

Protected by PDF Anti-Copy Free

(Upgrade to Pro Version to Remove the Watermark)

**ANALISIS SENTIMEN ULASAN PENGGUNA APLIKASI
SIREKAP MENGGUNAKAN OPTIMASI *SUPPORT*
VECTOR MACHINE DENGAN *GRID SEARCH*
*CROSS VALIDATION***



SKRIPSI

**Diajukan Sebagai Syarat Untuk Menyelesaikan Pendidikan
Program Sarjana (S-1)
Pada Program Studi Informatika**

**Oleh :
SIDIK FAHRIANSYAH
NIM : 2102020044**

**PROGRAM STUDI INFORMATIKA
FAKULTAS ILMU TEKNIK
UNIVERSITAS BINA INSAN
2024**

Protected by PDF Anti-Copy Free

(Upgrade to Pro Version to Remove the Watermark)

HALAMAN PENGESAHAN SKRIPSI



**ANALISIS SENTIMEN ULASAN PENGGUNA APLIKASI
SIREKAP MENGGUNAKAN OPTIMASI *SUPPORT
VECTOR MACHINE* DENGAN *GRID SEARCH
CROSS - VALIDATION***

Oleh:

SIDIK FAHRIANSYAH

NIM : 2102020044

Pembimbing I

**Lubuklinggau, Januari 2025
Pembimbing II**

Dr. Muhamad Akbar, ST., M.IT

Muhammad Irvai, M.Kom

**Mengesahkan,
Dekan Fakultas Ilmu Teknik
Universitas Bina Insan**

Dr. Rudi Kurniawan, ST., M.Kom

Protected by PDF Anti-Copy Free

(Upgrade to Pro Version to Remove the Watermark)

HALAMAN PERSETUJUAN TIM PENGUJI SKRIPSI



Pada hari Kamis tanggal 23 bulan Januari tahun 2025 telah dilaksanakan Sidang Skripsi oleh Studi Informatika Fakultas Ilmu Teknik Universitas Bina Insan.

Nama : Sidik Fahriansyah
Nim : 2102020044
Judul Skripsi : Analisis Sentimen Ulasan Pengguna Aplikasi Sirekap Menggunakan Optimasi *Support Vector Machine* Dengan *Grid Search Cross-Validation*.

Komisi Penguji

1. Ketua : Dr. Muhamad Akbar, ST., M.IT (.....)
2. Sekretaris : Muhammad Irvai, M. Kom (.....)
3. Anggota : Armanto, M. Kom (.....)

Mengetahui,
Kepala Program Studi Informatika
Fakultas Ilmu Teknik
Universitas Bina Insan

Budi Santoso M.Kom

Protected by PDF Anti-Copy Free

(Upgrade to Pro Version to Remove the Watermark)

HALAMAN MOTTO DAN PERSEMBAHAN



MOTTO :

- ❖ Allah tidak akan membebani seseorang melainkan sesuai dengan kesanggupannya. (QS. Al-Baqarah: 286)
- ❖ Belajarlah dari kesalahan, tidak ada *deadline* untuk menjadi lebih baik, tetapi selalu ada waktu untuk memperbaiki diri.
- ❖ 1% Kerja Keras 99% Do'a Ibu

Persembahkan Kepada :

- ❖ *Bapak dan Mamak Tercinta, yang menjadi alasan Sidik untuk terus semangat. Terima kasih telah mengusahakan yang terbaik untuk anak pertamamu, memberikan kasih sayang, doa, dan pengorbanan tanpa batas. Semua ini tidak akan tercapai tanpa Bapak dan Mamak.*
- ❖ *Diriku, atas ketekunan, usaha, dan kerja kerasnya selama ini.*
- ❖ *Adik-Adikku Tersayang, yang menjadi alasan Sidik untuk selalu memberikan contoh terbaik.*
- ❖ *Teman-teman seperjuangan, khususnya A2 Informatika. Terima kasih atas kebersamaan, kerja sama, dan tawa selama ini. Semoga Bahagia dan sukses selalu.*
- ❖ *Almamaterku.*

Protected by PDF Anti-Copy Free

(Upgrade to Pro Version to Remove the Watermark)

HALAMAN PERNYATAAN



Saya yang bertanda tangan di bawah ini :

Nama Mahasiswa/I : Sidik Fahriansyah

NIM : 2102020044

Program Studi : Informatika

Fakultas : Ilmu Teknik

Menyatakan dengan sesungguhnya bahwa penelitian dan penulisan Skripsi yang saya susun sebagai persyaratan untuk memperoleh gelar Sarjana (S-1) Universitas Bina Insan, merupakan hasil kerja saya sendiri dan tidak menyuruh orang lain yang mengerjakannya. Ada pun bagian tertentu dalam penulisan skripsi ini yang saya kutip dari hasil karya orang lain dan telah saya tuliskan sumbernya secara jelas sesuai dengan norma, kaidah dan etika penulisan ilmiah.

Jika dikemudian hari ternyata terbukti bahwa penelitian dan tugas akhir ini bukan hasil kerja saya sendiri atau plagiat dalam bagian-bagian tertentu, maka saya bersedia dikenakan sanksi sesuai dengan peraturan perundangan yang berlaku.

Lubuklinggau, 23 Januari 2025

Peneliti,

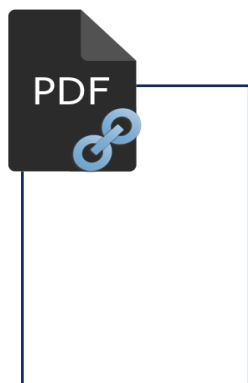
Sidik Fahriansyah

NIM.2102020044

Protected by PDF Anti-Copy Free

(Upgrade to Pro Version to Remove the Watermark)

DAFTAR RIWAYAT HIDUP



Biodata

Nama : Sidik Fahriansyah
Tempat / Tanggal Lahir : Tambahasri, 15 Februari 2004
Jenis Kelamin : Laki-Laki
Agama : Islam
Alamat : Jl. Mangga, Dusun 1, Desa Tambah Asri,
Kec. Tugumulyo, Kab. Musi Rawas, Prov.
Sumatera Selatan

Pendidikan

- SD : SD Negeri Tambah Asri
- SMP/MTs Sederajat : SMP Negeri L. Sidoharjo
- SMA/MAN/SMK Sederajat : SMK Negeri Tugumulyo

Pengalaman Organisasi dan Pelatihan (opsional)

1. Himpunan Mahasiswa Program Studi Informatika (HIMPI)

Prestasi Akademik dan Non-Akademik (opsional)

No	Prestasi Akademik dan Non-Akademik	Tahun
1.	-	-
2.	-	-
3.	-	-

Protected by PDF Anti-Copy Free

(Upgrade to Pro Version to Remove the Watermark)

ABSTRACT


The Recapitulation Information (Sirekap) application is an information technology-based tool available on the Google Play Store as a means of publishing and recapitulating the results of reporting. With more than 37,000 user reviews on the Google Play Store, analysis of unstructured reviews requires a special approach to effectively understand user perceptions. This study aims to analyze user review sentiments on the Sirekap application using the Support Vector Machine (SVM) algorithm optimized with Grid Search Cross-Validation. The initial dataset consisted of 20,000 reviews obtained by scraping techniques, after the preprocessing stage became 17,506 reviews, consisting of 14,296 negative reviews, 2,342 positive reviews, and 868 neutral reviews. Then the over-under sampling stage was carried out using SMOTEENN to balance the data. The amount of data increased to 20,215 reviews, with a distribution of 8,157 positive reviews, 6,394 neutral reviews, and 5,664 negative reviews. Model testing was conducted with five data comparison ratios: 50:50, 60:40, 70:30, 80:20, and 90:10. The best results were obtained at a ratio of 90:10 with an accuracy of 98%. The precision, recall, and f1-score values were 0.99 for the positive class, 0.97 for the negative class, and 0.98 for the neutral class, respectively. The best parameters at a ratio of 90:10 obtained through Grid Search Cross-Validation optimization were the RBF kernel, C value = 1000, and Gamma = 1. Based on the results of this study, the Support Vector Machine (SVM) algorithm optimized with Grid Search Cross-Validation can be used for user review sentiment classification with very good results.

Keywords: Sirekap Application, Sentiment Analysis, Support Vector Machine (SVM), Grid Search Cross-Validation, SMOTEENN

Protected by PDF Anti-Copy Free

(Upgrade to Pro Version to Remove the Watermark)

ABSTRAK



Aplikasi Sistem Informasi Revisi (Sirekap) adalah perangkat berbasis teknologi informasi yang tersedia di *Google Play Store* sebagai sarana publikasi dan rekapitulasi hasil perhitungan. Dengan lebih dari 37.000 ulasan pengguna di *Google Play Store*, analisis terhadap ulasan yang tidak terstruktur memerlukan pendekatan khusus untuk memahami persepsi pengguna secara efektif. Penelitian ini bertujuan menganalisis sentimen ulasan pengguna terhadap aplikasi Sirekap menggunakan algoritma *Support Vector Machine (SVM)* yang dioptimasi dengan *Grid Search Cross-Validation*. Dataset awal terdiri dari 20.000 ulasan yang diperoleh dengan teknik *scraping*, setelah dilakukan tahap *preprocessing* menjadi 17.506 ulasan, terdiri atas 14.296 ulasan negatif, 2.342 ulasan positif, dan 868 ulasan netral. Kemudian dilakukan tahap *over-under sampling* menggunakan *SMOTEENN* untuk menyeimbangkan data. Jumlah data meningkat menjadi 20.215 ulasan, dengan distribusi 8.157 ulasan positif, 6.394 ulasan netral, dan 5.664 ulasan negatif. Pengujian model dilakukan dengan lima rasio perbandingan data: 50:50, 60:40, 70:30, 80:20, dan 90:10. Hasil terbaik diperoleh pada rasio 90:10 dengan *accuracy* sebesar 98%. Nilai *precision*, *recall*, dan *f1-score* masing-masing adalah 0,99 untuk kelas positif, 0,97 untuk kelas negatif, dan 0,98 untuk kelas netral. Parameter terbaik pada rasio 90:10 yang diperoleh melalui optimasi *Grid Search Cross-Validation* adalah kernel *RBF*, nilai $C=1000$, dan $\text{Gamma}=1$. Berdasarkan hasil dari penelitian ini, algoritma *Support Vector Machine (SVM)* yang dioptimasi dengan *Grid Search Cross-Validation* dapat digunakan untuk klasifikasi sentimen ulasan pengguna dengan hasil yang sangat baik.

Kata Kunci : Aplikasi Sirekap, Analisis Sentimen, *Support Vector Machine (SVM)*, *Grid Search Cross-Validation*, *SMOTEENN*

Protected by PDF Anti-Copy Free

(Upgrade to Pro Version to Remove the Watermark)

KATA PENGANTAR



Alhamdulillah puji dan syukur peneliti ucapkan kepada Allah SWT atas segala rahmat dan karunia-Nya yang telah memberikan kekuatan dan kesempatan, sehingga Peneliti dapat menyelesaikan proposal skripsi ini dengan maksimal. Sholawat beserta salam semoga tetap tercurahkan kepada bagi Nabi Muhammad SAW, keluarga, sahabat, serta umatnya hingga akhir zaman. Sebagai bentuk rasa syukur, Peneliti ingin menyampaikan penghargaan setinggi-tingginya kepada semua pihak yang telah turut serta memberikan dukungan, bimbingan, serta dorongan dalam proses penyusunan proposal skripsi ini.

Kelancaran penyusunan proposal ini tidak terlepas dari bantuan berbagai pihak, baik secara langsung maupun tidak. Oleh karena itu, Peneliti mengucapkan terima kasih kepada segenap pihak yang telah membantu, yaitu :

1. Allah SWT yang telah memberikan segala nikmat dan Karunia-Nya sehingga Peneliti dapat menyelesaikan proposal ini.
2. Kedua Orang Tua Tercinta yang turut mendukung serta mendoakan Peneliti dalam pembuatan Proposal ini.
3. Bapak Dr. H. Sardiyo, M.M selaku Rektor Universitas Bina Insan
4. Bapak Dr. Muhammad Akbar, S.T., M.IT selaku wakil Rektor Universitas Bina Insan
5. Bapak Wakhid Nur Mukhlis, M.Pd., M.M selaku wakil II Rektor Universitas Bina Insan
6. Bapak Dr. Rudi Kurniawan, ST., M.Kom selaku Dekan Fakultas Ilmu Teknik Universitas Bina Insan
7. Bapak Budi Santoso, M.Kom selaku Ketua Program Studi Informatika Universitas Bina Insan.
8. Bapak Dr. Muhamad Akbar, ST., M.IT selaku dosen Pembimbing I yang telah banyak memberikan bimbingan dan arahan dalam Penulisan proposal ini.

Protected by PDF Anti-Copy Free

(Upgrade to Pro Version to Remove the Watermark)

9. Bapak Muhammad Irvai, M. Kom selaku dosen pembimbing II yang telah banyak memberikan bimbingan dan arahan dalam Penulisan proposal ini.
10. Seluruh Staf Dosen dan Pegawai Universitas Bina Insan yang telah banyak memberikan ilmu, pengetahuan dan bimbingan kepada penulis
11. Rekan-rekan Angkatan Tahun 2021 Program Studi Informatika yang telah saling membantu dalam penulisan proposal ini.

Peneliti menyadari dalam penelitian ini tentunya masih jauh dari kata sempurna. Hal ini dikarenakan keterbatasan ilmu yang dimiliki. Oleh karena itu, kiranya dapat diberikan kritik dan saran yang membangun untuk melengkapi kesempurnaan dari penelitian ini dan demi perbaikan di masa mendatang. Semoga Proposal ini dapat memberikan manfaat dan berdampak positif bagi Peneliti dan semua pihak yang membaca proposal ini.

Lubuklinggau, November 2024

Peneliti

Protected by PDF Anti-Copy Free

(Upgrade to Pro Version to Remove the Watermark)

DAFTAR ISI



Halaman

HALAMAN JUDUL	i
HALAMAN PENGESAHAN SKRIPSI.....	ii
HALAMAN PERSETUJUAN TIM PENGUJI SKRIPSI.....	iii
HALAMAN MOTTO DAN PERSEMBAHAN.....	iv
HALAMAN PERNYATAAN.....	v
DAFTAR RIWAYAT HIDUP	vi
<i>ABSTRACT</i>	vii
ABSTRAK	viii
KATA PENGANTAR.....	ix
DAFTAR ISI.....	xi
DAFTAR TABEL	xiii
DAFTAR GAMBAR	xv
DAFTAR LAMPIRAN	xvi
BAB 1 PENDAHULUAN	1
1.1 Latar Belakang Penelitian.....	1
1.2 Identifikasi Masalah	4
1.3 Rumusan Masalah	4
1.4 Batasan Masalah.....	4
1.5 Tujuan dan Manfaat Penelitian.....	5
1.5.1 Tujuan Penelitian.....	5
1.5.2 Manfaat Penelitian	5
1.6 Sistematika Penulisan.....	6
BAB II TINJAUAN PUSTAKA.....	8
2.1 Literatur	8
2.2 Penelitian Terdahulu Yang Relevan.....	18
2.3 Kerangka Berpikir	21
BAB III METODOLOGI PENELITIAN	22
3.1 Analisa Sistem	22
3.1.1 Analisa Sistem Yang Berjalan	22
3.1.2 Alternatif Pemecahan Masalah.....	23
3.1.3 Metode Analisa.....	24
3.2 Teknik Pemilihan Informan (Populasi, Sampel, dan Sampling)	26
3.2.1 Teknik Pengumpulan Data	26
3.2.2 Teknik Analisa Data	27
3.3 Tempat dan Waktu Penelitian	43
BAB IV HASIL PENELITIAN DAN PEMBAHASAN.....	44
4.1 Gambaran Umum Tempat Penelitian.....	44
4.1.1 Gambaran Umum Aplikasi Sistem Informasi Rekapitulasi ...	44
4.2 Hasil.....	46

Protected by PDF Anti-Copy Free

(Upgrade to Pro Version to Remove the Watermark)

4.3 Pembahasan	64
4.3.1 Penerapan Metode Analisa dan Validitas Data	64
4.3.2 Pengujian Hasil Analisa	68
BAB V KESIMPULAN DAN	72
5.1 Kesimpulan	72
5.2 Saran	72
DAFTAR PUSTAKA	75
LAMPIRAN	

Protected by PDF Anti-Copy Free

(Upgrade to Pro Version to Remove the Watermark)

DAFTAR TABEL



	Halaman
Tabel 2.1 Penelitian Relevan.....	18
Tabel 3. 1 <i>Script Data Cleaning</i>	29
Tabel 3.2 Hasil <i>Data Cleaning</i>	30
Tabel 3.3 <i>Script Case Folding</i>	31
Tabel 3.4 Hasil <i>Case Folding</i>	31
Tabel 3.5 <i>Script Normalization</i>	32
Tabel 3.6 Hasil <i>Normalization</i>	32
Tabel 3.7 <i>Script Tokenizing</i>	33
Tabel 3.8 Hasil <i>Tokenizing</i>	33
Tabel 3. 9 <i>Script Stopword Removal</i>	34
Tabel 3.10 Hasil <i>Stopword Removal</i>	35
Tabel 3.11 <i>Script Stemming</i>	36
Tabel 3.12 Hasil <i>Stemming</i>	36
Tabel 3.13 <i>Labeling</i>	37
Tabel 3.14 Pembagian Data Setelah <i>SMOTEENN</i>	39
Tabel 3.15 Tabel Pengujian <i>Confusion Matrix</i>	41
Tabel 3.16 Rencana Kegiatan Penelitian.....	43
Tabel 4.1 Tabel Split Data	55
Tabel 4.2 Hasil <i>GridSearchCV</i>	57

Protected by PDF Anti-Copy Free

(Upgrade to Pro Version to Remove the Watermark)

DAFTAR GAMBAR



	Halaman
Gambar 2.1 Logo Aplikasi Sirekap.....	9
Gambar 2.2 Logo Python.....	14
Gambar 2.3 Kerangka Berpikir	21
Gambar 3.1 Diagram Konsep Penelitian	25
Gambar 3.2 Tahap <i>Preprocessing</i>	28
Gambar 3.3 Tahap <i>Data Cleaning</i>	29
Gambar 3.4 Tahap <i>Case Folding</i>	31
Gambar 3.5 Tahap <i>Normalization</i>	32
Gambar 3.6 Tahap <i>Tokenizing</i>	33
Gambar 3.7 Tahap <i>Stopword Removal</i>	34
Gambar 3.8 Tahap <i>Stemming</i>	35
Gambar 3.9 Alur Penerapan Algoritma <i>SVM</i>	40
Gambar 4.1 Logo Sirekap.....	44
Gambar 4.2 Beberapa Ulasan Pengguna Sirekap	45
Gambar 4.3 Data Hasil <i>Scraping</i> Ulasan Pengguna Sirekap.....	47
Gambar 4.4 Data Hasil <i>Scraping</i> Ulasan Pengguna Sirekap.....	47
Gambar 4.5 Data Hasil <i>Scraping</i> Ulasan Pengguna Sirekap.....	47
Gambar 4.6 Hasil <i>Data Cleaning</i>	48
Gambar 4.7 Hasil <i>Case Folding</i>	48
Gambar 4.8 Hasil <i>Normalization</i>	49
Gambar 4.9 Hasil <i>Tokenizing</i>	49
Gambar 4.10 Hasil <i>Stopword Removal</i>	50
Gambar 4.11 Hasil <i>Stemming</i>	51
Gambar 4.12 Hasil <i>Labeling</i>	51
Gambar 4.13 Hasil Visualisasi <i>BarPlot</i>	52
Gambar 4.14 Hasil Visualisasi <i>WordCloud</i>	53
Gambar 4.15 Hasil <i>TF-IDF</i>	54
Gambar 4.16 Hasil Visualisasi Tahap <i>SMOTEENN</i>	54
Gambar 4.17 Proses <i>GridSearchCV</i>	56
Gambar 4.18 Proses <i>GridSearchCV</i>	56
Gambar 4.19 Proses <i>GridSearchCV</i>	56
Gambar 4.20 Hasil Pelatihan Model <i>SVM</i>	57
Gambar 4.21 <i>Script</i> Pengujian Model <i>SVM</i>	58
Gambar 4.22 Hasil Pengujian Model (Rasio 50:50).....	58
Gambar 4.23 Hasil Pengujian Model (Rasio 60:40).....	59
Gambar 4.24 Hasil Pengujian Model (Rasio 70:30).....	59
Gambar 4.25 Hasil Pengujian Model (Rasio 80:20).....	60
Gambar 4.26 Hasil Pengujian Model (Rasio 90:10).....	60

Protected by PDF Anti-Copy Free

(Upgrade to Pro Version to Remove the Watermark)


Gambar 4.27 Hasil Evaluasi Model (Rasio 50:50).....	61
Gambar 4.28 Hasil Evaluasi Model (Rasio 60:40).....	62
Gambar 4.29 Hasil Evaluasi Model (Rasio 70:30).....	62
Gambar 4.30 Hasil Evaluasi Model (Rasio 80:20).....	63
Gambar 4.31 Hasil Evaluasi Model (Rasio 90:10).....	64



Protected by PDF Anti-Copy Free

(Upgrade to Pro Version to Remove the Watermark)

DAFTAR LAMPIRAN

- 
- Lampiran 1.** Lembar Pengajuan
 - Lampiran 2.** Lembar Bimbingan al Skripsi Dosen Pembimbing 1
 - Lampiran 3.** Lembar Bimbingan al Skripsi Dosen Pembimbing 2
 - Lampiran 4.** Lembar Bimbingan Skripsi Dosen Pembimbing 1
 - Lampiran 5.** Lembar Bimbingan Skripsi Dosen Pembimbing 2
 - Lampiran 6.** Sampel Dataset Ulasan Pengguna Sirekap
 - Lampiran 7.** Tampilan Aplikasi Sirekap di *Google Play Store*
 - Lampiran 8.** *Source Code* Program

Protected by PDF Anti-Copy Free

(Upgrade to Pro Version to Remove the Watermark)

BAB I



DAHULUAN

1.1 Latar Belakang Penelitian

Demokrasi di Indonesia menjadi fondasi utama dalam pemerintahan. Negara demokrasi adalah negara yang diselenggarakan atas dasar kemauan dan kemauan rakyat. Sebagai bagian dari demokrasi, Pemilihan Umum (Pemilu) memberikan kesempatan yang seluas-luasnya kepada masyarakat untuk ikut aktif berpartisipasi dalam pengambilan kebijakan, yaitu dengan menggunakan haknya untuk memilih pemimpin sesuai keinginannya [1]. Pemilu merupakan wujud nyata dari demokrasi prosedural, di mana warga negara memiliki peran penting dalam menentukan arah pemerintahan [2].

Dengan kemajuan teknologi, proses pemungutan suara dan penghitungan hasil kini telah beralih ke sistem elektronik. Saat ini, teknologi dapat diterapkan di hampir semua tahapan pemilu. Penelitian yang dilakukan oleh *International IDEA* memperkuat hal ini, menunjukkan bahwa dari 106 negara yang menjadi subjek penelitian, banyak yang telah menggunakan teknologi dalam pemilu, termasuk untuk tabulasi hasil suara, pendaftaran serta verifikasi pemilih, dan *e-voting*. Komisi Pemilihan Umum (KPU) Republik Indonesia telah menerapkan sistem rekapitulasi elektronik baru bernama Sistem Informasi Rekapitulasi (Sirekap), untuk pertama kalinya di Pilkada Serentak 2020 [3].

Sistem Informasi Rekapitulasi (Sirekap) adalah perangkat aplikasi yang tersedia di *Google Play Store* berbasis teknologi informasi sebagai sarana publikasi hasil perhitungan suara dan proses rekapitulasi hasil perhitungan suara, serta alat bantu dalam pelaksanaan hasil perhitungan suara pemilu. Penggunaan Sirekap menjadi semakin relevan dan penting. Sirekap memiliki potensi untuk mempercepat proses pengumuman hasil pemilu. Adanya data yang tersedia langsung dalam bentuk digital, maka

Protected by PDF Anti-Copy Free

(Upgrade to Pro Version to Remove the Watermark)

proses analisis dan pelaporan dapat dilakukan dengan lebih cepat, serta dapat meningkatkan keyakinan terhadap transparansi dan keadilan pemilu [4]. Meskipun demikian, penggunaan aplikasi Sirekap ini, tentu tidak terlepas dari berbagai tantangan dan problematika yang perlu diatasi.

Saat ini, aplikasi Sirekap telah menerima sekitar 37.000 ulasan di *Google Play Store*. Ulasan ini merupakan sumber data penting bagi pengembang aplikasi untuk memahami persepsi pengguna, mengidentifikasi kebutuhan, serta menentukan langkah perbaikan atau pengembangan yang relevan. Mengingat sifat ulasan yang tidak terstruktur, diperlukan pendekatan khusus untuk memahami penilaian pengguna secara efektif [5]. Analisis sentimen dapat dimanfaatkan untuk memahami tanggapan masyarakat terhadap penggunaan aplikasi Sirekap [6]. Penelitian sebelumnya yang menggunakan optimasi *Support Vector Machine (SVM)* dengan *GridSearchCV* pada 1791 ulasan [7], menghasilkan analisis sentimen yang kurang akurat. Hal ini disebabkan oleh pemilihan nilai parameter yang tidak optimal. Selain itu, ketidakseimbangan data (*data imbalance*) pada ulasan pengguna aplikasi Sirekap juga menjadi tantangan dalam meningkatkan performa model klasifikasi.

Beberapa penelitian telah dilakukan untuk mencari metode yang efektif dalam analisis sentimen ulasan pengguna aplikasi. Penelitian pertama menganalisis ulasan aplikasi Sirekap di *Google Play Store* dengan menggunakan optimasi *Support Vector Machine (SVM)* melalui *GridSearchCV* pada 1791 ulasan, menghasilkan akurasi 81%, presisi 87%, dan recall 85% [7]. Penelitian serupa yang menggunakan algoritma *Random Forest* pada 5000 ulasan, menghasilkan akurasi 74%, presisi 75%, recall 74%, dan *f1-score* 74% [2]. Penelitian ketiga menganalisis ulasan aplikasi Indodax menggunakan *Support Vector Machine (SVM)* dengan pembobotan TF-IDF pada 1138 data, menghasilkan akurasi 85% dengan rasio data 80:20 [6]. Penelitian lain membandingkan *Support Vector*

Protected by PDF Anti-Copy Free

(Upgrade to Pro Version to Remove the Watermark)

Machine (SVM) dan *K-Nearest Neighbors (K-NN)* untuk ulasan aplikasi Gojek. *K-NN* dengan *SMOTEENN* mencatat akurasi 82,14%, presisi 82,28%, dan recall 95,43%, sementara *SVM* (kernel linear, $C=1$) mencapai akurasi 87,98%, presisi 88,55%, dan recall 95,43% [8].

Untuk meminimalisir kesalahan dalam mengidentifikasi sentimen pengguna aplikasi Sirekap dan meningkatkan efisiensi proses analisis, diperlukan metode yang dapat mengklasifikasikan sentimen pengguna secara lebih akurat. Penelitian ini menggunakan algoritma *Support Vector Machine (SVM)* yang dioptimalkan dengan *GridSearchCV* untuk memilih parameter terbaik dalam dataset dan penerapan metode *SMOTEENN (Synthetic Minority Over-sampling Technique Edited Nearest Neighbors)* untuk normalisasi data. *Support Vector Machine (SVM)* merupakan model pembelajaran mesin yang termasuk dalam kategori *Supervised Learning* [5] yaitu sebuah metode klasifikasi dimana sekumpulan data diberikan label untuk mengklasifikasikan kelas [8]. Penelitian yang dilakukan oleh Siji George C. G. dan B. Sumathi (2020), algoritma *Support Vector Machine (SVM)* yang dioptimalkan dengan *Grid Search Cross-Validation (GridSearchCV)* dapat digunakan untuk meningkatkan nilai akurasi [9].

