

**ANALISIS SENTIMEN PROGRAM KAMPANYE CALON
GUBERNUR JAWA BARAT MENGGUNAKAN METODE
SUPPORT VECTOR MACHINE (SVM)**

TUGAS AKHIR

Disusun sebagai salah satu syarat untuk kelulusan Kualifikasi penelitian
di Program Studi Teknik Informatika, Universitas Pasundan Bandung

Disusun Oleh :

Emilia Paradila S
NRP. 21.304.0043



**PROGRAM STUDI TEKNIK INFORMATIKA
FAKULTAS TEKNIK
UNIVERSITAS PASUNDAN BANDUNG
SEPTEMBER 2025**

LEMBAR PENGESAHAN

Telah disetujui dan disahkan Proposal Tugas Akhir, dari :

Nama : Emilia Paradila S

Npm : 21.304. 0043

Dengan Judul :

**“ANALISIS SENTIMEN PROGRAM KAMPANYE CALON GUBERNUR
JAWA BARAT MENGGUNAKAN METODE SUPPORT VECTOR
MACHINE (SVM)”**

Mengetahui, 28 September 2025

Menyetujui,

Pembimbing Utama


(Handoko Supeno, ST., MT.)

LEMBAR PERNYATAAN KEASLIAN TUGAS AKHIR

Saya menyatakan dengan sesungguhnya bahwa :

1. Tugas akhir ini adalah benar-benar asli dan belum pernah diajukan untuk mendapatkan gelar akademik, baik di Universitas Pasundan Bandung maupun di Perguruan Tinggi lainnya.
2. Tugas akhir ini merupakan gagasan, rumusan dan penelitian saya sendiri, tanpa bantuan dari orang lain kecuali arahan dari tim dosen pembimbing.
3. Dalam tugas akhir ini tidak terdapat karya atau pendapat orang lain, kecuali bagian-bagian tertentu dalam penulisan laporan Tugas Akhir yang saya kutip dari hasil karya orang lain telah dituliskan dalam sumbernya secara jelas dan sesuai dengan norma, kaidah, dan etika penulisan karya ilmiah, serta disebutkan dalam Daftar Pustaka pada tugas akhir ini.
4. Kakas, perangkat lunak, dan alat bantu kerja lainnya yang digunakan dalam penelitian ini sepenuhnya menjadi tanggung jawab saya, bukan tanggung jawab Universitas Pasundan Bandung.

Apabila di kemudian hari ditemukan seluruh atau sebagian laporan tugas akhir ini bukan hasil karya saya sendiri atau adanya plagiasi dalam bagian-bagian tertentu, saya bersedia menerima sanksi akademik, termasuk pencabutan gelar akademik yang saya sandang sesuai dengan norma yang berlaku di Universitas Pasundan, serta perundang-undangan lainnya.

Bandung, 28 September 2025

Yang membuat pernyataan

(Emilia Paradila S)

NRP : 21.304.0043

ABSTRAK

Calon gubernur memiliki peran penting dalam menentukan arah pembangunan dan pelayanan publik di suatu daerah. Dalam era digital saat ini, opini masyarakat terhadap calon pemimpin banyak disampaikan melalui media sosial. Hal ini mendorong perlunya penelitian untuk memahami persepsi publik sebagai bahan pertimbangan dalam evaluasi politik dan kebijakan.

Penelitian ini bertujuan untuk mengetahui persepsi masyarakat terhadap program kampanye calon gubernur Jawa Barat melalui analisis sentimen menggunakan metode Support Vector Machine (SVM). Data dikumpulkan melalui proses crawling dari media sosial, kemudian diproses melalui tahapan preprocessing, selanjutnya data dibagi menjadi data latih dan data uji, serta dihitung pembobotan kata menggunakan Term Frequency–Inverse Document Frequency (TF-IDF).

Hasil penelitian menunjukkan proses klasifikasi dengan SVM menghasilkan opini yang beragam, baik positif maupun negatif, dengan nilai F1-score 91%, precision 91%, recall 91%, dan akurasi 89,51% pada skenario pembagian data 80:20. Terdapat 485 ulasan positif dan 306 ulasan negatif. Penelitian ini juga menghasilkan aplikasi berbasis web yang memungkinkan pengguna melakukan analisis sentimen secara langsung, baik melalui input teks maupun unggahan dataset CSV. Dengan demikian, penelitian ini diharapkan dapat membantu memahami persepsi publik terhadap calon gubernur serta bermanfaat sebagai bahan evaluasi dalam konteks politik dan kebijakan publik.

Kata Kunci: Klasifikasi, Kampanye, TF-IDF, Support Vector Machine (SVM)

KATA PENGANTAR

Ucapan dan rasa syukur penulis layangkan ke hadirat Ilahi Robbi, yang telah berkenan menguatkan penulis untuk membuat Laporan Tugas Akhir dengan judul “Analisis Sentimen Program Kampanye Calon Gubernur Jawa Barat menggunakan Metode Support Vector Machine (SVM)”.

Adapun penulisan laporan ini bertujuan untuk memenuhi salah satu syarat keikutsertaan pada matakuliah TUGAS AKHIR, di Program Studi Teknik Informatika Universitas Pasundan.

Penulis menyadari laporan ini dapat terwujud berkat bantuan dan dorongan dari berbagai pihak. Maka pada kesempatan ini penulis sampaikan terima kasih yang sebesar besarnya atas segala bantuan yang penulis terima baik secara moral maupun materil, sehingga penulis dapat menyelesaikan laporan ini kepada :

1. Kepada Orang Tua tercinta, dan keluarga yang selalu memberikan motivasi seta do'anya dalam pembuatan tugas akhir ini.
2. Bapak Handoko Supeno, ST., MT. yang telah memnimbing penulis dalam menyelesaikan tugas akhir ini.
3. Bapak Wanda Gusdya Purnama, ST., MT. selaku Koordinator Tugas Akhir.
4. Seluruh civitas akademik Teknik Informatika di UNIVERSITAS PASUNDAN BANDUNG, yang telah memberikan bekal ilmu selama penulis menimba ilmu.
5. Kepada teman-teman seperjuangan Universitas Pasundan Bandung yang tidak bisa semua penulis sebutkan.

Segala kesalahan merupakan kelemahan dan kekurangan penulis. Oleh karena itu, penulis harapkan kritik dan saran dari semuua pihak demi perbaikan pada masa yang akan datang.

Akhir kata, semoga penulisan laporan ini dapat bermanfaat bagi penulis dan bagi perkembangan ilmu teknologi pada masa yang akan datang

Bandung, 28 September 2025

Penulis

DAFTAR ISI

LEMBAR PENGESAHAN.....	ii
LEMBAR PERNYATAAN KEASLIAN TUGAS AKHIR	iii
ABSTRAK	i
KATA PENGANTAR	ii
DAFTAR ISI.....	iii
DAFTAR GAMBAR	vi
DAFTAR TABEL	vii
DAFTAR LAMPIRAN	ix
DAFTAR SIMBOL	x
DAFTAR ISTILAH	xi
DAFTAR KODE PROGRAM	xii
BAB 1 PENDAHULUAN.....	1-1
1.1 Latar Belakang Tugas Akhir	1-1
1.2 Identifikasi Masalah	1-3
1.3 Tujuan Tugas Akhir	1-3
1.4 Lingkup Tugas Akhir	1-3
1.5 Metodologi Tugas Akhir	1-4
1.6 Sistematika Penulisan Tugas Akhir.....	1-8
BAB 2 LANDASAN TEORI	2-1
2.1 Teori yang digunakan.....	2-1
2.1.1 Analisis.....	2-1
2.1.2 Sentimen.....	2-1
2.1.3 Analisis Sentimen.....	2-2
2.1.4 <i>Support Vector Machine (SVM)</i>	2-3
2.1.5 Pengumpulan Data.....	2-5
2.1.6 Pre-processing	2-5
2.1.7 Python.....	2-6
2.1.8 JavaScript	2-7
2.1.9 Google Colab.....	2-7
2.1.10 Blackbox Testing.....	2-7
2.1.11 Confusion Matrix	2-8
2.1.12 Pembobotan Kata	2-9
2.1.13 Teori Penentuan Sentimen.....	2-10
2.2 Penelitian Terdahulu	2-11
BAB 3 METODE PENELITIAN	3-1
3.1 Alur Penyelesaian Tugas Akhir.....	3-1
3.2 Perumusan Masalah.....	3-4

3.3.1 Analisis Masalah	3-5
3.3.2 Solusi Masalah	3-6
BAB 4 ANALISIS DAN PERANCANGAN	4-1
4.1 Analisis.....	4-1
4.1.1 Analisis Kebutuhan dan Masalah	4-1
4.1.2 Penentuan karakteristik Data.....	4-1
4.1.3 Penetapan Tujuan	4-2
4.1.4 Penentuan Ide Sistem	4-2
4.1.5 Definisi Lingkungan.....	4-2
4.2 Pengumpulan Data	4-3
4.3 Pelabelan Data.....	4-5
4.4 Pre-Processing.....	4-7
4.4.1 Cleaning	4-8
4.4.2 Case Folding.....	4-10
4.4.3 Normalization.....	4-13
4.4.4 Tokenization.....	4-15
4.4.5 Stopword	4-18
4.4.6 Stemming	4-20
4.5 Split Data.....	4-22
4.6 Pembobotan Kata (TF-IDF)	4-23
4.6.1 Perhitungan TF (Term Frequency).....	4-23
4.6.2 Perhitungan IDF (Inverse Document Frequency)	4-26
4.6.3 Perhitungan TF-IDF (Term Frequency-Inverse Document Frequency).....	4-28
4.7 Klasifikasi Support Vector Machine	4-47
4.8 Confusion Matrix	4-61
4.9 Analisis Data	4-67
4.10 Perancangan	4-68
4.10.1 Flowchart.....	4-68
4.10.2 Penjelasan Skenario.....	4-69
4.10.3 Use case Diagram.....	4-70
4.10.4 Activity Diagram.....	4-70
4.10.5 Mock Up.....	4-72
BAB 5 IMPLEMENTASI	5-1
5.1 Implementasi Sistem	5-1
5.1.1 Kakas dan Kebutuhan perangkat lunak	5-1
5.1.1.1 Implementasi Tampilan Web	5-1
5.2 Tahap Pengujian.....	5-5
5.3 Evaluasi	5-7
BAB 6 KESIMPULAN	6-1
6.1 Kesimpulan	6-1

6.2	Saran.....	6-2
DAFTAR PUSTAKA		xi
LAMPIRAN.....		xiv

DAFTAR GAMBAR

Gambar 1. 1 Metodologi Tugas Akhir	1-5
Gambar 2. 1 Ilustrasi SVM untuk data linear.....	2-4
Gambar 2. 2 Tabel Confusion matrix sumber: medium.com	2-8
Gambar 3. 1 Fishbone Masalah.....	3-5
Gambar 3. 2 Fishbone Solusi	3-6
Gambar 4. 1 Proses Crawling.....	4-3
Gambar 4. 2 Tahapan Pre-processing	4-7
Gambar 4. 3 Tahapan Support Vector Machine	4-47
Gambar 4. 4 Alur Sistem.....	4-68
Gambar 4. 5 Usecase Diagram.....	4-70
Gambar 4. 6 Activity Diagram.....	4-71
Gambar 4. 7 Mockup Dashboard	4-72
Gambar 4. 8 Mockup Data Ulasan	4-72
Gambar 4. 9 Mockup Pre-Processing.....	4-73
Gambar 4. 10 Mockup Pembobotan.....	4-73
Gambar 4. 11 Mockup Klasifikasi	4-74
Gambar 5. 1 Website-Analisis Teks.....	5-2
Gambar 5. 2 Website-Dashboard	5-3
Gambar 5. 3 Website-Data Ulasan	5-3
Gambar 5. 4 Website-Pre-processing.....	5-4
Gambar 5. 5 Website-Pembobotan Kata	5-4
Gambar 5. 6 Website-Klasifikasi SVM.....	5-5

DAFTAR TABEL

Tabel 2. 1 Jenis Kernel.....	2-5
Tabel 2. 2 Penelitian Terdahulu	2-11
Tabel 3. 1 Alur Penyelesaian Tugas Akhir	3-1
Tabel 4. 1 Analisis kebutuhan dan masalah	4-1
Tabel 4. 2 Karakteristik data	4-1
Tabel 4. 3 Ide sistem	4-2
Tabel 4. 4 Lingkungan Klasifikasi	4-2
Tabel 4. 5 Hasil Dataset Pengumpulan data Paslon 1	4-4
Tabel 4. 6 Hasil Dataset Pengumpulan data Paslon 2	4-4
Tabel 4. 7 Hasil Dataset Pengumpulan data Paslon 3	4-4
Tabel 4. 8 Hasil Dataset Pengumpulan data Paslon 4	4-4
Tabel 4. 9 Hasil Labeling Data Paslon 1	4-5
Tabel 4. 10 Hasil Labeling Data Paslon 2	4-6
Tabel 4. 11 Hasil Labeling Data Paslon 3	4-6
Tabel 4. 12 Hasil Labeling data Paslon 4.....	4-6
Tabel 4. 13 Tahap Cleaning Paslon 1.....	4-8
Tabel 4. 14 Tahap Cleaning Paslon 2.....	4-9
Tabel 4. 15 Tahap Cleaning Paslon 3.....	4-9
Tabel 4. 16 Tahap Cleaning Paslon 4.....	4-10
Tabel 4. 17 Hasil Case Folding Paslon 1	4-11
Tabel 4. 18 Hasil Case Folding Paslon 2	4-11
Tabel 4. 19 Hasil Case Folding Paslon 3	4-12
Tabel 4. 20 Hasil Case Folding Paslon 4	4-12
Tabel 4. 21 Hasil Normalization Paslon 1.....	4-14
Tabel 4. 22 Hasil Normalization Paslon 2.....	4-14
Tabel 4. 23 Hasil Normalization Paslon 3.....	4-14
Tabel 4. 24 Hasil Normalization Paslon 4.....	4-15
Tabel 4. 25 Hasil Tokenization Paslon 1.....	4-16
Tabel 4. 26 Hasil Tokenization Paslon 2.....	4-16
Tabel 4. 27 Hasil Tokenization Paslon 3.....	4-17
Tabel 4. 28 Hasil Tokenization Paslon 4.....	4-18
Tabel 4. 29 Hasil Stopword Paslon 1	4-18
Tabel 4. 30 Hasil Stopword Paslon 2	4-19
Tabel 4. 31 Hasil Stopword Paslon 3.....	4-19
Tabel 4. 32 Hasil Stopword Paslon 4.....	4-20
Tabel 4. 33 Hasil Stemming Paslon 1	4-20
Tabel 4. 34 Hasil Stemming Paslon 2	4-21
Tabel 4. 35 Hasil Stemming Paslon 3	4-21
Tabel 4. 36 Hasil Stemming Paslon 4	4-22
Tabel 4. 37 Data Sample Perhitungan TF-IDF Paslon 1	4-23
Tabel 4. 38 Hasil Frekuensi Kemunculan Setiap Kata Dalam Dokumen	4-23
Tabel 4. 39 Hasil Perhitungan TF	4-25
Tabel 4. 40 Hasil Perhitungan DF	4-26
Tabel 4. 41 Hasil Perhitungan IDF.....	4-27
Tabel 4. 42 Hasil Perhitungan TF-IDF.....	4-28
Tabel 4. 43 Data Sample Perhitungan TF-IDF paslon 2	4-28
Tabel 4. 44 Hasil Frekuensi Kemunculan Setiap Kata Dalam Dokumen	4-29
Tabel 4. 45 Hasil Perhitungan TF	4-30
Tabel 4. 46 Hasil Perhitungan DF	4-31
Tabel 4. 47 Hasil Perhitungan IDF.....	4-32
Tabel 4. 48 Hasil Perhitungan TF-IDF.....	4-33
Tabel 4. 49 Data Sample Perhitungan TF-IDF Paslon 3.....	4-34
Tabel 4. 50 Hasil Frekuensi Kemunculan Setiap Kata Dalam Dokumen	4-34
Tabel 4. 51 Hasil Perhitungan TF	4-36

Tabel 4. 52 Hasil Perhitungan DF	4-37
Tabel 4. 53 Hasil Perhitungan IDF	4-38
Tabel 4. 54 Hasil Perhitungan TF-IDF.....	4-40
Tabel 4. 55 Data Sample Perhitungan TF-IDF Paslon 4.....	4-41
Tabel 4. 56 Hasil Frekuensi Kemunculan Setiap Kata Dalam Dokumen	4-41
Tabel 4. 57 Hasil Perhitungan TF	4-43
Tabel 4. 58 Hasil Perhitungan DF	4-44
Tabel 4. 59 Hasil Perhitungan IDF	4-45
Tabel 4. 60 Hasil Perhitungan TF-IDF	4-46
Tabel 4. 61 Data Vector Paslon 1.....	4-47
Tabel 4. 62 Dokumen Testing Paslon 1	4-50
Tabel 4. 63 Data Vector Paslon 2.....	4-52
Tabel 4. 64 Dokumen Testing Paslon 2	4-54
Tabel 4. 65 Data Vector Paslon 3.....	4-55
Tabel 4. 66 Dokumen Testing Paslon 3	4-57
Tabel 4. 67 Data Vector Paslon 4.....	4-58
Tabel 4. 68 Dokumen Testing Paslon 4	4-60
Tabel 4. 69 Confusion Matrix Paslon 1.....	4-62
Tabel 4. 70 Confusion Matrix Paslon 2.....	4-63
Tabel 4. 71 Confusion Matrix Paslon 3.....	4-64
Tabel 4. 72 Confusion Matrix Paslon 4.....	4-66
Tabel 5. 1 Spesifikasi Perangkat Keras	5-1
Tabel 5. 2 Perangkat Lunak.....	5-1
Tabel 5. 3 Rencana Pengujian	5-6
Tabel 5. 4 Kasus dan Hasil Uji.....	5-6

