

**ANALISIS SENTIMEN ULASAN FILM OPPENHEIMER PADA
SITUS IMDB (INTERNET MOVIE DATABASE) MENGGUNAKAN
ALGORITMA SUPPORT VECTOR MACHINE (SVM)**

SKRIPSI

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



disusun oleh

DIDI SETYOPRIBADI

18.11.2592

Kepada

**FAKULTAS ILMU KOMPUTER
UNIVERSITAS AMIKOM YOGYAKARTA
YOGYAKARTA
2024**

**ANALISIS SENTIMEN ULASAN FILM OPPENHEIMER PADA
SITUS IMDB (INTERNET MOVIE DATABASE) MENGGUNAKAN
ALGORITMA SUPPORT VECTOR MACHINE (SVM)**

SKRIPSI

untuk memenuhi salah satu syarat mencapai derajat Sarjana
Program Studi Informatika



disusun oleh

DIDI SETYOPRIBADI

18.11.2592

Kepada

**FAKULTAS ILMU KOMPUTER
UNIVERSITAS AMIKOM YOGYAKARTA
YOGYAKARTA
2024**

HALAMAN PERSETUJUAN

SKRIPSI

**ANALISIS SENTIMEN ULASAN FILM OPPENHEIMER PADA SITUS
IMDB (INTERNET MOVIE DATABASE) MENGGUNAKAN
ALGORITMA SUPPORT VECTOR MACHINE (SVM)**

yang disusun dan diajukan oleh

Didi Setyopribadi

18.11.2592

telah disetujui oleh Dosen Pembimbing Skripsi
pada tanggal 23 Februari 2024

Dosen Pembimbing,



Kusnawi S.Kom, M. Eng.

NIK. 190302112

HALAMAN PENGESAHAN

SKRIPSI

**ANALISIS SENTIMEN ULASAN FILM OPPENHEIMER PADA SITUS
IMDB (INTERNET MOVIE DATABASE) MENGGUNAKAN
ALGORITMA SUPPORT VECTOR MACHINE (SVM)**

yang disusun dan diajukan oleh

Didi Setyopribadi

18.11.2592

Telah dipertahankan di depan Dewan Penguji
pada tanggal 23 Februari 2024

Susunan Dewan Penguji

Nama Penguji

Tanda Tangan

Kusnawi, S.Kom, M.Eng
NIK. 190302112

Ali Mustopa, M.Kom
NIK. 190302192

Robert Marco, M.T., Ph.D
NIK. 190302228



Skripsi ini telah diterima sebagai salah satu persyaratan
untuk memperoleh gelar Sarjana Komputer
Tanggal 23 Februari 2024

DEKAN FAKULTAS ILMU KOMPUTER



Hanif Al Fatta, M.Kom., Ph.D.
NIK. 190302096

HALAMAN PERNYATAAN KEASLIAN SKRIPSI

Yang bertandatangan di bawah ini,

Nama mahasiswa : Didi Setyopribadi
NIM : 18.11.2592

Menyatakan bahwa Skripsi dengan judul berikut:

**ANALISIS SENTIMEN ULASAN FILM OPPENHEIMER PADA SITUS
IMDB (INTERNET MOVIE DATABASE) MENGGUNAKAN
ALGORITMA SUPPORT VECTOR MACHINE (SVM)**

Dosen Pembimbing: Kusnawi, 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, 23 Februari 2024