Dikarenakan keunggulan tersebut, *Support Vector Machine (SVM)* dengan *GridSearchCV* dinilai cocok untuk analisis sentimen ulasan pengguna aplikasi Sirekap, yang mungkin sebelumnya dilakukan secara manual atau kurang optimal. Penelitian ini bertujuan untuk menemukan metode analisis sentimen yang lebih akurat daripada pendekatan sebelumnya dan membantu memahami persepsi pengguna aplikasi Sirekap dengan lebih baik. Berdasarkan permasalahan dan kondisi yang ada, maka peneliti melakukan penelitian dengan judul “ **Analisis Sentimen Ulasan Pengguna Aplikasi Sirekap Menggunakan Optimasi *Support Vector Machine* Dengan *Grid Search Cross -Validation* “.**

Protected by PDF Anti-Copy Free

(Upgrade to Pro Version to Remove the Watermark)

1.2 Identifikasi Masalah

Berdasarkan latar belakang yang telah dijelaskan maka peneliti mengidentifikasi masalah yang ada yaitu :

- a. Hasil analisis penelitian sebelumnya yang menggunakan algoritma *Support Vector Machine (SVM)* dengan *Grid Search Cross-Validation* kurang optimal dalam mengklasifikasikan sentimen ulasan pengguna aplikasi Sirekap.
- b. Pemilihan parameter yang kurang tepat dalam algoritma *Support Vector Machine (SVM)* membuat performa model klasifikasi sentimen kurang maksimal.
- c. *Data Imbalance* pada ulasan aplikasi Sirekap dikarenakan data tidak seimbang, maka dibutuhkan metode *SMOTEENN (Synthetic Minority Oversampling Technique Edited Nearest Neighbors)* untuk normalisasi data.

1.3 Rumusan Masalah

Berdasarkan latar belakang yang telah dikemukakan, terdapat rumusan masalah dalam penelitian ini, yaitu : Bagaimana optimasi algoritma *Support Vector Machine (SVM)* menggunakan *GridSearchCV* dapat meningkatkan akurasi analisis sentimen ulasan pengguna aplikasi Sirekap ?

1.4 Batasan Masalah

Dengan luasnya ruang lingkup penelitian, maka peneliti memberikan batasan masalah yaitu parameter atau ruang lingkup yang mengikat fokus penelitian agar tetap terarah dan terukur. Diantaranya sebagai berikut :

- a. Data yang digunakan adalah data ulasan pengguna aplikasi Sirekap di *Google Play Store* yang diperoleh dengan cara *Scraping Data*.

Protected by PDF Anti-Copy Free

(Upgrade to Pro Version to Remove the Watermark)

- b. Penelitian ini hanya menggunakan metode *Support Vector Machine* (*SVM*) untuk klasifikasi sentimen tanpa melakukan perbandingan dengan algoritma lainnya.
- c. Optimasi model penelitian menggunakan *GridSearchCV* untuk menentukan parameter terbaik.
- d. Sistem ini dirancang menggunakan bahasa pemrograman *Python*.
- e. Sentimen yang di analisis dibagi menjadi tiga kategori, yaitu positif, negatif, dan netral.

1.5 Tujuan dan Manfaat Penelitian

1.5.1 Tujuan Penelitian

Tujuan dalam penelitian ini adalah :

- 1) Mengembangkan metode analisis sentimen yang lebih optimal untuk meningkatkan akurasi dalam mengklasifikasikan sentimen positif, negatif, dan netral pada ulasan aplikasi Sirekap.
- 2) Menentukan kombinasi parameter yang optimal pada algoritma *Support Vector Machine (SVM)* menggunakan optimasi *GridSearchCV* untuk memaksimalkan performa model klasifikasi sentimen.
- 3) Mengatasi ketidakseimbangan data (*data imbalance*) pada ulasan aplikasi Sirekap menggunakan metode *SMOTEENN (Synthetic Minority Oversampling Technique Edited Nearest Neighbors)* sehingga menghasilkan model klasifikasi yang lebih representatif.

1.5.2 Manfaat Penelitian

Adapun manfaat penelitian yaitu :

- 1) Memberikan pemahaman yang lebih cepat dan akurat mengenai sentimen pengguna aplikasi Sirekap melalui analisis otomatis ulasan pengguna.

Protected by PDF Anti-Copy Free

(Upgrade to Pro Version to Remove the Watermark)

- 2) Menghasilkan metode analisis sentimen yang lebih akurat dan efisien, yang dapat diterapkan pada ulasan pengguna aplikasi serupa.
- 3) Mengidentifikasi parameter yang signifikan untuk optimasi *Support Vector Machine (SVM)*, yang dapat diterapkan pada model klasifikasi sentimen lainnya untuk meningkatkan akurasi.
- 4) Membantu menghasilkan model klasifikasi sentimen yang lebih seimbang dengan mengatasi masalah ketidakseimbangan data, sehingga model dapat memberikan prediksi yang lebih representatif untuk setiap kelas sentimen.
- 5) Memberikan pemahaman yang lebih mendalam kepada pengembang aplikasi Sirekap berdasarkan hasil analisis sentimen, sehingga mereka dapat meningkatkan fitur, *user experience (UX)*, atau performa aplikasi sesuai dengan *feedback* pengguna.
- 6) Membantu pengembang dalam mengambil keputusan yang lebih tepat berdasarkan data sentimen yang dihasilkan, seperti memprioritaskan perbaikan fitur yang paling banyak dikritik atau mendapatkan *insight* untuk pengembangan fitur baru.

1.6 Sistematika Penulisan

Penulisan skripsi ini, yang merupakan laporan hasil penelitian, dirancang dalam lima bab dengan rincian sebagai berikut :

BAB I : PENDAHULUAN

Dalam bab ini berisi latar belakang masalah, identifikasi masalah, rumusan masalah, batasan masalah, tujuan dan manfaat penelitian, metode penelitian, dan sistematika penulisan.

Protected by PDF Anti-Copy Free

(Upgrade to Pro Version to Remove the Watermark)

BAB II : KAJIAN PUSTAKA

Dalam bab ini berisi teori-teori yang mendasari penelitian yang diteliti seperti literatur, penelitian referensi dan kerangka berpikir.

BAB III : METODOLOGI PENELITIAN

Dalam bab ini berisi tentang metode penelitian, metode pengumpulan data, metode analisa, tempat dan waktu penelitian, alat dan bahan yang digunakan penelitian, metode pengujian data dan pengolahan data.

BAB IV : HASIL DAN PEMBAHASAN

Dalam bab ini berisi tentang Gambaran umum tempat penelitian, hasil penelitian, pembahasan, pemecahan masalah yang dilakukan oleh peneliti.

BAB V : KESIMPULAN DAN SARAN

Dalam bab ini berisi tentang kesimpulan dari penelitian serta saran yang bermanfaat untuk pengembangan lebih lanjut.

DAFTAR PUSTAKA

LAMPIRAN



2.1 Literatur

2.1.1 Analisis Sentimen

Analisis sentimen adalah metode untuk mengidentifikasi dan mengklasifikasikan emosi dalam teks, dengan tujuan membedakan apakah sentimen tersebut positif, negatif, atau netral [2]. Analisis sentimen, atau yang juga disebut *opinion mining*, merupakan jenis penggalian data yang memanfaatkan pembelajaran mesin untuk mengekstraksi, merangkum, dan menganalisis informasi yang diperoleh dari teks [10].

Secara umum, analisis sentimen diterapkan dalam berbagai bidang seperti prediksi harga saham, isu politik, kepuasan terhadap produk dan layanan, analisis reputasi, dan lain-lain. Fokus analisis sentimen adalah untuk memvalidasi klasifikasi berdasarkan polaritas. Analisis sentimen juga mencakup proses bahasa alami dalam menganalisis opini, sentimen, apresiasi, dan emosi tentang produk, layanan, organisasi, individu, dan atributnya [11].

2.1.2 Ulasan Pengguna

Ulasan adalah respons yang diberikan oleh pengguna aplikasi, yang mencerminkan perasaan seperti kepuasan, kekecewaan, atau pendapat mereka tentang aplikasi tersebut [2]. Ulasan pengguna merupakan fitur di platform *Google Play Store* yang memberi kesempatan bagi pengguna untuk memberikan *feedback* atau umpan balik berupa rating dan pendapat tentang aplikasi yang telah mereka unduh. Fitur ini memungkinkan pengguna memberikan komentar tentang aplikasi yang mereka gunakan serta menyampaikan saran kepada pengembang.

Protected by PDF Anti-Copy Free
(Upgrade to Pro Version to Remove the Watermark)

2.1.3 Google Play Store

Google Play Store, sebelumnya disebut *Android Market*, merupakan layanan distribusi digital yang dikelola dan dikembangkan oleh *Google*. Sebagai toko aplikasi resmi untuk sistem operasi Android, platform ini memungkinkan pengguna untuk mencari dan mengunduh aplikasi yang dikembangkan menggunakan *Android Software Development Kit (SDK)* dan dirilis oleh *Google* [12]. *Google Play Store* adalah layanan yang dimiliki oleh *Google* yang menyediakan berbagai jenis konten digital, seperti aplikasi, *game*, dan lainnya. Di dalamnya, pengguna dapat melihat kolom penilaian dan memberikan ulasan untuk aplikasi yang tersedia [13].

2.1.4 Sistem Informasi Rekapitulasi (Sirekap)



Gambar 2.1 Logo Aplikasi Sirekap

Sistem Informasi Rekapitulasi (Sirekap) adalah perangkat aplikasi yang tersedia di *Google Play Store* berbasis teknologi informasi sebagai sarana publikasi hasil perhitungan suara dan proses rekapitulasi hasil perhitungan suara, serta alat bantu dalam pelaksanaan hasil perhitungan suara pemilu. Sistem Rekapitulasi Suara (Sirekap) dirancang untuk mengotomatiskan dan mempercepat proses pengumpulan dan pengolahan data suara [4]. Sistem Informasi Rekapitulasi (Sirekap) mulai dikembangkan oleh Komisi Pemilihan Umum (KPU) sejak awal tahun 2020 [3].

Protected by PDF Anti-Copy Free

(Upgrade to Pro Version to Remove the Watermark)

Aplikasi Sirekap merupakan alat yang sangat penting dalam pelaksanaan pemerintahan umum, memainkan peran sentral dalam mendukung para warga negara dalam proses demokrasi [2].



2.1.5 Machine Learning

Menurut Arthur Samuel, *Machine Learning* didefinisikan sebagai disiplin ilmu yang memungkinkan komputer untuk belajar dan berkembang tanpa harus diprogram secara langsung [14]. *Machine Learning* adalah bidang ilmu dari *Computer Science* yang mempelajari bagaimana memberikan kecerdasan pada komputer atau mesin. Fokus *Machine Learning* adalah pada pengembangan program komputer yang dapat mengakses dan mempelajari data untuk meningkatkan kinerjanya. Adanya proses pelatihan, pembelajaran, atau training merupakan ciri khas dari *Machine Learning*.

Machine Learning memerlukan data latih yang memadai untuk membangun model algoritma yang cerdas, serta membutuhkan verifikasi akurasi melalui data pengujian. Beberapa algoritma *Machine Learning* yang paling populer meliputi *Supervised Learning*, *Unsupervised Learning*, *Semi-supervised Learning*, *Reinforcement Learning*. Keempat jenis *Machine Learning* ini memiliki karakteristik dan mekanisme kerja yang berbeda-beda, yang dapat dijelaskan sebagai berikut :

- 1) ***Supervised ML*** : *output* diberikan label dan *input* sudah dipetakan sesuai dengan *output* yang sesuai.
- 2) ***Unsupervised ML*** : *output* tidak berlabel dan data berkelompok sesuai dengan pola atau kemiripan data.
- 3) ***Semi-supervised ML*** : *output* ada yang berlabel dan ada yang tidak berlabel.

Protected by PDF Anti-Copy Free

(Upgrade to Pro Version to Remove the Watermark)

4) **Reinforcement ML** : mirip dengan *Supervised ML*, namun

output yang diberikan terdiri dua jenis yaitu hadiah atau

hukuman



2.1.6 Support Vector Machine (SVM)

Support Vector Machine (SVM) merupakan model pembelajaran mesin yang termasuk dalam kategori *Supervised Learning* [5] yaitu sebuah metode klasifikasi dimana sekumpulan data diberikan label untuk mengklasifikasikan kelas [8]. *Support Vector Machine (SVM)* sering digunakan dalam klasifikasi data teks karena kemampuannya menangani data berdimensi tinggi dan bersifat kompleks [15]. Prinsip kerja *Support Vector Machine (SVM)* adalah mencari *hyperplane* optimal dengan margin maksimum yang berfungsi sebagai batas keputusan untuk memisahkan dua kelas yang berbeda [16].

Support Vector Machine (SVM) memiliki kemampuan untuk mentransformasikan data ke ruang dimensi yang lebih tinggi menggunakan fungsi kernel, yang dikenal sebagai *kernel trick*, sehingga memungkinkan data untuk dipisahkan dengan lebih baik dibandingkan dengan algoritma klasifikasi lainnya [17]. Algoritma *Support Vector Machine (SVM)* yang dioptimalkan dengan *Grid Search Cross-Validation (GridSearchCV)* dapat digunakan untuk meningkatkan nilai akurasi [9] dalam analisis sentimen ulasan pengguna aplikasi Sirekap, membantu mengidentifikasi sentimen positif, negatif, atau netral dengan lebih akurat sesuai kebutuhan analisis.

Menurut Feldman dan Sanger (2007), algoritma *Support Vector Machine (SVM)* adalah metode yang sangat efektif untuk menyelesaikan masalah klasifikasi. Selain itu, algoritma ini juga merupakan salah satu metode *machine learning* yang paling populer dan memberikan akurasi tinggi dalam klasifikasi teks [18].

Protected by PDF Anti-Copy Free

(Upgrade to Pro Version to Remove the Watermark)

Teori yang mendasari *Support Vector Machine (SVM)* telah berkembang sejak tahun 1960-an, namun baru diperkenalkan secara resmi oleh Boser, dan Guyon pada tahun 1992 [19]. Adapun kelebihan dan kekurangan algoritma *Support Vector Machine (SVM)* adalah sebagai berikut :

a) Kelebihan *Support Vector Machine (SVM)*

1. Tingkat Akurasi Tinggi

Support Vector Machine (SVM) seringkali memberikan hasil prediksi yang sangat baik, terutama untuk masalah klasifikasi biner, dengan kemampuan memisahkan kelas-kelas secara optimal.

2. Cocok untuk Dimensi Tinggi

Algoritma ini unggul dalam menangani dataset dengan jumlah fitur yang besar, karena dapat menemukan *hyperplane* yang memisahkan kelas-kelas secara efektif.

3. Penggunaan Memori Efisien

Model *Support Vector Machine (SVM)* hanya bergantung pada beberapa titik data penting (*support vectors*), sehingga lebih hemat dalam penggunaan memori.

4. Kemampuan Mengatasi Data Non-Linear

Dengan penerapan teknik kernel, *Support Vector Machine (SVM)* mampu bekerja pada data yang tidak linear dengan mengubah data ke ruang dimensi yang lebih tinggi.

5. Kemampuan Generalisasi yang Baik

Algoritma ini cenderung menghindari *overfitting*, terutama pada dataset yang kecil hingga menengah.

Protected by PDF Anti-Copy Free

(Upgrade to Pro Version to Remove the Watermark)

b) Kekurangan *Support Vector Machine (SVM)*

1. Komputasi yang Lambat

data dataset yang besar, baik dalam hal jumlah data dan fitur, proses pelatihan bisa menjadi sangat lambat.

2. Sensitif terhadap Skala Fitur

Fitur-fitur di dalam dataset perlu dinormalisasi sebelum menggunakan *Support Vector Machine (SVM)*.

3. Kurang Efisien untuk Dataset yang Besar

Meskipun efektif untuk dataset kecil, *Support Vector Machine (SVM)* bisa menjadi kurang efisien dalam hal komputasi dan memori pada dataset yang sangat besar.

2.1.7 *Grid Search Cross-Validation (GridSearchCV)*

GridSearchCV merupakan bagian dari modul *scikit-learn* yang melakukan validasi [9] untuk menentukan kombinasi parameter yang paling optimal secara sistematis melalui validasi silang [20]. Dengan menggunakan *Grid Search Cross-Validation*, pemilihan model dan penyetelan *hyperparameter* menjadi lebih efisien. Metode ini secara otomatis melakukan validasi pada setiap kombinasi model dan *hyperparameter*, sehingga dapat menghemat waktu dalam proses pengerjaan.

Menurut Marketbrew (2024), *Grid Search Cross-Validation* adalah metode optimasi yang memungkinkan pemilihan parameter algoritma secara sistematis dengan menguji berbagai kombinasi nilai yang telah ditentukan sebelumnya. Teknik ini bertujuan untuk meningkatkan akurasi model sekaligus meminimalkan risiko *overfitting* [21]. *Grid Search Cross-Validation* akan menghasilkan

Protected by PDF Anti-Copy Free

(Upgrade to Pro Version to Remove the Watermark)

skor terbaik atau *grid score* yang digunakan untuk menentukan parameter optimasi dalam proses klasifikasi [22].

2.1.8 Synthetic Minority Oversampling Technique Edited Nearest Neighbors (SMOTEENN)

SMOTEENN merupakan kombinasi dari metode *SMOTE* dan *ENN*, di mana *SMOTE* berfungsi sebagai teknik *oversampling*, sedangkan *ENN* digunakan untuk melakukan *undersampling*. *SMOTE* pertama kali diperkenalkan oleh Nithes V. Chawla, adalah teknik turunan dari *oversampling* yang berfungsi menyesuaikan dataset yang tidak seimbang melalui pembuatan data sintesis baru untuk kelas minoritas, sehingga mampu meningkatkan performa metode klasifikasi [23]. Dengan optimasi *SMOTE*, algoritma *Machine Learning* dapat bekerja lebih efektif, karena tidak hanya berfokus pada kata-kata dengan jumlah klasifikasi tertinggi, tetapi juga mengenali kata-kata yang termasuk dalam klasifikasi yang lebih rendah [24].

Sedangkan, *Edited Nearest Neighbors (ENN)* yang diperkenalkan oleh Wilson, merupakan teknik *undersampling* pada kelas mayoritas dilakukan dengan menghapus sampel dari kelas mayoritas yang labelnya berbeda dengan data-data tetangga terdekatnya. *Edited Nearest Neighbors (ENN)* digunakan untuk meningkatkan kualitas dataset dengan menghapus sampel yang dianggap *noise* atau ambigu [25].

2.1.9 Python



Gambar 2.2 Logo Python

Protected by PDF Anti-Copy Free

(Upgrade to Pro Version to Remove the Watermark)

Python adalah bahasa pemrograman berorientasi objek yang dapat digunakan untuk mengembangkan perangkat lunak dan dapat dilakukan di berbagai sistem operasi. Saat ini, *Python* juga populer di bidang ilmu data dan analisis. Ini karena oleh dukungan bahasa *Python* terhadap *library* yang mencakup fungsi analisis data dan fungsi pembelajaran mesin, alat *preprocessing* data, dan visualisasi data [26]. *Python* secara umum mendukung berbagai paradigma pemrograman, tidak terbatas pada pemrograman berorientasi objek, imperatif, dan pemrograman fungsional.

Salah satu fitur utama *Python* adalah sifatnya yang dinamis, serta dilengkapi dengan manajemen memori otomatis. Dalam melakukan analisis data, *Python* memiliki *library* yang sering digunakan yaitu:

1. *NumPy* merupakan *library* yang digunakan untuk melakukan perhitungan matematika maupun perhitungan ilmiah.
2. *Matplotlib* merupakan *library* yang digunakan untuk visualisasi data.
3. *Pandas* digunakan untuk memindahkan data saat melakukan analisis maupun *cleaning* data dan rangkuman statistic.
4. *Scraper* atau yang lebih dikenal *Scraping* adalah *tools* atau *library* pada *Python* yang digunakan untuk mengumpulkan data dari *Google Play Store*.

2.1.10 Text Mining

Text Mining adalah suatu teknik dalam ilmu komputer yang digunakan untuk menangani masalah informasi dalam jumlah besar dengan menggabungkan berbagai teknik dari *data mining*, *machine learning*, *natural language processing*, *information retrieval*, dan *knowledge management*. Seperti halnya *data mining*, *text mining* bertujuan untuk mengekstrak informasi yang berguna

Protected by PDF Anti-Copy Free

(Upgrade to Pro Version to Remove the Watermark)

dari sumber data dengan mengidentifikasi dan mengeksplorasi pola-pola yang tersembunyi. Sumber data dalam *text mining* berupa kumpulan dokumen seperti koran, majalah, artikel, surat, atau laporan penelitian termasuk jurnal, tugas akhir, atau tesis [18].

Proses dalam *text mining* mengikuti tahapan yang serupa dengan *Knowledge Discovery in Database (KDD)* dalam *data mining*. *Text Mining* memiliki kemampuan untuk mengumpulkan sejumlah besar data dan kemudian memprosesnya untuk mengekstrak informasi bermanfaat dari kumpulan teks. Analisis sentimen di media sosial, ekstraksi informasi untuk mengklasifikasi teks, dan peringkasan teks adalah beberapa aplikasi *text mining* [11].

2.1.11 Web Scraping

Web Scraping adalah proses ekstraksi data dari suatu situs web. Tujuan *web scraping* adalah mengumpulkan informasi dengan mencari dan mengekstraksi data dari sumber tertentu. Fokus utama dari *web scraping* adalah memperoleh data secara sistematis. Manfaat dari *web scraping* antara lain adalah untuk mendapatkan informasi yang lebih spesifik sehingga memudahkan dalam pencarian data atau informasi yang dibutuhkan [27]. *Web Scraping* atau *Web Crawling* di sisi lain jauh lebih cepat dan lebih efektif, dan dapat digunakan untuk mengumpulkan dan menyusun data dari ribuan, atau lebih dari jutaan, halaman untuk diproses dan diambil informasinya [28].

2.1.12 Preprocessing

Preprocessing adalah salah satu teknik dalam *data mining* yang digunakan untuk mentransformasikan data mentah menjadi format yang lebih mudah dipahami. Proses ini dilakukan untuk mengatasi berbagai masalah, seperti redundansi data, *noisy data*,

Protected by PDF Anti-Copy Free

(Upgrade to Pro Version to Remove the Watermark)

dan data yang hilang [29]. Tahap *preprocessing* merupakan langkah awal dalam pengolahan data agar data dapat diproses dan siap digunakan untuk keperluan penelitian, termasuk analisis sentimen [30].



2.1.13 Klasifikasi

Klasifikasi merupakan proses untuk menemukan model yang dapat membedakan antara kelas-kelas dari data yang tersedia, dengan tujuan agar model tersebut dapat digunakan untuk memprediksi kelas dari data yang belum diketahui. Klasifikasi adalah salah satu tugas penting dalam berbagai aplikasi, seperti kategorisasi teks, pengenalan nada, dan pengelompokan data. Seiring berkembangnya teknologi kecerdasan buatan, *Machine Learning* juga semakin berkembang dan digunakan secara luas untuk menyelesaikan berbagai masalah dalam membantu pekerjaan manusia.

Beberapa metode klasifikasi dalam *Machine Learning* meliputi *Random Forest*, *Naïve Bayes*, *Classification and Regression Trees (CART)*, *Support Vector Machines (SVM)*, dan lain-lain [9].

2.1.14 Confusion Matrix

Confusion matrix adalah tabel yang menunjukkan jumlah data uji yang diklasifikasikan dengan tepat serta jumlah data uji yang mengalami kesalahan klasifikasi [31]. *Confusion Matrix* juga dikenal sebagai *error matrix* merupakan tata letak tabel khusus yang memungkinkan visualisasi kinerja suatu algoritma, biasanya pembelajaran yang diawasi (dalam pembelajaran tanpa pengawasan biasanya disebut matriks yang cocok).

Protected by PDF Anti-Copy Free

(Upgrade to Pro Version to Remove the Watermark) 2.2 Penelitian Terdahulu Yang Relevan


Beberapa penelitian sebelumnya telah dilakukan oleh dosen maupun pakar lain untuk analisis sentimen ulasan pengguna. Dibawah ini merupakan penelitian yang relevan dengan penelitian yang akan dilakukan :

Tabel 2.1 Penelitian Relevan

No	Nama peneliti (Tahun)	Judul	Metode yang digunakan	Hasil Penelitian
1	Joko Setyanto, Theopilus Bayu Sasongko (2024) [7]	Sentiment Analysis of Sirekap Application Users Using the Support Vector Machine Algoritma	Support Vector Machine	Hasil penelitian menggunakan algoritma Support Vector Machine mendapatkan akurasi sebesar 81%, presisi 87%, dan recall 85% dengan rasio perbandingan 80:20 dari jumlah 1791 data ulasan.
2	Muhamad Fajar Yudhistira Herjanto, Carudin (2024) [2]	Analisis Sentimen Ulasan Pengguna Aplikasi Sirekap Pada Play Store Menggunakan Algoritma Random Forest Classifier	Random Forest Classifier	Hasil penelitian menggunakan algoritma Random Forest Classifier mendapatkan akurasi sebesar 74%, presisi 75%, recall 74%, f1-score 74%, dan rata-rata skor sebesar 74% dari 1000 data yang diuji dari total 5000 data. Data tersebut terdiri dari 500 ulasan dengan label negatif dan 500

Protected by PDF Anti-Copy Free

(Upgrade to Pro Version to Remove the Watermark)

				ulasan dengan label positif.
3	Alfio Kusur Ermatita, Helena Nurramdhani Irmanda (2022) [5]	 Analisis Sentimen Ulasan Aplikasi Indodax di Google Play Store Menggunakan Metode Support Vector Machine	Support Vector Machine	Hasil klasifikasi sentimen dengan SVM pada aplikasi Indodax menggunakan 1138 ulasan dan tiga rasio data latih-uji: 60:40, 70:30, dan 80:20. Akurasi tertinggi 85% dicapai pada rasio 80:20, dengan 196 dari 228 ulasan diklasifikasikan dengan benar.
4	Indah Aida Sapitri, Yusra, Muhammad Fikry (2023) [13]	Pengklasif ikasian Sentimen Ulasan Aplikasi Whatsapp Pada Google Play Store Menggunakan Metode Support Vector Machine	Support Vector Machine	Hasil penelitian menyatakan bahwa proses klasifikasi ulasan WhatsApp di Google Play Store dengan SVM menghasilkan akurasi 82% (rasio 90:10) dan 81% (rasio 80:20). Parameter terbaik pada rasio 90:10 adalah $C=1,0$ dan $\gamma=1,0$ dengan akurasi rata-rata 68%, sedangkan pada rasio 80:20 adalah $C=0,9$ dan $\gamma=0,7$ dengan akurasi 67%.
5	M. Nurul Muttaqin,	Analisis Sentimen Aplikasi	Support Vector Machine dan	Hasil penelitian menyatakan bahwa metode SVM

Protected by PDF Anti-Copy Free

(Upgrade to Pro Version to Remove the Watermark)

Iqbal Kharisudin (2021) [16]	Gojek Mengguna <i>Support Vector Machine</i> Dan <i>K - Nearest Neighbor</i>	<i>K -Nearest Neighbor</i>	dengan kernel linear dan $C=1$ menghasilkan akurasi 87,98%, lebih unggul dari KNN dengan $K=22$ yang memiliki akurasi 82,14%. SVM terbukti lebih baik dalam mengklasifikasikan ulasan pengguna Gojek di <i>Google Play Store</i> dibandingkan KNN.
------------------------------------	---	--------------------------------	--

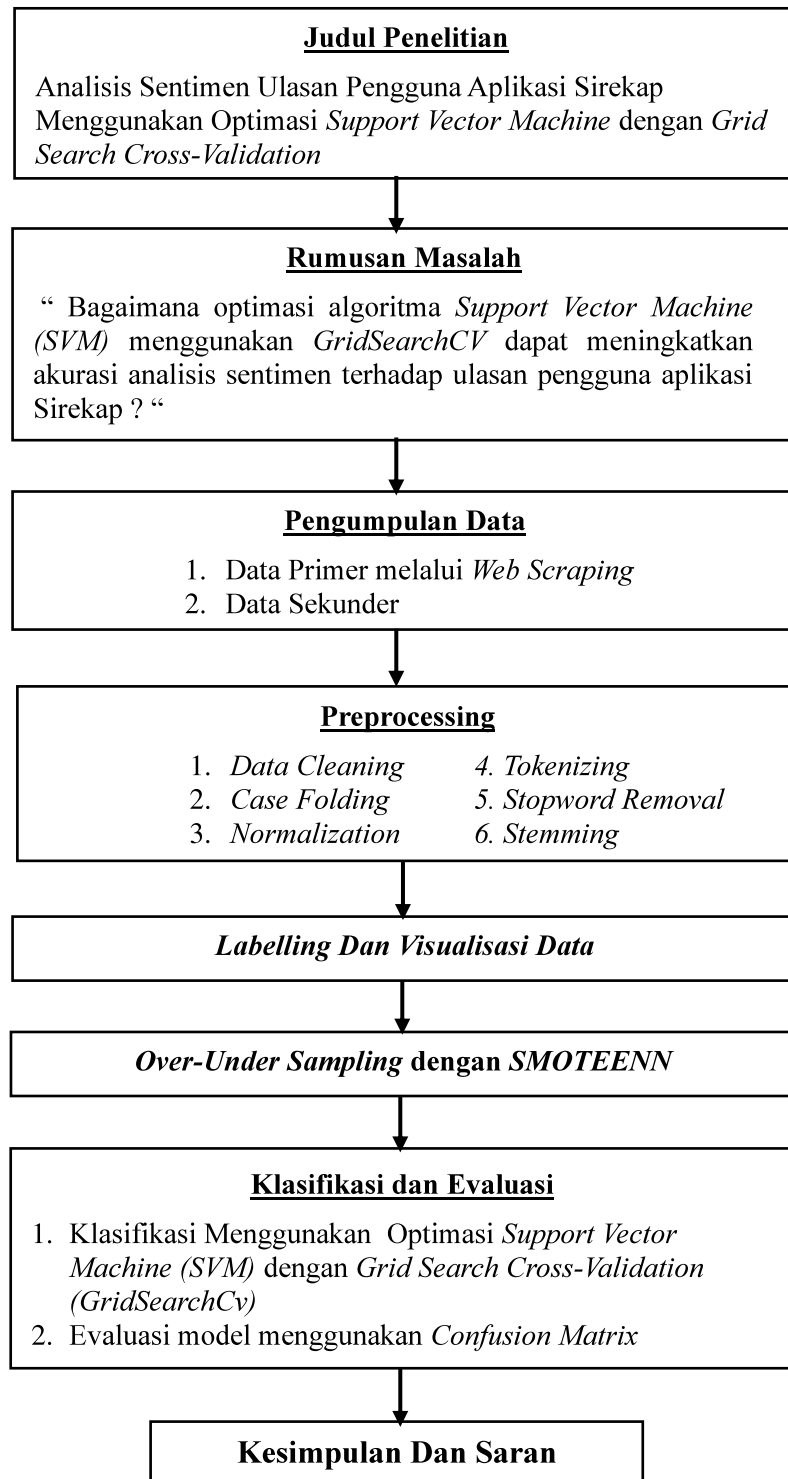
Dari hasil penelitian yang sudah dilakukan sebelumnya membuktikan bahwa algoritma klasifikasi *Support Vector Machine (SVM)* mendapatkan hasil akurasi yang baik. Sehingga, algoritma *Support Vector Machine (SVM)* dapat digunakan dalam penelitian ini. Terlebih lagi karena algoritma *Support Vector Machine (SVM)* sering digunakan dalam klasifikasi data teks karena kemampuannya menangani data berdimensi tinggi dan bersifat kompleks [15]. Pada penelitian ini, algoritma *Support Vector Machine (SVM)* yang dioptimalkan dengan *Grid Search Cross-Validation (GridSearchCV)* dapat digunakan untuk meningkatkan nilai akurasi [9] dalam analisis sentimen ulasan pengguna aplikasi Sirekap.