DAFTAR LAMPIRAN

Lampiran 1 Dataset Paslon 1.....	xv
Lampiran 2 Dataset Paslon 2.....	xv
Lampiran 3 Dataset Paslon 3.....	xvi
Lampiran 4 Dataset Paslon 4.....	xvi
Lampiran 5 Kode Proses Cleansing	xviii
Lampiran 6 Kode Proses Case Folding	xviii
Lampiran 7 Kode Proses Normalization	xx
Lampiran 8 Kode Proses Tokenization	xx
Lampiran 9 Kode Proses Stopword.....	xxi
Lampiran 10 Kode Proses Stemming.....	xxi
Lampiran 11 Kode Proses Split Data	xxii
Lampiran 12 Kode Proses TF-IDF	xxiv
Lampiran 13 Kode Proses Klasifikasi SVM	xxvi

DAFTAR SIMBOL

Simbol	Nama Simbol	Keterangan
	Activity	Penghubung antara Prosedur/Proses
	Transition	Simbol yang digunakan untuk mempresentasikan alur aktivitas
	Off Page Conector	Symbol yang digunakan untuk keluar/masuk pada prosedur atau proses dalam lembar/halaman yang lain

DAFTAR ISTILAH

Istilah	Keterangan
Analisis Sentimen	Proses mengevaluasi opini atau emosi dalam teks (tweet) menjadi kategori positif dan negatif. Dan digunakan untuk mengetahui pandangan masyarakat terhadap calon gubernur
Sentimen	Sikap, perasaan, atau opini yang terkandung dalam satu teks, dapat berupa positif dan negatif
Sentimen Positif	Kategori opini yang menunjukkan kepuasan, apresiasi, atau penilaian baik terhadap suatu objek.
Sentimen Negatif	Kategori opini yang menunjukkan ketidakpuasan, kritik, atau penilaian buruk terhadap suatu objek.
Support Vector Machine (SVM)	Algoritma klasifikasi berbasis margin dan hyperplane.
Natural Language Processing (NLP)	Bidang AI yang memproses bahasa alami manusia.
Machine Learning	Algoritma pembelajaran mesin untuk analisis data.
Crawling Data	Proses mengumpulkan data dari media sosial secara otomatis
Preprocessing	Proses pembersihan dan persiapan data sebelum dilakukan analisis
Cleaning	Menghapus URL, mention, emotikon, simbol, dan karakter khusus.
Case Folding	Mengubah semua huruf menjadi huruf kecil
Normalization	Menstandarkan kata tidak baku ke bentuk kata baku (KBBI)
Tokenization	Tahap memecah kata dalam sebuah kalimat menjadi tunggal
Stopword Removal	Menghapus kata umum yang tidak bermakna
Stemming	Mengembalikan kata ke bentuk dasarnya
TF-IDF	Metode pembobotan kata berdasarkan frekuensi kemunculan dan uniknya kata dalam dokumen.
Confusion Matrix	Matrix pada evaluasi model klasifikasi
Accuracy	Tingkat ketepatan prediksi model, dihitung dari rasio jumlah prediksi benar dibandingkan dengan total prediksi.
Precision	Rasio antara jumlah data positif yang diprediksi benar dibandingkan dengan seluruh data yang diprediksi positif.
Recall	Proporsi data relevan yang berhasil ditemukan
F1-Score	Rata-rata harmonik precision dan recall.
Twitter / X	Platform media sosial sumber data opini publik
Sentimen Positif / Negatif	Kategori hasil pada klasifikasi berdasarkan opini dalam tweet
WordCloud	Visualisasi kata yang sering muncul dalam dataset
Dataset	Kumpulan data tweet yang akan dianalisis
Polynomial	Jenis kernel yang digunakan untuk menangani data berpola polynomial
RBF Kernel	Jenis kernel yang efektif untuk data yang tidak dapat dipisahkan secara linear
Sigmoid	Kernel yang digunakan dalam SVM yang terinspirasi dari fungsi aktivitas di neural network

DAFTAR KODE PROGRAM

Kode Program 4. 3 Cleaning Data	4-8
Kode Program 4. 4 Case Folding	4-11
Kode Program 4. 5 Normalization	4-13
Kode Program 4. 6 Tokenization	4-15
Kode Program 4. 7 Stopword.....	4-18
Kode Program 4. 8 Stemming.....	4-21

BAB 1

PENDAHULUAN

Bab ini berisi penjelasan umum mengenai ulasan penelitian yang dilakukan dalam pengerjaan tugas akhir. Di dalamnya berisi latar belakang tugas akhir, identifikasi masalah, tujuan tugas akhir, lingkup tugas akhir, metodologi yang digunakan dalam pengerjaan tugas akhir, dan sistematika penulisan laporan tugas akhir.

1.1 Latar Belakang Tugas Akhir

Perkembangan teknologi saat ini memudahkan pengguna media sosial untuk dapat mengungkapkan isi hatinya atau pendapat tentang suatu hal melalui berbagai platform, dan Twitter adalah salah satu platform media sosial yang paling populer dikalangan masyarakat. Dengan banyaknya data yang tersebar dan tersedia di Twitter, analisis sentimen menjadi lebih mudah dilakukan. Data tweet yang dikumpulkan dapat digunakan untuk mengetahui opini masyarakat tentang calon gubernur. Analisis sentimen digunakan untuk mengevaluasi perasaan atau opini yang terkandung dalam tweet tentang calon gubernur [WIN23].

Dari berbagai metode yang digunakan dalam bidang data science untuk menganalisis data, analisis sentimen dapat menghemat waktu analisis data untuk mengetahui tingkat sentimen suatu objek. Hal ini karena analisis sentimen menggunakan perkembangan teknologi melalui berbagai tools khusus. Analisis sentimen terhadap calon gubernur adalah proses analisis pandangan, opini, dan perasaan masyarakat terhadap calon gubernur yang bersangkutan dengan menggunakan berbagai sumber seperti media sosial, ulasan berita dan platform lainnya. Tujuan dari analisis sentimen ini adalah untuk mengetahui respon dan perasaan publik terhadap Calon Gubernur, apakah sentimen tersebut cenderung positif dan negatif.

Analisis sentimen merupakan proses yang digunakan untuk mengevaluasi pendapat atau perasaan yang diungkapkan dalam teks. Pada penelitian ini analisis sentimen diterapkan untuk mengetahui sentimen publik terkait calon gubernur Jawa Barat. Data opini diperoleh dari aplikasi twitter dengan mengumpulkan data tweet selama rentang waktu tertentu.

Penelitian ini memilih Calon Gubernur Jawa Barat sebagai objek analisis sentimen karena posisi tersebut memiliki pengaruh yang signifikan terhadap pilihan kebijakan di provinsi ini, yang merupakan salah satu wilayah dengan jumlah penduduk terbesar di Indonesia. Salah satu alasan utama untuk memilih Calon Gubernur Jawa

Barat adalah banyaknya data yang tersedia, seperti tanggapan dan opini publik di berbagai media sosial yang mencakup berbagai masalah politik dan sosial. Data yang kaya dan representatif ini memungkinkan penelitian untuk mengumpulkan berbagai opini masyarakat tentang Calon Gubernur Jawa Barat, sehingga analisis yang dilakukan dapat memberikan gambaran yang komprehensif tentang preferensi dan persepsi publik tentang Calon Gubernur Jawa Barat.

Calon gubernur Jawa Barat memiliki peran sangat penting dalam mempermudah akses informasi dan pelayanan publik di provinsi ini. Dengan program yang dirancang oleh calon Gubernur, orang – orang dapat dengan mudah memenuhi kebutuhan pelayanan publik mereka, seperti akses ke layanan kesehatan, pendidikan, transportasi, dan infrastruktur lainnya. Jika calon Gubernur memiliki visi dan program yang jelas serta komitmen untuk meningkatkan kualitas hidup masyarakat Jawa Barat, mereka dapat memberikan pandangan yang lebih lengkap tentang prioritas pembangunan di provinsi ini. Program – program ini juga dapat membantu masyarakat dalam membuat keputusan yang lebih bijak berdasarkan informasi saat memilih pemimpin yang dapat membawa perubahan positif Jawa Barat.

Dalam penelitian ini, analisis sentimen digunakan untuk melakukan klasifikasi terhadap komentar, dan membaginya menjadi dua kelas yaitu sentimen positif dan negatif. Salah satu contoh penerapan praktis dari analisis sentimen digunakan untuk Calon Gubernur Jawa Barat ialah mengidentifikasi tren opini dan pandangan masyarakat terhadap calon tersebut. Ini dapat menunjukkan apakah persepsi publik terhadap calon tersebut cenderung positif atau negatif, yang dapat digunakan sebagai bahan pertimbangan saat membuat strategi kampanye dan komunikasi.

Metode Support Vector Machine (SVM) dipilih untuk analisis sentimen karena memiliki beberapa kelebihan yang sesuai dengan tujuan penelitian ini. Dan juga metode ini sangat relatif dan mudah diimplementasikan. Ini memungkinkan peneliti memproses data ulasan atau komentar dengan cepat. Support Vector Machine merupakan algoritma berbeda dengan dengan algoritma lainnya seperti misalnya *Neural Network (NN)*, dimana SVM menggunakan data terpilih saja yang akan berkontribusi pada model. Data-data yang akan berkontribusi pada model disebut dengan *support vector machine*. SVM memiliki kernel yang digunakan untuk mentransformasikan inputan ke bentuk yang diperlukan [SIT20].

Analisis sentimen adalah penambahan teks yang menganalisis dan mengekstraksi informasi yang bersifat subjektif dan dapat membantu pihak yang

membutuhkan [NOR21]. Analisis sentimen mampu mengubah data yang berbentuk tidak terstruktur menjadi data yang terstruktur [NAT20]. Namun, penerapan analisis sentimen ini akan secara otomatis mengambil data dari postingan atau tweet pengguna. Dan sistem akan melakukan klasifikasi untuk menilai apakah postingan tersebut mengandung kalimat positif dan negatif. Dengan menggunakan klasifikasi tersebut, pengguna dapat mengetahui dan menilai suatu topik tertentu serta dapat membuat keputusan.

1.2 Identifikasi Masalah

Berdasarkan latar belakang yang telah dipaparkan sebelumnya, maka dapat diperoleh permasalahan sebagai berikut :

1. Apakah klasifikasi sentimen berbasis machine learning dapat digunakan untuk memprediksi popularitas calon gubernur berdasarkan opini publik di media sosial?
2. Bagaimana hyperparameter serta pre-processing yang cocok untuk dataset sentimen untuk media sosial di Indonesia?
3. Bagaimana metode Support Vector Machine (SVM) dapat digunakan untuk mengetahui serta menganalisis opini masyarakat terhadap calon gubernur sekaligus menghitung akurasi hasil klasifikasinya?
4. Bagaimana membangun aplikasi berbasis web yang dapat digunakan untuk melakukan analisis sentimen secara langsung?

1.3 Tujuan Tugas Akhir

Berdasarkan latar belakang diatas maka tujuan tugas akhir sebagai berikut :

1. Memprediksi popularitas calon gubernur berdasarkan opini publik melalui klasifikasi sentimen berbasis machine learning.
2. Merancang hyperparameter serta pre-processing yang sesuai untuk dataset sentimen pada media sosial.
3. Mengetahui dan menganalisis opini masyarakat terhadap calon gubernur melalui analisis sentimen menggunakan metode *Support Vector Machine* (SVM) serta menghitung akurasi hasil klasifikasinya.
4. Membangun aplikasi berbasis web yang dapat digunakan untuk melakukan analisis sentimen secara langsung.

1.4 Lingkup Tugas Akhir

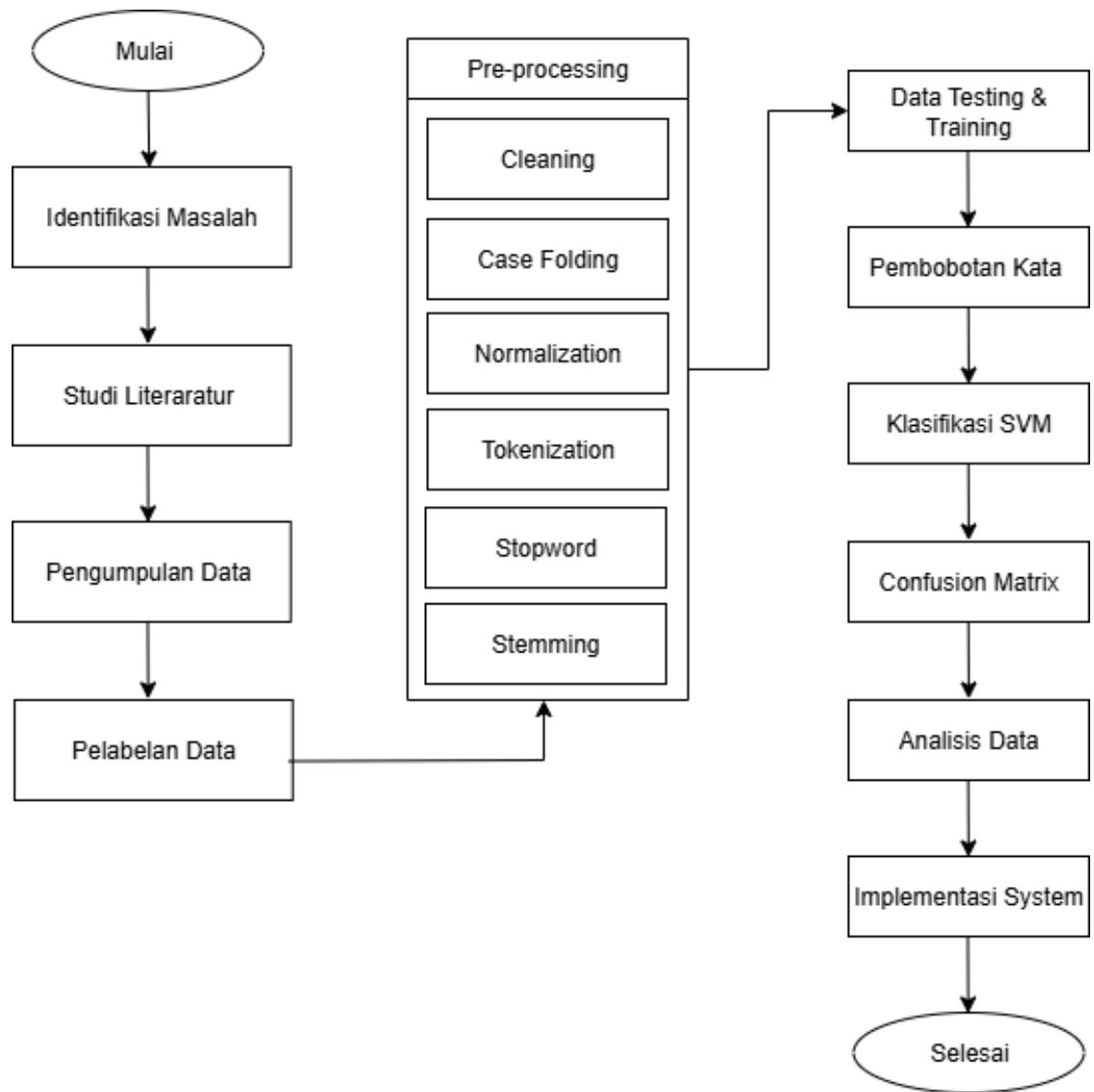
Agar dalam penyelesaian Tugas Akhir ini tidak menyimpang dari pokok permasalahan yang telah diidentifikasi maka penelitian ini memiliki lingkup sebagai

berikut :

1. Data yang digunakan berupa opini publik dari media sosial, yang diperoleh melalui *filtering* kata kunci terkait kampanye calon gubernur.
2. Melakukan evaluasi terhadap tingkat akurasi hasil klasifikasi SVM pada data opini publik yang digunakan.
3. Analisis sentimen dibatasi pada dua kelas yaitu sentimen positif dan sentimen negatif.
4. Sistem yang dibangun diwujudkan dalam bentuk aplikasi berbasis web yang menyediakan dua fitur utama, yaitu analisis cepat melalui input teks dan analisis batch melalui unggahan dataset CSV.

1.5 Metodologi Tugas Akhir

Bagian ini akan menjelaskan langkah – langkah metodologi penyelesaian yang dilakukan penulis dalam menyelesaikan tugas akhir. Langkah – langkah tersebut dijelaskan dalam bentuk diagram yang dapat dilihat pada gambar



Gambar 1. 1 Metodologi Tugas Akhir

Berikut merupakan rincian dari metodologi Tugas Akhir, yaitu :

1. Identifikasi Masalah

Pada tahap ini dilakukan identifikasi dan analisis permasalahan yang berkaitan dengan program kampanye calon gubernur Jawa Barat untuk mendapatkan solusi sementara dan penyelesaian dari permasalahan yang sedang terjadi.

2. Studi literatur

Pada tahap ini dilakukan riset setelah topik penelitian ditentukan. Proses mengkaji literatur melibatkan membaca dan membuat catatan mengenai jurnal atau tulisan ilmiah untuk mengumpulkan beberapa teori yang berkaitan dengan topik penelitian yang sedang diteliti. Hal ini dilakukan sebelum

melakukan proses pengumpulan data dan pengolahan data sebagai referensi dalam pembahasan dari hasil penelitian.

3. Pengumpulan Data

Tahap selanjutnya setelah melakukan studi literatur maka selanjutnya yaitu akan melakukan tahapan pengumpulan data. Pada proses pengumpulan data di media sosial disebut proses Crawling Data.

4. Pelabelan Data

Data yang diperoleh dari proses crawling akan dilakukan pelabelan, data tersebut nantinya akan diolah untuk dikategorikan menjadi 2 bagian yaitu negatif dan positif. Hal ini memudahkan peneliti untuk proses analisis sentimen dan memudahkan persepsi pengguna terhadap subjek yang dianalisis.

5. Pre-processing

Bagian ini bertujuan untuk membersihkan dan menyiapkan data agar siap dianalisis lebih lanjut, proses ini tidak hanya meningkatkan kualitas analisis, tetapi juga mengoptimalkan kinerja algoritma di tahap selanjutnya. Dalam bagian ini semua data yang diperoleh dari proses scrapping akan diolah sebagai berikut :

- a) Cleaning : Tahap ini adalah tahap pre-processing data bertujuan untuk menghilangkan informasi yang tidak relevan. Seperti karakter khusus, tanda baca, symbol, dan angka yang tidak diperlukan dalam data.
- b) Case Folding : Tahap case folding adalah proses mengubah seluruh huruf dalam teks menjadi huruf kecil (lowercase).
- c) Normalization : Tahap normalization adalah proses mengubah data teks kedalam bentuk yang sama atau seragam. Sehingga teks menjadi lebih konsisten dan mudah dianalisis.
- d) Tokenization : Ini adalah proses pemisahan data teks menjadi bagian-bagian kecil yang disebut token. Token berupa kata, frasa atau karakter.
- e) Stopword : adalah menghapus data atau token yang tidak relevan untuk analisis.
- f) Stemming : Tahap ini adalah proses mengubah kata-kata berimbuhan menjadi bentuk dasar. Stemming sendiri membantu dalam menormalisasikan teks sehingga membantu peneliti agar lebih efisien dan konsisten.

