


Protected by PDF Anti-Copy Free

(Upgrade to Pro Version to Remove the Watermark)

**ANALISIS SENTIMEN PENGGUNA APLIKASI QUR'AN
KEMENAG PADA  SAN GOOGLE PLAYSTORE
MENGUNAKAN METODE SUPPORT VECTOR MACHINE**



SKRIPSI

**Oleh :
YOGA YARKHAMSETIAWAN
NIM : 2102020164**

**PROGRAM STUDI INFORMATIKA
FAKULTAS ILMU TEKNIK
UNIVERSITAS BINA INSAN
2025**

Protected by PDF Anti-Copy Free

(Upgrade to Pro Version to Remove the Watermark)

HALAMAN PENGESAHAN SKRIPSI

PDF

**ANALISIS SENTIMEN BERGUNA APLIKASI QUR'AN
KEMENAG PADA ULASAN GOOGLE PLAY STORE
MENGUNAKAN METODE SUPPORT VECTOR MACHINE
(SVM)**

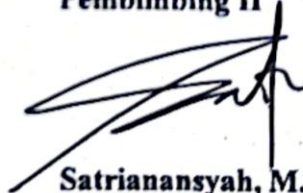
Oleh :
YOGA YARKHAMSETIAWAN
NIM : 2102020164

Pembimbing I



Dr. Muhammad Akbar, M.IT

Lubuklinggau, Januari 2025
Pembimbing II



Satrianansyah, M.Kom

Mengesahkan
Dekan Fakultas Ilmu Teknik
Universitas Bina Insan



UNIVERSITAS BINA INSAN
FAKULTAS ILMU TEKNIK

Dr. Rudi Kurniawan, ST, M.Kom

Protected by PDF Anti-Copy Free

(Upgrade to Pro Version to Remove the Watermark)

HALAMAN PERSETUJUAN TIM PENGUJI

PDF

Pada hari Sabtu Tanggal Dua Puluh Lima Januari Tahun Dua Ribu Dua Puluh Lima telah dilaksanakan Sidang Skripsi oleh Program Studi Informatika Fakultas Ilmu Teknik Universitas Bina Insan.

Nama : Yoga Yarkhamsetiawan
NIM : 2102020164
Judul Skripsi : Analisis Sentimen Pengguna Aplikasi Qur'an Kemenag Pada Ulasan Google Playstore Menggunakan Metode Support Vector Machine (SVM)

Komisi Penguji

1. Ketua : Dr. Muhammad Akbar, M.IT
2. Sekretaris : Satrianansyah, M.Kom
3. Anggota : Harma Oktavia Lingga W, 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 :

- ❖ “Orang lain gak akan bisa paham *struggle* dan masa sulitnya kita, yang mereka ingin tahu hanya bagian *success stories*. Berjuanglah untuk diri sendiri walaupun gak ada yang tepuk tangan. Kelak diri kita dimasa depan akan sangat bangga dengan apa yang kita perjuangkan hari ini!!!”

Persembahan kepada :

- ❖ Bapak Radi Ariyanto dan Ibu Kartini yang selalu mendoakan, mendukung, dan menjadi sumber inspirasi dalam setiap langkah saya.
- ❖ Adikku Nifa Nur Fusilla dan keluarga, yang selalu memberikan kasih sayang dan semangat tanpa henti.
- ❖ Teman-teman, yang senantiasa menemani dan membantu di saat suka maupun duka.
- ❖ Dan tak lupa, untuk diri saya sendiri, yang terus berjuang dan bertahan hingga mencapai titik ini.

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

HALAMAN PERNYATAAN

PDF

Saya yang bertanda tangan di bawah ini :

Nama Mahasiswa : Yoga Yarkhamsetiawan

NIM : 2102020164

Program Studi : Informatika

Fakultas : Ilmu Teknik

Menyatakan dengan sesungguhnya bahwa penelitian dan penulisan Skripsi yang saya susun sebagai persyaratan untuk memperoleh gelar sarjana (S1) 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, Januari 2025



Yoga
Yoga Yarkhamsetiawan
NIM.2102020164

Protected by PDF Anti-Copy Free

(Upgrade to Pro Version to Remove the Watermark)

DAFTAR RIWAYAT HIDUP



Biodata

Nama : Yoga Yarkhamsetiawan
Tempat/Tanggal Lahir : Air Beliti, 09 Januari 2002
Jenis Kelamin : Laki-laki
Agama : Islam
Alamat : Dusun I Desa Air Beliti Kec. Tuah Negeri
Kab. Musi Rawas Prov. Sumatera Selatan

Pendidikan

- SD : SDN Bandung Banpres
- SMP/MTS Sederajat : MTs Mazro'illah
- SMA/MAN/SMK Sederajat : MA Mazro'illah

Pengalaman Organisasi dan Pelatihan (Opsional)

1. ISKAM (Ikatan Santri Kulyatul 'Arifin Mazro'illah)
2. KARANG TARUNA Desa Air Beliti

Prestasi akademik dan Non-akademik (Opsional)

No	Prestasi Akademik dan Non-Akademik	Tahun
1.	Juara 1 Festival Hadrah Se-Sumatera Selatan	2018
2.	Juara 3 POSPEDA Cabang Hadang	2019
3.	Juara 1 di Kelas Formal / Non-Formal	2018-2019

Protected by PDF Anti-Copy Free

(Upgrade to Pro Version to Remove the Watermark)

ABSTRACT



The Kemenag Qur'an application was developed to make it easier for Muslims to read and study the Qur'an. Despite having a rating of 4.6 on the Google Play Store with a total download of more than 500 thousand, the application has received complaints regarding bugs and performance. The purpose of this study was to evaluate sentiment in user reviews of the Kemenag Qur'an application using the Support Vector Machine (SVM) algorithm. A total of 3,000 reviews were taken from the Google Play Store using the web scraping method. Reviews with a three-star rating were excluded from the labeling process because they were neutral and did not clearly represent positive or negative sentiments. After the preprocessing process to remove noise in the data, 2,677 reviews remained. Data imbalance was handled using the Synthetic Minority Oversampling Technique (SMOTE). Data was labeled into two categories: positive and negative. The classification process applied SVM using the Radial Basis Function (RBF) kernel and a training and testing data division of 80:20. The model showed an accuracy of 91% before SMOTE and increased to 97% after SMOTE was applied. The analysis shows that words such as 'good,' 'thank you,' and 'alhamdulillah' appear most often in positive reviews while negative reviews highlight issues such as 'open verse,' 'page,' and 'version.' The results of this study provide relevant recommendations for developers, especially in terms of bug fixes, feature improvements, and efforts to maintain service quality to meet user expectations and needs.

Keywords: Sentiment Analysis, Kemenag Qur'an, Support Vector Machine, SMOTE, Application Review.

Protected by PDF Anti-Copy Free

(Upgrade to Pro Version to Remove the Watermark)

ABSTRAK



Aplikasi Qur'an Kemenag dirancang untuk memudahkan umat Muslim membaca dan mempelajari Al-Qur'an. Meskipun memiliki rating 4.6 di *Google Play Store* dengan total unduhan melebihi 500 ribu, aplikasi ini menerima keluhan terkait bug dan performa. Tujuan dari penelitian ini adalah untuk mengevaluasi sentimen dalam ulasan pengguna terhadap aplikasi Qur'an Kemenag dengan menggunakan algoritma *Support Vector Machine* (SVM). Sebanyak 3.000 ulasan diambil dari *Google Play Store* menggunakan metode *web scraping*. Ulasan dengan rating bintang tiga dikeluarkan dari proses pelabelan karena bersifat netral dan tidak mewakili sentimen positif maupun negatif secara jelas. Setelah proses preprocessing untuk menghilangkan noise dalam data, tersisa 2.677 ulasan. Ketidakseimbangan data ditangani dengan *Synthetic Minority Oversampling Technique* (SMOTE). Data dilabeli menjadi dua kategori: positif dan negatif. Proses klasifikasi menerapkan SVM yang menggunakan kernel *Radial Basis Function* (RBF) dan pembagian data pelatihan serta pengujian sebesar 80:20. Model menunjukkan akurasi sebesar 91% sebelum SMOTE dan meningkat menjadi 97% setelah SMOTE diterapkan. Analisis menunjukkan bahwa Kata-kata seperti 'bagus', 'terima kasih', dan 'alhamdulillah' paling sering muncul dalam ulasan positif sedangkan ulasan negatif lebih banyak menyoroti isu seperti 'buka ayat,' 'halaman,' dan 'versi.' Hasil penelitian ini memberikan rekomendasi yang relevan bagi pengembang, khususnya dalam hal perbaikan bug, peningkatan fitur, dan upaya menjaga kualitas layanan agar sesuai dengan harapan serta kebutuhan pengguna.

Kata kunci: Analisis Sentimen, Qur'an Kemenag, *Support Vector Machine*, SMOTE, Ulasan Aplikasi.

Protected by PDF Anti-Copy Free

(Upgrade to Pro Version to Remove the Watermark)

KATA PENGANTAR



Alhamdulillah puji dan syukur mahasiswa ucapkan kepada Allah SWT atas segala rahmat dan karunia-Nya yang telah memberikan kekuatan dan kesempatan, sehingga mahasiswa dapat menyelesaikan skripsi ini dengan maksimal, Untuk diajukan sebagai syarat menyelesaikan pendidikan program Sarjana (S-1) Pada Program Studi Informatika Fakultas Ilmu Teknik Universitas Bina Insan. Sholawat beserta salam semoga tetap tercurahkan kepada baginda Nabi Muhammad SAW, keluarga, sahabat, serta umatnya hingga akhir zaman.

Selama proses penulisan dan penyusunan skripsi ini, mahasiswa telah berusaha sebaik-baiknya untuk dapat menyelesaikan skripsi ini baik tepat pada waktunya. Mahasiswa menyadari bahwa skripsi ini tentunya masih jauh dari sempurna dan mungkin terdapat kesalahan baik sengaja maupun tidak sengaja. Oleh karena itu, kritik dan saran yang membangun tentunya sangat diharapkan dari berbagai pihak.

Mahasiswa mengucapkan banyak terima kasih kepada pihak-pihak yang telah membantu selama proses penyelesaian skripsi ini diantaranya yaitu :

1. Tuhan Yang Maha Esa, yang senantiasa memberikan kesehatan, kekuatan, dan petunjuk-Nya sehingga proses penyelesaian skripsi ini dapat berjalan dengan lancar.
2. Bapak Radi Ariyanto dan Ibu Kartini selaku orang tua saya yang telah banyak memberikan dukungan dan bantuannya dalam penulisan skripsi ini.
3. Bapak Dr. H. Sardiyo, MM. selaku Rektor Universitas Bina Insan.
4. Bapak Dr. Muhamad Akbar, S.T., M.IT Selaku Wakil Rektor 1 Universitas Bina Insan.
5. Bapak Wakhid Nur Mukhlis, M.Pd., M.M selaku Wakil Rektor II Universitas Bina Insan.

Protected by PDF Anti-Copy Free

(Upgrade to Pro Version to Remove the Watermark)

6. Bapak Dr. Rudi Kurniawan, ST, M.Kom selaku Dekan Fakultas Ilmu Teknik Universitas Bina Insan yang telah banyak memberikan bimbingan dan arah dalam penulisan skripsi ini.
7. Bapak Budi Santoso, M. Kom selaku Kepala Program Studi Informatika Fakultas Ilmu Teknik Universitas Bina Insan yang telah banyak memberikan bimbingan dan arah dalam penulisan skripsi ini.
8. Bapak Dr. Muhamad Akbar, S.T., M.IT selaku Pembimbing I yang telah banyak memberikan bimbingan dan arah dalam penulisan skripsi ini.
9. Bapak Satrianansyah, M.Kom selaku Pembimbing II yang telah banyak memberikan bimbingan dan arah dalam penulisan skripsi ini.
10. Ibu Harma Oktavia Lingga W, M.Kom selaku Penguji yang telah banyak memberikan bimbingan dan arahan dalam penulisan skripsi ini.
11. Seluruh Staf Dosen dan Karyawan Universitas Bina Insan Lubuklinggau yang telah banyak memberikan ilmu pengetahuan dan bimbingan kepada mahasiswa.
12. Selamat untukmu, Aku. Yoga Yarkhamsetiawan, terima kasih telah melalui perjalanan yang penuh keraguan, ketegangan, dan pengorbanan. Dalam 3 tahun 4 bulan ini, engkau tidak hanya belajar hukum, tetapi juga kehidupan dan keberanian. Terima kasih telah mengizinkan dirimu jatuh, bangkit, dan terus berjuang meskipun tidak selalu sesuai harapan. Di balik semua rintangan, engkau membuktikan bahwa tidak pernah menyerah. Terima kasih untuk dirimu, Aku.

Akhir kata semoga penelitian ini dapat bermanfaat bagi untuk penelitian selanjutnya.

Lubuklinggau, Januari 2025

Yoga Yarkhamsetiawan

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	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	xiv
Daftar Lampiran	xv
BAB I PENDAHULUAN.....	1
1.1 Latar Belakang Penelitian	1
1.2 Identifikasi masalah	3
1.3 Rumusan Masalah	3
1.4 Batasan Masalah.....	3
1.5 Tujuan dan Manfaat Penelitian	4
1.5.1 Tujuan Penelitian.....	4
1.5.2 Manfaat Penelitian	4
BAB II KAJIAN PUSTAKA	5
2.1 Literatur	5
2.1.1 Analisis Sentimen.....	5
2.1.2 Qur'an Kemenag	5
2.1.3 <i>Google Playstore</i>	6
2.1.4 <i>Support Vector Machine (SVM)</i>	6
2.2 Penelitian Relevan	9
2.3 Kerangka Berfikir.....	14
BAB III METODOLOGI PENELITIAN	15
3.1 Analisa Sistem	15
3.1.1 Analisa Sistem Yang Berjalan	15
3.1.2 Alternatif Pemecahan Masalah.....	16
3.1.3 Metode Analisa.....	17
3.2 Teknik Pemilihan Informan (Populasi, sampel dan sampling).....	23
3.2.1 Teknik Pengumpulan Data	23
3.2.2 Teknik Analisa Data	24
3.3 Tempat dan Waktu Penelitian	28

Protected by PDF Anti-Copy Free

(Upgrade to Pro Version to Remove the Watermark)

	3.3.1 Tempat Penelitian	28
	3.3.2 Waktu Penelitian	29
	3.3.3 Alat dan I	29
BAB IV	PENELITIAN DAN PEMBAHASAN	31
	4.1 Gambaran Umum (Gambaran Penelitian).....	31
	4.1.1 Gambaran	31
	4.1.2 Gambaran Organisasi	32
	4.2 Hasil	32
	4.2.1 Hasil Pengumpulan Data.....	32
	4.2.2 Hasil <i>Preprocessing</i>	33
	4.2.3 Hasil <i>Labelling</i>	34
	4.2.4 Hasil Ekstrasi Fitur	35
	4.2.5 Hasil <i>Data Balancing</i>	35
	4.2.6 Hasil <i>Split Data</i>	36
	4.2.4 Hasil Klasifikasi	37
	4.2.5 Hasil Evaluasi.....	38
	4.3 Pembahasan.....	43
	4.3.1 Penerapan Metode Dan Validitas Data.....	43
	4.3.2 Pengujian Hasil Analisa	45
BAB V	KESIMPULAN DAN SARAN	47
	5.1 Kesimpulan	47
	5.2 Saran	47

DAFTAR PUSTAKA

LAMPIRAN

Protected by PDF Anti-Copy Free

(Upgrade to Pro Version to Remove the Watermark)

DAFTAR TABEL



Halaman

Tabel 2. 1 Penelitian Relevan.....	9
Tabel 3. 1 Waktu Penelitian.....	29
Tabel 3. 2 Spesifikasi Perangkat	29
Tabel 3. 3 Spesifikasi Perangkat Lunak	30
Tabel 3. 4 Bahan.....	30
Tabel 4. 1 <i>Hasil Labelling</i>	34

Protected by PDF Anti-Copy Free

(Upgrade to Pro Version to Remove the Watermark)

DAFTAR GAMBAR



Halaman

Gambar 2. 1 Ilustrasi Prinsip SVM Pada Data <i>Linear</i>	7
Gambar 2. 2 Ilustrasi Prinsip SVM pada Data <i>Non-Linear</i>	8
Gambar 2. 3 Kerangka Berpikir	14
Gambar 3. 1 Diagram Pengolahan Data.....	18
Gambar 4. 1 Aplikasi Qur'an Kemenag.....	31
Gambar 4. 2 Hasil Pengumpulan Data.....	32
Gambar 4. 3 Hasil <i>Preprocessing</i>	33
Gambar 4. 4 Hasil <i>Data Balancing</i>	35
Gambar 4. 5 Hasil <i>Split Data</i> Sebelum dan Sesudah SMOTE	36
Gambar 4. 6 Hasil Klasifikasi SVM	37
Gambar 4. 7 Hasil <i>Confusion Matrix</i> Sebelum dan Sesudah SMOTE	38
Gambar 4. 8 Hasil <i>Classification Report</i> Sebelum dan Sesudah SMOTE.....	39
Gambar 4. 9 Hasil Validasi K-Fold Sebelum dan Sesudah SMOTE	40
Gambar 4. 10 <i>Wordcloud</i> Sentimen Negatif dan Positif	42

Protected by PDF Anti-Copy Free

(Upgrade to Pro Version to Remove the Watermark)

DAFTAR LAMPIRAN



Lampiran 1. Surat Keputusan Pengesahan Skripsi.....	53
Lampiran 2. Surat Keputusan Pengujian Proposal Skripsi.....	55
Lampiran 3. Surat Keputusan Pengujian Skripsi.....	57
Lampiran 4. Lembar Perbaikan Seminar Proposal.....	59
Lampiran 5. Lembar Perbaikan Ujian Skripsi.....	60
Lampiran 6. Lembar Bimbingan Proposal.....	61
Lampiran 7. Lembar Bimbingan Skripsi.....	63
Lampiran 8. Tabulasi Data Proposal.....	65
Lampiran 9. Tabulasi Data Skripsi.....	66
Lampiran 10. Dokumentasi Proposal.....	67
Lampiran 11. Dokumentasi Skripsi.....	68
Lampiran 12. Surat Keterangan Bebas Plagiasi.....	69
Lampiran 13. Jurnal.....	70
Lampiran 14. Listing Program.....	80

Protected by PDF Anti-Copy Free

(Upgrade to Pro Version to Remove the Watermark)

BAB I

PRAHULUAN

1.1 Latar Belakang Penelitian

Kemajuan teknologi digital telah membawa perubahan signifikan dalam berbagai aspek kehidupan, termasuk cara umat Muslim mempelajari dan membaca Al-Qur'an. Berkat perkembangan teknologi, al-Qur'an kini dapat kita pelajari melalui media digital dengan menggunakan akses internet[1]. Pada akhir agustus 2016, Kementerian Agama, dalam hal ini Lajnah Pentashihan Mushaf Al-Qur'an, meluncurkan produk Al-Qur'an digital bernama "Qur'an Kemenag"[2]. Qur'an Kemenag adalah aplikasi atau *software* yang berbasis android yang mudah diunduh di *Google Play Store*[3]. Hal ini menjadi keuntungan tersendiri bagi umat islam, mengingat digitalisasi ini mempermudah dalam membaca, mencari ayat, tanda tajwid, dan asbabun nuzul dan lain sebagainya[4].

Pada *Google Play Store*, terdapat fitur ulasan bagi masing-masing aplikasi, termasuk aplikasi Qur'an Kemenag. Ulasan tersebut mencerminkan penilaian positif (kepuasan dan pujian) atau negatif (keluhan, kritik, dan saran) dari pengguna aplikasi. Ulasan pengguna dapat menjadi acuan calon pengguna dalam membuat keputusan memakai atau tidak suatu aplikasi atau layanan[5]. Bagi pihak pengelola aplikasi, ulasan pengguna bermanfaat untuk evaluasi aplikasi, meningkatkan kualitas layanan dan menyusun strategi dalam mempertahankan loyalitas pengguna[6].

Aplikasi Qur'an Kemenag di *Google Play Store* memiliki rating 4,6 bintang dari lebih 7 ribu ulasan dan telah diunduh lebih dari 500 ribu kali. Meskipun mendapatkan rating yang cukup tinggi, hal tersebut tidak dapat dijadikan tolak ukur kepuasan karena aplikasi Qur'an Kemenag masih menghadapi berbagai keluhan terkait pengalaman pengguna, salah satunya adalah bug pada fitur adzan. Berdasarkan alasan diatas, maka diperlukan penelitian analisis sentimen yang mengkaji ulasan pengguna[7] Qur'an Kemenag, baik berupa kepuasan (Positif) maupun kekecewaan (Negatif) penggunaannya. Untuk meneliti dan menganalisis objektivitas hal tersebut maka diperlukan suatu metode dan analisis untuk

Protected by PDF Anti-Copy Free

(Upgrade to Pro Version to Remove the Watermark)

mengklasifikasikan komentar pengguna kedalam beberapa kategori, dimana dalam penelitian ini metode yang digunakan yaitu metode *Support Vector Machine* (SVM).

Aplikasi Qur'an Kemenag dalam penelitian ini karena meskipun memiliki rating tinggi dan jumlah unduhan yang signifikan, masih terdapat keluhan yang menunjukkan adanya potensi masalah dalam pengalaman pengguna, yang perlu dianalisis lebih lanjut. Mengingat pentingnya aplikasi ini bagi umat Muslim di Indonesia, evaluasi dan perbaikan aplikasi ini menjadi hal yang krusial untuk memastikan kualitas pelayanan yang lebih baik di masa depan. Metode analisis sentimen menggunakan *Support Vector Machine* (SVM) dipilih karena kemampuannya dalam mengklasifikasikan teks secara akurat dan efisien, khususnya dalam membedakan sentimen positif dan negatif dari ulasan pengguna. Pada penelitian terdahulu, *Support Vector Machine* (SVM) merupakan algoritma dengan akurasi terbaik dalam melakukan analisis sentimen dibandingkan algoritma klasifikasi yang lain[8]. SVM telah terbukti efektif dalam menangani data yang tidak terstruktur seperti teks ulasan, sehingga sangat cocok untuk menganalisis data ulasan pengguna aplikasi Qur'an Kemenag guna mendapatkan wawasan yang lebih mendalam tentang kepuasan dan ketidakpuasan pengguna.

Meskipun analisis sentimen telah banyak digunakan di berbagai bidang, penelitian terkait aplikasi Qur'an Kemenag masih sangat terbatas. Berdasarkan studi sebelumnya, analisis sentimen menggunakan algoritma SVM terbukti lebih unggul dibandingkan dengan metode lainnya. Sebagai contoh, penelitian oleh Andry Novantika dan Sugiman yaitu, Penelitian ini menunjukkan bahwa algoritma *Support Vector Machine* (SVM) dengan *kernel linear* memiliki akurasi tertinggi sebesar 87,02%, dibandingkan dengan *Logistic Regression* yang hanya mencapai akurasi 85,17%, sehingga SVM lebih unggul dalam mengklasifikasikan sentimen ulasan aplikasi *Google Meet*[9]. Kemudian Ulfa Kusnia dan Fachrul Kurniawan Dalam penelitian ini, hasil eksperimen menunjukkan bahwa algoritma *Support Vector Machine* (SVM) memiliki tingkat akurasi yang lebih tinggi dibandingkan dengan algoritma *Naive Bayes*, yaitu sebesar 88% untuk SVM dan 87% untuk *Naive Bayes*, yang mengindikasikan bahwa SVM lebih unggul dalam mengklasifikasikan

Protected by PDF Anti-Copy Free

(Upgrade to Pro Version to Remove the Watermark)

data opini pengguna [10]. Selanjutnya oleh Rokhman et al Metode *Support Vector Machine* (SVM) menunjukkan hasil yang lebih baik dibandingkan dengan *Decision Tree*, dengan nilai akurasi tertinggi mencapai 90.20%, sedangkan akurasi yang diperoleh dari metode *Decision Tree* adalah 85.50% [11].

Penelitian ini memberikan solusi berupa informasi kepada pihak pengelola aplikasi Qur'an Kemenag untuk memahami dan menindaklanjuti keluhan pengguna secara tepat. Dengan mengetahui keluhan pengguna, pengembang dapat melakukan perbaikan yang sesuai dengan kebutuhan dan harapan pengguna. Hasil penelitian ini diharapkan dapat membantu meningkatkan kualitas aplikasi dan menjaga tingkat kepuasan pengguna untuk memenuhi kebutuhan umat Muslim di era digital. Maka dari itu, penulis tertarik untuk melakukan penelitian dengan judul “**Analisis Sentimen Pengguna Aplikasi Qur'an Kemenag Pada Ulasan Google Playstore Menggunakan Metode Support Vector Machine (SVM).**”

1.2 Identifikasi masalah

Meskipun aplikasi Qur'an Kemenag memiliki rating tinggi dan jumlah unduhan yang signifikan, terdapat berbagai keluhan dari pengguna, seperti bug pada fitur tertentu, yang menunjukkan potensi ketidakpuasan dalam pengalaman pengguna sehingga diperlukan analisis sentimen untuk mengevaluasi ulasan dan meningkatkan kualitas aplikasi.

1.3 Rumusan Masalah

Bagaimana analisis sentimen menggunakan metode *Support Vector Machine* (SVM) dapat membantu mengidentifikasi tingkat kepuasan dan keluhan pengguna terhadap aplikasi Qur'an Kemenag di *Google Play Store* untuk memberikan rekomendasi perbaikan yang tepat?

1.4 Batasan Masalah

Dalam mengidentifikasi dan menganalisis sentimen pada ulasan aplikasi Qur'an Kemenag menggunakan Metode SVM (*Support Vector Machine*), beberapa batasan masalah perlu diperhatikan untuk membatasi cakupan penelitian.

Protected by PDF Anti-Copy Free

(Upgrade to Pro Version to Remove the Watermark)

- 1.4.1 Data ulasan aplikasi Qur'an Kemenag diambil dari *Google Play Store* bulan agustus tahun 2016 sampai dengan bulan februari tahun 2025 menggunakan *google_play_scraper* dan *scrapy* untuk mendapatkan kata kunci kata kunci peraturan Indonesia dan urutan ulasan paling relevan.
- 1.4.2 Klasifikasi sentimen ulasan dibagi ke dalam dua kategori : positif dan negatif.
- 1.4.3 Metode yang digunakan adalah SVM (*Support Vector Machine*) dengan *kernel Radial Basis Function* (RBF).

1.5 Tujuan dan Manfaat Penelitian

1.5.1 Tujuan Penelitian

- a. Menganalisis sentimen ulasan pengguna aplikasi Qur'an Kemenag di *Google Play Store* untuk mengidentifikasi opini positif dan negatif terhadap aplikasi tersebut.
- b. Mengimplementasikan metode SVM (*Support Vector Machine*) dengan kernel RBF (*Radial Basis Function*).
- c. Memberikan rekomendasi yang relevan kepada pengembang aplikasi berdasarkan hasil analisis sentimen untuk meningkatkan kualitas layanan aplikasi Qur'an Kemenag.

1.5.2 Manfaat Penelitian

- a. Bagi pengembang aplikasi, hasil analisis dapat memberikan wawasan mengenai kebutuhan dan ekspektasi pengguna, sehingga membantu pengembangan fitur atau peningkatan kualitas aplikasi.
- b. Bagi peneliti, penelitian ini berkontribusi pada pengembangan literatur terkait analisis sentimen aplikasi al-Qur'an berbasis digital, khususnya "Qur'an Kemenag".
- c. Bagi masyarakat luas, penelitian ini diharapkan dapat meningkatkan kualitas aplikasi keagamaan yang diakses secara digital, sehingga memudahkan umat islam dalam mempelajari dan memahami al-Qur'an secara lebih efektif.

Protected by PDF Anti-Copy Free

(Upgrade to Pro Version to Remove the Watermark)

BAB II

KATA PENGANTAR



2.1 Literatur

2.1.1 Analisis Sentimen

Analisis sentimen dalam sebuah kalimat menggambarkan bagian pertimbangan suatu entitas atau peristiwa mengenai produk, layanan, topik, isu, orang, organisasi, atau peristiwa itu menjadi objek sasaran dalam kalimat sentimen[12]. Analisis sentimen digunakan sebagai gambaran umum media sosial untuk mengetahui pengetahuan perasaan lebih cenderung opini positif atau negatif[13].

Salah satu keuntungan dari analisis sentimen adalah dapat dilakukan secara otomatis, sehingga menghemat waktu dan sumber daya[14]. Hal ini memungkinkan penelitian dengan jumlah data yang besar dapat dilakukan lebih efisien. Selain itu, analisis sentimen membantu mengidentifikasi pola atau tren dalam data secara cepat, sehingga memberikan wawasan yang lebih mendalam untuk mendukung pengambilan keputusan.

Secara umum, analisis sentimen terbagi menjadi lima langkah yaitu *web scraping*, *pre-processing*, *labelling*, *classification*, dan *evaluation*. Analisis sentimen dapat mengubah data tidak beraturan menjadi data yang tersusun rapi. Manfaat adanya analisis sentimen yaitu sebagai evaluasi dan ide pada berbagai bidang. Analisis sentimen dapat menganalisis suatu kejadian, pernyataan, dan komentar yang kontroversi. Hasil dari analisis sentimen juga dapat menjadi sebuah gambaran bagi perusahaan, *public figure*, dan pemerintahan untuk menentukan langkah selanjutnya[15].