Protected by PDF Anti-Copy Free

(Upgrade to Pro Version to Remove the Watermark)

2.3 Kerangka Berpikir

Dalam melakukan penelitian ini, terdapat pola dalam berpikir untuk menentukan bentuk alur penelitian yang akan dilakukan oleh Peneliti. Adapun alur proses penelitian sebagai berikut.



Gambar 2.3 Kerangka Berpikir



3.1 Analisa Sistem

3.1.1 Analisa Sistem Yang Berjalan

Sistem Informasi Rekapitulasi (Sirekap) adalah perangkat aplikasi yang tersedia di *Google Play Store* berbasis teknologi informasi. Sirekap memiliki peranan penting dalam proses demokrasi, khususnya pada pemilu di Indonesia yakni sebagai sarana publikasi hasil perhitungan suara dan proses rekapitulasi hasil perhitungan suara. Saat ini, Sirekap telah menerima sekitar 37.000 ulasan di *Google Play Store*. Ulasan ini merupakan sumber data penting bagi pengembang aplikasi untuk memahami persepsi pengguna, mengidentifikasi kebutuhan, serta menentukan langkah perbaikan atau pengembangan yang relevan.

Namun, ulasan yang diterima belum dikelola secara terstruktur untuk menghasilkan analisis sentimen yang mendalam. Hal ini menyebabkan *feedback* dari pengguna belum dimanfaatkan secara optimal dalam pengembangan aplikasi. Selain itu, sifat ulasan yang tidak terstruktur memerlukan pendekatan khusus untuk memahami penilaian pengguna secara efektif. Salah satu pendekatan yang dapat diterapkan adalah analisis sentimen, yang dapat membantu memahami opini masyarakat terhadap aplikasi Sirekap. Namun, ada beberapa tantangan utama yang perlu diatasi diantaranya :

1. Ketidakseimbangan Kelas Data

Ulasan pengguna aplikasi Sirekap menunjukkan adanya ketidakseimbangan kelas yang signifikan, dimana ulasan negatif lebih banyak dibandingkan dengan ulasan positif dan netral.

Protected by PDF Anti-Copy Free

(Upgrade to Pro Version to Remove the Watermark)

2. Pemilihan Parameter Model Yang Kurang Optimal

Penelitian sebelumnya yang menggunakan optimasi *Support Vector Machine (SVM)* dengan *Grid Search Cross-Validation* [1] ulasan menghasilkan analisis sentimen yang kurang akurat. Hal ini disebabkan oleh pemilihan parameter model yang tidak optimal sehingga performa klasifikasi tidak maksimal.

3. Volume Data Yang Besar

Dengan jumlah ulasan mencapai puluhan ribu, proses analisis manual menjadi tidak efisien dan memakan waktu. Diperlukan pendekatan otomatis yang dapat mengolah data dalam jumlah besar dengan akurasi tinggi.

Melalui penelitian ini, pendekatan analisis sentimen dengan optimasi *Support Vector Machine (SVM)* menggunakan *Grid Search Cross-Validation* diharapkan mampu mengatasi tantangan-tantangan tersebut. Pendekatan ini diharapkan dapat meningkatkan akurasi klasifikasi sentimen ulasan pengguna dan memberikan wawasan yang lebih akurat kepada pengembang aplikasi Sirekap.

3.1.2 Alternatif Pemecahan Masalah

Penelitian ini menggunakan metode kuantitatif untuk menganalisis data secara numerik dan menguji hipotesis guna mengembangkan pemahaman ilmiah [32]. Beberapa alternatif solusi untuk mengatasi permasalahan pada aplikasi Sirekap meliputi :

1. Implementasi *Algoritma Support Vector Machine (SVM)*

Algoritma Support Vector Machine (SVM) digunakan untuk melakukan analisis sentimen secara otomatis terhadap ulasan pengguna aplikasi Sirekap menjadi tiga kelas sentimen, yaitu positif, negatif, dan netral.

Protected by PDF Anti-Copy Free

(Upgrade to Pro Version to Remove the Watermark)

2. Pemanfaatan Teknik *SMOTEENN* (*Synthetic Minority Oversampling Technique Edited Nearest Neighbors*)

SMOTEENN digunakan untuk mengatasi ketidakseimbangan kelas dengan meningkatkan jumlah sampel pada kelas minoritas (positif dan netral) dan menghilangkan sampel mayoritas (negatif) yang dinilai kurang representatif dan berpotensi mengurangi performa model.

3. Optimasi Model Dengan *GridSearchCV*

Mengoptimalkan parameter model *Support Vector Machine* (*SVM*) menggunakan *GridSearchCV* untuk mendapatkan kombinasi parameter terbaik yang dapat meningkatkan akurasi model dan performa dalam mengklasifikasikan ulasan.

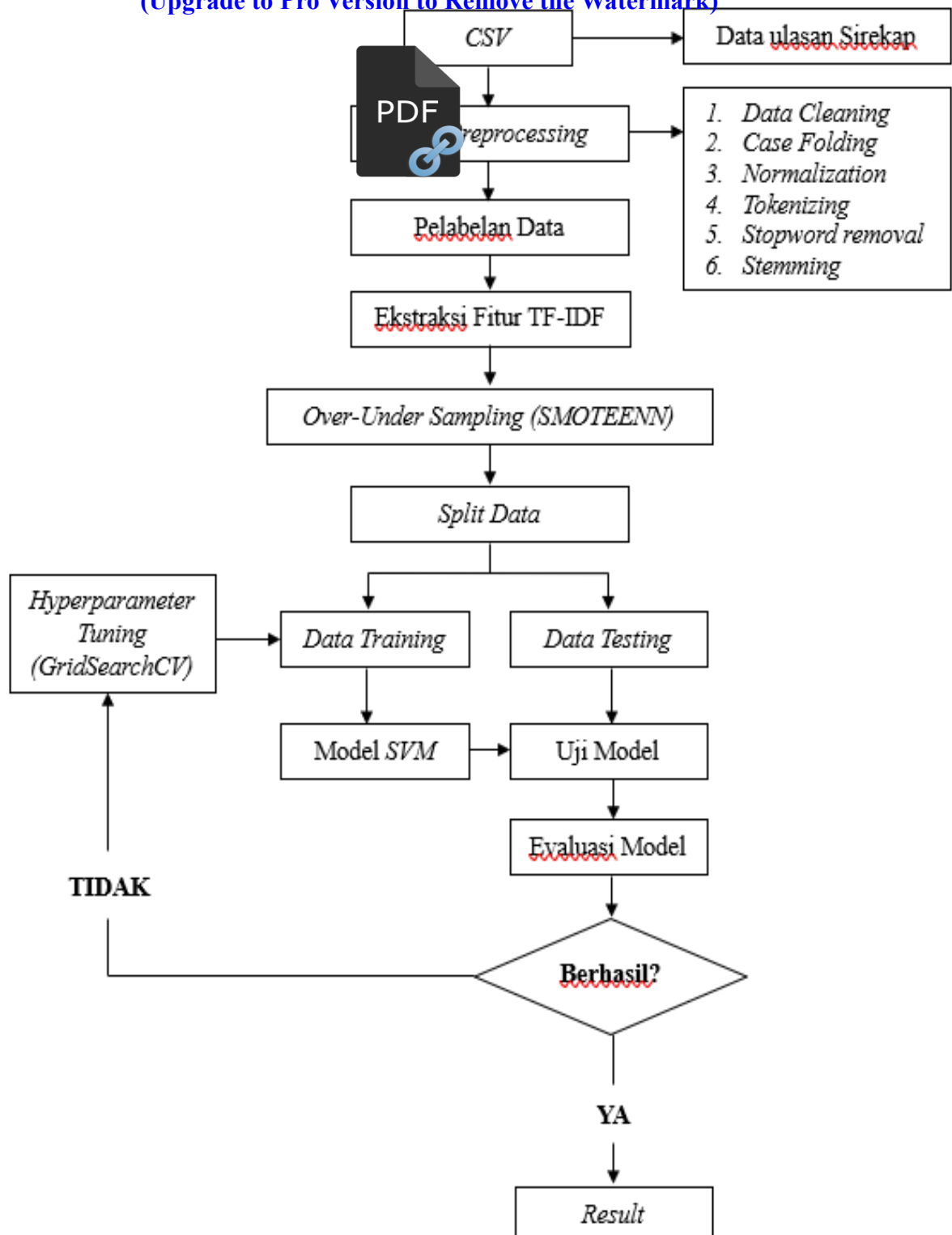
3.1.3 Metode Analisa

Pada penelitian ini, Peneliti menggunakan metode analisis sentimen melalui optimasi *Support Vector Machine* (*SVM*) dengan *Grid Search Cross-Validation* (*GridSearchCV*). Algoritma *Support Vector Machine* (*SVM*) sering digunakan dalam klasifikasi data teks karena kemampuannya menangani data berdimensi tinggi dan bersifat kompleks [15]. Sedangkan, *Grid Search Cross-Validation* (*GridSearchCV*) untuk menentukan kombinasi parameter yang paling optimal secara sistematis melalui validasi silang [20].

Algoritma *Support Vector Machine* (*SVM*) yang dioptimalkan dengan *Grid Search Cross-Validation* (*GridSearchCV*) dapat digunakan untuk meningkatkan nilai akurasi [9] dalam analisis sentimen ulasan pengguna aplikasi Sirekap. Terdapat diagram alur konsep penelitian yang menunjukkan tahapan dari pengumpulan data hingga evaluasi model, sebagaimana tertera pada Gambar 3.1.

Protected by PDF Anti-Copy Free

(Upgrade to Pro Version to Remove the Watermark)



Gambar 3.1 Diagram Konsep Penelitian

Protected by PDF Anti-Copy Free

(Upgrade to Pro Version to Remove the Watermark)

3.2 Teknik Pemilihan Informan (Populasi, Sampel, dan Sampling)

Populasi dalam penelitian ini mencakup seluruh ulasan pengguna aplikasi Sirekap yang di *platform Google Play Store*, yang jumlahnya mencapai 20.000 ulasan dari berbagai pengguna di Indonesia. Dari populasi ini, sebanyak 20.000 ulasan dipilih sebagai sampel penelitian. Teknik sampling yang digunakan adalah *random sampling*, di mana ulasan dipilih secara acak tanpa kriteria khusus selain menggunakan bahasa Indonesia. Pendekatan ini dilakukan untuk memastikan sampel dapat merepresentasikan populasi secara keseluruhan dan mencakup variasi sentimen pada ulasan.

3.2.1 Teknik Pengumpulan Data

Untuk mendapatkan informasi yang bermakna, Peneliti memerlukan data yang relevan dan dapat digunakan dalam penelitian ini. Oleh karena itu, Peneliti menerapkan beberapa metode pengumpulan data yang dapat digunakan dalam proses penelitian, yaitu sebagai berikut :

1. Data Primer

Data primer merupakan data yang diperoleh langsung dari sumber yang akan digunakan. Dalam penelitian ini, data primer yang digunakan adalah ulasan pengguna aplikasi Sistem Informasi Rekapitulasi (Sirekap) di *Google Play Store*, sebanyak 20.000 ulasan berbahasa Indonesia. Data tersebut dikumpulkan menggunakan teknik *scraping* menggunakan *library Google Play Scraper* di *Python* dengan bantuan *tools Jupyter Lab*. Ulasan ini berisi tanggapan atau komentar dari pengguna yang telah mengunduh aplikasi, mencakup berbagai jenis komentar, yakni komentar positif, netral, dan negatif.

2. Data Sekunder

Data sekunder diperoleh dari sumber-sumber yang telah tersedia sebelumnya untuk membantu Peneliti dalam penelitian. Data sekunder ini dapat mencakup wawasan literatur, seperti

Protected by PDF Anti-Copy Free

(Upgrade to Pro Version to Remove the Watermark)

buku, jurnal, artikel ilmiah, atau media sosial yang bisa diakses menggunakan internet [12].

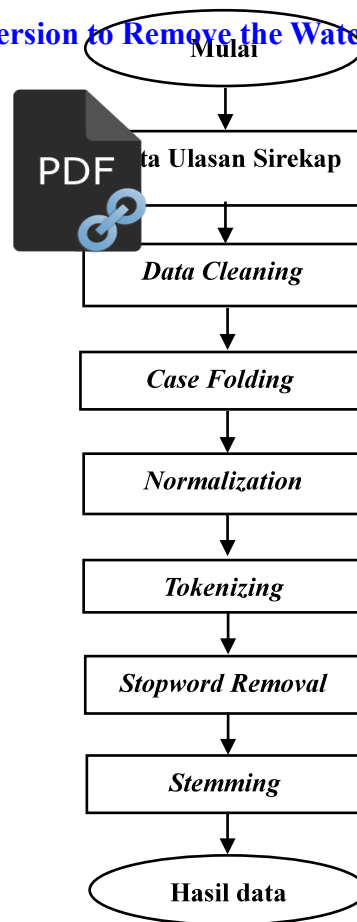
3.2.2 Teknik Analisa

Proses analisis data dalam penelitian ini dilakukan melalui beberapa tahapan sebagai berikut :

1. Tahap *Preprocessing*

Sebelum melakukan pengolahan data, tahap awal yang dilakukan adalah *Preprocessing* menggunakan bahasa pemrograman *Python*. Tahap ini meliputi beberapa langkah, salah satunya adalah pengambilan data ulasan pengguna aplikasi Sirekap dari *Google Play Store* melalui teknik *Web Scraping* dengan library *Google-Play-Scraper*. Setelah itu dilakukan proses *Data Cleaning*, *Case Folding*, *Normalization*, *Tokenizing*, *Stopword Removal*, dan *Stemming* yang tertera pada gambar berikut :

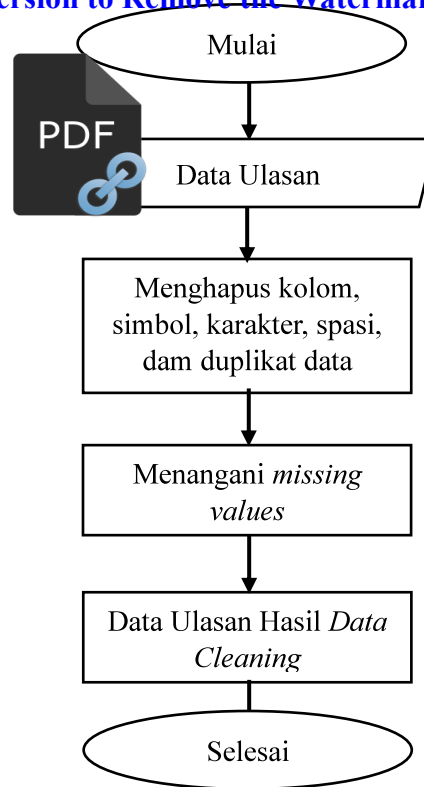
Protected by PDF Anti-Copy Free
 (Upgrade to Pro Version to Remove the Watermark)



Gambar 3.2 Tahap *Preprocessing*

- a) *Data Cleaning* merupakan proses menangani *missing values* menghapus duplikat berdasarkan teks ulasan, dan menghilangkan elemen-elemen yang tidak relevan, seperti simbol dan emotikon. Tujuan dari proses ini adalah untuk meningkatkan kualitas data dengan membersihkan data dari gangguan atau *noise* [2]. Tahap *Data Cleaning* dapat dilihat pada gambar berikut :

Protected by PDF Anti-Copy Free
 (Upgrade to Pro Version to Remove the Watermark)



Gambar 3.3 Tahap *Data Cleaning*

Berikut *script* dan hasil proses *Data Cleaning* dapat dilihat dari tabel berikut :

Tabel 3.1 *Script Data Cleaning*

<i>Data Cleaning</i>
<pre> # Hapus kolom yang tidak relevan df_clean = df.drop(columns=['reviewId', 'userName'], errors='ignore') # Tangani missing values df_clean = df_clean.dropna(subset=['content', 'rating']) # Hapus duplikat berdasarkan teks ulasan ('content') df_clean = df_clean.drop_duplicates(subset=['content']) def clean_text(text): # Menghapus spasi berlebih text = re.sub(r'\s+', ' ', text) # Menghapus simbol dan angka, hanya menyisakan huruf dan spasi text = re.sub(r'^[a-zA-Z\s]', '', text) # Menghapus emotikon emoji_pattern = re.compile("[" u"\U0001F600-\U0001F64F" u"\U0001F300-\U0001F5FF" </pre>

Protected by PDF Anti-Copy Free

(Upgrade to Pro Version to Remove the Watermark)

```
u"\U0001F680-\U0001F6FF"
u"\U0001F700-\U0001F77F"
u"\U0001F780-\U0001F7FF"
u"\U0001F800-\U0001F8FF"
u"\U0001F900-\U0001F9FF"
u"\U0001FA00-\U0001FA6F"
u"\U0001FA70-\U0001FAFF"
u"\U00002700-\U000027BF"
"]+", flags=re.UNICODE)
```

```
# Menghapus emotikon
text = emoji_pattern.sub(r'', text)
```

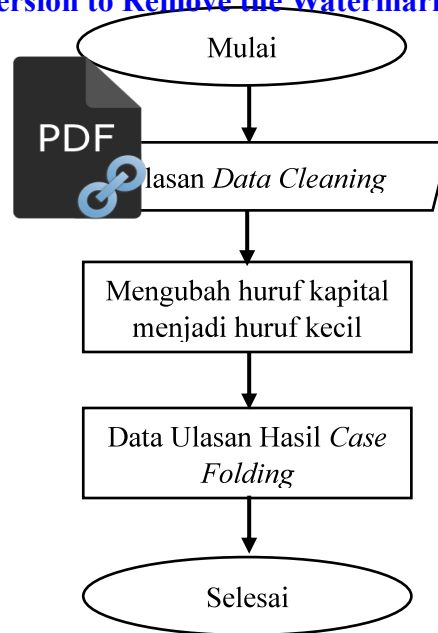
```
return text.strip()
```

Tabel 3.2 Hasil *Data Cleaning*

Ulasan	Data Cleaning
Beban.... Tambahh lah server yg banyak.... Satu indonesia ini loh y makee..... 🧑🏻 🧑🏻 Mempersulit petugas lapangan aja...	Beban Tambahh lah server yg banyak Satu indonesia ini loh y makee Mempersulit petugas lapangan aja

- b) *Case Folding* adalah proses di mana teks ulasan atau string diubah seluruhnya menjadi huruf kecil. Tujuannya adalah untuk memastikan setiap string memiliki konsistensi dalam penyusunan kata, sehingga setiap kata hanya memiliki satu bentuk fitur dalam susunan tersebut [12]. Tahap *Case Folding* dapat dilihat pada gambar berikut :

Protected by PDF Anti-Copy Free
 (Upgrade to Pro Version to Remove the Watermark)



Gambar 3.4 Tahap *Case Folding*

Berikut *script* dan hasil proses *Case Folding* dapat dilihat dari tabel berikut :

Tabel 3.3 *Script Case Folding*

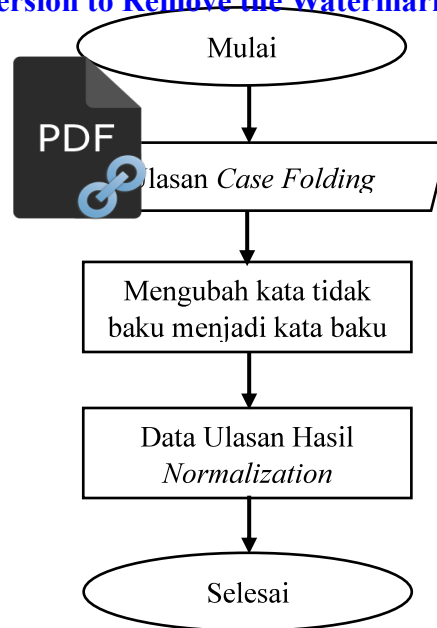
Case Folding
<code>df['content_case_folding'] = df['content'].str.lower()</code>

Tabel 3.4 Hasil *Case Folding*

Ulasan Data Cleaning	Case Folding
Beban Tambahh lah server yg banyak Satu indonesia ini loh h yg makee Mempersulit petugas lapangan aja	beban tambahh lah server yg banyak satu indonesia ini loh h yg makee mempersulit petugas lapangan aja

c) *Normalization* adalah proses mengubah dan memperbaiki singkatan atau kata tidak baku menjadi bentuk kata yang lengkap dan bermakna sesuai dengan Kamus Besar Bahasa Indonesia (KBBI) [12]. Pada tahap ini, digunakan daftar dalam format file CSV yang memuat bentuk kata-kata hasil koreksi.