6. Data Testing & Data Training

Bagian ini adalah dua data set yang digunakan dalam proses pengembangan machine learning. Data training digunakan untuk melatih model yang nantinya digunakan untuk mengevaluasi kinerja dan pola model, lalu di masukkan data baru yang disebut data testing untuk menguji seberapa baik model tersebut.

7. Pembobotan

Pembobotan dilakukan untuk memberikan nilai atau bobot tertentu dalam suatu variabel atau data. Setelah data memiliki bobot, nantinya data ini akan diklasifikasikan untuk mengetahui kelas dari kalimatnya. Pembobotan kata menggunakan TF-IDF, dimana TF-IDF memberikan bobot kata berdasarkan frekuensi term (TF) dan inverse document frequency (IDF).

8. Klasifikasi Support Vector Machine

Pada tahap Support Vector Machine akan melakukan klasifikasi data menjadi negative dan positif. Pada pengklasifikasian ini Support Vector Machine akan membagi kelas dengan menemukan hyperplane terbaik untuk mendapatkan hyperplane terbaik dilakukan dengan menghitung margin maksimal dari kedua kelas. Margin adalah jarak dari pattern terdekat kelas dengan hyperplane.

9. Confusion Matrix

Confusion Matrix adalah metode untuk machine learning dan statistic untuk mengevaluasi kinerja model klasifikasi. Dengan confusion matrix dapat membandingkan klasifikasi yang dilakukan oleh model dengan hasil klasifikasi yang seharusnya.

10. Analisis Data

pada tahapan penelitian ini, yaitu data yang telah di proses melalui tahap pre-processing akan di proses lagi dengan menggunakan metode atau algoritma yang sesuai untuk melakukan pengolahan data, sehingga data tersebut dapat menjadi informasi yang dibutuhkan dan berguna. Penelitian juga menggunakan metode pembobotan kata yaitu data yang akan digunakan dapat maksimal, proses klasifikasi juga menggunakan metode yang merupakan metode klasifikasi yang sudah terbukti telah mencapai hasil akurasi yang baik dan sesuai dengan perhitungan yang telah dilakukan.

11. Alur Sistem

Untuk memulai implementasi sistem, peneliti terlebih dahulu membuat alur fungsionalitas dari sistem serta tampilan antar muka. Tahapan pertama yang dilakukan yaitu dengan membuat desain sistem. Desain sistem ini terdiri dari flowchart, use case diagram, activity diagram, dan mockup yang menggambarkan alur proses, interaksi pengguna dengan sistem, serta rancangan tampilan antar muka. Dimana website akan menjadi output dari penelitian ini. Nantinya aplikasi akan berjalan sesuai dengan alur sistem yang telah dirancang tersebut.

1.6 Sistematika Penulisan Tugas Akhir

Laporan tugas akhir ini dibuat untuk mendokumentasikan pengerjaan tugas akhir. Oleh karena itu, diusulkan sistematika penulisan yang menjelaskan mengenai bab – bab pada laporan tugas akhir beserta deskripsinya. Serta keterkaitan antara bab sebelum dan sesudahnya berikut merupakan sistematika penulisan yang akan digunakan dalam penyusunan tugas akhir:

BAB 1 PENDAHULUAN

Bab ini berisi penjelasan umum mengenai usulan penelitian yang dilakukan dalam pengerjaan tugas akhir. Didalamnya berisi latar belakang, identifikasi masalah, tujuan tugas akhir, lingkup tugas akhir, metodologi penyelesaian tugas akhir serta sistematika penulisan tugas akhir.

BAB 2 LANDASAN TEORI

Bab ini menjelaskan mengenai teori – teori yang berkaitan dengan masalah yang diangkat dalam tugas akhir serta penelitian serta penelitian terdahulu yang berhubungan dengan permasalahan yang diangkat dalam tugas akhir.

BAB 3 SKEMA PENELITIAN

Bab ini menjelaskan tentang alur penyelesaian tugas akhir, perumusan masalah, kerangka pemikiran teoritis, dan profil sumber data yang digunakan.

BAB 4 ANALISIS DAN PERANCANGAN

Bab ini berisi mengenai konsep dan desain sistem yang akan dibangun. Di dalamnya akan dijelaskan deskripsi sistem, diagram perancangan, serta model prototipe yang akan digunakan dalam pengembangan sistem.

BAB 5 IMPLEMENTASI

Bab ini berisi mengenai penerapan metode Support Vector Machine (SVM) untuk melakukan analisis sentimen terhadap program kampanye calon gubernur Jawa Barat. Pada tahap ini, data yang dikumpulkan dari media sosial, khususnya Twitter, akan diproses dan dianalisis untuk mendapatkan gambaran sentimen publik terhadap calon gubernur.

BAB 6 KESIMPULAN DAN SARAN

Bab ini akan berisi kesimpulan yang dapat diperoleh dari saran yang dapat diterapkan kedepannya bila melakukan analisis sentimen atau opini publik di media sosial terkait dengan topik program kampanye.

BAB 2

LANDASAN TEORI

2.1 Teori yang digunakan

Berikut ini merupakan teori – teori yang digunakan sebagai teori pendukung pada penelitian tugas akhir ini.

2.1.1 Analisis

Analisis atau opini mining merupakan kajian tentang cara untuk memecahkan masalah opini masyarakat, sikap dan emosi suatu entitas, yang dapat mewakili individu, peristiwa atau topik [SUS20]. Sumber kata “analysis” berasal dari kata Yunani kuno "analusis", yang terdiri dari dua kata, "ana" yang berarti kembali dan "luein" yang berarti melepas atau mengurai, dan kata ini berarti mengurai kembali. Analisis berarti informasi atau persoalan yang kemudian akan dijelaskan dan diteliti agar menjadi informasi yang tersusun rapi dan mudah dipahami. Tujuan dari kegiatan berpikir analisis adalah untuk memecah suatu keseluruhan menjadi bagian-bagiannya sehingga dapat mengidentifikasi ciri-ciri setiap komponen, hubungan nya adalah satu sama lain, dan peran masing – masing dalam konteks untuk keseluruhan yang terintegrasi. Analisis dilakukan dengan tujuan untuk mengumpulkan suatu data agar dapat menyelesaikan masalah atau konflik.

Melalui suatu pendekatan analisis, berbagai persoalan yang kompleks dapat ditelaah secara lebih mendalam dan terstruktur. Proses ini memerlukan ketelitian dan kemampuan berpikir secara sistematis agar setiap unsur yang berkaitan dengan permasalahan dapat dikenali dan dan dipahami secara menyeluruh. Dalam praktiknya, analisis sering digunakan untuk mendukung kegiatan penelitian, pengambilan keputusan, serta melakukan evaluasi terhadap suatu kebijakan atau fenomena sosial. Oleh karena itu, analisis bukan hanya sekedar untuk memecahkan masalah tetapi juga untuk memperoleh pemahaman yang lebih utuh dan mendalam terhadap suatu objek kajian.

2.1.2 Sentimen

Sentimen adalah pendapat atau pandangan yang didasarkan pada perasaan yang berlebih-lebihan terhadap sesuatu [LING14]. Setiap orang memiliki pandangan yang berbeda-beda terhadap suatu objek tersebut, dari cara pandang seseorang kemudian dapat dinilai dengan cara pandang yang positif dan pandang

negatif. Dari perasaan yang timbul dalam menilai suatu objek sehingga akan menghasilkan suatu opini. Opini yang dihasilkan dari penilaian seseorang merupakan pengertian dari sentimen.

Salah satunya yaitu dalam dunia politik, banyak masyarakat yang menyampaikan opini atau tanggapan terhadap calon gubernur melalui media sosial. Opini-opini tersebut dapat dimanfaatkan untuk melakukan pengukuran terhadap sentimen publik. Pengukuran ini bertujuan untuk mengetahui dan mendeteksi bagaimana kondisi opini masyarakat secara umum. Teknik yang akan digunakan untuk melakukan pengukuran ini adalah analisis sentimen, yang mampu mengidentifikasi kecenderungan positif atau negatif dalam data sosial. Menurut KBBI sentimen merupakan pendapat atau pandangan yang didasarkan pada perasaan yang berlebih-lebihan terhadap sesuatu. Sikap sentimen ini dapat dilihat pada perasaan dengki, muncul perasaan yang emosi berlebih sehingga sentimen dapat berujung menjadi dampak yang merugikan dan tidak menguntungkan.

2.1.3 Analisis Sentimen

Analisis sentimen merupakan salah satu contoh dari bidang *Natural Language Processing* (NLP) yang paling populer. *Natural Language Processing* (NLP) adalah bidang ilmiah yang membahas tentang bagaimana caranya agar komputer bisa bekerja dan berpikir seperti manusia. *Natural Language Processing* (NLP) merupakan bagian dari *Artificial Intelligence* atau kecerdasan buatan. Dalam perkembangan data mining, *Artificial Intelligence* (AI) merupakan salah satu dari empat cabang ilmu data mining, yaitu statistika, database, dan pencarian informasi. Dalam penerapannya, *Artificial Intelligence* (AI) juga memerlukan machine learning sebagai algoritma penyelesaian. Adanya machine learning digunakan untuk menggantikan manusia dalam mengambil keputusan. Machine learning tidak mempunyai perasaan seperti manusia sehingga keputusan yang diambil berdasarkan data yang sudah diolah [SAP22].

Analisis sentimen merupakan proses komputasi menggunakan teknik analisis teks dalam bentuk data tekstual dengan cara mengelola, memahami, dan mengelompokkan emosional baik positif atau negatif. Analisis sentimen banyak digunakan karena meningkatnya kebutuhan individu atau kelompok dalam mengetahui pendapat seseorang terhadap sesuatu. Analisis sentimen juga dipengaruhi oleh data set yang digunakan akan mengalami penanganan yang berbeda [WID21]. Analisis sentimen ditujukan bukan hanya untuk pribadi tapi

juga untuk organisasi. Analisis sentimen juga disebut opinion mining yang merupakan proses menentukan emosional pengguna dengan cara menganalisis tulisan. Tak hanya tokoh, analisis sentimen juga dapat mengatasi masalah mengenai bisnis, program, produk, aplikasi, dan lain lain yang dapat dikomentari publik [HER20]. Keunggulan analisis sentimen ini adalah menghemat waktu dan tenaga dalam melakukan penelitian dengan jumlah data yang besar.

Berikut contoh penerapan analisis sentimen :

1. Bidang bisnis, misalnya mengetahui bagaimana reputasi merek produk baru di masyarakat sehingga dapat meningkatkan merek produk tersebut.
2. Bidang politik, misalnya mengetahui popularitas seorang tokoh sehingga dapat memberi pengetahuan mengenai tokoh tersebut.
3. Program, misalnya vaksinasi, covid 19, pilpres, dan lain-lain sehingga dapat memperbaiki kebijakan pemerintah tersebut.

Terdapat beberapa jenis analisis sentimen yaitu emotion detection, aspectbased sentiment analysis, dan fine grand sentiment analysis. Fine sentiment analysis adalah jenis analisis yang memiliki penilaian spesifik dan biasa digunakan pada bidang e-commerce. Emoticon detection adalah jenis analisis yang bertujuan untuk mengetahui emosi yang ada pada pesan misalnya emosi bahagia, sedih, marah, dan lain-lain. Aspect-based sentiment analysis merupakan jenis analisis untuk mengetahui aspek yang berpengaruh dan penilaian dari pelanggan [ARV21].

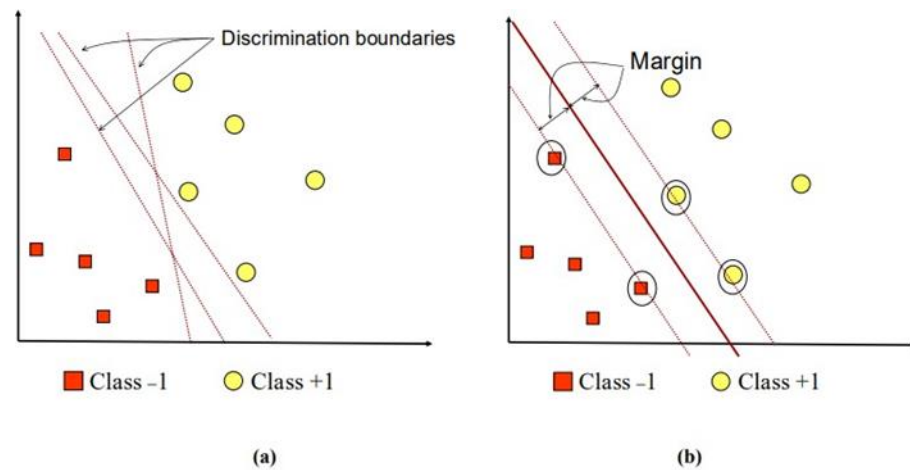
2.1.4 Support Vector Machine (SVM)

Support Vector Machine (SVM) merupakan salah satu teknik untuk melakukan suatu prediksi, baik dalam kasus klasifikasi atau regresi. Metode SVM memiliki prinsip dasar linier classifer yaitu kasus klasifikasi yang dapat dipisahkan secara linier, namun SVM yang dikembangkan dapat bekerja dengan problem non linier dengan memasukkan konsep kernel pada ruang berdimensi tinggi. Pada ruang berdimensi tinggi, hyperplane yang akan dapat memaksimalkan jarak (margin) antara kelas data. Metode *Support Vector Machine* (SVM) berakar dari teori pembelajaran statistik yang dapat memberikan hasil yang lebih baik dari metode yang lain [RIZ21].

Support Vector Machine (SVM) merupakan algoritma pembelajaran mesin yang bekerja sesuai prinsip *Structural Risk Minimization* (SRM). Dengan tujuan untuk menemukan hyperplane optimal yang memisahkan dua kelas pada ruang

input [GIF22]. Konsep klasifikasi menggunakan *Support Vector Machine* adalah untuk menemukan hyperplane terbaik yang berfungsi menjadi pemisah dua kelas pada data. *Support Vector Machine* juga mampu bekerja pada suatu dataset yang berdimensi tinggi dengan memakai trik kernel. *Support Vector Machine* akan menggunakan titik data tertentu yang berkontribusi (support vector) untuk menghasilkan model yang akan digunakan pada proses klasifikasi. Berikut adalah ilustrasi dari data support vector machine untuk data linear.

Klasifikasi support vector machine bekerja dengan mencari hyperplane terbaik, support vector machine bertujuan mencari hyperplane yang optimal yaitu dengan memaksimalkan margin atau pattern data (training data), dimana hyperplane terbaik yang berada di posisi tengah antara dua objek dari dua kelas. Margin atau objek terdekat disebut support vector [FAJ23].



Gambar 2. 1 Ilustrasi SVM untuk data linear

Permasalahan dalam dunia nyata pada umumnya bersifat non-linear, sedangkan permasalahan yang benar-benar linear cenderung jarang ditemukan. Untuk menyelesaikan masalah tersebut *Support Vector Machine* (SVM) memberikan kemudahan dengan adanya fungsi kernel [NUG11]. Berikut adalah jenis-jenis kernel :

Tabel 2. 1 Jenis Kernel

Jenis Kernel	Definisi
Polynomial	$ \vec{X}_i - \vec{X}_j = (\vec{X}_i - \vec{X}_j + 1)^P$
RBF	$K(\vec{X}_i - \vec{X}_j) = \exp(-\gamma(X, Y)^2)$
Sigmoid	$K(\vec{X}_i - \vec{X}_j) = \tanh \alpha \vec{X}_i - \vec{X}_j + \beta)$

2.1.5 Pengumpulan Data

Pengumpulan data teks media sosial pada penelitian ini dilakukan dengan menggunakan kombinasi metode crawling data dan pengambilan dari sumber lain yang relevan. Metode crawling digunakan untuk mengumpulkan data secara otomatis dari platform media sosial dengan menentukan kata kunci tertentu, menetapkan rentang waktu, serta menyesuaikan kebutuhan penelitian.

Selain itu, data juga dapat diperoleh melalui sumber lain yang sesuai dengan topik penelitian, sehingga dapat memperkaya variasi dan konteks analisis. Seluruh data yang dikumpulkan kemudian disimpan dalam format yang sesuai agar memudahkan tahap pengolahan lebih lanjut, termasuk proses pelabelan sentimen sebelum masuk ke tahap analisis menggunakan model yang telah ditentukan.

2.1.6 Pre-processing

Proses ini merupakan tahapan untuk mengubah data yang tidak terstruktur menjadi data yang rapi dan terstruktur. Pada tahap ini ada beberapa proses yang dilakukan yaitu :

1. Cleaning

Merupakan proses pembersihan data mencakup penghapusan mention, karakter khusus, link, hastag, nomor, dan pengelolaan data duplikat. Memastikan bahwa data yang akan digunakan untuk melakukan analisis sentimen bersih dan konsisten adalah tujuan dari proses ini. Hasilnya adalah dataset yang lebih bersih, kohesif, dan siap digunakan, yang memungkinkan untuk meningkatkan akurasi dan kepercayaan hasil analisis sentimen.

2. Case Folding

Case folding Merupakan proses yang berfungsi untuk mengubah huruf menjadi huruf kecil atau lowercase.

3. Normalization

Proses ini merupakan tahapan dimana pengubahan kata yang tidak baku atau singkatan menjadi kalimat baku yang sesuai dengan KBBI (Kamus Besar Bahasa Indonesia).

4. Tokenization

Tokenisasi memiliki fungsi untuk mengubah teks menjadi unit-unit kecil yang disebut token, seperti kata atau frasa.

5. Stopword Removal

Tahap ini memiliki fungsi untuk menghilangkan kata-kata umum yang dianggap tidak relevan dalam proses klasifikasi sentimen. Yaitu kata-kata seperti “yang”, “atau”, “ke”, “di”, “dengan”, dan sejenisnya dihapus dan tidak diikutsertakan dalam proses analisis.

