

**KLASIFIKASI DATA *REVIEW* IMDb BERDASARKAN
ANALISIS SENTIMEN MENGGUNAKAN
*SUPPORT VECTOR MACHINE***

SKRIPSI



disusun oleh

Gita Cahyani

17.11.1670

**PROGRAM SARJANA
PROGRAM STUDI INFORMATIKA
FAKULTAS ILMU KOMPUTER
UNIVERSITAS AMIKOM YOGYAKARTA
YOGYAKARTA
2021**

**KLASIFIKASI DATA *REVIEW* IMDb BERDASARKAN
ANALISIS SENTIMEN MENGGUNAKAN
*SUPPORT VECTOR MACHINE***

SKRIPSI

untuk memenuhi sebagian persyaratan
mencapai gelar Sarjana
pada Program Studi Informatika



disusun oleh

Gita Cahyani

17.11.1670

**PROGRAM SARJANA
PROGRAM STUDI INFORMATIKA
FAKULTAS ILMU KOMPUTER
UNIVERSITAS AMIKOM YOGYAKARTA
YOGYAKARTA
2021**

PERSETUJUAN

SKRIPSI

**KLASIFIKASI DATA *REVIEW* IMDb BERDASARKAN
ANALISIS SENTIMEN MENGGUNAKAN
*SUPPORT VECTOR MACHINE***

yang dipersiapkan dan disusun oleh

Gita Cahyani

17.11.1670

telah disetujui oleh Dosen Pembimbing Skripsi
pada tanggal 30 November 2020

Dosen Pembimbing,

Wiwi Widayani, M.Kom

NIK. 190302272

PENGESAHAN**SKRIPSI****KLASIFIKASI DATA *REVIEW* IMDB BERDASARKAN
ANALISIS SENTIMEN MENGGUNAKAN
*SUPPORT VECTOR MACHINE***

yang dipersiapkan dan disusun oleh

Gita Cahyani

17.11.1670

telah dipertahankan di depan Dewan Penguji
pada tanggal 19 Maret 2021

Susunan Dewan Penguji**Nama Penguji****Tanda Tangan**

Wiwi Widayani, M.Kom

NIK. 190302272

Irma Rofni Wulandari, S.Pd., M.Eng

NIK. 190302329

Alfie Nur Rahmi, M.Kom

NIK. 190302240

Skripsi ini telah diterima sebagai salah satu persyaratan
untuk memperoleh gelar Sarjana Komputer
Tanggal 19 Maret 2021

DEKAN FAKULTAS ILMU KOMPUTER

HANIF AL FATTA, M.Kom

NIK. 190302096

PERNYATAAN

Saya yang bertandatangan dibawah ini menyatakan bahwa, skripsi ini merupakan karya sayasendiri (ASLI), dan isi dalam skripsi ini tidak terdapat karya yang pernah diajukan oleh orang lain untuk memperoleh gelar akademis di suatu institusi pendidikan tinggi manapun, dan sepanjang pengetahuan saya juga tidak terdapat karya atau pendapat yang pernah ditulis dan/atau diterbitkan oleh orang lain, kecuali yang secara tertulis diacu dalam naskah ini dan disebutkan dalam daftar pustaka.

Segala sesuatu yang terkait dengan naskah dan karya yang telah dibuat adalah menjadi tanggungjawab saya pribadi.

Singawang, 5 Juni 2021



Gita Cahyani
17.11.1670

MOTTO

“Wahai orang-orang yang beriman! Mohonlah pertolongan (kepada Allah) dengan sabar dan salat. Sungguh, Allah beserta orang-orang yang sabar”

(Q.S Al-Baqarah Ayat 153)

“Banyak kegagalan dalam hidup ini dikarenakan orang-orang tidak menyadari betapa dekatnya mereka dengan keberhasilan saat mereka menyerah”

(Thomas Alva Edison)



