

**ANALISIS SENTIMEN ULASAN APLIKASI *STREAMING RCTI+*
MENGUNAKAN *SUPPORT VECTOR MACHINE*
PADA *GOOGLE PLAY STORE***



SKRIPSI

Diajukan untuk memenuhi salah satu syarat kelulusan Program Sarjana

ANNISA GHINA FAUZIYYAH

NIM : 19200496

**Program Studi Sistem Informasi
Fakultas Teknik dan Informatika
Universitas Bina Sarana Informatika
Jakarta
2024**

SURAT PERNYATAAN KEASLIAN SKRIPSI

Yang bertanda tangan dibawah ini:

Nama : Annisa Ghina Fauziyyah
NIM : 19200496
Jenjang : Sarjana (S1)
Program Studi : Sistem Informasi
Fakultas : Teknik dan Informatika
Perguruan Tinggi : Universitas Bina Sarana Informatika

Dengan ini menyatakan bahwa Skripsi yang telah saya buat dengan judul: "**Analisis Sentimen Ulasan Aplikasi Streaming Rcti+ Menggunakan Support Vector Machine pada Google Play Store**", adalah asli (orisinil) atau tidak plagiat (menjiplak) dan belum pernah diterbitkan/dipublikasikan dimanapun dan dalam bentuk apa pun.

Demikianlah surat pernyataan ini saya buat dengan sebenar-benarnya tanpa ada paksaan dari pihak manapun juga. Apabila di kemudian hari ternyata saya memberikan keterangan palsu dan atau ada pihak lain yang mengklaim bahwa Skripsi yang telah saya buat adalah hasil karya milik seseorang atau badan tertentu, saya bersedia diproses baik secara pidana maupun perdata dan kelulusan saya dari **Universitas Bina Sarana Informatika** dicabut/dibatalkan.

Dibuat di : Jakarta

Pada tanggal : 1 Juli 2024

Yang menyatakan,



A handwritten signature in black ink, appearing to read 'Annisa', is written over the stamp and QR code area.

Annisa Ghina Fauziyyah

SURAT PERNYATAAN PERSETUJUAN PUBLIKASI

KARYA ILMIAH UNTUK KEPENTINGAN AKADEMIS

Yang bertanda tangan dibawah ini:

Nama : Annisa Ghina Fauziyyah
NIM : 19200496
Jenjang : Sarjana (S1)
Program Studi : Sistem Informasi
Fakultas : Teknik dan Informatika
Perguruan Tinggi : Universitas Bina Sarana Informatika

Dengan ini menyatakan bahwa seluruh data, informasi, interpretasi serta pernyataan yang terdapat dalam karya ilmiah Penulis dengan judul "**Analisis Sentimen Ulasan Aplikasi Streaming Rcti+ Menggunakan Support Vector Machine pada Google Play Store**" ini, kecuali yang disebutkan sumbernya adalah hasil pengamatan, penelitian, pengelolaan, serta pemikiran saya.

Penulis menyetujui untuk memberikan ijin kepada pihak **Universitas Bina Sarana Informatika** untuk mendokumentasikan karya ilmiah saya tersebut secara internal dan terbatas, serta tidak untuk mengunggah karya ilmiah Penulis pada repository Universitas Bina Sarana Informatika.

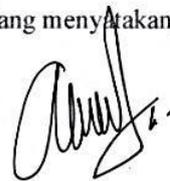
Penulis bersedia untuk bertanggung jawab secara pribadi, tanpa melibatkan pihak **Universitas Bina Sarana Informatika**, atas materi/isi karya ilmiah tersebut, termasuk bertanggung jawab atas dampak atau kerugian yang timbul dalam bentuk akibat tindakan yang berkaitan dengan data, informasi, interpretasi serta pernyataan yang terdapat pada karya ilmiah saya ini.

Demikian pernyataan ini saya buat dengan sebenarnya.

Dibuat di: Jakarta

Pada tanggal: 1 Juli 2024

Yang menyatakan,



Annisa Ghina Fauziyyah

PERSETUJUAN DAN PENGESAHAN SKRIPSI

PERSETUJUAN DAN PENGESAHAN SKRIPSI

Skripsi ini diajukan oleh:

Nama : Annisa Ghina Fauziyyah
NIM : 19200496
Jenjang : Sarjana (S1)
Program Studi : Sistem Informasi
Fakultas : Teknik dan Informatika
Perguruan Tinggi : Universitas Bina Sarana Informatika
Judul Skripsi : Analisis Sentimen Ulasan Aplikasi Streaming RCTI+ Menggunakan Support Vector Machine Pada Google Play Store

Telah dipertahankan pada periode 2024-1 dihadapan penguji dan diterima sebagai bagian persyaratan yang diperlukan untuk memperoleh Sarjana Komputer (S.Kom) pada Program Sarjana (S1) Program Studi Sistem Informasi di Universitas Bina Sarana Informatika.

Jakarta, 19 Agustus 2024

PEMBIMBING SKRIPSI

Pembimbing I : Mochamad Nandi Susila, M.Kom.

Pembimbing II : Nur Ali Farabi, M.Kom.

UNIVERSITAS DEWAN PENGUJI

Penguji I : Omar Pahlevi, M.Kom.

Penguji II : Yunita, M.Kom.

PEDOMAN PENGGUNAAN HAK CIPTA

Skripsi Sarjana yang berjudul “**Analisis Sentimen Ulasan Aplikasi *Streaming Rcti+* Menggunakan *Support Vector Machine* pada *Google Play Store***” adalah hasil karya tulis asli ANNISA GHINA FAUZIYYAH dan bukan hasil terbitan sehingga peredaran karya tulis hanya berlaku dilingkungan akademik saja, serta memiliki hak cipta. Oleh karena itu, dilarang keras untuk menggandakan baik sebagian maupun seluruhnya karya tulis ini, tanpa seizin penulis.

Referensi kepustakaan diperkenankan untuk dicatat tetapi pengutipan atau peringkasan isi tulisan hanya dapat dilakukan dengan seizin penulis dan disertai ketentuan pengutipan secara ilmiah dengan menyebutkan sumbernya.

Untuk keperluan perizinan pada pemilik dapat menghubungi informasi yang tertera di bawah ini:

Nama	: Annisa Ghina Fauziyyah
Alamat	: Perumahan Taman Walet Blok SA 11 No. 16, Kab. Tangerang, 15560
No. Telp / HP	: 087851928802
Email	: annisagf15@gmail.com



LEMBAR KONSULTASI SKRIPSI

UNIVERSITAS BINA SARANA INFORMATIKA

NIM : 19200496
Nama Lengkap : Annisa Ghina Fauziyyah
Dosen Pembimbing I : Mochammad Nandi Susila M.Kom
Judul Skripsi : Analisis Sentimen Ulasan Aplikasi *Streaming RCTI+*
Menggunakan *Support Vector Machine* pada *Google Play Store*

No.	Tanggal Bimbingan	Pokok Bahasan	Paraf Dosen Pembimbing I
1.	5 April 2024	Bimbingan Perdana	dr
2.	26 April 2024	Bimbingan BAB I (Latar Belakang)	dr
3.	6 Mei 2024	ACC BAB I	dr
4.	17 Mei 2024	Bimbingan BAB II (Metode Penelitian)	dr
5.	31 Mei 2024	ACC BAB II	dr
6.	10 Juni 2024	Bimbingan BAB III	dr
7.	21 Juni 2024	Bimbingan BAB IV	dr
8.	26 Juni 2024	Cek Keseluruhan BAB	dr
9.	1 Juli 2024	ACC Seluruh BAB	dr

Catatan untuk Dosen Pembimbing
Bimbingan Skripsi

Dimulai pada tanggal : 5 April 2024
Diakhiri pada tanggal : 1 Juli 2024
Jumlah pertemuan bimbingan : 9

Disetujui oleh,
Dosen Pembimbing I

(Mochammad Nandi Susila, M.Kom)



LEMBAR KONSULTASI SKRIPSI

UNIVERSITAS BINA SARANA INFORMATIKA

NIM : 19200496
Nama Lengkap : Annisa Ghina Fauziyyah
Dosen Pembimbing II : Nur Ali Farabi M.Kom
Judul Skripsi : Analisis Sentimen Ulasan Aplikasi *Streaming RCTI+* Menggunakan *Support Vector Machine* pada *Google Play Store*

No.	Tanggal Bimbingan	Pokok Bahasan	Paraf Dosen Pembimbing II
1.	5 April 2024	Bimbingan Perdana	
2.	26 April 2024	Bimbingan BAB I	
3.	13 Mei 2024	Revisi BAB I (Latar Belakang)	
4.	31 Mei 2024	ACC BAB I	
5.	3 Juni 2024	Bimbingan BAB II dan BAB III	
6.	21 Juni 2024	ACC BAB II dan BAB III	
7.	24 Juni 2024	Cek Keseluruhan BAB	
8.	28 Juni 2024	ACC Seluruh BAB	

Catatan untuk Dosen Pembimbing
Bimbingan Skripsi

Dimulai pada tanggal : 5 April 2024
Diakhiri pada tanggal : 28 Juli 2024
Jumlah pertemuan bimbingan : 8

Disetujui oleh,
Dosen Pembimbing II


(Nur Ali Farabi, M.Kom)

PERSEMBAHAN

Alhamdulillah Rabbil Aalamin, sujud serta syukur kepada Allah SWT.

Terimakasih atas karunia-Mu yang telah memberikan kemudahan dan kelancaran sehingga skripsi ini dapat terselesaikan dengan baik.

Skripsi ini saya persembahkan untuk diri saya sendiri yang telah berjuang dan berusaha selama ini. Terimakasih atas kerja kerasnya. Mari tetap berdoa dan berusaha serta jangan menyerah untuk kedepannya.

Halaman persembahan ini juga ditujukan sebagai ungkapan terimakasih kepada keluarga saya yang telah mendoakan dan memberikan dukungan penuh selama perjuangan menempuh pendidikan.

Terimakasih banyak untuk semuanya yang telah mendukung dan meyemangati dalam perjuangan ini.



Tuhan tau waktu yang tepat, tempat yang tepat dan jawaban yang tepat untuk semua doa – doa kita – Rony Parulian

KATA PENGANTAR

Alhamdulillah, Dengan mengucapkan puji syukur kehadirat Allah SWT, yang telah melimpahkan rahmat dan karunia-Nya, sehingga pada akhirnya penulis dapat menyelesaikan tugas ini dengan baik. Skripsi pada Program Sarjana ini penulis sajikan dalam bentuk buku yang sederhana. Adapun judul Skripsi, yang penulis ambil sebagai berikut, “**Analisis Sentimen Ulasan Aplikasi Streaming Rcti+ Menggunakan Support Vector Machine pada Google Play Store**”.

Tujuan penulisan Skripsi pada Program Sarjana ini dibuat sebagai salah satu syarat kelulusan Sarjana Universitas Bina Sarana Informatika. Sebagai bahan penulisan diambil berdasarkan hasil penelitian (eksperimen), observasi dan beberapa sumber literatur yang mendukung penulisan ini. Penulis menyadari bahwa tanpa bimbingan dan dorongan dari semua pihak, maka penulisan Skripsi ini tidak akan berjalan lancar. Oleh karena itu pada kesempatan ini, izinkanlah penulis menyampaikan ucapan terima kasih kepada:

1. Rektor Universitas Bina Sarana Informatika.
2. Dekan Fakultas Teknik dan Informatika.
3. Ketua Program Studi Sistem Informasi Universitas Bina Sarana Informatika.
4. Bapak Mochamad Nandi Susila, M.Kom., selaku dosen pembimbing I yang telah membimbing dan mengarahkan penulis dalam penyusunan skripsi.
5. Bapak Nur Ali Farabi, M.Kom., selaku dosen pembimbing II yang telah membimbing dan mengarahkan penulis dalam penyusunan skripsi.
6. Dosen – dosen program studi sistem informasi yang telah memberikan ilmu dan bimbingannya selama masa perkuliahan.
7. Kedua orang tua penulis yang selalu memberikan dukungan dan doa tanpa henti untuk kesuksesan anak – anaknya.
8. Keluarga besar penulis yang selalu memberikan dukungan dan doa kepada penulis.
9. Syifa, Zahra, Putra, Abel, dan teman – teman peneliti yang tidak bisa disebutkan satu persatu, yang selalu memberikan semangat, hiburan dan bantuan kepada peneliti selama pengerjaan skripsi ini.

10. Nyoman Paul, Nabila Taqiyyah, Rony Parulian dan Salma Salsabil alias PANAROMA, yang selalu menemani peneliti dengan musik kalian yang luar biasa. Terima kasih telah menjadi sumber inspirasi dan hiburan, membantu peneliti melewati setiap tantangan dengan lebih ringan dan penuh semangat.

Peneliti menyadari bahwa skripsi ini, masih jauh dari kata sempurna, oleh karena itu peneliti sangat mengharapkan kritik dan saran yang membangun dari semua pihak demi perbaikan dan penyempurnaan skripsi di masa yang akan datang. Semoga skripsi ini dapat bermanfaat, khususnya bagi peneliti dan umumnya bagi semua pembaca. *Aamiin ya Rabbal Alamiin.*

Jakarta, 1 Juli 2024

Penulis



Annisa Ghina Fauziyyah



ABSTRAK

Annisa Ghina Fauziyyah (19200496), Analisis Sentimen Ulasan Aplikasi Streaming Rcti+ Menggunakan Support Vector Machine pada Google Play Store.

Perkembangan inovasi teknologi menyebabkan munculnya berbagai macam layanan *streaming online* seperti RCTI+. Aplikasi RCTI+ berisi berbagai program dan tayangan televisi secara *online* dan *live streaming*, serta fitur lain seperti video, serial TV, film, berita, radio, dan lain – lain. Aplikasi ini telah di unduh lebih dari 10 juta kali dan memiliki *rating* 4,2 dengan 160 ribu ulasan. Dengan banyaknya pengguna aplikasi RCTI+, jumlah komentar dan ulasan yang dibagikan oleh masyarakat bersifat tidak terstruktur, membuat pemahaman terhadap opini pengguna menjadi terhambat. Oleh karena itu, analisis terhadap ulasan aplikasi RCTI+ diperlukan untuk mengetahui opini masyarakat tentang aplikasi tersebut. Dalam penelitian ini, analisis sentimen dilakukan menggunakan metode *Support Vector Machine (SVM)* dengan dataset sebanyak 1500, terdiri dari 948 sentimen positif dan 552 sentimen negatif. Setelah tahap *preprocessing* dan pembobotan *tf-idf* selesai, akan dilakukan pembagian data dengan rasio 80%:20%, menghasilkan 1200 data latih dan 300 data uji. Kemudian dilakukan klasifikasi sentimen menggunakan SVM. Hasilnya menunjukkan bahwa metode SVM memiliki *accuracy* sebesar 90,33%, *precision* sebesar 90,86%, dan *recall* sebesar 94,21%. Hasil tersebut menunjukkan bahwa algoritma *Support Vector Machine* efektif dan akurat dalam mengklasifikasikan sentimen ulasan aplikasi RCTI+ di *Google Play Store*.