Protected by PDF Anti-Copy Free
 (Upgrade to Pro Version to Remove the Watermark)



Gambar 3.5 Tahap *Normalization*

Berikut *script* dan hasil proses *Normalization* dapat dilihat dari tabel dibawah ini :

Tabel 3.5 Script *Normalization*

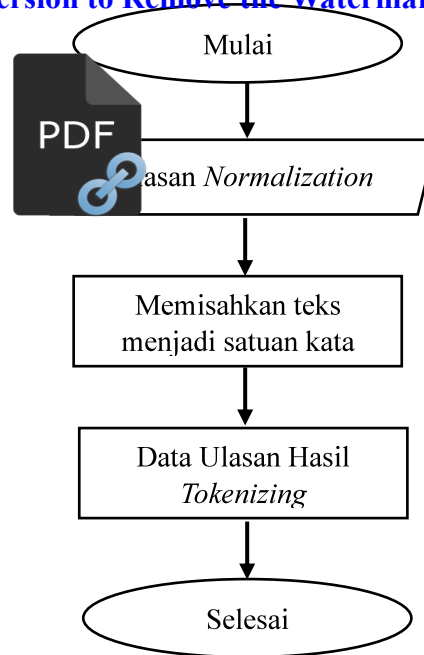
<i>Normalization</i>
<pre> data['content_normalization'], data['Kata_Baku'], data['Kata_Tidak_Baku'], data['Kata_Tidak_Baku_Hash'] = zip(*data['content_case_folding'].apply(lambda x: replace_taboo_words(x, kamus_tidak_baku))) </pre>

Tabel 3.6 Hasil *Normalization*

<i>Ulasan Case Folding</i>	<i>Normalization</i>
beban tambahh lah server yg banyak satu indonesia ini loh h yg makee mempersulit petugas lapangan aja	beban tambah lah server yang banyak satu indonesia ini loh yang makee mempersulit petugas lapangan saja

d) *Tokenizing* adalah proses memecah sebuah kalimat pada teks menjadi beberapa satuan kata yang terpisah [17]. Tahap *Tokenizing* dapat dilihat pada gambar berikut :

Protected by PDF Anti-Copy Free
 (Upgrade to Pro Version to Remove the Watermark)



Gambar 3.6 Tahap *Tokenizing*

Berikut *script* dan hasil proses *Tokenizing* dapat dilihat dari tabel dibawah ini :

Tabel 3.7 *Script Tokenizing*

<i>Tokenizing</i>	
<i># Tokenizing teks ulasan (pecah menjadi kata-kata)</i>	
<code>def safe_word_tokenize(text):</code>	
<code>if isinstance(text, str): # Pastikan text adalah string</code>	
<code>return word_tokenize(text)</code>	
<code>return [] # Kembalikan list kosong jika bukan string</code>	
<code>df_clean['tokens']</code>	=
<code>df_clean['stemmed_content'].apply(safe_word_tokenize)</code>	

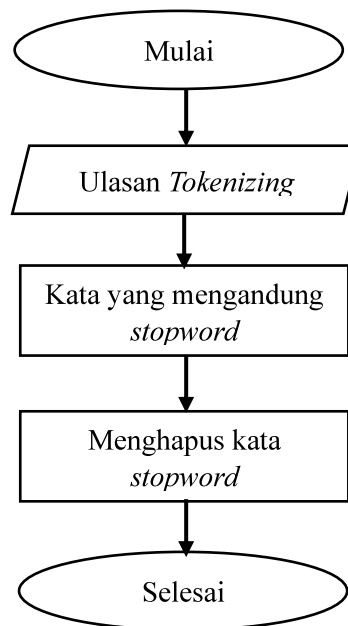
Tabel 3.8 Hasil *Tokenizing*

<i>Ulasan Normalization</i>	<i>Tokenizing</i>
beban tambah lah server	['beban', 'tambah', 'lah',
yang banyak satu indonesia	'server', 'yang', 'banyak',
ini loh yang makee	'satu', 'indonesia', 'ini', 'loh',
mempersulit petugas	'yang', 'makee', 'mempersulit',
lapangan saja	'petugas', 'lapangan', 'saja']

Protected by PDF Anti-Copy Free

(Upgrade to Pro Version to Remove the Watermark)

e) *Stopword Removal* adalah proses menghilangkan kata-kata yang sering muncul tetapi tidak memiliki makna penting dalam proses analisis [5]. Tahap *Stopword Removal* dapat dilihat pada gambar berikut :



Gambar 3.7 Tahap *Stopword Removal*

Berikut *script* dan hasil proses *Stopword Removal* dapat dilihat dari tabel dibawah ini :


Tabel 3.9 *Script Stopword Removal*

<i>Stopword Removal</i>
<pre> def remove_stopwords(text): if isinstance(text, str): return ' '.join([word for word in text.split() if word not in stop_words]) return " # Kembalikan string kosong jika bukan string (termasuk NaN) # Terapkan stopwords removal pada kolom 'content_filtered' df_clean['content_stopword_removal'] = df_clean['content_filtered'].apply(remove_stopwords) </pre>

Protected by PDF Anti-Copy Free

(Upgrade to Pro Version to Remove the Watermark)

Tabel 3.11 *Script Stemming*

<i>Stemming</i>	
<pre># Inialisasi ps = PorterS # Fungsi untuk an stemming def stemming_text(text): if isinstance(text, str): # Pastikan text adalah string # Stemming stemmed_tokens = [ps.stem(word) for word in text.split()] return ' '.join(stemmed_tokens) return text # Kembalikan text asli jika bukan string # Terapkan fungsi stemming pada kolom 'content' df_clean['stemmed_content'] = df_clean['content_stopword_removal'].apply(stemming_text)</pre>	

Tabel 3.12 Hasil *Stemming*

<i>Ulasan Stopword Removal</i>	<i>Stemming</i>
['beban', 'tambah', 'lah', 'server', 'yang', 'banyak', 'satu', 'indonesia', 'ini', 'loh', 'yang', 'makee', 'mempersulit', 'lapang saja', 'petugas', 'lapangan', 'saja']	beban tambah lah server yang banyak satu indonesia ini loh yang makee sulit tugas lapang saja


2. Tahap *Labeling*

Labeling atau pelabelan data adalah proses klasifikasi di mana kata atau kalimat diberi tag sesuai dengan kategori tertentu, seperti positif, netral, atau negatif, berdasarkan sentimen yang diungkapkan. Dalam konteks analisis sentimen, proses ini berfungsi untuk menentukan apakah suatu pernyataan menunjukkan emosi atau opini yang menyenangkan (positif), tidak memihak (netral), atau tidak menyenangkan (negatif). Klasifikasi ini membantu mengorganisir dan memahami data teks sehingga bisa digunakan untuk analisis lebih lanjut, seperti dalam pengukuran kepuasan pengguna, evaluasi produk, atau memahami umpan balik secara keseluruhan. Pelabelan data dapat dilihat dari tabel berikut :

Protected by PDF Anti-Copy Free

(Upgrade to Pro Version to Remove the Watermark)

Tabel 3.13 *Labeling*

	Sentimen
	Positif
	Positif
	Netral
2	Negatif
1	Negatif

3. Tahap Visualisasi Data

Setelah proses pelabelan, data akan diubah menjadi bentuk grafik untuk visualisasi yang lebih menarik menggunakan *library Matplotlib*. *Matplotlib* adalah pustaka *Python* yang bersifat lintas *platform*, dirancang khusus untuk menghasilkan plot 2D dengan kualitas tinggi. *Library* ini memfasilitasi pembuatan berbagai jenis grafik, seperti histogram, spektrum, dan grafik batang, menjadikannya lebih mudah dan efisien untuk digunakan [12]. Visualisasi ini penting dalam analisis data karena membantu dalam mengidentifikasi pola dan tren, serta menyajikan informasi yang relevan dengan cara yang lebih jelas dan mudah dipahami.

4. Tahap Pembobotan Kata *TF – IDF*

Setelah tahap pelabelan data, dilakukan pembobotan kata dengan menggunakan metode *Term Frequency – Inverse Document Frequency (TF-IDF)*. Metode *TF-IDF (Term Frequency-Inverse Document Frequency)* adalah teknik yang digunakan untuk menilai seberapa penting sebuah kata dalam sebuah dokumen relatif terhadap kumpulan dokumen lainnya. Bobot setiap kata dihitung berdasarkan frekuensinya dalam satu dokumen (*Term Frequency*) serta seberapa sering kata tersebut muncul di seluruh dokumen (*Inverse Document Frequency*). Semakin sering kata muncul dalam satu dokumen namun jarang ditemukan di dokumen lain, semakin tinggi nilai bobotnya,

Protected by PDF Anti-Copy Free

(Upgrade to Pro Version to Remove the Watermark)

menandakan bahwa kata tersebut memiliki signifikansi dalam konteks dokumen tersebut [12].

5. Tahap *Over-sampling* dengan *Synthetic Minority Oversampling Technique Edited Nearest Neighbors (SMOTEENN)*

Penerapan metode *under-over sampling* lebih efektif dibandingkan penggunaan *oversampling* atau *undersampling* secara terpisah. Pendekatan ini memanfaatkan kelebihan dari kedua teknik, yaitu meningkatkan representasi kelas minoritas melalui *oversampling* dan mengurangi *noise* atau data tidak relevan dengan *undersampling*. Hal ini memungkinkan model untuk lebih fokus pada pola penting dalam dataset, sehingga meningkatkan akurasi dan keandalan klasifikasi secara keseluruhan.

Pada tahap *oversampling* ini menggunakan metode *Synthetic Minority Oversampling Technique (SMOTE)*. Metode ini berfungsi dengan menyesuaikan dataset yang tidak seimbang melalui pembuatan data sintesis baru untuk kelas minoritas, sehingga mampu meningkatkan performa metode klasifikasi [23]. Sedangkan, pada tahap *undersampling* menggunakan metode *Edited Nearest Neighbors (ENN)*, yaitu untuk meningkatkan kualitas dataset dengan menghapus sampel yang dianggap *noise* atau ambigu [25].

6. Tahap Pembagian Data Latih dan Data Uji

Selanjutnya pembagian data latih dan data uji menggunakan lima rasio perbandingan, yaitu 50:50, 60:40, 70:30, 80:20, dan 90:10. Total data sebanyak 17.506 data ulasan setelah dilakukan proses *preprocessing*. Namun, setelah melalui tahap *over-under sampling* menggunakan *SMOTEENN* jumlah

Protected by PDF Anti-Copy Free

(Upgrade to Pro Version to Remove the Watermark)

data menjadi 20.215 data ulasan. Berikut pembagian data dapat dilihat pada tabel dibawah ini :

Tabel 3.14 Pembagian Data Setelah *SMOTEENN*

Rasio Pembagian	Jumlah Data Uji	Jumlah Data Latih	Jumlah Data Uji
50%	50%	10.107	10.108
60%	40%	12.129	8.086
70%	30%	14.150	6.065
80%	20%	16.172	4.043
90%	10%	18.193	2.022

7. Tahap *Hyperparameter Tuning* dengan *GridSearchCV*

Hyperparameter Tuning adalah proses untuk memilih setelan terbaik dari *hyperparameter* dalam sebuah model *machine learning*. Pada penelitian ini menggunakan *Grid Search Cross-Validation (GridSearchCV)*. *Hyperparameter tuning* dengan *GridSearchCV* bertujuan untuk menemukan kombinasi optimal dari parameter seperti *complexity (C)*, *gamma (γ)*, dan tipe kernel yang memberikan hasil terbaik pada model *Support Vector Machine (SVM)*. Proses ini melibatkan pengujian berbagai nilai *hyperparameter* untuk menentukan konfigurasi yang menghasilkan kinerja model paling optimal.

Pada penelitian [7] yang dilakukan oleh Joko Setyanto dan Theopilus Bayu Sasongko, *Hyperparameter Tuning* dengan *GridSearchCV* menggunakan algoritma *Support Vector Machine (SVM)* hanya melibatkan tiga kernel (*Linear, Polynomial, RBF*) dan parameter *C (complexity)* dengan jangkauan nilai (0.01, 0.05, 0.25, 0.5, 0.75, 1, dan 10). Namun, pada penelitian ini, Peneliti menggunakan empat kernel (*RBF, Linear, Polynomial, dan Sigmoid*), parameter *C (complexity)* dengan jangkauan nilai

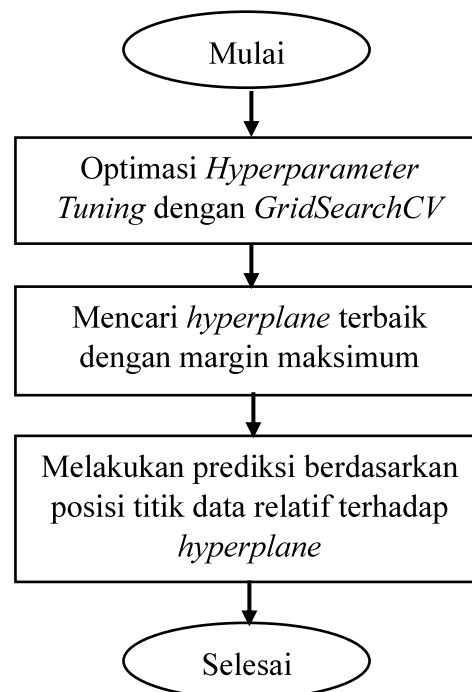
Protected by PDF Anti-Copy Free

(Upgrade to Pro Version to Remove the Watermark)

(0.1, 1, 10, 100, dan 1000), dan parameter *Gamma* dengan jangkauan nilai 1, 0.01, 0.001, dan 0.0001).

8. Implementasi *Support Vector Machine (SVM)*

Algoritma *Support Vector Machine (SVM)* berfungsi untuk melakukan klasifikasi suatu data berdasarkan data latih dan data uji menggunakan kombinasi nilai parameter kernel, *C*, dan *Gamma* yang paling optimal. Kombinasi nilai diperoleh dengan cara *Hyperparameter Tuning* menggunakan *GridSearchCV*. Algoritma *Support Vector Machine (SVM)* mencari *hyperplane* terbaik yang memisahkan kelas-kelas dengan margin maksimum. *Support Vector Machine (SVM)* melakukan prediksi berdasarkan posisi titik data relatif terhadap *hyperplane* yang ditentukan. Alur penerapan *Support Vector Machine (SVM)* dapat dilihat pada gambar berikut ini :



Gambar 3.9 Alur Penerapan Algoritma *SVM*

Protected by PDF Anti-Copy Free

(Upgrade to Pro Version to Remove the Watermark)

9. Evaluasi Model Support Vector Machine (SVM)

Dalam penelitian ini, evaluasi model dilakukan menggunakan pengujian *library Confusion Matrix* yang terdapat pada *Support Vector Machine (SVM)*. Metode pengujian ini digunakan untuk mengukur performa pada metode klasifikasi. *Confusion matrix* dapat digambarkan sebagai matriks 2x2 yang menampilkan hasil klasifikasi *biner* pada suatu *dataset* dengan empat istilah didalamnya.

Tabel 3.15 Tabel Pengujian *Confusion Matrix*

<i>Classification</i>	<i>Predicted : True</i>	<i>Predicted : False</i>
<i>Actual : True</i>	<i>True Positive</i> (TP)	<i>False Negative</i> (FN)
<i>Actual : False</i>	<i>False Positive</i> (FP)	<i>True Negative</i> (TN)

Keterangan :

- 1) *Actual* merupakan nilai asli dari label/kelas.
- 2) *Predicted* merupakan nilai prediksi hasil pemodelan *Machine Learning*.
- 3) *TP* adalah *True Positive*. Nilai aktual benar sesuai dengan nilai prediksi benar.
- 4) *FP* adalah *False Positive*. Nilai aktual salah sedangkan nilai prediksi benar. (*Error Tipe 1*)
- 5) *FN* adalah *False Negative*. Nilai aktual benar sedangkan nilai prediksi salah. (*Error Tipe 2*)
- 6) *TN* adalah *True Negative*. Nilai aktual salah sesuai dengan nilai prediksi salah.

Hasil dari tabel pengujian *confusion matrix* dapat dimanfaatkan untuk menghitung kinerja klasifikasi dengan menentukan nilai *accuracy*, *precision*, *recall*, dan *F1-Score*, seperti berikut :

Protected by PDF Anti-Copy Free

(Upgrade to Pro Version to Remove the Watermark)

a. Accuracy

Accuracy adalah jumlah rasio prediksi yang benar seperti pada Persamaan 1.

$$Accuracy = \frac{TP + TN}{TP + FP + TN + FN} \times 100\% \quad (1)$$

b. Precision

Precision adalah rasio jumlah antara data bernilai benar dengan data prediksi benar seperti pada Persamaan 2.

$$Precision = \frac{TP}{TP + FP} \times 100\% \quad (2)$$

c. Recall

Recall adalah perbandingan rasio jumlah prediksi benar dengan seluruh rasio jumlah data yang benar seperti pada Persamaan 3.

$$Recall = \frac{TP}{TP + FN} \times 100\% \quad (3)$$

d. F1 Score

F1 Score adalah perhitungan dengan membandingkan antara *precision* dan *recall* seperti Persamaan 4.

$$F1\ Score = \frac{2 * Precision * Recall}{Precision + Recall} \times 100\% \quad (4)$$

HASIL PENELITIAN DAN PEMBAHASAN



4.1 Gambaran Umum Tempat Penelitian

Penelitian ini dilakukan di rumah Peneliti yang beralamat di Jalan Mangga, Dusun 1, Desa Tambahasri, Kecamatan Tugumulyo. Meskipun penelitian ini dilaksanakan di rumah, lingkungan penelitian didukung oleh fasilitas yang memungkinkan Peneliti untuk mengakses data dan menggunakan perangkat yang diperlukan untuk menjalankan penelitian secara efektif. Tempat penelitian ini dipilih karena memberikan kenyamanan dan fokus yang diperlukan untuk mengolah data dan melakukan analisis secara mandiri. Penelitian ini menggunakan data ulasan pengguna aplikasi Sistem Informasi Rekapitulasi (Sirekap) yang diperoleh dari *platform Google Play Store*.

4.1.1 Gambaran Umum Aplikasi Sistem Informasi Rekapitulasi (Sirekap)



Gambar 4.1 Logo Sirekap

Sistem Informasi Rekapitulasi (Sirekap) adalah sebuah aplikasi yang dikembangkan oleh Komisi Pemilihan Umum (KPU) sejak awal tahun 2020. Aplikasi ini dirancang untuk mempermudah dan mempercepat proses rekapitulasi hasil perhitungan suara pada Pemilu. Sirekap berfungsi sebagai alat bantu bagi petugas KPPS (Kelompok Penyelenggara Pemungutan Suara) dalam mengumpulkan dan menghitung suara secara lebih efisien dan

Protected by PDF Anti-Copy Free

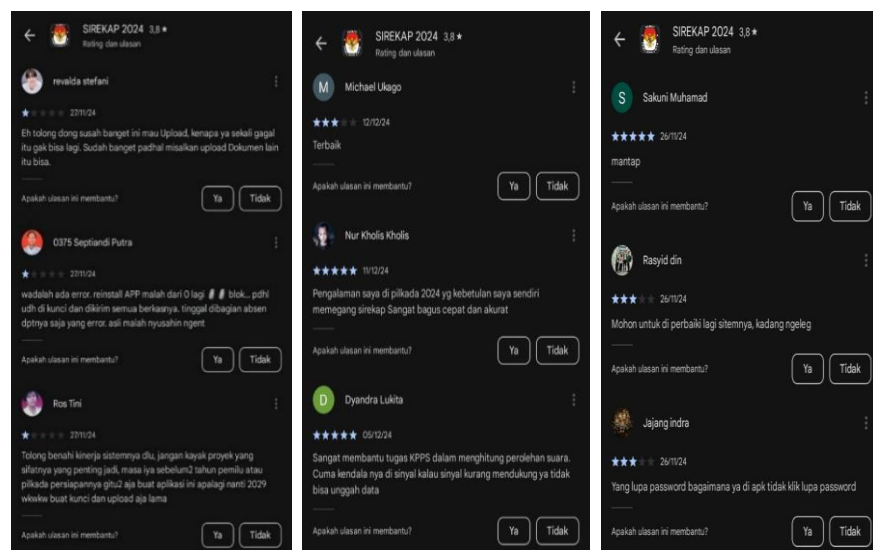
(Upgrade to Pro Version to Remove the Watermark)

transparan. Aplikasi ini hanya dapat digunakan oleh petugas KPPS yang memiliki akses untuk melakukan rekapitulasi hasil suara, sehingga tidak akses oleh masyarakat umum. Dengan demikian,



Aplikasi Sirekap memiliki peranan penting dalam memastikan akurasi dan kelancaran proses penghitungan suara di tingkat TPS (Tempat Pemungutan Suara). Aplikasi Sirekap dapat diunduh melalui *Google Play Store* dengan mengakses link : <https://play.google.com/store/apps/details?id=id.go.kpu.sirekap2024>. Aplikasi ini juga berfungsi untuk mempublikasikan hasil rekapitulasi suara secara langsung dan memastikan pengiriman data suara ke pusat pengolahan informasi secara *real-time*.

Berikut merupakan beberapa ulasan pengguna pada aplikasi Sistem Informasi Rekapitulasi (Sirekap) di *Google Play Store* dapat dilihat pada Gambar 4.2.



Gambar 4.2 Beberapa Ulasan Pengguna Sirekap

Protected by PDF Anti-Copy Free

(Upgrade to Pro Version to Remove the Watermark)

4.2 Hasil

Dalam penelitian implementasi dilakukan menggunakan *Python* dan *Jupyter Lab*. Untuk menjalankan program *Python* di *Google Colab*, diperlukan beberapa *Library* adalah kumpulan kode yang berfungsi untuk menyediakan fungsionalitas tertentu yang bisa digunakan berulang kali dalam berbagai program. Dalam penelitian ini, *library* yang digunakan antara lain :

- a) ***Google-Play-Scraper*** untuk melakukan *web scraping* pada data ulasan di *Google Play Store*.
- b) ***Pandas*** untuk pengolahan dan pembersihan data serta penyimpanan dataset.
- c) ***Numpy*** dan ***Scikit-learn*** untuk pemrosesan data dan implementasi model *machine learning*, termasuk *TF-IDF*, pembagian data, *SMOTEENN*, dan optimasi model menggunakan *GridSearchCV*.
- d) ***NLTK*** dan ***Sastrawi*** untuk pengolahan teks, seperti tokenisasi, penghapusan *stopword*, dan *stemming*.
- e) ***Matplotlib*** dan ***Seaborn*** untuk visualisasi hasil evaluasi model dan perbandingan metrik kinerja.

Penginstalan *library* ini dapat dilakukan dengan perintah “*pip install*” di *Jupyter Lab* . Setelah *library* terinstal, *library* tersebut dapat diimpor ke dalam kode dengan menggunakan perintah “*import*”. Untuk mencapai hasil yang diinginkan dalam penelitian ini, terdapat beberapa tahapan yang harus dilakukan, mulai dari tahap pengumpulan data, tahap persiapan data, tahap transformasi dan penyeimbangan data, tahap pembagian dan optimasi data, serta tahap pelatihan, uji, dan evaluasi model.

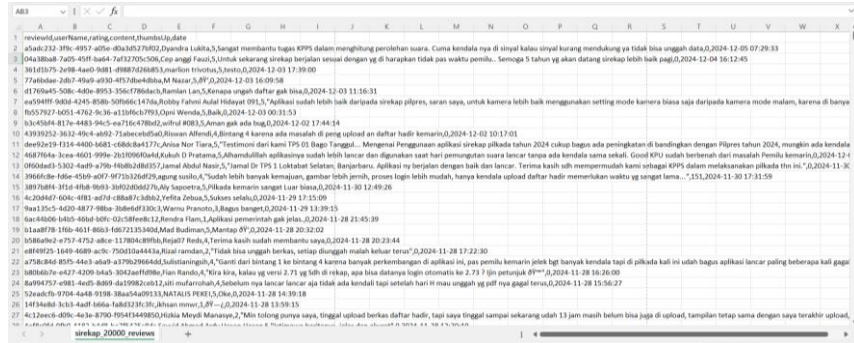
1. Tahap Pengumpulan Data

Pada penelitian ini, Peneliti menggunakan 20.000 data ulasan pengguna. Data tersebut dikumpulkan secara acak (*random*) menggunakan teknik *scraping* dengan *library Google-Play-*

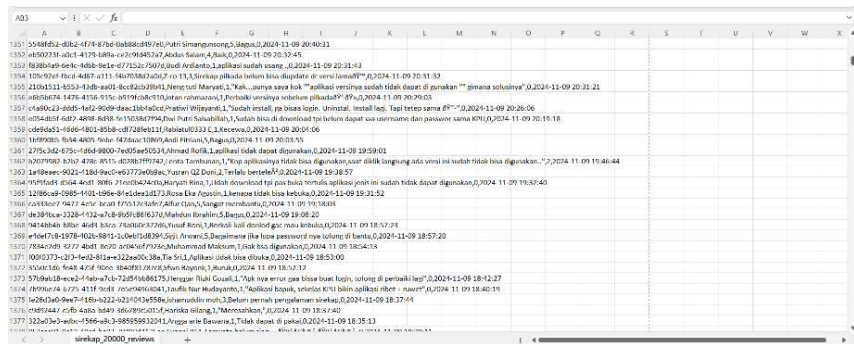
Protected by PDF Anti-Copy Free

(Upgrade to Pro Version to Remove the Watermark)

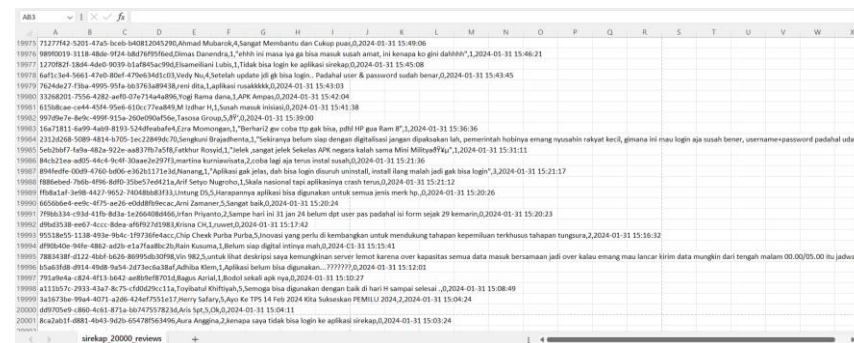
Scraper. Selanjutnya hasil dari pengambilan data akan diunduh dan disimpan kedalam file csv (*comma-separated values*) untuk digunakan pada PDF selanjutnya. Berikut merupakan hasil dari tahap *scraping* dan pengguna Aplikasi Sistem Informasi Rekapitulasi (Sirekap) dapat dilihat pada Gambar 4.3, 4.4, dan 4.5.



Gambar 4.3 Data Hasil Scraping Ulasan Pengguna Sirekap



Gambar 4.4 Data Hasil Scraping Ulasan Pengguna Sirekap



Gambar 4.5 Data Hasil Scraping Ulasan Pengguna Sirekap

Protected by PDF Anti-Copy Free

(Upgrade to Pro Version to Remove the Watermark)

2. Tahap Persiapan Data

a) Tahap *Preprocessing*

Tahap *Preprocessing* merupakan tahap mempersiapkan data agar dapat digunakan pada langkah selanjutnya.