PERSEMBAHAN

Dengan mengucap syukur Alhamdulillah kepada Allah SWT atas dilancarkannya penyelesaian dari skripsi ini, tidaklah bijak bila penulis mengatakan jika penelitian ini terselesaikan tanpa bantuan dan dukungan dari berbagai pihak. Maka dari itu pada lembar ini izinkan penulis mengucapkan sedikit rasa terimakasih atas terselesaikannya skripsi ini kepada:

1. Allah SWT yang dengan ridhonya penulis diberi kemudahan dan kelancaran dalam pengerjaan skripsi sehingga dapat terselesaikan,
2. Kedua orang tua tercinta penulis, Alm. Bapak Sunan dan Ibu Emi Utami. Rasa terimakasih yang luar biasa penulis berikan kepada orang tua yang telah bersusah payah membiayai penulis dalam menyelesaikan pendidikan, semoga dengan terselesaikannya skripsi ini dapat menjadi kebahagiaan sekaligus kebanggaan bagi kedua orang tua dan menjadi pembuka pintu berkah rezki bagi penulis dan keluarga, Aamiin.
3. Ketiga saudara kandung saya Nina Putriyani, Prima Setya, dan Agung Budiman yang selalu memberi dukungan dan motivasi agar tetap semangat,
4. Terimakasih kepada dosen pembimbing Ibu Wiwi Widayani yang sudah membimbing dengan sabar selama pengerjaan skripsi sehingga dapat terselesaikan dengan baik,
5. Teman – teman Kalimantan Chairul Amali, Amiruddin Istiqomah, dan Qory Aprilistania Putri yang sangat banyak membantu saya selama di Jogja.
6. Teman - teman seperjuangan saya Novitha, Zuhair, Imeh, Bahrul, Bopeng, Aang, Rizki, Rangga, Tito, Abi, Kelvin, Ananda, dan semua teman kelas 17-S1 Informatika 12 lainnya yang telah menemani semasa perkuliahan. Dan terimakasih untuk mas vanul yang sudah memberi kisi kisi sebelum pendadaran. Dan terimakasih juga untuk Syahrizal Azzuri yang sudah memberi dukungan dan memberi saran dalam pengerjaan skripsi penulis,
7. Terimakasih kepada Reza Hidayat Juliandry yang sudah mendoakan dan memberi dukungan demi kelancaran sidang pendadaran.

KATA PENGANTAR

Alhamdulillah, segala puji syukur kepada Allah SWT atas segala limpahan rahmat dan hidayahnya sehingga penulis dapat menyelesaikan skripsi yang berjudul “Klasifikasi Data *Review* IMDb Berdasarkan Sentimen Analisis menggunakan *Support Vector Machine*” ini dengan lancar. Skripsi ini disusun untuk memenuhi persyaratan memperoleh gelar Sarjana Komputer di Universitas Amikom Yogyakarta. Dengan segala keterbatasan yang penulis miliki masih terdapat kekurangan yang dapat diperbaiki. Semoga hasil dari penelitian ini dapat bermanfaat bagi masyarakat luas dan bagi bidang pendidikan, riset, teknologi.

Dalam penulisan skripsi ini, penulis mendapatkan banyak bantuan dari berbagai pihak baik itu dari segi moral maupun spiritual. Maka dari itu, ucapan terimakasih penulis sampaikan kepada:

Kedua orang tua penulis.

Kepada Dosen Pembimbing, Wiwi Widayani, M.Kom

Kepada teman-teman seperjuangan yang telah membantu dalam bertukar pikiran.

Kepada kampus tercinta, Universitas Amikom Yogyakarta.