Kata kunci : Analisis sentimen, RCTI+, *Support Vector Machine*

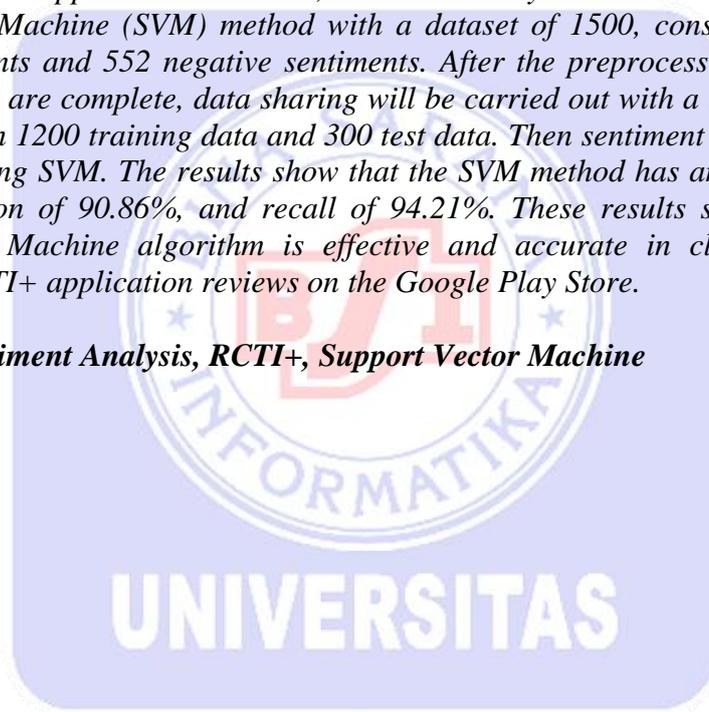


ABSTRACK

Annisa Ghina Fauziyyah (19200496), Sentiment Analysis of Rcti + Streaming Application Reviews Using Support Vector Machine on Google Play Store.

The development of technological innovation has led to the emergence of various online streaming services such as RCTI+. The RCTI+ application contains various television programmes and shows online and live streaming, as well as other features such as videos, TV series, films, news, radio, and others. This application has been downloaded more than 10 million times and has a rating of 4.2 with 160 thousand reviews. With the large number of RCTI+ app users, the number of comments and reviews shared by the community is unstructured, making understanding user opinions hampered. Therefore, an analysis of RCTI+ app reviews is needed to find out people's opinions about the app. In this research, sentiment analysis is conducted using the Support Vector Machine (SVM) method with a dataset of 1500, consisting of 948 positive sentiments and 552 negative sentiments. After the preprocessing and tf-idf weighting stages are complete, data sharing will be carried out with a ratio of 80%: 20%, resulting in 1200 training data and 300 test data. Then sentiment classification is performed using SVM. The results show that the SVM method has an accuracy of 90.33%, precision of 90.86%, and recall of 94.21%. These results show that the Support Vector Machine algorithm is effective and accurate in classifying the sentiment of RCTI+ application reviews on the Google Play Store.

Keywords : Sentiment Analysis, RCTI+, Support Vector Machine



DAFTAR ISI

SURAT PERNYATAAN KEASLIAN SKRIPSI	i
SURAT PERNYATAAN PERSETUJUAN PUBLIKASI KARYA ILMIAH.....	ii
PERSETUJUAN DAN PENGESAHAN SKRIPSI	iii
PEDOMAN PENGGUNAAN HAK CIPTA	iv
LEMBAR KONSULTASI SKRIPSI	v
PERSEMBAHAN.....	vii
KATA PENGANTAR.....	viii
ABSTRAK	x
DAFTAR ISI.....	xii
DAFTAR GAMBAR.....	xiv
DAFTAR TABEL	xv
BAB I PENDAHULUAN.....	1
1.1 Latar Belakang Masalah	1
1.2 Rumusan Masalah	4
1.3 Tujuan dan Manfaat Penelitian.....	4
1.4 Batasan Penelitian	5
1.5 Penelitian Terdahulu.....	5
1.6 Gambaran Metode Usulan.....	8
BAB II METODOLOGI PENELITIAN.....	10
2.1 Landasan Teori	10
2.1.1 Analisis Sentimen	10
2.1.2 Aplikasi <i>Mobile</i>	11
2.1.3 Aplikasi RCTI+	12
2.1.4 <i>Web Scraping</i>	13
2.1.5 <i>Text Mining</i>	13
2.1.6 <i>Preprocessing</i>	14
2.1.7 Pembobotan Kata.....	15
2.1.8 <i>Support Vector Machine (SVM)</i>	16
2.1.9 <i>Confusion Matrix</i>	17
2.1.10 <i>Python</i>	18
2.1.11 RapidMiner	19
2.2 Tahapan Penelitian	19
2.2.1 Pengumpulan Data.....	21

2.2.2	Pelabelan Data	21
2.2.3	<i>Preprocessing Data</i>	22
2.2.4	Pembobotan Kata.....	23
2.2.5	Pembagian Data.....	23
2.2.6	Klasifikasi <i>Support Vector Machine (SVM)</i>	23
2.2.7	Pengujian Evaluasi	23
BAB III HASIL DAN PEMBAHASAN		24
3.1	Proses Umum.....	24
3.1.1	Pengumpulan Data.....	24
3.1.2	Pelabelan Data	25
3.1.3	<i>Preprocessing Data</i>	26
3.1.4	Pembobotan Kata.....	35
3.1.5	<i>Split Data</i>	36
3.1.6	Klasifikasi <i>Support Vector Machine (SVM)</i>	36
3.1.7	Evaluasi Model.....	37
BAB IV PENUTUP		39
4.1	Kesimpulan.....	39
4.2	Saran.....	40
DAFTAR PUSTAKA		42
DAFTAR RIWAYAT HIDUP		44
SURAT PERNYATAAN KEBENARAN/KEABSAHAN DATA		45
BUKTI HASIL PENGECEKAN PLAGIARISME		46



DAFTAR GAMBAR

Gambar II. 1 Alur Tahapan Penelitian	20
Gambar II. 2 Tahapan <i>Preprocessing Data</i>	22
Gambar III. 1 <i>Install library web scraping</i>	24
Gambar III. 2 Proses <i>scraping data</i>	24
Gambar III. 3 Hasil <i>scraping data</i>	25
Gambar III. 4 Penyimpanan dataset.....	25
Gambar III. 5 Pelabelan data	26
Gambar III. 6 Parameter <i>read excel</i>	26
Gambar III. 7 Hasil <i>import data</i>	27
Gambar III. 8 Penentuan <i>role</i>	27
Gambar III. 9 Mengubah <i>type</i>	28
Gambar III. 10 Parameter <i>filter examples</i>	28
Gambar III. 11 <i>Add filter</i>	29
Gambar III. 12 Parameter <i>remove duplicates</i>	29
Gambar III. 13 Parameter <i>nominal to text</i>	30
Gambar III. 14 Proses <i>cleaning</i>	31
Gambar III. 15 Hasil <i>cleaning</i>	31
Gambar III. 16 Sub-Operator <i>process document from data</i>	31
Gambar III. 17 Hasil <i>case folding</i>	32
Gambar III. 18 Hasil <i>tokenizing</i>	33
Gambar III. 19 Hasil <i>stopwords removal</i>	34
Gambar III. 20 Hasil <i>filter tokens (by length)</i>	34
Gambar III. 21 Parameter TF-IDF.....	35
Gambar III. 22 Hasil pembobotan TF-IDF.....	35
Gambar III. 23 Rasio <i>split data</i>	36
Gambar III. 24 Proses implementasi SVM.....	37
Gambar III. 25 Hasil implementasi SVM.....	37

DAFTAR TABEL

Tabel I. 1 Penelitian Terdahulu	6
Tabel II. 1 <i>Confusion Matrix</i>	17
Tabel III. 1 <i>Confusion Matrix</i>	38



BAB I

PENDAHULUAN

1.1 Latar Belakang Masalah

Saat ini, perkembangan teknologi informasi dan komunikasi berkembang semakin pesat. Hal ini menyebabkan terjadinya pergeseran preferensi dan perilaku masyarakat dalam menggunakan teknologi informasi dan komunikasi. Salah satu media komunikasi yang mengalami pergeseran adalah televisi. Televisi mengalami perubahan yang cukup besar, dari televisi dengan pinggiran besar dan layar hitam putih menjadi televisi dengan layar besar dan berwarna, bahkan semakin lama muncul jenis – jenis baru seperti televisi HD, televisi interaktif, televisi satelit, kabel multi-saluran, dan bahkan televisi *streaming* dan *video on demand*. Namun, dengan perkembangan teknologi internet saat ini, layanan siaran tidak hanya ditampilkan secara eksklusif di televisi seperti sebelumnya. Pada era internet ini, orang dapat menonton layanan siaran televisi kapan saja dan di mana saja. Bahkan, mereka juga tidak perlu pergi ke pesawat televisi untuk menonton acara TV karena hal tersebut dapat dilakukan melalui perangkat seperti *smartphone* atau laptop (R. R. Putra & Hidayat, 2022).

Selain itu, terjadinya digitalisasi, peningkatan jumlah jaringan internet, dan perkembangan inovasi teknologi pada saat ini menyebabkan munculnya berbagai macam layanan *streaming online* yang membuat penonton dapat melihat layanan siaran televisi tanpa terpaku jam dan waktu siaran. Hal ini digambarkan sebagai revolusi media baru, yaitu pergeseran semua budaya ke arah produksi, distribusi, dan komunikasi yang dimediasi komputer. Layanan *streaming* merupakan salah satu contoh dari revolusi media baru. Kemudahan dalam mengakses layanan siaran tanpa

terikat waktu dan tempat membuat penonton mulai beralih dari media televisi tradisional ke layanan *streaming online*. Aplikasi *streaming* dapat mengirimkan audio, video, dan multimedia secara *real time* atau *on demand* ke jaringan. Mereka juga dapat mengirim media data digital dari *server* dan diterima oleh pengguna secara bersamaan. Ini berarti bahwa pengguna bisa langsung menikmati konten tanpa harus menunggu seluruh file selesai diunduh karena *server* mengirimkan data yang diperlukan dalam interval waktu tertentu (R. R. Putra & Hidayat, 2022).

Salah satu platform *video streaming* yang populer di Indonesia adalah RCTI+ yang berada di bawah naungan bisnis MNC Group, perusahaan media besar di Indonesia yang berdiri sejak tahun 1989 dan sekarang menjadi grup bisnis terkemuka di Indonesia. Sebelum era *streaming*, RCTI hanya berfungsi sebagai stasiun televisi bebas udara atau *Free to Air* (FTA) yang memiliki permasalahan utama terkait keterbatasan akses program dan tayangan televisi yang hanya dapat dinikmati melalui siaran televisi tradisional. Namun, dengan adanya aplikasi *streaming* RCTI+, pengguna dapat mengakses berbagai program dan tayangan televisi secara *online* dan *streaming* langsung, serta berbagai fitur lain, seperti video, serial TV, film, berita, radio, dan lain-lain (R. R. Putra & Hidayat, 2022).

Berdasarkan informasi dari situs *Google Play*, sejak peluncurannya di tahun 2019, aplikasi RCTI+ telah di unduh lebih dari 10 juta kali dan memiliki *rating* 4,2 dengan 160 ribu ulasan. Namun, pemberian *rating* yang disertai dengan ulasan tidak selalu mencerminkan bahwa kebutuhan atau masalah pengguna telah terpenuhi sepenuhnya, atau bahwa evaluasi untuk peningkatan layanan dapat dihentikan. Hal ini disebabkan oleh fakta bahwa beberapa pengguna memberikan *rating* tinggi, tetapi ulasan yang mereka berikan justru bersifat negatif (Nadhif et al., 2022). Ulasan pengguna pada aplikasi memiliki pengaruh besar terhadap calon pengguna dalam menentukan apakah

mereka akan mengunduh dan menggunakan aplikasi tersebut. Meskipun jumlah ulasan di *Google Play Store* cukup banyak, sifatnya yang tidak terstruktur membuat pemahaman terhadap opini pengguna menjadi terhambat. Oleh karena itu, diperlukan teknik khusus untuk menganalisis dan mengelompokkan ulasan pengguna terhadap aplikasi tersebut, yang dikenal dengan analisis sentimen (Diki Hendriyanto et al., 2022).

Tujuan dilakukannya analisis sentimen pada aplikasi RCTI+ adalah untuk mengidentifikasi dan memahami lebih dalam persepsi pengguna terkait berbagai aspek aplikasi yang mempengaruhi pengalaman mereka. Meskipun aplikasi ini telah memiliki popularitas yang cukup tinggi, ulasan pengguna menunjukkan adanya masalah – masalah seperti ketidakpuasan terhadap performa aplikasi, keluhan tentang fitur yang tidak berfungsi dengan baik, serta saran untuk peningkatan konten dan layanan pelanggan. Pemahaman terhadap masalah – masalah ini melalui analisis sentimen sangat penting untuk membantu pengembang aplikasi memahami kebutuhan dan ekspektasi pengguna secara mendalam, serta mengidentifikasi aspek – aspek yang perlu diperbaiki (Nadhif et al., 2022).

Berbagai algoritma dapat digunakan dalam analisis sentimen, terdapat berbagai pilihan algoritma yang bisa digunakan, seperti *K-Nearest Neighbor* (KNN), *Naïve Bayes*, dan *Support Vector Machine* (SVM). Berdasarkan beberapa referensi, SVM memiliki performa unggul dalam analisis sentimen dibandingkan dengan algoritma klasifikasi lainnya (Diki Hendriyanto et al., 2022). SVM adalah algoritma cerdas yang mempelajari pola data untuk mengelompokkannya (klasifikasi) atau memprediksi nilainya (regresi). Oleh karena itu, peneliti merekomendasikan penggunaan SVM sebagai metode yang efektif untuk menganalisis ulasan pengguna pada aplikasi RCTI+. Dengan melakukan analisis ini, diharapkan pengembang aplikasi RCTI+ dapat

memperoleh wawasan berharga yang dapat digunakan untuk meningkatkan kualitas aplikasi dan memberikan pengalaman pengguna yang lebih baik.

1.2 Rumusan Masalah

Berdasarkan penjelasan latar belakang diatas, maka masalah penelitian dapat dirumuskan sebagai berikut:

1. Bagaimana langkah – langkah melakukan analisis sentimen ulasan aplikasi RCTI+ di *Google Play Store* menggunakan algoritma klasifikasi *Support Vector Machine*?
2. Bagaimana tingkat akurasi algoritma *Support Vector Machine* dalam mengklasifikasikan sentimen ulasan aplikasi RCTI+ di *Google Play Store*?

1.3 Tujuan dan Manfaat Penelitian

Adapun tujuan dari penelitian ini adalah sebagai berikut:

1. Sebagai syarat kelulusan dari program Sarjana (S1) jurusan Sistem Informasi, Fakultas Teknik Informatika, Universitas Bina Sarana Informatika.
2. Melakukan analisis sentiment terhadap ulasan aplikasi RCTI+ di *Google Play Store* dengan memanfaatkan algoritma klasifikasi *Support Vector Machine*.
3. Mengevaluasi tingkat akurasi algoritma *Support Vector Machine* menggunakan *confusion matrix*.

Adapun Manfaat penelitian ini adalah sebagai berikut:

1. Bagi Mahasiswa
 - a) Meningkatkan dan menerapkan pengetahuan yang dipelajari selama kuliah secara langsung.
 - b) Mengetahui analisis sentimen pada masyarakat terhadap aplikasi RCTI+ menggunakan *Support Vector Machine (SVM)*.

2. Bagi Pembaca

- a) Memberikan informasi mengenai sentimen pengguna terhadap aplikasi RCTI+.
- b) Sebagai bahan referensi untuk penelitian selanjutnya mengenai analisis sentimen ulasan aplikasi *mobile*.

3. Bagi Universitas

Untuk mengetahui tolak ukur pengetahuan mahasiswa dalam menguasai materi yang sudah dipelajari.

1.4 Batasan Penelitian

Penelitian ini memiliki beberapa batasan, yaitu:

1. Data yang digunakan dalam penelitian ini diambil dari ulasan aplikasi RCTI+ di *Google Play Store* melalui metode *web scrapping* dan analisis dilakukan menggunakan RapidMiner.
2. Algoritma klasifikasi yang digunakan dalam penelitian ini adalah *Support Vector Machine*.
3. Klasifikasi sentimen dibagi menjadi dua, yaitu sentimen negatif dan sentimen positif.
4. Dataset yang digunakan berupa ulasan dalam Bahasa Indonesia.

