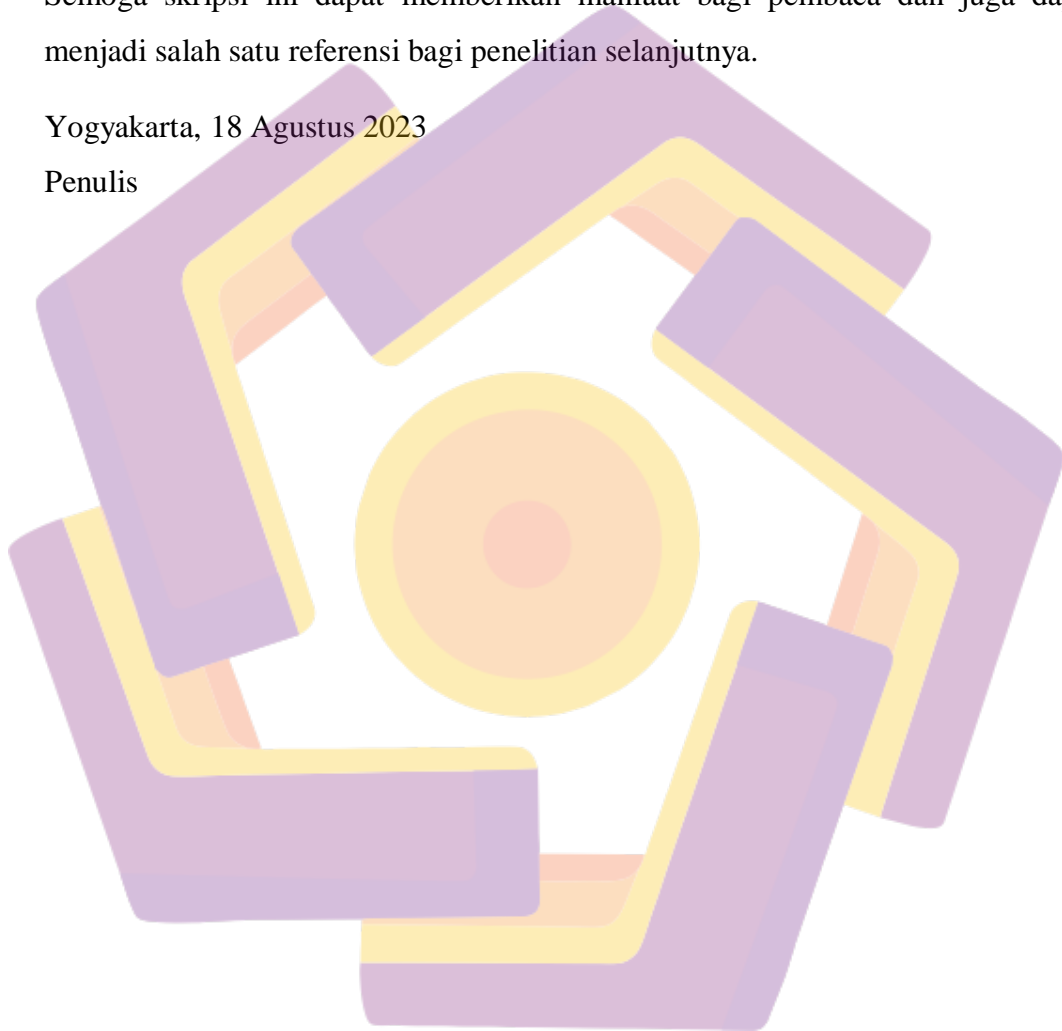
1. Bapak Prof. Dr. M. Suyanto, M.M selaku Rektor Universitas AMIKOM Yogyakarta.
2. Theopilus Bayu Sasongko, S.Kom, M.Eng sebagai dosen pembimbing yang telah memberikan bimbingan, masukan, dan waktunya selama penyusunan skripsi ini.
3. Hanif Al-Fatta, M.Kom selaku dekan Fakultas Ilmu Komputer.
4. Windha Mega Pradnya Duhita, M.Kom selaku Ketua Program Studi Informatika.
5. Bapak dan Ibu dosen yang telah memberikan ilmu dan pengetahuan yang sangat berguna bagi penulis selama perkuliahan.
6. Kedua orang tua saya, Bapak Ali Sulthoni dan Ibu Azizah yang senantiasa mendoakan setiap langkah.
7. Coach Justinus Lhaksana, yang telah memberikan edukasi terkait perkembangan sepakbola di Eropa & pandangannya tentang Manchester United melalui konten Youtube.
8. Seluruh teman-teman angkatan 2019, terutama untuk kelas IF 10, terimakasih atas kerjasamanya selama ini.