1) *Data Cleaning*

Pada tahap *Data Cleaning* data ulasan dilakukan proses menangani *missing values*, menghapus kolom yang tidak relevan, menghapus data duplikat, serta menghapus spasi, simbol, karakter, dan emotikon. Hasil dari tahap *Data Cleaning* dapat dilihat pada Gambar 4.6.

content	content_cleaning
Sangat membantu tugas KPPS dalam menghitung perolehan suara. Cuma kendala nya di sinyal kalau sinyal kurang mendukung ya tidak bisa unggah data	Sangat membantu tugas KPPS dalam menghitung perolehan suara Cuma kendala nya di sinyal kalau sinyal kurang mendukung ya tidak bisa unggah data
Untuk sekarang sirekap berjalan sesuai dengan yg di harapkan tidak pas waktu pemilu. Semoga 5 tahun yg akan datang sirekap lebih baik pagi	Untuk sekarang sirekap berjalan sesuai dengan yg di harapkan tidak pas waktu pemilu Semoga tahun yg akan datang sirekap lebih baik pagi
testo	testo
Kenapa ungha daftar gak bisa	Kenapa ungha daftar gak bisa

Gambar 4.6 Hasil *Data Cleaning*

2) *Case Folding*

Pada tahap *Case Folding* data ulasan diubah seluruhnya menjadi huruf kecil. Hasil dari tahap *Case Folding* dapat dilihat pada Gambar 4.7.

content_cleaning	content_case_folding
Sangat membantu tugas KPPS dalam menghitung perolehan suara Cuma kendala nya di sinyal kalau sinyal kurang mendukung ya tidak bisa unggah data	sangat membantu tugas kpps dalam menghitung perolehan suara cuma kendala nya di sinyal kalau sinyal kurang mendukung ya tidak bisa unggah data
Untuk sekarang sirekap berjalan sesuai dengan yg di harapkan tidak pas waktu pemilu Semoga tahun yg akan datang sirekap lebih baik pagi	untuk sekarang sirekap berjalan sesuai dengan yg di harapkan tidak pas waktu pemilu semoga tahun yg akan datang sirekap lebih baik pagi
testo	testo
NaN	NaN
Kenapa ungha daftar gak bisa	kenapa ungha daftar gak bisa

Gambar 4.7 Hasil *Case Folding*

Protected by PDF Anti-Copy Free

(Upgrade to Pro Version to Remove the Watermark)

3) Normalization

Pada tahap *Normalization* dilakukan perbaikan pada kata yang mengandung singkatan dan kata tidak baku menjadi kata yang sebenarnya dan bermakna sesuai dengan Kamus Besar Bahasa Indonesia (KBBI) menggunakan “kamuskatabaku.csv” yang diperoleh dari *Kaggle*. Hasil dari tahap *Normalization* dapat dilihat pada Gambar 4.8.

content_case_folding	content_normalization
sangat membantu tugas kpps dalam menghitung perolehan suara cuma kendala nya di sinyal kalau sinyal kurang mendukung ya tidak bisa unggah data	sangat membantu tugas kpps dalam menghitung perolehan suara cuma kendala ya di sinyal kalau sinyal kurang mendukung ya tidak bisa unggah data
untuk sekarang sirekap berjalan sesuai dengan yg di harapkan tidak pas waktu pemilu semoga tahun yg akan datang sirekap lebih baik pagi	untuk sekarang sirekap berjalan sesuai dengan yang di harapkan tidak pas waktu pemilu semoga tahun yang akan datang sirekap lebih baik pagi
testo	testo
NaN	
kenapa ungh daftar gak bisa	kenapa ungh daftar tidak bisa

Gambar 4.8 Hasil *Normalization*

4) Tokenizing

Pada tahap *Tokenizing* dilakukan tokenisasi atau memisahkan kalimat menjadi satuan kata. Hasil dari tahap *Tokenizing* dapat dilihat pada Gambar 4.9.

content_normalization	content_tokenizing
sangat membantu tugas kpps dalam menghitung perolehan suara cuma kendala ya di sinyal kalau sinyal kurang mendukung ya tidak bisa unggah data	[sangat, membantu, tugas, kpps, dalam, menghitung, perolehan, suara, cuma, kendala, ya, di, sinyal, kalau, sinyal, kurang, mendukung, ya, tidak, bisa, unggah, data]
untuk sekarang sirekap berjalan sesuai dengan yang di harapkan tidak pas waktu pemilu semoga tahun yang akan datang sirekap lebih baik pagi	[untuk, sekarang, sirekap, berjalan, sesuai, dengan, yang, di, harapkan, tidak, pas, waktu, pemilu, semoga, tahun, yang, akan, datang, sirekap, lebih, baik, pagi]

Gambar 4.9 Hasil *Tokenizing*

Protected by PDF Anti-Copy Free

(Upgrade to Pro Version to Remove the Watermark)

5) *Stopword Removal*

Pada tahap *Stopword Removal* dilakukan proses menghilangkan kata-kata yang sering muncul tetapi tidak memiliki bobot penting dalam proses klasifikasi. Hasil tahap *Stopword Removal* dapat dilihat pada Gambar 4.1

content_tokenizing	content_stopword_removal
['sangat', 'membantu', 'tugas', 'kpps', 'dalam', 'menghitung', 'perolehan', 'suara', 'cuma', 'kendala', 'ya', 'di', 'sinyal', 'kalau', 'sinyal', 'kurang', 'mendukung', 'ya', 'tidak', 'bisa', 'unggah', 'data']	['sangat', 'membantu', 'tugas', 'kpps', 'dalam', 'menghitung', 'perolehan', 'suara', 'cuma', 'kendala', 'ya', 'di', 'sinyal', 'kalau', 'sinyal', 'kurang', 'mendukung', 'ya', 'tidak', 'bisa', 'unggah', 'data']
['untuk', 'sekarang', 'sirekap', 'berjalan', 'sesuai', 'dengan', 'yang', 'di', 'harapkan', 'tidak', 'pas', 'waktu', 'pemilu', 'semoga', 'tahun', 'yang', 'akan', 'datang', 'sirekap', 'lebih', 'baik', 'pagi']	['untuk', 'sekarang', 'sirekap', 'berjalan', 'sesuai', 'dengan', 'yang', 'di', 'harapkan', 'tidak', 'pas', 'waktu', 'pemilu', 'semoga', 'tahun', 'yang', 'akan', 'datang', 'sirekap', 'lebih', 'baik', 'pagi']

Gambar 4.10 Hasil *Stopword Removal*

6) *Stemming*

Pada tahap *Stemming* proses mengubah kata-kata dalam data ulasan menjadi bentuk kata dasarnya. Hasil tahap *Stemming* dapat dilihat pada Gambar 4.11

Protected by PDF Anti-Copy Free

(Upgrade to Pro Version to Remove the Watermark)

content_stopword_removal	content_stemming
 <p>['smbantu', 'tugas', 'k', 'menghitung', 'suara', 'cuma', 'kendala', 'ya', 'di', 'sinyal', 'kalau', 'sinyal', 'kurang', 'mendukung', 'ya', 'tidak', 'bisa', 'unggah', 'data']</p>	<p>sangat bantu tugas kpps dalam hitung oleh suara cuma kendala ya di sinyal kalau sinyal kurang dukung ya tidak bisa unggah data</p>
<p>['untuk', 'sekarang', 'sirekap', 'berjalan', 'sesuai', 'dengan', 'yang', 'di', 'harapkan', 'tidak', 'pas', 'waktu', 'pemilu', 'semoga', 'tahun', 'yang', 'akan', 'datang', 'sirekap', 'lebih', 'baik', 'pagi']</p>	<p>untuk sekarang sirekap jalan sesuai dengan yang di harap tidak pas waktu milu moga tahun yang akan datang sirekap lebih baik pagi</p>

Gambar 4.11 Hasil *Stemming*

b) Tahap *Labeling*

Data ulasan pengguna yang sudah dilakukan tahapan *preprocessing*, akan dilakukan proses pelabelan data menjadi tiga kelas sentimen, yaitu positif, netral, dan negatif. Hasil tahap *Labeling* dapat dilihat pada Gambar 4.12.

rating	content_stemming	sentiment
5	steady	positif
1	aplikasi ya bapuk	negatif
3	bug buat dokumen lama dan foto halaman capres tidak bisa buka edit sudah ada maju aplikasi saran untuk server bawa ya di besar agar bisa upload cepat	netral

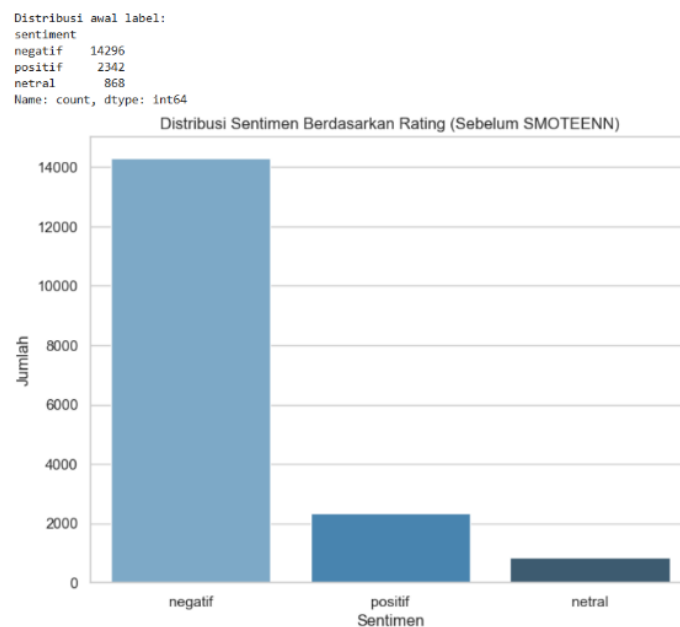
Gambar 4.12 Hasil *Labeling*

Protected by PDF Anti-Copy Free

(Upgrade to Pro Version to Remove the Watermark)

c) Tahap Visualisasi Data

Setelah ulasan melalui tahapan *labeling*, selanjutnya akan dilakukan visualisasi data agar lebih mudah dipahami dan memberikan informasi yang menarik. Pada tahap visualisasi ini digunakan *library Matplotlib* yang akan memvisualisasikan kedalam bentuk diagram untuk melihat banyak nya data dengan label positif, netral dan negatif. Hasil visualisasi dapat dilihat pada Gambar 4.13.



Gambar 4.13 Hasil Visualisasi *BarPlot*

Dari gambar hasil visualisasi data dalam bentuk diagram batang, menunjukkan bahwa ulasan dengan label tertinggi terdapat pada label negatif dengan sentimen berjumlah 14.296. Kemudian pada label positif berjumlah 2.342 dan terakhir ulasan dengan label netral berjumlah 868. Selain dalam bentuk diagram batang, visualisasi ulasan pada aplikasi Sistem Informasi Rekapitulasi (Sirekap) juga disajikan dalam bentuk *WordCloud* yaitu adalah gambar yang menampilkan katakata dari teks, dengan kata-kata yang paling sering muncul lebih

Protected by PDF Anti-Copy Free

(Upgrade to Pro Version to Remove the Watermark)

Hasil TF-IDF untuk 10 data teratas:

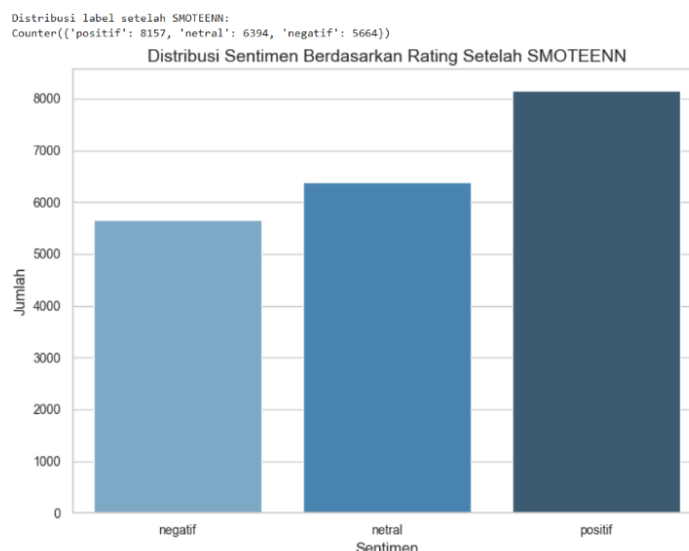
	aplikasi	belum	bisa	dan	di	guna	ini	login	malah	\
0	0.000000	0.321513	0.0	0.356959	0.000000	0.0	0.0	0.0	0.0	
1	0.000000	0.000000	0.0	0.352981	0.000000	0.0	0.0	0.0	0.0	
2	0.000000	0.000000	0.0	0.000000	0.000000	0.0	0.0	0.0	0.0	
3	0.000000	0.2010	0.0	0.000000	0.000000	0.0	0.0	0.0	0.0	
4	0.220886	0.000000	0.0	0.283075	0.728224	0.0	0.0	0.0	0.0	

	saja	sudah	tidak	untuk	ya	yang
0	0.000000	0.000000	0.281812	0.000000	0.830537	0.000000
1	0.000000	0.000000	0.278672	0.443278	0.000000	0.775404
2	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000
3	0.000000	0.000000	0.659151	0.000000	0.000000	0.000000
4	0.347378	0.306144	0.000000	0.355488	0.000000	0.000000

Gambar 4.15 Hasil TF-IDF

b) Tahap *Over-Under Sampling* dengan *SMOTEENN*

Pada tahap *SMOTEENN* ini dilakukan proses menangani ketidakseimbangan data (*data imbalance*) menggunakan teknik *oversampling* dengan *SMOTE* (*Synthetic Minority Oversampling Technique*) dan teknik *undersampling* dengan *ENN* (*Edited Nearest Neighbors*). Dari total data sebanyak 17.506 data ulasan menjadi 20.215 data ulasan setelah dilakukan proses *SMOTEENN*. Hasil *SMOTEENN* dapat dilihat pada Gambar 4.16.




Gambar 4.16 Hasil Visualisasi Tahap *SMOTEENN*

Protected by PDF Anti-Copy Free

(Upgrade to Pro Version to Remove the Watermark)

4. Tahap Pembagian dan Optimasi Data

a) Tahap Split

Pada  dilakukan pembagian data menjadi dua bagian, yaitu data latih dan data uji. Data latih digunakan untuk membuat model klasifikasi, sedangkan data uji digunakan untuk mengevaluasi model tersebut. Penelitian ini menggunakan lima rasio perbandingan, yaitu 50:50, 60:40, 70:30, 80:20, dan 90:10. Hasil tahap split data dapat dilihat pada Tabel 4.1.

Tabel 4.1 Tabel Split Data

Rasio Pembagian		Jumlah Data Latih	Jumlah Data Uji
Data Latih	Data Uji		
50%	50%	10.107	10.108
60%	40%	12.129	8.086
70%	30%	14.150	6.065
80%	20%	16.172	4.043
90%	10%	18.193	2.022

b) Tahap *Hyperparameter Tuning* dengan *GridSearchCV*

Tahap ini dilakukan untuk menentukan kombinasi parameter yang paling optimal untuk model klasifikasi. Parameter yang di *tuning* melibatkan empat kernel, yaitu *Linear*, *Polynomial*, *RBF*, dan *Sigmoid*. Sedangkan, nilai parameter C (*complexity*) = 0.1, 1, 10, 100, 1000 dan $\text{Gamma } (\gamma)$ = 1, 0.1, 0.01, 0.001, 0.0001. Berikut *script* dan proses *Hyperparameter Tuning* dengan *GridSearchCV* dapat dilihat pada Gambar 4.17 dan 4.18.

Protected by PDF Anti-Copy Free (Upgrade to Pro Version to Remove the Watermark)

```

: from sklearn.svm import SVC
  from sklearn.model_selection import GridSearchCV

# Definisi param_grid
param_grid = {'C': [0.1, 1, 10, 100, 1000],
              'gamma': [0.1, 0.01, 0.001, 0.0001],
              'kernel': ['rbf', 'linear', 'poly', 'sigmoid']}

# Inisialisasi model SVM
svc = SVC()

# Inisialisasi GridSearchCV
grid = GridSearchCV(SVC(), param_grid, refit = True, verbose = 3, cv=3)

# fitting the model for grid search
grid.fit(X_train, y_train)

```

Gambar 4.17 Script *GridSearchCV*

```

Fitting 3 folds for each of 100 candidates, totalling 300 fits
[CV 1/3] END .....C=0.1, gamma=1, kernel=rbf;, score=0.740 total time= 6.7s
[CV 2/3] END .....C=0.1, gamma=1, kernel=rbf;, score=0.724 total time= 5.9s
[CV 3/3] END .....C=0.1, gamma=1, kernel=rbf;, score=0.756 total time= 6.1s
[CV 1/3] END ....C=0.1, gamma=1, kernel=linear;, score=0.705 total time= 2.5s
[CV 2/3] END ....C=0.1, gamma=1, kernel=linear;, score=0.678 total time= 2.4s
[CV 3/3] END ....C=0.1, gamma=1, kernel=linear;, score=0.718 total time= 2.6s
[CV 1/3] END .....C=0.1, gamma=1, kernel=poly;, score=0.746 total time= 4.2s
[CV 2/3] END .....C=0.1, gamma=1, kernel=poly;, score=0.725 total time= 4.6s
[CV 3/3] END .....C=0.1, gamma=1, kernel=poly;, score=0.760 total time= 3.5s
[CV 1/3] END ....C=0.1, gamma=1, kernel=sigmoid;, score=0.684 total time= 4.7s
[CV 2/3] END ....C=0.1, gamma=1, kernel=sigmoid;, score=0.656 total time= 4.2s
[CV 3/3] END ....C=0.1, gamma=1, kernel=sigmoid;, score=0.695 total time= 3.9s
[CV 1/3] END .....C=0.1, gamma=0.1, kernel=rbf;, score=0.706 total time= 7.6s
[CV 2/3] END .....C=0.1, gamma=0.1, kernel=rbf;, score=0.682 total time= 7.5s
[CV 3/3] END .....C=0.1, gamma=0.1, kernel=rbf;, score=0.717 total time= 7.7s

```

Gambar 4.18 Proses *GridSearchCV*

```

[CV 1/3] END C=1000, gamma=0.001, kernel=sigmoid;, score=0.706 total time= 3.9s
[CV 2/3] END C=1000, gamma=0.001, kernel=sigmoid;, score=0.677 total time= 3.6s
[CV 3/3] END C=1000, gamma=0.001, kernel=sigmoid;, score=0.715 total time= 3.8s
[CV 1/3] END ..C=1000, gamma=0.0001, kernel=rbf;, score=0.706 total time= 6.4s
[CV 2/3] END ..C=1000, gamma=0.0001, kernel=rbf;, score=0.677 total time= 6.4s
[CV 3/3] END ..C=1000, gamma=0.0001, kernel=rbf;, score=0.716 total time= 6.6s
[CV 1/3] END C=1000, gamma=0.0001, kernel=linear;, score=0.707 total time= 3.4min
[CV 2/3] END C=1000, gamma=0.0001, kernel=linear;, score=0.680 total time= 3.0min
[CV 3/3] END C=1000, gamma=0.0001, kernel=linear;, score=0.717 total time= 3.5min
[CV 1/3] END .C=1000, gamma=0.0001, kernel=poly;, score=0.415 total time= 4.5s
[CV 2/3] END .C=1000, gamma=0.0001, kernel=poly;, score=0.415 total time= 4.6s
[CV 3/3] END .C=1000, gamma=0.0001, kernel=poly;, score=0.415 total time= 4.4s
[CV 1/3] END C=1000, gamma=0.0001, kernel=sigmoid;, score=0.705 total time= 3.8s
[CV 2/3] END C=1000, gamma=0.0001, kernel=sigmoid;, score=0.678 total time= 3.5s
[CV 3/3] END C=1000, gamma=0.0001, kernel=sigmoid;, score=0.718 total time= 3.7s

```

```

20]: > GridSearchCV ⓘ ⓘ
      > best_estimator_: SVC
          > SVC ⓘ

```

Gambar 4.19 Proses *GridSearchCV*

Protected by PDF Anti-Copy Free

(Upgrade to Pro Version to Remove the Watermark)

Hyperparameter Tuning ini dilakukan masing-masing pada lima rasio pembagian data, yaitu 50:50, 60:40, 70:30, 80:20, dan 90:10. Berikut hasil dari *Hyperparameter Tuning* dengan *GridSearchCV* dapat dilihat pada Tabel 4.2.

Tabel 4.2 Hasil *GridSearchCV*

Rasio	Kombinasi Parameter Terbaik		
	Kernel	C	Gamma
50 : 50	RBF	1000	1
60 : 40	RBF	1000	1
70 : 30	RBF	100	1
80 : 20	RBF	100	1
90 : 10	RBF	1000	1

5. Tahap Pelatihan, Pengujian, dan Evaluasi Model

a) Tahap Pelatihan Model

Pada tahap ini, dilakukan pelatihan (*training*) terhadap model *Support Vector Machine (SVM)*. Model *Support Vector Machine (SVM)* dilatih menggunakan data latih dengan parameter terbaik hasil dari *tuning* parameter menggunakan *GridSearchCV*. Berikut hasil dari *training* model dapat dilihat pada Gambar 4.20.

```
best_params = grid.best_params_
best_svm = SVC(**best_params, random_state=42)

best_svm.fit(X_train, y_train)

print("Model telah dilatih dengan parameter terbaik.")
```

Model telah dilatih dengan parameter terbaik.

Gambar 4.20 Hasil Pelatihan Model *SVM*

Protected by PDF Anti-Copy Free

(Upgrade to Pro Version to Remove the Watermark)

b) Tahap Pengujian Model

Pada bagian ini, dilakukan pengujian (*testing*) terhadap model *Support Vector Machine (SVM)* menggunakan data uji masing-masing rasio perbandingan data. Model *Support Vector Machine (SVM)* di uji menggunakan parameter terbaik hasil dari *tuning* parameter menggunakan *GridSearchCV*. Hasil pengujian model divisualisasikan dalam bentuk tabel *Classification Report (CR)* yang memuat metrik evaluasi utama, yaitu *accuracy*, *precision*, *recall*, dan *f1-score* untuk masing-masing kelas (Positif, Netral, dan Negatif).

Berikut adalah *script* yang digunakan untuk pengujian model, sebagaimana ditampilkan pada Gambar 4.21.

```
from sklearn.metrics import classification_report, confusion_matrix

y_pred = best_svm.predict(X_test)

print("\nClassification Report:")
print(classification_report(y_test, y_pred))
```

Gambar 4.21 Script Pengujian Model *SVM*

Hasil pengujian untuk masing-masing rasio pembagian data (50:50, 60:40, 70:30, 80:20, dan 90:10) menunjukkan performa model *Support Vector Machine (SVM)* dalam mengklasifikasikan setiap kelas berdasarkan distribusi data yang berbeda. Berikut adalah hasil pengujian untuk setiap rasio yang dapat dilihat pada Gambar 4.22, 4.23, 4.24, 4.25, dan 4.26.

```
Classification Report:
              precision    recall  f1-score   support

negatif      0.94      0.95      0.95      2902
netral       0.97      0.97      0.97      3246
positif      0.97      0.97      0.97      3960

accuracy                0.96      10108
macro avg      0.96      0.96      0.96      10108
weighted avg   0.96      0.96      0.96      10108
```

Gambar 4.22 Hasil Pengujian Model (Rasio 50:50)

Protected by PDF Anti-Copy Free

(Upgrade to Pro Version to Remove the Watermark)

Pada rasio 50:50 akurasi model *Support Vector Machine* (SVM) mencapai 96%, dengan performa yang sangat baik pada setiap kelas. *Precision*, *recall*, dan *f1-score* untuk semua kelas berkisar antara 0.94 hingga 0.97.

```
Classification Report:
              precision    recall  f1-score   support

negatif      0.94         0.95         0.95         2317
netral       0.97         0.97         0.97         2588
positif      0.97         0.97         0.97         3181

accuracy                0.96         8086
macro avg              0.96         0.96         0.96         8086
weighted avg           0.96         0.96         0.96         8086
```

Gambar 4.23 Hasil Pengujian Model (Rasio 60:40)

Akurasi tetap 96% pada rasio 60:40, dengan kelas Positif dan Netral menunjukkan performa yang sedikit lebih baik dibandingkan kelas Negatif. Nilai *precision*, *recall*, dan *f1-score* berada pada kisaran 0.94 hingga 0.97.

```
Classification Report:
              precision    recall  f1-score   support

negatif      0.95         0.96         0.96         1711
netral       0.97         0.97         0.97         1978
positif      0.98         0.96         0.97         2376

accuracy                0.97         6065
macro avg              0.96         0.97         0.97         6065
weighted avg           0.97         0.97         0.97         6065
```

Gambar 4.24 Hasil Pengujian Model (Rasio 70:30)

Pada rasio 70:30, model *Support Vector Machine* (SVM) mencapai akurasi 97%, dengan performa yang sangat baik pada semua kelas. *Precision*, *recall*, dan *f1-score* berkisar antara 0.95 hingga 0.98, menunjukkan keseimbangan yang baik di seluruh kelas.

Protected by PDF Anti-Copy Free

(Upgrade to Pro Version to Remove the Watermark)

Classification Report:

	precision	recall	f1-score	support
n	0.96	0.97	0.97	1160
p	0.97	0.98	0.98	1278
p	0.98	0.97	0.98	1605
accuracy			0.97	4043
macro avg	0.97	0.97	0.97	4043
weighted avg	0.97	0.97	0.97	4043

Gambar 4.25 Hasil Pengujian Model (Rasio 80:20)

Akurasi tetap 97% pada rasio 80:20, dengan performa lebih baik pada kelas Positif dan Netral. Nilai *precision*, *recall*, dan *f1-score* untuk ketiga kelas berkisar antara 0.96 hingga 0.98.

Classification Report:

	precision	recall	f1-score	support
negatif	0.97	0.97	0.97	591
netral	0.98	0.98	0.98	613
positif	0.99	0.99	0.99	818
accuracy			0.98	2022
macro avg	0.98	0.98	0.98	2022
weighted avg	0.98	0.98	0.98	2022

Gambar 4.26 Hasil Pengujian Model (Rasio 90:10)

Pada rasio 90:10, model *Support Vector Machine (SVM)* mencapai akurasi tertinggi sebesar 98%, dengan nilai *precision*, *recall*, dan *f1-score* yang sangat tinggi, terutama pada kelas Positif, dengan nilai 0.99.

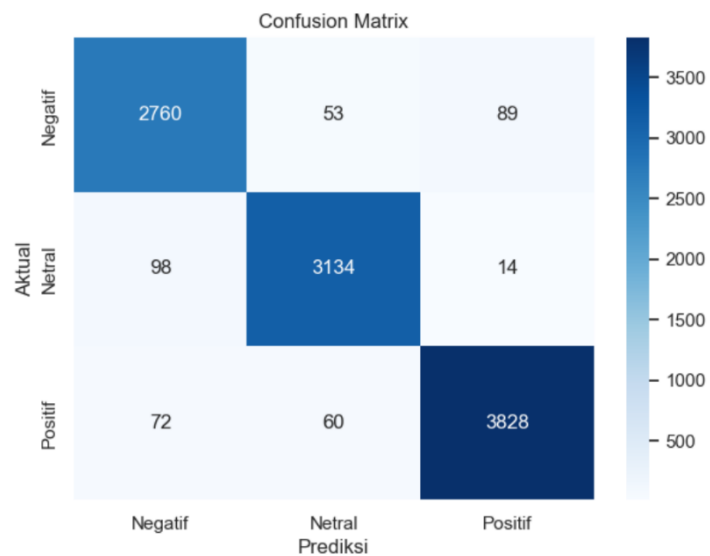
c) Tahap Evaluasi Model

Setelah dilakukan pengujian (*testing*) model, tahap selanjutnya yaitu evaluasi terhadap performa model *Support Vector Machine (SVM)* menggunakan *Confusion Matrix*. *Confusion Matrix* memberikan gambaran lebih mendetail mengenai distribusi prediksi model terhadap data aktual, termasuk bagaimana model berhasil atau gagal

Protected by PDF Anti-Copy Free

(Upgrade to Pro Version to Remove the Watermark)

mengklasifikasikan data ke dalam kategori yang benar. Berikut hasil evaluasi model pada masing-masing lima rasio perbandingan menggunakan *Confusion Matrix* dapat dilihat pada Gambar 4.27, 4.28, 4.29, 4.30, dan 4.31.

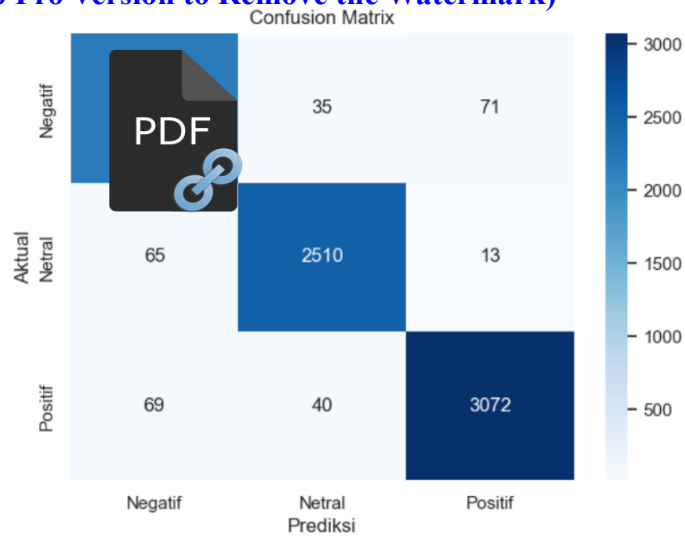


Gambar 4.27 Hasil Evaluasi Model (Rasio 50:50)

Confusion Matrix untuk rasio 50:50 menunjukkan bahwa model mampu mengklasifikasikan data dengan tingkat akurasi yang baik. Sebagian besar data berada pada diagonal utama, menunjukkan klasifikasi yang benar untuk setiap kelas (Positif, Netral, dan Negatif). Kesalahan klasifikasi terbesar terjadi pada kelas Negatif, dimana sebanyak 53 data salah diprediksi sebagai Netral dan 89 data salah diprediksi sebagai Positif.

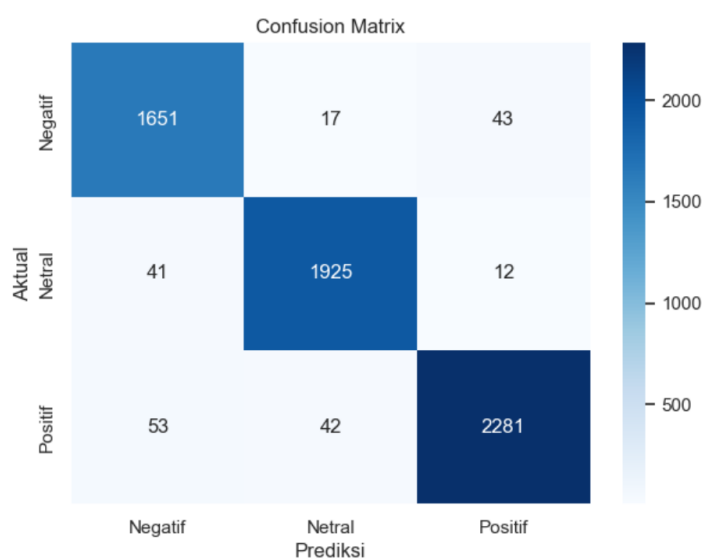
Protected by PDF Anti-Copy Free

(Upgrade to Pro Version to Remove the Watermark)



Gambar 4.28 Hasil Evaluasi Model Rasio 60:40)

Confusion Matrix untuk rasio 60:40 menunjukkan distribusi klasifikasi yang konsisten dengan rasio sebelumnya. Sebagian besar data tetap berada pada diagonal utama, mencerminkan klasifikasi yang akurat. Namun, kesalahan klasifikasi pada kelas Positif sedikit lebih besar, yakni sebanyak 69 data diprediksi salah sebagai Negatif dan 40 diprediksi salah sebagai Netral.