6. Stemming

Proses stemming ini merupakan proses untuk mengubah kata-kata yang berimbuhan menjadi kata baku, contoh “dibukakan” akan diubah menjadi kata baku yaitu “buka”.

2.1.7 Python

Python adalah bahasa pemrograman yang paling populer dan banyak digunakan di berbagai bidang, terutama dalam analisis data. Python didirikan pada tahun 1990 oleh Guido van Rossum, Python menjadi favorit banyak ilmuwan data, peneliti, dan pengembang perangkat lunak karena kemudahan penggunaan dan fleksibilitasnya. Kelebihan Python melibatkan ketersediaan berbagai pustaka dan kerangka kerja kuat seperti NumPy, Pandas, dan Matplotlib, yang mempermudah proses pengolahan, manipulasi, analisis, dan visualisasi data dengan efisien [REG23].

Python merupakan bahasa pemrograman interpretatif (setiap baris kode dieksekusi secara berurutan saat program dijalankan) yang sering digunakan para developer di dunia pemrograman, terutama dalam pengembangan *website*, *software*, aplikasi berbasis web, data learning untuk merancang aplikasi dan *software* di berbagai bidang. *Stack Overflow* membuktikan bahwa *python* adalah bahasa pemrograman yang populer peringkat keempat setelah *JavaScript*, *HTML/CSS* dan juga *SQL* pada tahun 2022 dari 70.000 responden, 48% menyatakan bahwa mereka menggunakan bahasa pemrograman ini [NAF24]

2.1.8 JavaScript

JavaScript adalah bahasa pemrograman yang populer dipakai saat membuat website, game dan aplikasi .Jika memahami dan menguasainya, Anda dapat membuat halaman website yang menarik atau mengembangkan game online berbasis web.selama proses pembuatan website, JavaScript memiliki kemampuan untuk membuat website tampak lebih menarik dan interaktif, dengan konten-konten dinamisnya [FRI24].

JavaScript merupakan sebuah bahasa pemrograman tingkat tinggi yang saat ini sudahh menjadi bahasa pemrograman utama bagi web developer di samping HTML (HyperText Markup Language) dan CSS (Cascading Style Sheet). JavaScript adalah bahasa pemrograman yang digunakan dalam pengembangan website agar lebih dinamis dan interaktif [PRA22]

2.1.9 Google Colab

Google colab atau google colaboratory adalah layanan google yang berbasis cloud dengan menggunakan bahasa python, google collab biasanya digunakan untuk machine learning dengan analisis data yang dimana layanan ini memungkinkan untuk mengerjakan bersama tim [FEB24].

Google collab juga memberikan akses gratis ke unit pemrosesan tensor dan unit pemrosesan grafis (TPU), yang akan mempercepat komputasi intensif seperti pelatihan model pembelajaran mesin dan *deep learning*.

2.1.10 Blackbox Testing

Menurut buku yang berjudul The Art of Testing yang ditulis oleh Glenford J. Myers (2004) menyebutkan bahwa black box testing muncul pada tahun 1979. Di dalam buku tersebut menyebutkan bahwa black box testing tidak berkonsentrasi pada perilaku internal dan struktur program tetapi black box testing adalah salah satu strategi pengujian penting dalam software testing yang berkonsentrasi menemukan keadaan bagaimana program tidak berjalan sesuai dengan spesifikasinya. (Myers et al.,2004)

Blackbox Testing adalah metode pengujian perangkat lunak yang berfokus pada pengujian fungsionalitas aplikasi. Proses pengujian blackbox testing yaitu dengan mencoba aplikasi atau program yang sudahh dibuat apakah sudah sesuai dengan hasil yang di harapkan [SHA21]. Adapun tujuan dilakukan pengujian dengan blackbox testing yaitu untuk menemukan kesalahan atau cacat

dalam fungsionalitas perangkat lunak, serta memastikan bahwa sistem berperilaku sesuai dengan yang diharapkan.

2.1.11 Confusion Matrix

Confusion Matrix merupakan sebuah teknik yang digunakan dalam data mining dan Machine Learning untuk menghitung seberapa baik sebuah model yang memprediksi label dari sebuah data. Teknik ini sering digunakan dalam evaluasi model classification dimana model harus memprediksi label dari sebuah data berdasarkan atribut-atribut yang ada. Confusion matrix membantu mengukur kinerja dimana output algoritma dapat berada dalam dua kategori yang sering disebut Positif atau Negatif, Ya atau Tidak. Dalam tabel terdapat empat sel, yang mewakili kombinasi unik dari nilai prediksi atau actual [KAN23].

		Actual Values	
		1 (Positive)	0 (Negative)
Predicted Values	1 (Positive)	TP (True Positive)	FP (False Positive) <i>Type I Error</i>
	0 (Negative)	FN (False Negative) <i>Type II Error</i>	TN (True Negative)

Gambar 2. 2 Tabel Confusion matrix sumber: medium.com

Komponen utama dari confusion matrix adalah sebagai berikut:

TP = *True Positive* (data positif yang diprediksi benar sebagai positif)

TN = *True Negative* (data negatif yang diprediksi benar sebagai negatif)

FP = *False Positive* (data negatif namun diprediksi sebagai data positif)

FN = *False Negative* (data positif namun diprediksi sebagai data negatif)

1. Accuracy

Accuracy merupakan metode pengujian berdasarkan tingkat kedekatan antara nilai prediksi dengan nilai aktual. Dalam mengetahui jumlah data yang diklasifikasikan secara benar maka dapat diketahui akurasi hasil prediksi [KAN23]. Persamaan dari *accuracy* dapat dilihat

pada rumus 2.1

$$Accuracy = \frac{TP+TN}{TP+FP+TN+FN} \dots\dots\dots(2.1)$$

2. Precision

Precision merupakan sebuah metode pengujian dengan melakukan perbandingan jumlah informasi relevan yang didapatkan sistem dengan jumlah seluruh informasi yang terambil oleh sistem baik yang relevan maupun tidak [KAN23]. Persamaan dari precision dapat dilihat pada rumus 2.2

$$Precision = \frac{TP}{TP+FP} \dots\dots\dots(2.2)$$

3. Recall

Recall merupakan metode pengujian yang membandingkan jumlah informasi relevan yang didapatkan sistem dengan jumlah seluruh informasi relevan yang ada dalam koleksi informasi (baik yang terambil atau tidak terambil oleh sistem) [KAN23]. Persamaan dari *recall* dapat dilihat pada rumus 2.3

$$Recall = \frac{TP}{TP+FN} \dots\dots\dots(2.3)$$

4. F1-Score

F1-Score menggambarkan perbandingan rata-rata precision dan *recall* yang dibobotkan. *Accuracy* tepat digunakan sebagai acuan performansi algoritma jika data set kita memiliki jumlah data *false* negatif dan *false* positif yang sangat mendekati (symmetric). Namun jika jumlahnya tidak mendekati, maka sebaiknya menggunakan F1-Score [KAN23]. Persamaan dari F1-Score dapat dilihat pada rumus 2.4

$$F1 = \frac{2 \times Recall \times Precision}{Recall + Precision} \dots\dots\dots(2.4)$$

2.1.12 Pembobotan Kata

Pada proses ini dilakukan setelah tahap preprocessing selesai. Proses pembobotan ini yaitu menggunakan algoritma Term Frequency- Inverse Document Frequency (TF-IDF). Algoritma ini bekerja dengan cara menghitung semua kata yang ada di dokumennya dan kemudian menggunakan rumus yang sudah ada untuk menghitung nilai IDF dari setiap kata yang muncul. Nilai TF dan IDF digabungkan untuk memberikan skor pembobotan kata TF-IDF. Skor tersebut menunjukkan seberapa penting kata tersebut dalam konteks dokumen

dan korpus secara keseluruhan [SUR21].

Algoritma TF-IDF pertama kali dicetuskan oleh salton dan Buckley pada tahun 1988 dan digunakan untuk kepentingan pencairan informasi, yang kemudian turut dimanfaatkan sebagai salah satu algoritma dalam metode pembobotan fitur dalam text mining [PRA19]. Algoritma TF-IDF (*Term Frequency – Invers Document Frequency*) adalah salah satu algoritma yang digunakan untuk menganalisa hubungan antara sebuah frase/kalimat dengan sekumpulan dokumen. Inti pada algoritma ini adalah akan melakukan perhitungan TF dan nilai IDF dari sebuah setiap kata kunci masing-masing dokumen. TF-IDF memiliki formula sebagai berikut:

$$TF - IDF = TF \times IDF \dots\dots\dots(2.5)$$

Rumus tersebut dapat diuraikan menjadi term frequency dari fitur i pada dokumen j dikalikan dengan IDF dari fitur i pada dokumen j. dimana IDF sendiri merupakan kepanjangan dari Inverse Document Frequency. Berikut merupakan rumus untuk TF dan IDF :

$$TF = \frac{\text{Jumlah kemunculan term pada satu dokumen}}{\text{jumlah seluruh term dalam dokumen}} \dots\dots(2.6)$$

$$IDF = \log \frac{\text{Jumlah seluruh dokumen}}{\text{Jumlah dokumen surat term muncul}}$$

2.1.13 Teori Penentuan Sentimen

Penentuan sentimen dalam analisis teks adalah proses mengklasifikasikan suatu pernyataan atau komentar ke dalam kategori positif maupun negatif. Dari sisi bahasa, hal ini sangat bergantung pada konteks kalimat yang digunakan serta pilihan kata yang muncul.

Sentimen positif biasanya mencerminkan perasaan puas dan senang, yang ditunjukkan kata atau frasa bernuansa pujian, apresiasi, maupun dukungan. Sementara sentimen negatif mencerminkan ketidakpuasan, kekecewaan, atau kemarahan, yang tampak melalui kosakata bernuansa kritik maupun penolakan. Dengan demikian, analisis sentimen tidak hanya mengandalkan teknik komputasi, tetapi juga pemahaman terhadap makna bahasa yang berfungsi sebagai penanda orientasi sentimen.

2.2 Penelitian Terdahulu

Tabel 2. 2 Penelitian Terdahulu

No	Judul	Peneliti	Alasan Digunakan	Hasil Penelitian
1	Analisis Sentimen Bakal Calon Presiden 2024 Menggunakan Algoritma Support Vector Machine Pada Twitter.	Wilson Winarto, Izmy Alwiah Musdar, Hasniati	Jurnal ini relevan dengan penelitian karena membahas penerapan algoritma Support Vector Machine (SVM) untuk analisis sentimen berbasis media sosial, terutama twitter. Proses yang dijelaskan sesuai dengan kebutuhan penelitian, termasuk pengumpulan data, preprocessing, pelabelan sentimen, dan evaluasi model dengan cross-validation k-fold. Selain itu, metrik akurasi dan presisi yang digunakan membantu dalam penilaian efektivitas metode SVM.	Penelitian ini menunjukkan bahwa dari 16.936 tweet yang dikumpulkan, 10.448 tweet lolos preprocessing. Sentimen terhadap Anies Baswedan didominasi negatif, Ganjar Pranowo seimbang, dan Prabowo Subianto lebih banyak positif. Akurasi tertinggi pada data Prabowo (83,904%) dan terendah pada Anies (78,28%). Wordcloud menunjukkan tema utama seperti "presiden", "dukung", dan "pilpres". Hasilnya membuktikan efektivitas SVM dalam analisis sentimen media sosial.
2	Prediksi Risiko Perjalanan Transportasi Online Dari Data Telematik Menggunakan Algoritma Support Vector Machine	Christ Memory Sitorus, Adhi Rizal, Muhammad Jajuli	Jurnal ini membahas penerapan algoritma SVM (Support Vector Machine) untuk mengklasifikasikan risiko perjalanan kedalam kategori berdasarkan data telematik. Proses penelitian ini dimulai dengan pengumpulan data, preprocessing, pengembangan fitur, hingga pemodelan dan penilaian yang dilakukan dengan cross-validation. Selain itu, pembahasan parameter C dalam SVM menjadi pertimbangan dalam desain model yang optimal.	Penelitian ini menghasilkan model SVM dengan akurasi terbaik 65,63% dan skor AUC 71,79 menggunakan parameter C=100. Model dikembangkan dari data telematik Grab dengan total 16 juta lebih data pergerakan. Penelitian ini menunjukkan bahwa data telematik efektif digunakan untuk memprediksi risiko perjalanan.
3	Seleksi Fitur Support Vector Machine pada Analisis Sentimen Keberlanjutan Pembelajaran Daring.	Amar P. Natasuwarna	Referensi ini digunakan karena membahas analisis sentimen terhadap pernyataan Mendikbud tentang kelanjutan pembelajaran daring pasca-pandemi Covid-19, yang merupakan masalah penting dalam konteks pendidikan saat ini. Penelitian ini menggunakan algoritma Support Vector Machine (SVM) dan dua metode seleksi fitur (Term Frequency dan TF-IDF), yang memberikan gambaran mendalam tentang bagaimana masyarakat menanggapi kebijakan melalui data Twitter yang valid.	Hasil penelitian ini menunjukkan bahwa algoritma SVM berhasil mengklasifikasikan komentar masyarakat menjadi positif dan negatif. Dengan k=8, algoritma ini mencapai akurasi tertinggi 86,00% dalam validasi cross-fold k.
4	Analisis Sentiment Cyberbullying Pada Sosial Media Twitter Menggunakan Metode	Muh. Fitra Rizki, Karina Auliasari, Renaldi Primaswara Prasetya	Jurnal ini relevan karena membahas penggunaan algoritma Support Vector Machine (SVM) linier dalam	Penelitian ini menunjukkan bahwa metode Support Vector Machine (SVM) mampu digunakan untuk

	Support Vector Machine		memprediksi kelulusan siswa. Jurnal ini menjelaskan secara rinci proses pengolahan data, pemilihan fitur, dan evaluasi akurasi model, sehingga dapat menjadi referensi penting dalam penggunaan SVM untuk prediksi.	melakukan klasifikasi dan prediksi dengan baik, termasuk dalam menangani kasus data yang linier maupun non-linier menggunakan konsep kernel.
5	Metode Algoritma Support Vector Machine (SVM) Linier Dalam Memprediksi kelulusan Mahasiswa	Oktaviana Bangun, Herman Mawengkang, Syahril Efendi	Jurnal ini relevan karena membahas penggunaan algoritma Support Vector Machine (SVM) linier dalam memprediksi kelulusan siswa. Jurnal ini menjelaskan secara rinci proses pengolahan data, pemilihan fitur, dan evaluasi akurasi model, sehingga dapat menjadi referensi penting dalam penggunaan SVM untuk prediksi.	Penelitian tersebut menunjukkan bahwa masa studi mahasiswa dapat dilihat dengan nilai IPK yang akan diprediksi dan apakah mahasiswa tersebut akan lulus atau tidak. Dengan menggunakan algoritma Support Vector machine (SVM) dan menggunakan 792 data training.
6	Penggunaan Bahasa Pemrograman Python dalam Menganalisis Hubungan Kualitas Kopi dengan Lokasi Pertanian Kopi	Regina Lo ¹ , Alfred Edbert Yunanto ² , Rasya Nuhaifa Movia ³ , Lambert Aditama Soehardjianto ⁴ , Ferdinand Wangsa ⁵ , Natalie Asdyo Lidjaja ⁶ , Rahmi Yulia Ningsih ⁷	Jurnal ini relevan karena menggunakan metode Support Vector Machine (SVM) sebagai bagian dari pendekatan analisis data kuantitatif. Selain itu, jurnal ini menggunakan bahasa pemrograman Python dan teknik Exploratory Data Analysis (EDA) untuk memahami hubungan antara variabel kualitas kopi dengan lokasi pertaniannya, khususnya dari segi ketinggian.	Hasil penelitian menunjukkan bahwa semakin tinggi lokasi kopi ditanam, semakin baik kualitas kopi yang dihasilkan. Python membantu mempermudah analisis data serta menghasilkan visualisasi yang memperlihatkan korelasi positif antara ketinggian dan kualitas kopi. Temuan ini bermanfaat untuk petani dan pelaku usaha kopi dalam memilih lokasi tanam yang optimal.

BAB 3

METODE PENELITIAN

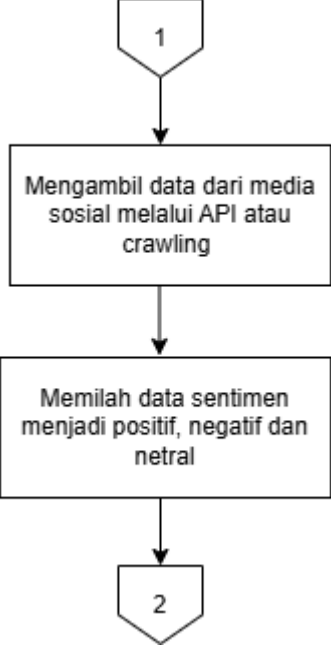
Bab ini berisi penjelasan alur alur penyelesaian tugas akhir, perumusan masalah, kerangka pemikiran teoritis.