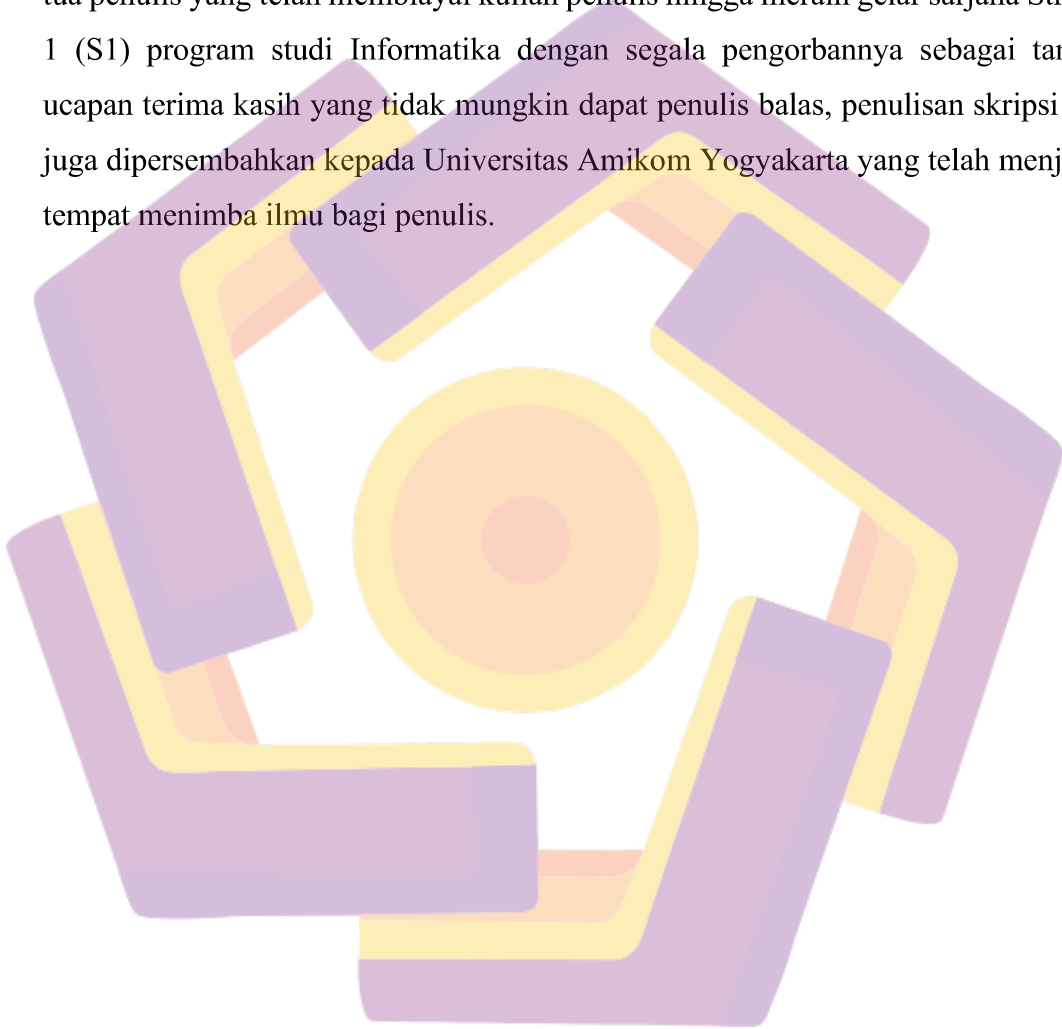
Yang Menyatakan,



Didi Setyopribadi

HALAMAN PERSEMBAHAN

Puji syukur penulis haturkan kepada Allah SWT atas segala rahmat, karunia dan hidayah-Nya yang telah memberikan kelancaran kepada penulis dalam menyelesaikan skripsi. Penulisan skripsi ini dipersembahkan kepada kedua orang tua penulis yang telah membiayai kuliah penulis hingga meraih gelar sarjana Strata 1 (S1) program studi Informatika dengan segala pengorbannya sebagai tanda ucapan terima kasih yang tidak mungkin dapat penulis balas, penulisan skripsi ini juga dipersembahkan kepada Universitas Amikom Yogyakarta yang telah menjadi tempat menimba ilmu bagi penulis.