2.1.2 Qur'an Kemenag

Mengenal Aplikasi Qur'an Kemenag di mulai rilis dan dikembangkan pada tahun 2015 dan diluncurkan pada tahun 2016 oleh LPMQ Kementerian agama. Pada tahun 2018 aplikasi ini mengalami perbaikan pada sistem dan penambahan beberapa fitur baru. Kemudian diluncurkan pada tahun 2019. Sekarang Aplikasi

Protected by PDF Anti-Copy Free

(Upgrade to Pro Version to Remove the Watermark)

Qur'an Kemenag sudah mulai dikembangkan pada versi 2.1 dimana dilakukan perbaikan kembali dan diluncurkan pada tahun 2020[16].

Aplikasi *Mobile Qur'an* merupakan salah satu aplikasi digital dari Kementerian Agama Indonesia yang dirancang untuk memudahkan akses pembacaan dan pemahaman Quran[17]. Aplikasi Al-Qur'an versi kemenag ini menyajikan banyak sekali fitur yang sesuai dengan perkembangan waktu. Disamping menyajikan teks al-Qur'an 30 juz, aplikasi ini juga dilengkapi dengan terjemahan tafsir dalam dua varian: tahlil dan ringkas asbabun nuzul, dan suara murottal al-Qur'an Syeikh Mahmud Khalil al-Hushary[18].

2.1.3 Google Play Store

Google Play Store (dulunya bernama *Android Market*) merupakan toko aplikasi resmi milik *Google* untuk perangkat yang menggunakan sistem operasi Android[19]. Yang menyediakan berbagai konten digital, diantaranya yaitu aplikasi, permainan, dan lain sebagainya[20]. *Google Playstore* memiliki *Rating*, *Review*, dan *Installs*. *Rating* merupakan nilai yang diberikan pengguna kepada aplikasi yang mereka gunakan, tinggi rendahnya *Rating* dapat memberikan gambaran terhadap aplikasi tersebut. *Review* berisikan pendapat pengguna terhadap aplikasi dapat berbentuk kritik maupun saran. *Installs* merupakan indikasi yang menunjukkan keseluruhan dari total aplikasi telah diunduh oleh pengguna[21].

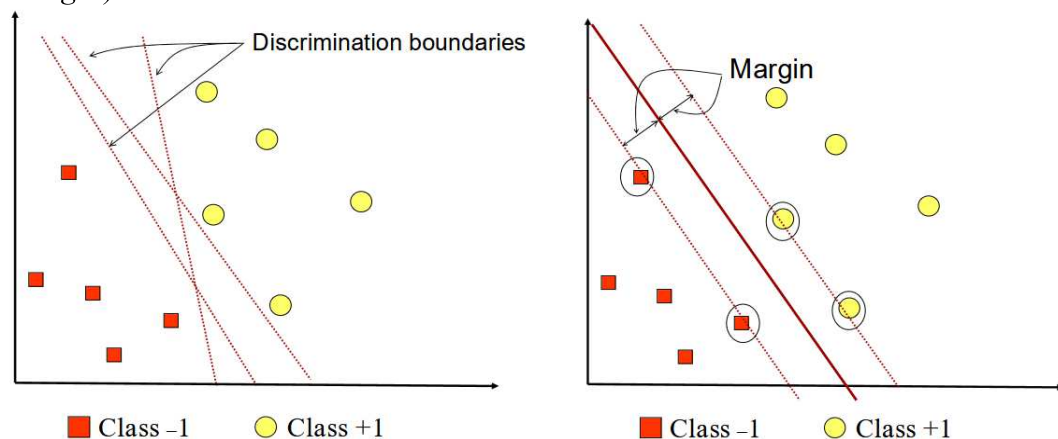
2.1.4 Support Vector Machine (SVM)

Support Vector Machine (SVM) pertama kali diperkenalkan oleh Vapnik pada tahun 1992 sebagai rangkaian harmonis konsep-konsep unggulan dalam bidang *pattern recognition*. Sebagai salah satu metode *pattern recognition*, usia SVM terbilang masih relatif muda. Walaupun demikian, evaluasi kemampuannya dalam berbagai aplikasinya menempatkannya sebagai *state of the art* dalam *pattern recognition*, dan dewasa ini merupakan salah satu tema yang berkembang dengan pesat. SVM adalah metode *machine learning* yang bekerja atas prinsip SRM (*Structural Risk Minimization*) dengan tujuan menemukan *hyperplane* terbaik yang memisahkan dua buah *class* pada *input space*[22]. SVM bekerja dengan

Protected by PDF Anti-Copy Free

(Upgrade to Pro Version to Remove the Watermark)

memanfaatkan *support vector*, yaitu data-data yang berada paling dekat dengan *hyperplane* dan berperan penting dalam menentukan posisi *hyperplane* terbaik. SVM menggunakan hipotesis *linier* pada ruang fitur berdimensi tinggi, didukung oleh teori pembelajaran statistik. Metode ini memiliki performansi tinggi di berbagai bidang aplikasi, di mana tingkat akurasinya sangat bergantung pada fungsi kernel, parameter yang digunakan, serta pengaturan margin maksimum (*max margin*).



Gambar 2. 1 Ilustrasi Prinsip SVM Pada Data *Linear*

Gambar tersebut menunjukkan konsep SVM (*Support Vector Machine*), di mana garis merah tebal adalah *hyperplane* yang berfungsi sebagai garis pemisah antara dua kelas, yaitu kelas -1 (kotak merah) dan kelas +1 (lingkaran kuning). *Margin* adalah jarak antara *hyperplane* dan titik data terdekat dari masing-masing kelas, yang ditunjukkan oleh garis putus-putus, dan tujuan SVM adalah memaksimalkan *margin* ini agar pemisahan lebih optimal. Titik data yang terletak paling dekat dengan *hyperplane* disebut *support vectors* (ditandai dengan lingkaran di sekitar titik-titik tertentu), karena mereka menentukan posisi *hyperplane* dan memengaruhi *margin*. Gambar kiri menunjukkan berbagai *hyperplane* potensial dengan *margin* kecil, sedangkan gambar kanan menampilkan *hyperplane* optimal dengan *margin* terbesar, menunjukkan hasil dari proses optimasi SVM. Berdasarkan karakteristiknya, SVM dibagi menjadi dua jenis:

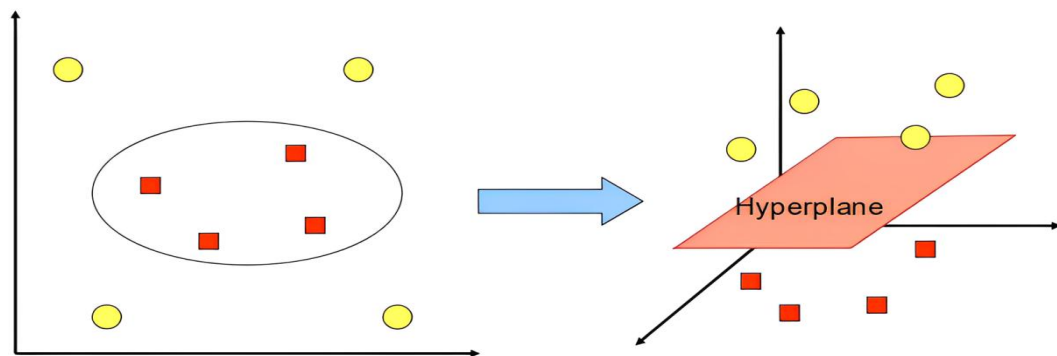
- a. SVM *Linier* : Memisahkan data dengan *hyperplane linier*.
- b. SVM *Non-Linier* : Menggunakan fungsi *kernel trick* untuk memisahkan data pada ruang berdimensi tinggi.

Protected by PDF Anti-Copy Free

(Upgrade to Pro Version to Remove the Watermark)

Karena tidak semua data dapat dipisahkan secara linear, SVM menggunakan *kernel trick* untuk menangani data yang bersifat *non-linear* dengan cara mentransformasikan data ke ruang yang berdimensi lebih tinggi, sehingga memungkinkan pencarian *hyperplane* yang dapat memisahkan data tersebut secara optimal. Beberapa jenis kernel yang umum digunakan diantaranya :

1. *linear kernel* (untuk data *linear*)
 2. *polynomial kernel* (untuk hubungan *non-linear* berbentuk *polinomial*)
 3. *radial basis function (RBF) kernel* atau *Gaussian kernel* (untuk data dengan pola *non-linear* kompleks)
- sigmoid kernel* (terkait dengan fungsi aktivasi *sigmoid*).



Gambar 2. 2 Ilustrasi Prinsip SVM pada Data *Non-Linear*

Gambar ini menunjukkan bagaimana SVM menangani data *non-linear* dengan menggunakan *kernel trick*. Pada sisi kiri, data tampak tidak dapat dipisahkan secara *linear* dalam ruang dua dimensi (2D), ditandai dengan adanya kelas-kelas yang saling tumpang tindih dalam sebuah pola melingkar. Untuk mengatasi hal ini, SVM mentransformasikan data ke dalam ruang berdimensi lebih tinggi (3D) menggunakan fungsi *kernel*. Setelah transformasi, seperti yang terlihat di sisi kanan gambar, data menjadi lebih mudah dipisahkan dengan *hyperplane*, yaitu bidang datar yang membagi dua kelas secara jelas. Proses ini memungkinkan SVM untuk bekerja secara efektif meskipun data awalnya bersifat *non-linear*.

SVM memiliki kelebihan seperti efektif dalam menangani klasifikasi *linear* maupun *non-linear*, *fleksibel* dengan berbagai pilihan *kernel*, serta mampu menghasilkan model yang akurat dengan pengendalian risiko *overfitting* melalui pengaturan *margin*. Namun, metode ini juga memiliki beberapa kekurangan, seperti

Protected by PDF Anti-Copy Free

(Upgrade to Pro Version to Remove the Watermark)

kebutuhan akan *normalisasi data*, sering memerlukan tuning *hyperparameter* yang cermat, kurang mudah diinterpretasikan dibandingkan model lain, serta kurang efisien ketika diterapkan pada dataset yang sangat besar.



2.2 Penelitian Relevan

Berikut merupakan beberapa penelitian terdahulu yang relevan dan menjadi landasan dalam penelitian ini, khususnya yang berkaitan dengan analisis sentimen pengguna aplikasi berbasis keagamaan, ulasan pengguna di *Google Play Store*, serta penerapan metode SVM (*Support Vector Machine*) dalam pengolahan teks untuk klasifikasi sentimen. Penelitian-penelitian ini memberikan gambaran tentang pendekatan, metode, dan hasil yang telah dicapai, sehingga menjadi acuan penting dalam pelaksanaan penelitian ini.

Tabel 2. 1 Penelitian Relevan

No	Judul	Tujuan	Metode	Hasil
1.	Analisis Sentimen pada review Aplikasi Grab di <i>Google Play Store</i> Menggunakan <i>Support Vector Machine</i> [23]	Menganalisis review pengguna aplikasi Grab di <i>Google Play</i> menggunakan SVM untuk meningkatkan kualitas layanan dan evaluasi perbaikan.	<i>Support Vector Machine</i> (SVM)	Analisis sentimen ulasan aplikasi Grab menggunakan SVM menghasilkan akurasi 85,54% dengan kata dominan "ovo" (positif) dan "driver" (negatif)
2.	Analisis Sentimen Aplikasi Glints Berdasarkan Ulasan	Penelitian ini menganalisis persepsi pengguna terhadap aplikasi Glints melalui analisis <i>sentimen</i>	<i>Support Vector Machine</i> (SVM) & <i>TF-IDF</i> (<i>Term</i>	Penelitian ini menggunakan SVM dengan akurasi 87,12% untuk menganalisis <i>sentimen</i> pengguna

Protected by PDF Anti-Copy Free

(Upgrade to Pro Version to Remove the Watermark)

No	Judul	Tujuan	Metode	Hasil
	<i>Google Play Store</i> Menggunakan Metode <i>Support Vector Machine</i> [24]	untuk mengidentifikasi kekuatan dan kelemahan, dan masalah yang dihadapi pengguna.	<i>Frequency-Inverse Document Frequency</i> .	aplikasi Glints dan mengidentifikasi perbaikan yang diperlukan, seperti layanan pencarian lowongan dan pembaruan aplikasi.
3.	Sentimen Analisis Review Aplikasi Digital Korlantas Pada <i>Google Play Store</i> Menggunakan Metode SVM[25]	Penelitian ini menganalisis sentimen ulasan pengguna aplikasi Digital Korlantas Polri menggunakan SVM untuk mengklasifikasikan opini menjadi kelas positif dan negatif.	<i>Support Vector Machine (SVM) & TF-IDF (Term Frequency-Inverse Document Frequency)</i> .	Penelitian ini menunjukkan bahwa metode SVM efektif untuk analisis sentimen ulasan aplikasi Digital Korlantas Polri, dengan akurasi 75% untuk sentimen negatif dan 77% untuk sentimen positif pada pembagian data 70:30.
4.	Analisis Sentimen Pada Ulasan Aplikasi Threads Di <i>Google Play Store</i> Menggunakan	Penelitian ini bertujuan untuk menganalisis sentimen ulasan pengguna aplikasi WhatsApp di <i>Google Play Store</i> menggunakan	<i>Support Vector Machine (SVM) & TF-IDF (Term Frequency-Inverse</i>	Hasil analisis sentimen menggunakan SVM pada aplikasi Threads menunjukkan akurasi 88%, dengan ulasan

Protected by PDF Anti-Copy Free

(Upgrade to Pro Version to Remove the Watermark)

No	Judul	Tujuan	Metode	Hasil
	Algoritma <i>Support Vector Machine</i> [26]	metode untuk mengukur dan meningkatkan kualitas layanan aplikasi.	<i>Support Document Vector Machine Frequency</i>).	positif lebih dominan dan hasil evaluasi yang baik pada sentimen positif.
5.	Analisis Sentimen Aplikasi MyXL menggunakan Metode <i>Support Vector Machine</i> berdasarkan Ulasan Pengguna di <i>Google Play Store</i> [27]	Penelitian ini menganalisis sentimen ulasan aplikasi MyXL menggunakan <i>SVM</i> untuk mengklasifikasikan sentimen, menguji pengaruh <i>kernel</i> , dan membandingkan data seimbang dan tidak seimbang.	ini <i>Support Vector Machine (SVM) & TF-IDF (Term Frequency-Inverse</i>	Hasil penelitian menunjukkan bahwa algoritma <i>SVM</i> mencapai akurasi, presisi, recall, dan <i>f-measure</i> 88% dalam mengklasifikasikan sentimen ulasan aplikasi MyXL dengan data seimbang dan <i>kernel linear</i> .
6.	Analisis Sentimen Aplikasi Novel Online Di <i>Google Play Store</i> Menggunakan Algoritma <i>Support</i>	Penelitian ini menganalisis umpan balik pengguna aplikasi novel online di <i>Google Play Store</i> menggunakan <i>SVM</i> untuk menilai kualitas aplikasi	ini <i>Support Vector Machine (SVM) & TF-IDF (Term Frequency-Inverse</i>	Penelitian ini menggunakan <i>SVM</i> untuk analisis sentimen aplikasi novel online, mencapai akurasi 90,67% dan memberikan wawasan tentang

Protected by PDF Anti-Copy Free

(Upgrade to Pro Version to Remove the Watermark)

No	Judul	Tujuan	Metode	Hasil
	<i>Vector Machine</i> (SVM)[28]	berda <i>sentimen</i> pengg	<i>Document Frequency</i> .	kepuasan pengguna serta area yang perlu diperbaiki.
7.	<i>Support Vector Machine: Analisis Sentimen Aplikasi Saham di Google Play Store</i> [29]	Menganalisis ulasan pengguna aplikasi investasi saham di Google Play Store untuk mengidentifikasi aplikasi terbaik menggunakan metode SVM.	<i>Support Vector Machine</i> (SVM) & <i>Rapid Miner</i>	aplikasi HSB dengan akurasi 88,70%, diikuti oleh Ajaib akurasi 61,89%), Pluang akurasi 68,25%), Bibit akurasi 64,89%), dan Stockbit akurasi 66,95%)
8.	<i>Analisis Sentimen Aplikasi Tiket Online Di Google Play Store</i> Menggunakan Metode <i>Support Vector Machine</i> (SVM)[30]	Penelitian ini menganalisis <i>sentimen</i> aplikasi tiket online di Google Play Store menggunakan SVM untuk menentukan aplikasi terbaik dan memberikan masukan peningkatan layanan.	<i>Support Vector Machine</i> (SVM) & <i>TF-IDF</i> (<i>Term Frequency-Inverse Document</i>).	Pegipegi: 78,21% Agoda: 77,00% Traveloka: 75,03% Mister Aladin: 64,00% Tiket.com: 58,68%
9.	<i>Analisis Sentimen</i>	Penelitian ini menganalisis	<i>Support Vector</i>	Ruang Guru: 83,33% (Tingkat

Protected by PDF Anti-Copy Free

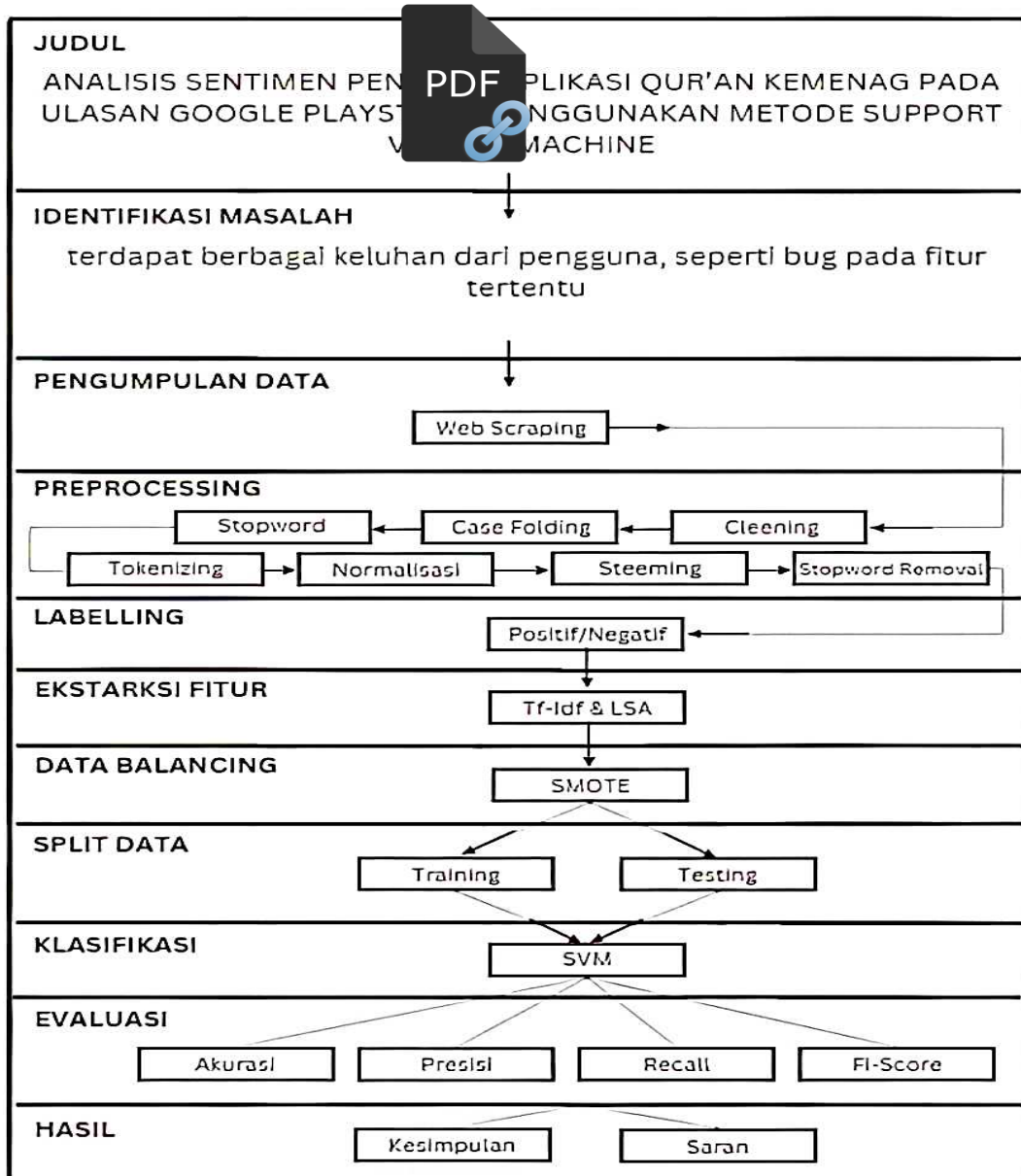
(Upgrade to Pro Version to Remove the Watermark)

No	Judul	Tujuan	Metode	Hasil
	Aplikasi Pembelajaran Online Di Play Store	<i>sentimen</i> analisis	<i>Machine</i> (SVM) & TF-IDF (<i>Term Frequency-Inverse Frequency</i>)	akurasi tertinggi) Zenius: 82,67% Kelas Pintar: 82,00% Edmodo: 74,67% <i>Google Classroom</i> : 64,33% (Tingkat akurasi terendah)
	Pada Masa Pandemi Covid-19 Menggunakan Algoritma <i>Support Vector Machine</i> (SVM)[31]	menggunakan <i>SVM</i> untuk menentukan aplikasi terbaik secara lebih akurat.	<i>Frequency-Inverse Frequency</i>	
10.	Analisis <i>Sentimen</i> Masyarakat Terhadap Kenaikan Harga BBM Di Media Sosial Twitter Menggunakan Metode <i>Support Vector Machine</i> [32].	Tujuan penelitian ini adalah untuk menganalisis <i>sentimen</i> masyarakat terhadap kebijakan kenaikan harga BBM di Indonesia pada September 2022 melalui tweet di media sosial Twitter.	<i>Support Vector Machine</i> (SVM) & <i>Tool CountVectorizer</i> .	Hasil penelitian menunjukkan bahwa metode <i>Support Vector Machine</i> (SVM) mencapai akurasi 76% dalam menganalisis <i>sentimen</i> masyarakat terhadap kenaikan harga BBM.

Protected by PDF Anti-Copy Free

(Upgrade to Pro Version to Remove the Watermark)

2.3 Kerangka Berfikir



Gambar 2. 3 Kerangka Berpikir

Gambar 2.3 menjelaskan bagaimana beberapa faktor dalam ulasan pengguna menjadi kunci untuk menentukan sentimen pengguna terhadap aplikasi Qur'an Kemenag. Faktor-faktor tersebut akan dihitung dan diklasifikasikan menggunakan metode *Support Vector Machine* (SVM) untuk menganalisis kepuasan, keluhan, dan kebutuhan pengguna berdasarkan ulasan di *Google Play Store*.



3.1 Analisa Sistem

3.1.1 Analisa Sistem Yang Berjalan

Sistem yang berjalan saat ini adalah pengelolaan ulasan pengguna aplikasi Qur'an Kemenag yang tersedia melalui platform *Google Playstore*. Ulasan ini mencakup berbagai tanggapan dari pengguna terkait pengalaman mereka menggunakan aplikasi, yang dapat berupa komentar, rating bintang, atau kombinasi keduanya. *Google Playstore* menyediakan dashboard untuk pengelola aplikasi, di mana ulasan pengguna dapat dibaca secara langsung.

Akan tetapi, proses analisis ulasan tersebut dilakukan secara manual tanpa adanya dukungan sistematis untuk mengelompokkan sentimen pengguna. Hal ini membuat proses analisis membutuhkan waktu yang cukup lama, terutama mengingat jumlah ulasan yang terus bertambah seiring dengan waktu. Informasi yang diperoleh dari ulasan juga belum dimanfaatkan secara maksimal untuk mendukung pengambilan keputusan strategis dalam pengembangan aplikasi.

Saat ini, beberapa kelebihan dari sistem yang berjalan adalah ulasan tersedia dalam jumlah besar dan mencerminkan pengalaman pengguna secara langsung. Informasi ini memberikan gambaran awal terkait persepsi pengguna terhadap aplikasi. Selain itu, dashboard *Google Playstore* memungkinkan pengelola aplikasi mengakses ulasan secara cepat.

Meskipun demikian, terdapat beberapa kekurangan yang signifikan pada sistem saat ini. Proses manual dalam membaca dan memahami ulasan membutuhkan waktu yang lama dan rentan terhadap bias subjektif. Tidak adanya sistem analisis sentimen menyebabkan kesulitan dalam mengidentifikasi pola atau tren sentimen pengguna, baik itu positif, negatif, maupun netral. Selain itu, pengelolaan data ulasan dalam jumlah besar menjadi tantangan tersendiri, terutama ketika diperlukan penilaian menyeluruh untuk perbaikan aplikasi.

Akibat dari kekurangan tersebut, pengelola aplikasi kehilangan kesempatan untuk secara proaktif memahami kebutuhan pengguna dan meningkatkan kualitas

Protected by PDF Anti-Copy Free

(Upgrade to Pro Version to Remove the Watermark)

aplikasi sesuai dengan masukan yang diberikan. Ulasan negatif yang tidak tertangani juga berpotensi menorehkan reputasi aplikasi, terutama di mata calon pengguna baru yang menjadikan ulasan sebagai salah satu pertimbangan utama dalam mengunduh aplikasi.



Dengan sistem yang berjalan seperti ini, diperlukan solusi yang mampu melakukan analisis ulasan secara otomatis dan sistematis. Solusi tersebut diharapkan dapat mengelompokkan ulasan berdasarkan *sentimen*, sehingga pengelola dapat dengan mudah mengetahui tingkat kepuasan pengguna dan mengidentifikasi aspek-aspek aplikasi yang memerlukan perbaikan atau pengembangan lebih lanjut.

3.1.2 Alternatif Pemecahan Masalah

Berdasarkan permasalahan yang diidentifikasi pada analisis sistem yang berjalan, terdapat kebutuhan untuk mengotomatisasi proses analisis sentimen ulasan pengguna aplikasi Qur'an Kemenag di *Google Playstore*. Permasalahan utama yang dihadapi adalah waktu yang lama untuk melakukan analisis manual, bias subjektif, dan keterbatasan dalam memanfaatkan data secara efektif untuk peningkatan kualitas aplikasi. Untuk mengatasi permasalahan tersebut, berikut adalah beberapa alternatif pemecahan masalah yang dapat diterapkan:

a. Menggunakan Analisis Manual

Alternatif pertama adalah membaca dan mengategorikan sentimen setiap ulasan secara manual. Metode ini tidak memerlukan alat atau teknologi khusus. Namun, pendekatan ini sangat memakan waktu, rentan terhadap bias subjektif, dan sulit diterapkan untuk jumlah ulasan yang besar.

b. Menggunakan Algoritma Machine Learning Lainnya

Beberapa algoritma seperti Naive Bayes atau Random Forest dapat digunakan untuk menganalisis sentimen. Alternatif ini memungkinkan pengolahan data yang lebih efisien dibandingkan metode manual. Namun, algoritma lain sering kali kurang optimal dalam menangani data teks dibandingkan *Support Vector Machine (SVM)*.

Protected by PDF Anti-Copy Free

(Upgrade to Pro Version to Remove the Watermark)

c. Menggunakan Metode *Support Vector Machine* (SVM)

Metode ini menawarkan akurasi tinggi untuk analisis sentimen. SVM dikenal mampu menangani data dengan baik, memiliki kemampuan generalisasi yang kuat, serta efisien dalam mengklasifikasikan ulasan positif dan negatif. Selain itu, SVM relatif efisien dalam penggunaan sumber daya dibandingkan metode yang lebih kompleks seperti *deep learning*.

d. Menggunakan Layanan Cloud untuk Analisis Sentimen

Alternatif ini mencakup pemanfaatan *platform* berbasis *cloud* seperti *Google Cloud Natural Language* atau IBM Watson. Solusi ini memungkinkan analisis cepat tanpa pengembangan model dari awal. Namun, biaya yang dibutuhkan cukup tinggi, dan ada keterbatasan dalam kustomisasi model.