1.5 Penelitian Terdahulu

Berbagai penelitian sebelumnya telah dilakukan terkait analisis sentimen pada ulasan aplikasi *mobile*. Penelitian tersebut memanfaatkan berbagai teknik machine learning, seperti Support Vector Machine (SVM), Naïve Bayes, K-Nearest Neighbor, Random Forest, dan sebagainya. Penulis mengemukakan temuan berikut dari studi sebelumnya sebagai sebuah contoh.

Penelitian yang mendasari skripsi ini dilakukan dengan mengkaji kembali penelitian – penelitian terdahulu yang relevan, terutama yang dipublikasikan dalam jurnal ilmiah. Berikut merupakan penelitian terdahulu yang sejenis dengan analisis sentimen menggunakan algoritma SVM yang dapat dilihat pada Tabel 1.1.

Tabel I. 1
Penelitian Terdahulu

No	Judul Penelitian	Nama Peneliti	Metode dan Tools	Hasil Penelitian
1	Analisis Sentimen Ulasan Aplikasi Mola Pada <i>Google Play Store</i> Menggunakan Algoritma <i>Support Vector Machine</i>	(Diki Hendriyanto et al., 2022)	<i>Support Vector Machine</i>	Data yang digunakan pada penelitian ini terdiri dari 520 ulasan pengguna aplikasi Mola. Diperoleh hasil terbaik pada skenario 1 dengan ratio perbandingan 90:10 menghasilkan <i>accuracy</i> 92,31%, <i>precision</i> 96,3%, dan <i>recall</i> 89,66%.
2	Analisis Sentimen Ulasan Pengguna Aplikasi Myim3 pada Situs <i>Google Play</i> Menggunakan <i>Support Vector Machine</i>	(Aditiya et al., 2022)	<i>Support Vector Machine</i>	Penelitian ini menerapkan algoritma <i>support vector machine</i> dengan 2 kernel, yaitu kernel <i>linear</i> dan <i>RBF</i> dengan menguji 4 skenario. Hasil terbaik dicapai pada skenario 3 dengan perbandingan 70:30 untuk kernel <i>linear</i> dengan akurasi 87% dan skenario 1 (90:10) dengan akurasi 87% di kernel <i>RBF</i> .
3	Analisis Sentimen pada <i>review</i> Aplikasi Grab di	(Wahyudi et al., 2021)	<i>Support Vector Machine</i>	Penelitian ini menganalisis 900 data ulasan pengguna dari aplikasi Grab Indonesia di

	<i>Google Play Store</i> Menggunakan <i>Support Vector</i> <i>Machine</i>			<i>Google Play Store</i> . Metode SVM yang digunakan berhasil mencapai akurasi sebesar 85,54%.
4	Perbandingan Algoritma <i>Support Vector Machine</i> dan <i>Random Forest</i> untuk Analisis Sentimen Terhadap Kebijakan Pemerintah Indonesia Terkait Kenaikan Harga BBM Tahun 2022	(Samantri & Afiyati, 2024)	<i>Support Vector Machine</i> dan <i>Random Forest</i>	Penelitian ini mengumpulkan 4256 tweet yang menggunakan kata kunci "bbm naik" untuk dianalisis. Dari hasil pengujian, algoritma SVM mencapai akurasi 77%, sedikit lebih baik dibandingkan algoritma <i>Random Forest</i> yang mencapai 76%. Meski sama – sama menunjukkan performa yang baik, disimpulkan bahwa algoritma SVM memiliki tingkat akurasi yang jauh lebih baik untuk analisis sentiment.
5	Analisis Sentimen <i>Zoom Cloud Meetings</i> di <i>Play Store</i> Menggunakan <i>Naïve Bayes</i> dan <i>Support Vector Machine</i>	(Herlinawati et al., 2020)	<i>Support Vector Machine (SVM)</i> dan <i>Naïve Bayes</i>	Penelitian ini menbandingkan dua algoritma klasifikasi yaitu <i>Support Vector Machine (SVM)</i> dan <i>Naïve Bayes</i> dengan menggunakan dataset sebanyak 1.007 data. Menghasilkan nilai akurasi yang lebih tinggi pada algoritma SVM sebesar 81,22% dan nilai <i>AUC</i> sebesar 0.886, sementara untuk algoritma <i>Naïve Bayes</i> sebesar 74,37%.

1.6 Gambaran Metode Usulan

Analisis sentimen ulasan aplikasi RCTI+ menggunakan *Support Vector Machine (SVM)* pada *Google Play Store* dilakukan dengan tujuan untuk memahami pendapat pengguna terhadap aplikasi *streaming video* tersebut. Berikut adalah gambaran metode yang akan digunakan dalam penelitian ini:

1. Pengumpulan Data

Data yang digunakan dalam penelitian ini ialah ulasan aplikasi RCTI+ pada *Google Play Store*. Data ini dikumpulkan menggunakan metode *web scraping*, yang memungkinkan pengumpulan data secara otomatis dan efisien.

2. Pelabelan Data

Data yang sudah terkumpulkan kemudian diberi label dan dikelompokkan ke dalam dua kategori, yaitu sentimen positif dan sentimen negatif.

3. *Preprocessing Data*

Data yang telah dilakukan pelabelan kemudian dilakukan *preprocessing* untuk menghilangkan *noise* dan memperbaiki kualitas data. Terdapat beberapa langkah dalam proses ini, yaitu *cleaning*, *case folding*, *tokenizing*, *stopword removal*, dan *filter tokens (by length)*.

4. Pembobotan Kata

Data yang telah di proses, kemudian dilakukan pembobotan kata untuk mengubah nilai term dan memberikan indikator untuk setiap kata sesuai dengan tingkat kepentingannya dengan metode TF-IDF.

5. Pembagian Data

Setelah dilakukan pembobotan, data kemudian dibagi menjadi dua bagian: data *training* dan data *testing*. Data *training* digunakan untuk mempelajari pola sentimen

dalam ulasan, sedangkan data *testing* digunakan untuk menguji akurasi model yang dibuat.

6. Klasifikasi *Support Vector Machine (SVM)*

Klasifikasi dilakukan melalui penggunaan data latih untuk memahami pola data. Algoritma SVM menggunakan data uji dengan ulasan positif dan negatif untuk mempelajari pola data berdasarkan karakteristik data masing – masing kelas.

7. Evaluasi Model

Kinerja model SVM dievaluasi menggunakan *confusion matrix* untuk menentukan *accuracy*, *precision*, dan *recall*.



BAB II

METODOLOGI PENELITIAN

2.1 Landasan Teori

2.1.1 Analisis Sentimen

Analisis sentimen adalah salah satu teknik *Natural Language Processing (NLP)* dan *machine learning* untuk mengevaluasi secara otomatis teks untuk perasaan yang dirasakan penulis (positif, negatif, dan netral). Ini adalah area pemrosesan bahasa alami (NLP) yang berkembang karena interaksi manusia dengan komputer, ekstraksi informasi, dan penyulingan perasaan (*distillation of feelings*) dari data yang ada. Analisis sentimen mencakup mengidentifikasi kata atau frasa dalam teks yang menunjukkan sikap netral, positif, atau negatif. Analisis sentimen biasanya digunakan untuk mengekstrak berbagai karakteristik dari data tekstual terstruktur atau tidak terstruktur dan menganalisisnya untuk mengekstrak pikiran, pendapat, dan perasaan dari data tersebut. Seiring dengan banyaknya pengguna internet, semakin banyak juga orang yang mengungkapkan perasaan dan pendapat mereka secara bebas di berbagai media sosial elektronik. Sentimen bisa menjadi sumber daya yang berharga bagi perusahaan yang ingin memahami pandangan masyarakat terhadap produk atau layanan mereka (Purnamasari et al., 2023).

Analisis sentimen menjadi sangat populer karena kebutuhan individu atau kelompok yang meningkat untuk mengetahui pendapat seseorang tentang sesuatu. Penanganan yang berbeda untuk analisis sentimen juga dipengaruhi oleh dataset yang digunakan (Widayat, 2021). Analisis sentimen tidak hanya ditujukan untuk individu saja, tetapi juga kepada organisasi.

Analisis sentimen, juga dikenal sebagai opini mining, adalah usaha untuk mengenali perasaan pengguna dengan menganalisis tulisan mereka. Analisis sentimen dapat digunakan untuk mengatasi masalah bisnis, program, produk, aplikasi, dan lain – lain yang dapat dikritik oleh masyarakat (Herlinawati et al., 2020).

Secara umum, terdapat lima langkah dalam analisis sentimen, yaitu pengambilan data, *pre-processing*, *feature selection*, *classification*, dan *evaluation*. Data yang tidak terstruktur dapat diubah menjadi data yang terstruktur rapi melalui analisis sentimen. Analisis sentimen juga dapat menganalisis peristiwa, pernyataan, dan komentar yang kontroversial. Selain itu, analisis sentimen dapat membantu perusahaan, *public figure*, dan pemerintahan memutuskan apa yang harus mereka lakukan (Natasuwarna, 2019).

2.1.2 Aplikasi Mobile

Aplikasi *mobile*, juga dikenal sebagai "aplikasi seluler", adalah sejenis perangkat lunak yang dirancang untuk digunakan pada perangkat seluler, seperti tablet PC atau *smartphone*. Aplikasi seluler sering berfungsi untuk memenuhi kebutuhan pengguna dengan layanan serupa dengan yang ada di PC, namun umumnya berupa *software* kecil yang memiliki kapasitas terbatas. Sebaliknya, setiap aplikasi menawarkan fitur yang terbatas dan terpisah seperti *game*, kalkulator, atau navigasi *web* seluler. Walaupun aplikasi tidak bisa multitasking karena keterbatasan sumber daya perangkat keras pada perangkat seluler awal, keunggulan mereka sekarang memungkinkan pengguna untuk memilih fungsi perangkat mereka (Y. W. S. Putra et al., 2023).

Aplikasi terbagi dalam dua kategori utama, yaitu aplikasi *native* atau asli dan aplikasi *web*. Aplikasi *native* memiliki performa yang lebih baik dan tampilan antarmuka pengguna yang lebih lancar, serta umumnya memerlukan proses

pengembangan dan jaminan kualitas yang lebih ketat sebelum diluncurkan. Aplikasi *web* dijalankan melalui *browser* dan menggunakan HTML5 atau CSS. Semua aplikasi *web* membutuhkan koneksi internet yang kuat agar dapat digunakan, karena saat pengguna dialihkan ke halaman *web* tertentu, semua data disimpan di *database* berbasis *server* (Y. W. S. Putra et al., 2023).

2.1.3 Aplikasi RCTI+

Aplikasi RCTI+ adalah platform *streaming* yang dibuat oleh PT MNC Digital Indonesia, bagian dari Media Nusantara Citra Group. Aplikasi ini diluncurkan pada tahun 2019 dan telah menjadi salah satu aplikasi *streaming* yang paling populer di Indonesia. Aplikasi ini menawarkan berbagai fitur, termasuk *streaming* video yang sudah terlewatkan atau *live streaming* dari empat stasiun televisi MNC, yaitu RCTI, GTV, MNCTV, dan iNews. RCTI+ juga memiliki layanan *streaming* unggulan yang sedang berkembang. RCTI+ diambil dari nama stasiun televisi milik MNC Group, RCTI, yang merupakan stasiun televisi swasta pertama di Indonesia. Sedangkan, tanda “+” pada RCTI+ menunjukkan pergeseran media televisi *free-to-air* menjadi digital (R. R. Putra & Hidayat, 2022).

Aplikasi ini juga memiliki lima konten pilar yang dibuat khusus, yaitu *Video+*, *News+*, *Radio+*, *HOT*, dan *Games+*, untuk mempermudah pengguna mencari konten yang sesuai dengan kategori mereka. Selain itu, RCTI+ juga menawarkan layanan *Audio Video on Demand (AVOD)* yang memungkinkan pengguna untuk menonton konten secara gratis tanpa perlu berlangganan. Aplikasi ini sangat mudah digunakan, dengan registrasi tanpa dipungut biaya dan *streaming* 4 stasiun TV populer tanpa *buffer*.

Aplikasi RCTI+ terus meningkatkan kualitas dan jumlah kontennya. Selain itu, aplikasi ini terus berkembang dengan menambahkan fitur baru, seperti agregasi berita

dari ratusan penerbit terkemuka yang dapat dibaca secara langsung dan terkini. Selain itu, RCTI+ menawarkan fitur *download* film sepuasnya dan berbagai *game* terbaik yang tidak *lag*. Meskipun aplikasi RCTI+ masih baru, namun konten dan ide kreatifnya mampu mencetak prestasi.

2.1.4 Web Scraping

Web scraping adalah teknik pengumpulan data melalui internet. Meskipun teknik ini tidak terlalu baru, saat ini sangat populer untuk pemenuhan data mining. Sebelum ini, teknik ini disebut dengan beberapa istilah, seperti *screen scraping*, *data mining*, *web harvesting*, dan teknik lain yang serupa. Pada dasarnya, *web scraping* merupakan proses pengambilan data menggunakan beragam metode melalui pemanfaatan API (*Application Programming Interface*). Cara ini biasanya dimulai dengan menulis kode program yang digunakan sebagai otomatisasi *query* untuk melakukan permintaan data ke server. Anda dapat mengekstraksi data dari hasil permintaan tersebut untuk mendapatkan informasi yang dicari. Manfaat *web scraping* adalah mengumpulkan informasi, yang membuat pencarian lebih mudah (Sahria, 2020).

2.1.5 Text Mining

Text mining adalah metode untuk mengumpulkan informasi berkualitas tinggi dari sejumlah besar data *unsupervised* terkait dengan topik tertentu. Teknik ini memiliki kemampuan untuk mengidentifikasi aspek emosional dari suatu pernyataan. Proses utama dalam *text mining* adalah untuk menemukan kata yang bisa mewakili isi konten dari dokumen. Setelah itu, proses pengkajian akan dilakukan untuk melihat bagaimana setiap dokumen berhubungan satu sama lain (Afdal & Rahma Elita, 2022).

2.1.6 *Preprocessing*

Preprocessing adalah tahapan yang merubah data mentah menjadi terstruktur guna mempermudah pemahaman, karena data mentah seringkali tidak memiliki format yang teratur. Proses ini perlu dilakukan untuk menghasilkan data yang berkualitas karena, data yang berkualitas buruk akan menghasilkan kualitas mining yang buruk pula. *Preprocessing* menghabiskan lebih banyak waktu daripada tahapan lain karena data harus dipersiapkan untuk digunakan dalam setiap algoritma *data mining* yang disarankan (Swastika et al., 2023). Dalam proses *preprocessing* dilakukan dengan beberapa tahap, yaitu:

1. *Cleaning*

Cleaning adalah sebuah langkah yang dilakukan untuk menghapus link, mention, hashtag, URL, tanda baca, angka, dan satu huruf. Tahap ini dilakukan agar data yang digunakan lebih rapi untuk tahap *preprocessing* teks selanjutnya.

2. *Case Folding*

Case folding merupakan langkah yang dilakukan untuk memudahkan pencarian dengan mengubah semua huruf kapital dalam dokumen menjadi huruf kecil.

3. *Tokenizing*

Tokenizing merupakan proses untuk menghilangkan spasi kosong, tanda baca, dan simbol non-huruf serta karakter dari setiap dokumen ulasan, serta memecah kalimat menjadi perkata.

4. *Stopwords Removal*

Stopwords Removal merupakan proses untuk menghapus kata – kata yang tidak memiliki arti penting atau tidak relevan.

5. Filter Tokens (by Length)

Filter Tokens (by Length) merupakan penghapusan kata – kata dengan panjang antara 4 hingga 25 karakter.