KATA PENGANTAR

Puji syukur penulis haturkan kepada Allah SWT atas segala rahmat, karunia dan hidayah-Nya sehingga penulis dapat menyelesaikan skripsi dengan judul **“Analisis Sentimen Ulasan Film Oppenheimer Pada Situs IMDB (Internet Movie Database) Menggunakan Algoritma Support Vector Machine (SVM)”** dengan lancar sebagai salah satu syarat untuk meraih gelar sarjana Strata 1 (S1) program studi Informatika pada Fakultas Ilmu Komputer di Universitas Amikom Yogyakarta, penulisan skripsi ini juga diharapkan dapat memberikan manfaat bagi para pembaca atau peneliti dalam melakukan penelitian.

Dalam proses pembuatan skripsi ini penulis mengucapkan terima kasih kepada pelbagai pihak yang telah memberikan doa, dukungan, bimbingan, saran, arahan dan bantuan; kepada:

1. Kedua orang tua dan adik tercinta di provinsi Kalimantan Selatan, khususnya Ibu dan Bapak yang telah memberikan segalanya kepada penulis.
2. Keluarga besar Ibu di Bantul; Mbah Putri, Mbah Kakung, Pakdhe, Budhe, Paklik, Bulik, saudara dan keponakan yang telah memberikan doa dan dukungan kepada penulis.
3. Bapak Kusnawi, S.Kom., M.Eng. selaku dosen pembimbing dan penguji skripsi yang telah memberikan bimbingan, saran, arahan dan pelbagai ilmu kepada penulis selama proses pembuatan skripsi.
4. Muhammad Benz Junior Hartono, Muhammad Lubis Notangsi dan Alfian Ady Chandra yang telah menemani dan membantu penulis dalam proses pembuatan skripsi.
5. Pakdhe Subar dan keluarga besar selaku pemilik kos yang telah memberikan doa, dukungan dan bantuan kepada penulis.
6. Semua teman-teman di Asrama Mahasiswa Saijaan Kotabaru Kalimantan Selatan Yogyakarta yang telah memberikan dukungan dan bantuan kepada penulis.

7. Semua teman-teman perkuliahan di Universitas Amikom Yogyakarta yang telah berbagi pelbagai ilmu.
8. Bapak Ali Mustopa, M.Kom dan Bapak Robert Marco, M.T., Ph.D selaku dosen penguji skripsi.
9. Bapak Prof. Dr. M. Suyanto, M.M selaku rektor Universitas Amikom Yogyakarta.
10. Bapak Hanif Al Fatta, S.Kom., M.Kom., Ph.D selaku Dekan Fakultas Ilmu Komputer di Universitas Amikom Yogyakarta.
11. Ibu Windha Mega Pradnya Dhuhita, M.Kom selaku Kaprodi Informatika pada Fakultas Ilmu Komputer di Universitas Amikom Yogyakarta.
12. Ibu Rifda Faticha Alfa Aziza, S.Kom., M.Kom selaku dosen wali kelas 11 program studi Informatika angkatan 2018.
13. Semua pihak yang telah membantu dalam proses pembuatan skripsi yang tidak bisa penulis sebutkan satu per satu.

Akhir kata, penulis memohon maaf jika penulisan skripsi ini masih banyak kekurangan dan penulis berharap semoga kekurangan pada penelitian ini dapat diperbaiki di penelitian selanjutnya oleh peneliti yang lain.