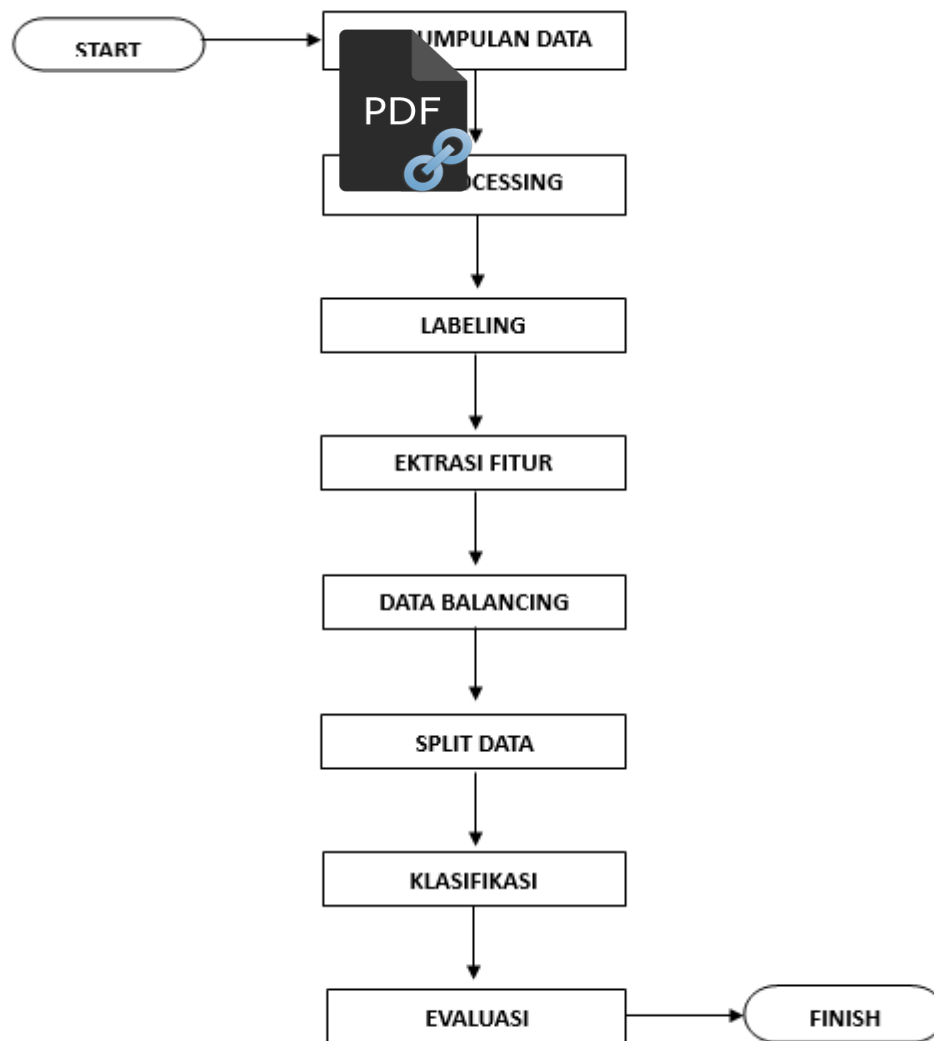
Setelah mempertimbangkan alternatif yang ada, penelitian ini memilih metode *Support Vector Machine* (SVM) sebagai solusi terbaik. Pemilihan ini didasarkan pada kemampuan SVM dalam memberikan akurasi tinggi pada klasifikasi teks, efisiensi waktu dalam pengolahan data ulasan, dan kemudahan implementasi dibandingkan metode yang lebih kompleks. Dengan demikian, metode ini diharapkan mampu menghasilkan analisis sentimen yang akurat dan mendukung pengambilan keputusan berbasis data untuk pengembangan aplikasi Qur'an Kemenag.

3.1.3 Metode Analisa

Metode analisis yang digunakan dalam penelitian ini adalah analisis sentimen pada ulasan pengguna aplikasi qur'an kemenag yang tersedia di *platform google play store*. penelitian ini mengadopsi pendekatan *machine learning* untuk mengidentifikasi dan mengklasifikasikan sentimen ulasan pengguna. tahapan proses analisis digambarkan secara rinci dalam alur pada gambar 3.1 sebagai berikut

Protected by PDF Anti-Copy Free

(Upgrade to Pro Version to Remove the Watermark)



Gambar 3. 1 Diagram Pengolahan Data

a. Pengumpulan Data

Tahap awal pengolahan data dimulai dengan pengumpulan data berupa ulasan pengguna aplikasi Qur'an Kemenag di *Google Play Store*. Pengumpulan data dilakukan menggunakan teknik *web scraping* dengan pustaka seperti *BeautifulSoup* atau *Selenium*. *Web scraping* adalah proses pengumpulan data web terstruktur secara otomatis yang dilakukan oleh aplikasi khusus atau kode pemrograman. Data yang terkumpul bisa berasal dari ribuan, jutaan bahkan miliaran data di dunia maya[33]. Data yang diambil meliputi teks ulasan, skor bintang, dan metadata lainnya yang relevan. Setelah data terkumpul, dilakukan pra-pemrosesan data (data

Protected by PDF Anti-Copy Free

(Upgrade to Pro Version to Remove the Watermark)

preprocessing) untuk membersihkan dan menyusun data agar siap untuk analisis lebih lanjut.

b. *Preprocessing*



Setelah data ulasan per hasil dikumpulkan, langkah berikutnya adalah melakukan *preprocessing*. *Preprocessing* adalah tahapan dimana aplikasi melakukan seleksi data yang akan diproses pada setiap dokumen[34]. Pengambilan data pada penelitian ini dilakukan melalui proses *web scraping* yang mengambil ulasan dari aplikasi Qur'an Kemenag di *Google Play Store*. Data mentah yang diperoleh kemudian disimpan dalam format *Comma Separated Values* (CSV) agar mudah diolah lebih lanjut. Untuk klasifikasi, diperlukan serangkaian proses awal untuk membersihkan data dari kata-kata yang tidak relevan. Proses ini dikenal dengan istilah *preprocessing*, yang mencakup tahapan-tahapan berikut:

- 1) *Cleaning* adalah proses menghilangkan semua karakter di dalam tweet yang tidak termasuk alfabet, agar dapat mengurangi karakter atau simbol yang tidak dikehendaki atau tidak memiliki makna dalam analisis sentimen[35].
- 2) *Case folding* adalah tahap untuk mengkonversi karakter alfabet yang telah melalui tahap cleansing ke huruf kecil (*lower case*)[36].
- 3) *Stopword* adalah mengubah kata yang tidak penting, dalam hal ini kata-kata yang tidak mengandung sentimen dengan contoh kata hubung[37].
- 4) *Tokenizing* adalah tahap pemotongan string input berdasarkan tiap kata yang menyusunnya. Karakter selain huruf akan dianggap delimiter dan akan dihilangkan atau dihapus untuk proses mendapat kata-kata penyusun teks[38].
- 5) *Normalisasi* proses yang digunakan untuk mengubah kata-kata tidak baku atau singkatan dalam ulasan aplikasi menjadi bentuk yang lebih baku dan mudah dipahami[39].
- 6) *Stemming* merupakan salah satu proses dari mengubah token yang berimbuhan menjadi kata dasar, dengan menghilangkan semua imbuhan yang ada pada token tersebut. Pentingnya *stemming* dalam proses pembuatan sistem adalah untuk menghilangkan imbuhan pada awalan dan akhiran[40].
- 7) *Stopword Removal* adalah penghapusan kata sambung, kata depan dan juga kata ganti. Proses ini dilakukan dengan mencari kata pada data dalam kamus

Protected by PDF Anti-Copy Free

(Upgrade to Pro Version to Remove the Watermark)

stopword, jika kata tersebut ditemukan pada kamus maka akan dihilangkan[41].

c. *Labelling*



Setelah data siap, langkah selanjutnya adalah memberikan label, *Labelling* atau pelabelan adalah tahap yang penting dalam analisis sentimen yang mencakup pemberian label atau kategori tertentu pada data ulasan untuk mengindikasikan sentimen yang terkandung dalam ulasan tersebut[42]. Ulasan dengan bintang 4-5 diberi label positif, sedangkan ulasan dengan bintang 1-2 diberi label negatif. Ulasan dengan rating bintang 3 tidak digunakan dalam analisis ini dan akan dihapus, karena rating tersebut dianggap netral dan tidak memberikan informasi yang jelas terkait sentimen positif atau negatif. Proses pelabelan ini dilakukan secara otomatis menggunakan skrip *Python*, dan validasi manual dilakukan pada sebagian data untuk memastikan kualitas serta akurasi pelabelan yang dihasilkan.

d. *Ekstrasi Fitur*

Ekstraksi fitur adalah metode untuk mengidentifikasi fitur yang relevan dengan klasifikasi[43]. Pada penelitian ini, dua metode ekstraksi fitur yang digunakan adalah TF-IDF (*Term Frequency-Inverse Document Frequency*) dan LSA (*Latent Semantic Analysis*). Metode pertama, TF-IDF, digunakan untuk mengubah data teks menjadi representasi numerik berdasarkan frekuensi kata dalam dokumen dan invers frekuensi kata dalam seluruh korpus. TF-IDF memberi bobot lebih tinggi pada kata-kata yang sering muncul dalam dokumen, tetapi jarang ditemukan di seluruh korpus, sehingga memungkinkan algoritma untuk fokus pada kata-kata yang relevan. Proses ini menghasilkan matriks kata-dokumen yang siap digunakan dalam tahap klasifikasi.

Selain itu, LSA juga diterapkan sebagai metode ekstraksi fitur tambahan. LSA mengurangi dimensi matriks kata-dokumen yang dihasilkan oleh TF-IDF menggunakan teknik SVD (*Singular Value Decomposition*). Dengan mereduksi dimensi, LSA mengungkapkan struktur tersembunyi dalam data teks dan menangani masalah sparsity (data yang jarang) yang sering muncul dalam teks. LSA mengidentifikasi hubungan tersembunyi antara kata-kata yang sering muncul dalam konteks yang sama, sehingga dapat meningkatkan representasi fitur dengan

Protected by PDF Anti-Copy Free

(Upgrade to Pro Version to Remove the Watermark)

mengurangi noise. Matriks fitur hasil dari LSA selanjutnya digunakan sebagai input untuk algoritma klasifikasi. Kedua metode ini, TF-IDF dan LSA, saling melengkapi untuk menghasilkan representasi yang lebih efisien dan efektif dalam klasifikasi sentimen.

e. *Data Balancing*

Pada proses analisis sentimen ulasan aplikasi Qur'an Kemenag, salah satu tantangan utama yang sering dihadapi adalah ketidakseimbangan jumlah data antara kelas sentimen positif dan negatif. Ketidakseimbangan data (*imbalanced data*) dapat memengaruhi performa model pembelajaran mesin, terutama jika jumlah data pada salah satu kelas jauh lebih besar dibandingkan kelas lainnya. Oleh karena itu, diperlukan proses data balancing untuk memastikan bahwa distribusi data antar kelas menjadi lebih seimbang. Proses ini merupakan pengecekan terhadap data yang akan digunakan, apakah data tersebut sudah balance atau belum, karena berdasarkan penelitian sebelumnya hal tersebut mempengaruhi terhadap proses data mining yang akan dilakukan[44].

Dalam penelitian ini, metode SVM SMOTE (*Support Vector Machine Synthetic Minority Oversampling Technique*) digunakan untuk menangani ketidakseimbangan data. Teknik ini mensintesis sampel baru dari kelas minoritas untuk menyeimbangkan kumpulan data dengan membuat instance baru dari kelas minoritas dengan membentuk kombinasi *konveks* dari *instance* tetangga. Kemudian, level oversampling yang dibutuhkan dipilih secara acak. Menarik garis secara efektif antara titik-titik minoritas di ruang fitur dan sampel di sepanjang garis ini. Contoh minoritas baru ini ditambahkan ke data pelatihan, dan pengklasifikasi dilatih dengan data tambahan[45]. Jika dilihat dari hasil perbandingan tersebut, nampak terlihat bahwa dataset yang digunakan tidak pada kondisi balance (seimbang). Setelah proses balancing selesai, data yang telah diresample ini kemudian digunakan dalam tahap pembagian data (*split data*) dan klasifikasi. Dengan adanya proses balancing ini, model yang dihasilkan diharapkan memiliki performa yang lebih baik dalam mengklasifikasikan data dari kelas minoritas.

Protected by PDF Anti-Copy Free

(Upgrade to Pro Version to Remove the Watermark)

f. *Split Data*

Split data adalah cara membagi data set agar menjadi dua bagian terpisah yaitu subset utama yang digunakan untuk machine learning dan analisis data. Data training adalah subset dari data yang digunakan untuk melatih model machine learning. Pada tahap ini, model belajar pola dan relasi dalam data untuk memahami dan memprediksi hasil tertentu. Tujuan dari data training digunakan untuk mengejar model. Data Testing adalah data yang tidak digunakan selama proses pelatihan. Setelah model selesai dilatih dengan data training, data testing digunakan untuk menguji sejauh mana model dapat melakukan prediksi yang akurat pada data yang belum pernah dilihat sebelumnya. Dalam penelitian ini, pembagian dataset mengikuti rasio 80/20, di mana 80% digunakan untuk data training dan 20% untuk data testing[46].

Proses ini bertujuan memastikan model dapat diuji menggunakan data yang belum pernah dilihat sebelumnya, sehingga performanya dapat dievaluasi secara objektif. Tahap ini juga penting untuk mencegah overfitting, yaitu kondisi di mana model terlalu menyesuaikan diri dengan data pelatihan sehingga tidak dapat bekerja dengan baik pada data baru. Dengan pembagian data yang tepat, model diharapkan mampu menghasilkan performa yang optimal pada data pengujian dan memiliki kemampuan generalisasi yang baik dalam menghadapi data baru.

g. **Klasifikasi**

Tahap selanjutnya adalah klasifikasi, klasifikasi adalah proses pengelompokan objek yang memiliki karakteristik atau ciri yang sama ke dalam beberapa kelas[47]. Pada penelitian ini penulis memilih metode SVM (*Support Vector Machine*). Proses dimulai dengan ekstraksi fitur menggunakan metode TF-IDF (*Term Frequency-Inverse Document Frequency*), yang mengubah data teks menjadi representasi numerik untuk memudahkan analisis. Data kemudian dibagi menjadi dua bagian, yaitu *training set* dan *testing set* dengan perbandingan 80:20. Model SVM dilatih menggunakan *training set* untuk mempelajari pola sentimen dalam data, dan selanjutnya diuji menggunakan *testing set* untuk mengevaluasi kinerjanya dalam melakukan klasifikasi sentimen secara akurat.

Protected by PDF Anti-Copy Free

(Upgrade to Pro Version to Remove the Watermark)

h. Evaluasi

Tahap akhir adalah evaluasi model, evaluasi model kesenjangan (*discrepancy model*) menurut (dalam Fernandes, 1984) adalah untuk mengetahui tingkat kesesuaian baku (*standard*) yang sudah ditentukan dalam program dengan kinerja (*performance*) sesungguhnya dari program tersebut[48]. Evaluasi dilakukan dengan mengukur performa algoritma menggunakan *Confusion Matrix*, yang mencakup beberapa metrik utama, yaitu akurasi, presisi, recall, dan *F1-score*. Hasil dari evaluasi ini digunakan untuk menentukan efektivitas model dalam mengklasifikasikan sentimen positif dan negatif serta memberikan analisis mengenai kekuatan dan kelemahan model yang dikembangkan.

3.2 Teknik Pemilihan Informan (Populasi, sampel dan sampling)

3.2.1 Teknik Pengumpulan Data

Untuk mendapatkan data yang akurat, relevan, dan valid, penelitian ini mengumpulkan data dengan cara berikut:

a. Data Primer

Data primer diperoleh langsung dari sumber utama melalui metode berikut:

1) Teknik *Web Scraping*

Penelitian ini menggunakan teknik *web scraping* untuk mengumpulkan data ulasan pengguna aplikasi Qur'an Kemenag yang terdapat pada platform *Google Play Store*. Teknik ini diawali dengan identifikasi sumber data, yaitu halaman ulasan aplikasi Qur'an Kemenag di *Google Play Store*, di mana ulasan yang diambil mencakup teks komentar, nilai rating, dan metadata lain yang relevan. Peneliti menggunakan alat *web scraping* seperti pustaka *Python*, misalnya *Selenium* atau *BeautifulSoup*, yang memiliki kemampuan untuk menangani struktur HTML secara fleksibel. Proses pengambilan data dilakukan dengan mengakses halaman ulasan, menavigasi melalui beberapa halaman jika diperlukan, dan mengekstrak informasi yang sesuai untuk disimpan dalam format terstruktur seperti CSV atau *database*.

Setelah data berhasil diambil, dilakukan proses pembersihan untuk memastikan kualitas data yang akan dianalisis. Langkah pembersihan ini meliputi

Protected by PDF Anti-Copy Free

(Upgrade to Pro Version to Remove the Watermark)

penghapusan duplikasi, penghilangan nilai kosong, serta penyesuaian format data. Data yang telah bersih kemudian dipecah menjadi dua kelompok, yaitu data training sebesar 80% dan data testing sebesar 20%, dengan tetap mempertahankan distribusi sentimen yang seimbang. Teknik *Text Wrapping* yang digunakan dalam penelitian ini dirancang untuk efisiensi dan akurasi, sehingga data yang diperoleh dapat mewakili sentimen pengguna aplikasi Qur'an Kemenag secara keseluruhan.

b. Data Sekunder

Data sekunder yang digunakan dalam penelitian ini meliputi berbagai referensi teoritis dan pendukung, seperti jurnal yang relevan dengan analisis sentimen serta penerapan algoritma *Support Vector Machine (SVM)*. Dalam penelitian ini, peneliti mengutip sebanyak 51 referensi untuk mendukung analisis dan pengembangan penelitian.

3.2.2 Teknik Analisa Data

Teknik analisis data yang digunakan dalam penelitian ini berfokus pada ekstraksi fitur dari teks ulasan untuk mendukung proses klasifikasi sentimen. Beberapa teknik utama yang digunakan adalah *Term Frequency-Inverse Document Frequency (TF-IDF)* dan *Latent Semantic Analysis (LSA)*. Berikut adalah penjelasan tentang teknik-teknik tersebut:

a. *Term Frequency-Inverse Document Frequency (TF-IDF)*

TF-IDF adalah teknik yang digunakan untuk menilai pentingnya kata dalam suatu dokumen relatif terhadap koleksi dokumen lainnya. TF-IDF mengukur seberapa sering kata muncul dalam dokumen (*Term Frequency*) dan seberapa jarang kata tersebut muncul dalam seluruh koleksi dokumen (*Inverse Document Frequency*). Metode TF-IDF merupakan metode untuk menghitung bobot setiap kata yang paling umum digunakan pada *information retrieval*. Metode ini juga terkenal efisien, mudah dan memiliki hasil yang akurat[49]. Kata-kata yang memiliki nilai TF-IDF tinggi dianggap lebih penting dalam konteks analisis sentimen.

Protected by PDF Anti-Copy Free

(Upgrade to Pro Version to Remove the Watermark)

b. *Latent Semantic Analysis (LSA)*

LSA adalah teknik untuk mengurangi dimensi data dengan mengungkapkan hubungan tersembunyi antara kata dalam dokumen. *Latent Semantic Analysis (LSA)* adalah teori dan metode yang berkaitan pada himpunan dokumen yang banyak dan terstruktur untuk mengekstraksi dan mewakili penggunaan arti kata dengan perhitungan statistik dan aljabar linier[50]. LSA menggunakan metode aljabar linier untuk mengidentifikasi pola-pola semantik dalam teks dan mengurangi redundansi data. Teknik ini memungkinkan untuk menghasilkan representasi fitur yang lebih ringkas dan bermakna, yang selanjutnya digunakan untuk analisis sentimen.

c. *Synthetic Minority Oversampling Technique (SMOTE)*

SMOTE adalah salah satu teknik oversampling paling populer dan pendekatan pertama yang mencoba menyeimbangkan dataset dengan menghasilkan instance minoritas sintetis[51]. SMOTE bekerja dengan menghasilkan data sintetis baru berdasarkan sampel yang ada dari kelas minoritas tanpa melakukan duplikasi, sehingga dapat menghindari masalah overfitting. Proses SMOTE dimulai dengan memilih setiap sampel dari kelas minoritas sebagai sampel dasar untuk sintesis sampel tambahan. Sampel baru ini kemudian dihasilkan dengan cara interpolasi linier antara sampel dasar dan sampel tetangga terdekat, dan proses ini diulang hingga menghasilkan jumlah sampel yang diinginkan. SMOTE secara efektif mengatasi masalah overfitting yang biasa terjadi pada teknik oversampling acak yang hanya mereplikasi data secara identik. Berbagai pengembangan teknik SMOTE telah banyak dieksplorasi untuk meningkatkan kinerjanya dalam menangani ketidakseimbangan data.

Dalam penelitian ini, peneliti hanya menggunakan teknik *Support Vector Machine SMOTE (SVM-SMOTE)* untuk menangani ketidakseimbangan data pada dataset analisis sentimen. Teknik SVM-SMOTE menggabungkan metode interpolasi SMOTE dengan algoritma *Support Vector Machine (SVM)* untuk menghasilkan sampel sintetis dari kelas minoritas dan memperbaiki keseimbangan data. Teknik ini diterapkan untuk meningkatkan akurasi model dalam mengklasifikasikan sentimen pada dataset yang memiliki distribusi kelas yang tidak seimbang.

Protected by PDF Anti-Copy Free

(Upgrade to Pro Version to Remove the Watermark)

Dataset yang telah diseimbangkan menggunakan SMOTE kemudian digunakan sebagai input untuk klasifikasi dengan algoritma *Support Vector Machine* (SVM). Dengan data lebih seimbang, SVM diharapkan dapat menghasilkan hyperplane yang optimal untuk memisahkan kelas sentimen, sehingga meningkatkan akurasi model pada data *testing*, terutama untuk kelas minoritas.

d. *Support Vector Machine* (SVM)

Setelah fitur diekstraksi menggunakan TF-IDF dan LSA, model SVM (*Support Vector Machine*) digunakan untuk melakukan klasifikasi sentimen ulasan. SVM adalah algoritma *machine learning* yang mencari hyperplane yang memisahkan dua kelas data dengan margin terbesar. Dalam penelitian ini, SVM digunakan untuk memisahkan sentimen positif dan negatif pada ulasan pengguna aplikasi Qur'an Kemenag.

Tahap selanjutnya adalah pengklasifikasian sentimen menggunakan algoritma *Support Vector Machine* (SVM). Model SVM dilatih menggunakan data latih yang telah dilabeli sebagai sentimen positif atau negatif. Kinerja algoritma SVM diukur menggunakan *Confusion Matrix*, yang mencakup beberapa metrik utama, yaitu:

Tabel 1. *Confusion Matrix*

	<i>Actual Values</i>	
	1 (<i>Positive</i>)	0 (<i>Negative</i>)
Prediksi	<i>TP</i> (<i>True Positive</i>)	<i>FP</i> (<i>False Positive</i>)
	<i>FN</i> (<i>False Negative</i>)	<i>TN</i> (<i>True Negative</i>)

Protected by PDF Anti-Copy Free

(Upgrade to Pro Version to Remove the Watermark)

True Positives (TP): Ini adalah contoh yang diklasifikasikan dengan benar sebagai positif.

False Positive (FP): Ini adalah contoh yang salah diklasifikasikan sebagai positif.

True Negatives (TN): Ini adalah contoh yang diklasifikasikan dengan benar sebagai negatif.

False Negative (FN): Ini adalah contoh yang salah diklasifikasikan sebagai negatif.

Yang kemudian akan digunakan untuk menghitung berbagai metrik lainnya guna mengukur performa dalam analisis sentimen pengguna aplikasi qur'an kemenag pada ulasan *google playstore*, seperti :

1) Accuracy

Accuracy adalah ukuran seberapa baik model memprediksi kelas sebenarnya dari data uji. *Akurasi* dihitung dengan membagi jumlah prediksi yang benar dengan total jumlah prediksi.

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN} \dots\dots\dots(1)$$

2) Precision

Precision adalah ukuran seberapa baik model memprediksi kelas positif dengan benar. *Presisi* dihitung dengan membagi jumlah prediksi positif yang benar dengan total jumlah prediksi positif.

$$Precision = \frac{TP}{TP+FP} \dots\dots\dots(2)$$

3) Recall

Recall adalah ukuran seberapa baik model mendeteksi kelas positif. *Recall* dihitung dengan membagi jumlah prediksi positif yang benar dengan total jumlah kelas positif.

$$Recall = \frac{TP}{TP+FN} \dots\dots\dots(3)$$

4) F-1 Score

F-1 score adalah rata-rata harmonik dari presisi dan recall. Ini memberikan keseimbangan antara kedua metrik ini. *F-1 score* berguna ketika kelas memiliki

Protected by PDF Anti-Copy Free

(Upgrade to Pro Version to Remove the Watermark)

distribusi yang tidak seimbang. Secara matematis, F-1 score dihitung dengan rumus.

$$F-1 \text{ Score} = \frac{2 \times \text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}} \dots\dots\dots (4)$$



5) Validasi K-Fold

Validasi K-Fold adalah teknik validasi yang membagi dataset menjadi k bagian atau lipatan (*folds*) yang sama besar. Dalam setiap iterasi, satu lipatan digunakan sebagai data validasi, sementara lipatan lainnya digunakan sebagai data pelatihan. Proses ini diulang hingga semua lipatan berperan sebagai data validasi sekali, dan rata-rata performa model dari semua iterasi digunakan untuk mengevaluasi akurasi secara keseluruhan. Dalam analisis sentimen pada ulasan pengguna Aplikasi Quran Kemenag, validasi *K-Fold* membantu memastikan bahwa model *SVM (Support Vector Machine)* yang digunakan tidak mengalami masalah *overfitting* atau *underfitting*.

Overfitting terjadi ketika model terlalu "menghafal" data pelatihan sehingga performanya sangat baik pada data pelatihan tetapi buruk pada data validasi atau data baru. Hal ini dapat dihindari dengan menggunakan validasi *K-Fold*, karena model diuji pada berbagai subset data.

Underfitting di sisi lain, terjadi ketika model tidak mampu mempelajari pola yang cukup dari data pelatihan, sehingga tidak memberikan hasil yang akurat baik pada data pelatihan maupun validasi. Dengan validasi *K-Fold*, performa model dapat dievaluasi secara menyeluruh pada berbagai subset data untuk memastikan model cukup kompleks menangkap pola dari ulasan pengguna.

3.3 Tempat dan Waktu Penelitian

3.3.1 Tempat Penelitian

Penelitian dilakukan di laboratorium komputer universitas bina insan. Dengan memanfaatkan data yang telah diambil dari *scraping web* pengguna aplikasi *qur'an kemenag* pada ulasan *google playstore*.

Protected by PDF Anti-Copy Free

(Upgrade to Pro Version to Remove the Watermark)

3.3.2 Waktu Penelitian

Estimasi pelaksanaan penelitian akan dimulai pada bulan September 2024, seperti yang terperinci dalam Tabel 3.1. Rencana ini mencakup jadwal kerja yang telah disusun untuk proyek penelitian yang diharapkan dimulai pada tanggal tersebut.

Tabel 3. 1 Waktu Penelitian

No	Jenis Kegiatan	Waktu Kegiatan																							
		Sep-24				Okt-24				Nov-24				Des-24				Jan-25				Feb-25			
	Minggu Ke-	1	2	3	4	1	2	3	4	1	2	3	4	1	2	3	4	1	2	3	4	1	2	3	4
1	Pengajuan Judul																								
2	Pengumpulan Data																								
3	Penulisan Proposal																								
4	Bimbingan Proposal																								
5	Ujian Proposal																								
6	Revisi Ujian Proposal																								
7	Pengolahan Dan Pengujian Data																								
8	Bimbingan Skripsi																								
9	Ujian Skripsi																								

3.3.3 Alat dan Bahan

Berikut ini adalah beberapa alat dan bahan yang diperlukan dalam penelitian ini:

Tabel 3. 2 Spesifikasi Perangkat

Perangkat/OS	Spesifikasi
ASUS VIVOBOOK M1403QA	AMD Ryzen 5 5600H with Radeon Graphics 3.30 GHz, 8,00 GB

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

Tabel 3. 3 Spesifikasi Perangkat

Sistem	Tools
Interpreter	JupyterLab

Tabel 3. 4 Bahan

No	Bahan
1.	Dataset <i>Scriping</i> Dari <i>Google Playstore</i>
2.	Jurnal
3.	Kertas A4
4.	Tinta

BAB IV

HASIL PENELITIAN DAN PEMBAHASAN

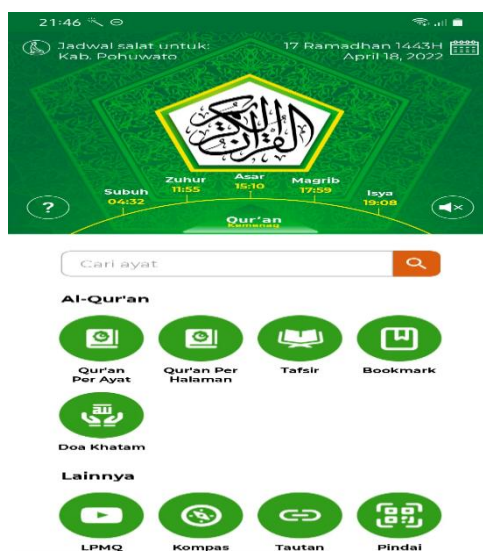
PDF

4.1 Gambaran Umum (Tempat dan Waktu Penelitian)

4.1.1 Gambaran Umum

Penelitian ini dilakukan pada aplikasi Qur'an Kemenag, sebuah aplikasi yang dikembangkan oleh Kementerian Agama Republik Indonesia. Aplikasi ini dirancang untuk memenuhi kebutuhan umat Muslim dalam membaca dan mempelajari Al-Qur'an secara digital. Fitur utama aplikasi meliputi:

- Teks Al-Qur'an yang disertai terjemahan dalam Bahasa Indonesia.
- Audio murattal untuk memudahkan pengguna mendengarkan lantunan ayat-ayat Al-Qur'an.
- Pencarian ayat yang memudahkan pengguna menemukan ayat berdasarkan kata kunci tertentu.
- Penanda bacaan, yang membantu pengguna melanjutkan bacaan Al-Qur'an di waktu berikutnya.