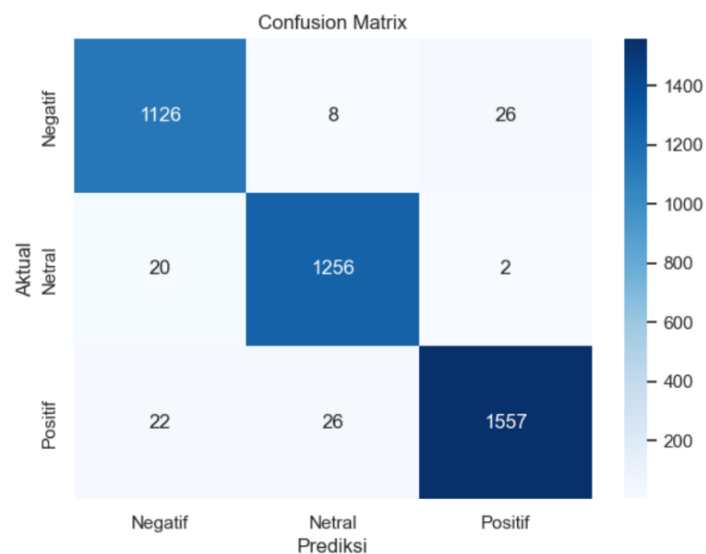


Gambar 4.29 Hasil Evaluasi Model (Rasio 70:30)

Protected by PDF Anti-Copy Free

(Upgrade to Pro Version to Remove the Watermark)

Confusion Matrix untuk rasio 70:30 menunjukkan distribusi prediksi yang baik, dengan berkurangnya kesalahan klasifikasi kesalahan. Namun, kesalahan klasifikasi terbesar masih terdapat pada kelas Positif, yakni sebanyak 53 data diprediksi salah sebagai Negatif dan 42 diprediksi salah sebagai Netral.

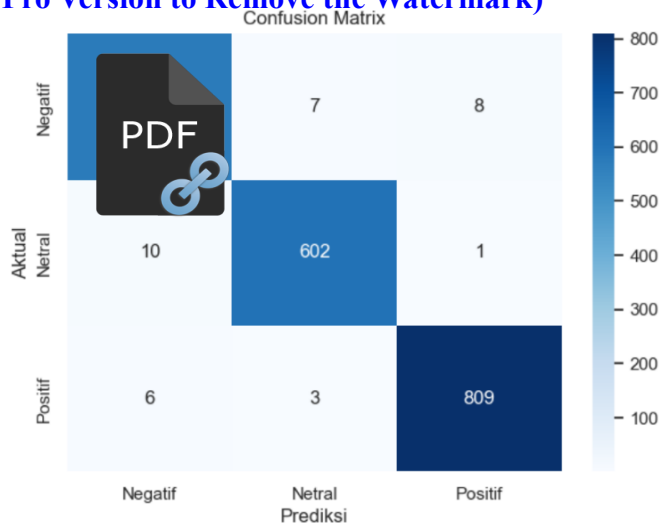


Gambar 4.30 Hasil Evaluasi Model (Rasio 80:20)

Confusion Matrix untuk rasio 80:20 menunjukkan perbaikan performa model. Distribusi klasifikasi semakin mendekati diagonal utama. Jumlah data yang salah diklasifikasikan tiap kelas menurun. Seperti pada rasio sebelumnya, kesalahan klasifikasi terbesar masih terdapat pada kelas Positif, yakni sebanyak 22 data diprediksi salah sebagai Negatif dan 26 data diprediksi salah sebagai Netral.

Protected by PDF Anti-Copy Free

(Upgrade to Pro Version to Remove the Watermark)



Gambar 4.31 Hasil Evaluasi Model (Rasio 90:10)

Confusion Matrix untuk rasio 90:10 menunjukkan performa terbaik dengan kesalahan klasifikasi yang sangat minim untuk di setiap kelas. Kesalahan klasifikasi terbesar terjadi pada kelas Negatif, dengan hanya 15 data yang salah diprediksi yakni 7 data diprediksi sebagai Netral dan 8 data diprediksi sebagai Positif. Hasil ini mencerminkan kemampuan model *Support Vector Machine (SVM)* yang optimal dalam menangkap pola dari data pelatihan untuk menghasilkan klasifikasi yang akurat.

4.3 Pembahasan

4.3.1 Penerapan Metode Analisa dan Validitas Data

Penelitian ini dilakukan untuk menganalisis sentimen ulasan pengguna terhadap Aplikasi Sistem Informasi Rekapitulasi (Sirekap) di *Google Play Store* dan mengembangkan metode analisis pada penelitian sebelumnya. Metode yang digunakan adalah optimasi *Support Vector Machine (SVM)* dengan *Grid Search Cross-Validation*, sebuah teknik pembelajaran mesin, untuk

Protected by PDF Anti-Copy Free

(Upgrade to Pro Version to Remove the Watermark)

mengklasifikasikan ulasan ke dalam tiga kategori sentimen, yaitu Positif, Negatif, dan Netral.

1. Tahap **Preprocessing** dan *Preprocessing* Data

Tahap *Preprocessing* data dalam penelitian ini adalah melakukan pengambilan data ulasan pengguna di *Google Play Store* dengan cara *Scraping* data ulasan menggunakan *library Google-Play-Scraper* di *Python* melalui *tools Jupyter Lab*. *Library* tersebut diinstall melalui *Jupyter Lab* dengan memasukkan perintah “*pip install google-play-scraper*”. Setelah itu dapat dilakukan proses pengambilan data ulasan pengguna pada aplikasi Sistem Informasi Rekapitulasi (Sirekap) yang terdiri dari beberapa fitur, yaitu *reviewId*, *username*, *rating*, *content*, *thumbsUp*, dan *date*.

Data ulasan diambil secara acak (*random*) sebanyak 20.000 ulasan. Setelah data ulasan pengguna dikumpulkan, dilakukan tahap *Preprocessing* data atau mempersiapkan data agar dapat digunakan pada tahap selanjutnya. Adapun tahap-tahap *Preprocessing*, yaitu melakukan *Data Cleaning*, *Case Folding*, *Normalization*, *Tokenizing*, *Stopword Removal*, dan *Stemming*. Setelah dilakukan *Preprocessing*, jumlah data berkurang menjadi 17.506 ulasan dari total data sebanyak 20.000 ulasan. Kemudian data disimpan dalam format file *CSV (Comma-Separated Values)* yang siap digunakan pada tahap selanjutnya.

2. Tahap *Labeling* dan Visualisasi Data

Tahap *Labeling* dilakukan untuk mengklasifikasikan ulasan pengguna menjadi tiga kategori sentimen, yaitu Positif, Negatif, dan Netral. Ulasan dengan rating 1 atau 2 dikategorikan sebagai ulasan Negatif, rating 3 sebagai ulasan Netral, dan rating 4 atau 5 sebagai ulasan Positif. Dari 17.506

Protected by PDF Anti-Copy Free

(Upgrade to Pro Version to Remove the Watermark)

data ulasan yang telah diberi label, terdapat 2.342 ulasan positif, 14.000 ulasan negatif, dan 868 ulasan netral. Hal ini menunjukkan adanya ketidakseimbangan distribusi data, dengan jumlah ulasan negatif jauh lebih dominan dibandingkan dengan ulasan positif dan netral.

Selanjutnya, data tersebut divisualisasikan menggunakan *library Matplotlib* untuk mempermudah interpretasi dan meningkatkan daya tarik penyajian. Visualisasi dilakukan dalam bentuk diagram batang untuk menampilkan distribusi data, serta *wordcloud* untuk menggambarkan kumpulan kata yang paling sering muncul dalam ulasan. Secara keseluruhan, ulasan lebih banyak menggarisbawahi kebutuhan untuk memperbaiki performa aplikasi, khususnya dalam aspek aksesibilitas dan stabilitas sistem.

3. Tahap Pembobotan Kata dan Penanganan Ketidakseimbangan Data

Setelah tahap pelabelan data, dilakukan pembobotan *TF-IDF (Term Frequency – Inverse Document Frequency)* untuk memberikan bobot pada setiap kata yang ada pada dokumen. Sebagai contoh, pada baris pertama, kata “**aplikasi**” memiliki nilai *TF-IDF* 0.0, yang menunjukkan bahwa kata ini tidak muncul atau tidak relevan dalam dokumen tersebut. Tahap selanjutnya adalah penerapan metode *SMOTEENN (Synthetic Minority Over-sampling Technique Edited Nearest Neighbors)* untuk menangani ketidakseimbangan data (*data imbalance*).

Cara kerja *SMOTEENN* yaitu melakukan *oversampling* pada kelas minoritas dengan *Synthetic Minority Oversampling Technique (SMOTE)* untuk menambah jumlah data pada kelas

Protected by PDF Anti-Copy Free

(Upgrade to Pro Version to Remove the Watermark)

tersebut dan *undersampling* pada kelas mayoritas menggunakan *Edited Nearest Neighbors (ENN)* untuk mengurangi data pada kelas mayoritas. *Oversampling* dilakukan kelas Positif dan Netral, sedangkan *undersampling* dilakukan terhadap kelas Negatif. Proses ini menghasilkan dataset dengan total 20.215 ulasan, terdiri dari 8.157 ulasan positif, 5.664 ulasan negatif, dan 6.394 ulasan netral, dengan distribusi data yang lebih seimbang.

4. Tahap Pembagian Data Latih dan Data Uji

Pada tahap ini dilakukan pembagian data menjadi data latih dan data uji menggunakan lima rasio berbeda, yaitu : rasio 50:50, terdapat 10.107 data latih dan 10.108 data uji; rasio 60:40, terdapat 12.129 data latih dan 8.086 data uji; rasio 70:30, terdapat 14.150 data latih dan 6.065 data uji; rasio 80:20, terdapat 16.172 data latih dan 4.043 data uji; dan rasio 90:10, terdapat 18.193 data latih dan 2.022 data uji. Pembagian data ini dilakukan untuk memastikan model diuji pada berbagai proporsi data, yang dapat memberikan gambaran yang lebih komprehensif mengenai kinerja model dalam kondisi yang berbeda.

5. Tahap Optimasi dengan *GridSearchCV* dan Pelatihan Model

Tahap selanjutnya, yaitu proses optimasi data menggunakan *Hyperparameter Tuning* dengan *GridSearchCV*. Proses ini dilakukan untuk menentukan kombinasi parameter yang paling optimal untuk model *Support Vector Machine (SVM)*. Adapun parameter yang di-*tuning*, yaitu kernel (*RBF, Linear, Polynomial, Sigmoid*), *C* (0.1, 1, 10, 100, 1000), dan *Gamma* (1, 0.1, 0.01, 0.001, 0.0001). Berdasarkan nilai parameter tersebut, dihasilkan 100 kombinasi kandidat yang

Protected by PDF Anti-Copy Free

(Upgrade to Pro Version to Remove the Watermark)

akan di-*tuning* dengan *cross-validation* ($cv=3$), menghasilkan total 300 *fit*

Hasil proses *GridSearchCV* adalah kombinasi parameter terbaik untuk semua rasio pembagian data menggunakan kernel *RBF* dengan nilai $\Gamma = 1$. Sedangkan nilai C bervariasi antara 100 dan 1000, dengan rasio 70:30 dan 80:20 menggunakan $C = 100$, sementara rasio 50:50, 60:40, dan 90:10 menggunakan $C = 1000$. Hal ini menunjukkan bahwa pada rasio pembagian yang lebih seimbang dan rasio dengan lebih sedikit data uji (seperti 90:10), model memerlukan nilai C yang lebih besar (1000) untuk menangani lebih baik penalti kesalahan. Sementara pada rasio yang lebih condong ke data latih (70:30 dan 80:20), nilai C yang lebih kecil (100) bekerja lebih baik untuk menghindari *overfitting* dan menghasilkan model yang lebih generalisasi.

Selanjutnya, dilakukan pelatihan (*training*) model *Support Vector Machine* (*SVM*) menggunakan data latih dan parameter terbaik hasil dari proses *GridSearchCV*.

4.3.2 Pengujian Hasil Analisa

Tahap pengujian dilakukan untuk mengevaluasi performa model *Support Vector Machine* (*SVM*) yang telah dioptimasi menggunakan *Grid Search Cross-Validation*. Pengujian ini bertujuan untuk memastikan bahwa model mampu mengklasifikasikan sentimen ulasan pengguna aplikasi Sirekap dengan baik berdasarkan tiga kategori sentimen, yaitu Positif, Negatif, dan Netral.

1. Pengujian Model

Setelah pelatihan model, tahap selanjutnya yaitu dilakukan pengujian (*testing*) model *Support Vector Machine* (*SVM*). Pengujian model divisualisasikan dalam bentuk tabel

Protected by PDF Anti-Copy Free

(Upgrade to Pro Version to Remove the Watermark)

Classification Report (CR) yang memuat metrik evaluasi utama, yaitu *accuracy*, *precision*, *recall*, dan *f1-score*. Berikut adalah hasilujian model berdasarkan *Classification Report (CR)* setiap rasio pembagian data.

- 1) Rasio 50:50
 - a. Positif : *precision* 0.97, *recall* 0.97, dan *f1-score* 0.97.
 - b. Negatif : *precision* 0.94, *recall* 0.95, dan *f1-score* 0.95.
 - c. Netral : *precision* 0.97, *recall* 0.97, dan *f1-score* 0.97.
 - d. Akurasi Keseluruhan : 96%.
- 2) Rasio 60:40
 - a. Positif : *precision* 0.97, *recall* 0.97, dan *f1-score* 0.97.
 - b. Negatif : *precision* 0.94, *recall* 0.95, dan *f1-score* 0.95.
 - c. Netral : *precision* 0.97, *recall* 0.97, dan *f1-score* 0.97.
 - d. Akurasi Keseluruhan : 96%.
- 3) Rasio 70:30
 - a. Positif : *precision* 0.98, *recall* 0.96, dan *f1-score* 0.97.
 - b. Negatif : *precision* 0.95, *recall* 0.96, dan *f1-score* 0.96.
 - c. Netral : *precision* 0.97, *recall* 0.97, dan *f1-score* 0.97.
 - d. Akurasi Keseluruhan : 97%.
- 4) Rasio 80:20
 - a. Positif : *precision* 0.98, *recall* 0.97, dan *f1-score* 0.98.

Protected by PDF Anti-Copy Free

(Upgrade to Pro Version to Remove the Watermark)

b. Negatif : *precision* 0.96, *recall* 0.97, dan *f1-score* 0.97.

c. Netral : *precision* 0.97, *recall* 0.98, dan *f1-score*

d. Akurasi Keseluruhan : 97%.

5) Rasio 90:10

a. Positif : *precision* 0.99, *recall* 0.99, dan *f1-score* 0.99.

b. Negatif : *precision* 0.97, *recall* 0.97, dan *f1-score* 0.97.

c. Netral : *precision* 0.98, *recall* 0.98, dan *f1-score* 0.98.

d. Akurasi Keseluruhan : 98%.

Berdasarkan hasil *Classification Report*, kinerja model meningkat secara signifikan seiring dengan bertambahnya proporsi data latih. Hal ini terlihat pada rasio 90:10 yang mencapai akurasi tertinggi, yaitu sebesar 98%.

2. Evaluasi Model

Setelah pengujian model, tahap terakhir adalah evaluasi model *Support Vector Machine (SVM)* menggunakan *Confusion Matrix*. *Confusion Matrix* bertujuan untuk menganalisis performa model secara lebih mendetail dengan memetakan prediksi terhadap label sebenarnya. *Confusion Matrix* menunjukkan empat kategori utama, yaitu *True Positive (TP)*, *True Negative (TN)*, *False Positive (FP)*, dan *False Negative (FN)*. Berikut hasil evaluasi berdasarkan *Confusion Matrix* untuk setiap rasio pembagian data :

Protected by PDF Anti-Copy Free

(Upgrade to Pro Version to Remove the Watermark)

1) Rasio 50:50

a. Positif : TP = 3.828, FP_(neg) = 72, FP_(net) = 60
 b. Negatif : TP = 2.760, FP_(pos) = 89, FP_(net) = 53
 c. Netral : TP = 3.134, FP_(pos) = 14, FP_(neg) = 98

2) Rasio 60:40

a. Positif : TP = 3.072, FP_(neg) = 69, FP_(net) = 40
 b. Negatif : TP = 2.211, FP_(pos) = 71, FP_(net) = 35
 c. Netral : TP = 2.510, FP_(pos) = 13, FP_(neg) = 65

3) Rasio 70:30

a. Positif : TP = 2.281, FP_(neg) = 53, FP_(net) = 42
 b. Negatif : TP = 1.651, FP_(pos) = 43, FP_(net) = 17
 c. Netral : TP = 1.925, FP_(pos) = 12, FP_(neg) = 41

4) Rasio 80:20

a. Positif : TP = 1.557, FP_(neg) = 22, FP_(net) = 26
 b. Negatif : TP = 1.126, FP_(pos) = 26, FP_(net) = 8
 c. Netral : TP = 1.256, FP_(pos) = 2, FP_(neg) = 20

5) Rasio 90:10

a. Positif : TP = 809, FP_(neg) = 6, FP_(net) = 3
 b. Negatif : TP = 576, FP_(pos) = 8, FP_(net) = 7
 c. Netral : TP = 602, FP_(pos) = 1, FP_(neg) = 10



5.1 KESIMPULAN

Dari penelitian yang telah dilakukan mengenai Analisis Sentimen Ulasan Pengguna Aplikasi Sirekap Menggunakan Optimasi *Support Vector Machine* Dengan *Grid Search Cross-Validation*, dapat disimpulkan sebagai berikut :

1. Penerapan klasifikasi dengan optimasi *Support Vector Machine (SVM)* menggunakan *GridSearchCV* menghasilkan akurasi di atas 90% untuk semua rasio pembagian data. Rasio 90:10 memberikan hasil terbaik dengan akurasi 98%. Klasifikasi sentimen menunjukkan *precision*, *recall*, dan *f1-score* masing-masing 0.99 untuk kelas Positif, 0.97 untuk kelas Negatif, dan 0.98 untuk kelas Netral. Kombinasi parameter terbaik pada rasio 90:10 adalah kernel *RBF*, $C=1000$, dan $\text{Gamma}=1$, yang terbukti efektif meningkatkan akurasi dan kinerja model.
2. Penerapan *Grid Search Cross-Validation* berhasil mengidentifikasi kombinasi parameter optimal untuk model *Support Vector Machine (SVM)*, yaitu kernel *RBF* dengan $\text{Gamma} = 1$ dan nilai C yang bervariasi sesuai rasio pembagian data. Pada rasio 50:50, 60:40, dan 90:10, nilai C yang optimal adalah 1000, sementara pada rasio 70:30 dan 80:20, nilai C yang optimal adalah 100. Teknik ini memastikan model bekerja secara maksimal pada setiap rasio data.
3. Dibandingkan dengan penelitian sebelumnya yang juga menggunakan *Support Vector Machine (SVM)* dan *Grid Search Cross-Validation*, metode yang digunakan dalam penelitian ini dengan penambahan *SMOTEENN* menunjukkan performa yang lebih baik. Penelitian sebelumnya mencapai akurasi 81% dengan 1.791 data ulasan, sementara penerapan *SMOTEENN* secara

Protected by PDF Anti-Copy Free

(Upgrade to Pro Version to Remove the Watermark)

- signifikan mengatasi ketidakseimbangan data dan mengoptimalkan parameter model yang menghasilkan akurasi yang lebih tinggi dan peningkatan model secara keseluruhan.
4. Setelah dilakukan *labeling*, dihasilkan tiga kelas sentimen, yaitu Positif, Negatif, dan Netral. Dari total 17.506 ulasan, terdapat 14.296 ulasan dengan sentimen Negatif, 2.342 ulasan dengan sentimen Positif, dan 868 ulasan dengan sentimen Netral. Hasil ini menunjukkan bahwa persepsi pengguna terhadap Aplikasi Sistem Informasi Rekapitulasi (Sirekap) didominasi oleh ulasan negatif.
 5. Ulasan secara umum menekankan pentingnya perbaikan performa aplikasi, terutama terkait aksesibilitas dan stabilitas sistem. Berdasarkan visualisasi menggunakan *WordCloud*, masalah yang sering disoroti meliputi kesulitan login, gangguan server, *bug* atau *crash*, serta belum tersedianya aplikasi untuk perangkat iOS.

5.2 SARAN

Untuk meningkatkan hasil pada penelitian selanjutnya, beberapa saran yang dapat diterapkan antara lain :

1. Disarankan agar peneliti selanjutnya menambahkan metode klasifikasi lain seperti *Decision Tree*, *K-Nearest Neighbors*, dan *Naïve Bayes* untuk melakukan perbandingan dengan hasil yang diperoleh menggunakan *Support Vector Machine (SVM)*. Hal ini dapat memberikan gambaran lebih komprehensif mengenai performa masing-masing algoritma dalam mengklasifikasikan sentimen.
2. Selain itu, penggunaan metode *labeling* alternatif seperti *TextBlob*, *VADER Sentiment*, *NLTK*, atau Kamus *Lexicon* juga dapat diperhitungkan. Dengan berbagai pendekatan *labeling*, hasil klasifikasi dapat lebih terverifikasi dan dibandingkan untuk meningkatkan keakuratan dan kualitas model.

Protected by PDF Anti-Copy Free


(Upgrade to Pro Version to Remove the Watermark)

3. Peneliti selanjutnya juga dapat mengeksplorasi penggunaan arsitektur pembelajaran mendalam (*Deep Learning*) atau model berbasis transformasi seperti *BERT* untuk meningkatkan performa dalam mengatasi masalah yang lebih kompleks.
4. Disarankan agar peneliti selanjutnya melakukan pembagian data menjadi tiga bagian, yaitu Data Latih, Data Uji dan Data Validasi untuk meningkatkan performa model dan mengurangi resiko *overfitting*.

Protected by PDF Anti-Copy Free

(Upgrade to Pro Version to Remove the Watermark)

DAFTAR PUSTAKA

- 
- [1] S. A. Hasibuan, “Sistem [PDF] n Umum Dalam Kaitan Dengan Negara Demokrasi,” *War. Dharma*, vol. 17, no. 2, pp. 602–609, 2023, doi: 10.46576/wdw.v17i2.311
- [2] M. F. Y. Herjanto and C. Carudin, “Analisis Sentimen Ulasan Pengguna Aplikasi Sirekap Pada Play Store Menggunakan Algoritma Random Forest Classifier,” *J. Inform. dan Tek. Elektro Terap.*, vol. 12, no. 2, pp. 1204–1210, 2024, doi: 10.23960/jitet.v12i2.4192.
- [3] C. C. Gauru, R. Martini, and L. K. Alfirdaus, “Implementasi Sirekap Dalam Pilkada 2020 Kabupaten Semarang,” *Reformasi*, vol. 12, no. 2, pp. 224–230, 2022, doi: 10.33366/rfr.v12i2.3874.
- [4] I. A. Pradesa, “Analisis Penggunaan Sistem Rekapitulasi Suara (Sirekap) Dalam Menghadapi Problematika Pemilu 2024,” *Triwikrama J. Multidisiplin Ilmu Sos.*, vol. 03, no. 04, pp. 47–57, 2024.
- [5] A. Kusuma, E. Ermatita, and H. N. Irmada, “Analisis Sentimen Pada Ulasan Aplikasi Indodax di Google Play Store Menggunakan Metode Support Vector Machine,” *Pros. Semin. Nas. Mhs. Bid. Ilmu Komput. dan Apl.*, vol. 3, no. 2, pp. 773–782, 2022.
- [6] G. Natalianus, V. Kurniawan, and N. R. Feta, “Perbandingan Akurasi Algoritma Naïve Bayes Dan Support Vector Machine Dalam Analisis Sentimen Pengguna Twitter Terhadap Aplikasi Sirekap Abstrak,” vol. 9, no. 2, 2024.
- [7] J. Setyanto and T. B. Sasongko, “Sentiment Analysis of Sirekap Application Users Using the Support Vector Machine Algorithm,” *J. Appl. Informatics Comput.*, vol. 8, no. 1, pp. 71–76, 2024, doi: 10.30871/jaic.v8i1.7772.
- [8] A. Roihan, P. A. Sunarya, and A. S. Rafika, “Pemanfaatan Machine Learning dalam Berbagai Bidang: Review paper,” *IJCIT (Indonesian J. Comput. Inf. Technol.)*, vol. 5, no. 1, pp. 75–82, 2020, doi: 10.31294/ijcit.v5i1.7951.
- [9] E. M. Z. Darmawan and A. Fauzan Dianta, “Implementasi Optimasi Hyperparameter GridSearchCV Pada Sistem Prediksi Serangan Jantung Menggunakan SVM,” *Teknol. J. Ilm. Sist. Inf.*, vol. 13, no. 1, pp. 8–15, 2023, [Online]. Available: <https://doi.org/10.26594/teknologi.v13i1.3098>Tersediaonline di www.jurnal.unipdu.ac.idHalamanjurnal di www.jurnal.unipdu.ac.id/index.php/teknologi
- [10] N. B. Sidauruk and N. Riza, “Sentimen Analisis Data Pengguna Terhadap Kai Access Systematic Literature Review,” *J. Mhs. Tek. Inform.*, vol. 7, no. 2, pp. 1297–1303, 2023.
- [11] A. Hermawan, I. Jowensen, J. Junaedi, and Edy, “Implementasi Text-Mining

Protected by PDF Anti-Copy Free

(Upgrade to Pro Version to Remove the Watermark)

- untuk Analisis Sentimen pada Twitter dengan Algoritma Support Vector Machine,” *JST (Jurnal Sistem dan Teknol.*, vol. 12, no. 1, pp. 129–137, 2023, doi: 10.23887/jstundiksha.v12i1.52358.
- [12] R. Kurniawan, H. O. L. Y., and R. P. Aprisusanti, “Sentiment Analysis of Google Play Store User Reviews on Digital Population Identity App Using K-Nearest Neighbors,” *J. Sisfokom (Sistem Inf. dan Komputer)*, vol. 13, no. 2, pp. 170–178, 2024, doi: 10.32736/sisfokom.v13i2.2071.
- [13] I. Aida Sapitri and M. Fikry, “Pengklasifikasian Sentimen Ulasan Aplikasi Whatsapp Pada Google Play Store Menggunakan Support Vector Machine,” *J. TEKINKOM*, vol. 6, no. 1, pp. 1–7, 2023, doi: 10.37600/tekinkom.v6i1.773.
- [14] B. Mahesh, “Machine Learning Algorithms - A Review,” *Int. J. Sci. Res.*, vol. 9, no. 1, pp. 381–386, 2020, doi: 10.21275/art20203995.
- [15] S. W. Iriananda, R. W. Budiawan, and A. Y. Rahman, “Optimasi Klasifikasi Sentimen Komentar Pengguna Game Bergerak Menggunakan Svm , Grid Search Dan Kombinasi N-Gram Optimizing Sentiment Classification Of Mobile Game User Reviews Using Svm , Grid Search And N-Gram Combinations,” vol. 11, no. 4, 2024, doi: 10.25126/jtiik.1148244.
- [16] M. N. Muttaqin and I. Kharisudin, “Analisis Sentimen Pada Ulasan Aplikasi Gojek Menggunakan Metode Support Vector Machine dan K Nearest Neighbor,” *UNNES J. Math.*, vol. 10, no. 2, pp. 22–27, 2021, [Online]. Available: <http://journal.unnes.ac.id/sju/index.php/ujm>
- [17] M. Diki Hendriyanto, A. A. Ridha, and U. Enri, “Analisis Sentimen Ulasan Aplikasi Mola Pada Google Play Store Menggunakan Algoritma Support Vector Machine Sentiment Analysis of Mola Application Reviews on Google Play Store Using Support Vector Machine Algorithm,” *J. Inf. Technol. Comput. Sci.*, vol. 5, no. 1, pp. 1–7, 2022.
- [18] S. Widaningsih, U. Suryakencana, A. Suheri, and U. Suryakencana, “Klasifikasi Jurnal Ilmu Komputer Berdasarkan Pembagian Web of,” vol. 2018, no. March, pp. 23–24, 2018.
- [19] F. S. Jumeilah, “Penerapan Support Vector Machine (SVM) untuk Pengkategorian Penelitian,” *J. RESTI (Rekayasa Sist. dan Teknol. Informasi)*, vol. 1, no. 1, pp. 19–25, 2017, doi: 10.29207/resti.v1i1.11.
- [20] M. Beasiswa and P. I. P. Sdn, “Implementasi Metode MABAC Untuk Menentukan Siswa Yang Layak,” vol. 5, no. 4, pp. 1342–1351, 2024, doi: 10.47065/josh.v5i4.5347.
- [21] A. Nadroh, D. N. Triwibowo, and R. B. B. Sumantri, “Klasifikasi Status Gizi Balita Menggunakan Algoritma Support Vector Machine Dengan Optimasi Grid Search Cross-Validation,” vol. 8, no. 2, pp. 250–257, 2024.
- [22] C. Chairunnisa, I. Ernawati, and M. M. Santoni, “Klasifikasi Sentimen