Singkawang, 25 Mei 2021

Penulis,
Gita Cahyani

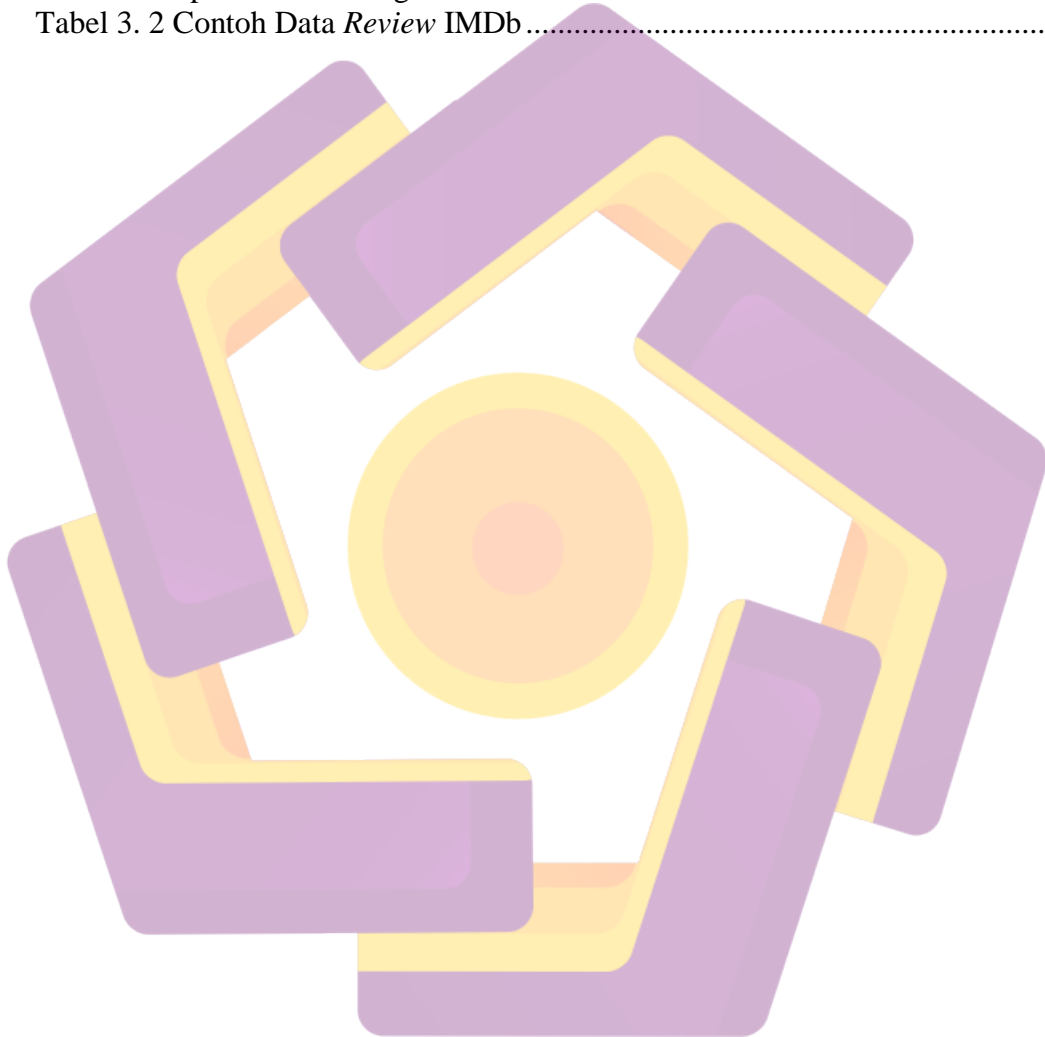
DAFTAR ISI

JUDUL	i
PERSETUJUAN	ii
PENGESAHAN	iii
PERNYATAAN.....	iv
MOTTO.....	v
PERSEMBAHAN	vi
KATA PENGANTAR	vii
DAFTAR ISI.....	viii
DAFTAR TABEL	x
DAFTAR GAMBAR	xi
DAFTAR PERSAMAAN	xii
INTISARI.....	xiii
ABSTRACT	xiv
BAB I PENDAHULUAN	1
2.1 Latar Belakang.....	1
2.2 Rumusan Masalah	4
2.3 Batasan Masalah.....	4
2.4 Tujuan Penelitian.....	5
2.5 Manfaat Penelitian.....	5
2.6 Metode Penelitian.....	6
2.7 Sistematika Penulisan.....	7
BAB II LANDASAN TEORI	8
2.1 Kajian Pustaka	8
2.2 Internet Movie Database (IMDb)	11
2.3 <i>Text Mining</i>	13
2.4 <i>Text Pre-processing</i>	13
2.5 Seleksi Fitur <i>Information Gain</i>	14
2.6 Klasifikasi.....	15
2.7 <i>Support Vector Machine</i>	16
2.8 Bahasa pemrograman Python	23