Gambar 4. 1 Aplikasi Qur'an Kemenag

Aplikasi Qur'an Kemenag telah tersedia di *Google Play Store* dengan jumlah unduhan lebih dari 500 ribu kali dan memiliki rating tinggi, yaitu 4.6 dari ratusan ribu ulasan pengguna. Meskipun demikian, ulasan pengguna mencatat

Protected by PDF Anti-Copy Free

(Upgrade to Pro Version to Remove the Watermark)

beberapa keluhan terkait bug pada fitur tertentu, seperti fitur *bookmark* yang tidak bekerja secara optimal, dan masalah lainnya seperti aplikasi yang terkadang *crash*.



4.1.2 Gambaran Organisasi

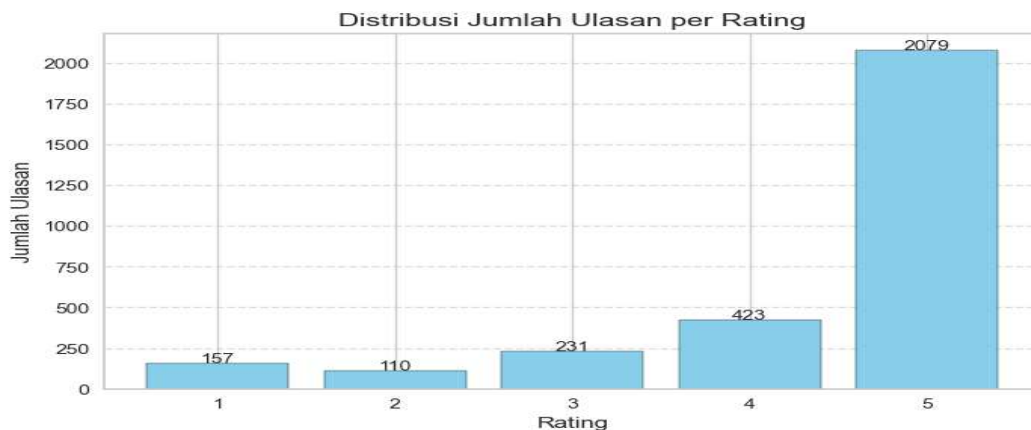
Aplikasi Qur'an Kemenag dikembangkan dan dikelola oleh Kementerian Agama Republik Indonesia melalui Direktorat Jenderal Bimbingan Masyarakat Islam.

- Visi dan Misi: Aplikasi ini mencerminkan visi Kementerian Agama untuk meningkatkan pemahaman, pengamalan, dan penyebaran ajaran Islam di Indonesia.
- Tim Pengembang: Terdiri dari pengembang aplikasi dan tim teknis yang bertanggung jawab atas pembaruan, pemeliharaan, dan penyelesaian masalah teknis.
- Pengelolaan Ulasan: Kementerian Agama menggunakan ulasan pengguna sebagai salah satu cara untuk mengevaluasi dan meningkatkan kualitas aplikasi.

Dengan dukungan pemerintah, aplikasi ini menjadi salah satu solusi teknologi bagi umat Islam di Indonesia untuk mengakses Al-Qur'an secara mudah dan praktis melalui perangkat digital.

4.2 Hasil

4.2.1 Hasil Pengumpulan Data



Gambar 4. 2 Hasil Pengumpulan Data

Protected by PDF Anti-Copy Free

(Upgrade to Pro Version to Remove the Watermark)

Gambar 4.2 menunjukkan distribusi jumlah ulasan pengguna terhadap aplikasi Qur'an Kemenag berdasarkan rating yang diberikan di *Google Play Store*. Penelitian ini berhasil memperoleh sebanyak 3.000 ulasan, yang dikumpulkan mulai dari akhir bulan Agustus 2024 hingga bulan Februari tahun 2025. Data tersebut selanjutnya digunakan untuk proses analisis sentimen. Dari grafik terlihat bahwa sebagian besar pengguna memberikan rating 5 bintang, yaitu sebanyak 2.079 ulasan, yang menunjukkan mayoritas pengguna memiliki kesan positif terhadap aplikasi tersebut. Rating 4 bintang diberikan oleh 423 pengguna, sementara 231 pengguna memberikan 3 bintang. Adapun rating 2 bintang dan 1 bintang masing-masing diberikan oleh 110 dan 157 pengguna. Distribusi ini menunjukkan bahwa meskipun terdapat sejumlah ulasan dengan rating rendah, namun ulasan dengan rating tinggi mendominasi. Hal ini mengindikasikan bahwa secara umum aplikasi Qur'an Kemenag mendapatkan respons yang baik dari para penggunanya.

4.2.2 Hasil *Preprocessing*

content	text_clean	text_case	text_stopward	text_tokens	normalisasi	text_steamindo	text_Stopword
Kok setelah update malah error.. Fungsi bookma...	Kok setelah update malah error Fungsi bookmark...	kok setelah update malah error fungsi bookmark...	update error fungsi bookmark bermasalah downlo...	[update, error, fungsi, bookmark, bermasalah, ...	[update, error, fungsi, bookmark, bermasalah, ...	update error fungsi bookmark masalah download ...	update error fungsi bookmark download resource...
Perbaiki kompas dan adakan fitur GPS dengan di...	Perbaiki kompas dan adakan fitur GPS dengan di...	perbaiki kompas dan adakan fitur gps dengan di...	perbaiki kompas adakan fitur gps disesuaikan j...	[perbaiki, kompas, adakan, fitur, gps, disesuaikan...	[perbaiki, kompas, adakan, fitur, gps, disesuaikan...	baik kompas adakan fitur gps sesuai jam bmkg	kompas adakan fitur gps sesuai jam bmkg
Bagus yang dulu bisa didownload, sekarang harus...	Bagus yang dulu bisa didownload sekarang harus...	bagus yang dulu bisa didownload sekarang harus...	bagus didownload pake data	[bagus, didownload, pake, data]	[bagus, didownload, pakai, data]	bagus didownload pakai data	bagus didownload pakai data
Alhamdulillah, app yg sangat bermanfaat bagi u...	Alhamdulillah app yg sangat bermanfaat bagi um...	alhamdulillah app yg sangat bermanfaat bagi um...	alhamdulillah app yg bermanfaat umat islam yg ...	[alhamdulillah, app, yg, bermanfaat, umat, isl...	[alhamdulillah, aplikasi, yang, bermanfaat, um...	alhamdulillah aplikasi yang manfaat umat islam...	alhamdulillah aplikasi manfaat umat islam lanc...
Mantap nih aplikainya top markotop	Mantap nih aplikainya top markotop	mantap nih aplikainya top markotop	mantap nih aplikainya top markotop	[mantap, nih, aplikainya, top, markotop]	[mantap, nih, aplikainya, top, markotop]	mantap nih aplikainya top markotop	mantap nih aplikainya top markotop
Baik dan mudah dibaca	Baik dan mudah dibaca	baik dan mudah dibaca	mudah dibaca	[mudah, dibaca]	[mudah, dibaca]	mudah baca	mudah baca

Gambar 4. 3 Hasil *Preprocessing*

Gambar 4.3 menunjukkan tahapan preprocessing data teks ulasan pengguna aplikasi Qur'an Kemenag yang dilakukan sebelum masuk ke proses analisis sentimen. Tahapan ini terdiri dari beberapa langkah penting, dimulai dari kolom content yang berisi data ulasan mentah seperti yang diambil dari *Google Play Store*.

Protected by PDF Anti-Copy Free

(Upgrade to Pro Version to Remove the Watermark)

Selanjutnya, dilakukan pembersihan teks pada kolom `text_clean` dengan menghapus karakter-karakter yang tidak diperlukan seperti simbol dan angka. Kemudian, pada kolom `text_clean` teks diubah menjadi huruf kecil untuk menyeragamkan format data. Selanjutnya, dilakukan penghapusan stopword pada kolom `text_stopward`, yang bertujuan untuk menghilangkan kata-kata umum yang tidak memiliki makna signifikan dalam analisis, seperti “dan”, “yang”, “di”, dan sebagainya. Pada kolom `text_tokens`, data diubah menjadi bentuk token atau daftar kata-kata yang telah dipisahkan.

Tahapan berikutnya adalah proses normalisasi, yaitu mengganti kata-kata tidak baku atau singkatan menjadi bentuk baku agar konsisten. Setelah normalisasi, dilakukan proses stemming, yang ditunjukkan pada kolom `text_stemming`, dengan mengubah kata menjadi bentuk dasar menggunakan algoritma stemmer bahasa Indonesia. Terakhir, pada kolom `text_Stopword` ditampilkan hasil akhir setelah penghapusan stopword pasca proses normalisasi dan stemming. Keseluruhan proses preprocessing ini bertujuan untuk meningkatkan kualitas data teks sehingga algoritma machine learning, dalam hal ini *Support Vector Machine* (SVM), dapat mengolah dan menganalisis sentimen secara lebih akurat dan efektif.

4.2.3 Hasil Labelling

Tabel 4.1 Hasil Labelling

No.	Jumlah Ulasan	Kategori Sentimen	Presentase Sentimen
1	2.414	Positif	90,18%
2	263	Negatif	9,82%
Total	2.677		100%

Berdasarkan Tabel 4.1 proses labelling sentimen menghasilkan dua kategori: positif dan negatif. Dataset awal mencakup kurang dari 3.000 ulasan, di mana penilaian berbintang 4 dan 5 dianggap sentimen positif, sementara berbintang 1 dan 2 dianggap negatif. Penilaian berbintang 3 dihapus karena dianggap netral dan tidak mewakili sentimen positif atau negatif. Setelah penyaringan, diperoleh 2.770 ulasan dengan komposisi 90,5% positif dan 9,6% negatif. Setelah pembersihan data, jumlah ulasan yang tersisa adalah 2.677. Dari total tersebut,

Protected by PDF Anti-Copy Free

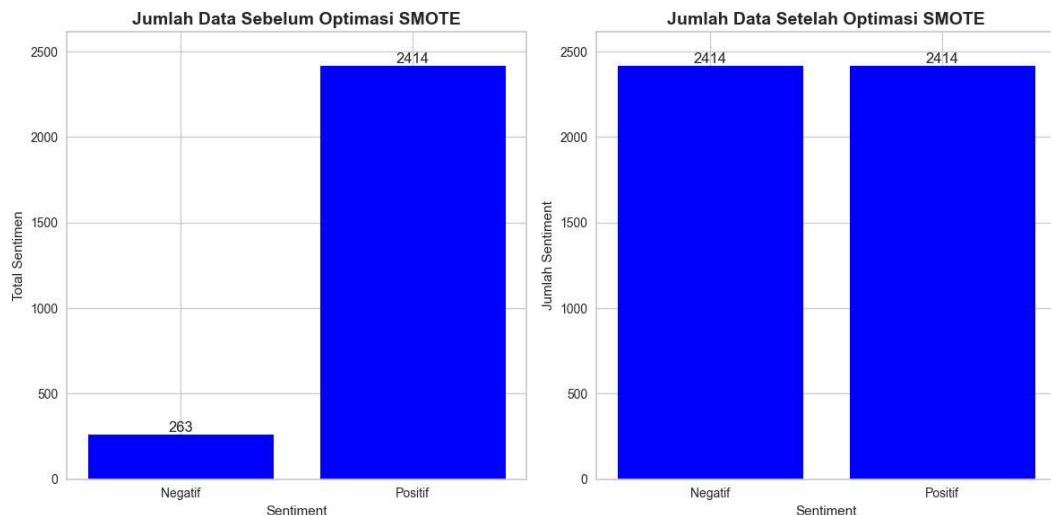
(Upgrade to Pro Version to Remove the Watermark)

2.414 ulasan (90,18%) dikategorikan sebagai sentimen positif dan 263 ulasan (9,82%) sebagai negatif. Proses *data balancing* dilakukan manual untuk memastikan akurasi sebelum digunakan dalam pelatihan model. Ketimpangan jumlah data antara sentimen positif dan negatif menjadi pertimbangan untuk penerapan teknik penyeimbangan data seperti SMOTE.

4.2.4 Hasil Ekstraksi Fitur

Pada tahap ekstraksi fitur, teks ulasan diubah menjadi representasi numerik untuk mendukung proses klasifikasi sentimen. Proses ini dilakukan dalam dua langkah utama, yaitu transformasi dengan TF-IDF dan reduksi dimensi dengan LSA. Transformasi TF-IDF mengubah teks menjadi vektor numerik dengan batasan 1.500 fitur, menghasilkan matriks bobot kata per dokumen. Selanjutnya, reduksi dimensi dilakukan menggunakan LSA melalui dekomposisi SVD dengan pemilihan 500 komponen utama. Hasilnya adalah vektor fitur yang ringkas namun tetap mempertahankan informasi penting, sehingga memberikan fondasi kuat untuk analisis sentimen dalam mengklasifikasikan ulasan ke dalam kategori positif dan negatif.

4.2.5 Hasil Data Balancing



Gambar 4.4 Hasil Data Balancing

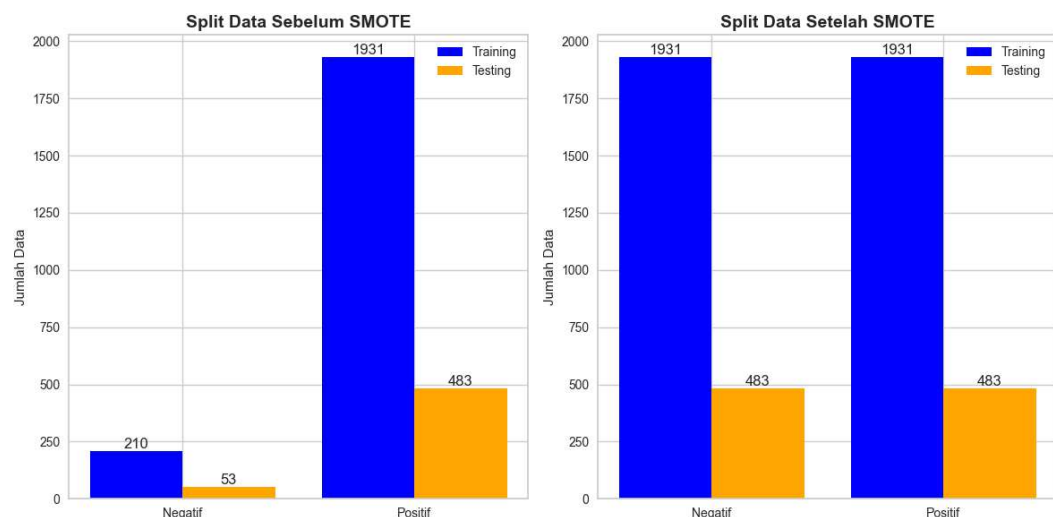
Gambar 4.4 menunjukkan perbandingan jumlah data sentimen sebelum dan sesudah dilakukan optimasi menggunakan metode *Synthetic Minority Over-sampling Technique* (SMOTE). Pada grafik sebelah kiri, terlihat adanya

Protected by PDF Anti-Copy Free

(Upgrade to Pro Version to Remove the Watermark)

ketidakseimbangan jumlah data antara kelas sentimen positif dan negatif, di mana data sentimen positif berjumlah 1.931, sedangkan data sentimen negatif hanya berjumlah 263. Ketidakseimbangan ini dapat menyebabkan bias pada proses pelatihan model, karena algoritma cenderung lebih banyak mempelajari pola dari kelas yang jumlah datanya lebih besar. Untuk mengatasi permasalahan tersebut, dilakukan proses oversampling menggunakan metode SMOTE, sebagaimana ditampilkan pada grafik sebelah kanan. Setelah dilakukan SMOTE, jumlah data pada kedua kelas menjadi seimbang, masing-masing sebanyak 2.414 data untuk kelas positif dan negatif. Dengan distribusi data yang seimbang ini, diharapkan kinerja algoritma *Support Vector Machine* (SVM) dalam mengklasifikasikan sentimen menjadi lebih optimal dan tidak berpihak pada salah satu kelas saja.

4.2.6 Hasil *Split Data*



Gambar 4.5 Hasil *Split Data* Sebelum/Setelah SMOTE

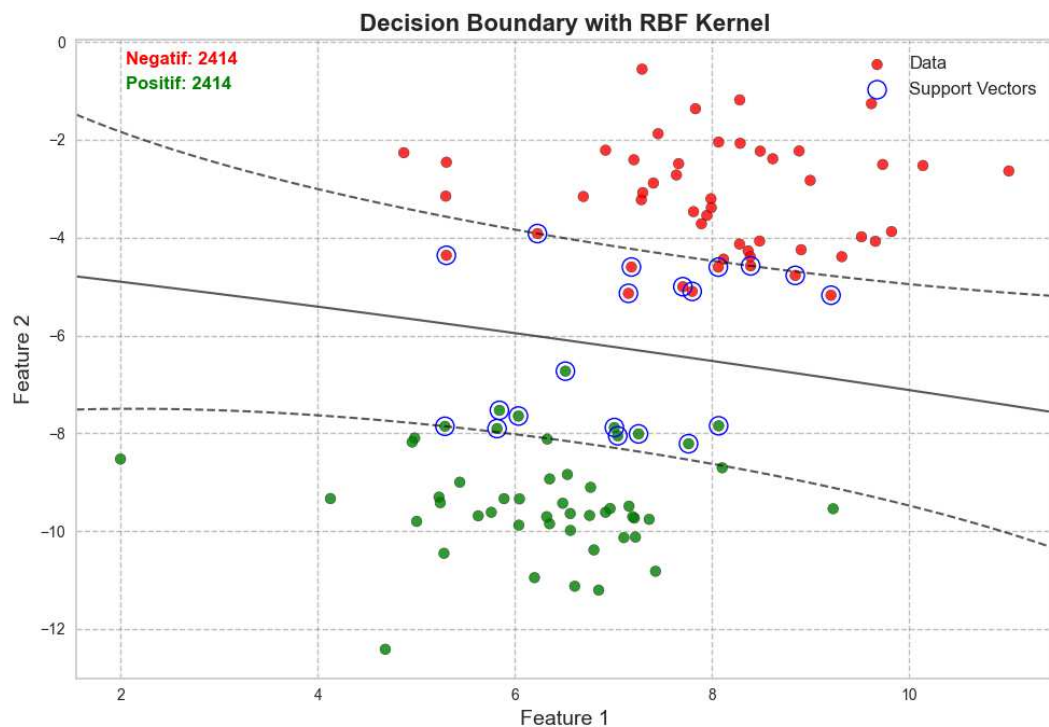
Berdasarkan gambar 4.5 proses pembagian data (*split data*) dilakukan sebelum dan sesudah penerapan teknik SMOTE. Sebelum SMOTE diterapkan, data pelatihan (*training*) dan pengujian (*testing*) terbagi tidak seimbang antara kelas negatif dan positif. Untuk kelas negatif hanya terdapat 210 data pelatihan dan 53 data pengujian, sementara untuk kelas positif terdapat 1.931 data pelatihan dan 483 data pengujian. Ketidakseimbangan ini dapat menyebabkan model cenderung bias terhadap kelas mayoritas. Oleh karena itu, dilakukan penyeimbangan data menggunakan teknik SMOTE (*Synthetic Minority Over-sampling Technique*), yang

Protected by PDF Anti-Copy Free

(Upgrade to Pro Version to Remove the Watermark)

menghasilkan jumlah data yang seimbang pada kedua kelas. Setelah SMOTE diterapkan, baik kelas negatif maupun positif memiliki jumlah data pelatihan yang sama, yaitu 1.931, dan jumlah data uji yang juga sama, yaitu 483. Hal ini bertujuan untuk meningkatkan kinerja model dalam mengenali kedua kelas dengan lebih baik dan adil, terutama pada kelas minoritas.

4.2.7 Hasil Klasifikasi



Gambar 4. 4 Hasil Klasifikasi SVM

Gambar 4.6 menunjukkan *Decision Boundary* hasil klasifikasi menggunakan model *Support Vector Machine* (SVM) dengan kernel *Radial Basis Function* (RBF). Dataset terdiri dari dua kelas, yaitu kelas negatif yang ditandai dengan warna merah dan kelas positif yang ditandai dengan warna hijau. Jumlah data masing-masing kelas adalah 2.414, yang diinformasikan melalui keterangan teks pada sudut kiri atas gambar.

Hyperplane yang dihasilkan oleh model digambarkan sebagai garis solid hitam di tengah. Garis putus-putus di kedua sisi *hyperplane* menunjukkan margin (+1 dan -1), yang merupakan batas jarak terdekat dari data terhadap *hyperplane*.

Protected by PDF Anti-Copy Free

(Upgrade to Pro Version to Remove the Watermark)

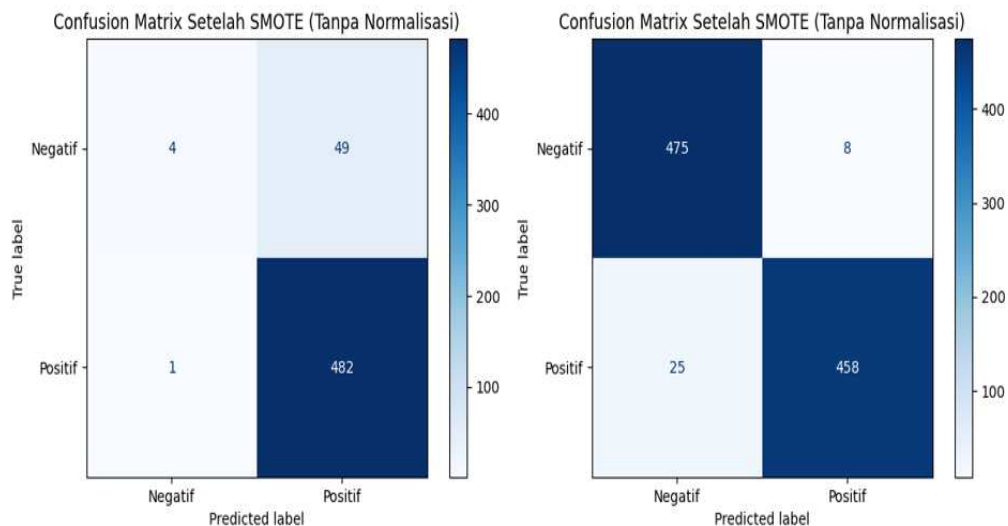
Data yang dikelilingi lingkaran biru disebut *support vectors*, yaitu data yang berada di margin dan memiliki pengaruh besar dalam menentukan posisi *hyperplane*.

Penggunaan kernel RBF memungkinkan SVM untuk menghasilkan *decision boundary* yang non-linear seperti terlihat pada pola garis batas yang melengkung mengikuti distribusi data. Hal ini membuat model SVM dengan kernel RBF efektif dalam memisahkan data yang tidak dapat dipisahkan secara linear di ruang fitur awal.

Secara keseluruhan, gambar ini menunjukkan bahwa model SVM dengan kernel RBF mampu memisahkan dua kelas data secara optimal. Titik-titik data pada margin menjadi titik kritis yang digunakan model untuk mempelajari pola pemisahan, sehingga menghasilkan prediksi yang akurat. Sumbu horizontal merepresentasikan fitur pertama (*Feature 1*), sedangkan sumbu vertikal merepresentasikan fitur kedua (*Feature 2*).

4.2.8 Hasil Evaluasi

a. Confusion Matrix



Gambar 4. 5 Hasil *Confusion Matrix* Sebelum dan Sesudah SMOTE

Evaluasi performa model dilakukan menggunakan *Confusion Matrix*, yang merupakan matriks untuk menggambarkan perbandingan antara prediksi model dengan label asli data uji. *Confusion Matrix* pada penelitian ini ditunjukkan pada Gambar 4.7 yang terdiri dari dua bagian. Gambar sebelah kiri menunjukkan hasil

Protected by PDF Anti-Copy Free

(Upgrade to Pro Version to Remove the Watermark)

setelah dilakukan SMOTE pada model pertama, sedangkan gambar sebelah kanan menunjukkan hasil model kedua setelah SMOTE. Pada *confusion matrix* kiri, terdapat 482 data ulasan Positif yang berhasil diklasifikasikan dengan benar sebagai Positif (*True Positive*), dan hanya 4 ulasan Positif yang salah diklasifikasikan sebagai Negatif (*False Negative*). Namun, model ini menunjukkan kelemahan dalam mengklasifikasikan data Negatif, di mana hanya 4 ulasan Negatif yang berhasil diklasifikasikan dengan benar (*True Negative*), sedangkan 49 ulasan Negatif lainnya salah diklasifikasikan sebagai Positif (*False Positive*).

Sementara itu, pada *confusion matrix* kanan, performa model terlihat lebih seimbang. Terdapat 458 data ulasan Positif yang diklasifikasikan dengan benar sebagai Positif (*True Positive*) dan 475 ulasan Negatif yang juga berhasil diklasifikasikan dengan benar sebagai Negatif (*True Negative*). Kesalahan klasifikasi juga relatif rendah, yaitu 25 ulasan Positif yang diklasifikasikan sebagai Negatif (*False Negative*) dan 8 ulasan Negatif yang diklasifikasikan sebagai Positif (*False Positive*). Berdasarkan hasil tersebut, model kedua menunjukkan performa yang lebih baik secara keseluruhan dalam mengklasifikasikan kedua label data secara seimbang.

b. Classification Report untuk Model SVM

Classification Report Sebelum SMOTE:					Classification Report Setelah SMOTE:				
	precision	recall	f1-score	support		precision	recall	f1-score	support
negatif	0.80	0.08	0.14	53	negatif	0.95	0.98	0.97	483
positif	0.91	1.00	0.95	483	positif	0.98	0.95	0.97	483
accuracy			0.91	536	accuracy			0.97	966
macro avg	0.85	0.54	0.54	536	macro avg	0.97	0.97	0.97	966
weighted avg	0.90	0.91	0.87	536	weighted avg	0.97	0.97	0.97	966

Gambar 4. 6 *Classification Report* Sebelum dan Sesudah SMOTE

Classification Report digunakan untuk memberikan gambaran lengkap mengenai performa model *Support Vector Machine* (SVM) dalam melakukan klasifikasi sentimen ulasan pengguna aplikasi Qur'an Kemenag pada *Google Play Store*. Gambar 4.8 menunjukkan perbandingan hasil klasifikasi sebelum dan sesudah penerapan metode SMOTE (*Synthetic Minority Over-sampling Technique*).

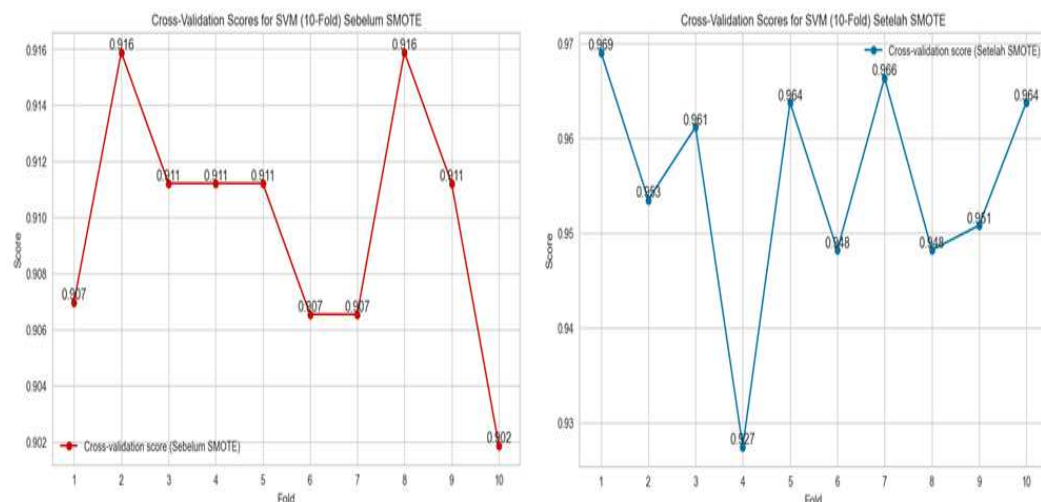
Protected by PDF Anti-Copy Free

(Upgrade to Pro Version to Remove the Watermark)

Sebelum SMOTE diterapkan, model menunjukkan ketidakseimbangan performa antara kelas negatif dan positif. Hal ini terlihat dari nilai *precision* untuk kelas negatif yang hanya sebesar 0.80, dengan *recall* yang sangat rendah yaitu 0.08, dengan *F1-score* hanya 0.14. Sementara itu, *precision* dan *recall* untuk kelas positif masing-masing adalah 0.91 dan 1.00, dengan *F1-score* sebesar 0.95. Nilai akurasi keseluruhan sebelum SMOTE adalah 0.91, namun performa yang timpang antara kedua kelas menunjukkan bahwa model lebih condong mengenali ulasan positif dan mengabaikan yang negatif.