Yogyakarta, 23 Februari 2024

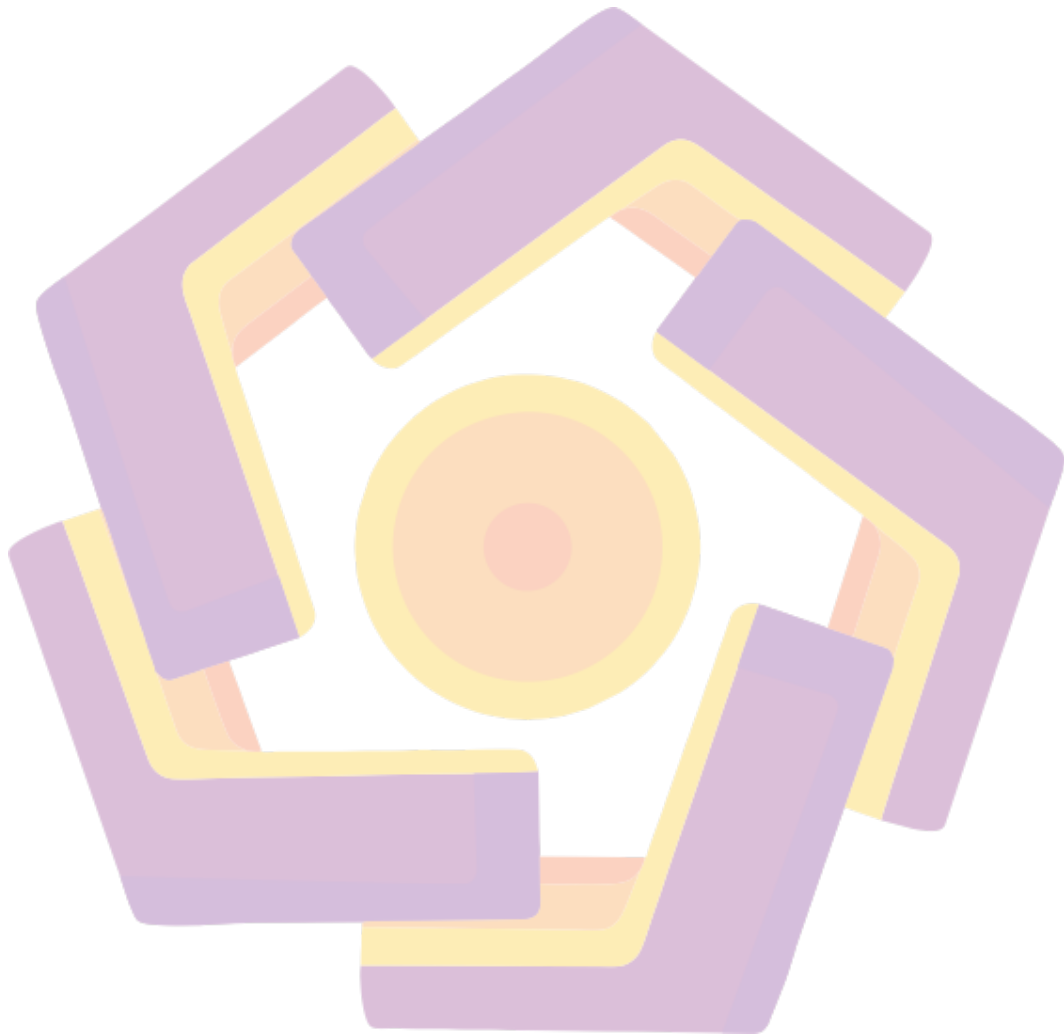
Penulis

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
INTISARI	xv
ABSTRACT.....	xvi
BAB I PENDAHULUAN	1
1.1 Latar Belakang Masalah.....	1
1.2 Rumusan Masalah.....	1
1.3 Batasan Masalah.....	2
1.4 Tujuan Penelitian	2
1.5 Manfaat Penelitian	3
1.6 Sistematika Penulisan	3
BAB II TINJAUAN PUSTAKA	5
2.1 Studi Literatur	5
2.2 Data Mining.....	9
2.3 Text Mining	10
2.4 Text Preprocessing.....	12
2.5 TF-IDF (Term Frequency-Inverse Document Frequency).....	13
2.6 Support Vector Machine (SVM)	15
2.7 K-Fold Cross Validation	20
2.8 Confusion Matrix	22
BAB III METODE PENELITIAN.....	26
3.1 Alat dan Bahan.....	26
3.1.1 Bahan	26
3.1.2 Alat.....	27