2.9	Pengujian dengan <i>Confusion Matrix</i>	24
BAB III METODE PENELITIAN.....		27
3.1	Alat dan Bahan Penelitian	27
3.1.1	Alat Penelitian.....	27
3.1.2	Bahan Penelitian.....	28
3.2	Alur Penelitian.....	29
3.2.1	Pengumpulan Data	30
3.2.2	Perancangan Sistem	31
3.3	Perancangan Antar Muka	42
BAB IV IMPLEMENTASI DAN PEMBAHASAN		48
4.1	Implementasi	48
4.1.1	Pengumpulan Data	48
4.1.2	<i>Input Dataset</i>	50
4.1.3	<i>Pre-processing Data</i>	54
4.1.4	Pembototan <i>Information Gain</i>	57
4.1.5	Klasifikasi SVM.....	58
4.2	Pengujian dan Pembahasan	60
BAB V KESIMPULAN DAN SARAN.....		61
5.1	Kesimpulan.....	61
5.2	Saran.....	61

DAFTAR TABEL

Tabel 2. 1 Persamaan dan Perbedaan dengan Penelitian Sebelumnya.....	10
Tabel 2. 2 <i>Confusion Matrix</i>	25
Tabel 3. 1 Spesifikasi Perangkat Keras.....	27
Tabel 3. 2 Contoh Data <i>Review</i> IMDb.....	38



DAFTAR GAMBAR

Gambar 1. 1 Tampilan <i>User Review</i> di IMDb	3
Gambar 2. 1 Tampilan <i>Web</i> IMDb.....	12
Gambar 2. 2 Menu Pada IMDb	12
Gambar 2. 3 Proses <i>Pre-processing</i>	14
Gambar 2. 4 SVM Model Linear dan Non-Linear	17
Gambar 3. 1 Tahapan Alur Penelitian.....	29
Gambar 3. 2 Contoh <i>Data Review</i> IMDb.....	30
Gambar 3. 3 <i>Flowchart</i> Sistem	31
Gambar 3. 4 Diagram Alir <i>Tokenizing</i>	33
Gambar 3. 5 Diagram Alir <i>Filtering</i>	34
Gambar 3. 6 Diagram Alir <i>Stopword Removal</i>	35
Gambar 3. 7 Diagram Alir <i>Stemming</i>	36
Gambar 3. 8 <i>Flowchart Information Gain</i>	37
Gambar 3. 9 Diagram Tahap Proses Klasifikasi <i>Support Vector Machine</i>	41
Gambar 3. 10 Perancangan Tampilan Halaman <i>Dashboard</i>	43
Gambar 3. 11 Perancangan Tampilan Halaman <i>Dataset</i>	44
Gambar 3. 12 Perancangan Tampilan Halaman <i>Text Pre-processing</i>	45
Gambar 3. 13 Perancangan Tampilan Halaman <i>Information Gain</i>	46
Gambar 3. 14 Perancangan Tampilan Halaman Klasifikasi SVM.....	47