9. Teman-teman seperjuangan (Ivanda, Budi, Aqsal, Fajar, Sofyan, Nando, Farhan, dan Samudra) yang telah memberikan dukungan dan terimakasih atas momen-momen yang tak pernah terlupakan.

Penulis menyadari bahwa skripsi ini masih jauh dari sempurna, oleh karena itu, penulis sangat mengharapkan saran dan kritik yang membangun dari pembaca. Semoga skripsi ini dapat memberikan manfaat bagi pembaca dan juga dapat menjadi salah satu referensi bagi penelitian selanjutnya.

Yogyakarta, 18 Agustus 2023

Penulis

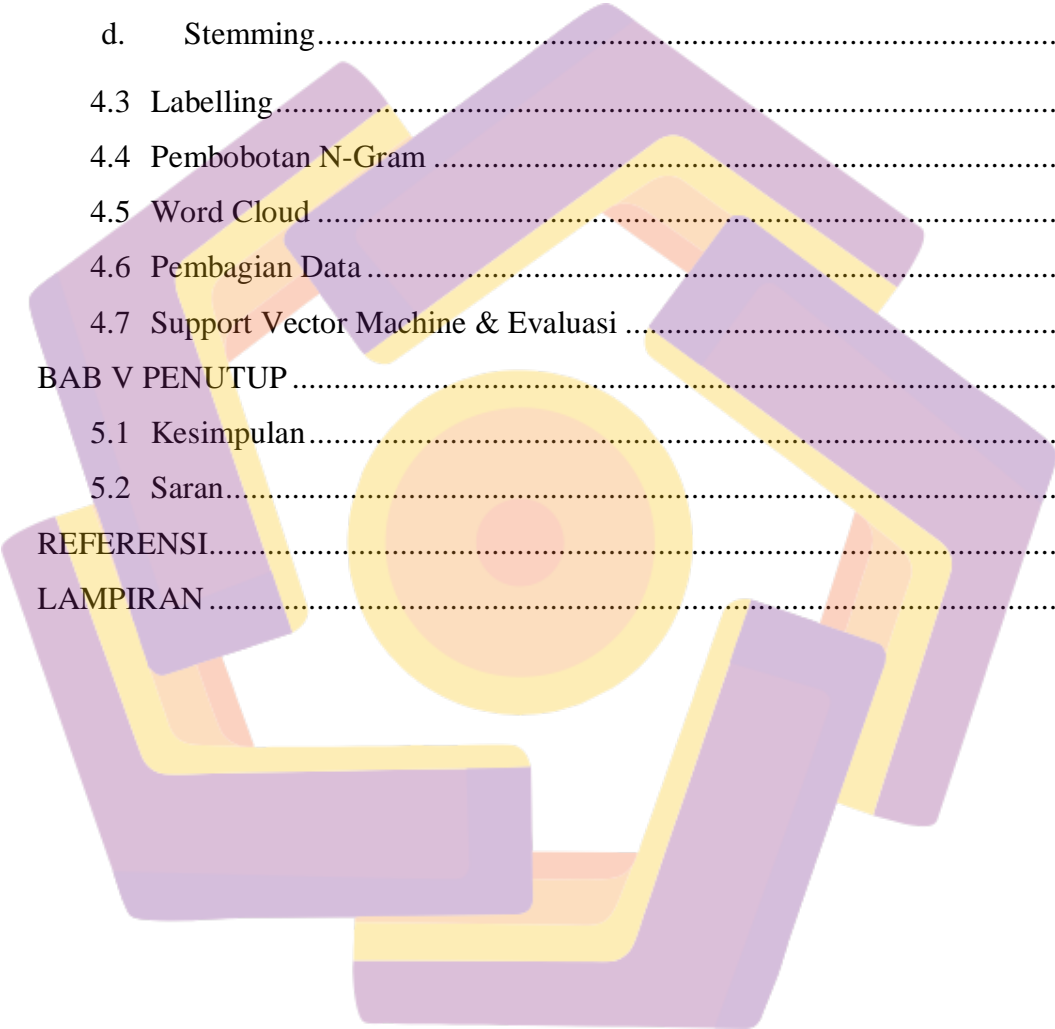




## DAFTAR ISI

HALAMAN JUDUL .....	i
HALAMAN PERSETUJUAN .....	ii
HALAMAN PENGESAHAN .....	iii
HALAMAN PERNYATAAN KEASLIAN SKRIPSI .....	iv
HALAMAN PERSEMBAHAN .....	v
KATA PENGANTAR .....	vi
DAFTAR ISI .....	viii
DAFTAR TABEL .....	xi
DAFTAR GAMBAR .....	xii
DAFTAR LAMPIRAN .....	xiv
DAFTAR LAMBANG DAN SINGKATAN .....	xv
DAFTAR ISTILAH .....	xvi
INTISARI .....	xviii
ABSTRACT .....	xix
BAB I PENDAHULUAN .....	20
1.1 Latar Belakang .....	20
1.2 Rumusan Masalah .....	21
1.3 Batasan Masalah .....	21
1.4 Tujuan Penelitian .....	22
1.5 Manfaat Penelitian .....	22
1.6 Sistematika Penulisan .....	22
BAB II TINJAUAN PUSTAKA .....	24
2.1 Studi Literatur .....	24
2.2 Dasar Teori .....	33
2.2.1 Data Mining .....	33
2.2.2 Natural Language Processing (NLP) .....	34