Setelah SMOTE diterapkan, performa model meningkat secara signifikan dan menjadi lebih seimbang dalam mengenali kedua kelas. *Precision* untuk kelas negatif meningkat menjadi 0.95 dan *recall* menjadi 0.98, sedangkan untuk kelas positif, *precision* dan *recall* masing-masing adalah 0.98 dan 0.95. *F1-score* untuk kedua kelas menjadi sama, yaitu sebesar 0.97, menunjukkan bahwa model memiliki performa yang sangat baik dan seimbang. Nilai akurasi keseluruhan pun meningkat menjadi 0.97. Peningkatan ini menunjukkan bahwa teknik SMOTE sangat efektif dalam menangani ketidakseimbangan data, khususnya dalam memperbaiki kemampuan model dalam mengenali kelas minoritas (ulasan negatif). Dengan demikian, model SVM yang telah dioptimalkan menggunakan SMOTE mampu memberikan hasil klasifikasi yang lebih akurat, adil, dan andal dalam menganalisis sentimen ulasan pengguna.

c. Validasi K-Fold



Gambar 4. 7 Validasi K-Fold Sebelum dan Sesudah SMOTE

Protected by PDF Anti-Copy Free

(Upgrade to Pro Version to Remove the Watermark)

Gambar 4.9 menunjukkan hasil evaluasi model *Support Vector Machine* (SVM) menggunakan metode validasi silang (*cross-validation*) dengan 10 fold, baik sebelum maupun sesudah diterapkan metode SMOTE (*Synthetic Minority Over-sampling Technique*). Dalam proses validasi ini, dataset dibagi menjadi 10 bagian (*fold*), di mana masing-masing bagian secara bergantian digunakan sebagai data uji, sementara 9 bagian lainnya digunakan sebagai data latih. Tujuan dari validasi silang ini adalah untuk memastikan kestabilan performa model, menghindari *overfitting*, dan mengevaluasi kemampuan model dalam menghasilkan prediksi yang konsisten. Pada grafik sebelah kiri (sebelum SMOTE), skor validasi bervariasi antara 0.902 hingga 0.916. Skor tertinggi tercatat pada fold ke-2 dan ke-8 sebesar 0.916, sedangkan skor terendah terdapat pada fold ke-10 sebesar 0.902. Selain itu, skor pada fold ke-6 dan ke-7 juga menunjukkan nilai yang lebih rendah yaitu 0.907. Fluktuasi skor ini menandakan bahwa model SVM belum sepenuhnya stabil dalam menangani distribusi data yang tidak seimbang.

Setelah diterapkannya SMOTE (grafik kanan), terlihat peningkatan dan kestabilan performa model. Skor validasi berkisar antara 0.927 hingga 0.969. Skor tertinggi tercatat pada fold ke-1 sebesar 0.969, sementara skor terendah berada pada fold ke-4 sebesar 0.927. Meskipun terjadi sedikit variasi, skor pada sebagian besar fold berada di atas 0.95, yang menunjukkan bahwa model memiliki kemampuan generalisasi yang lebih baik setelah distribusi data diseimbangkan. Secara keseluruhan, hasil validasi K-Fold ini memperlihatkan bahwa penerapan SMOTE berhasil meningkatkan dan menstabilkan performa model SVM dalam memprediksi sentimen ulasan aplikasi Qur'an Kemenag di *Google Play Store*. Model menjadi lebih andal dan tidak terlalu sensitif terhadap variasi data pada setiap fold, sehingga dapat diandalkan untuk melakukan prediksi pada data baru.

Protected by PDF Anti-Copy Free

(Upgrade to Pro Version to Remove the Watermark)

4.2.9 Hasil Analisis *Wordcloud*



Gambar 4. 8 *Wordcloud* Sentimen Negatif dan Positif

Gambar 4.10 menampilkan *wordcloud* dari hasil analisis sentimen ulasan pengguna aplikasi Qur'an Kemenag di *Google Play Store* yang terbagi menjadi dua kategori, yaitu sentimen negatif (kiri) dan sentimen positif (kanan). *Wordcloud* ini menggambarkan kata-kata yang paling sering muncul dalam masing-masing kategori sentimen, sehingga dapat memberikan wawasan mengenai persepsi dan pengalaman pengguna terhadap aplikasi. Pada *wordcloud* sentimen negatif, terlihat bahwa kata-kata seperti "aplikasi", "buka", "halaman", "ayat", dan "tolong" muncul secara dominan. Kata-kata tersebut mengindikasikan adanya keluhan atau masalah yang dirasakan pengguna, seperti kesulitan membuka halaman ayat, tampilan yang tidak sesuai, atau fitur yang tidak berjalan sebagaimana mestinya. Selain itu, muncul juga kata "versi", "offline", dan "error", yang menunjukkan adanya ketidakpuasan terhadap versi tertentu atau fitur offline yang tidak berfungsi optimal.

Sementara itu, pada *wordcloud* sentimen positif, kata-kata yang paling menonjol adalah "bagus", "alhamdulillah", "aplikasi", "terima kasih", dan "bantu". Ini menunjukkan bahwa banyak pengguna merasa puas terhadap manfaat dan kemudahan penggunaan aplikasi ini. Kata-kata seperti "manfaat", "mudah", "baca", dan "lengkap" mencerminkan bahwa aplikasi ini dinilai efektif dalam membantu pengguna membaca dan memahami Al-Qur'an. Kemunculan kata "kemenag" juga menunjukkan adanya apresiasi terhadap Kementerian Agama sebagai pengembang

Protected by PDF Anti-Copy Free

(Upgrade to Pro Version to Remove the Watermark)

aplikasi. Secara keseluruhan, *wordcloud* ini memberikan wawasan penting bagi pengembang dalam mengidentifikasi kekuatan aplikasi yang perlu dipertahankan serta permasalahan yang perlu diperbaiki. Dengan memperhatikan kata-kata yang sering muncul dalam sentimen negatif dan positif, pengembang dapat menyusun strategi pembaruan yang lebih tepat sasaran agar aplikasi semakin baik di mata pengguna.

4.3 Pembahasan

4.3.1 Penerapan Metode Dan Validitas Data

Pada penelitian ini, algoritma *Support Vector Machine* (SVM) dipilih untuk analisis *sentimen* terhadap ulasan pengguna aplikasi Qur'an Kemenag. Algoritma ini dipilih karena memiliki keunggulan dalam menangani data dengan dimensi tinggi dan mampu memberikan hasil yang optimal dalam analisis teks. Sebelum penerapan SVM, data ulasan yang diperoleh melalui *web scraping* menjalani serangkaian proses *preprocessing* untuk meningkatkan kualitas data. Tahapan *preprocessing* meliputi :

- a. *Cleaning*: Menghapus karakter tidak relevan seperti tanda baca, angka, dan simbol khusus.
- b. *Case Folding*: Mengubah seluruh teks menjadi huruf kecil untuk konsistensi.
- c. *Stopword Removal*: Menghilangkan kata-kata umum yang tidak signifikan.
- d. *Tokenizing*: Memecah teks ulasan menjadi token atau kata-kata individual.
- e. *Normalisasi*: Mengubah kata-kata tidak baku menjadi baku menggunakan kamus normalisasi.
- f. *Stemming*: Mengubah kata-kata menjadi bentuk dasarnya.
- g. *Stopword Removal (Akhir)*: Menghapus kembali stopwords yang tersisa setelah tahap stemming.

Setelah *preprocessing*, data teks direpresentasikan ke dalam bentuk vektor menggunakan metode *Term Frequency-Inverse Document Frequency* (TF-IDF). TF-IDF memberikan bobot pada setiap kata berdasarkan frekuensi kemunculannya dalam dokumen dan seluruh korpus, sehingga meningkatkan kemampuan model untuk memahami relevansi kata-kata tertentu terhadap ulasan. Untuk mengatasi

Protected by PDF Anti-Copy Free

(Upgrade to Pro Version to Remove the Watermark)

dimensi tinggi yang dihasilkan oleh TF-IDF, digunakan *Latent Semantic Analysis* (LSA). LSA melakukan reduksi dimensi melalui dekomposisi nilai singular (SVD), yang menghasilkan representasi yang lebih sederhana dan semantik yang tetap terjaga.

Data yang telah diproses kemudian dilabeli menjadi dua kategori sentimen, yaitu positif dan negatif, berdasarkan skor ulasan yang diberikan oleh pengguna. Masalah utama yang ditemukan dalam dataset adalah ketidakseimbangan jumlah data antara sentimen positif dan negatif. Untuk mengatasi hal ini, penelitian ini menggunakan teknik *Synthetic Minority Over-sampling Technique* (SMOTE). SMOTE adalah metode yang secara sintetik menambah data pada kelas minoritas sehingga distribusi data menjadi lebih seimbang. Dengan penerapan teknik ini, model memiliki peluang yang lebih baik untuk mempelajari pola pada kedua kelas sentimen, sehingga mengurangi bias dalam proses pelatihan.

Data ulasan kemudian dipisahkan menjadi data pelatihan (*training set*) dan data pengujian (*testing set*) dengan perbandingan 80:20. Pemisahan ini dilakukan secara acak menggunakan metode *stratified sampling* untuk memastikan distribusi sentimen dalam kedua set tetap proporsional. Selain itu, validitas label diuji untuk memastikan bahwa tidak ada kesalahan dalam pelabelan data yang dapat memengaruhi performa model.

Model SVM dilatih menggunakan kernel *Radial Basis Function* (RBF). Pemilihan kernel ini didasarkan pada kemampuan RBF untuk menangani data yang tidak dapat dipisahkan secara *linear*. Kernel RBF memetakan data ke dalam dimensi yang lebih tinggi sehingga memungkinkan pemisahan yang lebih baik antara kelas sentimen positif dan negatif. *Hyperparameter model*, seperti parameter C dan γ , dioptimalkan menggunakan *Grid Search* untuk memastikan performa model berada pada tingkat terbaik.

Validitas model juga diuji dengan menggunakan *k-fold cross-validation* dengan nilai $k=10$. Teknik ini digunakan untuk menghindari overfitting dan memastikan generalisasi model terhadap data yang belum terlihat. Hasil validasi menunjukkan bahwa model memiliki stabilitas performa pada setiap iterasi, yang

Protected by PDF Anti-Copy Free

(Upgrade to Pro Version to Remove the Watermark)

mengindikasikan bahwa model yang dibangun cukup andal untuk analisis sentimen ini.



4.3.2 Pengujian Hasil Analisis



Pengujian dilakukan untuk mengukur efektivitas algoritma *Support Vector Machine* (SVM) dalam mengklasifikasikan sentimen ulasan pengguna terhadap aplikasi Qur'an Kemenag. Evaluasi dilakukan melalui empat metode utama, yaitu *confusion matrix*, *classification report*, validasi *K-Fold*, dan analisis *wordcloud* sentimen. Berdasarkan *confusion matrix*, model berhasil mengklasifikasikan dengan benar sebanyak 458 data sentimen positif (*True Positive*) dan 475 data sentimen negatif (*True Negative*). Sementara itu, terdapat 8 data negatif yang salah diklasifikasikan sebagai positif (*False Positive*) dan 25 data positif yang salah diklasifikasikan sebagai negatif (*False Negative*). Hasil ini menunjukkan bahwa prediksi yang benar jauh lebih dominan dibandingkan prediksi yang salah, menandakan performa model yang baik.

Lebih lanjut, *classification report* memperlihatkan bahwa model SVM mencapai nilai presisi sebesar 0,95 untuk sentimen negatif dan 0,98 untuk sentimen positif. Nilai recall yang diperoleh adalah 0,98 untuk sentimen negatif dan 0,95 untuk sentimen positif, sehingga menghasilkan nilai F1-score yang seimbang yaitu 0,97 untuk kedua kelas. Secara keseluruhan, akurasi model mencapai 97%, menunjukkan kemampuan model dalam membedakan sentimen positif dan negatif dengan sangat baik. Untuk memastikan kestabilan model, dilakukan validasi menggunakan teknik *K-Fold cross-validation* dengan $K=10$. Hasil rata-rata akurasi pada setiap *fold* mencapai 95,6%, dengan variasi yang relatif kecil antara *fold* satu dengan yang lain, menunjukkan bahwa model memiliki kemampuan generalisasi yang baik dan tidak mengalami *overfitting*.

Selain itu, analisis *wordcloud* dilakukan untuk memahami kata-kata dominan dalam setiap sentimen secara visual. Pada sentimen positif, kata-kata seperti “bagus”, “alhamdulillah”, “terima kasih”, “bantu”, dan “mudah” muncul secara dominan, menggambarkan kepuasan pengguna terhadap manfaat dan kemudahan penggunaan aplikasi. Sedangkan pada sentimen negatif, kata-kata

Protected by PDF Anti-Copy Free

[\(Upgrade to Pro Version to Remove the Watermark\)](#)

seperti “aplikasi”, “buka”, “halaman”, “ayat”, “offline”, dan “error” sering muncul, menunjukkan adanya kendala yang dialami pengguna, seperti kesulitan membuka aplikasi dan masalah . Secara keseluruhan, hasil pengujian ini menunjukkan bahwa algoritma  mampu mengklasifikasikan sentimen ulasan dengan tingkat akurasi tinggi serta memberikan gambaran yang jelas mengenai persepsi pengguna terhadap aplikasi Qur’an Kemenag.



5.1 Kesimpulan

Penelitian ini berhasil menganalisis sentimen ulasan pengguna aplikasi Qur'an Kemenag di *Google Play Store* dengan mengelompokkan opini menjadi sentimen positif dan negatif, sehingga memberikan gambaran mengenai tingkat kepuasan dan ketidakpuasan pengguna terhadap aplikasi tersebut. Algoritma *Support Vector Machine* (SVM) dengan kernel *Radial Basis Function* (RBF) berhasil diimplementasikan untuk memproses dataset ulasan dan menunjukkan performa yang baik dalam klasifikasi sentimen dengan tingkat akurasi yang memuaskan. Sementara itu, ulasan positif mencerminkan kepuasan pengguna terhadap kemudahan penggunaan dan manfaat aplikasi. Algoritma SVM yang digunakan menghasilkan akurasi sebesar 91% pada data awal, dan meningkat menjadi 97% setelah diterapkannya metode SMOTE untuk mengatasi ketidakseimbangan data. Hasil analisis ini memberikan rekomendasi yang relevan bagi pengembang, khususnya dalam hal perbaikan bug, peningkatan fitur, dan upaya menjaga kualitas layanan agar sesuai dengan harapan serta kebutuhan pengguna.


5.2 Saran

Berdasarkan hasil penelitian, pengembang disarankan untuk melakukan perbaikan pada fitur aplikasi, khususnya dengan menangani bug dan fitur yang sering menjadi keluhan pengguna sebagaimana tercermin dalam ulasan bernada negatif. Selain itu, disarankan agar pengembang mengimplementasikan sistem monitoring berbasis analisis sentimen secara real-time guna memantau ulasan pengguna dan merespons masukan dengan lebih cepat dan tepat. Untuk penelitian selanjutnya, peneliti disarankan untuk mengelompokkan data sentimen ke dalam tiga label, yaitu positif, negatif, dan netral, serta mencoba menerapkan metode terbaru yang lebih sesuai dan relevan guna meningkatkan akurasi serta pemahaman yang lebih mendalam terhadap persepsi pengguna.

Protected by PDF Anti-Copy Free

(Upgrade to Pro Version to Remove the Watermark)

DAFTAR PUSTAKA

- 
- [1] H. Hastani, “Metode Pem Al-Qur’an Berbasis Digital Learning,” *Al Ghazali*, vol. 6, no. 1, pp. 2023, doi: 10.52484/al_ghazali.v6i1.403.
- [2] Afif, Z., Fadlly, H., Akbar, A., Syatri, J., Mustopa, M., Jaeni, A., ... & Musadad, M. (2018). PREFERENSI MASYARAKAT DALAM PENGGUNAAN AL-QUR'AN DIGITAL. *SUHUF*, 11(2), 185-214.
- [3] S. O. Sriani and L. Ukhti, “Uji Akurasi Arah Kiblat Menggunakan Fitur Kompas Kiblat Pada Aplikasi Quran Kemenag Versi 2.1.4,” *Astroislamica J. Islam. Astron.*, vol. 1, no. 2, pp. 213–231, 2022, doi: 10.47766/astroislamica.v1i2.951.
- [4] I. N. Istiqomah, U. Ruslan, and J. Abidin, “Pemahaman Remaja Terhadap Penggunaan Aplikasi Al Qur’an Digital Di Masjid Nurul Huda Cikarang Bekasi Jawa Barat,” *Al-Ulum J. Pemikir. dan Penelit. ke Islam.*, vol. 11, no. 3, pp. 233–244, 2024, doi: 10.31102/alulum.11.3.2024.233-244.
- [5] A. Nabillah *et al.*, “Twitter User Sentiment Analysis Of TIX ID Applications Using Support Vector Machine Algorithm,” *RISTEC Res. Inf. Syst. Technol.*, vol. 3, no. 1, pp. 14–27, 2022, doi: 10.31980/ristec.v3i1.1898.
- [6] O. Irnawati and K. Solecha, “Analisis Sentimen Ulasan Aplikasi Flip Menggunakan Naïve Bayes dengan Seleksi Fitur PSO,” *J. Ilm. Intech Inf. Technol. J. UMUS*, vol. 4, no. 02, pp. 189–199, 2022, doi: 10.46772/intech.v4i02.868.
- [7] A. Saepulrohman, S. Saepudin, and D. Gustian, “Analisis Sentimen Kepuasan Pengguna Aplikasi Whatsapp Menggunakan Algoritma Naïve Bayes Dan Support Vector Machine,” *is Best Account. Inf. Syst. Inf. Technol. Bus. Enterp. this is link OJS usf@*, vol. 6, no. 2, pp. 91–105, 2021, doi: 10.34010/aisthebest.v6i2.4919.
- [8] F. M. Mochamad Tri Anjasmoros, Istiadi, “Analisis Sentimen Aplikasi Go-Jek Menggunakan Metode SVM Dan NBC (Studi Kasus: Komentar Pada Play Store),” *Conf. Innov. Appl. Sci. Technol. (CIASTECH 2020)*, no. Ciastech, pp. 489–498, 2020, [Online]. Available: <http://publishing-widyagama.ac.id/ejournal-v2/index.php/ciastech/article/view/1905>
- [9] A. Novantika, “Analisis sentimen ulasan pengguna aplikasi video conference google meet menggunakan metode svm dan logistic regression,” *Prism. Pros. Semin. Nas. Mat.*, vol. 5, pp. 808–813, 2022, [Online]. Available: <https://journal.unnes.ac.id/sju/index.php/prisma/>
- [10] U. Kusnia, F. Kurniawan, and S. Artikel, “Analisis Sentimen Review Aplikasi Media Berita Online Pada Google Play menggunakan Metode Algoritma Support Vector Machines (SVM) Dan Naive Bayes INFO ARTIKEL ABSTRAK,” *J. Keilmuan dan Apl. Tek. Inform.*, vol. 14(1), no.

Protected by PDF Anti-Copy Free

(Upgrade to Pro Version to Remove the Watermark)

- 36, pp. 24–25, 2022, [Online]. Available: <https://doi.org/10.35891/explorit>
- [11] K. A. Rokhman, B. Berliana, and P. Arsi, “Perbandingan Metode Support Vector Machine Dan Decision Tree Untuk Analisis Sentimen Review Komentar Pada Aplikasi Layanan Online,” *J. Inf. Syst. Manag.*, vol. 3, no. 1, pp. 1–7, 2021, doi: [10.30605/joism.2021v3i1.341](https://doi.org/10.30605/joism.2021v3i1.341).
- [12] R. P. I. Putra, M. Akbar, and R. Amalia, “Analisis Sentimen Masyarakat Terhadap Kinerja Persatuan Sepakbola Seluruh Indonesia Menggunakan Metode Backpropagation,” *J. Inf. Technol. Ampera*, vol. 1, no. 2, pp. 106–118, 2020, doi: [10.51519/journalita.volume1.issue2.year2020.page106-118](https://doi.org/10.51519/journalita.volume1.issue2.year2020.page106-118).
- [13] A. Nurian, “Analisis Sentimen Ulasan Pengguna Aplikasi Google Play Menggunakan Naïve Bayes,” *J. Inform. dan Tek. Elektro Terap.*, vol. 11, no. 3s1, pp. 829–835, 2023, doi: [10.23960/jitet.v11i3s1.3348](https://doi.org/10.23960/jitet.v11i3s1.3348).
- [14] F. R. Irawan, A. Jazuli, and T. Khotimah, “Analisis Sentimen Terhadap Pengguna Gojek Menggunakan Metode K-Nearset Neighbors Sentiment Analysis of Gojek Users Using K-Nearest Neighbor,” *JIKO (Jurnal Inform. dan Komputer)*, vol. 5, no. 1, pp. 62–68, 2022, doi: [10.33387/jiko](https://doi.org/10.33387/jiko).
- [15] A. P. Natasuwarna, “Seleksi Fitur Support Vector Machine pada Analisis Sentimen Keberlanjutan Pembelajaran Daring,” *Techno.Com*, vol. 19, no. 4, pp. 437–448, 2020, doi: [10.33633/tc.v19i4.4044](https://doi.org/10.33633/tc.v19i4.4044).
- [16] D. A. Puspitasari, “Kebijakan Pentashihan Aplikasi Al-Qur’an Digital di Indonesia,” *J. Pendidik. Agama Islam*, vol. 8, no. 1, pp. 12–22, 2021, [Online]. Available: <http://ejournal.uin-malang.ac.id/index.php/jpai>
- [17] A. B. Susanto, “Penggunaan Aplikasi Al Qur ’ an Digital Dalam Pembelajaran Pada Mata Pelajaran Qur ’ an Hadist”.
- [18] M. Rohmati, “Menguak Nilai-Nilai Hubb al-Wathan dalam Al-Qashas: 85 (Studi Terjemah dan Tafsir Qur’an Kemenag Digital),” *J. Ushuluddin Media Dialog Pemikir. Islam*, vol. 22, no. 1, pp. 37–51, 2020, doi: [10.24252/jumdpi.v22i1.11602](https://doi.org/10.24252/jumdpi.v22i1.11602).
- [19] L. B. Ilmawan, “Membangun Web Crawler Berbasis Web Service Untuk Data Crawling Pada Website Google Play Store,” *Ilk. J. Ilm.*, vol. 10, no. 2, pp. 215–224, 2018, doi: [10.33096/ilkom.v10i2.282.215-224](https://doi.org/10.33096/ilkom.v10i2.282.215-224).
- [20] 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](https://doi.org/10.37600/tekinkom.v6i1.773).
- [21] R. S. Oktavian and S. Budi, “Analisis Dataset Google Playstore Menggunakan Metode Exploratory Data Analysis,” *J. Strateg.*, vol. 2, no. 2, pp. 636–649, 2020.
- [22] Y. X. Chu, X. G. Liu, and C. H. Gao, “Multiscale models on time series of

Protected by PDF Anti-Copy Free

(Upgrade to Pro Version to Remove the Watermark)

- silicon content in blast furnace hot metal based on Hilbert-Huang transform,” *Proc. 2011 Chinese Conf. Precis. Conf. CCDC 2011*, pp. 842–847, 2011, doi: 10.1109/CCDC.2011.5946190.
- [23] M. M. Maarif and N. [REDACTED], “Analisis Sentimen Review Aplikasi LinkedIn di Google Play Menggunakan Support Vector Machine,” *Progresif J. Ilm. Komput.*, vol. 20, no. 1, p. 454, 2024, doi: 10.35889/progresif.v20i1.1614.
- [24] A. Sukmawati, dian eka Ratnawati, and nanang yudi Setiawan, “Analisis Sentimen Aplikasi Glints Berdasarkan Ulasan Google Play Store Menggunakan Metode Support Vectormachine,” *J. Pengembangan Teknol. Inf. dan Ilmu Komput.*, vol. 1, no. 2548–964X, 2017, [Online]. Available: <https://j-ptiik.ub.ac.id/index.php/j-ptiik/article/view/14107/6300>
- [25] E. R. Kaburuan and N. R. Setiawan, “Sentimen Analisis Review Aplikasi Digital Korlantas Pada Google Play Store Menggunakan Metode SVM,” *J. Sisfokom (Sistem Inf. dan Komputer)*, vol. 12, no. 1, pp. 105–116, 2023, doi: 10.32736/sisfokom.v12i1.1614.
- [26] M. D. Hendriyanto, A. A. Ridha, and U. Enri, “Analisis Sentimen Ulasan Aplikasi Mola Pada Google Play Store Menggunakan Algoritma Support Vector Machine,” *INTECOMS J. Inf. Technol. Comput. Sci.*, vol. 5, no. 1, pp. 1–7, 2022, doi: 10.31539/intecomsv5i1.3708.
- [27] D. Diandra Audiansyah, D. Eka Ratnawati, and B. Trias Hanggara, “Analisis Sentimen Aplikasi MyXL menggunakan Metode Support VectorMachine berdasarkan Ulasan Pengguna di Google Play Store,” *J. Pengemb. Teknol. Inf. dan Ilmu Komput.*, vol. 6, no. 8, pp. 3987–3994, 2022, [Online]. Available: <http://j-ptiik.ub.ac.id>
- [28] S. I. Nurhafida and F. Sembiring, “Analisis Sentimen Aplikasi Novel Online Di Google Play Store Menggunakan Algoritma Support Vector Machine (SVM),” *J. Sains Komput. Inform. (J-SAKTI)*, vol. 6, no. 1, pp. 317–327, 2022.
- [29] S. Lestari and S. Saepudin, “Support Vector Machine: Analisis Sentimen Aplikasi Saham di Google Play Store,” *JUSIFO (Jurnal Sist. Informasi)*, vol. 7, no. 2, pp. 81–90, 2021, doi: 10.19109/jusifo.v7i2.9825.
- [30] F. Bei and S. Sudin, “Analisis Sentimen Aplikasi Tiket Online Di Play Store Menggunakan Metode Support Vector Machine (Svm),” *Sismatik*, vol. 01, no. 01, pp. 91–97, 2021.
- [31] A. Erfina, E. S. Basryah, A. Saepulrohman, and D. Lestari, “Analisis Sentimen Aplikasi Pembelajaran Online Di Play Store Pada Masa Pandemi Covid-19 Menggunakan Algoritma Support Vector Machine (Svm),” *Semin. Nas. Inform.*, vol. 1, no. 1, pp. 145–152, 2020, [Online]. Available: <http://jurnal.upnyk.ac.id/index.php/semnasif/article/view/4094>

Protected by PDF Anti-Copy Free

(Upgrade to Pro Version to Remove the Watermark)

- [32] M. R. Fauzan, H. Oktafia, L. Wijaya, and J. Karman, "KENAIKAN HARGA BBM DI MEDIA SOSIAL TWITTER," vol. 1, no. 1, pp. 82–89, 2023.
- [33] D. Rudini, D. G. Purnama, and A. A. Khan, "Penggunaan Teknik Web Scraping dalam Aplikasi Pengambilan Data dari Google Maps untuk Menunjang Digital Marketing," *Lentera Multidiscip. Stud.*, vol. 2, no. 1, pp. 10–19, 2023, doi: 10.57096/lentera.v2i1.61.
- [34] L. Hermawan and M. Bellanar Ismiati, "Pembelajaran Text Preprocessing berbasis Simulator Untuk Mata Kuliah Information Retrieval," *J. Transform.*, vol. 17, no. 2, p. 188, 2020, doi: 10.26623/transformatika.v17i2.1705.
- [35] H. Syah and A. Witanti, "Analisis Sentimen Masyarakat Terhadap Vaksinasi Covid-19 Pada Media Sosial Twitter Menggunakan Algoritma Support Vector Machine (Svm)," *J. Sist. Inf. dan Inform.*, vol. 5, no. 1, pp. 59–67, 2022, doi: 10.47080/simika.v5i1.1411.
- [36] A. Salam, J. Zeniarja, and R. S. U. Khasanah, "Analisis Sentimen Data Komentar Sosial Media Facebook Dengan K-Nearest Neighbor (Studi Kasus Pada Akun Jasa Ekspedisi Barang J&T Ekpress Indonesia)," *Pros. SINTAK*, pp. 480–486, 2018.
- [37] L. A. Andika, P. A. N. Azizah, and R. Respatiwan, "Analisis Sentimen Masyarakat terhadap Hasil Quick Count Pemilihan Presiden Indonesia 2019 pada Media Sosial Twitter Menggunakan Metode Naive Bayes Classifier," *Indones. J. Appl. Stat.*, vol. 2, no. 1, p. 34, 2019, doi: 10.13057/ijas.v2i1.29998.
- [38] A. Riyani, M. Zidny Naf'an #2, and A. Burhanuddin, "Penerapan Cosine Similarity dan Pembobotan TF-IDF untuk Mendeteksi Kemiripan Dokumen," *Jlk*, vol. 2, no. 1, pp. 23–27, 2019.
- [39] N. Z. Rania and R. D. Syah, "Analisis Sentimen Terhadap Aplikasi Gojek Pada Play Store Menggunakan Metode Random Forest Classifier," *J. Ilm. Inform. Komput.*, vol. 29, no. 2, pp. 144–153, 2024, doi: 10.35760/ik.2024.v29i2.11877.
- [40] S. Ernawati and R. Wati, "Penerapan Algoritma K-Nearest Neighbors Pada Analisis Sentimen Review Agen Travel," *J. Khatulistiwa Inform.*, vol. 6, no. 1, pp. 64–69, 2018.
- [41] M. T. Diwandanu and L. M. Wisudawati, "Analisis Sentimen Terhadap Twit Maxim Pada Twitter Menggunakan R Programming Dan K Nearest Neighbors," *J. Ilm. Inform. Komput.*, vol. 28, no. 1, pp. 1–16, 2023, doi: 10.35760/ik.2023.v28i1.7909.
- [42] N. Cahyono and Anggista Oktavia Praneswara, "Analisis Sentimen Ulasan Aplikasi TikTok Shop Seller Center di Google Playstore Menggunakan Algoritma Naive Bayes," *Indones. J. Comput. Sci.*, vol. 12, no. 6, pp. 3925–

Protected by PDF Anti-Copy Free

(Upgrade to Pro Version to Remove the Watermark)

3940, 2023, doi: 10.33022/ijcs.v12i6.3473.