2.1.7 Pembobotan Kata

Pembobotan kata adalah Proses mengubah nilai suatu kata (kata, frase, atau unit indeks lainnya) dan memberikan indikator kata yang sesuai berdasarkan tingkat kepentingannya (Diandra Audiansyah et al., 2022). Salah satu teknik pembobotan kata yang umum digunakan adalah TF-IDF (*Term Frequency – Inverse Document Frequency*). TF-IDF menggabungkan konsep frekuensi adanya hubungan antar kata dan inverse frekuensi yang ada dalam dokumen untuk menentukan seberapa banyak hubungan antara kata dan dokumen yang akan diberikan bobot (Aditiya et al., 2022). Pembobotan kata TF mengacu pada frekuensi kemunculan kata dalam suatu dokumen, sedangkan pembobotan kata dengan IDF didasarkan pada frekuensi kemunculan kata tersebut dalam dokumen lain. Dalam pembobotan TF-IDF ini, kata yang jarang muncul diberi bobot lebih tinggi dibandingkan kata yang sering muncul (Devi A et al., 2021).

Metode TF-IDF memiliki rumus untuk menentukan bobot (W) setiap dokumen yang terkait dengan kata kunci. Rumus untuk menghitung bobot (W) adalah sebagai berikut:

$$W_{(t.d)} = Wtf_{(t.d)} \times idf_t$$

Keterangan:

$W_{(t.d)}$ = Bobot TF-IDF

$Wtf_{(t.d)}$ = Bobot kata dalam setiap dokumen

idf_t = Bobot *Inverse Document Frequency* dari nilai ($\log(N/df)$)

N = Jumlah seluruh dokumen

df = Jumlah seluruh dokumen yang mengandung *term*

2.1.8 *Support Vector Machine (SVM)*

Algoritma *Support Vector Machine (SVM)* didesain oleh Boser, Guyon, dan Vapnik. Algoritma ini pertama kali ditunjukkan kepada publik pada tahun 1992. Algoritma SVM dapat menemukan *hyperplane* terbaik dari ruang input melalui kombinasi teori – teori komputasi sebelumnya. Teori – teori ini termasuk *margin hyperplane* dan *kernel* yang dibuat oleh Aronszajn pada tahun 1950, serta konsep pendukung lainnya. *Hyperplane* adalah bidang yang membedakan kelas satu dari kelas lainnya. Prinsip dasar SVM adalah *classifier linear*, yang berarti hanya dapat digunakan untuk mengklasifikasi data antara dua kelas. Namun, karena kasus di dunia nyata biasanya lebih dari dua kelas, maka algoritma ini dikembangkan lebih lanjut untuk menangani masalah data linear atau non-linear dengan memasukkan konsep trik *kernel* dan menggunakan fungsi Φ untuk memetakan data ke dalam ruang yang besar (Hilda Kusumahadi et al., 2019).

Dalam SVM, titik ekstrim digunakan untuk membentuk *hyperplane*, titik inilah yang disebut sebagai *support vector*. Setiap elemen data dalam kumpulan data terdiri dari sebuah titik dalam ruang berdimensi n (n adalah jumlah fitur), dimana koordinatnya mewakili nilainya. Algoritma SVM digunakan untuk menemukan *hyperplane* yang membedakan kedua kelas secara jelas dan menyelesaikan masalah klasifikasi dan regresi. *Hyperplane* merupakan bidang pemisah antara satu kelas dengan kelas lainnya. Salah satu keunggulan SVM adalah bahwa mereka selalu dapat digunakan untuk kumpulan data pembelajaran yang diawasi (*supervised*). SVM dapat bekerja secara klasifikasi *non-linear* dengan memetakan *input* ke fitur dimensi tinggi (Purnamasari et al., 2023). Untuk menangani kasus *non-linear*, SVM diubah untuk memasukkan fungsi kernel, yang memungkinkan pencarian hasil dengan cepat (Pamungkas et al., 2021).

2.1.9 Confusion Matrix

Confusion matrix adalah instrumen yang dipergunakan untuk mengevaluasi kualitas model berdasarkan perbandingan antara hasil prediksi dengan data sebenarnya. Matrix ini terdiri dari empat kategori: *true positif (TP)*, *true negative (TN)*, *false positif (FP)*, dan *false negative (FN)*. TP dan TN menggambarkan data yang sama dalam kelas aktual dan prediksi, sedangkan FP dan FN menunjukkan data yang tidak sesuai dalam kelas aktual dan prediksi, sehingga dianggap sebagai kesalahan (Putri & Ridwan, 2023). Berikut adalah tabel dari *confusion matrix*:

Tabel II. 1
Confusion Matrix

<i>Prediction Class</i>	<i>Actual Class</i>	
	<i>Negative</i>	<i>Positive</i>
<i>Predict Negative</i>	TN	FN
<i>Predict Positive</i>	FP	TP

Keterangan:

- 1) TP (*True Positive*) = Jumlah data nilai aktual kelas positif dan nilai prediksi kelas positif
- 2) TN (*True Negative*) = Jumlah data nilai aktual kelas negatif dan nilai prediksi kelas negatif
- 3) FP (*False Positive*) = Jumlah data nilai aktual kelas negatif dan nilai prediksi kelas positif
- 4) FN (*False Negative*) = Jumlah data nilai aktual kelas positif dan nilai prediksi kelas negatif

Berdasarkan tabel 2.1, evaluasi dengan metode *confusion matrix* memberikan nilai untuk *accuracy*, *precision*, dan *recall* (Ginting et al., 2020).

1) *Accuracy*

Accuracy adalah ukuran yang menunjukkan seberapa sering model klasifikasi membuat prediksi yang benar. Dengan demikian, *accuracy* dapat memberikan gambaran umum tentang kualitas prediksi model tersebut. Berikut rumus perhitungannya:

$$Accuracy = \frac{(TP+TN)}{(TP+TN+FN+FP)}$$

2) *Precision*

Precision adalah metrik yang digunakan untuk mengukur kualitas prediksi positif. Ini menunjukkan seberapa akurat model dalam membuat prediksi positif. Berikut rumus perhitungannya:

$$Precision = \frac{TP}{(TP+FP)}$$

3) *Recall*

Recall adalah ukuran yang menilai seberapa baik model klasifikasi dapat menemukan indikasi positif yang sesungguhnya. *Recall* menunjukkan seberapa banyak indikasi positif yang dikenali model. Berikut rumus perhitungannya:

$$Recall = \frac{TP}{(TP+FN)}$$

2.1.10 *Python*

Python merupakan bahasa pemrograman yang diciptakan oleh Guido van Rossum pada tahun 1991. *Python* adalah bahasa pemrograman tingkat tinggi yang menggunakan semantik dinamis untuk meningkatkan keterbacaan kode atau *syntax* dan dapat melakukan eksekusi berbagai instruksi berorientasi objek secara langsung. Walaupun *python* dianggap sebagai bahasa pemrograman tingkat tinggi, tetapi *python*

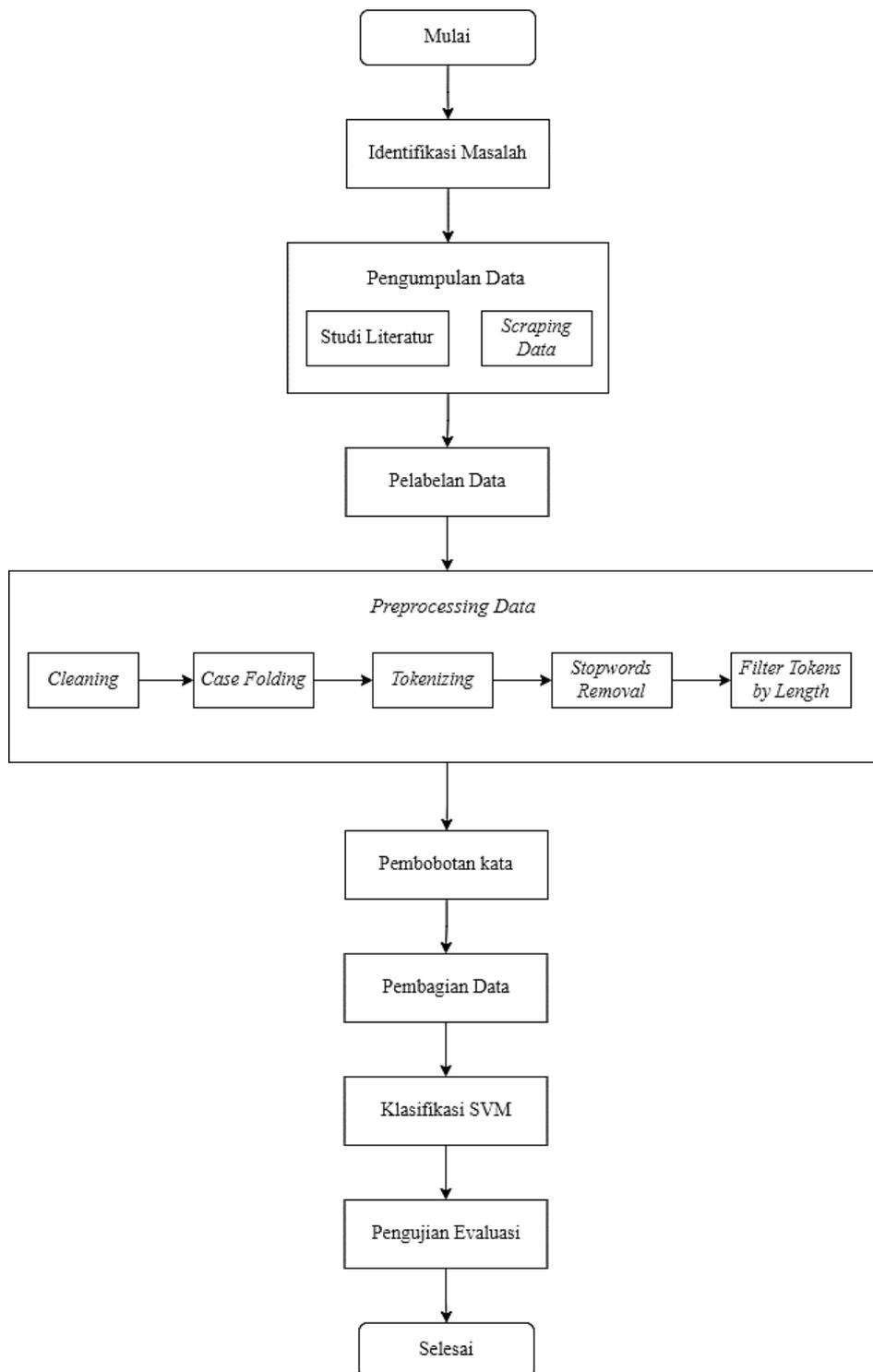
tetap dirancang sedemikian rupa untuk mudah dipahami dan dipelajari. *Python* dilengkapi dengan *library* yang memuat fungsi analisis data, *machine learning*, alat pra-prosesing data, serta alat visualisasi. Ketersediaan ini menjadikan *Python* sebagai bahasa pemrograman yang sangat populer dalam bidang ilmu data dan analisis (Pane & Saputra, 2020).

2.1.11 RapidMiner

RapidMiner merupakan sebuah alat yang dibuat oleh Dr. Markus Hofmann dan Ralf Klinkenberg dari Institut Teknologi Blanchardstown. RapidMiner adalah perangkat lunak berbasis *open source* yang memfasilitasi pemrograman berbasis *GUI* (*Graphical User Interface*). Perangkat lunak ini dikembangkan menggunakan bahasa pemrograman Java yang dapat berjalan di berbagai sistem operasi tanpa memerlukan keterampilan pemrograman khusus untuk menggunakannya (Vidiya & Testiana, 2023). *Platform* ini dirancang oleh perusahaan yang berfokus pada aplikasi yang melibatkan data besar dalam berbagai industri, termasuk bisnis, penelitian, pendidikan, pelatihan, dan pembelajaran. *Platform* ini menawarkan solusi pembelajaran yang meliputi pengelompokan, klasifikasi, dan analisis regresi. Selain itu, RapidMiner juga dapat membaca dan mengolah berbagai format file, termasuk .xls, .csv, dan lain – lain (Prasetyo et al., 2021).

2.2 Tahapan Penelitian

Tahapan yang akan dilakukan pada penelitian ini digambarkan dalam gambar 2.1 sebagai berikut:



Sumber : Penelitian

Gambar II. 1
Alur Tahapan Penelitian

2.2.1 Pengumpulan Data

Dalam proses penyusunan penelitian ini, peneliti mengumpulkan data dan informasi yang relevan untuk mendukung proses pengumpulan data yang terkait dengan penelitian sebagai berikut:

1. Studi Literatur

Dalam melakukan studi literatur, peneliti mengumpulkan dan mempelajari berbagai sumber ilmiah yang relevan dengan penulisan skripsi, termasuk teori analisis sentimen, *Support Vector Machine (SVM)*, serta penggunaan *Python* dalam analisis sentimen. Selain itu, peneliti juga menggunakan jurnal, *e-book*, dan sumber *website* resmi sebagai referensi, yang membantu dalam menyelesaikan penelitian ini.

2. *Scraping Data*

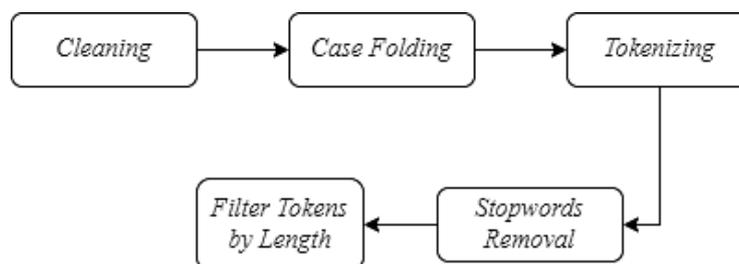
Ulasan aplikasi RCTI+ dikumpulkan dari *Google Play Store* dengan menggunakan teknik pengambilan data secara otomatis atau biasa disebut *scraping data*. *Scraping data* dilakukan menggunakan pemrograman bahasa *Python* yang dijalankan dengan bantuan *website Google Collab*. Data yang diperoleh kemudian disimpan dalam bentuk CSV.

2.2.2 Pelabelan Data

Setelah data berhasil dikumpulkan menggunakan teknologi *scraping*, selanjutnya setiap ulasan pengguna diberi label secara manual dan dikategorikan ke dalam dua jenis, yaitu sentimen positif dan sentimen negatif.

2.2.3 Preprocessing Data

Langkah selanjutnya setelah dilakukan pelabelan adalah *preprocessing data*. Proses ini menggunakan perangkat lunak RapidMiner untuk mengorganisir data yang semula tidak terstruktur menjadi data terstruktur dengan cara menghapus data yang tidak relevan. Tahapan yang digunakan peneliti untuk melakukan *preprocessing data* dapat dilihat pada gambar 2.2 dibawah ini:



Sumber : Penelitian

Gambar II. 2
Tahapan *Preprocessing Data*

1. *Cleaning* dilakukan untuk menghilangkan elemen yang tidak relevan seperti *noise* atau simbol pada teks.
2. *Case Folding* dilakukan untuk menyelaraskan penulisan huruf, dengan merubah huruf kapital menjadi huruf kecil (*lowercase*).
3. *Tokenizing* dilakukan untuk memecah kalimat menjadi potongan kata atau token untuk mengetahui dari mana kata tersebut berasal.
4. *Stopwords Removal* dilakukan untuk untuk menghilangkan kata dan frasa yang umum digunakan, namun tidak memengaruhi sentimen dalam kalimat.
5. *Filter Tokens (by Length)* dilakukan untuk menghapus kata – kata yang memiliki panjang antara 4 hingga 25 karakter.

2.2.4 Pembobotan Kata

Setelah melakukan *preprocessing data* pada teks, tahapan berikutnya yaitu pembobotan kata menggunakan metode TF-IDF. Metode TF-IDF akan menghitung nilai *Term Frequency* (TF) dan *Inverse Document Frequency* (IDF) untuk masing – masing token (kata) di setiap dokumen di dalam *Dataframe*. Pembobotan kata dengan metode TF-IDF akan dilakukan dengan operator *Process Document from Data*, Yang termasuk bagian dari *ekstens Text Processing* pada *software* RapidMiner.

2.2.5 Pembagian Data

Setelah melakukan pembobotan kata, data akan dipecah menjadi data latih dan data uji. Data diklasifikasikan dengan menggunakan operator *Split Data* dengan perbandingan rasio 80%:20%, yang mana 80% digunakan sebagai data latih dan 20% sebagai data uji.