2.2.3 Analisis Sentimen .....	34

2.2.4	Pelabelan Kata.....	35
2.2.5	TextBlob .....	36
2.2.6	Confusion Matrix .....	36
2.2.7	Preprocessing .....	38
2.2.8	N-Gram.....	38
2.2.9	TF-IDF.....	39
2.2.10	WordCloud.....	39
2.2.11	Support Vector Machine .....	40
<b>BAB III METODE PENELITIAN.....</b>		<b>42</b>
3.1	Objek Penelitian.....	42
3.2	Alur Penelitian .....	43
3.2.1	Pengumpulan Data (Data Collection).....	44
3.2.2	Preprocessing .....	45
3.2.2.1	Case Folding .....	45
3.2.2.2	Tokenisasi .....	46
3.2.2.3	Stopwords .....	47
3.2.2.4	Stemming .....	48
3.2.3	Labelling .....	49
3.2.4	Pembobotan N-Gram.....	50
3.2.5	Pembagian Data.....	50
3.2.6	Support Vector Machine .....	51
3.2.7	Evaluasi.....	51
3.3	Alat dan Bahan.....	51
3.3.1	Alat .....	51
3.3.2	Bahan .....	52
<b>BAB IV</b>	<b>HASIL DAN PEMBAHASAN .....</b>	<b>54</b>



4.1 Pengumpulan Data (Data Collection) .....	54
4.2 Preprocessing .....	56
a. Case Folding.....	56
b. Tokenisasi.....	56
c. Stopwords.....	57
d. Stemming.....	58
4.3 Labelling.....	60
4.4 Pembobotan N-Gram .....	62
4.5 Word Cloud .....	64
4.6 Pembagian Data .....	66
4.7 Support Vector Machine & Evaluasi .....	70
<b>BAB V PENUTUP .....</b>	<b>75</b>
5.1 Kesimpulan.....	77
5.2 Saran.....	77
<b>REFERENSI.....</b>	<b>79</b>
<b>LAMPIRAN .....</b>	<b>84</b>