- [43] G. F. Grandis, Y. Arum, and Indriati, "Seleksi Fitur Gain Ratio pada Analisis Sentimen Kebijakan Pemerintah Mengenai Pembelajaran Jarak Jauh dengan K-Nearest Neighbor," *J. Pengemb. Teknol. Inf. dan Ilmu Komput.*, vol. 5, no. 8, pp. 514, 2021.
- [44] K. Akbar and M. Hayaty, "Data Balancing untuk Mengatasi Imbalance Dataset pada Prediksi Produksi Padi," *J. Ilm. Intech Inf. Technol. J. UMUS*, vol. 2, no. 02, 2020, doi: 10.46772/intech.v2i02.283.
- [45] R. Fahlapi *et al.*, "Analisa Sentimen Vaksinasi Covid-19 Dengan Metode Support Vector Machine Dan Naïve Bayes Berbasis Teknik Smote," *J. Inform. Kaputama*, vol. 6, no. 1, pp. 57–63, 2022, doi: 10.59697/jik.v6i1.136.
- [46] S. Paaban Simanjuntak, S. Sandino Berutu, and G. C. Setyawan, "Implementasi Metode CNN pada Klasifikasi Sentimen terhadap Pelaksanaan Piala Dunia U-17 (Implementation of the CNN Method in Classifying Sentiments Regarding the Implementation of the U-17 World Cup)," *J. Eng. Emerg. Technol.*, vol. 02, no. 01, pp. 23–32, 2024, [Online]. Available: www.jeet.unram.ac.id
- [47] N. I. Widiastuti, E. Rainarli, and K. E. Dewi, "Peringkasan dan Support Vector Machine pada Klasifikasi Dokumen," *J. Infotel*, vol. 9, no. 4, p. 416, 2017, doi: 10.20895/infotel.v9i4.312.
- [48] Muryadi, Agustanico Dwi. "Model evaluasi program dalam penelitian evaluasi." *Jurnal Ilmiah Penjas (Penelitian, Pendidikan Dan Pengajaran)* 3.1 (2017).
- [49] A. Azis Maarif, "Penerapan Algoritma Fuzzy," *Univ. Dian Nusant.*, 2015, [Online]. Available: [repository.unair.ac.id/29371/3/15 BAB II.pdf](http://repository.unair.ac.id/29371/3/15%20BAB%20II.pdf)
- [50] D. A. Perkasa *et al.*, "Sistem Ujian Online Essay Dengan Penilaian Menggunakan Metode Latent Sematic Analysis (Lsa)," *J. Ilm. Rekayasa dan Manaj. Sist. Inf.*, vol. 1, no. 1, pp. 1–9, 2015, [Online]. Available: <https://ejournal.uin-suska.ac.id/index.php/RMSI/article/view/1313>
- [51] S. Fatika, N. Halim, and D. Aktuaria, "Analisis Perbandingan Klasifikasi dan Penerapan," vol. 12, no. 2, 2023.

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

LAMPIRAN

1. Surat Keputusan Pembimbing Skripsi



SURAT KEPUTUSAN
DEKAN FAKULTAS ILMU TEKNIK
UNIVERSITAS BINA INSAN
 Nomor 0834/UNIV.BI/FTT.1/SK/2024

Tentang

DOSEN PEMBIMBING SKRIPSI
PROGRAM STUDI INFORMATIKA TAHUN AKADEMIK 2024/2025
UNIVERSITAS BINA INSAN

DEKAN FAKULTAS ILMU TEKNIK
UNIVERSITAS BINA INSAN

- Menimbang** :
- Bahwa untuk kelancaran penyelesaian skripsi mahasiswa pada Program Studi Informatika Tahun Akademik 2024/2025 pada Universitas Bina Insan, perlu ditunjuk pembimbing skripsi yang bertanggung jawab penuh pada pelaksanaan bimbingan tersebut;
 - Bahwa untuk keperluan sebagaimana poin satu tersebut di atas perlu ditetapkan dengan surat keputusan Dekan.
- Mengingat** :
- Undang-undang No. 12 tahun 2012 tentang Pendidikan Nasional,
 - Peraturan Pemerintah Republik Indonesia No. 4 tahun 2014 tentang penyelenggaraan Pendidikan Tinggi;
 - Keputusan Menteri Pendidikan Nasional Republik Indonesia No. 232/U/2000 tentang Pedoman Penyusunan Kurikulum Pendidikan Tinggi dan Penilaian Hasil Belajar Mahasiswa,
 - Keputusan Menteri Pendidikan Nasional Republik Indonesia No. 184/U/2001 tentang pedoman Pengawasan-Pengendalian dan Pembinaan Program Diploma, Sarjana dan Pascasarjana di Perguruan Tinggi;
 - SK Menteri Riset, Teknologi Dan Pendidikan Tinggi Republik Indonesia Nomor 223/KPT/I/2019 Tentang Izin Penggabungan Sekolah Tinggi Ilmu Ekonomi Musi Rawas dan Sekolah Tinggi Manajemen Ilmu Komputer Musi Rawas Menjadi Universitas Bina Insan;
 - SK Ketua Yayasan Nomor 01.113/YPDT-Plg/KP/SK/IV/2019 Tentang Pengangkatan Dekan Universitas Bina Insan Lubuklinggau;
 - SK Dekan Universitas Bina Insan Nomor 1235/UNIV.BI/R/KP/SK/2020 Tentang Pengangkatan Pejabat Pada Universitas Bina Insan Lubuklinggau;
 - Statuta Universitas Bina Insan Lubuklinggau;

MEMUTUSKAN

- Menetapkan** :
- Pertama** : Mengangkat nama-nama yang tercantum pada lampiran surat keputusan ini sebagai Dosen Pembimbing 1 dan Pembimbing 2 Skripsi Program Studi Informatika Tahun Akademik 2024/2025 pada Universitas Bina Insan.
- Kedua** : Semua biaya yang timbul akibat dikeluarkannya surat keputusan ini dibebankan pada anggaran Universitas Bina Insan.
- Ketiga** : Surat keputusan ini berlaku sejak tanggal ditetapkan, dengan ketentuan apabila ternyata dikemudian hari terdapat kekeliruan dalam penetapan surat keputusan ini akan diperbaiki sebagaimana mestinya;

Ditetapkan di : Lubuklinggau
 Pada tanggal : 29 Agustus 2024
 Dekan Fakultas Ilmu Teknik,


Dr. Rudi Kurniawan, S.T., M.Kom
 UNIVERSITAS BINA INSAN
 FAKULTAS ILMU TEKNIK

- Tembusan Yth:
- Ketua Yayasan Pendidikan Dwi Tunggal Palembang (sebagai laporan)
 - Rektor Universitas Bina Insan (sebagai laporan)
 - Arsip

Protected by PDF Anti-Copy Free

(Upgrade to Pro Version to Remove the Watermark)



No	NIM	Nama Mahasiswa	Pembimbing 1	Pembimbing 2
102	2102020151	Rocky Putra A	Dr. Rudi Kurniawan, St., M. Kom	Tri Hasanah Bimastari Aviani, M. Kom
103	2102020152	Muhammad Ayasha Alfiah Ajerti	Ahmad Sobri, M. Kom	Muhammad Nur Alamsyah, M. Kom
104	2102020154	Daniel Adless Erch	Lukman Sunardi, M. Kom	Nelly Khairani Daulay, M. Kom
105	2102020157	Defri Afandy Saputra	Harma Oktavia Lingga Wijaya, M. Kom	Bunga Intan, M. Kom
106	2102020159	Hawmelia Pontesyah	Dr. Susanto, M. Kom	A. Taqwa Martadinata, M. Kom
107	2102020160	Rima Amanda Putri	Elmayati, M. Kom	Lukman Hakim, M. Kom
108	2102020162	Berly Ramadhan	Cindi Wulandari, M. Kom	Fido Rizki, M. Kom
109	2102020163	Setiyo Adi Wibowo	Dr. Rudi Kurniawan, St., M. Kom	Budi Santoso, M. Kom
110	2102020164	Yoga Yarkhamsetiawan	Dr. Muhamad Akbar, ST., M. IT	Satrianansyah, M. Kom
111	2102020167	Riska Almira	Davit Irawan, M. Kom	Harma Oktavia Lingga Wijaya, M. Kom
112	2102020171P	Yoga Pahrullah	Asep Toyib Hidayat, M. Kom	Bunga Intan, M. Kom

Ditetapkan di : Lubuklinggau
 Pada tanggal : 29 Agustus 2024
 Dekan Fakultas Ilmu Teknik,

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

- Tembusan Yth:
1. Ketua Yayasan Pendidikan Dwi Tunggal Palembang (sebagai laporan)
 2. Rektor Universitas Bina Insan (sebagai laporan)
 3. Arsip

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

Lampiran Surat Keputusan Penguji Proposal Skripsi



SURAT KEPUTUSAN
DEKAN FAKULTAS ILMU TEKNIK
Nomor : 0008 /UNIV.BI/FIT.3/SK/2025

TENTANG
PENGGANGKATAN DOSEN PENGUJI PROPOSAL SKRIPSI MAHASISWA
PROGRAM STUDI INFORMATIKA FAKULTAS ILMU TEKNIK
UNIVERSITAS BINA INSAN LUBUKLINGGAU

DENGAN RAHMAT TUHAN YANG MAHA ESA, UNIVERSITAS BINA INSAN LUBUKLINGGAU

- Memperhatikan** : Bahwa dengan selesainya mahasiswa menyusun Proposal Skripsi Program Studi Informatika Fakultas Ilmu Teknik Tahun Akademik 2024/2025, maka perlu menunjuk dan mengangkat Dosen Penguji Proposal Skripsi untuk menguji mahasiswa tersebut dalam menyelesaikan kuliahnya di lingkungan Universitas Bina Insan Lubuklinggau;
- Menimbang** : 1. Bahwa dalam upaya menyelenggarakan pendidikan tinggi yang berkualitas dipandang perlu mengangkat Dosen Penguji Proposal Skripsi di lingkungan Universitas Bina Insan Lubuklinggau;
2. Sehubungan dengan Butir 1 (satu) tersebut di atas, maka dipandang perlu mengeluarkan Surat Keputusan sebagai landasan hukumnya;
- Mengingat** : 1. Undang-Undang No. 20 Tahun 2003 tentang Sistem Pendidikan Nasional;
2. Peraturan Pemerintah Republik Indonesia No. 60 Tahun 1999 tentang Pendidikan Tinggi;
3. Keputusan Menteri Pendidikan Nasional Republik Indonesia No. 232/U/2000 tentang Pedoman Penyusunan Kurikulum Pendidikan Tinggi dan Penilaian Hasil Belajar Mahasiswa;
4. Keputusan Menteri Pendidikan Nasional Republik Indonesia No. 184/U/2001 tentang Pedoman Pengawasan-pengendalian dan Pembinaan Program Diploma, Sarjana dan Pascasarjana di Perguruan Tinggi;
5. SK Menteri Riset, Teknologi Dan Pendidikan Tinggi Republik Indonesia Nomor 223/KPT/I/2019 Tentang Izin Penggabungan Sekolah Tinggi Ilmu Ekonomi Musi Rawas dan Sekolah Tinggi Manajemen dan Ilmu Komputer Musi Rawas Menjadi Universitas Bina Insan;
6. SK Ketua Yayasan Nomor 01.113/YPDT-PtG/KP/SK/IV/2019 Tentang Pengangkatan Rektor Universitas Bina Insan Lubuklinggau;
7. SK Rektor Universitas Bina Insan Nomor 1235/UNIV.BI/R/KP/SK/2020 Tentang Pengangkatan Pejabat Pada Universitas Bina Insan Lubuklinggau;
8. Statuta Universitas Bina Insan Lubuklinggau;

MEMUTUSKAN

- Menetapkan Pertama** : Mengangkat Saudara yang namanya tercantum pada lampiran ini, sebagai penguji Proposal Skripsi Program Studi Informatika Fakultas Ilmu Teknik Tahun Akademik 2024/2025 di Universitas Bina Insan Lubuklinggau;
- Kedua** : Semua biaya yang timbul akibat dikeluarkannya Surat Keputusan ini dibebankan kepada anggaran Universitas Bina Insan Lubuklinggau atau dana khusus yang disediakan untuk itu;
- Ketiga** : Kepada yang bersangkutan diberikan honorarium yang besarnya sesuai dengan peraturan Universitas Bina Insan Lubuklinggau;
- Keempat** : Surat Keputusan ini berlaku sejak tanggal ditetapkan, dengan ketentuan apabila ternyata dikemudian hari terdapat kekeliruan dalam penetapan surat keputusan ini, akan diperbaiki sebagaimana mestinya;

Demikian Surat Keputusan ini ditetapkan untuk dilaksanakan sebagaimana mestinya.

Ditetapkan di : Lubuklinggau
Pada tanggal : 03 Januari 2025
Dekan Fakultas Ilmu Teknik

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

- Tembusan Yth.
1. Ketua Yayasan Pendidikan Dwi Tunggal Palembang (sebagai laporan)
 2. Rektor Universitas Bina Insan (sebagai laporan)
 3. Arsip.

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



Dosen Dekan Fakultas Ilmu Teknik Universitas Bina Insan Lubuklinggau
: 0808 ANIV.BI/FTI.3/SK/2025
: 03 Januari 2025
: Susunan Pengangkatan Dosen Perang Sidang Proposal Skripsi
Program Studi Informatika IA. 2024/2025

No	Nama Mahasiswa	NIM	Ketua	Sekretaris	Anggota	Hari	Tanggal	Jam	Ruang
1	Ramadhan Istihario	2102020131	Harna Oktavia Lingga W. M.Kom	Budi Santoso, M.Kom	Nelly Khairani Dauley, M.Kom	Sabtu	04/01/2025	08.00-09.00	Ruang Sidang 1
2	Randy Marsevin	2102020098	Tri Hasanah BA. M.Kom	A. Taqwa Martadinata, M.Kom	Dr. M. Agus Syamsul A. S.St., M.Kom	Sabtu	04/01/2025	08.00-09.00	Ruang Sidang 2
3	Krisna Riski Pratama	2102020105	Dr. Susanto, M.Kom	Andi Anto Tri susilo, M.Kom	Elmayati, M.Kom	Sabtu	04/01/2025	08.00-09.00	Ruang Sidang 3
4	Sigit Suseno	2102020108	Harna Oktavia Lingga W. M.Kom	Tri Hasanah BA. M.Kom	Dr. Muhamad Akbar, S.T., MIT	Sabtu	04/01/2025	09.00-10.00	Ruang Sidang 1
5	Rahmat Panulloh	2102020118	Nelly Khairani Dauley, M.Kom	Muhammad Irvai, M.Kom	Arnanto, M.Kom	Sabtu	04/01/2025	09.00-10.00	Ruang Sidang 2
6	Rima Amanda Putri	2102020160	Elmayati, M.Kom	Lukman Hakim, M.Kom	Budi Santoso, M.Kom	Sabtu	04/01/2025	09.00-10.00	Ruang Sidang 3
7	Yoga Yarkham Setiawan	2102020164	Dr. Muhamad Akbar, S.T., MIT	Satrianansyah, M.Kom	Harna Oktavia Lingga W. M.Kom	Sabtu	04/01/2025	10.00-11.00	Ruang Sidang 1
8	Ully Kurniati	2102020022	Dr. M. Agus Syamsul A. S.St., M.Kom	Nelly Khairani Dauley, M.Kom	Elmayati, M.Kom	Sabtu	04/01/2025	10.00-11.00	Ruang Sidang 2
9	Rocky Putra A.	2102020151	Dr. Rudi Kurniawan, S.T., M.Kom	Tri Hasanah BA. M.Kom	Budi Santoso, M.Kom	Sabtu	04/01/2025	10.00-11.00	Ruang Sidang 3
10	M. Azyaraka Pradana	2102020119	Arnanto, M.Kom	Fido Rizki, M.Kom	Cindi Wulandari, M.Kom	Sabtu	04/01/2025	11.00-12.00	Ruang Sidang 1
11	Riski Anabela	2102020004	Joni Karman, M.Kom	Harna Oktavia Lingga W. M.Kom	Dr. Rudi Kurniawan, S.T., M.Kom	Sabtu	04/01/2025	11.00-12.00	Ruang Sidang 2
12	M. Tengku Rafliansyah	2102020143	Budi Santoso, M.Kom	Arnanto, M.Kom	Tri Hasanah BA. M.Kom	Sabtu	04/01/2025	11.00-12.00	Ruang Sidang 3

Protected by PDF Anti-Copy Free

(Upgrade to Pro Version to Remove the Watermark)

Lampiran 3. Surat Keputusan Penguji Skripsi



SURAT KEPUTUSAN
DEKAN FAKULTAS ILMU TEKNIK
 Nomor : 0166 /UNIV.BI/F.IT.3/SK/2025

TENTANG
PENGGANGKATAN DOSEN PENGUJI SKRIPSI MAHASISWA
PROGRAM STUDI INFORMATIKA FAKULTAS ILMU TEKNIK
UNIVERSITAS BINA INSAN LUBUKLINGGAU

DENGAN RAHMAT TUHAN YANG MAHA ESA, UNIVERSITAS BINA INSAN LUBUKLINGGAU

- Memperhatikan** : Bahwa dengan selesainya mahasiswa menyusun Skripsi Program Studi Informatika Fakultas Ilmu Teknik Tahun Akademik 2024/2025, maka perlu menunjuk dan mengangkat Dosen Penguji Skripsi untuk menguji mahasiswa tersebut dalam menyelesaikan kuliahnya di lingkungan Universitas Bina Insan Lubuklinggau;
- Menimbang** : 1. Bahwa dalam upaya menyelenggarakan pendidikan tinggi yang berkualitas dipandang perlu mengangkat Dosen Penguji Skripsi di lingkungan Universitas Bina Insan Lubuklinggau;
 2. Sehubungan dengan Butir 1 (satu) tersebut di atas, maka dipandang perlu mengeluarkan Surat Keputusan sebagai landasan hukumnya;
- Mengingat** : 1. Undang-Undang No. 20 Tahun 2003 tentang Sistem Pendidikan Nasional;
 2. Peraturan Pemerintah Republik Indonesia No. 60 Tahun 1999 tentang Pendidikan Tinggi;
 3. Keputusan Menteri Pendidikan Nasional Republik Indonesia No. 232/U/2000 tentang Pedoman Penyusunan Kurikulum Pendidikan Tinggi dan Penilaian Hasil Belajar Mahasiswa;
 4. Keputusan Menteri Pendidikan Nasional Republik Indonesia No. 184/U/2001 tentang Pedoman Pengawasan-pengendalian dan Pembinaan Program Diploma, Sarjana dan Pascasarjana di Perguruan Tinggi;
 5. SK Menteri Riset, Teknologi Dan Pendidikan Tinggi Republik Indonesia Nomor 223/KPT/I/2019 Tentang Izin Penggabungan Sekolah Tinggi Ilmu Ekonomi Musi Rawas dan Sekolah Tinggi Manajemen dan Ilmu Komputer Musi Rawas Menjadi Universitas Bina Insan;
 6. SK Ketua Yayasan Nomor 01.113/YPDT-Plg/KP/SK/IV/2019 Tentang Pengangkatan Rektor Universitas Bina Insan Lubuklinggau;
 7. SK Rektor Universitas Bina Insan Nomor 1235/UNIV.BI/R/KP/SK/2020 Tentang Pengangkatan Pejabat Pada Universitas Bina Insan Lubuklinggau;
 8. Statuta Universitas Bina Insan Lubuklinggau;

MEMUTUSKAN

- Menetapkan Pertama** : Mengangkat Saudara yang namanya tercantum pada lampiran ini, sebagai penguji Skripsi Program Studi Informatika Fakultas Ilmu Teknik Tahun Akademik 2024/2025 di Universitas Bina Insan Lubuklinggau;
- Kedua** : Semua biaya yang timbul akibat dikeluarkannya Surat Keputusan ini dibebankan kepada anggaran Universitas Bina Insan Lubuklinggau atau dana khusus yang disediakan untuk itu;
- Ketiga** : Kepada yang bersangkutan diberikan honorarium yang besarnya sesuai dengan peraturan Universitas Bina Insan Lubuklinggau;
- Keempat** : Surat Keputusan ini berlaku sejak tanggal ditetapkan, dengan ketentuan apabila ternyata dikemudian hari terdapat kekeliruan dalam penetapan surat keputusan ini, akan diperbaiki sebagaimana mestinya;

Demikian Surat Keputusan ini ditetapkan untuk dilaksanakan sebagaimana mestinya.

Ditetapkan di : Lubuklinggau
 Pada tanggal : 24 Januari 2025
 Dekan Fakultas Ilmu Teknik,

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

Tembusan Yth.
 1. Ketua Yayasan Pendidikan Dwi Tunggal Palembang (sebagai laporan)
 2. Arsip.

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

13	Pamiza	2102020078	Dr.M.Agus Syamsul A.S.S	M.Kom	Dr. Rudi Kurniawan, S.T, M.Kom	Sabtu	25/01/2025	11.00-12.00	Ruang Sidang 2
14	Yoga Yarkhamsetiawan	2102020164	Dr. Muhammad Akbar, MIT	Satrianansyah, M.Kom	Harna oktavia Lingga W,M.Kom	Sabtu	25/01/2025	11.00-12.00	Ruang Sidang 3
15	Muhammad Salman	2102020057	Dr. Rudi Kurniawan, S.T, M.Kom	Bunga Intan, M.Kom	Budi Santoso, M.Kom	Sabtu	25/01/2025	11.00-12.00	Ruang Sidang 4
16	Hestin Nurdiana	2102020116	Novi Lestari, M. Kom	Rusdiyanto, M. Kom	Ahmad Sobri, M.Kom	Sabtu	25/01/2025	11.00-12.00	Ruang Sidang 5
17	Putri Cheria Adhany	2102020070	Cindi Wulandari, M.Kom	Bunga Intan, M.Kom	Budi Santoso, M.Kom	Sabtu	25/01/2025	13.00-14.00	Ruang Sidang 2
18	Rizki Anabela	2102020004	Joni Karman, M.Kom	Harna oktavia Lingga W,M.Kom	Dr. Rudi Kurniawan, S.T, M.Kom	Sabtu	25/01/2025	13.00-14.00	Ruang Sidang 3
19	Fitra Marshanda	2102020035	Budi Santoso, M.Kom	Armanto, M.Kem	Dr. M. Agus Syamsul .A.ST.,M.Kom	Sabtu	25/01/2025	13.00-14.00	Ruang Sidang 4
20	Bintang Deni Syaputra	2102020103	Dr. M. Agus Syamsul Arifin, M.Kom	Budi Santoso, M.Kom	Dr. Susanto, M.Kom	Sabtu	25/01/2025	13.00-14.00	Ruang Sidang 5
21	Erik kurniawan	2102020106	Elmayati,M.Kom	Harna oktavia Lingga W,M.Kom	Budi Santoso, M.Kom	Sabtu	25/01/2025	14.00-15.00	Ruang Sidang 1
22	Orya Abi Dunya	2102020021	Budi Santoso, M.Kom	Antoni Zulius M.Kom	Armanto, M.Kom	Sabtu	25/01/2025	14.00-15.00	Ruang Sidang 2
23	Aulia Putri	2102020005	Budi Santoso, M.Kom	Bunga Intan, M.Kom	Harna oktavia Lingga W,M.Kom	Sabtu	25/01/2025	14.00-15.00	Ruang Sidang 3
24	Angel reka poetri	2102020095	Asep Toyib Hidayat, M.Kom	Muhammad Irvai, M.Kom	Novi Lestari, M.Kom	Sabtu	25/01/2025	14.00-15.00	Ruang Sidang 4
25	Ulya kurniati	2102020022	Dr.M.Agus Syamsul Arifin,SSI, M.Kom	Nelly Khairani Doulay,M.Kom	Elmayati, M.Kom	Sabtu	25/01/2025	14.00-15.00	Ruang Sidang 5
26	M.Tengku raffiansyah	2102020143	Budi santoso M.Kom	Armanto M.Kom	Tri Hasanah BA, M.Kom	Sabtu	25/01/2025	15.00-16.00	Ruang Sidang 1
27	Sigit Suseno	2102020108	Harna oktavia Lingga W,M.Kom	Tri Hasanah BA, M.Kom	Dr. Muhammad Akbar, MIT	Sabtu	25/01/2025	15.00-16.00	Ruang Sidang 2

Protected by PDF Anti-Copy Free

(Upgrade to Pro Version to Remove the Watermark)

Lampiran 4. Lembar Perbaikan Seminar Proposal



LEMBAR PERBAIKAN SEMINAR PROPOSAL SKRIPSI

Nama Mahasiswa : Yoga Yarkhamsetiawan
 NIM : 2102020164
 Jenjang Pendidikan : Strata 1 (S1)
 Fakultas : Ilmu Teknik
 Program Studi : Informatika
 Konsentrasi :

Judul : Analisis Sentimen Pengguna Aplikasi Qur'an Kemenag Pada Ulasan Google Play Store Menggunakan Metode Support Vector Machine(SVM)

No	Dosen Penguji	Komentar Perbaikan	Tanda Tangan Ujian	Tanda Tangan Revisi
1	Dr. Muhammad Akbar, M.IT	-		
2	Satrianansyah, M.Kom	-		
3	Harna. Oktavia. L.W. M.Kom.	Perbaiki Daftar Pustaka • Kalimat Bahasa China • Tambah Jurnal Oospem		

Lubuklinggau, 09 Januari2025
 Ketua Program Studi Informatika

(Budi Santoso, M.Kom)

Protected by PDF Anti-Copy Free

(Upgrade to Pro Version to Remove the Watermark)

Lampiran 5. Lembar Perbaikan Ujian Skripsi



LEMBAR PERBAIKAN SIDANG SKRIPSI

Nama Mahasiswa : Yoga Yarkhamsetiawan
 NIM : 2102020164
 Jenjang Pendidikan : Strata 1 (S1)
 Fakultas : Ilmu Teknik
 Program Studi : Informatika
 Konsentrasi :
 Judul : Analisis Sentimen Pengguna Aplikasi Qur'an Kemenag Pada Ulasan Google Play Store Menggunakan Metode Support Vector Machine(SVM)

No	Dosen Penguji	Komentar Perbaikan	Tanda Tangan Ujian	Tanda Tangan Revisi
1	Dr. Muhammad Akbar, M.IT	-		
2	Satrianansyah, M.Kom	-		
3	Harna Oktavia Lingga W, M.Kom	Perbaiki ser. Referensi		

Lubuklinggau, 03 Februari2025
 Ketua Program Studi Informatika

(Budi Santoso, M.Kom)

Protected by PDF Anti-Copy Free

(Upgrade to Pro Version to Remove the Watermark)

Lampiran 6. Lembar Bimbingan Proposal



LEMBAR BIMBINGAN PROPOSAL SKRIPSI

Nama : YOGA YARKHAMSETIAWAN
 Nim : 2102020164
 Program Studi : INFORMATIKA
 Pembimbing 1 : DR. MUHAMMAD AKBAR, M.IT
 Pembimbing 2 : SATRIANANSYAH, M.KOM
 Judul : ANALISIS SENTIMEN PENGGUNA APLIKASI QUR'AN KEMENAG PADA ULASAN GOOGLE PLAY STORE MENGGUNAKAN METODE SUPPORT VECTOR MACHINE (SVM)


NO	TANGGAL	TOPIK	KOMENTAR PEMBIMBING	TANDA TANGAN PEMBIMBING	
				1	2
1.	3/12-29.	proposal	perbaiki sesuai catatan Pd dapat proposal.		
2.	07/12-29.	proposal.	* Perbaiki label. * Perbaiki penulisan gambar / foto. * Istilah asing akan miring. * ACC lanjut pl.		