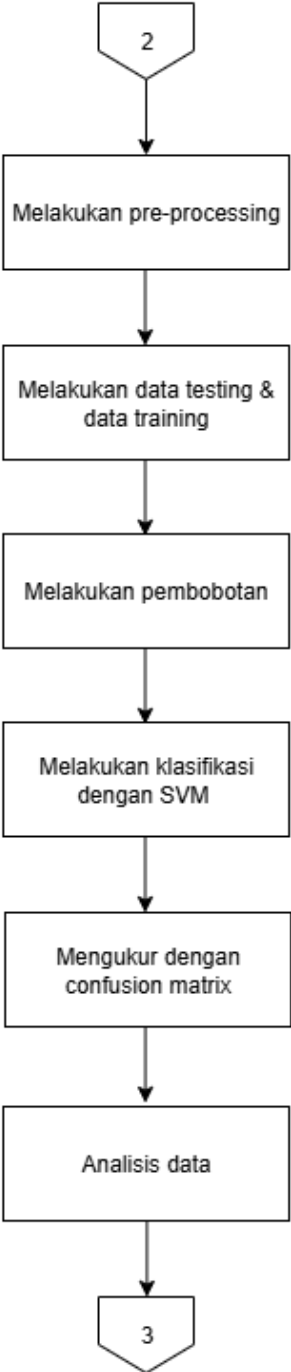
3.1 Alur Penyelesaian Tugas Akhir

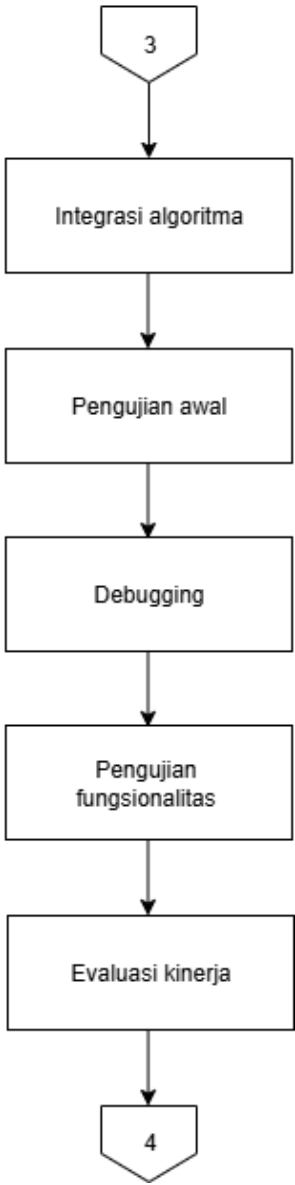
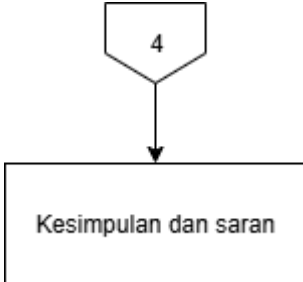
Alur penelitian merupakan kerangka tugas akhir untuk menjelaskan mengenai tahapan atau langkah-langkah dalam menyelesaikan tugas akhir ini.

Tabel 3. 1 Alur Penyelesaian Tugas Akhir

Tahap dan Hasil	Langkah penelitian	Literatur
Tahap 1 : Menentukan topik tugas akhir, mengidentifikasi masalah tugas akhir dan menentukan tujuan serta lingkup tugas akhir Hasil : Menetapkan rumusan masalah Kontribusi : Untuk menentukan ruang lingkup dan tujuan dari penelitian tugas akhir	<pre> graph TD A[Menentukan topik tugas akhir] --> B[Menentukan maslaah yang akan diselesaikan di tugas akhir] B --> C[Menentukan tujuan dan lingkup tugas akhir] C --> D{{1}} </pre>	1. Latar belakang 2. Identifikasi masalah 3. Tujuan dan manfaat tugas akhir 4. Lingkup tugas akhir
Tahap 2 : Pengumpulan data Hasil : Dataset yang relevan untuk analisis. Kontribusi : Menyediakan data untuk proses analisis.		

	 <pre> graph TD 1{{1}} --> A[Mengambil data dari media sosial melalui API atau crawling] A --> B[Memilah data sentimen menjadi positif, negatif dan netral] B --> 2{{2}} </pre>	
Tahap 3: Pengolaan data Hasil : Sistem analisis sentimen berhasil dibangun dengan tahapan preprocessing, pembagian data, pembobotan, klasifikasi SVM, dan evaluasi confusion matrix sehingga dapat mengklasifikasikan sentimen positif dan negatif dengan akurasi tertentu.		

	 <pre> graph TD 2{{2}} --> A[Melakukan pre-processing] A --> B[Melakukan data testing & data training] B --> C[Melakukan pembobotan] C --> D[Melakukan klasifikasi dengan SVM] D --> E[Mengukur dengan confusion matrix] E --> F[Analisis data] F --> 3{{3}} </pre>	
Tahap 4 : Implementasi sistem Hasil : Sistem website analisis sentimen berhasil dibangun dengan fitur utama berupa analisis teks, data ulasan, preprocessing, pembobotan TF-IDF, dan klasifikasi		

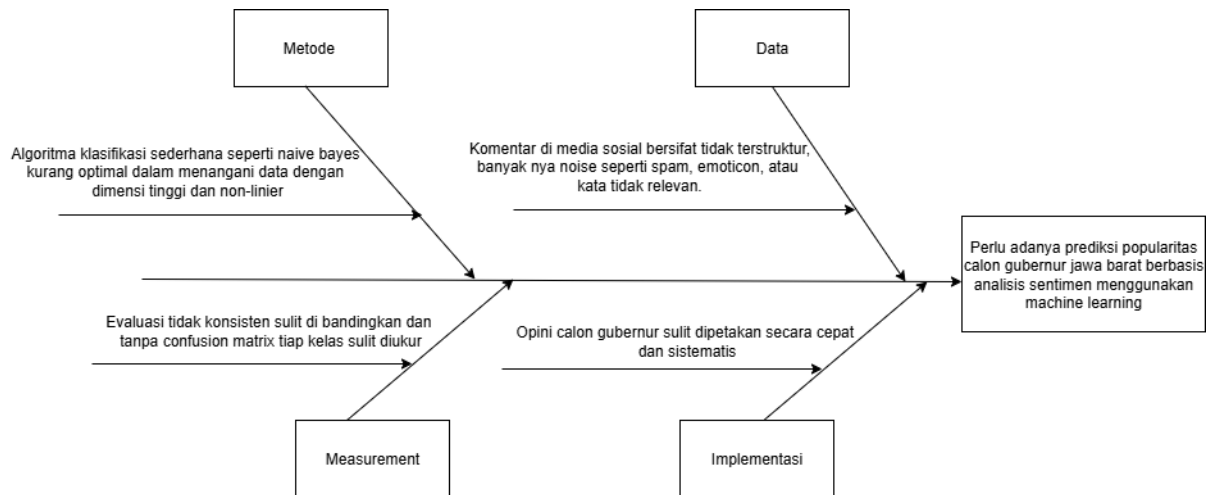
menggunakan SVM. Hasil prediksi ditampilkan pada antarmuka website dalam bentuk label sentimen dan visualisasi distribusi sentimen sehingga memudahkan pengguna dalam memahami kecenderungan opini pada program kampanye calon gubernur Jawa Barat.	 <pre> graph TD 3{{3}} --> A[Integrasi algoritma] A --> B[Pengujian awal] B --> C[Debugging] C --> D[Pengujian fungsionalitas] D --> E[Evaluasi kinerja] E --> 4{{4}} </pre>	
Tahap 6 : Kesimpulan dan Saran Hasil : Kesimpulan dan hasil tugas akhir serta saran dari tugas akhir	 <pre> graph TD 4{{4}} --> F[Kesimpulan dan saran] </pre>	

3.2 Perumusan Masalah

Analisis masalah dilakukan untuk mengidentifikasi penyebab utama dari suatu permasalahan serta mengetahui hubungan sebab akibat permasalahan. Hubungan dari masalah dengan solusi tugas akhir untuk mengetahui ketepatan solusi dalam menyelesaikan permasalahan tugas akhir.

3.3.1 Analisis Masalah

Analisis faktor-faktor permasalahan yang terjadi dalam sistem menjadi fokus analisis. Di bawah ini merupakan deskripsi dari penyebab masalah yang terjadi.



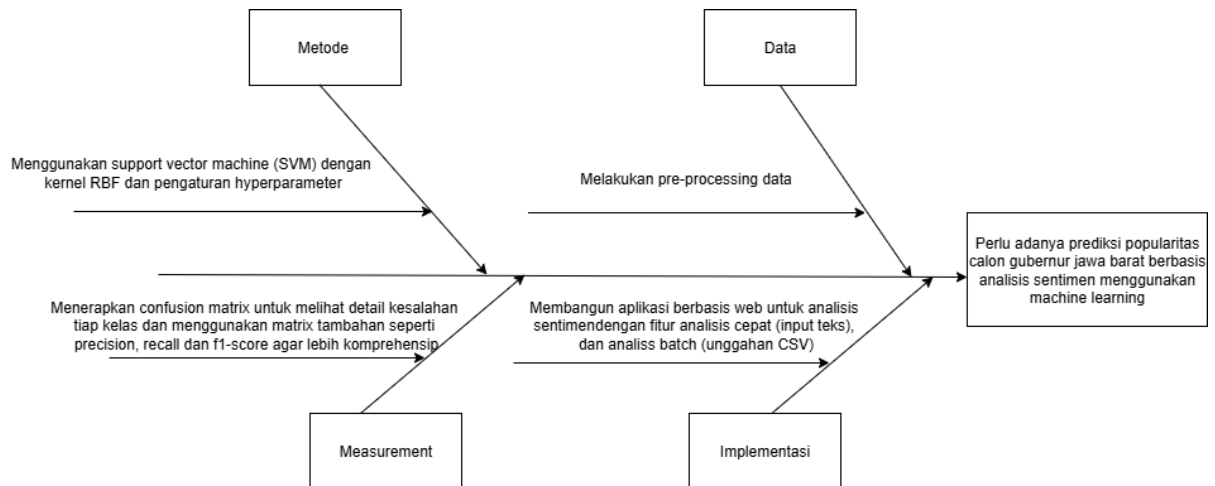
Gambar 3. 1 Fishbone Masalah

Terdapat 3 komponen utama pada diagram *fishbone*, yaitu sebagai berikut :

1. Metode : Algoritma klasifikasi sederhana seperti naive bayes kurang optimal Dalam menangani data dengan dimensi tinggi dan non-linier.
2. Measurement : Evaluasi tidak konsisten sulit di bandingkan dan tanpa confusion matrix tiap kelas sulit diukur.
3. Data : Komentar di media sosial bersifat tidak terstruktur, banyak nya noise seperti spam, emoticon, atau kata tidak relevan.
4. Implementasi : Opini calon gubernur sulit dipetakan secara cepat dan sistematis

3.3.2 Solusi Masalah

Permasalahan yang ada harus memiliki sebuah solusi yang dapat menyelesaikan masalahnya, hal ini saling terhubung untuk dapat bisa menjawab setiap masalah dengan solusi yang baik, berikut adalah solusinya:



Gambar 3. 2 Fishbone Solusi

Terdapat 3 komponen utama pada diagram *fishbone*, yaitu sebagai berikut :

1. Metode : Menggunakan Support Vector Machine (SVM) dengan kernel RBF dan pengaturan hyperparameter
2. Measurement, Menerapkan confusion matrix untuk melihat detail kesalahan tiap kelas dan menggunakan matrix tambahan seperti precision, recall dan F1-score agar lebih evaluasi lebih komprehensif.
3. Data, Melakukan Pre-processing data.
4. Implementasi : Membangun aplikasi berbasis web untuk analisis sentimen dengan fitur analisis cepat (input teks), dan analisis batch (unggah CSV).

BAB 4

ANALISIS DAN PERANCANGAN

4.1 Analisis

4.1.1 Analisis Kebutuhan dan Masalah

Analisis kebutuhan dan masalah didapat dari hasil pencairan data tentang opini masyarakat terhadap program kampanye calon gubernur Jawa Barat di media sosial, khusus nya Twitter (X), yang dimana menjadi permasalahan adalah belum tersedianya media yang mengelompokkan opini berdasarkan sentimen secara otomatis, sehingga penulis membuat sebuah rancangan sistem untuk menganalisis dan memetakan tanggapan masyarakat terhadap program kampanye tersebut.

Tabel 4. 1 Analisis kebutuhan dan masalah

Kebutuhan	Masalah
Sistem untuk mengumpulkan dan menganalisis opini publik terhadap program kampanye	Data dari media sosial tidak terstruktur dan sulit diolah secara manual karena mengandung symbol, singkatan, dan Bahasa tidak baku
Algoritma klasifikasi yang mampu menangani teks pendek dan informal	Komentar masyarakat bersifat ambigu dan tidak eksplisit, sehingga sulit dianalisis secara manual
Representasi teks yang sesuai agar bisa diolah oleh model machine learning	Diperlukan Teknik pembobotan kata atau TF-IDF yang sesuai
Visualisasi hasil analisis dalam bentuk yang mudah dipahami	Sulit mengetahui tren opini publik terhadap topik tertentu tanpa visualisasi

4.1.2 Penentuan karakteristik Data

Langkah penentuan karakteristik data adalah mengumpulkan data dari media sosial Twitter/X, dengan menggunakan kata kunci yang digunakan dalam pengumpulan data mencakup topik program-program yang berkaitan dengan program kampanye calon gubernur Jawa Barat. Karakteristik ini meliputi sumber, format, jumlah data, bahasa, label sentimen dan proses labeling. Penentuan karakteristik dapat dilihat pada tabel 4.2

Tabel 4. 2 Karakteristik data

Karakteristik	Deskripsi
Sumber	Twitter/X
Format	CSV
Jumlah Data	Sekitar 4000 tweet (1000 per paslon)
Bahasa	Bahasa Indonesia
Label Sentimen	Positif dan Negatif
Proses Labeling	Dilakukan secara manual oleh pakar

4.1.3 Penetapan Tujuan

Tujuan dalam konteks klasifikasi sentimen terhadap program kampanye calon gubernur Jawa Barat adalah pernyataan yang ingin dicapai melalui sistem. Tujuan klasifikasi sentimen adalah sebagai berikut:

- Mengklasifikasikan opini publik terhadap program kampanye calon gubernur Jawa Barat menjadi sentimen positif dan negatif.
- Menyediakan antarmuka web yang memudahkan pengguna dalam menginput teks atau melihat hasil klasifikasi.
- Menyediakan visualisasi data seperti grafik distribusi sentimen dan wordcloud.

4.1.4 Penentuan Ide Sistem

Setelah menetapkan tujuan pengembangan sistem akan mulai menentukan ide sistem yang akan dibangun. Penentuan ide ini mencakup beberapa aspek utama yang sesuai dengan kebutuhan pengguna dan tujuan yang ingin dicapai. Ide sistem dapat dilihat pada tabel 4.3

Tabel 4. 3 Ide sistem

Aspek	Deskripsi
Fitur	Input teks opini → proses pre-processing → klasifikasi dengan SVM → ditampilkan pada web
Visualisasi	Diagram pie chart dan grafik batang
Akurasi	Model diuji dengan metrik akurasi, precision, recall, dan F1-Score
Teknologi	Teknologi yang akan digunakan berupa webiste

4.1.5 Definisi Lingkungan

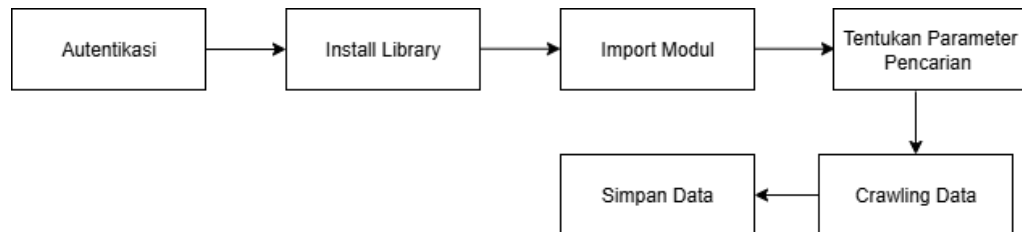
Tahap terakhir analisis adalah mendefinisikan lingkungan, dalam tahap selanjutnya yaitu pembuatan sistem klasifikasi sentimen dengan menyesuaikan karakteristik lingkungan pengembangan yang sudah dirancang seperti yang dapat dilihat pada tabel 4.4

Tabel 4. 4 Lingkungan Klasifikasi

Karakteristik Lingkungan	Deskripsi
Platform Pengembangan	Aplikasi berbasis web, dijalankan melalui browser
Bahasa Pemrograman	HTML, CSS dan Python menggunakan framework Flask
Library	Scikit-learn, Pandas, Matplotlib, Wordcloud, Seaborn
Metode pengembangan	Prototyping → mulai dari Pre-processing dan pelatihan model
Pengguna	Peneliti, mahasiswa dan tim kampanye

4.2 Pengumpulan Data

Pengumpulan data komentar yang akan digunakan dalam melakukan penelitian ini dapat dilakukan dengan menggunakan aplikasi Tweet Harvester, data penelitian ini diperoleh melalui proses crawling data dari platform Twitter. Memanfaatkan API Twitter, program ini dilakukan secara otomatis mengumpulkan tweet berdasarkan kata kunci tertentu, seperti “acep adang cagub”, “jeje wiradinata cagub”, “ahmad syaikhu cagub”, dan “dedi mulyadi cagub”. Data yang dikumpulkan dalam pengaturan Bahasa Indonesia dan dalam jangka waktu tertentu. Hasil crawling tersebut di simpan dalam file `acep_adang_cagub.csv`, yang berisikan berupa kumpulan tweet mentah yang berkaitan dengan program kampanye cagub jawa barat. File ini kemudian digunakan sebagai dataset untuk proses analisis sentimen lebih lanjut. Berikut adalah proses crawling untuk pengumpulan data pada penelitian ini.



Gambar 4. 1 Proses Crawling

Dimana untuk memperoleh data ulasan pengguna di media sosial proses yang harus dilakukan seperti gambar di atas, seperti

- a. Autentikasi : Memasukan token autentikasi twitter sebagai akses awal
- b. Install Library : Install library *tweet-harvest* dan *Node.js* untuk proses crawling data
- c. Import Modul: Impor modul yang digunakan seperti *pandas*.
- d. Tentukan Parameter Pencarian : Menentukan keyword serta jumlah tweet yang akan diambil.
- e. Crawling Data : jalankan proses pengambilan data tweet sesuai parameter pencarian.
- f. Simpan data : menyimpan data yang sudah diolah dan dipilih ke dalam format yang sesuai dalam bentuk file *.CSV*.

Berikut merupakan hasil dari dataset pada pengumpulan data setiap paslon.