## DAFTAR TABEL

Tabel 2. 1 Tabel Confusion Matrix .....	36
Tabel 3. 1 Video pertandingan untuk analisis sentimen .....	52
Tabel 4. 1 Contoh hasil <i>crawling</i> komentar Youtube.....	55
Tabel 4. 2 Hasil Pengurangan Data .....	55
Tabel 4. 3 Contoh Case Folding.....	56
Tabel 4. 4 Data Tokenisasi.....	57
Tabel 4. 5 Contoh Stopwords Bahasa Indonesia.....	58
Tabel 4. 6 Contoh Stopwords Bahasa Inggris .....	58
Tabel 4. 7 Hasil Stemming .....	59
Tabel 4. 8 Hasil Penghapusan Tanda Baca pada bahasa Indonesia .....	59
Tabel 4. 9 Hasil Penghapusan Tanda Baca pada bahasa Inggris.....	60
Tabel 4. 10 Hasil Labelling Data dengan Lexicon Resource.....	60
Tabel 4. 11 Hasil Labelling Data dengan TextBlob .....	60
Tabel 4. 12 Perbandingan kategori label di setiap video bahasa Indonesia.....	61
Tabel 4. 13 Perbandingan kategori label di setiap video bahasa Inggris .....	62
Tabel 4. 14 Jumlah Data Latih & Data Uji .....	66
Tabel 4. 15 Hasil Accuracy, Precision, Recall, f1_score dari video yang memiliki komentar Bahasa Indonesia .....	75
Tabel 4. 16 Hasil Accuracy, Precision, Recall, f1_score dari video yang memiliki komentar Bahasa Inggris.....	75

## DAFTAR GAMBAR

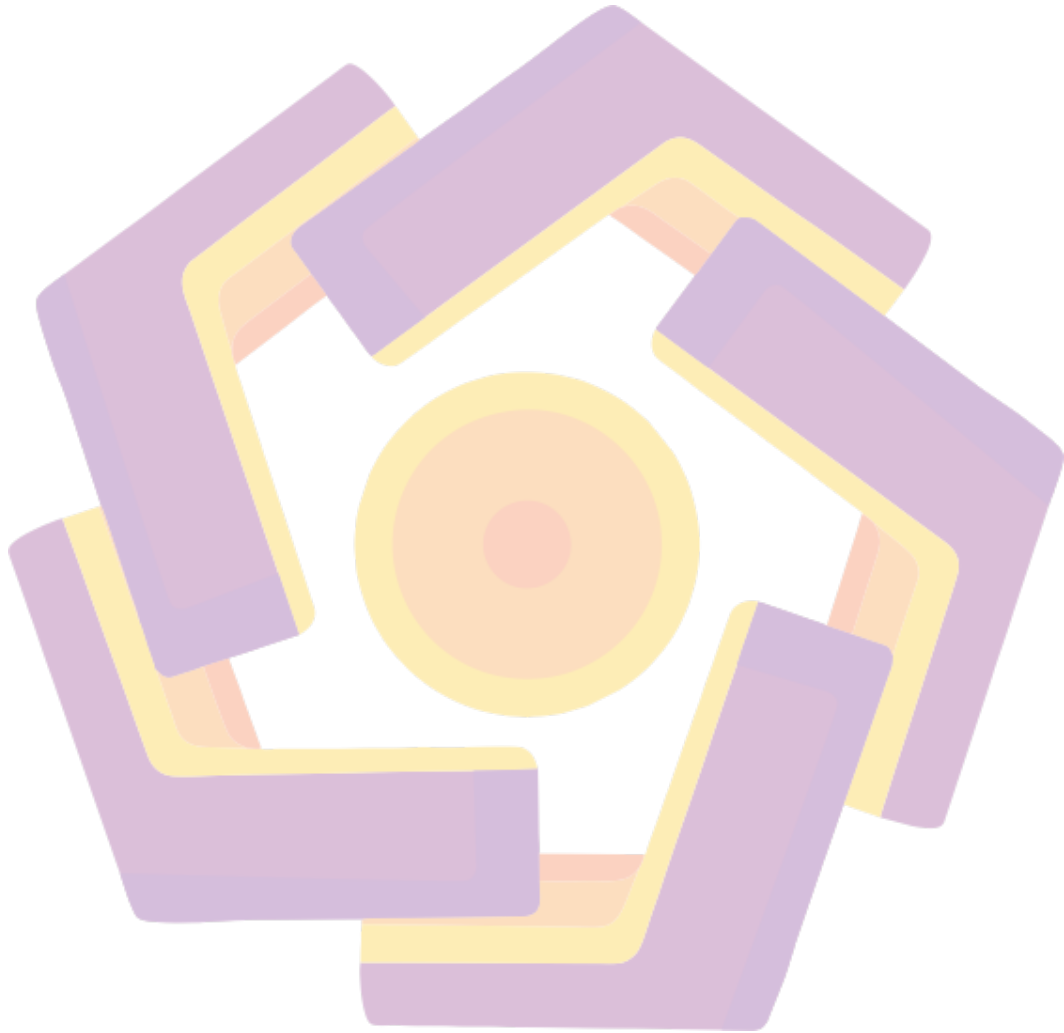
Gambar 3. 1 Halaman <i>Channel</i> Youtube Manchester United.....	42
Gambar 3. 2 Halaman <i>Channel</i> Youtube Vidio .....	42
Gambar 3. 3 Daftar video yang digunakan .....	43
Gambar 3. 4 Alur Penelitian.....	44
Gambar 3. 5 Alur Pengumpulan Data.....	44
Gambar 3. 6 Alur Case Folding.....	45
Gambar 3. 7 Tokenisasi .....	46
Gambar 3. 8 Alur Filtering.....	47
Gambar 3. 9 Alur Stemming .....	48
Gambar 3. 10 Alur Labelling .....	49
Gambar 3. 11 Alur Pembobotan N-Gram .....	50
Gambar 4. 1 Contoh data komentar Youtube.....	54
Gambar 4. 2 Hasil Sampel Unigram.....	63
Gambar 4. 3 Hasil Sampel Bigram.....	63
Gambar 4. 4 Hasil Sampel Trigram.....	64
Gambar 4. 5 Contoh Word Cloud Manchester United vs Brighton.....	65
Gambar 4. 6 Contoh Word Cloud Manchester United vs Liverpool.....	65
Gambar 4. 7 Jumlah Data Manchester United vs Brighton sebelum proses SMOTE.....	67
Gambar 4. 8 Jumlah Data Manchester United vs Brighton sesudah proses SMOTE .....	67
Gambar 4. 9 Jumlah Data Manchester United vs Liverpool sebelum proses SMOTE.....	67
Gambar 4. 10 Jumlah Data Manchester United vs Liverpool sesudah proses SMOTE.....	68
Gambar 4. 11 Jumlah Data Manchester United vs Arsenal sebelum proses SMOTE.....	68
Gambar 4. 12 Jumlah Data Manchester United vs Arsenal sesudah proses SMOTE .....	68