Lubuklinggau, 30 Desember 2024
 Ketua Program Studi Informatika

(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
 FAKULTAS ILMU TEKNIK
Lubuklinggau, Klaten, Lubuklinggau Selatan 1 Kota Lubuklinggau Provinsi Sumatera Selatan

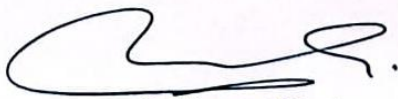
PDF

LEMBAR BUKU PROPOSAL SKRIPSI

Nama : YOGA YARKHANNURRAHMAN
Nim : 2102020164
Program Studi : INFORMATIKA
Pembimbing 1 : DR. MUHAMMAD AKBAR, M.IT
Pembimbing 2 : SATRIANANSYAH, M.KOM
Judul : ANALISIS SENTIMEN PENGGUNA APLIKASI QUR'AN KEMENAG PADA ULASAN GOOGLE PLAY STORE MENGGUNAKAN METODE *SUPPORT VECTOR MACHINE (SVM)*

NO	TANGGAL	TOPIK	KOMENTAR PEMBIMBING	TANDA TANGAN PEMBIMBING	
				1	2
	19/12		Perbaiki latar belakang - Cari permasalahan - 2 why?	d	
			Acc bagian 2 & 3	d	
	30/12		Acc Usian	d	

Lubuklinggau, 30 Desember 2024
Ketua Program Studi Informatika


 (Budi Santoso, M.Kom)

0733-4553932 (Rektorat Universitas) 0812-1826-6228 (Marketing UNIVBI)
 0733-3280300 (Bina Insan) 0852-3151-5800 (Admin UNIVBI)
 0733-3280200 (Pascasarjana) Admin@univbinainsan.ac.id univbinainsan.ac.id - pasca.univbinainsan.ac.id

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

Lampiran 7. Lembar Bimbingan Skripsi



LEMBAR BIMBINGAN SKRIPSI

Nama : YOGA YARKHAMSETIAWAN
 Nim : 2102020164
 Program Studi : INFORMATIKA
 Pembimbing 1 : DR. MUHAMMAD AKBAR, M.IT
 Pembimbing 2 : SATRIANANSYAH, M.KOM
 Judul : ANALISIS SENTIMEN PENGGUNA APLIKASI QUR'AN KEMENAG PADA ULASAN GOOGLE PLAY STORE MENGGUNAKAN METODE *SUPPORT VECTOR MACHINE* (SVM)

NO	TANGGAL	TOPIK	KOMENTAR PEMBIMBING	TANDA TANGAN PEMBIMBING	
				1	2
1.	19-01-25	Skripsi	<ul style="list-style-type: none"> * perbaiki sesuai catatan. * perbaiki hasil dan pembahasan. Detail harus berada di paragraf. * tambahkan jurnal pembimbing. 		
2.	15-01-25	Skripsi	<ul style="list-style-type: none"> * perbaiki Abstrak. * lengkapi semuanya; <ul style="list-style-type: none"> - Data pendahuluan - dll. * Aa. Lanjut pl. 		

Lubuklinggau, ..16. Januari.....2025
Ketua Program Studi Informatika

(Budi Santoso, M.Kom)

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



LEMbaran KEMERIAAN SKRIPSI

Nama : YOGA YARKHAMUSLAWAN
 Nim : 2102020164
 Program Studi : INFORMATIKA
 Pembimbing 1 : DR. MUHAMMAD AKBAR, M.IT
 Pembimbing 2 : SATRIANANSYAH, M.KOM
 Judul : ANALISIS SENTIMEN PENGGUNA APLIKASI QUR'AN KEMENAG PADA
 ULASAN GOOGLE PLAY STORE MENGGUNAKAN METODE *SUPPORT VECTOR
 MACHINE (SVM)*

NO	TANGGAL	TOPIK	KOMENTAR PEMBIMBING	TANDA TANGAN PEMBIMBING	
				1	2
1.	16-01-25	skripsi	Penentuan label $\begin{array}{r l} 54 & 3 \ 21 \\ + & N \end{array} = 3 \text{ kelas}$ Kohen kappa = 2 kelas perlu ada jurnal		
			Perbaiki, CM. $\begin{array}{r l} 47 & 3 \\ 1 & 532 \end{array} \checkmark$		
2.	17-01-25	Skripsi	CM. 04. Ane ujian		

Lubuklinggau, 18 Januari 2024
 Ketua Program Studi Informatika

(Budi Santoso, M.Kom)

Protected by PDF Anti-Copy Free

(Upgrade to Pro Version to Remove the Watermark)

Lampiran 8. Tabulasi Data Proposal

	precis	recall	f1-score	support
negatif	0.62	0.28	0.39	100
positif	0.87	0.96	0.91	483
accuracy			0.85	583
macro avg	0.74	0.62	0.65	583
weighted avg	0.82	0.85	0.82	583

Protected by PDF Anti-Copy Free

(Upgrade to Pro Version to Remove the Watermark)

Lampiran 9. Tabulasi Data Skripsi

Classification Report for SMOTE:

	precision	recall	f1-score	support
negatif	0.95	0.98	0.97	483
positif	0.98	0.95	0.97	483
accuracy			0.97	966
macro avg	0.97	0.97	0.97	966
weighted avg	0.97	0.97	0.97	966

Protected by PDF Anti-Copy Free

(Upgrade to Pro Version to Remove the Watermark)

Lampiran 10. Dokumentasi Proposal



Protected by PDF Anti-Copy Free

(Upgrade to Pro Version to Remove the Watermark)

Lampiran 11. Dokumentasi Skripsi



Protected by PDF Anti-Copy Free

(Upgrade to Pro Version to Remove the Watermark)

Lampiran 12. Surat Keterangan Bebas Plagiasi



SURAT KETERANGAN BEBAS PLAGIASI

Menerangkan bahwa mahasiswa :

Nama : Yoga Yarkhamsetiawan

NIM : 2102020164

Fakultas : Fakultas Ilmu Teknik

Program Studi : Informatika

Memiliki jurnal dengan Judul “**Analisis Sentimen Pengguna Aplikasi Qur'an Kemenag Pada Ulasan Google Playstore Menggunakan Metode Support Vector Machine (SVM)**” Telah dilakukan uji plagiasi dengan turmin sebesar 27% dari batas maksimal 30% dan dinyatakan memenuhi standar bebas plagiasi dari Universitas Bina Insan.

Demikian surat keterangan ini disampaikan dengan sebenarnya untuk dapat dipergunakan sebagaimana mestinya.

Lubuklinggau, 16 Mei 2025

Kepala LPPM



Dr. Ahmad Fahrudin, M.Pd

Protected by PDF Anti-Copy Free

(Upgrade to Pro Version to Remove the Watermark)

Lampiran 13. Jurnal



ANALISIS SENTIMEN PENGGUNA APLIKASI QUR'AN KEMENAG PADA ULASAN GOOGLE PLAY STORE MENGGUNAKAN METODE SUPPORT VECTOR MACHINE (SVM)

Yoga_Yarkhamsetiawan¹, Muhamad_Akbar², Satrianansyah³

¹Program Studi, Universitas Bina Insan, LubukLinggau, Indonesia

^{2,3}Program Studi, Universitas Bina Insan, LubukLinggau, Indonesia

Email: ¹realyogays@gmail.com, ²muhamad.akbar@univbinainsan.ac.id,

³Satrianansyah@univbinainsan.ac.id

Abstrak

Aplikasi Qur'an Kemenag dirancang untuk memudahkan umat Muslim membaca dan mempelajari Al-Qur'an. Meskipun memiliki rating 4.6 di Google *Play Store* dengan total unduhan melebihi 500 ribu, aplikasi ini menerima keluhan terkait bug dan performa. Tujuan dari penelitian ini adalah untuk mengevaluasi sentimen dalam ulasan pengguna terhadap aplikasi Qur'an Kemenag dengan menggunakan algoritma *Support Vector Machine* (SVM). Sebanyak 3.000 ulasan diambil dari Google Play Store menggunakan metode web scraping. Ulasan dengan rating bintang tiga dikeluarkan dari proses pelabelan karena bersifat netral dan tidak mewakili sentimen positif maupun negatif secara jelas. Setelah proses preprocessing untuk menghilangkan noise dalam data, tersisa 2.677 ulasan. Ketidakseimbangan data ditangani dengan *Synthetic Minority Oversampling Technique* (SMOTE). Data dilabeli menjadi dua kategori: positif dan negatif. Proses klasifikasi menerapkan SVM yang menggunakan kernel *Radial Basis Function* (RBF) dan pembagian data pelatihan serta pengujian sebesar 80:20. Model menunjukkan akurasi sebesar 91% sebelum SMOTE dan meningkat menjadi 97% setelah SMOTE diterapkan. Analisis menunjukkan bahwa Kata-kata seperti 'bagus', 'terima kasih', dan 'alhamdulillah' paling sering muncul dalam ulasan positif sedangkan ulasan negatif lebih banyak menyoroti isu seperti 'buka ayat', 'halaman,' dan 'versi.' Hasil penelitian ini merekomendasikan pengembang untuk meningkatkan kualitas aplikasi melalui optimalisasi fitur, perbaikan tampilan halaman, dan peningkatan pengalaman pengguna.

Kata kunci— Analisis Sentimen, Qur'an Kemenag, *Support Vector Machine*, SMOTE, Ulasan Aplikasi.

Abstract

The *Kemenag Qur'an* application is designed to make it easier for Muslims to read and study the *Qur'an*. Despite having a rating of 4.6 on the *Google Play Store* with a total download of more than 500 thousand, this application has received complaints regarding bugs and performance. The purpose of this study was to evaluate sentiment in user reviews of the *Kemenag Qur'an* application using the *Support Vector Machine* (SVM) algorithm. A total of 3,000 reviews were taken from the *Google Play Store* using the web scraping method. Reviews with a three-star rating were excluded from the labeling process because they were neutral and did not clearly represent positive or negative sentiments. After the preprocessing process to remove noise in the data, 2,677 reviews remained. Data imbalance was handled using the *Synthetic Minority Oversampling Technique* (SMOTE). Data was labeled into two categories: positive and negative. The classification process applied SVM using the *Radial Basis Function* (RBF) kernel and a training and testing data division of 80:20. The model showed an accuracy of 91% before SMOTE and increased to 97% after SMOTE was applied. The analysis shows that words such as 'good,' 'thank you,' and 'alhamdulillah' appear most often in positive reviews while negative reviews highlight issues such as 'open verse,' 'page,' and 'version.' The results of this study recommend that developers improve the quality of applications through feature optimization, improving page appearance, and enhancing user experience.

Protected by PDF Anti-Copy Free

(Upgrade to Pro Version to Remove the Watermark)



Keywords— *Sentiment Analysis*, *Application Review*



Application Review, *Support Vector Machine*, *SMOTE*,

I. PENDAHULUAN

Kemajuan teknologi digital telah mengubah cara umat Muslim mengakses dan mempelajari Al-Qur'an. "Pada akhir Agustus 2016, Kementerian Agama, dalam hal ini Lajnah Pentashihan Mushaf Al-Qur'an, meluncurkan produk Al-Qur'an digital bernama "Qur'an Kemenag"[1]. Aplikasi berbasis Android ini, yang mudah diunduh dari *Google Play Store*, mempermudah proses membaca, mencari ayat, serta mencari informasi tanda tajwid dan asbabun nuzul. Meskipun memperoleh rating tinggi (4,6 bintang dari lebih dari 7 ribu ulasan) dan jumlah unduhan telah melebihi 500 ribu kali, aplikasi ini tetap mendapatkan sejumlah keluhan terkait pengalaman pengguna, seperti adanya bug pada beberapa fitur. Oleh karena itu, dibutuhkan penelitian analisis sentimen terhadap ulasan pengguna untuk mengetahui tingkat kepuasan maupun kekecewaan, dengan pendekatan algoritma *Support Vector Machine* (SVM).

Meskipun analisis sentimen telah banyak digunakan di berbagai bidang, penelitian khusus mengenai aplikasi Qur'an Kemenag masih sangat terbatas. Berbagai penelitian sebelumnya mengindikasikan bahwa algoritma SVM memiliki kinerja yang lebih baik dalam mengelompokkan sentimen pengguna dibandingkan pendekatan lainnya. Sebagai contoh, riset yang dilakukan oleh Andry Novantika dan Sugiman mengungkap bahwa penerapan SVM dengan kernel linear mampu mencapai tingkat akurasi sebesar 87,02%, melebihi *Logistic Regression* yang hanya memperoleh 85,17%, sehingga membuktikan efektivitas SVM dalam klasifikasi ulasan pada aplikasi *Google Meet*. [2]. Selanjutnya, eksperimen oleh Ulfa Kusnia dan Fachrul Kurniawan mengindikasikan bahwa SVM memiliki akurasi 88% dibandingkan dengan 87% yang dicapai oleh *Naive Bayes*[3]. Di sisi lain, studi yang dilakukan oleh Rokhman dan rekan-rekannya mengindikasikan bahwa algoritma SVM berhasil mencapai tingkat akurasi sebesar 90,20%, sedangkan *Decision Tree* hanya memperoleh 89,80%. [4]. Studi-studi tersebut menggambarkan *state of the art* dalam penerapan SVM untuk analisis sentimen, namun penerapan metode ini pada aplikasi Qur'an Kemenag masih jarang dilakukan, sehingga terdapat celah penelitian yang perlu diisi.

Penelitian ini difokuskan pada pengkajian sentimen pengguna terhadap aplikasi Qur'an milik Kementerian Agama berdasarkan ulasan yang tersedia di platform *Google Play Store*, menggunakan pendekatan algoritma SVM. Tidak seperti penelitian sebelumnya yang umumnya hanya menggunakan metode klasifikasi dasar, studi ini menyempurnakan model dengan menerapkan teknik SMOTE untuk menangani distribusi label yang tidak seimbang, sehingga hasil klasifikasinya diharapkan mampu membedakan sentimen positif dan negatif secara lebih presisi. *Fenomena gap* yang menjadi fokus adalah bahwa, meskipun aplikasi Qur'an Kemenag memiliki rating tinggi serta telah mencapai lebih dari 500 ribu unduhan, masih terdapat keluhan dari pengguna terkait bug, performa, dan fitur aplikasi. Oleh karena itu, penelitian ini berupaya memberikan wawasan mendalam kepada pengembang untuk meningkatkan kualitas dan pengalaman pengguna. Berdasarkan latar belakang tersebut, penulis memutuskan untuk mengangkat topik ini sebagai tugas akhir dengan judul: "**Analisis Sentimen Pengguna Aplikasi Qur'an Kemenag Pada Ulasan Google Play Store Menggunakan Metode Support Vector Machine (SVM)**".

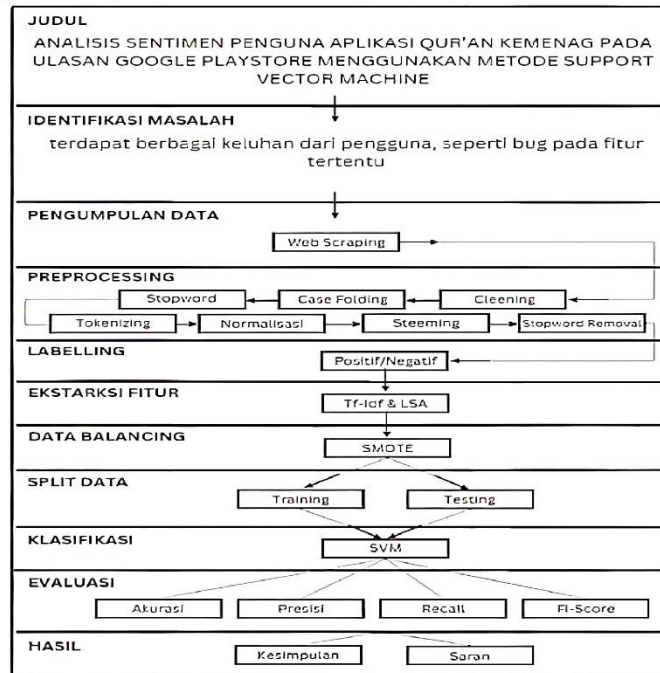
Protected by PDF Anti-Copy Free

(Upgrade to Pro Version to Remove the Watermark)



II. METODOLOGI PENELITIAN

“Dalam penelitian ini, penulis menggunakan beberapa metode untuk memperoleh data dan informasi yang digunakan untuk memecahkan masalah yang ada”[5]. Berikut adalah tahapan penelitian yang telah dilakukan untuk menyelesaikan permasalahan yang ada:



Sumber: Data Diolah, 2025

Gambar 1. Tahapan Penelitian

2.1 Pengumpulan Data

Langkah awal dalam proses pengolahan data dimulai dengan menghimpun ulasan dari pengguna aplikasi Qur'an Kemenag yang tersedia di platform Play Store. Proses pengambilan data ini dilakukan melalui metode scraping menggunakan pustaka seperti *BeautifulSoup* maupun *Selenium*. “*Web scraping* adalah proses pengumpulan data web terstruktur secara otomatis yang dilakukan oleh aplikasi khusus atau kode pemrograman. Data yang terkumpul bisa berasal dari ribuan, jutaan bahkan miliaran data di dunia maya”[6].

2.2 Preprocessing dan Labelling

Setelah data ulasan pengguna berhasil dikumpulkan, langkah berikutnya adalah melakukan *preprocessing*. Tahap *preprocessing* merupakan proses di mana aplikasi memilih data yang akan diproses dari setiap dokumen. Beberapa langkah yang termasuk dalam tahapan *preprocessing* adalah sebagai berikut:

- a) *Cleaning* adalah penghapusan karakter non-alfanumerik untuk mengurangi kebisingan dalam teks. Hal ini mencakup penghapusan tanda baca, simbol seperti @ pada nama pengguna, tag hash (#), emotikon, dan URL.

Protected by PDF Anti-Copy Free

(Upgrade to Pro Version to Remove the Watermark)



- b) *Case Folding* adalah tahap dimana karakter dalam teks dikonversi menjadi huruf kecil, serta menghilangkan tanda baca dan angka.
- c) *Stopword* adalah langkah yang menghapus kata-kata yang tidak memiliki makna penting, seperti kata sambung atau kata penjelas yang tidak berperan dalam analisis.
- d) *Tokenizing* adalah proses pemecahan teks menjadi bagian-bagian terkecil, yaitu kata-kata, berdasarkan pemisahan spasi dalam kalimat.
- e) *Normalisasi* adalah proses memperbaiki kata-kata yang tidak baku, dengan tujuan untuk memperbaiki kesalahan ejaan atau bentuk kata agar memiliki makna yang serupa.
- f) *Stemming* adalah proses penghapusan imbuhan pada kata. Imbuhan dalam bahasa Indonesia seperti me-, mem-, meng-, di-, dan sebagainya, dihilangkan untuk mendapatkan bentuk dasar kata.
- g) *Stopword Removal* adalah langkah yang bertujuan untuk menghapus kata-kata yang tidak berkontribusi pada makna utama, seperti kata penghubung atau kata penjelas.

Setelah preprocessing, dilakukan labelling untuk mengklasifikasikan sentimen ulasan[8]. Ulasan dengan bintang 4-5 diberi label positif, sedangkan bintang 1-2 diberi label negatif. Bintang 3 dihapus karena dianggap netral.

2.3 Ekstraksi Fitur

Proses ekstraksi fitur bertujuan untuk memilih elemen-elemen yang paling relevan dalam analisis sentimen. Dalam penelitian ini, dua pendekatan yang diterapkan untuk ekstraksi fitur adalah *Term Frequency-Inverse Document Frequency* (TF-IDF) dan *Latent Semantic Analysis* (LSA). "Pembobotan TF-IDF (*Term Frequency-Inverse Document Frequency*) adalah metode untuk mengkonversi data teks ke dalam format numerik dan memberikan bobot pada setiap kata atau fitur"[8]. "LSA digunakan untuk mengurangi dimensi data yang dihasilkan oleh TF-IDF dengan teknik *Singular Value Decomposition* (SVD)"[9].

2.4 Data Balancing dan Split Data

Untuk menangani perbedaan proporsi kelas dalam dataset, metode oversampling berbasis SVM-SMOTE diterapkan. Ketidakseimbangan kelas terjadi ketika terdapat distribusi yang tidak merata antara kelas mayoritas dan kelas minoritas. Apabila hal ini tidak diperhatikan, model bisa lebih condong kepada kelas mayoritas, yang mengakibatkan ketidaktepatan dalam mengenali kelas minoritas yang mungkin menyimpan informasi penting untuk prediksi. "Karena kelas mayoritas berkuasa, akurasi model dapat sangat tinggi, tetapi hasil prediksi untuk kelas minoritas akan sangat rendah"[8]. Selanjutnya, Pembagian dataset menjadi data latih dan data uji merupakan langkah penting dalam proses pengembangan model sentimen. Data latih digunakan untuk membangun model, sementara data uji berfungsi untuk mengevaluasi kinerjanya. "Penelitian ini menerapkan rasio pembagian 80:20, mengikuti studi sebelumnya yang menggunakan algoritma SVM dan pembobotan TF-IDF"[8].

2.5 Klasifikasi dan Evaluasi

"*Support Vector Machine* dapat dengan dijelaskan bahwa algoritma yang digunakan sebagai upaya untuk menemukan hyperplane optimal yang bertindak sebagai pembatas untuk dua kelas di ruang input"[10]. "Penelitian ini mengadopsi SVM sebagai metode klasifikasi utama, mengingat keunggulannya yang berakar pada teori pembelajaran statistika dan performanya yang sering melampaui metode lain"[8]. Model SVM bekerja dengan menghitung fungsi keputusan:

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



$$f(x) = \text{sign}(w \cdot x + b) \quad (1)$$

Penjelasan :

- $f(x)$: Merupakan fungsi keputusan yang digunakan untuk menentukan apakah ulasan tersebut termasuk kategori positif atau negatif.
- w : adalah vektor bobot yang diperoleh setelah proses pelatihan model.
- x : adalah vektor fitur yang merepresentasikan teks ulasan setelah proses ekstraksi fitur menggunakan TF-IDF.
- b : adalah bias yang membantu menentukan margin pemisah antara kelas.
- $\text{sign}(\cdot)$: adalah fungsi tanda yang mengeluarkan label kelas berdasarkan nilai yang dihitung.

“Tahap evaluasi bertujuan mengukur akurasi model yang telah dilatih pada dataset pelatihan” [8]. Untuk mengevaluasi kinerja model, digunakan *Confusion Matrix*, *Classification Report*, dan *Word Cloud*. *Confusion Matrix* berfungsi untuk mengukur akurasi model, sementara *Classification Report* menyediakan informasi tentang akurasi, presisi, recall, dan F1-score. *Word Cloud* digunakan sebagai alat untuk menggambarkan kata-kata yang paling sering muncul dalam ulasan positif dan negatif. Metrik evaluasi dihitung sebagai berikut:

Tabel 1. Confusion Matrix

		Actual Values	
		0 (Negatif)	1 (Positif)
Prediksi		TN <i>(True Negative)</i>	FP <i>(False Positive)</i>
		FN <i>(False Negative)</i>	TP <i>(True Positive)</i>

Sumber: Data Diolah, 2025

Keterangan :

- TP : Ini adalah contoh yang diklasifikasikan dengan benar sebagai positif.
- FP : Ini adalah contoh yang salah diklasifikasikan sebagai positif.
- TN : Ini adalah contoh yang diklasifikasikan dengan benar sebagai negatif.
- FN : Ini adalah contoh yang salah diklasifikasikan sebagai negatif.

Berdasarkan nilai-nilai tersebut, perhitungan metrik evaluasi dilakukan sebagai berikut:

- a) Akurasi : Mengukur seberapa baik model memprediksi kelas sebenarnya dari data uji.

$$\text{Akurasi} = \frac{TP + TN}{TP + TN + FP + FN} \quad (2)$$

- b) Presisi : Mengukur seberapa baik model memprediksi kelas positif dengan benar.

$$\text{Presisi} = \frac{TP}{TP + FP} \quad (3)$$

- c) Recall : Mengukur seberapa baik model mendeteksi kelas positif.

$$\text{Recall} = \frac{TP}{TP + FN} \quad (4)$$

- d) F1-Score : Merupakan rata-rata harmonik antara presisi dan recall, memberikan keseimbangan antara kedua metrik tersebut.

$$\text{F1-Score} = 2 \times \frac{\text{Presisi} \times \text{Recall}}{\text{Presisi} + \text{Recall}} \quad (5)$$

Hasil evaluasi ini membantu mengidentifikasi kekuatan dan kelemahan model dalam mengklasifikasikan sentimen secara optimal.

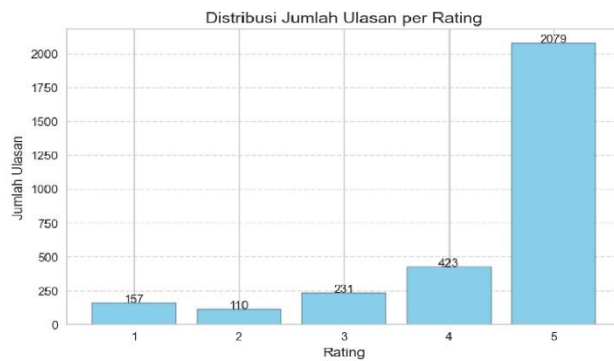
Protected by PDF Anti-Copy Free
(Upgrade to Pro Version to Remove the Watermark)



III. HASIL DAN PEMBAHASAN

3.1 Pengumpulan Data

Ulasan pengguna aplikasi Qur'an Kemenag diperoleh dengan memanfaatkan metode pengambilan data otomatis (*web scraping*) menggunakan pustaka *google_play_scraper* dalam bahasa pemrograman Python. Dengan memanfaatkan parameter seperti jumlah ulasan, bahasa, negara, dan metode penyortiran, penelitian ini berhasil memperoleh 3.000 ulasan dari akhir bulan agustus tahun 2016 sampai bulan februari tahun 2025 yang akan digunakan untuk analisis sentimen. Gambar berikut menunjukkan distribusi ulasan yang dikumpulkan dari Google Play Store:



Sumber: Data Diolah, 2025

Gambar 2. Hasil Pengumpulan Data

3.2 Preprocessing dan Labelling

Dalam penelitian ini, dilakukan tahap praproses dan pelabelan terhadap data tanggapan pengguna dari aplikasi Qur'an milik Kementerian Agama yang tersedia di platform distribusi aplikasi Android, guna keperluan analisis sentimen dengan algoritma *Support Vector Machine* (SVM). *Preprocessing* bertujuan meningkatkan kualitas data, sementara *labelling* dilakukan berdasarkan skor ulasan: skor 4-5 diberi label "positif", skor 1-2 diberi label "negatif", dan skor 3 dihapus karena dianggap ambigu antara positif dan negatif. Gambar berikut menunjukkan hasil *preprocessing* dan *labelling* data.

content	text_clean	text_case	text_stopward	text_tokens	normalisasi	text_stemindo	text_Stopward	label
Kok setelah update malah error. fungsi bookmark...	Kok setelah update malah error. fungsi bookmark...	kok setelah update malah error. fungsi bookmark...	update error fungsi bookmark bermasalah downlo...	[update, error, fungsi, bookmark, bermasalah, downlo...	[update, error, fungsi, bookmark, bermasalah, downlo...	update error fungsi bookmark masalah download	update error fungsi bookmark download resource...	negatif
Perbaiki kompas dan adakan fitur GPS dengan di...	Perbaiki kompas dan adakan fitur GPS dengan di...	perbaiki kompas dan adakan fitur gps dengan di...	perbaiki kompas adakan fitur gps disesuaikan j...	[perbaiki, kompas, adakan, fitur, gps, disesuaikan...	[perbaiki, kompas, adakan, fitur, gps, disesuaikan...	baik kompas adakan fitur gps sesuai jam bmkq	kompas adakan fitur gps sesuai jam bmkq	negatif
Bagus yang dulu bisa didownload, sekarang harus...	Bagus yang dulu bisa didownload, sekarang harus...	bagus yang dulu bisa didownload sekarang harus...	bagus didownload pake data	[bagus, didownload, pake, data]	[bagus, didownload, pake, data]	bagus didownload pake data	bagus didownload pake data	negatif
Alhamdulillah, app yg sangat bermanfaat bagi u...	Alhamdulillah, app yg sangat bermanfaat bagi um...	alhamdulillah app yg sangat bermanfaat bagi um...	alhamdulillah app yg bermanfaat umat islam yg	[alhamdulillah, app, yg, bermanfaat, umat, isl...]	[alhamdulillah, aplikasi, yang, bermanfaat, um...	alhamdulillah aplikasi yang manfaat umat islam...	alhamdulillah aplikasi manfaat umat islam lan...	positif
Mantap nih aplikainya top markotop	Mantap nih aplikainya top markotop	mantap nih aplikainya top markotop	mantap nih aplikainya top markotop	[mantap, nih, aplikainya, top, markotop]	[mantap, nih, aplikainya, top, markotop]	mantap nih aplikainya top markotop	mantap nih aplikainya top markotop	positif
Baik dan mudah dibaca	Baik dan mudah dibaca	baik dan mudah dibaca	mudah dibaca	[mudah, dibaca]	[mudah, dibaca]	mudah baca	mudah baca	positif

Sumber: Data Diolah, 2025

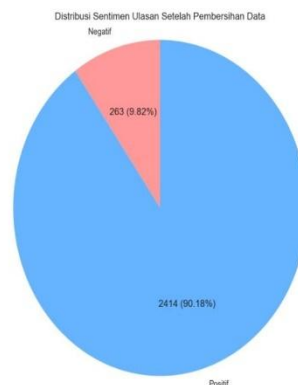
Gambar 3. Hasil Preprocessing dan Labelling

Protected by PDF Anti-Copy Free

(Upgrade to Pro Version to Remove the Watermark)



Dataset awal mencakup 2.770 entri ulasan. Oleh karena itu, hanya penilaian berbintang 4 dan 5 yang diklasifikasikan sebagai berisi sentimen positif, sementara penilaian berbintang 1 dan 2 dianggap merepresentasikan sentimen negatif. Setelah penyaringan, diperoleh total 2.770 ulasan dengan komposisi sekitar 90,5% positif dan 9,6% negatif. Setelah dilakukan pembersihan data dengan menghapus baris-baris yang mengandung *missing values*, jumlah data tersisa menjadi 2.677 ulasan. Berikut Gambar 4. menunjukkan distribusi sentimen setelah pembersihan data:



Sumber: Data Diolah, 2025

Gambar 4. Distribusi Sentimen Ulasan Setelah Pembersihan Data

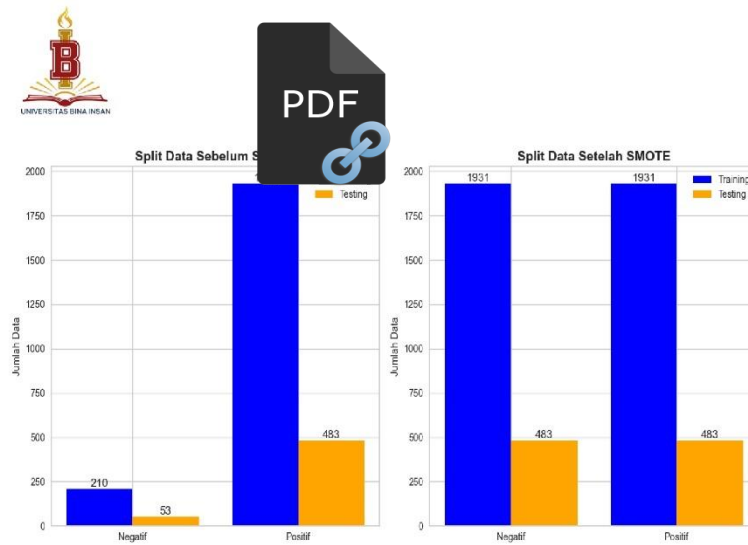
3.3 Ekstraksi Fitur

Pada tahap ekstraksi fitur, teks ulasan diubah menjadi representasi numerik untuk mendukung proses klasifikasi sentimen. Proses ini dilakukan dalam dua langkah utama, yaitu transformasi dengan TF-IDF dan reduksi dimensi dengan LSA. Representasi teks diubah menjadi bentuk numerik menggunakan pendekatan TF-IDF, yang dibatasi pada 1.500 fitur. Proses ini menghasilkan matriks yang menggambarkan bobot setiap kata dalam tiap dokumen. Tahap berikutnya adalah reduksi dimensi yang dilakukan menggunakan metode LSA melalui dekomposisi SVD, dengan mempertahankan 500 komponen utama. Hasil dari proses tersebut berupa vektor fitur yang lebih ringkas namun tetap mengandung informasi penting, yang kemudian digunakan sebagai dasar dalam proses analisis sentimen untuk membedakan antara tanggapan positif dan negatif.