Tabel 4. 5 Hasil Dataset Pengumpulan data Paslon 1

No	Tweet
1	@Telkomsel Signal lup lep ga bener. Di dusun patinggen 1. Desa Karangpawitan. kec Padaherang kab Pangandaran Jabar. Tolong perbaiki
2	@ekonomi_oke Acep janjiin listrik gratis, tapi nggak ada rincian biayanya. Jangan cuma omong doang!
3	@anti_nepot Tim sukses Acep-Adang isinya keluarga semua. Emang mau jadi dinasti politik?
4	27 November nanti jangan lupa coblos Acep - Gita ya! Kita semua berharap Jabar bisa menjadi lebih maju dan lebih bahagia. Yuk kita sukseskan bersama! https://t.co/BlpiK5AhK3

Tabel 4. 6 Hasil Dataset Pengumpulan data Paslon 2

No	Tweet
1	Jadi calon Wakil Gubernur Jabar dampingi Jeje Wiradinata pose tak biasa Ronal Surapradja di surat suara curi perhatian https://t.co/hmS5vwiTaH
2	@rini_warga Lagi-lagi janji manis dari Jeje, tapi realisasi nihil. Udah capek percaya, kenyataan gak sesuai ucapan.
3	@dadan_smi Salut buat Pak Jeje yang selalu turun langsung ke lapangan, bukan hanya duduk di kantor. Pemimpin seperti ini langka.
4	Satu suara untuk kang Jeje (cagub jabar) Dan satu untuk kang Karna (cabup Majalengka) https://t.co/d73SbVPxgMT topik sekarang kok nggak dibahas?

Tabel 4. 7 Hasil Dataset Pengumpulan data Paslon 3

No	Tweet
1	@BangPino_ Pria lames @PKSejahtera cuma wakil aja di DKI warga Jakarta ga percaya lagi setelah berhianat @MardaniAliSera dibutakan akan kebodohnya sendiri. Jabar siapa yang endorse @syaikhu_ahmad kenapa Depok kalah? dah ga dipercaya. Suswono ga menjual
2	@aagus.suherman Maa syaa Allah. Panutan sekali ustadz. Memberikan contoh bahwa kekalahan dalam kontestasi bukan berarti tidak berkontribusi. Saya yakin Allah siapkan posisi yg terbaik pasca pilkada ini. ????
3	@PKSejahtera @syaikhu_ahmad Sudah Lelah 10 Tahun Jadi Oposisi Yang digaji sama Rakyat aja Lelah trus Bagaimana Rakyat yang suaranya dijual ke Rezim Dinasti Oligarki Mulyono. https://t.co/ORopQRGUKv .
4	@dwiwahyanto013bahagiaBerjiwa besar & Legowo... mantaps Ustd. Syaikhu, cerminan seorang negarawan sejati. Menang kalah itulah kompetisi, yg terpenting kita lihat dan dukung jalannya roda pemerintahan yg dipimpin oleh pak Dedi mendatang... selamat bertugas kepada pak Dedi & tetap semangat Ustd. Syaikhu terus tebar kebaikan serta tetap membangun negeri ini ??????????

Tabel 4. 8 Hasil Dataset Pengumpulan data Paslon 4

No	Tweet
1	@BayuAngora Sebaiknya Dedi Mulyadi fokus soal mensejahterakan rakyat Jabar. Soal keyakinannya jadikan itu urusan pribadi dan tidak perlu diumbar mau nyembah patung kek mau sesajen kek. Umat Islam itu unik kalau agamanya disenggol bahkan yang islam KTP saj
2	Dedi Mulyadi Kedatangan Keluarga Besar Bos Persib Umuh Muchtar di Subang https://t.co/KLA5a73AI2

- 3 Satu-satunya video Dedy Mulyadi yang jelek adalah ketika Dedi Mulyadi membunuh karakter mahasiswa HMI yang menanyakan tupoksi Dedy Mulyadi yg seorang DPR RI tapi malah sibuk bersihin sampah pasar. Selain itu apik kabeh. Terlebih lagu-lagu EMKA 9.
- 4 @tekarok007 Ga salah ayy coblos kang Dedi Mulyadi kemarin <https://t.co/CznK1ondYD>

4.3 Pelabelan Data

Setelah data sudah melewati tahap reduksi dan memenuhi kriteria yang diperlukan dalam penelitian, langkah selanjutnya adalah melakukan pelabelan data. Dalam proses pelabelan pada analisis sentimen data Media Sosial, data yang didapat belum mempunyai label positif dan negatif. Pelabelan dilakukan sebagai acuan pada proses klasifikasi menggunakan metode Support Vector Machine, yang nantinya akan menjadi tolak ukur pada perhitungan dalam proses klasifikasi. Untuk proses pelabelan peneliti dibantu oleh pakar Bahasa yaitu Ibu Lani Purniawati, S.S yang merupakan Guru Bahasa Indonesia yang membantu memvalidasi data. Hal ini dilakukan untuk menjaga keakuratan data. Adapun data yang divalidasi oleh pakar yaitu sebanyak 4000 data sentimen, yang terdiri dari masing-masing 1000 data setingan pasangan calon yaitu paslon 1 “Bapak Drs. K.H. Acep Adang Ruhiat, M.Si.”, paslon 2 “Bapak Jeje Wiradinata”, paslon 3 “Bapak Ahmad Syaikhul”, paslon 4 “Bapak H. Dedi Mulyadi, S.H., M.M”. Sehingga keempat paslon tersebut mewakili seluruh data pada pemilihan Gubernur Jawa Barat terwakili secara proporsional dalam proses analisis sentimen. Berikut merupakan contoh data yang sudah dilabelisasi oleh pakar pada sentimen program kampanye Calon Gubernur Jawa Barat pada setiap pasangan calon (paslon).

Tabel 4. 9 Hasil Labeling Data Paslon 1

Sentimen	Label
@Telkomsel Signal lup lep ga bener. Di dusun patinggen 1. Desa Karangpawitan. kec Padaherang kab Pangandaran Jabar. Tolong perbaiki	Negatif
@ekonomi_oke Acep janjiin listrik gratis, tapi nggak ada rincian biayanya. Jangan cuma omong doang!	Positif
@anti_nepot Tim sukses Acep-Adang isinya keluarga semua. Emang mau jadi dinasti politik?	Negatif
27 November nanti jangan lupa coblos Acep - Gita ya! Kita semua berharap Jabar bisa menjadi lebih maju dan lebih bahagia. Yuk kita sukseskan bersama! https://t.co/BlpiK5AhK3	Positif

Tabel 4. 10 Hasil Labeling Data Paslon 2

Sentimen	Label
Jadi calon Wakil Gubernur Jabar dampingi Jeje Wiradinata pose tak biasa Ronal Surapradja di surat suara curi perhatian https://t.co/hmS5vwiTaH	Positif
@rini_warga Lagi-lagi janji manis dari Jeje, tapi realisasi nihil. Udah capek percaya, kenyataan gak sesuai ucapan.	Negatif
@dadan_smi Salut buat Pak Jeje yang selalu turun langsung ke lapangan, bukan hanya duduk di kantor. Pemimpin seperti ini langka.	Positif
@fitri_mahasiswa Setiap musim hujan, daerah saya banjir. Mana perhatian dari Jeje Wiradinata? Gak kelihatan aksi nyatanya.	Negatif

Tabel 4. 11 Hasil Labeling Data Paslon 3

Sentimen	Label
@BangPino__ Pria lames @PKSejahtera cuma wakil aja di DKI warga Jakarta ga percaya lagi setelah berkhianat @MardaniAliSera dibutakan akan kebodohnya sendiri. Jabar siapa yang endorse @syaikhu_ahmad kenapa Depok kalah? dah ga dipercaya. Suswono ga menjual	Negatif
@aagus.suherman Maa syaa Allah. Panutan sekali ustadz. Memberikan contoh bahwa kekalahan dalam kontestasi bukan berarti tidak berkontribusi. Saya yakin Allah siapkan posisi yg terbaik pasca pilkada ini. ????	Positif
@PKSejahtera @syaikhu_ahmad Sudah Lelah 10 Tahun Jadi Oposisi Yang digaji sama Rakyat aja Lelah trus Bagaimana Rakyat yang suaranya dijual ke Rezim Dinasti Oligarki Mulyono. https://t.co/ORopQRGUkv .	Negatif
@dwiwahyanto013bahagiaBerjiwa besar & Legowo... mantaps Ustd. Syaikhu, cerminan seorang negarawan sejati. Menang kalah itulah kompetisi, yg terpenting kita lihat dan dukung jalannya roda pemerintahan yg dipimpin oleh pak Dedi mendatang... selamat bertugas kepada pak Dedi & tetap semangat Ustd. Syaikhu terus tebar kebaikan serta tetap membangun negeri ini ??????????	Positif

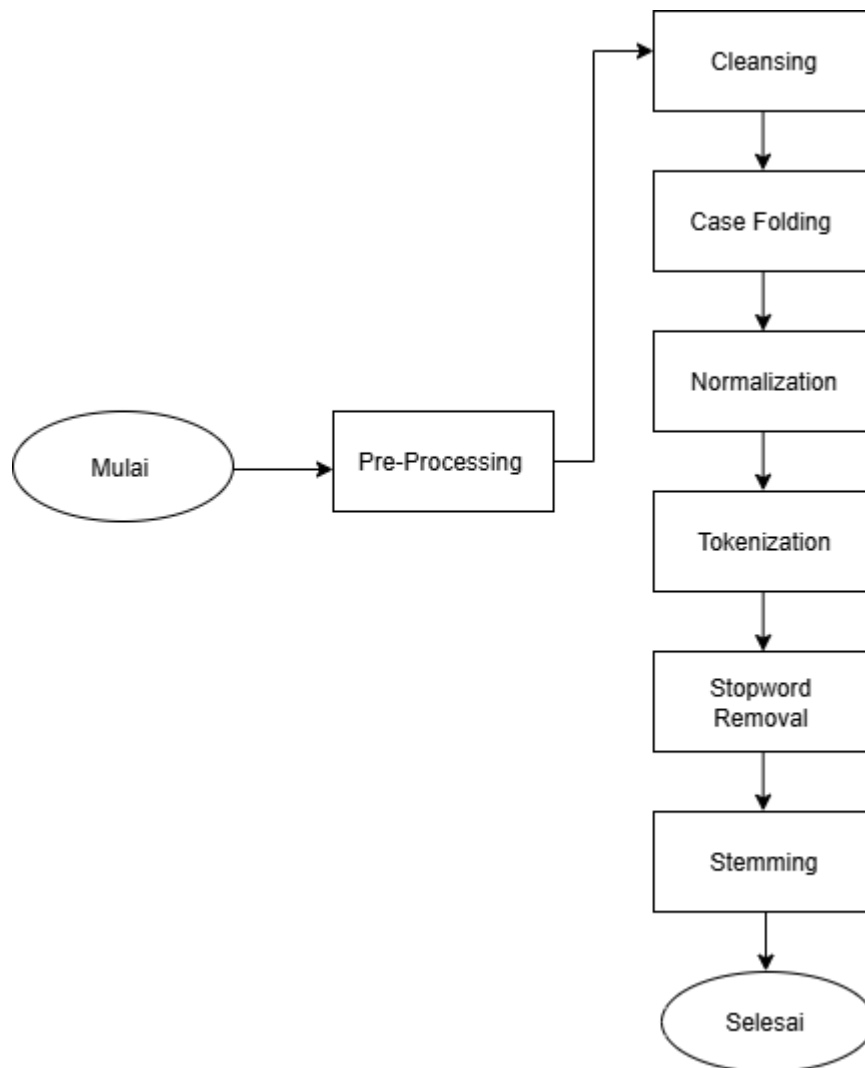
Tabel 4. 12 Hasil Labeling data Paslon 4

Sentimen	Label
@BayuAngora Sebaiknya Dedi Mulyadi fokus soal mensejahterakan rakyat Jabar. Soal keyakinannya jadikan itu urusan pribadi dan tidak perlu diumbar mau nyembah patung kek mau sesajen kek. Umat Islam itu unik kalau agamanya disenggol bahkan yang islam KTP saj	Negatif
Dedi Mulyadi Kedatangan Keluarga Besar Bos Persib Umuh Muchtar di Subang https://t.co/KLA5a73AI2	Positif
Satu-satunya video Dedy Mulyadi yang jelek adalah ketika Dedi Mulyadi membunuh karakter mahasiswa HMI yang menanyakan tupoksi Dedy Mulyadi yg seorang DPR RI tapi malah sibuk bersihin sampah pasar. Selain iku apik kabeh. Terlebih lagu-lagu EMKA 9.	Negatif
@tekarok007 Ga salah ayy coblos kang Dedi Mulyadi kemarin https://t.co/CznK1ondYD	Positif

4.4 Pre-Processing

Tahap pre-processing adalah tahap dimana mempersiapkan data sebelum dilakukan analisis. Pre-processing akan melakukan olah data meliputi merapikan data, membersihkan data serta mengubah kata ke huruf kecil. Adapun tahapan yang akan digunakan dalam melakukan pre-processing yaitu *cleaning*, *case folding*, *normalization*, *tokenization*, *stopword removal*, dan *stemming*. Tujuan dari tahapan ini adalah untuk membuat teks lebih mudah dipahami dan diolah oleh algoritma mesin. Dalam kebanyakan kasus, data yang diperoleh dari X masih dalam bentuk data mentah dan mengandung elemen yang tidak relevan. Contohnya seperti tanda baca, tautan (URL), mention, dan emoticon, yang dapat menyulitkan analisis.

Berikut adalah tahapan dari pre-processing.



Gambar 4. 2 Tahapan Pre-processing

4.4.1 Cleaning

Proses cleaning bertujuan untuk membersihkan kata dengan menghilangkan atau penghapusan simbol yang dianggap tidak penting seperti hastag, mention, emoticon, karakter dan menyeragamkan bentuk data yang dianggap tidak memiliki arti sehingga dapat memudahkan data untuk di proses lebih lanjut.

Kode di bawah ini dibersihkan secara bertahap dengan menggunakan fungsi khusus yang dirancang untuk menghilangkan berbagai elemen tersebut

Kode Program 4. 1 Cleaning Data

```
df['cleaning'] = df['full_text '].apply(lambda x: remove_URL(x))
df['cleaning'] = df['cleaning'].apply(lambda x: remove_html(x))
df['cleaning'] = df['cleaning'].apply(lambda x: remove_numbers(x))
df['cleaning'] = df['cleaning'].apply(lambda x: remove_username(x))
df['cleaning'] = df['cleaning'].apply(lambda x: remove_symbols(x))
```

Pada kode program diatas terdapat beberapa fungsi pembersihan data, baris pertama digunakan untuk menghapus URL atau tautan yang biasanya muncul dalam teks tweet . Kemudian, kode mengambil full_text yang berisikan tweet mentah sebagai data utama. Akun X yang biasanya diawali dengan tanda (@) tidak dapat digunakan untuk analisis, baris berikutnya membersihkan akun tersebut. Selanjutnya, komponen seperti kode HTML yang kadang-kadang muncul dalam teks juga dihilangkan. Emoji yang menunjukkan emosi kadang-kadang dianggap mengganggu dan dihapus untuk analisis yang konsisten, selain itu juga symbol dan karakter yang tidak termasuk huruf dan angka dalam tweet dihapus. Dengan proses cleaning ini, teks tweet menjadi lebih bersih, rapi, dan siap untuk dilakukan ke tahap berikutnya

Tabel 4. 13 Tahap Cleaning Paslon 1

Sebelum Cleaning	Setelah Cleaning
@Telkomsel Signal lup lep ga bener. Di dusun patinggen 1. Desa Karangpawitan. kec Padaherang kab Pangandaran Jabar. Tolong perbaiki	Signal lup lep ga bener Di dusun patinggen Desa Karangpawitan kec Padaherang kab Pangandaran Jabar Tolong perbaiki
@ekonomi_oke Acep janjiin listrik gratis, tapi nggak ada rincian biayanya. Jangan cuma omong doang!	Acep janjiin listrik gratis tapi nggak ada rincian biayanya Jangan cuma omong doang
@anti_nepot Tim sukses Acep-Adang isinya keluarga semua. Emang mau jadi dinasti politik?	Tim sukses AcepAdang isinya keluarga semua Emang mau jadi dinasti politik

27 November nanti jangan lupa coblos Acep - Gita ya! Kita semua berharap Jabar bisa menjadi lebih maju dan lebih bahagia. Yuk kita sukseskan bersama! <https://t.co/BlpiK5AhK3>

November nanti jangan lupa coblos Acep Gita ya Kita semua berharap Jabar bisa menjadi lebih maju dan lebih bahagia Yuk kita sukseskan bersama

Tabel 4. 14 Tahap Cleaning Paslon 2

Sebelum Cleaning	Setelah Cleaning
Jadi calon Wakil Gubernur Jabar dampingi Jeje Wiradinata pose tak biasa Ronal Surapradja di surat suara curi perhatian https://t.co/hmS5vwiTaH	Jadi calon Wakil Gubernur Jabar dampingi Jeje Wiradinata pose tak biasa Ronal Surapradja di surat suara curi perhatiin
@rini_warga Lagi-lagi janji manis dari Jeje, tapi realisasi nihil. Udah capek percaya, kenyataan gak sesuai ucapan.	Lagilagi janji manis dari Jeje tapi realisasi nihil Udah capek percaya kenyataan gak sesuai ucapan
@dadan_smi Salut buat Pak Jeje yang selalu turun langsung ke lapangan, bukan hanya duduk di kantor. Pemimpin seperti ini langka.	Salut buat Pak Jeje yang selalu turun langsung ke lapangan bukan hanya duduk di kantor Pemimpin seperti ini langka
@fitri_mahasiswa Setiap musim hujan, daerah saya banjir. Mana perhatian dari Jeje Wiradinata? Gak kelihatan aksi nyatanya.	Setiap musim hujan daerah saya banjir Mana perhatian dari Jeje Wiradinata Gak kelihatan aksi nyatanya