Gambar 4. 13 Jumlah Data Manchester United vs Chelsea sebelum proses SMOTE.....	69
Gambar 4. 14 Jumlah Data Manchester United vs Chelsea sesudah proses SMOTE .....	69
Gambar 4. 15 Jumlah Data Manchester United vs Manchester City sebelum proses SMOTE.....	69
Gambar 4. 16 Jumlah Data Manchester United vs Manchester City sesudah proses SMOTE.....	70
Gambar 4. 17 Hasil Manchester United vs Brighton tanpa SMOTE .....	70
Gambar 4. 18 Hasil Manchester United vs Brighton dengan SMOTE.....	71
Gambar 4. 19 Hasil Manchester United vs Liverpool tanpa SMOTE .....	71
Gambar 4. 20 Hasil Manchester United vs Liverpool dengan SMOTE .....	72
Gambar 4. 21 Hasil Manchester United vs Arsenal tanpa SMOTE .....	72
Gambar 4. 22 Hasil Manchester United vs Arsenal dengan SMOTE .....	73
Gambar 4. 23 Hasil Manchester United vs Chelsea Tanpa SMOTE.....	73
Gambar 4. 24 Hasil Manchester United vs Chelsea dengan SMOTE .....	74
Gambar 4. 25 Hasil Manchester United vs Manchester City tanpa SMOTE.....	74
Gambar 4. 26 Hasil Manchester United vs Manchester City dengan SMOTE.....	75