2.2.6 Klasifikasi *Support Vector Machine* (SVM)

Setelah selesai melakukan pembagian data, tahap berikutnya yaitu klasifikasi menggunakan metode *Support Vector Machine* menggunakan aplikasi RapidMiner untuk memudahkan proses klasifikasi.

2.2.7 Pengujian Evaluasi

Setelah mendapatkan hasil dari klasifikasi *Support Vector Machine*, langkah terakhir adalah pengujian evaluasi. Pada tahap ini, dilakukan evaluasi untuk mengetahui seberapa efektif kinerja dari algoritma *Support Vector Machine* dalam proses klasifikasi. Perhitungan evaluasi ini menggunakan *confusion matrix* untuk menghitung nilai *accuracy*, *precision*, dan *recall*.

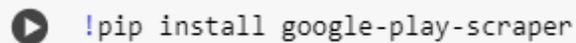
BAB III

HASIL DAN PEMBAHASAN

3.1 Proses Umum

3.1.1 Pengumpulan Data

Pengumpulan data pada penelitian ini menggunakan teknologi *scraping* dengan *library* “*google-play-scraper*” di *Google Colab* untuk mengumpulkan data ulasan aplikasi RCTI+ secara otomatis dari situs *Google Play Store*.



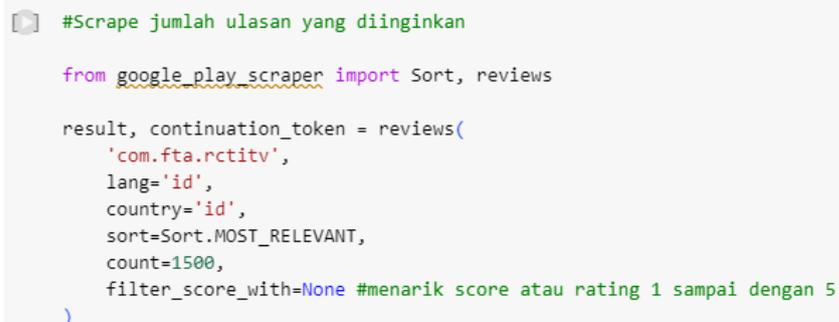
```
!pip install google-play-scraper
```

Sumber: Data olahan *Google Colab*

Gambar III. 1

Install library web scraping

Dalam proses pengumpulan data, *library Google-Play-Scraper* akan melakukan pengumpulan data berdasarkan parameter yang ditentukan. Peneliti memasukkan alamat *url id* RCTI+ pada *Google Play Store* seperti "com.fta.rctitv" untuk mengambil data dari aplikasi RCTI+, dengan hanya mengambil teks dalam Bahasa Indonesia. Pengumpulan data dilakukan dengan mengambil ulasan paling relevan terhadap aplikasi RCTI+ sebanyak 1500 ulasan.



```
#Scrape jumlah ulasan yang diinginkan

from google_play_scraper import Sort, reviews

result, continuation_token = reviews(
    'com.fta.rctitv',
    lang='id',
    country='id',
    sort=Sort.MOST_RELEVANT,
    count=1500,
    filter_score_with=None #menarik score atau rating 1 sampai dengan 5
)
```

Sumber: Data olahan *Google Colab*

Gambar III. 2

Proses scraping data

	userName	score	at	content
112	sri setiowati	5	2024-05-26 13:20:21	Anak aku suka banget doraemon dan sekarang tia...
339	diana putri	5	2024-05-26 11:39:02	Aku suka banget nonton doraemon, dan berkat ap...
66	Amalia amel	5	2024-05-26 11:30:30	Aplikasinya bagus banget, apalagi banyak kart...
331	Sriana Yulianti	5	2024-05-26 10:29:26	Akhirnya aku menemukan aplikasi live streaming...
342	JAL	5	2024-05-26 09:51:11	aplikasi yang cocok digunakan saat liburan non...

Sumber: Data olahan *Google Colab*

Gambar III. 3
Hasil *scraping data*

Setelah dataset berhasil dikumpulkan, data tersebut kemudian akan disimpan ke dalam sebuah file dengan ekstensi CSV yang sesuai dengan jumlah ulasan yang diinginkan.

```
[22] my_df.to_csv("rcti_dataset.csv", index = False)
```

Sumber: Data olahan Google Colab

Gambar III. 4
Penyimpanan dataset

3.1.2 Pelabelan Data

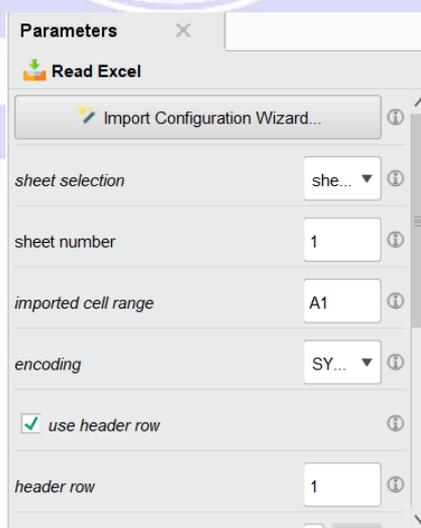
Pada tahap ini, data yang telah dikumpulkan sebelumnya akan diberikan label. Pelabelan data ini akan dilakukan secara manual oleh peneliti dengan menggunakan *Microsoft Excel 2019*. Kumpulan data ini diberi label untuk memeriksa komentar yang terdapat dalam ulasan aplikasi RCTI+ dan membagi dataset menjadi dua kategori, yaitu sentimen positif dan sentimen negatif. Setelah proses pelabelan data selesai, didapatkan hasil bahwa ada 948 data menunjukkan sentimen positif dan 552 data lainnya menunjukkan sentimen negatif.

Text	Sentimen
Langganan kok ada iklannya emang semuanya dimonopoli... gimana sih... Bukan masalah mengurangi vid	Negatif
Masih banyak iklan padahal berbayar	Negatif
Dikit2 bayar... nonton piala asia bayar, mau nonton bulutangkis jg harus bayar lg	Positif
wahhh gimana min udah bayar premium tetep gabisa nonton wkwk apk kocak lol	Negatif
Udah bayar langganan,,, muncul iklan :(Negatif
Tolong d perbaiki lagi masa bru d update belum lama sekarang d suruh update lagi	Negatif
Sebelumnya udh pernah downlod pas nonton g bayar kenapa sekarang nonton timnas bayar.sangat men	Negatif
Wkwkwkwk aplikasi payah udh bayar malah gabisa di tonton	Negatif
RCTI+ saya di hp bisa, tapi di smarttv kok gabisa diarahkan pakai visionplus, login akun yg sama Aku login	Negatif
Aplikasi butut nonton timnas aja gak bisa..uinstal butut butut	Negatif
Buat apa ada aplikasi ini, pengen nonton timnas sambil jualan lewat hp aja ga bisa, harus berlangganan d	Negatif
sy kira bs nonton free. ternyata hrs bayar paket. better nonton di tv	Negatif

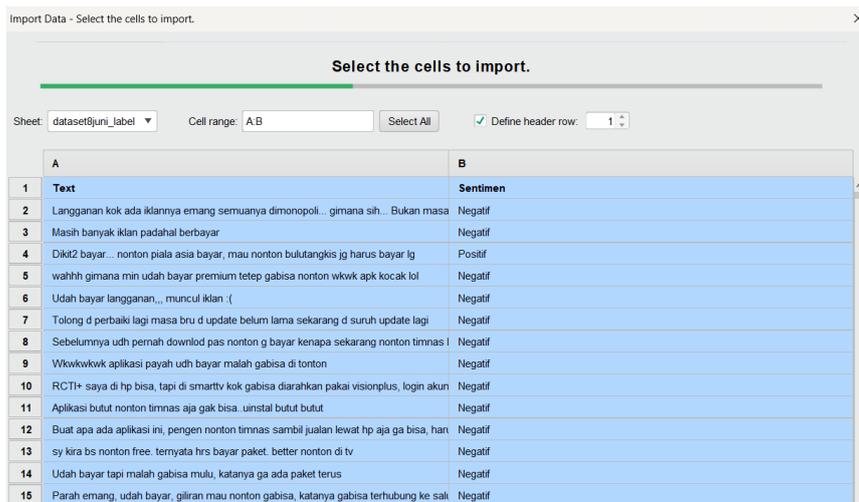
Gambar III. 5
Pelabelan data

3.1.3 Preprocessing Data

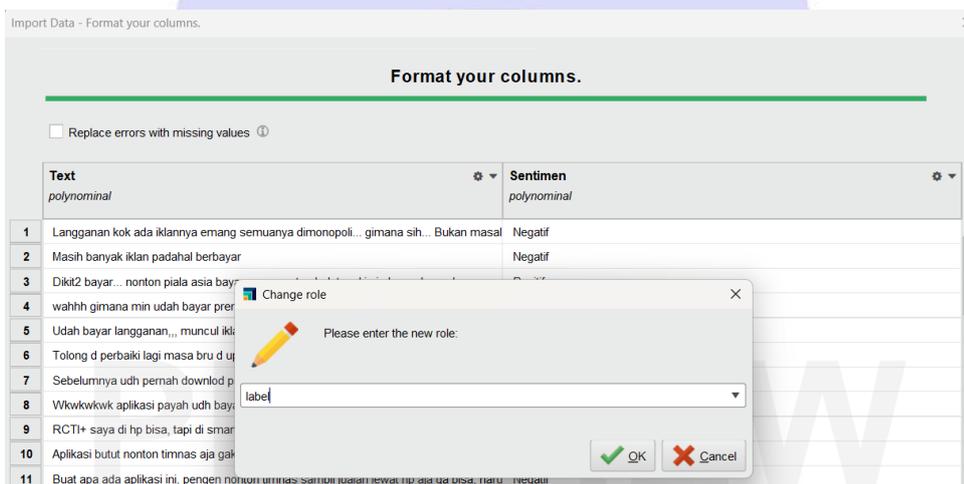
Sebelum memulai proses *preprocessing*, beberapa langkah penting harus dilakukan untuk mempersiapkan data. Pertama, dataset harus diimpor terlebih dahulu menggunakan operator *Read Excel*. Pada parameter *Read Excel*, untuk mengimpor data dapat dilakukan dengan mengklik “*Import Configuration Wizard*” dan pilih dataset yang akan digunakan. Selanjutnya, data yang telah di *import* akan diberikan *role* berupa “label” dan mengubah tipe data sentimen menjadi *binominal*. Hal ini karena sentimen pada data hanya memiliki dua kategori, yakni positif dan negatif. Berikut merupakan tampilan dari proses *import data*.



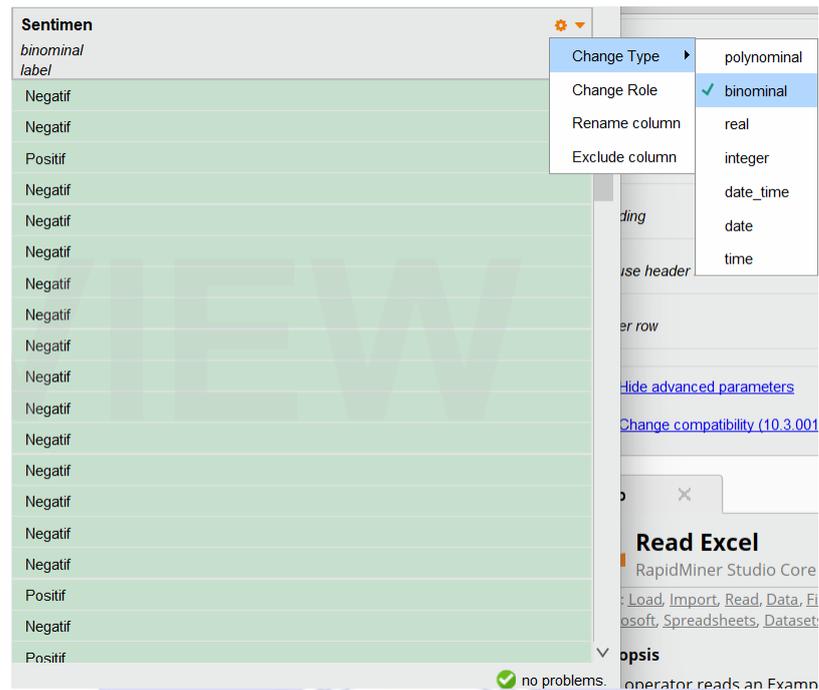
Gambar III. 6
Parameter *read excel*



Gambar III. 7
Hasil *import data*

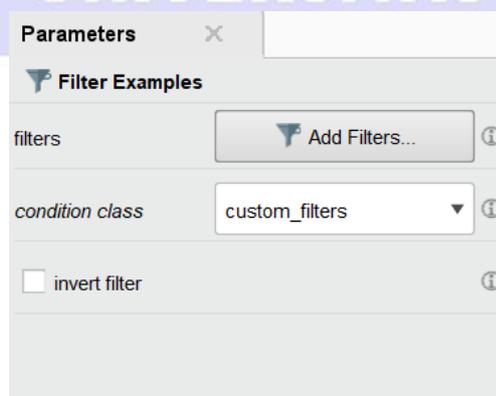


Gambar III. 8
Penentuan *role*

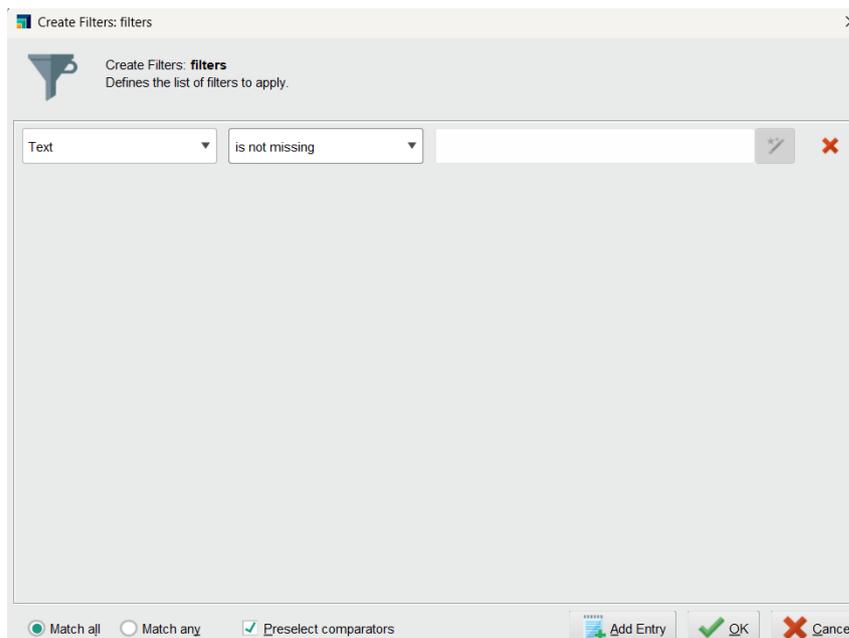


Gambar III. 9
Mengubah *type*

Setelah dataset selesai di *import*, langkah berikutnya adalah menggunakan operator *Filter Examples* untuk memilih data yang tidak bersifat *is not missing* pada *Text*. *Filter Examples* bertujuan untuk memilih dan menghapus contoh data yang tidak sesuai dengan kondisi yang ditentukan. Berikut adalah tampilan pada operator *Filter Examples*.