Tabel 4. 15 Tahap Cleaning Paslon 3

Sebelum Cleaning	Setelah Cleaning
@BangPino__ Pria lames @PKSejahtera cuma wakil aja di DKI warga Jakarta ga percaya lagi setelah berhianat @MardaniAliSera dibutakan akan kebodohnya sendiri. Jabar siapa yang endorse @syaikhu_ahmad kenapa Depok kalah? dah ga dipercaya. Suswono ga menjual @aagus.suherman Maa syaa Allah. Panutan sekali ustadz. Memberikan contoh bahwa kekalahan dalam kontestasi bukan berarti tidak berkontribusi. Saya yakin Allah siapkan posisi yg terbaik pasca pilkada ini. ????	Pria lames cuma wakil aja di DKI warga Jakarta ga percaya lagi setelah berhianat dibutakan akan kebodohnya sendiri Jabar siapa yang endorse kenapa Depok kalah dah ga dipercaya Suswono ga menjual
@PKSejahtera @syaikhu_ahmad Sudah Lelah 10 Tahun Jadi Oposisi Yang digaji sama Rakyat aja Lelah trus Bagaimana Rakyat yang suaranya dijual ke Rezim Dinasti Oligarki Mulyono. https://t.co/ORopQRGUkv	suherman Maa syaa Allah Panutan sekali ustadz Memberikan contoh bahwa kekalahan dalam kontestasi bukan berarti tidak berkontribusi Saya yakin Allah siapkan posisi yg terbaik pasca pilkada ini
@dwiwahyanto013bahagiaBerjiwa besar & Legowo... mantaps Ustd. Syaikhu, cerminan seorang negarawan sejati. Menang kalah	Sudah Lelah Tahun Jadi Oposisi Yang digaji sama Rakyat aja Lelah trus Bagaimana Rakyat yang suaranya dijual ke Rezim Dinasti Oligarki Mulyono
	besar Legowo mantaps Ustd Syaikhu cerminan seorang negarawan sejati Menang kalah itulah kompetisi yg terpenting

itulah kompetisi, yg terpenting kita lihat dan dukung jalannya roda pemerintahan yg dipimpin oleh pak Dedi mendatang... selamat bertugas kepada pak Dedi & tetap semangat Ustd. Syaikhul terus tebar kebaikan serta tetap membangun negeri ini ??????????	kita lihat dan dukung jalannya roda pemerintahan yg dipimpin oleh pak Dedi mendatang selamat bertugas kepada pak Dedi tetap semangat Ustd Syaikhul terus tebar kebaikan serta tetap membangun negeri ini
--	---

Tabel 4. 16 Tahap Cleaning Paslon 4

Sebelum Cleaning	Setelah Cleaning
@BayuAngora Sebaiknya Dedi Mulyadi fokus soal mensejahterakan rakyat Jabar. Soal keyakinannya jadikan itu urusan pribadi dan tidak perlu diumbar mau nyembah patung kek mau sesajen kek. Umat Islam itu unik kalau agamanya disenggol bahkan yang islam KTP saj Dedi Mulyadi Kedatangan Keluarga Besar Bos Persib Umuh Muchtar di Subang https://t.co/KLA5a73AI2 Satu-satunya video Dedy Mulyadi yang jelek adalah ketika Dedi Mulyadi membunuh karakter mahasiswa HMI yang menanyakan tupoksi Dedy Mulyadi yg seorang DPR RI tapi malah sibuk bersihin sampah pasar. Selain itu apik kabeh. Terlebih lagu-lagu EMKA 9. @tekarok007 Ga salah ayy coblos kang Dedi Mulyadi kemarin https://t.co/CznK1ondYD	Sebaiknya Dedi Mulyadi fokus soal mensejahterakan rakyat Jabar Soal keyakinannya jadikan itu urusan pribadi dan tidak perlu diumbar mau nyembah patung kek mau sesajen kek Umat Islam itu unik kalau agamanya disenggol bahkan yang islam KTP saj Dedi Mulyadi Kedatangan Keluarga Besar Bos Persib Umuh Muchtar di Subang Satusatunya video Dedy Mulyadi yang jelek adalah ketika Dedi Mulyadi membunuh karakter mahasiswa HMI yang menanyakan tupoksi Dedy Mulyadi yg seorang DPR RI tapi malah sibuk bersihin sampah pasar Selain itu apik kabeh Terlebih lagulagu EMKA Ga salah ayy coblos kang Dedi Mulyadi kemarin

4.4.2 Case Folding

Setelah proses pembersihan kata selesai, tahap berikutnya adalah tahap semua teks yang terdapat pada dataset akan di samakan semua hurufnya menjadi huruf kecil atau lowercase. Proses ini bertujuan untuk menyamakan format penulisan suatu kata untuk memastikan bahwa kata-kata yang memiliki makna yang sama tetapi ditulis dengan kombinasi huruf besar dan kecil tidak dianggap sebagai kata yang berbeda, misalnya seperti kata “Signal”, “signal”, dan “SIGNAL” akan diubah menjadi “signal” agar model tidak menganggapnya sebagai tiga kata yang berbeda. Hal ini menunjukkan analisis lebih akurat dan konsisten. Berikut adalah kode program case folding.

Kode Program 4. 2 Case Folding

```
def case_folding(text):
    if isinstance(text, str):
        lowercase_text = text.lower()
        return lowercase_text
    else:
        return text
df['case_folding'] = df['cleaning'].apply(case_folding)
```

Perintah untuk mengubah seluruh teks yang awalnya menggunakan huruf kapital menjadi huruf kecil atau lowercase. Dan langkah ini dilakukan dengan sangat baik dan berhasil pada tahap ini. Berikut adalah hasil data sebelum dan sesudah proses *case folding*.

Tabel 4. 17 Hasil Case Folding Paslon 1

Sebelum Case Folding	Setelah Case Folding
Signal lup lep ga bener Di dusun patinggen	signal lup lep ga bener di dusun patinggen
Desa Karangpawitan kec Padaherang kab	desa karangpawitan kec padaherang kab
Pangandaran Jabar Tolong perbaiki	pangandaran jabar tolong perbaiki
Acep janjiin listrik gratis tapi nggak ada rincian biayanya Jangan cuma omong doang	acep janjiin listrik gratis tapi nggak ada rincian biayanya jangan cuma omong doang
Tim sukses AcepAdang isinya keluarga semua Emang mau jadi dinasti politik	tim sukses acepadang isinya keluarga semua emang mau jadi dinasti politik
November nanti jangan lupa coblos Acep Gita ya Kita semua berharap Jabar bisa menjadi lebih maju dan lebih bahagia Yuk kita sukseskan bersama	november nanti jangan lupa coblos acep gita ya kita semua berharap jabar bisa menjadi lebih maju dan lebih bahagia yuk kita sukseskan bersama

Tabel 4. 18 Hasil Case Folding Paslon 2

Sebelum Case Folding	Setelah Case Folding
Jadi calon Wakil Gubernur Jabar dampingi Jeje Wiradinata pose tak biasa Ronal Surapradja di surat suara curi perhatiin	jadi calon wakil gubernur jabar dampingi jeje wiradinata pose tak biasa ronal surapradja di surat suara curi perhatiin
Lagilagi janji manis dari Jeje tapi realisasi nihil Udah capek percaya kenyataan gak sesuai ucapan	lagilagi janji manis dari jeje tapi realisasi nihil udah capek percaya kenyataan gak sesuai ucapan
Salut buat Pak Jeje yang selalu turun langsung ke lapangan bukan hanya duduk di kantor Pemimpin seperti ini langka	salut buat pak jeje yang selalu turun langsung ke lapangan bukan hanya duduk di kantor pemimpin seperti ini langka

Setiap musim hujan daerah saya banjir
 Mana perhatian dari Jeje Wiradinata Gak
 kelihatan aksi nyatanya

setiap musim hujan daerah saya banjir
 mana perhatian dari jeje wiradinata gak
 kelihatan aksi nyatanya

Tabel 4. 19 Hasil Case Folding Paslon 3

Sebelum Case Folding	Setelah Case Folding
Pria lames cuma wakil aja di DKI warga Jakarta ga percaya lagi setelah berhianat dibutakan akan kebodohnya sendiri Jabar siapa yang endorse kenapa Depok kalah dah ga dipercaya Suswono ga menjual	pria lames cuma wakil aja di dki warga jakarta ga percaya lagi setelah berhianat dibutakan akan kebodohnya sendiri jabar siapa yang endorse kenapa depok kalah dah ga dipercaya suswono ga menjual
suherman Maa syaa Allah Panutan sekali ustadz Memberikan contoh bahwa kekalahan dalam kontestasi bukan berarti tidak berkontribusi Saya yakin Allah siapkan posisi yg terbaik pasca pilkada ini	suherman maa syaa allah panutan sekali ustadz memberikan contoh bahwa kekalahan dalam kontestasi bukan berarti tidak berkontribusi saya yakin allah siapkan posisi yg terbaik pasca pilkada ini
Sudah Lelah Tahun Jadi Oposisi Yang digaji sama Rakyat aja Lelah trus Bagaimana Rakyat yang suaranya dijual ke Rezim Dinasti Oligarki Mulyono	sudah lelah tahun jadi oposisi yang digaji sama rakyat aja lelah trus bagaimana rakyat yang suaranya dijual ke rezim dinasti oligarki mulyono
besar Legowo mantaps Ustd Syaikh cerminan seorang negarawan sejati Menang kalah itulah kompetisi yg terpenting kita lihat dan dukung jalannya roda pemerintahan yg dipimpin oleh pak Dedi mendatang selamat bertugas kepada pak Dedi tetap semangat Ustd Syaikh terus tebar kebaikan serta tetap membangun negeri ini	besar legowo mantaps ustd syaikh cerminan seorang negarawan sejati menang kalah itulah kompetisi yg terpenting kita lihat dan dukung jalannya roda pemerintahan yg dipimpin oleh pak dedi mendatang selamat bertugas kepada pak dedi tetap semangat ustd syaikh terus tebar kebaikan serta tetap membangun negeri ini

Tabel 4. 20 Hasil Case Folding Paslon 4

Sebelum Case Folding	Setelah Case Folding
Sebaiknya Dedi Mulyadi fokus soal mensejahterakan rakyat Jabar Soal keyakinannya jadikan itu urusan pribadi dan tidak perlu diumbar mau nyembah patung kek mau sesajen kek Umat Islam itu unik kalau agamanya disenggol bahkan yang islam KTP saj	sebaiknya dedi mulyadi fokus soal mensejahterakan rakyat jabar soal keyakinannya jadikan itu urusan pribadi dan tidak perlu diumbar mau nyembah patung kek mau sesajen kek umat islam itu unik kalau agamanya disenggol bahkan yang islam ktp saj
Dedi Mulyadi Kedatangan Keluarga Besar Bos Persib Umuh Muchtar di Subang	dedi mulyadi kedatangan keluarga besar bos persib umuh muchtar di subang

Satusatunya video Dedy Mulyadi yang jelek adalah ketika Dedi Mulyadi membunuh karakter mahasiswa HMI yang menanyakan tupoksi Dedy Mulyadi yg seorang DPR RI tapi malah sibuk bersihin sampah pasar Selain itu apik kabeh Terlebih lagulagu EMKA Ga salah ayy coblos kang Dedi Mulyadi kemarin	satusatunya video dedy mulyadi yang jelek adalah ketika dede mulyadi membunuh karakter mahasiswa hmi yang menanyakan tupoksi dede mulyadi yg seorang dpr ri tapi malah sibuk bersihin sampah pasar selain itu apik kabeh terlebih lagulagu emka ga salah ayy coblos kang dede mulyadi kemarin
---	--

4.4.3 Normalization

Tahap ini merupakan tahapan yang akan digunakan untuk mengubah kalimat yang tidak baku atau singkatan menjadi kalimat baku yang sesuai dengan kamus KBBI (Kamus Besar Bahasa Indonesia). Berikut adalah contoh dari tahapan *normalization*.

Kode Program 4.3 Normalization

```
def replace_taboo_words(text, kamus_tidak_baku):
    if isinstance(text, str):
        words = text.split()
        hasil_kalimat = []
        kata_baku = []
        kata_tidak_baku = []
        kata_tidak_baku_hash = []
```

Proses normalisasi merupakan langkah yang sangat penting, terutama saat mengolah data teks dari X yang umumnya mengandung banyak kata tidak baku, singkatan, serta penulisan yang tidak sesuai kaidah bahasa konvensional. Jika kata-kata tidak dinormalisasi, jumlah variasi kata dalam dataset akan meningkat, ini membuat proses analisis menjadi lebih sulit. seperti pelatihan model analisis sentimen dan perhitungan TF-IDF adalah sebagai acuan untuk proses normalisasi kata dalam penelitian ini. Peneliti menggunakan kamus kata baku yang diperoleh dari platform Kaggle, pada variasi penulisan kata yang sebenarnya dengan makna yang sama dapat diseragamkan dengan menggunakan normalisasi. Ini akan memungkinkan model untuk belajar dengan lebih baik dan hasil analisis menjadi lebih akurat. Berikut adalah hasil dari tahapan *normalization*.

Tabel 4. 21 Hasil Normalization Paslon 1

Sebelum Normalisasi	Setelah Normalisasi
Signal lup lep ga bener Di dusun patinggen Desa Karangpawitan kec Padaherang kab Pangandaran Jabar Tolong perbaiki	signal lup lep tidak benar di dusun patinggen desa karangpawitan kec padaherang kab pangandaran jabar tolong perbaiki
Acep janjiin listrik gratis tapi nggak ada rincian biayanya Jangan cuma omong doang	acep janjiin listrik gratis tapi tidak ada rincian biayanya jangan cuma omong doang
Tim sukses Acepadang isinya keluarga semua Emang mau jadi dinasti politik	tim sukses acepadang isinya keluarga semua memang mau jadi dinasti politik
November nanti jangan lupa coblos Acep Gita ya Kita semua berharap Jabar bisa menjadi lebih maju dan lebih bahagia Yuk kita sukseskan bersama	november nanti jangan lupa coblos acep gita ya kita semua berharap jabar bisa menjadi lebih maju dan lebih bahagia yuk kita sukseskan bersama

Tabel 4. 22 Hasil Normalization Paslon 2

Sebelum Normalisasi	Setelah Normalisasi
jadi calon wakil gubernur jabar dampingi jeje wiradinata pose tak biasa ronol surapradja di surat suara curi perhatiin	jadi calon wakil gubernur jabar dampingi jeje wiradinata pose tak biasa ronol surapradja di surat suara curi perhatiin
lagilagi janji manis dari jeje tapi realisasi nihil udah capek percaya kenyataan gak sesuai ucapan	lagilagi janji manis dari jeje tapi realisasi nihil sudah capek percaya kenyataan tidak sesuai ucapan
salut buat pak jeje yang selalu turun langsung ke lapangan bukan hanya duduk di kantor pemimpin seperti ini langka	salut buat pak jeje yang selalu turun langsung ke lapangan bukan hanya duduk di kantor pemimpin seperti ini langka
setiap musim hujan daerah saya banjir mana perhatian dari jeje wiradinata gak kelihatan aksi nyatanya	setiap musim hujan daerah saya banjir mana perhatian dari jeje wiradinata tidak kelihatan aksi nyatanya

Tabel 4. 23 Hasil Normalization Paslon 3

Sebelum Normalisasi	Setelah Normalisasi
pria lames cuma wakil aja di dki warga jakarta ga percaya lagi setelah berhianat dibutakan akan kebodohnya sendiri jabar siapa yang endorse kenapa depok kalah dah ga dipercaya suswono ga menjual	pria lames cuma wakil saja di dki warga jakarta tidak percaya lagi setelah berhianat dibutakan akan kebodohnya sendiri jabar siapa yang endorse kenapa depok kalah sudah tidak dipercaya suswono tidak menjual
suherman maa syaa allah panutan sekali ustadz memberikan contoh bahwa kekalahan dalam kontestasi bukan berarti tidak berkontribusi saya yakin allah siapkan posisi yg terbaik pasca pilkada ini	suherman maa syaa allah panutan sekali ustadz memberikan contoh bahwa kekalahan dalam kontestasi bukan berarti tidak berkontribusi saya yakin allah siapkan posisi yang terbaik pasca pilkada ini

sudah lelah tahun jadi oposisi yang digaji sama rakyat aja lelah trus bagaimana rakyat yang suaranya dijual ke rezim dinasti oligarki mulyono besar legowo mantaps ustd syaikh cerminan seorang negarawan sejati menang kalah itulah kompetisi yg terpenting kita lihat dan dukung jalannya roda pemerintahan yg dipimpin oleh pak dedi mendatang selamat bertugas kepada pak dedi tetap semangat ustd syaikh terus tebar kebaikan serta tetap membangun negeri ini

sudah lelah tahun jadi oposisi yang digaji sama rakyat saja lelah terus bagaimana rakyat yang suaranya dijual ke rezim dinasti oligarki mulyono besar legowo mantaps ustd syaikh cerminan seorang negarawan sejati menang kalah itulah kompetisi yang terpenting kita lihat dan dukung jalannya roda pemerintahan yang dipimpin oleh pak dedi mendatang selamat bertugas kepada pak dedi tetap semangat ustd syaikh terus tebar kebaikan serta tetap membangun negeri ini

Tabel 4. 24 Hasil Normalization Paslon 4

Sebelum Normalisasi					Setelah Normalisasi				
sebaiknya	dedi	mulyadi	fokus	soal	sebaiknya	dedi	mulyadi	fokus	soal
mensejahterakan		rakyat	jabar	soal	mensejahterakan		rakyat	jabar	soal
keyakinannya	jadikan	itu	urusan	pribadi dan	keyakinannya	jadikan	itu	urusan	pribadi dan
tidak perlu	diumbang	mau	nyembah	patung kek	tidak perlu	diumbang	mau	nyembah	patung kayak
mau sesajen	kek	umat islam	itu	unik kalau	mau sesajen	kayak	umat islam	itu	unik kalau
agamanya	disenggol	bahkan	yang islam	ktp saj	agamanya	disenggol	bahkan	yang islam	ktp saj
dedi mulyadi	kedatangan	keluarga	besar	bos	dedi mulyadi	kedatangan	keluarga	besar	bos
persib umuh	muchtar	di	subang		persib umuh	muchtar	di	subang	
satusatunya	video	dedy	mulyadi	yang jelek	satusatunya	video	dedy	mulyadi	yang jelek
adalah ketika	dedi mulyadi	membunuh	karakter		adalah ketika	dedi mulyadi	membunuh	karakter	
mahasiswa hmi	yang menanyakan	tupoksi	dedy		mahasiswa hmi	yang menanyakan	tupoksi	dedy	
mulyadi yg	seorang dpr	ri	tapi	malah sibuk	mulyadi yang	seorang dpr	ri	tapi	malah sibuk
bersihin sampah	pasar	selain	iku	apik kabeh	bersihin sampah	pasar	selain	iku	apik kabeh
terlebih lagulagu	emka				terlebih lagulagu	emka			
ga salah ayy	coblos	kang	dedi mulyadi	kemarin	tidak salah ayy	coblos	kang	dedi mulyadi	kemarin

4.4.4 Tokenization

Tahap ini merupakan proses untuk memecahkan teks atau kalimat menjadi token-token atau potongan kata yang terpisah, biasanya berdasarkan spasi. Dengan kata lain, setiap kata dalam kalimat akan dipisahkan menjadi unit kata terpisah yang akan memungkinkan analisis setiap kata secara terpisah. Berikut adalah contoh dari tahapan *tokenization*.