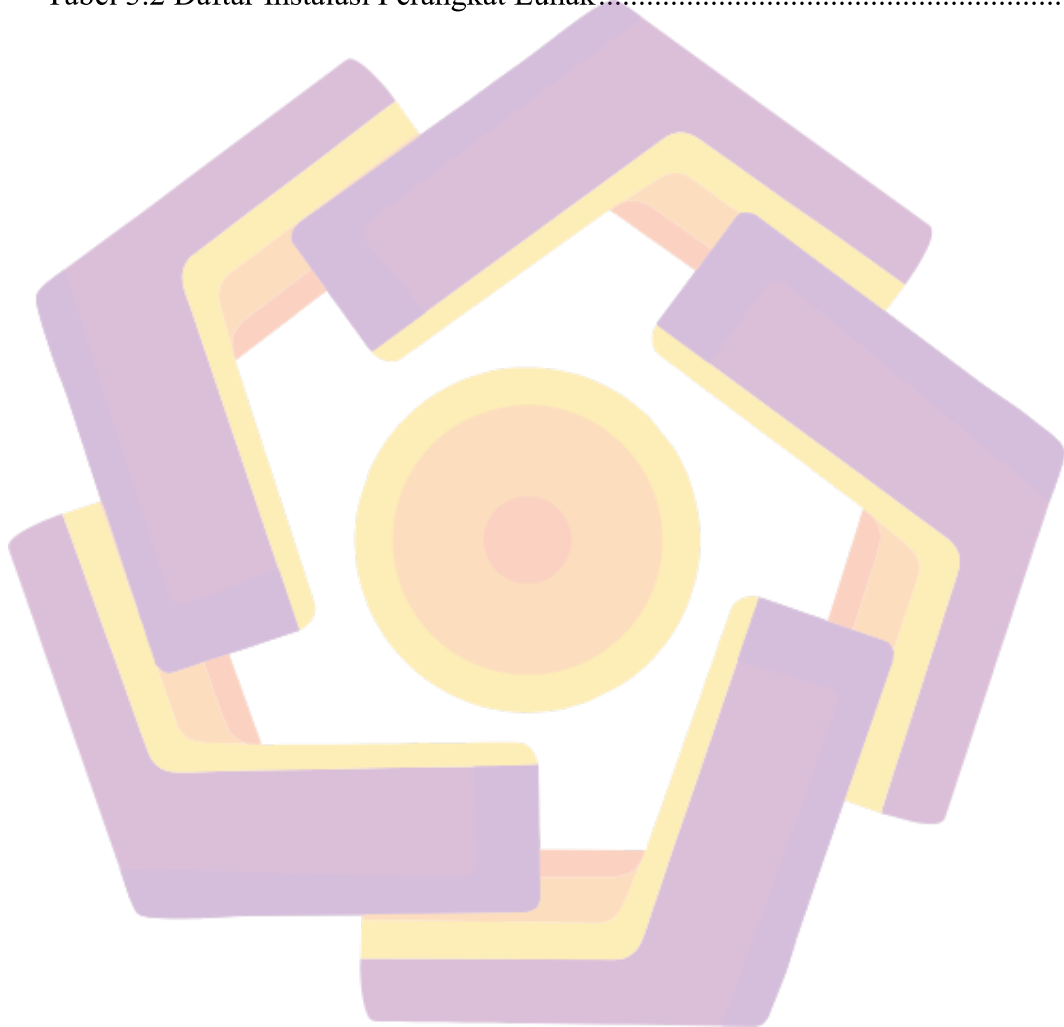
3.2	IMDB (Internet Movie Database).....	29
3.3	Alur Penelitian.....	32
3.3.1	Identifikasi Masalah	32
3.3.2	Studi Literatur	34
3.3.3	Analisa Kebutuhan Penelitian	35
3.3.4	Tahap Persiapan Proses Text Preprocessing.....	35
3.3.5	Text Preprocessing.....	37
3.3.6	WordCloud	40
3.3.7	TF-IDF (Term Frequency-Inverse Document Frequency).....	41
3.3.8	Tahap Persiapan Proses Pembuatan Model Algoritma Support Vector Machine (SVM)	43
3.3.9	Pembuatan Model Algoritma Support Vector Machine (SVM)	44
3.3.10	Implementasi Model Algoritma Support Vector Machine (SVM) Dalam Melakukan Proses Klasifikasi	44
3.3.11	Menguji & Mengevaluasi Model Algoritma Support Vector Machine (SVM)	45
BAB IV	HASIL DAN PEMBAHASAN.....	49
4.1	Pengumpulan Data	49
4.2	Text Preprocessing.....	49
4.2.1	Input Dataset	49
4.2.2	Lowercase	50
4.2.3	Fixes Contraction	51
4.2.4	Cleaning	52
4.2.5	Remove Stopword	53
4.2.6	Lemmatization	54
4.2.7	Filtering	56
4.3	WordCloud	57
4.4	TF-IDF (Term Frequency-Inverse Document Frequency).....	59
4.5	Support Vector Machine (SVM)	62
4.6	Menguji & Mengevaluasi	65
4.6.1	K-Fold Cross Validation	65
4.6.2	Confusion Matrix.....	67
BAB V	PENUTUP	70
5.1	Kesimpulan.....	70