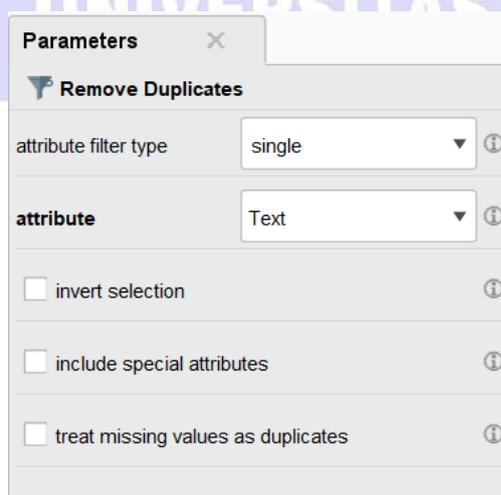


Gambar III. 10
Parameter *filter examples*



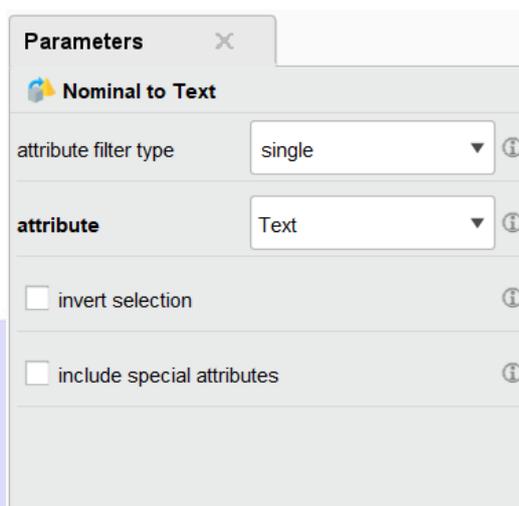
Gambar III. 11
Add filter

Setelah melakukan *Filter Examples*, langkah berikutnya melibatkan penggunaan operator *Remove Duplicates* untuk menghapus data yang sama atau ganda pada atribut *Text*. Dengan demikian, dataset menjadi lebih bersih dan efektif digunakan untuk analisis data dan proses *data mining*. Berikut merupakan tampilan dari operator *Remove Duplicates*.



Gambar III. 12
Parameter remove duplicates

Langkah selanjutnya adalah menggunakan operator *Nominal to Text* karena tidak semua data yang diambil memiliki format teks. Operator ini digunakan untuk mengubah tipe atribut *nominal* menjadi atribut teks. Berikut merupakan tampilan dari operator *Nominal to Text*.

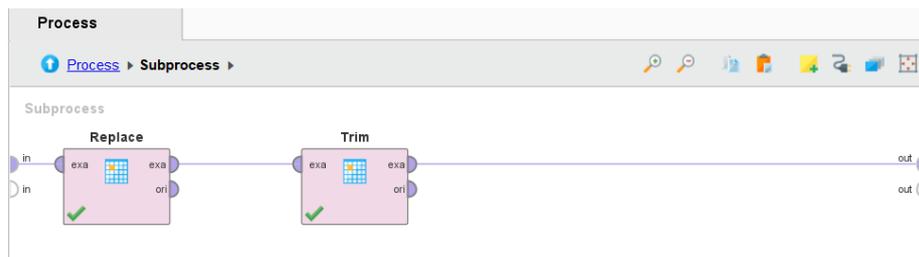


Gambar III. 13
Parameter *nominal to text*

Setelah melewati beberapa tahapan sebelumnya, maka langkah selanjutnya adalah melakukan *Preprocessing* pada dataset. Proses ini memiliki 5 tahapan, yaitu *Cleansing*, *Case Folding*, *Tokenizing*, *Stopwords Removal*, dan *Filter tokens (by length)*.

1. *Cleaning*

Pada tahap ini, proses *cleaning* dilakukan untuk menghapus *noise* atau karakter yang tidak diinginkan. Dalam proses ini, peneliti menggunakan operator *Replace* dan *Trim* dan mengintegrasikannya kedalam operator *subprocess*. Pada parameter operator *Replace* dan *Trim*, *atribut filter type* yang digunakan adalah *single*, dan *attribute* yang diisi adalah *Text*. Untuk parameter *Replace*, *replace what* diisi dengan komponen *Regular Expression* seperti *icon* [-!"#%&'()*+,-./:;<=>?@\$%\\$\$_`{|}~].



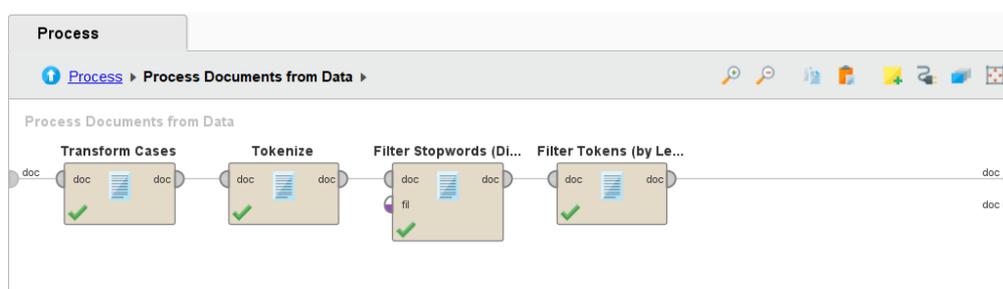
Gambar III. 14
Proses *cleaning*

Row No.	Sentimen	Text
1	Negatif	Langganan kok ada iklannya emang semuanya dimonopoli gimana sih Bukan masalah mengurangi video tapi mengganggu siaran fungsinya langganan kan biar nyama
2	Negatif	Masih banyak iklan padahal berbayar
3	Positif	Dikit2 bayar nonton piala asia bayar mau nonton bulutangkis jg harus bayar lg
4	Negatif	wahhh gimana min udah bayar premium tetep gabisa nonton wkwk apk kocak lol
5	Negatif	Udah bayar langganan muncul iklan
6	Negatif	Tolong d perbaiki lagi masa bru d update belum lama sekarang d suruh update lagi
7	Negatif	Sebelumnya udh pernah downlo d pas nonton g bayar kenapa sekarang nonton timnas bayarsangat mengecewakan
8	Negatif	Wkwkwkwk aplikasi payah udh bayar malah gabisa di tonton
9	Negatif	RCTI saya di hp bisa tapi di smarttv kok gabisa diarahkan pakai visionplus login akun yg sama Aku login pakai RCTI gabisa Gimana sih gajelas banget Udah bayar mah
10	Negatif	Aplikasi butut nonton timnas aja gak bisa instal butut butut

Gambar III. 15
Hasil *cleaning*

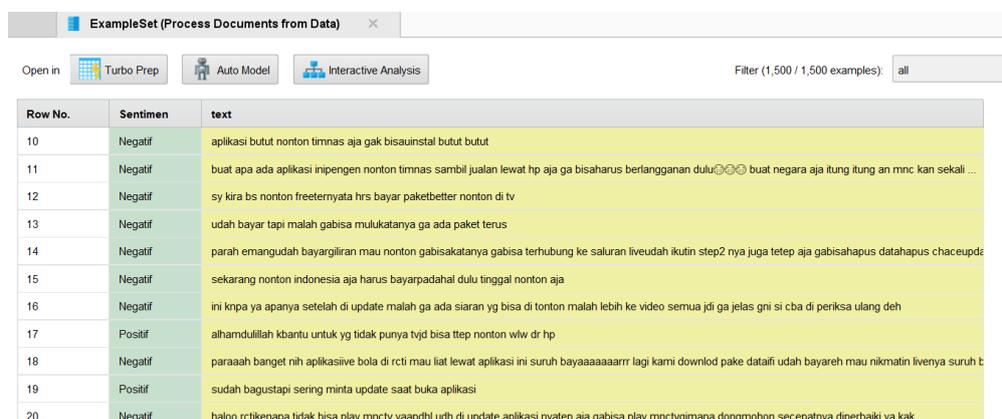
2. Case Folding

Pada tahap ini, proses *case folding* dilakukan untuk mengubah semua huruf besar menjadi huruf kecil (*lower case*). Saat melaksanakan proses *case folding*, *tokenizing*, *stopwords removal*, dan *filter tokens by length*, peneliti menggunakan sebuah operator yang disebut "*Process Documents from Data*" untuk menerapkan semua langkah – langkah tersebut secara bersamaan. Berikut merupakan tampilan di dalam operator *Process Documents from Data*.



Gambar III. 16
Sub-Operator *process document from data*

Untuk proses *case folding*, peneliti menggunakan operator *Transform Case*, dengan parameter *transform to* diubah menjadi *lower case*. Berikut adalah hasil dari proses *case folding*.

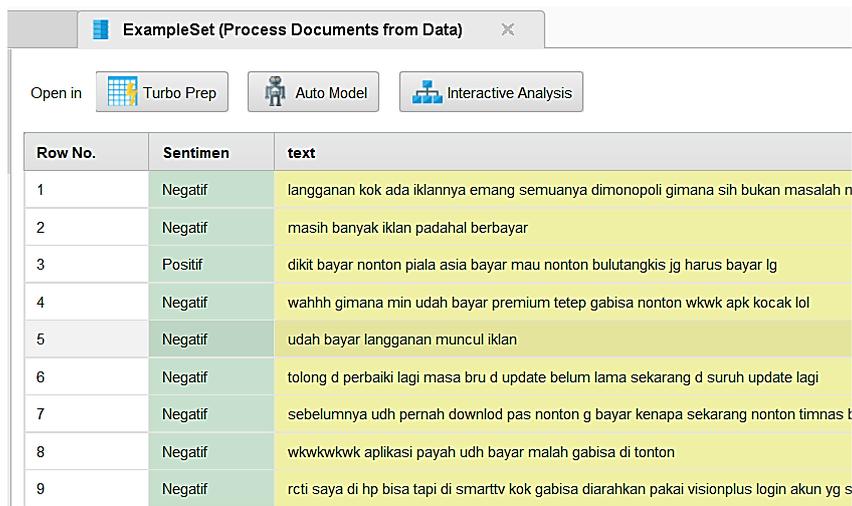


Row No.	Sentimen	text
10	Negatif	aplikasi butut nonton timnas aja gak bisa instal butut butut
11	Negatif	buat apa ada aplikasi impengen nonton timnas sambil jualan lewat hp aja ga bisaharus berlangganan dulu🙄🙄🙄 buat negara aja itung itung an mnc kan sekali ...
12	Negatif	sy kira bs nonton freeternyata hrs bayar paketbetter nonton di tv
13	Negatif	udah bayar tapi malah gabisa mulukatnya ga ada paket terus
14	Negatif	parah emangudah bayargiliran mau nonton gabisakatanya gabisa terhubung ke saluran liveudah ikutin step2 nya juga tetep aja gabisahapus datahapus chaceupda
15	Negatif	sekarang nonton indonesia aja harus bayarpadahal dulu tinggal nonton aja
16	Negatif	ini knpa ya apanya setelah di update malah ga ada siaran yg bisa di tonton malah lebih ke video semua jdi ga jelas gni si cba di periksa ulang deh
17	Positif	alhamdulillah kbantu untuk yg tidak punya tvjd bisa ttep nonton wlv dr hp
18	Negatif	paraaah banget nih aplikasiive bola di rcti mau liat lewat aplikasi ini suruh bayaaaaaaarrr lagi kami downlo d pake datafi udah bayareh mau nikmatin livenya suruh b
19	Positif	sudah bagustapi sering minta update saat buka aplikasi
20	Negatif	haloo rctikenapa tidak bisa play mnc tv yaapdhi udh di update aplikasi nyatep aja gabisa play mnc tv gimana dongmohon secepatnya diperbaiki ya kak

Gambar III. 17
Hasil *case folding*

3. *Tokenizing*

Proses *tokenizing* merupakan proses pemisahan setiap kata dari dokumen yang akan digunakan. Tujuannya untuk menghilangkan spasi kosong, tanda baca, dan simbol non-huruf serta karakter dari setiap dokumen ulasan, serta memecah kalimat menjadi perkata. Pada proses ini, peneliti menggunakan operator *Transform Case* dengan parameter menggunakan mode *non-letters*.



Row No.	Sentimen	text
1	Negatif	langganan kok ada iklannya emang semuanya dimonopoli gimana sih bukan masalah n
2	Negatif	masih banyak iklan padahal berbayar
3	Positif	dikit bayar nonton piala asia bayar mau nonton bulutangkis jg harus bayar lg
4	Negatif	wahhh gimana min udah bayar premium tetep gabisa nonton wkwk apk kocak lol
5	Negatif	udah bayar langganan muncul iklan
6	Negatif	tolong d perbaiki lagi masa bru d update belum lama sekarang d suruh update lagi
7	Negatif	sebelumnya udh pernah downlo d pas nonton g bayar kenapa sekarang nonton timnas t
8	Negatif	wkwkwkwk aplikasi payah udh bayar malah gabisa di tonton
9	Negatif	rcti saya di hp bisa tapi di smarttv kok gabisa diarahkan pakai visionplus login akun yg s

Gambar III. 18
Hasil *tokenizing*

4. *Stopwords Removal*

Proses *stopwords removal* dilakukan untuk menghilangkan kata – kata di dalam dataset yang tidak mempunyai arti atau tidak relevan pada teks. Kata – kata yang umumnya dihilangkan melalui proses ini adalah kata ganti, kata sambung, dan kata – kata yang tidak menyampaikan informasi penting. Dalam proses ini, peneliti menggunakan operator *Filter Stopwords (Dictionary)* pada RapidMiner dan parameter yang digunakan diambil dari file *stopwords* Bahasa Indonesia yang telah disiapkan oleh peneliti. Berikut merupakan hasil dari proses *stopwords removal*.

Row No.	Sentimen	text
1	Negatif	langganan iklannya emang dimonopoli gimana sih mengurangi video mengganggu siaran fungsinya langga
2	Negatif	iklan berbayar
3	Positif	dikit bayar nonton piala asia bayar nonton bulutangkis jg bayar lg
4	Negatif	wahhh gimana min udah bayar premium tetep gabisa nonton wkwk apk kocak lol
5	Negatif	udah bayar langganan muncul iklan
6	Negatif	tolong d perbaiki bru d update d suruh update
7	Negatif	udh downlod pas nonton g bayar nonton timnas bayarsangat mengecewakan
8	Negatif	wkwkwkwk aplikasi payah udh bayar gabisa tonton
9	Negatif	rcti hp smarttv gabisa diarahkan pakai visionplus login akun yg login pakai rcti gabisa gimana sih gajelas ba

Gambar III. 19
Hasil *stopwords removal*

5. *Filter Tokens (by length)*

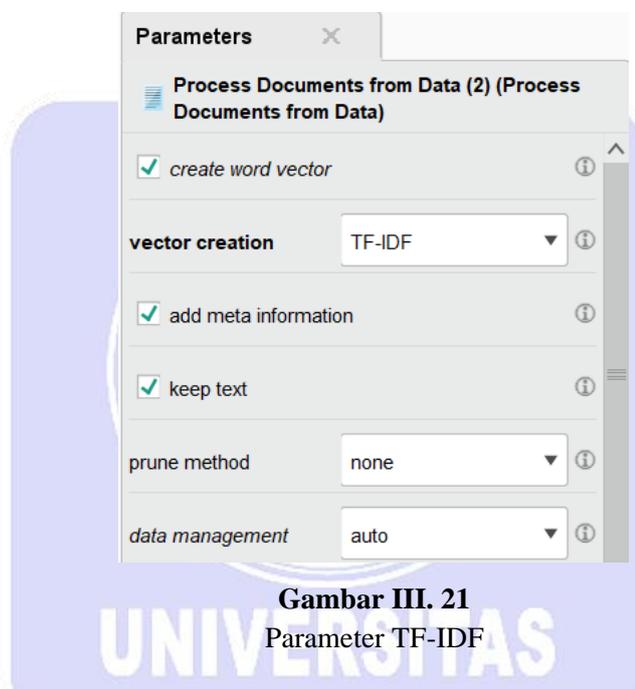
Proses *filter tokens (by length)* dilakukan untuk menghilangkan kata – kata dengan jumlah karakter dibawah ambang yang ditentukan, sehingga kata – kata yang dihasilkan pada dataset tidak memiliki kata yang terlalu panjang atau terlalu pendek. Pada proses ini, peneliti memanfaatkan operator *Filter Tokens (by Length)*, dengan pengaturan parameter yang digunakan yaitu minimal 4 karakter hingga 25 karakter. Berikut merupakan hasil dari proses *filter tokens (by length)*.