3.4 Data Balancing dan Split Data

Distribusi ulasan setelah pembersihan (Gambar 4.) menunjukkan dataset yang tidak seimbang antara sentimen positif dan negatif. Dari total 2.677 data, terdapat 2.414 ulasan positif dan hanya 263 ulasan negatif. Pembagian data dilakukan dengan rasio 80% untuk *training* dan 20% untuk *testing*, di mana sentimen positif mendominasi (1.931 data *training* dan 210 data *testing*) sementara sentimen negatif hanya terdiri dari 483 data *training* dan 53 data *testing*. Ketidakseimbangan ini berpotensi menyebabkan model bias terhadap kelas mayoritas dan mengurangi akurasi prediksi pada data negatif. Sebagai solusi atas ketidakseimbangan data, digunakan metode SMOTE (*Synthetic Minority Over-sampling Technique*) yang bertujuan menghasilkan data sintesis pada kelompok dengan jumlah data lebih sedikit, sehingga distribusi menjadi lebih proporsional dan pelatihan model dapat berlangsung secara lebih optimal. Berikut Gambar 5. menunjukkan perbandingan *split data* sebelum dan sesudah SMOTE:

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

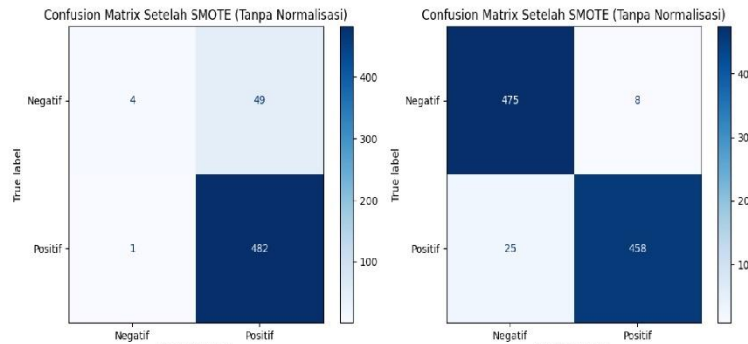


Sumber: Data Diolah, 2025
Gambar 5. Perbandingan *Split Data* Sebelum dan Sesudah SMOTE

3.5 Klasifikasi dan Evaluasi

Proses klasifikasi sentimen terhadap ulasan aplikasi Qur'an Kemenag memanfaatkan algoritma SVM yang menggunakan kernel *Radial Basis Function* (RBF), mengingat kemampuannya yang unggul dalam mengelola data dengan banyak fitur serta cocok untuk tugas klasifikasi berbasis teks. Model dikembangkan menggunakan data latih dan kemudian divalidasi dengan data uji guna menilai kinerjanya. Proses penilaian terhadap model dilakukan melalui:

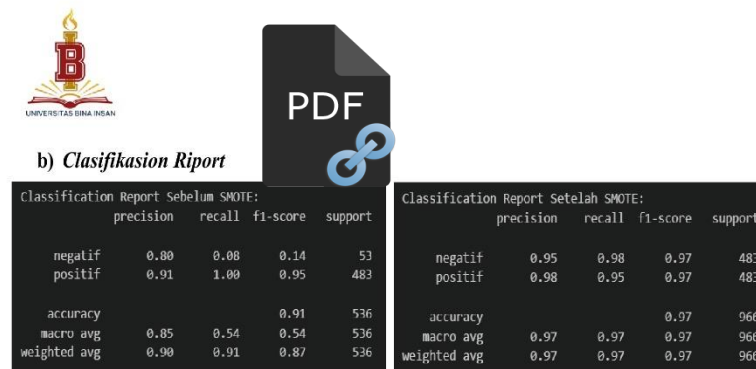
a) Confusion Matrix



Sumber: Data Diolah, 2025
Gambar 6. Confusion Matrix Sebelum dan Sesudah SMOTE

Gambar 6. menunjukkan bahwa sebelum SMOTE, model kesulitan mengenali kelas negatif, dengan hanya 4 prediksi benar, sementara 49 sampel negatif salah diklasifikasikan sebagai positif. Sebaliknya, model cukup akurat dalam mengenali kelas positif dengan 482 prediksi benar dan 1 kesalahan. Setelah SMOTE, prediksi benar kelas negatif meningkat drastis menjadi 475, meskipun terjadi peningkatan kesalahan klasifikasi pada kelas positif (25 sampel salah diklasifikasikan sebagai negatif). Temuan ini mengindikasikan bahwa penerapan SMOTE berhasil mengatasi masalah ketidakseimbangan data, serta meningkatkan akurasi pada kelas negatif, tanpa mengorbankan performa keseluruhan model.

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



Sumber: Data Diolah, 2025

Gambar 7. Classification Riport Sebelum dan Sesudah SMOTE

Gambar 7. menunjukkan perbandingan *Classification Report* sebelum dan sesudah penerapan SMOTE. Sebelum penerapan SMOTE, model menunjukkan ketidakseimbangan dalam mendeteksi kelas negatif, dengan nilai *precision* sebesar 0,80, *recall* yang rendah yaitu 0,08, dan *F1-Score* yang hanya mencapai 0,14. Di sisi lain, untuk kelas positif, nilai *precision* tercatat 0,91, *recall* mencapai 1,00, dan *F1-score* sebesar 0,95, dengan akurasi total model sebesar 0,91. Setelah SMOTE, performa model dalam mengenali kelas negatif meningkat signifikan, dengan *recall* naik menjadi 0,98, *F1-score* menjadi 0,97, dan *precision* meningkat menjadi 0,95. Pada kelas positif, *precision* naik menjadi 0,98, meskipun *recall* sedikit menurun menjadi 0,95. Akurasi total model mengalami peningkatan dari 0,91 menjadi 0,97, yang mengindikasikan bahwa penerapan SMOTE berhasil mengatasi ketidakseimbangan data tanpa menurunkan performa pada kelas mayoritas.

c) *Word Cloud*



Sumber: Data Diolah, 2025

Gambar 8. Word Cloud Sentimen Negatif dan Positif

Gambar 8. menunjukkan *Word Cloud* sentimen pengguna aplikasi Qur'an Kemenag. Pada *Word Cloud* merah (negatif), kata-kata seperti "aplikasi", "buka", "ayat", "halaman", "tolong", dan "offline" mendominasi, mengindikasikan kendala teknis seperti kesulitan membuka halaman dan fitur offline yang bermasalah. Kehadiran kata "mohon" dan "versi" mencerminkan permintaan pengguna untuk perbaikan aplikasi. Sebaliknya, *Word Cloud* hijau (positif) didominasi oleh kata "bagus", "alhamdulillah", "terima kasih", dan "manfaat", yang mencerminkan kepuasan pengguna terhadap fitur aplikasi. Kata "baca", "ayat", dan "quran" menunjukkan bahwa aplikasi ini memberikan manfaat dalam membaca Al-Qur'an. Meskipun aplikasi mendapatkan apresiasi, pengembang tetap perlu memperbaiki kendala teknis agar pengalaman pengguna semakin optimal.

Protected by PDF Anti-Copy Free

(Upgrade to Pro Version to Remove the Watermark)



IV. KESIMPULAN

Penelitian ini berhasil melakukan analisis sentimen terhadap ulasan yang diberikan oleh pengguna aplikasi Qur'an Kemnag di platform Google Play, menggunakan metode SVM yang dikombinasikan dengan kernel RBF. Hasil menunjukkan bahwa sentimen negatif didominasi oleh keluhan terkait fitur aplikasi, seperti kesulitan membuka halaman, keterbatasan fitur offline, dan permintaan perbaikan. Sementara itu, sentimen positif mencerminkan kepuasan terhadap fitur yang disediakan. Metode SVM mencapai akurasi 91% sebelum penerapan SMOTE dan meningkat menjadi 97% setelah SMOTE diterapkan untuk menangani ketidakseimbangan data. Hasil ini memberikan wawasan bagi pengembang dalam memperbaiki bug, meningkatkan performa aplikasi, serta mengoptimalkan fitur agar lebih sesuai dengan kebutuhan pengguna. Untuk pengembangan lebih lanjut, disarankan agar klasifikasi sentimen tidak hanya terbatas pada kategori positif dan negatif, tetapi juga mencakup sentimen netral agar memperoleh wawasan yang lebih komprehensif. Selain itu, penggunaan teknik *deep learning* seperti LSTM atau BERT bisa menjadi opsi yang menarik untuk meningkatkan akurasi dalam klasifikasi dan memperoleh pemahaman yang lebih mendalam mengenai pola sentimen.

V. DAFTAR PUSTAKA

- [1] Afif, Z., Fadly, H., Akbar, A., Syatri, J., Mustopa, M., Jaeni, A., ... & Musadad, M. (2018). PREFERENSI MASYARAKAT DALAM PENGGUNAAN AL-QUR'AN DIGITAL. *SUHUF*, 11(2), 185-214.
- [2] S. Sisnawati, R. Astuti, and F. Muhamad basysyar, "ANALISIS SENTIMEN PADA APLIKASI KFCKU Di GOOGLE PLAYSTORE MENGGUNAKAN NAÏVE BAYES," *JATI (Jurnal Mhs. Tek. Inform.,* vol. 8, no. 3, pp. 3010–3016, 2024, doi: 10.36040/jati.v8i3.8382.
- [3] I. H. Kusuma and N. Cahyono, "Analisis Sentimen Masyarakat Terhadap Penggunaan E-Commerce Menggunakan Algoritma K-Nearest Neighbor," *J. Inform. J. Pengemb. IT*, vol. 8, no. 3, pp. 302–307, 2023, doi: 10.30591/jpit.v8i3.5734.
- [4] H. Hidayatullah, P. Purwantoro, and Y. Umaidah, "Penerapan Naïve Bayes Dengan Optimasi Information Gain Dan Smote Untuk Analisis Sentimen Pengguna Aplikasi Chatgpt," *JATI (Jurnal Mhs. Tek. Inform.,* vol. 7, no. 3, pp. 1546–1553, 2023, doi: 10.36040/jati.v7i3.6887.
- [5] M. R. Fauzan, H. Oktafia, L. Wijaya, and J. Karman, "KENAIKAN HARGA BBM DI MEDIA SOSIAL TWITTER," vol. 1, no. 1, pp. 82–89, 2023.
- [6] D. Rudini, D. G. Purnama, and A. A. Khan, "Penggunaan Teknik Web Scraping dalam Aplikasi Pengambilan Data dari Google Maps untuk Menunjang Digital Marketing," *Lentera Multidiscip. Stud.*, vol. 2, no. 1, pp. 10–19, 2023, doi: 10.57096/lentera.v2i1.61.
- [7] R. Abdillah, E. Haerani, and R. M. Candra, "Analisis Sentimen Ulasan Aplikasi Wetu Untuk Peningkatan Layanan Menggunakan Metode Support Vector Machine," *J. Inf. Syst. Res.*, vol. 4, no. 3, pp. 865–873, 2023, doi: 10.47065/josh.v4i3.3353.
- [8] Wahyudi, Rudiman, and N. A. Verdikha, "Klasifikasi Sentimen X-Twitter Perihal Pemindahan Ibu Kota Indonesia Menggunakan Ekstraksi Fitur TF-IDF dan Metode Support Vector Machine (SVM)," *J. Teknol. Inf.*, vol. 18, no. 2, pp. 185–199, 2024.
- [9] D. A. Perkasa *et al.*, "Sistem Ujian Online Essay Dengan Penilaian Menggunakan Metode Latent Semantic Analysis (Lsa)," *J. Ilm. Rekayasa dan Manaj. Sist. Inf.*, vol. 1, no. 1, pp. 1–9, 2015, [Online]. Available: <https://ejournal.uin-suska.ac.id/index.php/RMSI/article/view/1313>
- [10] J. Teknik Elektro dan Komputasi *et al.*, "Implementasi Algoritma Naïve Bayes dan Support Vector Machine (SVM) Pada Klasifikasi Penyakit Kardiovaskular," *J. Tek. Elektro dan Komputasi*, vol. 4, no. 2, pp. 207–214, 2022, [Online]. Available: <http://jurnal.unmuhjember.ac.id/index.php/ELKOM/article/view/7691>

Protected by PDF Anti-Copy Free

(Upgrade to Pro Version to Remove the Watermark)

Lampiran 14. Listing Program



```

# 1. WEB SCRAPING
from google_play_scraper import reviews, Sort # Mengambil ulasan dari Google Play
import pandas as pd # Untuk pengolahan data
import numpy as np # Untuk operasi numerik
result, continuation_token = reviews(
    'com.quran.kemenag',
    lang='id',
    country='id',
    sort=Sort.MOST_RELEVANT,
    count=3000,
    filter_score_with=None
)

# 2. Membuat DataFrame dari hasil scraping
df_busu = pd.DataFrame(np.array(result), columns=['review'])
df_busu = df_busu.join(pd.DataFrame(df_busu.pop('review').tolist()))
new_df = df_busu[['userName', 'score', 'at', 'content']]
sorted_df = new_df.sort_values(by='at', ascending=False)

len(df_busu.index)

df_busu[['userName', 'score', 'at', 'content']].head()

new_df=df_busu[['userName', 'score', 'at', 'content']]
sorted_df = new_df.sort_values(by='at', ascending=False)
sorted_df.head()

my_df=sorted_df[['userName', 'score', 'at', 'content']]
my_df.head()

my_df.to_csv("cobascraping.csv", index = False) #Save the file as

# Membaca file CSV
my_df = pd.read_csv("cobascraping.csv")
# Menampilkan informasi DataFrame awal
print("Original DataFrame:")
print(my_df.info())
# Tahap pembersihan data
# 1. Mengatasi nilai yang hilang
my_df.dropna(inplace=True)
# 2. Menghapus duplikat
my_df.drop_duplicates(inplace=True)
# 3. Mengubah tipe data jika diperlukan
# Contoh: Mengubah kolom 'date_column' menjadi datetime
#my_df['date_column'] = pd.to_datetime(my_df['date_column'])
# 4. Langkah pembersihan lainnya sesuai kebutuhan (misalnya, menghapus outlier, memperbaiki typo, dll.)
# Menampilkan informasi tentang DataFrame yang telah dibersihkan
print("\nCleansed DataFrame:")
print(my_df.info())

# 1. CLEANING
import re
def clean_text(df, text_field, new_text_field_name):
    # Menghapus mention, URL, karakter khusus, angka, dan spasi berlebih
    df[new_text_field_name] = df[text_field].apply(lambda elem: re.sub(r"(@[A-Za-z0-9_]+)|(^0-9A-Za-z
\)|(\w+:\w+\S+)|\rt|http.+?", "", elem)) # Menghapus mention, URL, karakter khusus
    df[new_text_field_name] = df[new_text_field_name].apply(lambda elem: re.sub(r"\d+", "", elem)) # Menghapus angka
    df[new_text_field_name] = df[new_text_field_name].apply(lambda elem: re.sub(r"\s+", " ", elem).strip()) # Menghapus
    spasi berlebih
    # Menampilkan hasil output setelah cleaning
    return df

```

Protected by PDF Anti-Copy Free

(Upgrade to Pro Version to Remove the Watermark)

```
# Melakukan pembersihan teks
data_clean = clean_text(my_df, 'content', 'text_clean')
# Menampilkan hasil setelah pembersihan
data_clean.head(5)

# 2. CASE FOLDING
my_df['text_case'] = my_df['text_clean'].str.lower()
data_clean.head(5)

# 3. STOPWORD
import nltk.corpus
nltk.download('stopwords')
from nltk.corpus import stopwords
stop = stopwords.words('Indonesian')
data_clean['text_stopward'] = data_clean['text_case'].apply(lambda x: ''.join([word for word in x.split() if word not in (stop)]))
data_clean.head(5)

# 4. TOKENIZING
import nltk
nltk.download('punkt')
from nltk.tokenize import sent_tokenize, word_tokenize
data_clean['text_tokens'] = data_clean['text_stopward'].apply(lambda x: word_tokenize(x))
data_clean.head(5)

# 5. NORMALISASI
normalized_word = pd.read_csv("normalisasi.csv")
normalized_word_dict = {}
for index, row in normalized_word.iterrows():
    if row[0] not in normalized_word_dict:
        normalized_word_dict[row[0]] = row[1]
def normalized_term(document):
    return [normalized_word_dict[term] if term in normalized_word_dict else term for term in document]
data_clean['normalisasi'] = data_clean['text_tokens'].apply(normalized_term)
data_clean.head(5)

pip install sastrawi

# 6. STEMMING
from Sastrawi.Stemmer.StemmerFactory import StemmerFactory
# Membuat objek factory dan stemmer
factory = StemmerFactory()
stemmer = factory.create_stemmer()
# apply stemmed term to dataframe
def get_stemmed_term(document):
    return [stemmer.stem(term) for term in document] # Menggunakan stemmer langsung pada term
# Gunakan kolom 'normalisasi' untuk proses stemming
data_clean['text_steamindo'] = data_clean['normalisasi'].apply(lambda x: ''.join(get_stemmed_term(x)))
data_clean.head(5)
data_clean.to_csv('hasil_TextPreProcessing.csv', index=False)

# 7. STOPWORD REMOVAL
nltk.download("stopwords")
stop = stopwords.words('indonesian')
data_clean['text_Stopword'] = data_clean['text_steamindo'].apply(lambda x: ''.join([word for word in x.split() if word not in (stop)]))
data_clean.head(5)
df = pd.read_csv('hasil_TextPreProcessing.csv')
df.head(5)

#LABELLING
sentimen = []
for index, row in df.iterrows():
    if row['score'] in [4, 5]:
```



Protected by PDF Anti-Copy Free

(Upgrade to Pro Version to Remove the Watermark)

```

sentimen.append("positif")
elif row['score'] in [1, 2]:
    sentimen.append("negatif")
else:
    sentimen.append(None) # Untuk skor lain

data_clean['label'] = sentimen
data_clean.head(5)
sentimen_counts = data_clean.groupby(['label']).size()
print(sentimen_counts)
data_clean.to_csv('hasil_label.csv', index=False)
import matplotlib.pyplot as plt

# 1. Memuat dan membersihkan data
def load_data():
    data = pd.read_csv('hasil_label.csv')
    data = data.dropna(axis='rows') # Menghapus baris dengan nilai kosong
    return data

df = load_data()
df = df[['text_Stopword', 'label']] # Memastikan hanya kolom yang dibutuhkan

# 2. Memisahkan fitur (x) dan label (y)
x = df['text_Stopword']
y = df['label']
# 2. Mengonversi label ke dalam format numerik
df['label'] = df['label'].map({'negatif': 0, 'positif': 1})

from sklearn.model_selection import train_test_split
from imblearn.over_sampling import SVMSMOTE
from sklearn.feature_extraction.text import TfidfVectorizer
from sklearn.decomposition import TruncatedSVD
from sklearn.svm import SVC
from sklearn.metrics import classification_report, confusion_matrix, accuracy_score, ConfusionMatrixDisplay
# 4. Mengubah teks menjadi representasi numerik dengan TF-IDF
tfidf = TfidfVectorizer(max_features=1500) # Batasi jumlah fitur
x_tfidf = tfidf.fit_transform(x) # Terapkan TF-IDF ke seluruh dataset
# 5. Terapkan LSA untuk reduksi dimensi
lsa = TruncatedSVD(n_components=500) # Tentukan jumlah komponen untuk LSA
x_lsa = lsa.fit_transform(x_tfidf)
# Membagi data menjadi data training dan testing sebelum SMOTE
x_train_original, x_test_original, y_train_original, y_test_original = train_test_split(
    x_lsa, y, test_size=0.2, random_state=42, stratify=y
)
# Menampilkan jumlah data training dan testing sebelum SMOTE
print(f"Jumlah data training sebelum SMOTE: {x_train_original.shape[0]}")
print(f"Jumlah data testing sebelum SMOTE: {x_test_original.shape[0]}")
# Menampilkan distribusi label di data training dan testing sebelum SMOTE
from collections import Counter
print("Distribusi label di data training sebelum SMOTE:", Counter(y_train_original))
print("Distribusi label di data testing sebelum SMOTE:", Counter(y_test_original))

# 6. Terapkan SVMSMOTE untuk menangani ketidakseimbangan data
svm_smote = SVMSMOTE(random_state=42)
x_resampled, y_resampled = svm_smote.fit_resample(x_lsa, y)
# 7. Membagi data menjadi data latih dan data uji
x_train, x_test, y_train, y_test = train_test_split(
    x_resampled, y_resampled, test_size=0.2, random_state=42, stratify=y_resampled
)

# Menampilkan jumlah data training dan testing sebelum SMOTE
print(f"Jumlah data training setelah SMOTE: {x_train.shape[0]}")
print(f"Jumlah data testing setelah SMOTE: {x_test.shape[0]}")
# Menampilkan distribusi label di data training dan testing sebelum SMOTE

```



Protected by PDF Anti-Copy Free

(Upgrade to Pro Version to Remove the Watermark)

```

print("Distribusi label di data training setelah SMOTE:", Counter(y_train))
print("Distribusi label di data testing setelah SMOTE:", Counter(y_test))

# KLASIFIKASI SVM
svm_model = SVC(kernel='rbf', C=1, gamma=0.001, probability=True, class_weight='balanced')

# Melatih model dengan data latih
svm_model.fit(x_train, y_train)

# Memprediksi data pengujian
y_pred_test = svm_model.predict(x_test) # SVM

#EVALUASI
# Membuat confusion matrix
clr = confusion_matrix(y_test, y_pred_test)
clr

from sklearn.metrics import ConfusionMatrixDisplay
import matplotlib.pyplot as plt
# Menampilkan confusion matrix menggunakan hasil prediksi dari SVM dan data uji LSA
titles_options = [("Confusion matrix, tanpa normalisasi", None),
                  ("Confusion matrix, dengan normalisasi", 'true')]
for title, normalize in titles_options:
    disp = ConfusionMatrixDisplay.from_estimator(
        svm_model, x_test, y_test,
        display_labels=["Negatif", "Positif"],
        cmap=plt.cm.Blues,
        normalize=normalize
    )
    disp.ax_.set_title(title)
    plt.show()

# Menghitung akurasi untuk setiap kelas
import numpy as np
clrr = clr.astype('float') / clr.sum(axis=1)[:, np.newaxis]
clrr.diagonal()

# Menghitung Akurasi Total
from sklearn.metrics import accuracy_score
acc_score = accuracy_score(y_test, y_pred_test)
acc_score

# Impor library yang diperlukan
from sklearn.metrics import confusion_matrix
# Menghitung confusion matrix untuk data training
cm = confusion_matrix(y_test.ravel(), y_pred_test.ravel())
# Menghitung True Negative, False Negative, True Positive, dan False Positive
tn = cm[0][0] # True Negative
fn = cm[1][0] # False Negative
tp = cm[1][1] # True Positive
fp = cm[0][1] # False Positive
# Menghitung Precision, Recall, Specificity, FPR, FNR, dan F1-Score
pre = tp / (tp + fp) if (tp + fp) > 0 else 0 # Precision
rec = tp / (tp + fn) if (tp + fn) > 0 else 0 # Recall (Sensitivity)
spe = tn / (tn + fp) if (tn + fp) > 0 else 0 # Specificity
fal = fp / (fp + tn) if (fp + tn) > 0 else 0 # Fall Out (False Positive Rate)
mis = fn / (fn + tp) if (fn + tp) > 0 else 0 # Miss Rate (False Negative Rate)
f1 = (2 * rec * pre) / (rec + pre) if (rec + pre) > 0 else 0 # F1-Score
# Menampilkan hasil
print("Precision untuk setiap kelas: ", pre)
print("Recall (Sensitivity) untuk setiap kelas: ", rec)
print("Specificity (True Negative Rate) untuk setiap kelas: ", spe)
print("Fall Out (False Positive Rate) untuk setiap kelas: ", fal)
print("Miss Rate (False Negative Rate) untuk setiap kelas: ", mis)

```

Protected by PDF Anti-Copy Free

(Upgrade to Pro Version to Remove the Watermark)

```

print("F1-Score untuk setiap kelas: ", f1)

print(classification_report(y_test, y_pred_test))

from sklearn.preprocessing import LabelEncoder
from yellowbrick.classifier import ROCAUC
# Konversi label string ke numerik
label_encoder = LabelEncoder()
y_train_numeric = label_encoder.fit_transform(y_train) # 0 = negatif, 1 = positif
y_test_numeric = label_encoder.transform(y_test)
# Latih model dengan label numerik
svm_model.fit(x_train, y_train_numeric)
# 5. Visualisasi ROC Curve
visualizer = ROCAUC(svm_model, classes=["negatif", "positif"])
visualizer.fit(x_train, y_train_numeric) # Data pelatihan
visualizer.score(x_test, y_test_numeric) # Data pengujian
visualizer.show()

# Membuat grafik cross-validation
from sklearn.model_selection import cross_val_score
import matplotlib.pyplot as plt
import numpy as np
# Perform cross-validation with 10 folds.
cv_scores = cross_val_score(svm_model, x_train, y_train, cv=10)
# Plot the cross-validation scores
plt.figure(figsize=(10, 6))
plt.plot(np.arange(1, 11), cv_scores, marker='o', linestyle='-', color='b', label='Cross-validation score')
# Annotating the score at each point (fold).
for i, score in enumerate(cv_scores):
    plt.text(i + 1, score, f'{score:.3f}', ha='center', va='bottom')
# Plot formatting
plt.title('Cross-Validation Scores for Support Vector Machine (10-Fold)')
plt.xlabel('Fold')
plt.ylabel('Score')
plt.xticks(np.arange(1, 11))
plt.grid(True)
plt.legend()
plt.show()

pip install wordcloud matplotlib
from wordcloud import WordCloud
import matplotlib.pyplot as plt

# Filter data positif
positive_data = data_clean[data_clean['label'] == 'positif']
# Gabungkan semua teks dari kolom 'text_steamindo' yang sudah melalui preprocessing
positive_text = " ".join(positive_data['text_Stopword'])
# Buat Word Cloud dengan warna hijau untuk kata-kata dan latar belakang hitam
wordcloud = WordCloud(width=800, height=400, background_color='black', colormap='Greens').generate(positive_text)
# Tampilkan Word Cloud tanpa margin putih
plt.figure(figsize=(10, 5), dpi=100)
plt.imshow(wordcloud, interpolation='bilinear')
plt.axis('off') # Hilangkan sumbu
plt.margins(0) # Hilangkan margin putih
plt.title("Word Cloud untuk Sentimen Positif", color='white') # Judul dengan warna putih
plt.show()

from wordcloud import WordCloud
import matplotlib.pyplot as plt
# Filter data negatif
negative_data = data_clean[data_clean['label'] == 'negatif']
# Gabungkan semua teks dari kolom 'text_steamindo' yang sudah melalui preprocessing
negative_text = " ".join(negative_data['text_Stopword'])
# Buat Word Cloud

```

Protected by PDF Anti-Copy Free

(Upgrade to Pro Version to Remove the Watermark)

```
wordcloud = WordCloud(width=800, height=400, background_color='black', colormap='Reds').generate(negative_text)
# Tampilkan Word Cloud
plt.figure(figsize=(10, 5))
plt.imshow(wordcloud, interpolation='bilinear')
plt.axis('off')
plt.title("Word Cloud untuk Sentimen Negatif")
plt.show()
```