5.2	Saran	70
DAFTAR PUSTAKA		72



DAFTAR TABEL

Tabel 2.1 Keaslian Penelitian	7
Tabel 2.2 Struktur Confusion Matrix	23
Tabel 3.1 Daftar Perangkat Lunak	28
Tabel 3.2 Daftar Instalasi Perangkat Lunak	36



DAFTAR GAMBAR

Gambar 2.1 Proses <i>data mining</i> . (Jiawei Han, Jian Pei, Hanghang Tong, 2023). ...	9
Gambar 2.2 Proses <i>text classification</i>	11
Gambar 2.3 Proses <i>text classification</i> pada dokumen. (Chengqing Zong, Rui Xia, Jiajun Zhang, 2021).	11
Gambar 2.4 Proses <i>text clustering</i>	12
Gambar 2.5 Proses <i>text preprocessing</i>	12
Gambar 2.6 (a) Proses menemukan <i>hyperplane</i> , (b) Contoh <i>hyperplane</i> terbaik. (Anto Satriyo Nugroho, Arief Budi Witarto, Dwi Handoko, 2003).	16
Gambar 2.7 Proses transformasi dari <i>input space</i> ke <i>feature space</i> pada non-linear <i>Support Vector Machine</i> (SVM). (Krisantus Sembiring, 2007).	19
Gambar 2.8 Proses <i>K-Fold Cross Validation</i> menggunakan 500 dataset training dan 10-fold.	22
Gambar 3.1 Tampilan halaman situs Kaggle yang berisi dataset ulasan film Oppenheimer pada situs IMDB (<i>Internet Movie Database</i>).	26
Gambar 3.2 Tampilan aplikasi Microsoft Excel yang berisi dataset ulasan film Oppenheimer pada situs IMDB (<i>Internet Movie Database</i>).	27
Gambar 3.3 Tampilan halaman utama pada situs IMDB (<i>Internet Movie Database</i>).	30
Gambar 3.4 Tampilan halaman film Oppenheimer pada situs IMDB (<i>Internet Movie Database</i>) yang berisi pelbagai informasi.	30
Gambar 3.5 Tampilan halaman rating film Oppenheimer pada situs IMDB (<i>Internet Movie Database</i>).	31
Gambar 3.6 Tampilan halaman ulasan pengguna film Oppenheimer pada situs IMDB (<i>Internet Movie Database</i>).	31
Gambar 3.7 Alur Penelitian.	32
Gambar 3.8 Tampilan halaman Google Trends Worldwide film Oppenheimer pada bulan Juli-September 2023.	33
Gambar 3.9 Tampilan halaman Google's Year in Search Global pada tahun 2023.	34