DAFTAR PERSAMAAN

(Persamaan 2. 1).....	15
(Persamaan 2. 2).....	15
(Persamaan 2. 3).....	15
(Persamaan 2. 4).....	17
(Persamaan 2. 5).....	17
(Persamaan 2. 6).....	17
(Persamaan 2. 7).....	18
(Persamaan 2. 8).....	18
(Persamaan 2. 9).....	18
(Persamaan 2. 10).....	18
(Persamaan 2. 11).....	18
(Persamaan 2. 12).....	18
(Persamaan 2. 13).....	19
(Persamaan 2. 14).....	19
(Persamaan 2. 15).....	19
(Persamaan 2. 16).....	19
(Persamaan 2. 17).....	20
(Persamaan 2. 18).....	20
(Persamaan 2. 19).....	20
(Persamaan 2. 20).....	20
(Persamaan 2. 21).....	21
(Persamaan 2. 22).....	25
(Persamaan 2. 23).....	26
(Persamaan 2. 24).....	26

INTISARI

Internet Movie Database (IMDb) adalah situs *web* yang menyediakan informasi mengenai film dari seluruh dunia, termasuk orang-orang yang terlibat di dalamnya mulai dari aktor/aktris, sutradara, penulis, penata rias, *soundtrack*, nominasi yang pernah didapat, dan *review* dari para pengunjung. Ini mempengaruhi bagaimana cara kita mendapatkan informasi, dimana informasi dan hiburan dengan mudahnya kita akses lewat media digital. Besarnya peran sebuah media (IMDb) untuk mengangkat serta suatu film agar dapat mempengaruhi keputusan calon penonton untuk menentukan pilihan menonton serta menilai sebuah kualitas film.

Penelitian ini bertujuan membuat sistem analisis sentimen secara otomatis untuk mengklasifikasi *review* film yang ada di IMDb kedalam sentimen positif, dan sentimen negatif. Sehingga kebutuhan user untuk menentukan akan menonton suatu film dengan melihat *review* penonton menjadi lebih cepat dan efisien. Metode *Machine Learning* yang digunakan peneliti untuk klasifikasi teks adalah *Support Vector Machine* (SVM) dengan Seleksi Fitur *Information Gain*. Masukan sistem ini berupa dataset *review user* IMDb, dan hasil keluaran sistem ini berupa klasifikasi sentimen positif dan sentimen negatif.

Pembuatan sistem berbasis *web* yang mengimplementasikan metode *Support Vector Machine* berhasil digunakan untuk menganalisis sentimen secara otomatis. *Output* dari sistem yaitu menunjukkan jumlah dari sentimen positif dan sentimen negatif yang berhasil di klasifikasi oleh sistem. Hasil akurasi menggunakan metode *Support Vector Machine* dengan seleksi fitur *Information Gain* mencapai hingga 86.5% dan data diklasifikasi menjadi sentimen positif dan sentimen negatif dengan menggunakan data sebanyak 2000 data.

Kata Kunci : Analisis Sentimen, *Support Vector Machine*, Klasifikasi Teks, *Information Gain*

ABSTRACT

Internet Movie Database (IMDb) is a website that provides information about films from around the world, including the people involved in it ranging from actors/actresses, directors, writers, makeup artists, soundtracks, nominations, and reviews from actors. visitors. This affects how we get information, where we can easily access information and entertainment through digital media. The magnitude of the role of a media (IMDb) to raise and a film in order to influence the decision of prospective audiences to determine the choice of watching and assessing the quality of a film.

This study aims to create a sentiment analysis system automatically to classify film reviews on IMDb into positive sentiment and negative sentiment. So that the user's need to determine whether to watch a film by looking at audience reviews becomes faster and more efficient. Machine Learning method used by researchers for text classification is Support Vector Machine (SVM) with Information Gain Feature Selection. The input of this system is in the form of an IMDb user review dataset, and the output of this system is a classification of positive sentiment and negative sentiment.

The creation of a web-based system that implements the Support Vector Machine method has been successfully used to analyze sentiment automatically. The output of the system is showing the number of positive and negative sentiments that have been successfully classified by the system. Accuracy results using the Support Vector Machine method with Information Gain feature selection reach up to 86.5% and the data is classified into positive sentiment and negative sentiment using 2000 data.

Keywords : *Sentiment Analysis, Support Vector Machine, Text Classification, Information Gain*