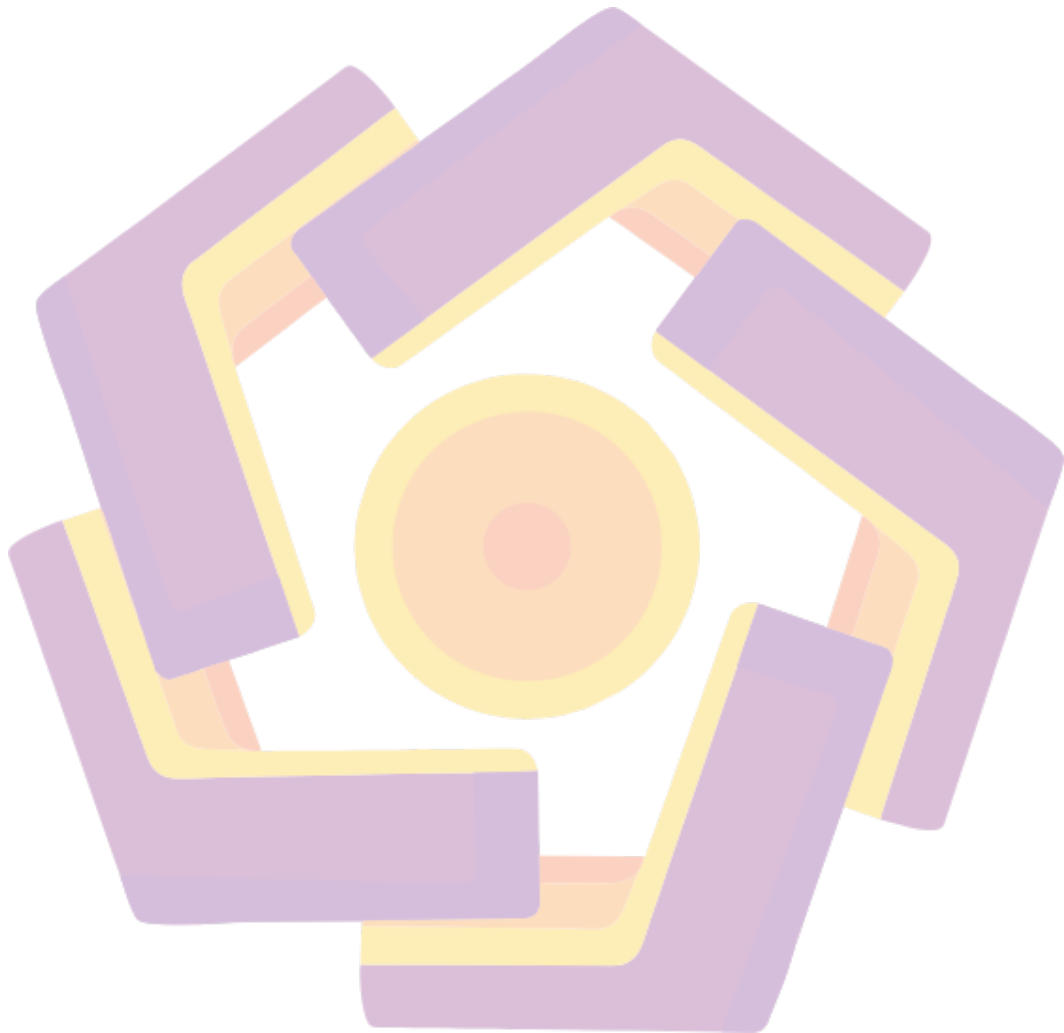
## DAFTAR LAMPIRAN

Lampiran 1. Source Code.....	84
Lampiran 2. Sample Dataset .....	98



## DAFTAR LAMBANG DAN SINGKATAN

SVM	Support Vector Machines
AUC	Area Under the ROC Curve
FA	Football Association

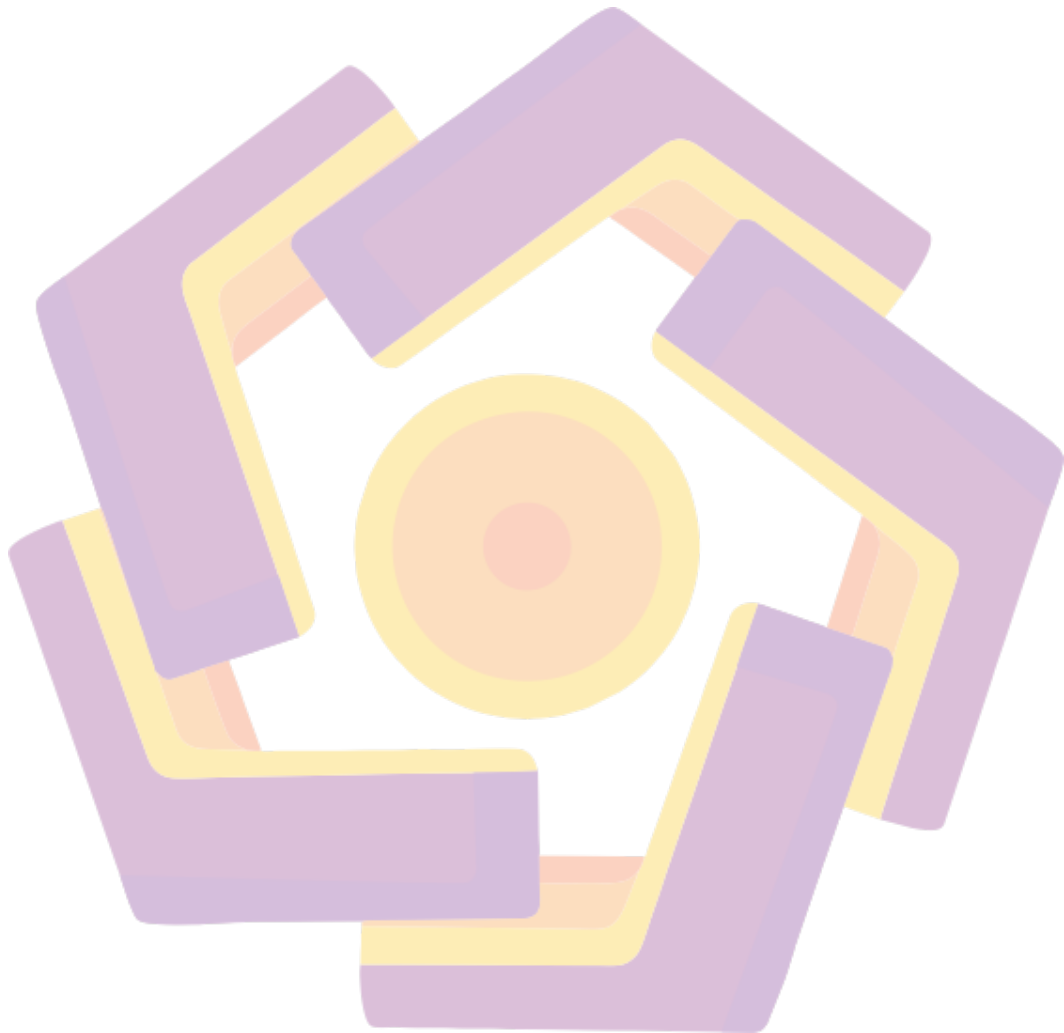


## DAFTAR ISTILAH

Klub	Perkumpulan yang kegiatannya mengadakan persekutuan untuk maksud tertentu
Klasemen	Kedudukan suatu tim/kelompok dalam mengikuti ajang perlombaan/pertandingan
Streaming	Teknologi pengiriman data, video atau audio dalam bentuk yang telah dikompresi melalui jaringan internet yang ditampilkan oleh suatu player secara realtime.
Platform	Wadah yang dipakai untuk menjalankan sebuah sistem sesuai dengan rencana program yang telah dibuat.
Cuplikan	Sebagian dari keseluruhan.
Model	Struktur matematis yang digunakan untuk mengklasifikasikan teks atau data menjadi kategori yang sesuai.
Fitur	Karakteristik khusus yang terdapat pada suatu alat.
Matchweek	Pertandingan yang dimainkan oleh suatu tim di setiap pekan.
Klasifikasi	Proses menempatkan teks atau data ke dalam kategori yang sesuai berdasarkan model yang digunakan.
Pelatihan	Proses mengoptimalkan model dengan memberikan data latih yang dilabel.
Data Latih	Data yang digunakan untuk melatih model, terdiri dari input (teks) dan label (kategori yang sesuai).
Data Uji	Data yang digunakan untuk menguji kemampuan model dalam menangkap pola dalam data teks.
<i>Lexicon resource</i>	Kumpulan kata atau frasa yang digunakan sebagai dasar untuk melakukan pelabelan otomatis.
Akurasi	Tingkat keakuratan model dalam mengklasifikasikan teks atau data ke dalam kategori yang sesuai.