Gambar 3.10 Tampilan aplikasi Microsoft Excel yang berisi dataset ulasan film Oppenheimer pada situs IMDB (<i>Internet Movie Database</i>) yang sudah diberi label sentimen positif dan negatif secara manual.	35
Gambar 3.11 Tampilan isi dataframe Pandas.	37
Gambar 3.12 Tampilan hasil proses <i>lowercase</i> kepada kolom Text.....	38
Gambar 3.13 Tampilan hasil proses <i>fixes contraction</i> kepada kolom Lowercase. 38	
Gambar 3.14 Tampilan hasil proses <i>cleaning</i> kepada kolom Contraction.	38
Gambar 3.15 Tampilan hasil proses <i>remove stopword</i> kepada kolom Cleaning. ...	39
Gambar 3.16 Tampilan hasil proses <i>lemmatization</i> kepada kolom Stopword.....	39
Gambar 3.17 Tampilan hasil proses <i>filtering</i> kepada kolom Lemmatize.	39
Gambar 3.18 Tampilan hasil proses pembuatan kolom Preprocessing.	40
Gambar 3.19 Tampilan hasil proses visualisasi WordCloud pada kolom Preprocessing yang memiliki label sentimen positif pada kolom Sentiment.....	40
Gambar 3.20 Tampilan hasil proses visualisasi WordCloud pada kolom Preprocessing yang memiliki label sentimen negatif pada kolom Sentiment.....	41
Gambar 3.21 Tampilan hasil proses TF-IDF pada kolom Preprocessing.	42
Gambar 3.22 Tampilan hasil proses IDF pada kolom Preprocessing.	43
Gambar 3.23 Tampilan hasil proses TF-IDF pada kolom Preprocessing index ke-0.....	43
Gambar 3.24 Tampilan hasil proses pembagian data pada kolom Sentiment dan kolom Preprocessing hasil proses TF-IDF.....	44
Gambar 3.25 Tampilan hasil proses pembuatan model algoritma <i>Support Vector Machine</i> (SVM) kernel <i>linear</i>	44
Gambar 3.26 Tampilan hasil proses klasifikasi model algoritma <i>Support Vector Machine</i> (SVM) kernel <i>linear</i> kepada 20% data testing pada kolom Preprocessing hasil proses TF-IDF.	45
Gambar 3.27 Tampilan hasil proses <i>K-Fold Cross Validation</i>	46
Gambar 3.28 Tampilan nilai rata-rata fit time, score time dan test score hasil proses <i>K-Fold Cross Validation</i>	47
Gambar 3.29 Tampilan nilai <i>accuracy</i> , <i>precision</i> , <i>recall</i> dan <i>f1-score</i> hasil proses <i>Confusion Matrix</i>	47

Gambar 3.30 Tampilan diagram tabel <i>Confusion Matrix</i>	48
Gambar 4.1 Tampilan kode program input dataset.	49
Gambar 4.2 Tampilan kode program <i>lowercase</i>	50
Gambar 4.3 Tampilan kode program <i>fixes contraction</i>	51
Gambar 4.4 Tampilan kode program <i>cleaning</i>	52
Gambar 4.5 Tampilan kode program <i>remove stopwords</i>	53
Gambar 4.6 Tampilan kode program <i>lemmatization</i>	55
Gambar 4.7 Tampilan kode program <i>filtering</i>	56
Gambar 4.8 Tampilan kode program membuat kolom Preprocessing dengan menggunakan kolom Filtering pada variabel Dataframe.	57
Gambar 4.9 Tampilan kode program visualisasi WordCloud.	58
Gambar 4.10 Tampilan kode program TF-IDF.	59
Gambar 4.11 Tampilan kode program IDF.	60
Gambar 4.12 Tampilan kode program TF-IDF index ke-0.	61
Gambar 4.13 Tampilan kode program pembagian 80% data training dan 20% data testing.	62
Gambar 4.14 Tampilan kode program pembuatan model algoritma <i>Support Vector Machine</i> (SVM) kernel linear.	64
Gambar 4.15 Tampilan kode program implementasi model algoritma <i>Support Vector Machine</i> (SVM) kernel linear dalam melakukan proses klasifikasi.	65
Gambar 4.16 Tampilan kode program <i>K-Fold Cross Validation</i>	66
Gambar 4.17 Tampilan kode program menghitung rata-rata nilai <i>fit_time</i> , <i>score_time</i> dan <i>test_score</i> pada <i>K-Fold Cross Validation</i>	67
Gambar 4.18 Tampilan kode program <i>Confusion Matrix</i>	68
Gambar 4.19 Tampilan kode program visualisasi <i>Confusion Matrix</i>	69