Protected by PDF Anti-Copy Free

(Upgrade to Pro Version to Remove the Watermark)

- Ulasan Pengguna Aplikasi PeduliLindungi di Google Play Menggunakan Algoritma Support Vector Machine dengan Seleksi Fitur Chi-Square,” *Inform. J. Ilmu Ke* vol. 18, no. 1, p. 69, 2022, doi: 10.52958/iftk.v17i4.4594
- [23] C. Cahyaningtyas, Y. N. and I. R. Widiyanti, “Analisis Sentimen Pada Rating Aplikasi Shopee Menggunakan Metode Decision Tree Berbasis SMOTE,” *Aiti*, vol. 18, no. 2, pp. 173–184, 2021, doi: 10.24246/aiti.v18i2.173-184.
- [24] B. Ramadhani and R. R. Suryono, “Komparasi Algoritma Naïve Bayes dan Logistic Regression Untuk Analisis Sentimen Metaverse,” *J. Media Inform. Budidarma*, vol. 8, no. 2, p. 714, 2024, doi: 10.30865/mib.v8i2.7458.
- [25] A. Indrawati, “Penerapan Teknik Kombinasi Oversampling Dan Undersampling Untuk Mengatasi Permasalahan Imbalanced Dataset,” *JIKO (Jurnal Inform. dan Komputer)*, vol. 4, no. 1, pp. 38–43, 2021, doi: 10.33387/jiko.v4i1.2561.
- [26] G. G. Dodi Aleksander Manalu, “Implementasi Metode Data Mining K-Means Clustering Terhadap Data Pembayaran Transaksi Menggunakan Bahasa Pemrograman Python Pada Cv Digital Dimensi,” *Raspberry Pi OS Syst. Adm. with Syst. Python*, vol. 8, no. 1, pp. 175–305, 2023, doi: 10.1201/b23421-3.
- [27] D. Rika Widianita, “Pencarian Data Quick Count Pilpres Dengan Teknik Web Scraping,” vol. VIII, no. I, pp. 1–19, 2023.
- [28] M. A. Khder, “Web scraping or web crawling: State of art, techniques, approaches and application,” *Int. J. Adv. Soft Comput. its Appl.*, vol. 13, no. 3, pp. 144–168, 2021, doi: 10.15849/ijasca.211128.11.
- [29] N. Agustina, D. H. Citra, W. Purnama, C. Nisa, and A. R. Kurnia, “Implementasi Algoritma Naive Bayes untuk Analisis Sentimen Ulasan Shopee pada Google Play Store,” *MALCOM Indones. J. Mach. Learn. Comput. Sci.*, vol. 2, no. 1, pp. 47–54, 2022, doi: 10.57152/malcom.v2i1.195.
- [30] R. Maulana, A. Voutama, and T. Ridwan, “Analisis Sentimen Ulasan Aplikasi MyPertamina pada Google Play Store menggunakan Algoritma NBC,” *J. Teknol. Terpadu*, vol. 9, no. 1, pp. 42–48, 2023, doi: 10.54914/jtt.v9i1.609.
- [31] D. Normawati and S. A. Prayogi, “Implementasi Naïve Bayes Classifier Dan Confusion Matrix Pada Analisis Sentimen Berbasis Teks Pada Twitter,” *J. Sains Komput. Inform. (J-SAKTI)*, vol. 5, no. 2, pp. 697–711, 2021.
- [32] F. W. Nalle, “Analisis Faktor-Faktor Yang Mempengaruhi Tingkat Kemiskinan Di Kecamatan Insana Kabupaten Timor Tengah Utara (Ttu),” *Jurnal Dinamika Ekonomi Pembangunan*, vol. 1, no. 3. p. 35, 2019. doi: 10.14710/jdep.1.3.35-45.

Protected by PDF Anti-Copy Free
(Upgrade to Pro Version to Remove the Watermark)



LAMPIRAN

Protected by PDF Anti-Copy Free

(Upgrade to Pro Version to Remove the Watermark)

1. Lembar Pengajuan Judul



Formulir Pengajuan Judul Skripsi Program Studi Informatika

Nama : Sidik Fahriansyah
NIM : 2102020044
Alamat : Jln. Mangga, Dusun 1, Desa Tambah Asri, Kecamatan Tugumulyo
No.Hp : 0831-7884-3120

Rumusan Masalah 1 : Bagaimana algoritma *Random Forest* dan *SVM* dalam mengklasifikasikan sentimen ulasan pengguna aplikasi Sirekap ~~2024~~ ?
Judul 1 *Acc* : Perbandingan Algoritma *Random Forest* dan *SVM* (*Support Vector Machine*) dalam Analisis Sentimen Ulasan Pengguna Aplikasi Sirekap ~~2024~~ pada *Google PlayStore*

Rumusan Masalah 2 : Bagaimana algoritma *Decision Tree* dalam mendeteksi email spam pada dataset *Enron* ?
Judul 2 : Implementasi Algoritma *Decision Tree* untuk Deteksi Email Spam : Studi Kasus pada Dataset *Enron*


Rumusan Masalah 3 : Bagaimana efektivitas algoritma *Random Forest* dalam mendeteksi penyakit Typhoid berdasarkan data Rekam Medis pasien di Klinik Pratama Rawat Inap "UMMI" ?
Judul 3 : Implementasi Algoritma *Random Forest* untuk Deteksi Penyakit *Typhoid* Berdasarkan Data Rekam Medis Pasien : Studi Kasus di Klinik Pratama Rawat Inap " UMMI "

Lubuklinggau, 10 September 2024
Mahasiswa yang mengusulkan,



(Sidik Fahriansyah)

Menyetujui Dosen Pembimbing,

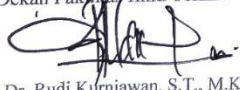
Pembimbing 1 (Dr. Muhamad Akbar, ST., M.IT)

 11/9/2024
(.....)

Pembimbing 2 (Muhammad Irvai, M. Kom)

 10/9/2024
(.....)

Mengesahkan,
Dekan Fakultas Ilmu Teknik


(Dr. Rudi Kurniawan, S.T., M.Kom)

Mengetahui,
Ketua Program Studi,


(Budi Santoso, M. Kom)

Protected by PDF Anti-Copy Free

(Upgrade to Pro Version to Remove the Watermark)

YAYASAN PENDIDIKAN DWI TUNGGAL PALEMBANG
UNIVERSITAS BINA INSAN
Soeharto KM.13 Kel. Lubuk Kupang Kec. Lubuklinggau Selatan I Kota Lubuklinggau Provinsi Sumatera Selatan

PDF **Surat Pengajuan Judul Skripsi**
Program Studi Informatika

Nama : Sidik Fahriansyah
NIM : 2102020044
Alamat : Jln. Mangga, Dusun 1, Desa Tambah Asri, Kecamatan Tugumulyo
No.Hp : 0831-7884-3120

Rumusan Masalah 1 : Bagaimana algoritma *Random Forest* dan *SVM* dalam mengklasifikasikan sentimen ulasan pengguna aplikasi Sirekap 2024 ?
Judul 1 : Perbandingan Algoritma *Random Forest* dan *SVM (Support Vector Machine)* dalam Analisis Sentimen Ulasan Pengguna Aplikasi Sirekap 2024 pada *Google PlayStore*

Rumusan Masalah 2 : Bagaimana algoritma *Decision Tree* dalam mendeteksi email spam pada dataset *Enron* ?
Judul 2 : Implementasi Algoritma *Decision Tree* untuk Deteksi Email Spam : Studi Kasus pada Dataset *Enron*

Rumusan Masalah 3 : Bagaimana efektivitas algoritma *Random Forest* dalam mendeteksi penyakit Typhoid berdasarkan data Rekam Medis pasien di Klinik Pratama Rawat Inap "UMMI" ?
Judul 3 : Implementasi Algoritma *Random Forest* untuk Deteksi Penyakit *Typhoid* Berdasarkan Data Rekam Medis Pasien : Studi Kasus di Klinik Pratama Rawat Inap "UMMI"

Lubuklinggau, 10 September 2024
Mahasiswa yang mengusulkan,

*Analisis Sentimen Ulasan pengguna Aplikasi Sirekap ~~2024~~
dan Menggambarkan Optimalisasi SVM dengan CV dan Search.*
(Sidik Fahriansyah)
- Merencanakan algoritma terbaik.

Menyetujui Dosen Pembimbing,

Pembimbing 1 (Dr. Muhamad Akbar, ST., M.IT) (.....)

Pembimbing 2 (Muhammad Irvai, M. Kom) (.....)

Mengesahkan,
Dekan Fakultas Ilmu Teknik

Mengetahui,
Ketua Program Studi,

(Dr. Rudi Kurniawan, S.T., M.Kom) (Budi Santoso, M. Kom)

Protected by PDF Anti-Copy Free

2. (Upgrade to Pro Version to Remove the Watermark) Lembar Bimbingan Proposal Skripsi Dosen Pembimbing 1



LEMBAR BIMBINGAN PROPOSAL SKRIPSI

Nama : Sidik Fahriansyah
Nim : 2102020044
Program Studi : Informatika
Pembimbing 1 : Dr. Muhammad Akbar, ST., M.IT
Pembimbing 2 : Muhammad Irvai, M.Kom
Judul : Analisis Sentimen Ulasan Pengguna Aplikasi Sircap Menggunakan Optimasi Support Vector Machine (svm) dengan Grid Search Cross-Validation (GridSearchCV)

NO	TANGGAL	TOPIK	KOMENTAR PEMBIMBING	TANDA TANGAN PEMBIMBING	
				1	2
	21/24 /20		Revisi latar belakang. Masukan		
			Tujuan Penelitian Revisi Sesuai dengan identifikasi masalah		
			Gambar 4. disesuaikan		
			Bab III Aca		
	23/2024. /11		Aca bab I, 2, 3 Sesuaikan proposal pertanyaan ujian proposal		

Lubuklinggau, 29 November 2024
Ketua Program Studi Informatika

(Budi Santoso, M.Kom)

Protected by PDF Anti-Copy Free

3. (Upgrade to Pro Version to Remove the Watermark) Lembar Bimbingan Proposal Skripsi Dosen Pembimbing 2



ASIAN PENDIDIKAN DWI TUNGGAL PALEMBANG
UNIVERSITAS BINA INSAN
FAKULTAS ILMU TEKNIK
Jl. Soeharto KM.13 Kel. Lubuk Kupang Kec. Lubuklinggau Selatan I Kota Lubuklinggau Provinsi Sumatera Selatan

LEMBAR BIMBINGAN PROPOSAL SKRIPSI

Nama : Sidik Fahriansyah
Nim : 2102020044
Program Studi : Informatika
Pembimbing 1 : Dr. Muhamad Akbar, ST., M.IT
Pembimbing 2 : Muhammad Irvai, M.Kom
Judul : Analisis Sentimen Ulasan Pengguna Aplikasi Sirekap Menggunakan Optimasi Support Vector Mechine (SVM) dengan GridSearch Cross Validation (GridSearchCV)



NO	TANGGAL	TOPIK	KOMENTAR PEMBIMBING	TANDA TANGAN PEMBIMBING	
				1	2
1.	27/2024 09	BAB I	<ul style="list-style-type: none">- penulisan bahara yang /inggris diminikan- format kutipan menggunakan IEEE- Koreksi identifikasi masalah dengan latar belakang- lengkap Bab 2 dan 3		07
2	20/2024 10	BAB I, II, III	<ul style="list-style-type: none">- penulisan bahara yang- perbaikan kerangka berpikir- perbaikan bab metode penelitian		07
3	25/2024 10	BAB I, II, III	sesuaikan dengan sistematika penulisan		07
4.	26/2024 10	BAB I, II, III	ACC lanjut ke p 2		07

Lubuklinggau, 29 November 2024
Ketua Program Studi Informatika

(Budi Santoso, M.Kom)


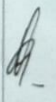
Protected by PDF Anti-Copy Free

4. (Upgrade to Pro Version to Remove the Watermark) Lembar Bimbingan Skripsi Dosen Pembimbing 1

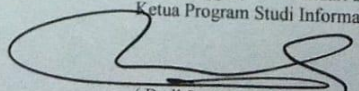
 **PDF**  **UNIVERSITAS BINA INSAN**
FAKULTAS ILMU TEKNIK
Jalan KM 1.3 Kel. Lubuk Kuning Kec. Lubuklinggau Selatan I Kota Lubuklinggau Provinsi Sumatera Selatan

LEMBAR BIMBINGAN SKRIPSI

Nama: Sidik Fahriansyah
Nim: 2102020044
Program Studi: Informatika
Pembimbing 1: Dr. Muhamad Akbar, ST., M.IT
Pembimbing 2: Muhammad Irvai, M.Kom
Judul: Analisis Sentimen Ulasan Pengguna Aplikasi Sirekap Menggunakan Optimasi Support Vector Machine Dengan Grid Search Cross-Validation

NO	TANGGAL	TOPIK	KOMENTAR PEMBIMBING	TANDA TANGAN PEMBIMBING	
				1	2
			<p>(+) komen DL.</p> <p>(+) <u>Struktur</u></p> <p>1. dikembangkan dengan arsitektur deep learn.</p> <p>2. Pembacaan data, Trainig. Testing Validasi.</p> <hr/> <p>App. ulasan. Struktur</p>		
					

Lubuklinggau, 07 Januari 2025
Ketua Program Studi Informatika


(Budi Santoso, M.Kom)

0271-42521400 (Marketing) Universitas Bina Insan
0271-42521400 (Admin) UNIVERSITAS BINA INSAN
0271-42521400 (Marketing) UNIVERSITAS BINA INSAN
0271-42521400 (Admin) UNIVERSITAS BINA INSAN

Protected by PDF Anti-Copy Free

5. (Upgrade to Pro Version to Remove the Watermark) Lembar Bimbingan Skripsi Dosen Pembimbing 2

UNIVERSITAS BINA INSAN
FAKULTAS ILMU TEKNIK
Jalan Raya K. 11 Kel. Lubuk Keping Kec. Lubuklinggau Selatan 1 Kota Lubuklinggau Prov. Sumatera Selatan

LEMBAR BIMBINGAN SKRIPSI

Nama :
Nim : 2102020044
Program Studi : Informatika
Pembimbing 1 : Dr. Muhamad Akbar, ST., M.IT
Pembimbing 2 : Muhammad Irvai, M.Kom
Judul : Analisis Sentimen Ulasan Pengguna Aplikasi Sirekap Menggunakan Optimasi Support Vector Machine Dengan Grid Search Cross-Validation

NO	TANGGAL	TOPIK	KOMENTAR PEMBIMBING	TANDA TANGAN PEMBIMBING	
				1	2
1.	07/2025 01	BAB 1 - $\sqrt{}$	<ul style="list-style-type: none">- Perbaiki bagian Pembahasan- Perbaiki coding terkait imbalansi data- Perbaiki grafik		
2.	08/2025 01	BAB 1 - $\sqrt{}$	<ul style="list-style-type: none">- Tambahkan daftar isi, gambar dll- Selesaikan dengan sistematika penulisan		
3.	08/2025 02	BAB 1 - $\sqrt{}$	<ul style="list-style-type: none">- Perbaiki kesimpulan- Penomoran- Cite penulisan b. Inggris / asing		
4.	08/2025 02	BAB 1 - $\sqrt{}$	<ul style="list-style-type: none">- BAB 1 - $\sqrt{}$ OK- ACC lanjut ke p1		

Lubuklinggau, 07 Januari 2025
Ketua Program Studi Informatika

(Budi Santoso, M.Kom)

Protected by PDF Anti-Copy Free

(Upgrade to Pro Version to Remove the Watermark)

6. Sampel Dataset Ulasan Aplikasi Sirekap

reviewid	rating	content	thumbsUp	date
1 05457824-07db-440a-8731-5be7f6b14db	5	Sy rasa mantull	0	2024-02-19 01:22:47
2 3adc3901-9993-4ade-9ca0-dcd179ca67cf	1	APK KENTANG	0	2024-02-10 18:57:55
3 ab11f456-9198-47f5-995d-61232bc16d6f	1	Enggak mutu	0	2024-02-25 07:10:41
4 39f464b7-4bb1-4f5e-aded-b6c1c81e57ae	1	data pelatihan tidak tereset, kolom di perbaiki	0	2024-02-12 22:28:53
5 5319ccca-3313-4945-9e72-76f98a0c95db	2	enapa ga bisa login. Alasannya device bim di lakukan inisialisasi. Mohon utk diperbaikannya.	0	2024-02-11 00:50:14
6 zbc1d809-e9ff-44d8-b95b-e10ad31f404	1	gak beres. Bikin aplikasi tidak slap. Cari developer yang handal, jangan developer abal abal.	1	2024-02-19 06:03:20
7 5b83b4e4-3b6c-4d29-a50a-032b25337e1e	5	Trims	0	2024-02-13 16:34:48
8 5558b6f5-4e28-47b6-920c-ce59721c0e04	1	Rhega Nata gagal login	0	2024-02-09 02:24:32
9 86f3ab9d-459d-4471-9b3a-5f3ce4eba336	5	Asyifa Ramadani Mdhwn lancar ketika d gunakan	0	2024-02-12 10:00:40
10 a5962e10-c4d1-4c5a-a0fb-b09f2638532f	1	Adi Purwanto Gagal masuk terus sudah ber hari han	0	2024-02-09 22:27:16
11 310132ea-5675-fc7f-a5b3-ca136e5ad80	1	Zhaa Maksas banget buat bikin aplikasi yang masih banyak minus, bug dan gabisa login	0	2024-02-09 18:02:42
12 b6d7f8c-094e-420b-8d57-9908262334f	1	Heim Maulana Saya curiga dengan aplikasi ini sdh ada yg mengendalikan	0	2024-02-11 13:23:56
13 f0cc195-927f-440a-9668-e36515298175	1	Indomobil Hino Nusra Aplikasi tidak berguna	0	2024-02-09 20:43:10
14 e62a1934-bb1d-4914-bb1d-51bf65edc5cc	1	Dedy Syahputra hancur negara... anggaran pemilu besar, tapi kualitas Nol... gak ada kredibilitas nya KPU ni...	0	2024-02-17 11:22:03
15 d7fa190b-13c1-410b-8f34-act1bde55be9	5	Rafka Iday jk kelancaran login dan penggunaan sebab susah masuk walau udah sesuai intruksi pps,ppk	0	2024-02-09 22:58:33
16 030432d9-cae9-462f-8347-9b127eca05e0	1	Sumiani Sumiani Belum bisa masuk	0	2024-02-10 05:01:22
17 2334e474-02a4-47c0-a17b-485ae285375	5	RIZWAN HDYT lgkatkan lagi, kalau nanti masih dipake di pilkada. Servemnya ditambah biar gk lemot wkwkw	0	2024-02-20 09:12:04
18 1c619964-1101-4746-9980-1280067cb1a2	1	KESW Apk tolok bukan mempermudah malah bikin resed	0	2024-02-15 19:09:34
19 fbc502c-5c0ff-4140-a111-ccccce2ae4f0	1	yun carne jin tapi muncul notifikasi "Please re-authenticate to continue Mohon diperbaiki. Terima kasih	0	2024-02-10 19:13:00
20 6001632d-478f-4d5c-8979-6e6da1b290c9	1	ngan Sekr. DPRD Medan Aplikasi yang sangat jelek, mau masuk pun tak bisa2, ada apa ihili	0	2024-02-15 19:24:22
21 fdef9eaf-7384-4092-ba0c-18f6590ddad2	1	Chindy Jiwanida Ariati Aplikasi burekiiiiiiii	0	2024-02-16 23:26:22
22 77eb7a70-1984-4d95-a40e-465fe15218ac	3	Zainal AE an data foto untuk rekap masih banyak yang tidak sesuai, proses verifikasi data terlalu lama	1	2024-02-10 07:43:37

reviewid	rating	content	thumbsUp	date
23 3bf57476-5fa4-4461-a3e8-675a5090342e	1	Rerung Yusuf paramisi Susah di perbaharui	0	2024-02-09 17:22:13
24 209fe074-e044-411a-b48c-06c934688dcd	1	M Badrul Anam Busuk aplikasinya sering error dan server sering down dan ga bisa buat edit jumlah suara	0	2024-02-23 23:23:32
25 ef54244b-c087-4076-b413-54f67d97e8db	5	Yohana bolo Baik	0	2024-02-13 13:33:35
26 81a82b62-c0ef-4ca8-a7c0-6101bfa529ac	5	Happi Gunawan Sya lom bisa msuk nh nihon d kondisikan	0	2024-02-10 13:16:43
27 37524eaf-4a6f-4445-a86a-02796a3238ca	1	Pak Dion mang disetting untuk memenangkan Paslon Prabowo Gibran!!!! cueh!!!! aplikasi sampah!!!!	1	2024-02-18 10:10:17
28 d71d16ef-2c6e-4469-9ac7-daf59f7e52d5	5	Mohammad soleh Trims	0	2024-02-10 09:38:46
29 e52c38f8-6627-417a-9b64-f1209e045ace	1	siti rakhmawati ana ini monedeati pemilu kok masih eror saja aplikasinya. Kami sirekap jadi bingung sendiri.	0	2024-02-09 04:36:56
30 d73b03c-0e62-4b52-b495-9a115e737ac	1	MIMI BJ asalon tertentu seperti memang di rancang untuk sebuah tujuan merugikan pihak 01 dan 03.	0	2024-02-17 06:47:04
31 6baed842-1407-419b-b184-c717a93c8c39	1	Dido Iskandar Aplikasi gagal	0	2024-02-16 23:21:59
32 f8b7d97e-2bb8-45ac-9a54-59b365ecbb7f	2	elvia Angraeni Rakhmadi tpi sebalikny. Tolong di perhatikan lagi, kerja jd terhambat kalau seperti ini. TERIMAKASIH	0	2024-02-09 18:55:34
33 f5a32aa5-2c8e-46d0-9e24-f11c6941459f	1	Ahmad Fahri ah di pegang admin 2 jadi kita admin 1 gabisa masuk ke menu nasional 1, tolonglah denya	0	2024-02-09 17:11:28
34 0002326e-8706-42ee-97ec-5040e3f329b	1	Novelia Kalimbe Saat foto kameranya ke zoom sndri dan hasilnya gelap	0	2024-02-11 17:02:36
35 c195285e-d3d8-4525-95ed-1ae713461666	1	Teguh Triono Gak bisa dibuka	0	2024-02-17 17:10:29
36 63edc08a-6b0e-4a35-b8cd-06db2d964442	1	Haishi Yalloh Aplikasi gak jelas	0	2024-02-17 15:55:53
37 db45915a-4cc7-4d36-a866-aef352cd5516	5	Faris Maulana Bim bisa juga nih, solusinya gimana ya	0	2024-02-09 14:49:51
38 26bb6b3d-c03c-4c30-a89c-ck779277a7	1	39_Dedi Sarif Hidayatulloh aplikasi yang error ini, pekerjaan kami juga terhambat karena ngakalin masalah aplikasi ini	0	2024-02-21 22:27:27
39 eec839b6-2324-4b7e-bf06-8b8c4033ab3	1	Kompom Mledug aplikasi bodol	0	2024-02-10 11:58:04
40 4998402b-b834-41b6-83eb-b31c927ed996	1	KAYRA CELL APLIKASI DONTOL!!	0	2024-02-15 03:36:03
41 f9d00c4d-712d-444e-9f8e-a2ec8b9f94d2	1	Dama Yanti hasil sudah selesai, gak peduli dengan si rekap, gak bisa terus cape coba2 terus. Wassalam	0	2024-02-22 18:06:15
42 3279556e-efb6-414b-a596-8f5dfabca15	1	Munfarid Munfarid Aplikasi terbaru yg pernah kupakai di smartphone	0	2024-02-29 06:23:07
43 d0c70755-e597-4e2b-b348-516d82ad440	1	Wayan Werdiana Aplikasi nasional kok sperti ini...	0	2024-02-13 17:26:29
44 49b6e411-4a83-47d8-e8d0-0a6f93b3a02b	1	Aplikasi yang jelek	0	2024-02-09 18:06:44

reviewid	rating	content	thumbsUp	date
44 d6064311-4c83-47d6-a650-03d7238c3e30	1	Ardiansyah lan APLIKASI SAMPAH HARI H GAK BISA DIPAKE	0	2024-02-16 09:51:41
45 03a9d130-d2f8-45b3-90e7-a626a748a5a	1	Ashari Moh The Real Beban Negara	1	2024-02-21 09:01:57
46 c36ffca2-1021-41ee-bd15-d814bd4f336e	1	yamin chennel Lancar jaya	0	2024-02-12 19:10:52
47 526d1837-88b2-48b1-a10d-29c0f94cb330	1	Andi Erik sandi Mau masuk gagal terus bukan mempermudah malah mempersulit	0	2024-02-09 12:39:08
48 80cd9c22-ea7c-451b-a930-8d102c69949d	1	Jam'ani Nor udah berhasil upload. Dana aja gede 2.8 Miliar bikin aplikasi ini kemana tuh lan uanganya	0	2024-02-22 20:59:32
49 2f7f1872-b629-474f-ac5c-a86e038f6b54	1	Dery Apriansyah Susah untuk login	0	2024-02-08 21:30:35
50 ff55721-28e2-4747-9de3-6b358ec2d4eb	1	Ito Chanel Krpa ini gk bisa dilogin banyak ribetnya	0	2024-02-10 09:20:50
51 c3929add-c631-4d0c-88da-dcc271ed5998	5	SUMADI Sampai sekarang bim bisa login	0	2024-02-10 07:00:33
52 2bcb13a5-7963-4d29-a077-94234f5299b6	1	Neng sili Nurbaeni Arrrrggghhhh bikin setres aja nih aplikasi	0	2024-02-08 18:43:04
53 59b5e594-cad0-468f-b660-1be86c511441	1	Nur Kholis ket banget kadang2 gagal apalagi nanti pada pelaksanaan tgl 14 ngga kebayang repotnya!!	0	2024-02-09 13:29:12
54 89689984-968a-49ce-88fe-101f90aa6411	1	Brama Kumbara /a, tsdak dapat dibuka, dan TIDAK LAYAK digunakan utk penghitungan suara pemilu saat ini	2	2024-02-19 07:12:38
55 bcdabdd0-3e7b-4b8a-abe5-51c895cc1a90	1	marselinus jonda Harus Mete terus demi update app, ditamba dgn jaringan hanya adanya di Hutan Kaya gini	0	2024-02-13 18:47:35
56 6c14507a-e3c7-42fa-9981-8ab970b468b7	1	Supardi Etendi ius nya pengguna sirekap. Login udah tidak bisa lagi, sampai saat ini pun belum ada solusi.	0	2024-02-09 02:46:08
57 ac44227e-3a41-4996-a2ed-48aa35056f3c	1	Ibnu Naba Tujuan aplikasi ini adalah untuk KECEURANGAN	0	2024-02-13 16:55:34
58 26c25171-e5b4-481a-9fa1-20f83400405	1	Doa restu ITUH BUKTI SILANG XXX DI BACA ANOKA 888 SAMA APLIKASI CATAT INI? EMAIL AJA	1	2024-02-21 09:16:18
59 d154de7d-3505-43c6-a977-5b2b3aa969e9	1	Indah Ayu Damahadi nya pada saat hari pemilu, semoga d pemilu yg akan datang aplikasi nya bener2 udah siap	0	2024-02-25 18:13:17
60 4c0683f5-052f-4d05-aee0-49ced25398f8	1	De Bungsu di gagal mendapatkan kunci digital. Pemilu H-4 belum ada perbaikan gimana nanti kirimnya	0	2024-02-10 00:29:23
61 b2967a04-7d71-42fa-ac6d-f4a2be06e01	1	Nwi Dian Aplikasi ini bayak kecurangan	0	2024-02-17 23:27:31
62 a11746cb-ea71-4382-9f9d-301a81ae918e	1	Haifz Altamis Aplikasi payah... Mending tidak usah pake aplikasi kalau belum siap	0	2024-02-10 05:45:44
63 c273025c-4e46-48e3-b500-5f384e809cbe	1	johan ashari yang tersubmit itu terjadi penggelembungan suara yang menguntungkan salah satu paslon.	0	2024-02-17 19:25:22
64 2198276d-be0b-465c-9753-527d21128855	1	Indra Fattah bisa di upgrade, lebih parah dari yang sebelumnya... Padahal hari sudah mendekati pemilu.	0	2024-02-11 17:42:19
65 d06e2a2f-df69-4873-9073-1b3c1e3672a7	1	BERMAIN ASYIK SIREKAP tahap awal sudah banyak masalah, apalagi nanti waktu hari H?	0	2024-02-09 05:03:15