Taktik

Suatu siasat atau akal yang dirancang dan akan dilaksanakan dalam permainan perorangan, kelompok, maupun tim untuk memenangkan suatu pertandingan secara sportif.



## INTISARI

Manchester United merupakan salah satu klub sepak bola yang sukses dengan sejarah prestasi yang mengesankan di tingkat domestik maupun internasional. Dalam era digital seperti sekarang, komentar tentang pertandingan banyak diungkapkan oleh penggemar klub. Untuk mengatasi masalah tersebut, teknologi yang digunakan adalah analisis sentimen, yang merupakan proses analisis yang dilakukan pada pendapat, sentimen, penilaian, sikap, dan emosi orang terhadap suatu entitas yang diungkap berupa teks, baik positif, negatif, atau netral.

Penelitian ini akan menganalisis opini penonton terhadap performa Manchester United di liga Inggris pada paruh awal musim 2022-2023 menggunakan metode N-Gram sebagai pembobotan kata dan metode Support Vector Machine sebagai metode klasifikasi data. Hasil dari kanal Youtube "Vidio" menggunakan metode support vector machine (SVM) dengan teknik balancing SMOTE menghasilkan rata-rata akurasi adalah 60%. Sedangkan, Hasil dari kanal Youtube "Manchester United" menggunakan metode support vector machine (SVM) dengan teknik balancing SMOTE menghasilkan rata-rata akurasi adalah 87%.

Kesimpulan penelitian ini diantaranya Akurasi Support Vector Machine meningkat pada saat melakukan pengujian menggunakan teknik balancing SMOTE analisis sentimen menggunakan klasifikasi Support Vector Machine akan efektif apabila memiliki dataset yang bersih. Perlu menambah dataset untuk video yang berasal dari kanal Youtube "Vidio". Membutuhkan filter lebih mendalam terhadap komentar spam atau komentar yang tidak berhubungan dengan sepakbola. Lebih baik menggunakan dataset bahasa Inggris dalam penelitian analisis sentimen ini.

**Kata kunci:** manchester united, support vector machine, n-gram, analisis sentimen, komentar.

## ABSTRACT

*Manchester United is a successful football club with an impressive history of achievements at both domestic and international levels. In the digital era like now, many comments about matches are expressed by club fans. To overcome this problem, the technology used is sentiment analysis, which is an analysis process carried out on people's opinions, sentiments, judgments, attitudes, and emotions towards an entity which is expressed in the form of text, whether positive, negative, or neutral.*

*This study will analyze audience opinion on Manchester United's performance in the English league in the early half of the 2022-2023 season using the N-Gram method as word weighting and the Support Vector Machine method as a data classification method. The results of the Youtube channel "Vidio" using the support vector machine (SVM) method with the SMOTE balancing technique produce an average accuracy of 60%. Meanwhile, the results from the "Manchester United" Youtube channel using the support vector machine (SVM) method with the SMOTE balancing technique produce an average accuracy of 87%.*

*The conclusion of this study is that the accuracy of the Support Vector Machine increases when testing using the SMOTE balancing technique, sentiment analysis using the Support Vector Machine classification will be effective if it has a clean dataset. Need to add a dataset for videos originating from the "Vidio" Youtube channel. Needs more in-depth filtering of spam comments or comments that are not related to football. It is better to use English dataset in this sentiment analysis research..*

**Keyword:** manchester united, support vector machine, n-gram, sentiment analysis, comment.