Row No.	Sentimen	text
1	Negatif	langganan iklannya emang dimonopoli gimana mengurangi video mengganggu siaran fungsinya langganan biar nyaman
2	Negatif	iklan berbayar
3	Positif	dikit bayar nonton piala asia bayar nonton bulutangkis bayar
4	Negatif	wahhh gimana udah bayar premium tetep gabisa nonton wkwk kocak
5	Negatif	udah bayar langganan muncul iklan
6	Negatif	tolong perbaiki update suruh update
7	Negatif	downlod nonton bayar nonton timnas bayarsangat mengecewakan
8	Negatif	wkwkwkwk aplikasi payah bayar gabisa tonton
9	Negatif	rcti smarttv gabisa diarahkan pakai visionplus login akun login pakai rcti gabisa gimana gajelas banget udah bayar mahal

Gambar III. 20
Hasil *filter tokens (by length)*

3.1.4 Pembobotan Kata

Setelah melalui tahap *preprocessing*, selanjutnya dataset akan dilakukan pembobotan kata. Dalam proses ini, setiap kata akan dinilai berdasarkan bobotnya menggunakan metode TF-IDF untuk menghitung pentingnya kata dalam dokumen. Untuk menerapkan TF-IDF, operator "*Process Documents from Data*" digunakan, dengan memilih opsi "*Create Word Vector*" dan mengaktifkan fitur TF-IDF untuk menghitung bobot kata.



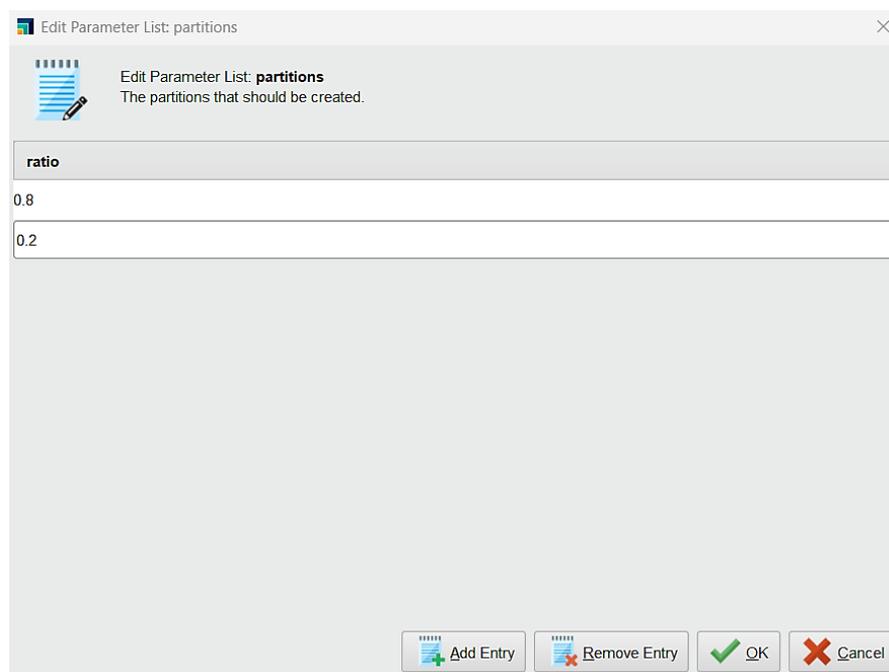
Gambar III. 21
Parameter TF-IDF

Sentimen	text	aamiin	aapalah	abal	abis	absen	aca
Positif	aplikasi acara sngt lengkap suka liat sinetron tukang bubur pengen haji film kartunnya	0	0	0	0	0	0
Positif	aplikasi terbaik nonton streaming cocok gapunya tontonan lengkap recomended dehi	0	0	0	0	0	0
Positif	suka banget aplikasinya nonton sinetron dimana suka banget sinetron mencintaimu	0	0	0	0	0	0
Negatif	minggu beli paket premium sinetron abis update sinetron akses	0	0	0	0.426	0	0
Negatif	update mulu update gajelas gaada perubahan signifikan jugaa update	0	0	0	0	0	0
Negatif	mengeluhkan update rcti iklan nonton tukang bubur hajipadahal episode ganti episod	0	0	0	0	0	0
Positif	gacor parah aplikasi keren banget notabene suka nonton family spongebob squarep	0	0	0	0	0	0
Positif	aplikasi streaming satunini terbaik mudah aplikasi pengalaman terbaik menonton ap	0	0	0	0	0	0
Negatif	error gimana beli premium pake suruh beli premium	0	0	0	0	0	0
Positif	suka banget acara rcti sinetron animasi super duper keren abis sinetron tukang bub	0	0	0	0.284	0	0
Negatif	kecewa jadwal sesuai tayangan doraemon kungfu panda tayangin gimana	0	0	0	0	0	0
Negatif	tampilannya berubah suka banget koneksi kvision	0	0	0	0	0	0

Gambar III. 22
Hasil pembobotan TF-IDF

3.1.5 *Split Data*

Sebelum menerapkan algoritma klasifikasi *Support Vector Machine*, dataset akan dibagi menjadi dua bagian, yaitu data latih dan data uji. Dalam langkah ini, peneliti memanfaatkan operator *Split Data* untuk membagi dataset dengan perbandingan 80% untuk data latih dan 20% untuk data uji. Dari pembagian rasio tersebut, menghasilkan 1200 data latih dan 300 data uji.

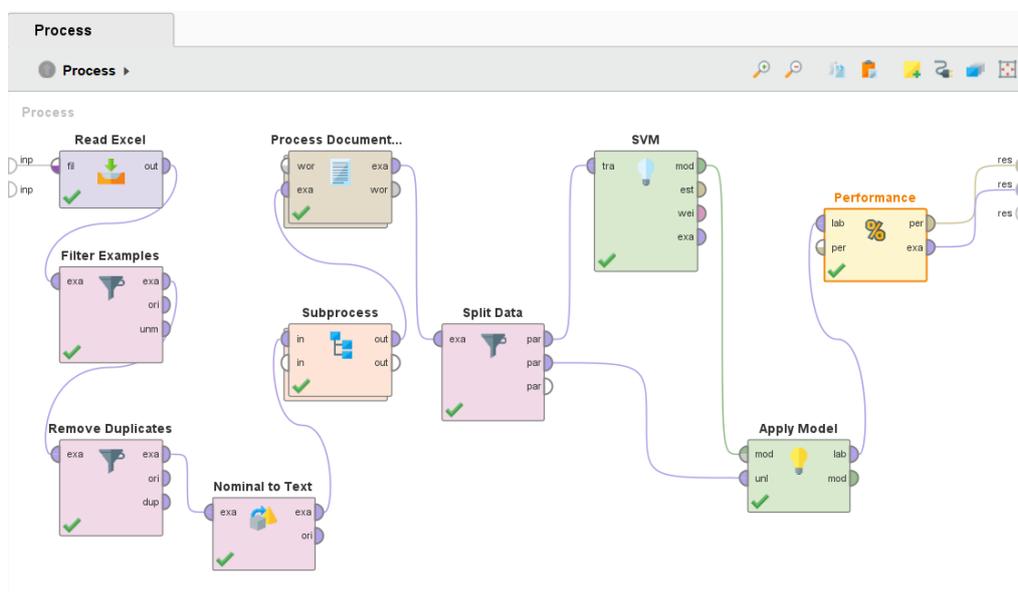


Gambar III. 23
Rasio split data

3.1.6 *Klasifikasi Support Vector Machine*

Setelah melakukan pembagian data latih dan data uji, langkah berikutnya yaitu mengimplementasikan algoritma *Support Vector Machine* di RapidMiner dengan cara menghubungkan tiga operator, yaitu *Support Vector Machine*, *Apply Model*, dan *Performance (classification)*. Dalam proses ini, data *training* akan diklasifikasikan menggunakan operator *Support Vector Machine* dan hasilnya akan dimasukkan ke dalam operator *Apply Model* untuk menghasilkan prediksi bersama dengan data uji. Kemudian, Operator *Performance* akan menilai model klasifikasi yang telah dibuat

dan memperlihatkan tingkat kinerjanya. Berikut merupakan tampilan keseluruhan dari proses klasifikasi algoritma *Support Vector Machine* pada RapidMiner.



Gambar III. 24
Proses implementasi SVM

Table View Plot View

accuracy: 90.33%

	true Negatif	true Positif	class precision
pred. Negatif	92	11	89.32%
pred. Positif	18	179	90.86%
class recall	83.64%	94.21%	

Gambar III. 25
Hasil implementasi SVM

3.1.7 Evaluasi Model

Setelah mendapatkan hasil dari klasifikasi menggunakan algoritma *Support Vector Machine*, maka langkah selanjutnya yaitu mengevaluasi model dengan menghitung nilai *confusion matrix* untuk mencari nilai *accuracy*, *precision*, dan *recall*. Data yang digunakan pada tahap ini diambil dari data uji sebanyak 300 data, terdiri dari 190 data dengan sentimen positif dan 110 data dengan sentimen negatif. Berdasarkan hasil pada Gambar 3.25 diatas, penerapan algoritma *support vector*

machine menghasilkan nilai *accuracy* sebesar 90,33% dengan pembagian data latih dan data uji dalam perbandingan 80%:20%. Langkah selanjutnya akan dilakukan perhitungan nilai *accuracy*, *precision*, dan *recall* berdasarkan tabel *confusion matrix* dibawah ini.

Tabel III. 1
Confusion Matrix

	<i>True Negatif</i>	<i>True Positif</i>	<i>Precision</i>
Pred. Negatif	92	11	89,32%
Pred. Positif	18	179	90,86%
Class Recall	83,64%	94,21%	

Berikut merupakan rumus perhitungan *confusion matrix*:

$$1. \text{ Accuracy} = \frac{(TP+TN)}{(TP+TN+FN+FP)} = \frac{(179+92)}{(179+92+11+18)} = \frac{271}{300} = 0,90$$

$$2. \text{ Precision} = \frac{TP}{(TP+FP)} = \frac{179}{(179+18)} = \frac{179}{197} = 0,90$$

$$3. \text{ Recall} = \frac{TP}{(TP+FN)} = \frac{179}{(179+11)} = \frac{179}{190} = 0,94$$

Berdasarkan hasil perhitungan evaluasi model menggunakan *confusion matrix* diatas, didapatkan nilai *accuracy* yang menggambarkan seberapa akurat penggunaan model SVM untuk klasifikasi data sebesar 90,33%, *precision* yang menggambarkan hasil dari prediksi positif yang benar sebesar 90,86%, dan *recall* yang menggambarkan sensitivitas model terhadap sampel positif sebesar 94,21%.

BAB IV

PENUTUP

4.1 Kesimpulan

Berdasarkan hasil penelitian dan pengujian yang telah dilakukan oleh peneliti, maka dapat diambil kesimpulan sebagai berikut:

1. Dalam proses pengumpulan data ulasan RCTI+ pada halaman *reviews* di *Google Play Store*, teknik *web scraping* diterapkan menggunakan *library* pemrograman *python* dan *website Google Colab*, diperoleh dataset sebanyak 1500 data berdasarkan ulasan yang relevan. Data ulasan tersebut kemudian terbagi menjadi dua kelas setelah dilakukan pelabelan secara manual, didapatkan 948 sentimen positif dan 552 sentimen negatif.
2. Tahapan pengolahan data dimulai dengan tahapan *preprocessing*, yang meliputi beberapa langkah utama seperti *Cleaning*, *Case Folding*, *Tokenizing*, *Stopword Removal*, dan *Filter Tokens (by Length)*. Setelah itu, tahapan pembobotan kata TF-IDF digunakan untuk melakukan konversi kata – kata menjadi vektor numerik yang dapat diterapkan pada model. Kemudian diimplementasikan dengan menggunakan metode *Support Vector Machine*, dengan pembagian data latih dan data uji sebesar 80%:20%. Dari pembagian tersebut, didapatkan data latih sebanyak 1200 dan data uji sebanyak 300.
3. Dalam penerapan algoritma klasifikasi *Support Vector Machine*, akan digunakan *confusion matrix* untuk mengevaluasi kinerja model tersebut. Tahap evaluasi ini menggunakan 300 data uji, dengan rincian 190 data memiliki sentimen positif dan 110 data lainnya memiliki sentimen negatif.

4. Berdasarkan hasil evaluasi model menggunakan *confusion matrix*, didapatkan nilai *accuracy* sebesar 90,33%, *precision* sebesar 90,86%, dan *recall* sebesar 94,21%. Hasil tersebut menunjukkan bahwa algoritma *Support Vector Machine* efektif dan akurat dalam mengklasifikasikan sentimen ulasan aplikasi RCTI+ di *Google Play Store*.

Dari hasil pengujian evaluasi model tersebut, dapat disimpulkan bahwa aplikasi RCTI+ lebih banyak mendapatkan ulasan positif sebesar 948 data dari pengguna di *Google Play Store*. Meskipun masih terdapat beberapa kritik, ulasan positif yang lebih banyak dan dominan menunjukkan bahwa aplikasi RCTI+ telah memberikan layanan yang memuaskan bagi sebagian besar pengguna di *Google Play Store*.

4.2 Saran

Adapun saran peneliti untuk penelitian selanjutnya:

1. Penelitian selanjutnya dapat memperluas dataset dengan ulasan dari platform lain untuk mendapatkan hasil yang lebih komprehensif.
2. Menambahkan sentimen netral pada dataset.
3. Melakukan pelabelan sentiment secara otomatis dengan metode pendekatan seperti *lexicon-based*, dan lain - lain.
4. Menambahkan atau mengurangi tahap *preprocessing* karna dapat mempengaruhi jumlah kata yang didapat.
5. Menggunakan algoritma klasifikasi lain, seperti *naïve bayes*, *Decision Tree*, *Random Forest Tree*, dan lainnya untuk membandingkan performa dengan metode *Support Vector Machine*.

6. Evaluasi lebih lanjut dapat dilakukan dengan menambahkan metrik evaluasi seperti *F1-score* dan *ROC-AUC* untuk memberikan gambaran yang lebih mendalam tentang kinerja model.
7. Menggunakan data ulasan selain Bahasa Indonesia.