Protected by PDF Anti-Copy Free

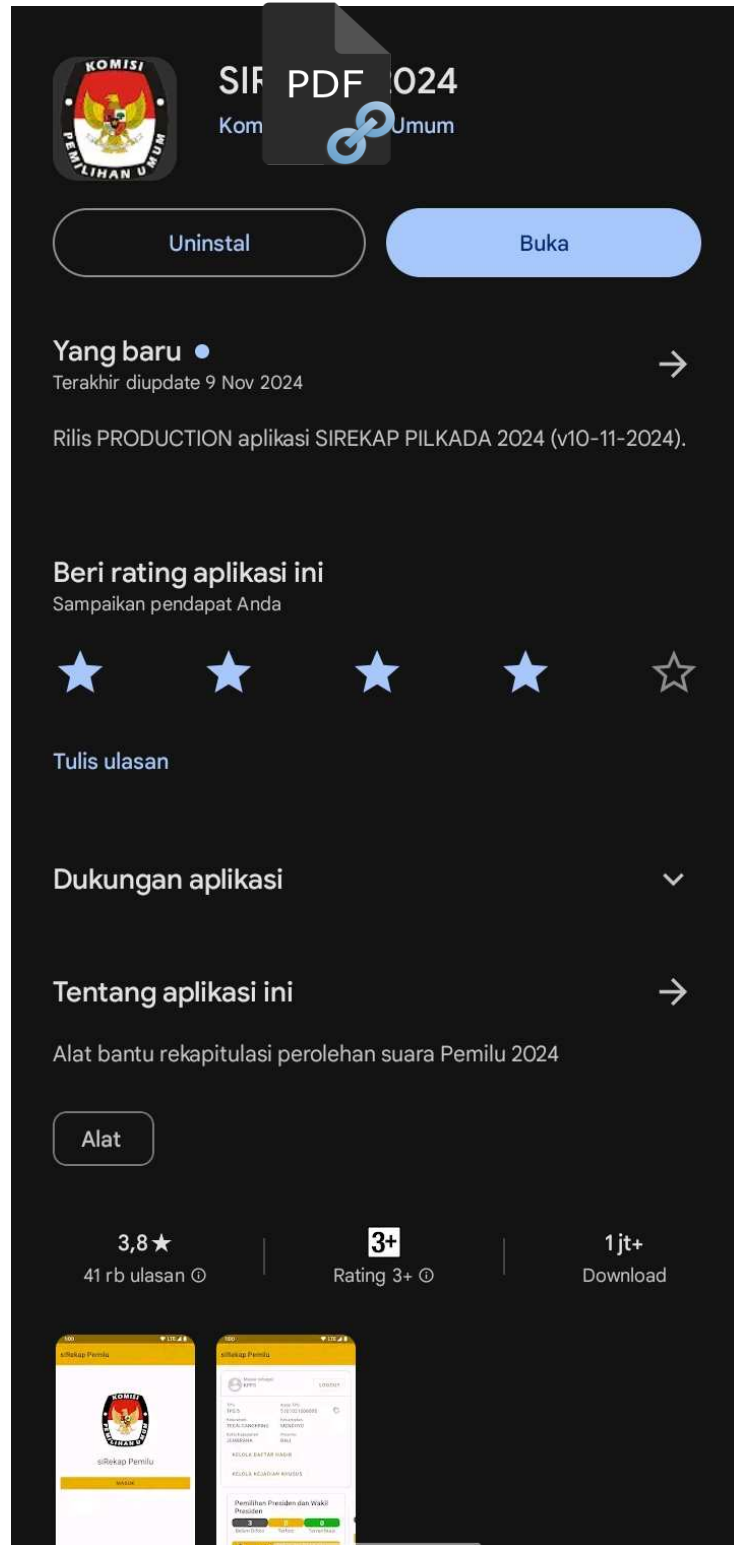
(Upgrade to Pro Version to Remove the Watermark)

reviewid	userName	rating	content	thumbsUp	date	
66	a4cc640-0aa3-4f90-8c09-4d2b040a312c	Jembut Ultraman	1	Mempersenjatai orang perang dengan pisau tumpul.	3	2024-02-15 11:19:32
67	6289a929-52df-42f3-ac0a-47195dc0039f	527	1	ik masuk clear data hapus apk ganti pw hp telep aja. ini gimanaa belum apa apa udah stres	0	2024-02-09 03:18:22
68	5bd3887-fa8c-464a-8364-e85e494939ff		1	Apk ngaco yg di input bisa berubah	0	2024-02-16 13:35:25
69	6e08act2-2603-4b8d-8e2b-1f425961c02d		1	g banyak... Satu indonesia ini lohh yg makee... 🙄 🙄 🙄 Mempersulit petugas lapangan aja...	0	2024-02-15 12:40:35
70	13a1224f-0241-4548-a721-4378625e75fa		1	Gausah maksain pake aplikasi deh, kl emng bin mampu	0	2024-02-08 22:01:41
71	47d3102b-16b-4795-ae50-3cc80ff14495		1	Sudah tidak bisa daftar	0	2024-02-16 04:35:25
72	f8c40baa-4772-473f-a7ea-3e18d7ca9681		1	Aplikasinya sangat membagongkan	0	2024-02-13 23:11:49
73	09d9292f-d07f-4c8d-98a4-fef538ab2e9f		1	bisa tembus tekan detail untuk masuk konek terverifikasi tidak bisa Mala dapat warna merah..	0	2024-02-11 23:18:08
74	bb41632c-864b-494b-9906-86c3e20a0a2d		1	aplikasi sangat buruk untuk digunakan	0	2024-02-09 13:16:32
75	0e6d3acd-e045-4529-83d9-dbeaf53d957d		1	Dari sebekum bintek sampai saat ini udh mau dekat pemilu blm juga dapat link WA dari Kpu	1	2024-02-10 09:10:55
76	1d360730-4e40-4768-8e7b-9be651890a5e	Ziunk Losari	1	lelet caww no optimal	0	2024-02-09 15:00:43
77	b72b371c-98ff-43c9-b11f-296607d25464	ia feri ardiansyah handika	1	Ga jelas . Susah login	0	2024-02-12 15:09:34
78	a35a9222-fa8e-4565-8b00-03f84974657a	eka tok	5	tingkatkan	0	2024-02-14 14:13:15
79	4de75d72-6291-4b19-86b8-7f029d164029	Auliya Zahra	1	1 fikat digital, harus bagaimana ya biar bisa kebuka aplikasinya, sudah hapus data dan cache	0	2024-02-09 12:42:45
80	1568d1f9-9e68-455c-b3df-f074b1688813	Abdi. Junaidi	1	Apk tak berguna sangat buruk	1	2024-02-11 16:44:17
81	c95369e8-3fc6-4404-b8b2-86e9430281c	Dayyan tsabitul azmi	1	Buruk	0	2024-02-17 23:50:44
82	385f82d1-b055-478c-bb24-b05186ccc5eb	chairil abdulah	1	Aplikasi nya jelek tidak bisa digunakan secara efisien	0	2024-02-08 17:23:00
83	4e77ba7e-1317-4047-b816-6e99436a320	Johan Saya	1	Bikin kesal gak bisa di gunakan berkali2 dicoba gagal kalau ada perbaikan di segerakan	0	2024-02-12 19:35:02
84	0e516326-c89f-44e6-a7bd-be7039fe723f	khoanu ddn	5	Sangat lancar	0	2024-02-11 08:57:49
85	17a9a233-36e7-4d1f-aa9a-13874168e069	Andi Mustamsiqin	1	Payah	1	2024-02-29 21:44:17
86	920511b5-3e2d-4073-b6d1-1136f6d78f9e	hammad Nurryan Afgani	1	Hajat nasional kok pakai aplikasi beta, sedih saya. Kasian hanya menghambat kerja kpps..	3	2024-02-14 19:37:15
87	4e03b0d-19f3-414f-b0e4-1007eae6b1f3	anggia dewi	3	a sudah bisa login malah keluar dengan sendirinya, belum bisa ikut simulasi sampai hari ini.	0	2024-02-09 12:34:36

reviewid	userName	rating	content	thumbsUp	date	
88	c8e92c2d-330c-4771-8e3-7cbe0581290	Adi Purnama	1	blong diperbaiki lagi itu saja sih nanti kalo ada tambahan lagi saya beri tahu lagi Trimakasih	0	2024-02-19 11:01:56
89	4f8aa58b-ada0-4a55-afb5-ed35ae9c4f50	ACCOUNT MASTER	1	ampe matanya pada berair dan lama lama bisa katarak nungguin aplikasi terbaik ini 🙏 🙏 🙏	0	2024-02-17 01:05:30
90	c144cc7e-0445-444c-b58e-65e9273dac4a	Combe Al	1	Aplikasi ini sangat membantu sekali, kerja jadi lebih lama 🙄 🙄 🙄	0	2024-02-12 02:30:07
91	881b065f-5413-45b7-8ff-35c06257b699	samsudin	1	proses aktivasi gagal, padahal sebelumnya bisa login, aneh	1	2024-02-09 06:46:33
92	e689ced9-e8d8-4e9a-80f6-6d053c480676	AGUS HARIADI	1	ni saya rasa sangat mengecewakan karna tidak bisa login dan di tidak bisa diakses kembali	0	2024-02-10 07:38:29
93	be8b4358-f115-42f9-8f52-1c6d085816da	Muda Wali	1	sangat rusak	0	2024-03-19 13:00:26
94	a87dad8b-8472-48a2-91d7-cd759e77483a	Winarto	5	Kenapa gak bisa di buka bos	0	2024-02-14 13:27:25
95	4229873-4273-481b-a4b1-fa534a551d32	Raka putra Sayuti	1	h emang kaga ada yg jelas masih alayyyyyy, dari jaman nya peduli lindungi juga kaga jelas	0	2024-02-10 07:59:58
96	e98acd4a-4384-4c1f-8638-c5dad530398	NELWAN SUBUHADI	5	Oke Aplikasi nya, semoga dapat digunakan pada 14 February 2024	0	2024-02-11 08:17:09
97	c326d27f-34ad-4b74-a2df-c308e020bba	Shofia Nufus	3	lap semangat buat semuanya. Dan yg buat Aplikasinya gambate pasti berhasil kami doakan	0	2024-02-09 23:49:29
98	cb55c650-d31c-4e0f-bc4d-9a13f3a01474	Andy Latifh	3	Mantap	0	2024-02-22 11:26:56
99	9caaad3-97c0-4d3f-b892-3b1519165218	hasbullah guru	5	Membantu kpps dalam melakukan hasil pemungutan suara.	0	2024-02-14 00:36:56
100	58c8009-8fca-4983-95d3-400ad9906ff	Melyi Farida	2	TDK BS login	0	2024-02-09 11:29:59
101	b1fa732c-cc6a-4e12-ab8e-ab596249c02a	syafria febri	1	Makin kesini makin banyak kendala	0	2024-02-11 09:55:45
102	fd9cd031-311e-487a-a61e-16e56101525	Martin D. Kurniawan	1	1egini? Mengecewakan. Kemana tim ahli dan yang lainnya untuk pengembangan aplikasi ini	0	2024-02-10 03:28:20
103	42dcd988-0c20-4aaf-a78a-181664b3ded8	zakafitriyadi akbar	1	Aplikasi parah. Gw kira udah pdf ga bisa di edit ternyata	0	2024-02-09 17:48:41
104	c8cb664-ba8b-4ac3-9806-1c95ab2c0caa	Santani	1	Aplikasi yg ribet	0	2024-02-10 11:19:35
105	a340428e-87a3-43d4-b344-fc930607066	yudihandika26	2	utama Sara adakan foto manual yang bisa ambil dari galeri untuk mempermudah pekerjaan	0	2024-02-23 10:24:01
106	6aa900f-0722-48df-8ca1-75f711c37a3	Yusuf Aswa	1	Buruk	0	2024-02-13 10:12:40
107	714ed58e-2a3d-4461-ae8c-d7b167a8cb3	Dodi Ansyah	5	Sangat bagus	0	2024-02-10 22:02:03
108	8031525a-7fc7-4b8a-adt7-e4841fcca46	Arif Rahman	1	Direkam mencla mende saya kpu	0	2024-02-17 06:56:16
109	8c544148-9924-4101-90ed-16ec4921920b	N SEMESTA INDONESIA	1	luk acara sepeenting ini... gak serius jadi pengembang, mundur aja memalukan anak bangsa	0	2024-02-15 03:55:55

Protected by PDF Anti-Copy Free

(Upgrade to Pro Version to Remove the Watermark)
7. Tampilan Aplikasi Sirekap di *Google Play Store*



Protected by PDF Anti-Copy Free

8. (Upgrade to Pro Version to Remove the Watermark) Source Code Program

“ANALISIS SENTIMEN ULOAN PENGGUNA APLIKASI SIREKAP MENGGUNAKAN OPTIMASI *SUPPORT VECTOR MACHINE* DENGAN *GRID SEARCH CROSS-VALIDATION*”

Berikut merupakan *source code* untuk proses pengambilan data ulasan Aplikasi Sistem Informasi Rekapitulasi (Sirekap) dari *Google Play Store* menggunakan *python* dan *Jupyter Lab* :

```
from google_play_scraper import Sort, reviews
import pandas as pd
import time

app_id = 'id.go.kpu.sirekap2024'
def get_all_reviews(app_id, lang='id', country='id', batch_size=100, max_reviews=20000, sleep_time=1):
    all_reviews = []
    next_token = None

    while len(all_reviews) < max_reviews:
        try:
            result, next_token = reviews(
                app_id,
                lang=lang,
                country=country,
                sort=Sort.NEWEST,
                count=batch_size,
                continuation_token=next_token
            )
            all_reviews.extend(result)
            if not next_token:
                break

            time.sleep(sleep_time)

        except Exception as e:
            print(f"Error occurred: {e}")
            break

    return all_reviews[:max_reviews]

app_reviews = get_all_reviews(app_id, max_reviews=20000)
data = []
for review in app_reviews:
    data.append({
        'reviewId': review['reviewId'],
        'userName': review['userName'],
        'rating': review['score'],
        'content': review['content'],
        'thumbsUp': review['thumbsUpCount'],
        'date': review['at'],
    })
df = pd.DataFrame(data)

output_file = 'sirekap_20000_reviews.csv'
df.to_csv(output_file, index=False)
print(f"Sebanyak {len(df)} ulasan berhasil disimpan ke '{output_file}'")
```

Protected by PDF Anti-Copy Free

(Upgrade to Pro Version to Remove the Watermark)

Berikut *source code* dari tahap *Data Cleaning* :

```
## Data Cleaning

import pandas as pd
import re
from IPython.display import

df = pd.read_csv('sirekap_20000_reviews.csv')

df = df.drop(columns=['reviewId', 'userName', 'thumbsUp'], errors='ignore')
df = df.dropna(subset=['content', 'rating'])
df = df.drop_duplicates(subset=['content'])
def clean_text(text):
    text = re.sub(r'\s+', ' ', text)
    text = re.sub(r'[^a-zA-Z\s]', '', text)

    emoji_pattern = re.compile(
        "[
        u'\U0001F600-\U0001F64F"
        u'\U0001F300-\U0001F5FF"
        u'\U0001F680-\U0001F6FF"
        u'\U0001F700-\U0001F77F"
        u'\U0001F780-\U0001F7FF"
        u'\U0001F800-\U0001F8FF"
        u'\U0001F900-\U0001F9FF"
        u'\U0001FA00-\U0001FA6F"
        u'\U0001FA70-\U0001FAFF"
        u'\U0002700-\U00027BF"
        ]+', flags=re.UNICODE)

    text = emoji_pattern.sub(r'', text)
    return text.strip()

df['content_cleaning'] = df['content'].apply(clean_text)
df['date'] = pd.to_datetime(df['date'], errors='coerce')
display(df[['date', 'rating', 'content', 'content_cleaning']].head())

print("\nJumlah data setelah cleaning:", df.shape[0])

df.to_csv('cleaning.csv', index=False)
print("\nData cleaning hasil disimpan ke 'cleaning.csv'")
```

Berikut *source code* dari tahap *Case Folding* :

```
## Case Folding

df = pd.read_csv('cleaning.csv')

df['content_case_folding'] = df['content_cleaning'].str.lower()
display(df[['date', 'rating', 'content', 'content_cleaning', 'content_case_folding']].head())
df.to_csv('case_folding.csv', index=False)
print("\nData hasil case folding disimpan ke 'case_folding.csv'")
```

Protected by PDF Anti-Copy Free

(Upgrade to Pro Version to Remove the Watermark)

Berikut *source code* dari tahap *Normalization* :

```
#Normalization
from IPython.display import display
data = pd.read_csv('case_folding')

data['content_normalization'], data['Kata_Tidak_Baku'], data['Kata_Tidak_Baku_Hash'] =
zip(*data['content_case_folding'].apply(lambda x: replace_taboo_words(x, kamus_tidak_baku)))
df = pd.DataFrame(data[['date', 'rating', 'content', 'content_cleaning', 'content_case_folding', 'content_normalization']])

display(df.head())

df.to_csv('normalization.csv', index=False)
print("\nData hasil case folding disimpan ke 'normalization.csv'")
```

Berikut *source code* dari tahap *Tokenizing* :

```
#Tokenizing
import pandas as pd
import nltk
from nltk.tokenize import word_tokenize

df = pd.read_csv('normalization.csv') # Ganti dengan nama file hasil stemming

# Tokenizing teks ulasan (pecah menjadi kata-kata)
def safe_word_tokenize(text):
    if isinstance(text, str): # Pastikan text adalah string
        return word_tokenize(text)
    return [] # Kembalikan list kosong jika bukan string

df['content_tokenizing'] = df['content_normalization'].apply(safe_word_tokenize)

# Tampilkan hasil tokenizing (misalnya, 10 baris pertama)
display(df[['date', 'rating', 'content', 'content_cleaning', 'content_case_folding', 'content_normalization', 'content_tokenizing']].head(10))

# Simpan hasil ke file CSV
df.to_csv('tokenizing.csv', index=False)
print("\nData hasil tokenizing disimpan ke 'tokenizing.csv'")
```

Berikut *source code* dari tahap *Stopword Removal* :

```
#Stopword Removal
import pandas as pd
from nltk.corpus import stopwords
from IPython.display import display
import nltk

# Ambil daftar stopwords dari NLTK
stop_words = set(stopwords.words('indonesian')) # Ganti dengan 'indonesian' jika menggunakan bahasa Indonesia

# Baca data yang sudah dibersihkan
df = pd.read_csv('tokenizing.csv') # Ganti dengan nama file hasil cleaning

# Stopword Removal
def remove_stopwords(text):
    if isinstance(text, str): # Pastikan input adalah string
        return ' '.join(word for word in text.split() if word not in stop_words)
    return '' # Kembalikan string kosong jika bukan string (termasuk NaN)

# Terapkan stopwords removal pada kolom 'content_normalization'
df['content_stopword_removal'] = df['content_tokenizing'].apply(remove_stopwords)

# Tampilkan hasil (10 data teratas)
display(df[['date', 'rating', 'content', 'content_cleaning', 'content_case_folding', 'content_normalization', 'content_tokenizing', 'content_stopword_removal']].head(10))

# Tampilkan jumlah data
print("\nJumlah data setelah stopwords removal:", df.shape[0])

# Simpan hasil ke file CSV
df.to_csv('stopword_removal.csv', index=False)
print("\nData cleaning hasil disimpan ke 'stopword_removal.csv'")
```

Protected by PDF Anti-Copy Free

(Upgrade to Pro Version to Remove the Watermark)

Berikut *source code* dari tahap *Stemming* :

```
#Stemming
|
import pandas as pd
from nltk.corpus import stopwords
from Sastrawi.Stemmer.StemmerFactory import StemmerFactory
from IPython.display import display
import nltk

stop_words = set(stopwords.words('Indonesian'))

df = pd.read_csv('stopword_removal.csv')

factory = StemmerFactory()
stemmer = factory.create_stemmer()

def stemming(text):
    if isinstance(text, str):
        return ' '.join([stemmer.stem(word) for word in text.split()])
    return ''

df['content_stemming'] = df['content_stopword_removal'].apply(stemming)

display(df[['date', 'rating', 'content', 'content_cleaning', 'content_case_folding', 'content_normalization', 'content_tokenizing', 'content_stopword_removal', 'content_stemming']].head(10))

print("\nJumlah data setelah stemming:", df.shape[0])

df.to_csv('stemming_result.csv', index=False)
print("\nData cleaning hasil disimpan ke 'stemming_result.csv'")
```

Berikut *source code* dari tahap *Labeling* :

```
#Labeling

import pandas as pd

df = pd.read_csv('Hasil_Preprocessing_Data.csv')

def label_sentiment(rating):
    if rating in [1, 2]:
        return 'negatif'
    elif rating == 3:
        return 'netral'
    elif rating in [4, 5]:
        return 'positif'

df['sentiment'] = df['rating'].apply(label_sentiment)

print("Available columns in the dataset:")
print(df.columns.tolist())

columns_to_save = [
    'date', 'rating', 'content', 'content_cleaning',
    'content_case_folding', 'content_normalization',
    'content_tokenizing', 'content_stopword_removal',
    'content_stemming', 'sentiment'
]

existing_columns = [col for col in columns_to_save if col in df.columns]
df[existing_columns].to_csv('Hasil_Labeling_Data.csv', encoding='utf8', index=False)

display(df[existing_columns].head())
print("\nData with sentiment labels saved to 'Hasil_Labeling_Data.csv'")
```

Berikut *source code* dari tahap *Visualisasi Data* :

```
#Visualisasi Data

import pandas as pd
import seaborn as sns
import matplotlib.pyplot as plt

df = pd.read_csv('Hasil_Labeling_Data.csv')

sentiment_counts_cleaned = df['sentiment'].value_counts()

sns.set(style='whitegrid')
plt.figure(figsize=(8, 6))
sns.barplot(x=sentiment_counts_cleaned.index, y=sentiment_counts_cleaned.values, palette="Blues_d")
plt.title('Distribusi Sentimen Berdasarkan Rating (Data Cleaned)')
plt.xlabel('Sentimen')
plt.ylabel('Jumlah')
```

Protected by PDF Anti-Copy Free

(Upgrade to Pro Version to Remove the Watermark)

```
plt.figure(figsize=(6, 6))
plt.pie(sentiment_counts_cleaned.values, labels=sentiment_counts_cleaned.index, autopct='%1.1f%%', colors=sns.color_palette("Blues_d", len(sentiment_counts_cleaned)))
plt.title('Proporsi Sentimen Berdasarkan Rating')
plt.show()

#WordCloud

from wordcloud import WordCloud
import matplotlib.pyplot as plt

teks_gabungan = " ".join(df['content_stemming'])

wordcloud = WordCloud(width=800, height=400, background_color='white', colormap='viridis').generate(teks_gabungan)

plt.figure(figsize=(10, 5))
plt.imshow(wordcloud, interpolation='bilinear')
plt.axis('off')
plt.title("Wordcloud Gabungan - Semua Kelas", fontsize=16)
plt.show()
```

Berikut *source code* dari tahap *TF-IDF* :

```
#TF-IDF

from sklearn.feature_extraction.text import TfidfVectorizer
import pandas as pd

df['content_stemming'] = df['content_stemming'].astype(str)

vectorizer = TfidfVectorizer(
    min_df=0.1,
    max_df=0.9,
    max_features=5000,
    stop_words=None
)

tfidf_matrix = vectorizer.fit_transform(df['content_stemming'])
tfidf_df = pd.DataFrame(
    tfidf_matrix.toarray(),
    columns=vectorizer.get_feature_names_out()
)

y = df['sentiment']

print("\nHasil TF-IDF untuk 10 data teratas:")
print(tfidf_df.head(10))

print("\nDistribusi awal label:")
print(y.value_counts())
```

Berikut *source code* dari tahap *SMOTEENN* :

```
#SMOTEENN

from imblearn.over_sampling import SMOTE

smote = SMOTE(random_state=42)
X_resampled, y_resampled = smote.fit_resample(tfidf_df, y)

print("Distribusi label setelah SMOTE:")
print(y_resampled.value_counts())
```

Protected by PDF Anti-Copy Free

(Upgrade to Pro Version to Remove the Watermark)

Berikut *source code* dari tahap *Split Data* :

```
#Split Data
from sklearn.model_selection import train_test_split
X_train, X_test, y_train, y_test = train_test_split(X_resampled, y_resampled, test_size=0.1, random_state=101)
print("\nDistribusi label pada training set:")
print(y_train.value_counts())
print("\nDistribusi label pada testing set:")
print(y_test.value_counts())
```

Berikut *source code* dari tahap *Hyperparameter Tuning* :

```
#GridSearchCV
from sklearn.svm import SVC
from sklearn.model_selection import GridSearchCV
param_grid = {'C': [0.1, 1, 10, 100, 1000],
              'gamma': [1, 0.1, 0.01, 0.001, 0.0001],
              'kernel': ['rbf', 'linear', 'poly', 'sigmoid']}
svc = SVC()
grid = GridSearchCV(SVC(), param_grid, refit = True, verbose = 3, cv=3)
grid.fit(X_train, y_train)
print("Parameter terbaik dari GridSearchCV:")
print(grid.best_params_)
```

Berikut *source code* dari tahap *Training Model* :

```
#Training Model
best_params = grid.best_params_
best_svm = SVC(**best_params, random_state=42)
best_svm.fit(X_train, y_train)
print("Model telah dilatih dengan parameter terbaik.")
```

Berikut *source code* dari tahap *Testing Model* :

```
#Testing Model
from sklearn.metrics import classification_report, confusion_matrix
y_pred = best_svm.predict(X_test)
print("\nClassification Report:")
print(classification_report(y_test, y_pred))
```

Protected by PDF Anti-Copy Free

(Upgrade to Pro Version to Remove the Watermark)

Berikut *source code* dari tahap Evaluasi Model :

```
#Evaluasi Model
|
import seaborn as sns
import matplotlib.pyplot as plt

cm = confusion_matrix(y_test, y_pred)

plt.figure(figsize=(7,5))
sns.heatmap(cm, annot=True, fmt='d', cmap='Blues', xticklabels=['Negatif', 'Netral', 'Positif'], yticklabels=['Negatif', 'Netral', 'Positif'])
plt.xlabel('Prediksi')
plt.ylabel('Aktual')
plt.title('Confusion Matrix')
plt.show()
```