INTISARI

Dengan pesatnya perkembangan teknologi komunikasi dan informasi pada zaman sekarang, film menjadi salah satu konten media yang paling banyak diminati warganet dengan beragam jenis genre dan judul; bahkan warganet juga dapat menontonnya dengan mudah secara online pada situs penyedia jasa streaming resmi berbayar seperti Netflix, Hulu, Apple TV+, Amazon Prime Video, Disney+, dll. Warganet yang sudah menonton sebuah film dapat memberikan ulasan positif dan negatif pada situs IMDB (*Internet Movie Database*) yang menjadi database online pelbagai informasi tentang film. Ulasan positif dan negatif sebuah film pada situs IMDB (*Internet Movie Database*) dapat menjadi referensi calon penonton yang ingin menonton atau membeli sebuah film.

Penelitian ini berfokus pada penerapan algoritma klasifikasi *Support Vector Machine* (SVM) dalam melakukan analisis sentimen positif dan negatif ulasan film *Oppenheimer* pada situs IMDB (*Internet Movie Database*) yang menjadi trending topik Google Trends Worldwide pada bulan Juli-September 2023 dan Google's Year in Search Global diurutan ke-2 dalam kategori movies pada tahun 2023 menggunakan dataset yang diunduh dari situs Kaggle sebanyak 3857 baris data.

Penerapan algoritma klasifikasi *Support Vector Machine* (SVM) dalam melakukan analisis sentimen positif dan negatif pada penelitian ini menghasilkan nilai *accuracy* 88%, *precision* 88%, *recall* 87% dan *f1-score* 87% menggunakan teknik *Confusion Matrix* serta nilai rata-rata test score 86% menggunakan teknik *K-Fold Cross Validation* dengan 10-fold.

Kata kunci: Analisis Sentimen, SVM, IMDB, *Oppenheimer* (2023)

ABSTRACT

With the rapid development of communication and information technology nowadays, films have become one of the content media that is most popular with netizens with various types of genres and titles; Netizens can even watch it easily online on official paid streaming service provider sites such as Netflix, Hulu, Apple TV+, Amazon Prime Video, Disney+, etc. Netizens who have watched a film can provide positive and negative reviews on the IMDB (Internet Movie Database) site, which is an online database of various information about films. Positive and negative reviews of a film on the IMDB (Internet Movie Database) site can be a reference for potential viewers who want to watch or buy a film.

This research focuses on the application of the Support Vector Machine (SVM) classification algorithm in analyzing the positive and negative sentiments of Oppenheimer film reviews on the IMDB (Internet Movie Database) site which is a trending topic on Google Trends Worldwide in July September 2023 and Google's Year in Search Global in order of 2nd in the film category in 2023 using a dataset downloaded from the Kaggle site of 3857 rows of data.

The application of the Support Vector Machine (SVM) classification algorithm in conducting positive and negative sentiment analysis in this research resulted in an accuracy value of 88%, precision 88%, recall 87% and f1-score 87% using the Confusion Matrix technique and an average test score value of 86 % using the K-Fold Cross Validation Technique with 10-fold.

Keyword: Sentiment Analysis, SVM, IMDB, Oppenheimer (2023)