DAFTAR PUSTAKA

- Aditiya, P., Enri, U., & Maulana, I. (2022). Analisis Sentimen Ulasan Pengguna Aplikasi Myim3 Pada Situs *Google Play* Menggunakan *Support Vector Machine*. *JURIKOM (Jurnal Riset Komputer)*, 9(4), 1020. <https://doi.org/10.30865/jurikom.v9i4.4673>
- Afdal, M., & Rahma Elita, L. (2022). Penerapan Text Mining Pada Aplikasi Tokopedia Menggunakan Algoritma K-Nearest Neighbor. *Jurnal Ilmiah Rekayasa Dan Manajemen Sistem Informasi*, 8(1).
- Devi A, A. P., A, L., A, A. Z., Maryamah, L, R., & S, R. W. (2021). Metode Pembobotan Kata Berbasis Cluster Untuk Perangkingan Dokumen Berbahasa Arab Cluster-Based Term Weighting Method for Ranking Arabic Documents. In *Techno.COM* (Vol. 20, Issue 2).
- Diandra Audiansyah, D., Eka Ratnawati, D., & Trias Hanggara, B. (2022). Analisis Sentimen Aplikasi MyXL menggunakan Metode Support Vector Machine berdasarkan Ulasan Pengguna di Google Play Store. *Jurnal Pengembangan Teknologi Informasi Dan Ilmu Komputer*, 6(8), 3987–3994. <http://j-ptiik.ub.ac.id>
- Diki Hendriyanto, M., Ridha, A. A., & Enri, U. (2022). Analisis Sentimen Ulasan Aplikasi Mola Pada Google Play Store Menggunakan Algoritma Support Vector Machine. *Journal of Information Technology and Computer Science (INTECOMS)*, 5(1).
- Ginting, V. S., Kusriani, & Taufiq, E. (2020). Implementasi Algoritma C4.5 Untuk Memprediksi Keterlambatan Pembayaran Sumbangan Pembangunan Pendidikan Sekolah Menggunakan Python. *Inspiration : Jurnal Teknologi Informasi Dan Komunikasi*, 10(1), 36–44.
- Herlinawati, N., Yuliani, Y., Faizah, S., & Gata, W. (2020). Analisis Sentimen Zoom Cloud Meetings Di Play Store Menggunakan Naïve Bayes Dan Support Vector Machine. *CESS (Journal of Computer Engineering System and Science)*, 5(2), 2502–2714.
- Hilda Kusumahadi, S., Junaedi, H., & Santoso, J. (2019). Klasifikasi Helpdesk Menggunakan Metode Support Vector Machine. *Jurnal Informatika: Jurnal Pengembangan IT*, 4(1), 54–60. <https://doi.org/10.30591/jpit.v4i1.1125>
- Natasuwarna, A. P. (2019). Tantangan Menghadapi Era Revolusi 4.0 - Big Data dan Data Mining. *Seminar Nasional Hasil Pengabdian Kepada Masyarakat 2019*, 29.
- Pamungkas, B., Purbaya, M. E., & Januarita, D. (2021). Analisis Sentimen Twitter Menggunakan Metode Support Vector Machine (SVM) pada Kasus Benih Lobster 2020. *Journal of Informatics, Information System, Software Engineering and Applications*, 3(2). <https://doi.org/10.20895/INISTA.V3I2>

- Pane, S. F., & Saputra, Y. A. (2020). *Big Data: Classification Behavior Menggunakan Python* (Y. A. Saputra, Ed.; 1st ed.). Kreatif Industri Nusantara.
- Prasetyo, V. R., Lazuardi, H., Mulyono, A. A., & Lauw, C. (2021). Penerapan Aplikasi RapidMiner Untuk Prediksi Nilai Tukar Rupiah Terhadap US Dollar Dengan Metode Linear Regression. *Jurnal Nasional Teknologi Dan Sistem Informasi*, 7(1), 8–17. <https://doi.org/10.25077/teknosi.v7i1.2021.8-17>
- Purnamasari, D., Bayu, A., Desy, A., Fanka, W. A. P., Reza, A., Safrila, M., Yanda, O. N., & Hidayati, U. (2023). *Pengantar Metode Analisis Sentimen*.
- Putra, R. R., & Hidayat, Z. (2022). Komunikasi Pemasaran Layanan Video Streaming Dan On Demand Mnc Group (Studi Kasus: Aplikasi Rcti+). *Jurnal Ilmu Sosial Dan Pendidikan (JISIP)*, 6(1), 2598–9944. <https://doi.org/10.36312/jisip.v6i1.2813/>
- Putra, Y. W. S., Dawis, A. M., Novi, N., Natsir, F., & Fitria, F. (2023). *Pengantar Aplikasi Mobile* (E. Erlangga & W. Andriyani, Eds.). Penerbit Haura Utama.
- Putri, D. S., & Ridwan, T. (2023). Analisis Sentimen Ulasan Aplikasi Pospay dengan Algoritma Support Vector Machine. *Jurnal Ilmiah Informatika*.
- Sahria, Y. (2020). Implementasi Teknik Web Scraping pada Jurnal SINTA Untuk Analisis Topik Penelitian Kesehatan Indonesia. *Prosiding University Research Colloquium*. <http://sinta2.ristekdikti.go.id/journals/detail>
- Swastika, R., Mukodimah, S., Susanto, F., Muslihudin, M., & Ipinuwati, S. (2023). *Implementasi Data Mining (Clustering, Association, Prediction, Estimation, Classification)* (D. Susianto & Suyono, Eds.; 1st ed.). Penerbit Adab.
- Vidiya, E. C., & Testiana, G. (2023). Analisis Pola Pembelian di Lathansa Cafe & Ramen dengan Menggunakan Algoritma FP-Growth Berbantuan RapidMiner. *G-Tech: Jurnal Teknologi Terapan*, 7(3), 1118–1126. <https://doi.org/10.33379/gtech.v7i3.2739>
- Wahyudi, R., Kusumawardhana, G., Purwokerto, A., Letjend, J., Soemarto, P., Purwanegara, K., Purwokerto, T., & Banyumas, K. (2021). Analisis Sentimen pada review Aplikasi Grab di Google Play Store Menggunakan Support Vector Machine. *Jurnal Informatika*, 8(2). <http://ejournal.bsi.ac.id/ejournal/index.php/ji>
- Widayat, W. (2021). Analisis Sentimen Movie Review menggunakan Word2Vec dan metode LSTM Deep Learning. *Jurnal Media Informatika Budidarma*, 5(3), 1018. <https://doi.org/10.30865/mib.v5i3.3111>
- Samantri, M., & Afiyati. (2024). Perbandingan Algoritma Support Vector Machine dan Random Forest untuk Analisis Sentimen Terhadap Kebijakan Pemerintah Indonesia Terkait Kenaikan Harga BBM Tahun 2022. *Jurnal JTik (Jurnal Teknologi Informasi Dan Komunikasi)*, 8(1), 1–9. <https://doi.org/10.35870/jtik.v8i1.1202>

DAFTAR RIWAYAT HIDUP

I. Biodata Mahasiswa

NIM : 19200496
Nama Lengkap : Annisa Ghina Fauziyyah
Tempat/ Tanggal Lahir : Tangerang, 15 Juli 2001
Alamat Lengkap : Perumahan Taman Walet Blok SA 11 No 16 RT03
RW11 Kel. Sindangsari Kec. Pasar Kemis Kab.
Tangerang Banten, 15560

II. Pendidikan

a. Formal

1. SDS Permata Sari 1, lulus tahun 2012
2. SMP Negeri 4 Pasar Kemis, lulus tahun 2016
3. SMK Negeri 2 Kabupaten Tangerang, lulus tahun 2019

III. Riwayat Pengalaman berorganisasi / pekerjaan

1. Bendahara Karang Taruna RT03/RW11. Tahun 2020
2. Admin Online, Mustika Collection. Tahun 2020
3. Magang Quality Control, PT Surya Toto Indonesia. Tahun 2023



Jakarta, 20 Agustus 2024

Annisa Ghina Fauziyyah

SURAT PERNYATAAN KEBENARAN/KEABSAHAN DATA

HASIL RISET UNTUK KARYA ILMIAH

Yang bertanda tangan dibawah ini:

Nama : Annisa Ghina Fauziyyah
NIM : 19200496
Jenjang : Sarjana (S1)
Program Studi : Sistem Informasi
Fakultas : Teknik dan Informatika
Perguruan Tinggi : Universitas Bina Sarana Informatika

Dengan ini menyatakan bahwa data dan atau informasi yang saya gunakan dalam penulisan karya ilmiah dengan judul "**Analisis Sentimen Ulasan Aplikasi Streaming Rcti+ Menggunakan Support Vector Machine pada Google Play Store**" merupakan data dan atau informasi yang saya peroleh melalui hasil penelitian sendiri dan tidak didasarkan pada data atau informasi hasil riset dari perusahaan/instansi/lembaga.

Saya bersedia untuk bertanggung jawab secara pribadi, tanpa melibatkan pihak **Universitas Bina Sarana Informatika**, atas materi/isi karya ilmiah tersebut, termasuk bertanggung jawab atas dampak atau kerugian yang timbul dalam bentuk akibat tindakan yang berkaitan dengan data dan atau informasi yang terdapat pada karya ilmiah saya ini.

Demikian pernyataan ini saya buat dengan sebenarnya.

Dibuat di: Jakarta

Pada tanggal: 1 Juli 2024

Mengetahui,

Yang Menyatakan,

Dosen Pembimbing I

Dosen Pembimbing II



Mochammad Nandi Susila, M.Kom.



Nur Ali Farabi, M.Kom.



Annisa Ghina Fauziyyah

BUKTI HASIL PENGECEKAN PLAGIARISME

ANALISIS SENTIMEN ULASAN APLIKASI *STREAMING RCTI+* MENGUNAKAN *SUPPORT VECTOR MACHINE* PADA *GOOGLE PLAY STORE*



SKRIPSI

Diajukan untuk memenuhi salah satu syarat kelulusan Program Sarjana

ANNISA GHINA FAUZIYAH
NIM : 19200496

Program Studi Sistem Informasi
Fakultas Teknik dan Informatika
Universitas Bina Sarana Informatika
Jakarta
2024

BAB I PENDAHULUAN

1.1 Latar Belakang Masalah

Saat ini, perkembangan teknologi informasi dan komunikasi berkembang semakin pesat. Hal ini menyebabkan terjadinya pergeseran preferensi dan perilaku masyarakat dalam menggunakan teknologi informasi dan komunikasi. Salah satu media komunikasi yang mengalami pergeseran adalah televisi. Televisi mengalami perubahan yang cukup besar, dari televisi dengan pinggir besar dan layar hitam putih menjadi televisi dengan layar besar dan berwarna, bahkan semakin lama muncul jenis-jenis baru seperti televisi HD, televisi interaktif, televisi satelit, kabel multi-saluran, dan bahkan televisi *streaming* dan *video on demand*. Namun, dengan perkembangan teknologi internet saat ini, layanan siaran tidak hanya ditampilkan secara eksklusif di televisi seperti sebelumnya. Pada era internet ini, orang dapat menonton layanan siaran televisi kapan saja dan di mana saja. Bahkan, mereka juga tidak perlu pergi ke pesawat televisi untuk menonton acara TV karena hal tersebut dapat dilakukan melalui perangkat seperti *smartphone* atau laptop (R. R. Putra & Hidayat, 2022).

Selain itu, terjadinya digitalisasi, peningkatan jumlah jaringan internet, dan perkembangan inovasi teknologi pada saat ini menyebabkan munculnya berbagai macam layanan *streaming online* yang membuat penonton dapat melihat layanan siaran televisi tanpa terpaku jam dan waktu siaran. Hal ini digambarkan sebagai revolusi media baru, yaitu pergeseran semua budaya ke arah produksi, distribusi, dan komunikasi yang dimediasi komputer. Layanan *streaming* merupakan salah satu contoh dari revolusi media baru. Kemudahan dalam mengakses layanan siaran tanpa

BAB II METODOLOGI PENELITIAN

2.1 Landasan Teori

2.1.1 Analisis Sentimen

Analisis sentimen adalah salah satu teknik *Natural Language Processing (NLP)* dan *machine learning* untuk mengevaluasi secara otomatis teks untuk perasaan yang dirasakan penulis (positif, negatif, dan netral). Ini adalah area pemrosesan bahasa alami (NLP) yang berkembang karena interaksi manusia dengan komputer, ekstraksi informasi, dan penyulingan perasaan (*distillation of feelings*) dari data yang ada. Analisis sentimen mencakup mengidentifikasi kata atau frasa dalam teks yang menunjukkan sikap netral, positif, atau negatif. Analisis sentimen biasanya digunakan untuk mengekstrak berbagai karakteristik dari data tekstual terstruktur atau tidak terstruktur dan menganalisisnya untuk mengekstrak pikiran, pendapat, dan perasaan dari data tersebut. Seiring dengan banyaknya pengguna internet, semakin banyak juga orang yang mengungkapkan perasaan dan pendapat mereka secara bebas di berbagai media sosial elektronik. Sentimen bisa menjadi sumber daya yang berharga bagi perusahaan yang ingin memahami pandangan masyarakat terhadap produk atau layanan mereka (Purnamasari et al., 2023).

Analisis sentimen menjadi sangat populer karena kebutuhan individu atau kelompok yang meningkat untuk mengetahui pendapat seseorang tentang sesuatu. Penanganan yang berbeda untuk analisis sentimen juga dipengaruhi oleh dataset yang digunakan (Widayat, 2021). Analisis sentimen tidak hanya ditujukan untuk individu saja, tetapi juga kepada organisasi.

BAB III
HASIL DAN PEMBAHASAN

3.1 Proses Umum

3.1.1 Pengumpulan Data

Pengumpulan data pada penelitian ini menggunakan teknologi *scraping* dengan library "google-play-scraper" di Google Colab untuk mengumpulkan data ulasan aplikasi RCTI+ secara otomatis dari situs *Google Play Store*.



Sumber: Data olahan Google Colab

Gambar III.1
Instal library web scraping

Dalam proses pengumpulan data, library *Google-Play-Scraper* akan melakukan pengumpulan data berdasarkan parameter yang ditentukan. Peneliti memasukkan alamat url id RCTI+ pada *Google Play Store* seperti "com.fta.rctiitv" untuk mengambil data dari aplikasi RCTI+, dengan hanya mengambil teks dalam Bahasa Indonesia. Pengumpulan data dilakukan dengan mengambil ulasan paling relevan terhadap aplikasi RCTI+ sebanyak 1500 ulasan.

```
#Scrape jumlah ulasan yang diinginkan
from google_play_scraper import Sort, reviews
result, continuation_token = reviews(
    app="com.fta.rctiitv",
    lang="id",
    country="ID",
    sort=Sort.MOST_RELEVANT,
    count=1500,
    filter_score_min=4)
#Filter score_min=4 memfilter score atau rating 4 sampai dengan 5
```

Sumber: Data olahan Google Colab

Gambar III.2
Proses scraping data

BAB IV
PENUTUP

4.1 Kesimpulan

Berdasarkan hasil penelitian dan pengujian yang telah dilakukan oleh peneliti, maka dapat diambil kesimpulan sebagai berikut:

1. Dalam proses pengumpulan data ulasan RCTI+ pada halaman *reviews* di *Google Play Store*, teknik *web scraping* diterapkan menggunakan library pemrograman *python* dan website *Google Colab*, diperoleh dataset sebanyak 1500 data berdasarkan ulasan yang relevan. Data ulasan tersebut kemudian terbagi menjadi dua kelas setelah dilakukan pelabelan secara manual, didapatkan 948 sentimen positif dan 552 sentimen negatif.
2. Tahapan pengolahan data dimulai dengan tahapan *preprocessing*, yang meliputi beberapa langkah utama seperti *Cleaning*, *Case Folding*, *Tokenizing*, *Stopword Removal*, dan *Filter Tokens (by Length)*. Setelah itu, tahapan pembobotan kata TF-IDF digunakan untuk melakukan konversi kata-kata menjadi vektor numerik yang dapat diterapkan pada model. Kemudian diimplementasikan dengan menggunakan metode *Support Vector Machine*, dengan pembagian data latih dan data uji sebesar 80%:20%. Dari pembagian tersebut, didapatkan data latih sebanyak 1200 dan data uji sebanyak 300.
3. Dalam penerapan algoritma klasifikasi *Support Vector Machine*, akan digunakan *confusion matrix* untuk mengevaluasi kinerja model tersebut. Tahap evaluasi ini menggunakan 300 data uji, dengan rincian 190 data memiliki sentimen positif dan 110 data lainnya memiliki sentimen negatif.



SKRIPSI ANNISA Turnitin-1728363926322

ORIGINALITY REPORT

20%
SIMILARITY INDEX

15%
INTERNET SOURCES

14%
PUBLICATIONS

7%
STUDENT PAPERS

PRIMARY SOURCES

1	Submitted to Universitas Brawijaya Student Paper	1%
2	www.researchgate.net Internet Source	1%
3	123dok.com Internet Source	1%
4	Submitted to Konsorsium Perguruan Tinggi Swasta Indonesia II Student Paper	1%
5	Muhammad Diki Hendriyanto, Azhari Ali Ridha, Ultach Enri. "Analisis Sentimen Ulasan Aplikasi Mola Pada Google Play Store Menggunakan Algoritma Support Vector Machine", INTECOMS: Journal of Information Technology and Computer Science, 2022 Publication	1%
6	etheses.uin-malang.ac.id Internet Source	1%
7	Dany Pratmanto, Fabriyan Fandi Dwi Imaniawan. "Analisis Sentimen Terhadap	<1%