Kode Program 4. 4 Tokenization

```
def tokenize(text):
    tokens = text.split()
    return tokens
df['tokenize'] = df['hasil_normalisasi'].apply(tokenize)
```

Tokenizing adalah proses memecah kalimat menjadi unit-unit kata (token) yang terpisah. Proses ini dilakukan dengan menggunakan library NLTK yang memungkinkan tokenisasi berjalan lebih cepat dan akurat, yang membantu model memahami struktur bahasa secara lebih baik. Proses ini sangat penting dalam analisis teks karena memungkinkan model untuk memahami dan mengolah setiap kata secara terpisah. Tanpa tokenization, teks akan dianggap sebagai satu kesatuan string yang tidak memiliki struktur, yang akan membuat perhitungan frekuensi kata atau analisis maknanya sulit dilakukan. Tokenization juga mempermudah proses ekstraksi fitur, yang sangat penting dalam pelatihan model machine learning. Berikut adalah hasil dari tahapan tokenization.

Tabel 4. 25 Hasil Tokenization Paslon 1

Sebelum Tokenisasi	Setelah Tokenisasi
signal lup lep tidak benar di dusun patinggen desa karangpawitan kec padaherang kab pangandaran jabar tolong perbaiki	['signal', 'lup', 'lep', 'tidak', 'benar', 'di', 'dusun', 'patinggen', 'desa', 'karangpawitan', 'kec', 'padaherang', 'kab', 'pangandaran', 'jabar', 'tolong', 'perbaiki']
acep janjiin listrik gratis tapi tidak ada rincian biayanya jangan cuma omong doang	['acep', 'janjiin', 'listrik', 'gratis', 'tapi', 'tidak', 'ada', 'rincian', 'biayanya', 'jangan', 'cuma', 'omong', 'doang']
tim sukses acepadang isinya keluarga semua memang mau jadi dinasti politik	['tim', 'sukses', 'acepadang', 'isinya', 'keluarga', 'semua', 'memang', 'mau', 'jadi', 'dinasti', 'politik']
november nanti jangan lupa coblos acep gita ya kita semua berharap jabar bisa menjadi lebih maju dan lebih bahagia yuk kita sukseskan bersama	['november', 'nanti', 'jangan', 'lupa', 'coblos', 'acep', 'gita', 'ya', 'kita', 'semua', 'berharap', 'jabar', 'bisa', 'menjadi', 'lebih', 'maju', 'dan', 'lebih', 'bahagia', 'yuk', 'kita', 'sukseskan', 'bersama']

Tabel 4. 26 Hasil Tokenization Paslon 2

Sebelum Tokenisasi	Setelah Tokenisasi
jadi calon wakil gubernur jabar dampingi jeje wiradinata pose tak biasa ronal surapradja di surat suara curi perhatiin	['jadi', 'calon', 'wakil', 'gubernur', 'jabar', 'dampingi', 'jeje', 'wiradinata', 'pose', 'tak', 'biasa', 'ronal', 'surapradja', 'di', 'surat', 'suara', 'curi', 'perhatian']
lagilagi janji manis dari jeje tapi realisasi nihil sudah capek percaya kenyataan tidak sesuai ucapan	['lagilagi', 'janji', 'manis', 'dari', 'jeje', 'tapi', 'realisasi', 'nihil', 'sudah', 'capek', 'percaya', 'kenyataan', 'tidak', 'sesuai', 'ucapan']

salut buat pak jeje yang selalu turun langsung ke lapangan bukan hanya duduk di kantor pemimpin seperti ini langka	['salut', 'buat', 'pak', 'jeje', 'yang', 'selalu', 'turun', 'langsung', 'ke', 'lapangan', 'bukan', 'hanya', 'duduk', 'di', 'kantor', 'pemimpin', 'seperti', 'ini', 'langka']
setiap musim hujan daerah saya banjir mana perhatian dari jeje wiradinata tidak kelihatan aksi nyatanya	['setiap', 'musim', 'hujan', 'daerah', 'saya', 'banjir', 'mana', 'perhatian', 'dari', 'jeje', 'wiradinata', 'tidak', 'kelihatan', 'aksi', 'nyatanya']

Tabel 4. 27 Hasil Tokenization Paslon 3

Sebelum Tokenisa	Setelah Tokenisasi
pria lames cuma wakil saja di dki warga jakarta tidak percaya lagi setelah berhianat dibutakan akan kebodohnya sendiri jabar siapa yang endorse kenapa depok kalah sudah tidak dipercaya suswono tidak menjual	['pria', 'lames', 'cuma', 'wakil', 'saja', 'di', 'dki', 'warga', 'jakarta', 'tidak', 'percaya', 'lagi', 'setelah', 'berhianat', 'dibutakan', 'akan', 'kebodohnya', 'sendiri', 'jabar', 'siapa', 'yang', 'endorse', 'kenapa', 'depok', 'kalah', 'sudah', 'tidak', 'dipercaya', 'suswono', 'tidak', 'menjual']
suherman maa syaa allah panutan sekali ustadz memberikan contoh bahwa kekalahan dalam kontestasi bukan berarti tidak berkontribusi saya yakin allah siapkan posisi yang terbaik pasca pilkada ini	['suherman', 'maa', 'syaa', 'allah', 'panutan', 'sekali', 'ustadz', 'memberikan', 'contoh', 'bahwa', 'kekalahan', 'dalam', 'kontestasi', 'bukan', 'berarti', 'tidak', 'berkontribusi', 'saya', 'yakin', 'allah', 'siapkan', 'posisi', 'yang', 'terbaik', 'pasca', 'pilkada', 'ini']
sudah lelah tahun jadi oposisi yang digaji sama rakyat saja lelah terus bagaimana rakyat yang suaranya dijual ke rezim dinasti oligarki mulyono	['sudah', 'lelah', 'tahun', 'jadi', 'oposisi', 'yang', 'digaji', 'sama', 'rakyat', 'saja', 'lelah', 'terus', 'bagaimana', 'rakyat', 'yang', 'suaranya', 'dijual', 'ke', 'rezim', 'dinasti', 'oligarki', 'mulyono']
besar legowo mantaps ustd syaikh cerminan seorang negarawan sejati menang kalah itulah kompetisi yang terpenting kita lihat dan dukung jalannya roda pemerintahan yang dipimpin oleh pak dedi mendatang selamat bertugas kepada pak dedi tetap semangat ustd syaikh terus tebar kebaikan serta tetap membangun negeri ini	['besar', 'legowo', 'mantaps', 'ustd', 'syaikh', 'cerminan', 'seorang', 'negarawan', 'sejati', 'menang', 'kalah', 'itulah', 'kompetisi', 'yang', 'terpenting', 'kita', 'lihat', 'dan', 'dukung', 'jalannya', 'roda', 'pemerintahan', 'yang', 'dipimpin', 'oleh', 'pak', 'dedi', 'mendatang', 'selamat', 'bertugas', 'kepada', 'pak', 'dedi', 'tetap', 'semangat', 'ustd', 'syaikh', 'terus', 'tebar', 'kebaikan', 'serta', 'tetap', 'membangun', 'negeri', 'ini']

Tabel 4. 28 Hasil Tokenization Paslon 4

Sebelum Tokenisasi	Setelah Tokenisasi
sebaiknya dedi mulyadi fokus soal mensejahterakan rakyat jabar soal keyakinannya jadikan itu urusan pribadi dan tidak perlu diumbar mau nyembah patung kayak mau sesajen kayak umat islam itu unik kalau agamanya disenggol bahkan yang islam ktp saj dedi mulyadi kedatangan keluarga besar bos persib umuh muchtar di subang	['sebaiknya', 'dedi', 'mulyadi', 'fokus', 'soal', 'mensejahterakan', 'rakyat', 'jabar', 'soal', 'keyakinannya', 'jadikan', 'itu', 'urusan', 'pribadi', 'dan', 'tidak', 'perlu', 'diumbar', 'mau', 'nyembah', 'patung', 'kayak', 'mau', 'sesajen', 'kayak', 'umat', 'islam', 'itu', 'unik', 'kalau', 'agamanya', 'disenggol', 'bahkan', 'yang', 'islam', 'ktp', 'saj'] ['dedi', 'mulyadi', 'kedatangan', 'keluarga', 'besar', 'bos', 'persib', 'umuh', 'muchtar', 'di', 'subang']
satusatunya video dedy mulyadi yang jelek adalah ketika dedi mulyadi membunuh karakter mahasiswa hmi yang menanyakan tupoksi dedy mulyadi yang seorang dpr ri tapi malah sibuk bersihin sampah pasar selain iku apik kabeh terlebih lagulagu emka	['satusatunya', 'video', 'dedy', 'mulyadi', 'yang', 'jelek', 'adalah', 'ketika', 'dedi', 'mulyadi', 'membunuh', 'karakter', 'mahasiswa', 'hmi', 'yang', 'menanyakan', 'tupoksi', 'dedy', 'mulyadi', 'yang', 'seorang', 'dpr', 'ri', 'tapi', 'malah', 'sibuk', 'bersihin', 'sampah', 'pasar', 'selain', 'iku', 'apik', 'kabeh', 'terlebih', 'lagulagu', 'emka']
tidak salah ayy coblos kang dedi mulyadi kemarin	['tidak', 'salah', 'ayy', 'coblos', 'kang', 'dedi', 'mulyadi', 'kemarin']

4.4.5 Stopword

Pada tahap ini semua kata yang termasuk kedalam stopwords akan terhapus, penghapusan ini dilakukan karena kata-kata tersebut dianggap tidak memiliki makna penting dalam kategori tersebut tidak digunakan. Berikut adalah contoh dari tahapan *stopword*.

Kode Program 4. 5 Stopword

```
def remove_stopwords(text):
    return [word for word in text if word not in stop_words]
df['stopword removal'] = df['tokenize'].apply(lambda x:
remove_stopwords(x))
```

Tabel 4. 29 Hasil Stopword Paslon 1

Sebelum Stopword	Setelah Stopword
['signal', 'lup', 'lep', 'tidak', 'benar', 'di', 'dusun', 'patinggen', 'desa', 'karangpawitan', 'kec', 'padaherang', 'kab', 'pangandaran', 'jabar', 'tolong', 'perbaiki']	['signal', 'lup', 'lep', 'dusun', 'patinggen', 'desa', 'karangpawitan', 'kec', 'padaherang', 'kab', 'pangandaran', 'jabar', 'tolong', 'perbaiki']
['acep', 'janjiin', 'listrik', 'gratis', 'tapi', 'tidak', 'ada', 'rincian', 'biayanya', 'jangan', 'cuma', 'omong', 'doang']	['acep', 'janjiin', 'listrik', 'gratis', 'rincian', 'biayanya', 'omong', 'doang']

['tim', 'sukses', 'acepadang', 'isinya', 'keluarga', 'semua', 'memang', 'mau', 'jadi', 'dinasti', 'politik']	['tim', 'sukses', 'acepadang', 'isinya', 'keluarga', 'dinasti', 'politik']
['november', 'nanti', 'jangan', 'lupa', 'coblos', 'acep', 'gita', 'ya', 'kita', 'semua', 'berharap', 'jabar', 'bisa', 'menjadi', 'lebih', 'maju', 'dan', 'lebih', 'bahagia', 'yuk', 'kita', 'sukseskan', 'bersama']	['november', 'lupa', 'coblos', 'acep', 'gita', 'ya', 'berharap', 'jabar', 'maju', 'bahagia', 'yuk', 'sukseskan']

Tabel 4. 30 Hasil Stopword Paslon 2

Sebelum Stopword	Setelah Stopword
['jadi', 'calon', 'wakil', 'gubernur', 'jabar', 'dampingi', 'jeje', 'wiradinata', 'pose', 'tak', 'biasa', 'ronal', 'surapradja', 'di', 'surat', 'suara', 'curi', 'perhatian']	['calon', 'wakil', 'gubernur', 'jabar', 'dampingi', 'jeje', 'wiradinata', 'pose', 'ronal', 'surapradja', 'surat', 'suara', 'curi', 'perhatian']
['lagilagi', 'janji', 'manis', 'dari', 'jeje', 'tapi', 'realisasi', 'nihil', 'sudah', 'capek', 'percaya', 'kenyataan', 'tidak', 'sesuai', 'ucapan']	['lagilagi', 'janji', 'manis', 'jeje', 'realisasi', 'nihil', 'capek', 'percaya', 'kenyataan', 'sesuai', 'ucapan']
['salut', 'buat', 'pak', 'jeje', 'yang', 'selalu', 'turun', 'langsung', 'ke', 'lapangan', 'bukan', 'hanya', 'duduk', 'di', 'kantor', 'pemimpin', 'seperti', 'ini', 'langka']	['salut', 'jeje', 'turun', 'langsung', 'lapangan', 'duduk', 'kantor', 'pemimpin', 'langka']
['setiap', 'musim', 'hujan', 'daerah', 'saya', 'banjir', 'mana', 'perhatian', 'dari', 'jeje', 'wiradinata', 'tidak', 'kelihatan', 'aksi', 'nyatanya']	['musim', 'hujan', 'daerah', 'banjir', 'perhatian', 'jeje', 'wiradinata', 'aksi']

Tabel 4. 31 Hasil Stopword Paslon 3

Sebelum Stopword	Setelah Stopword
['pria', 'lames', 'cuma', 'wakil', 'saja', 'di', 'dki', 'warga', 'jakarta', 'tidak', 'percaya', 'lagi', 'setelah', 'berhianat', 'dibutakan', 'akan', 'kebodohnya', 'sendiri', 'jabar', 'siapa', 'yang', 'endorse', 'kenapa', 'depok', 'kalah', 'sudah', 'tidak', 'dipercaya', 'suswono', 'tidak', 'menjual']	['pria', 'lames', 'wakil', 'dki', 'warga', 'jakarta', 'percaya', 'berhianat', 'dibutakan', 'kebodohnya', 'jabar', 'endorse', 'depok', 'kalah', 'dipercaya', 'suswono', 'menjual']
['suherman', 'maa', 'syaa', 'allah', 'panutan', 'sekali', 'ustadz', 'memberikan', 'contoh', 'bahwa', 'kekalahan', 'dalam', 'kontestasi', 'bukan', 'berarti', 'tidak', 'berkontribusi', 'saya', 'yakini', 'allah', 'siapkan', 'posisi', 'yang', 'terbaik', 'pasca', 'pilkada', 'ini']	['suherman', 'maa', 'syaa', 'allah', 'panutan', 'ustadz', 'contoh', 'kekalahan', 'kontestasi', 'berkontribusi', 'allah', 'siapkan', 'posisi', 'terbaik', 'pasca', 'pilkada']
['sudah', 'lelah', 'tahun', 'jadi', 'oposisi', 'yang', 'digaji', 'sama', 'rakyat', 'saja', 'lelah', 'terus', 'bagaimana', 'rakyat', 'yang', 'suaranya', 'dijual', 'ke', 'rezim', 'dinasti', 'oligarki', 'mulyono']	['lelah', 'oposisi', 'digaji', 'rakyat', 'lelah', 'rakyat', 'suaranya', 'dijual', 'rezim', 'dinasti', 'oligarki', 'mulyono']

['besar', 'legowo', 'mantaps', 'ustd', 'syaikhu', 'cerminan', 'seorang', 'negarawan', 'sejati', 'menang', 'kalah', 'itulah', 'kompetisi', 'yang', 'terpenting', 'kita', 'lihat', 'dan', 'dukung', 'jalannya', 'roda', 'pemerintahan', 'yang', 'dipimpin', 'oleh', 'pak', 'dedi', 'selamat', 'bertugas', 'kepada', 'ustd', 'syaikhu', 'tebar', 'kebaikan', 'membangun', 'pak', 'dedi', 'tetap', 'semangat', 'ustd', 'syaikhu', 'terus', 'tebar', 'kebaikan', 'serta', 'tetap', 'membangun', 'negeri', 'ini']	['legowo', 'mantaps', 'ustd', 'syaikhu', 'cerminan', 'negarawan', 'sejati', 'menang', 'kalah', 'kompetisi', 'terpenting', 'lihat', 'dukung', 'jalannya', 'roda', 'pemerintahan', 'dipimpin', 'dedi', 'selamat', 'bertugas', 'dedi', 'semangat', 'ustd', 'syaikhu', 'tebar', 'kebaikan', 'membangun', 'negeri']
--	--

Tabel 4. 32 Hasil Stopword Paslon 4

Sebelum Stopword	Setelah Stopword
['sebaiknya', 'dedi', 'mulyadi', 'fokus', 'soal', 'mensejahterakan', 'rakyat', 'jabar', 'soal', 'keyakinannya', 'jadikan', 'itu', 'urusan', 'pribadi', 'dan', 'tidak', 'perlu', 'diumbang', 'mau', 'nyembah', 'patung', 'kayak', 'mau', 'sesajen', 'kayak', 'umat', 'islam', 'itu', 'unik', 'kalau', 'agamanya', 'disenggol', 'bahkan', 'yang', 'islam', 'ktp', 'saj']	['dedi', 'mulyadi', 'fokus', 'mensejahterakan', 'rakyat', 'jabar', 'keyakinannya', 'jadikan', 'urusan', 'pribadi', 'diumbang', 'nyembah', 'patung', 'kayak', 'sesajen', 'kayak', 'umat', 'islam', 'unik', 'agamanya', 'disenggol', 'islam', 'ktp', 'saj']
['dedi', 'mulyadi', 'kedatangan', 'keluarga', 'besar', 'bos', 'persib', 'umuh', 'muchtar', 'di', 'subang']	['dedi', 'mulyadi', 'kedatangan', 'keluarga', 'bos', 'persib', 'umuh', 'muchtar', 'subang']
['satunya', 'video', 'dedi', 'mulyadi', 'yang', 'jelek', 'adalah', 'ketika', 'dedi', 'mulyadi', 'membunuh', 'karakter', 'mahasiswa', 'hmi', 'yang', 'menanyakan', 'tupoksi', 'dedi', 'mulyadi', 'yang', 'seorang', 'dpr', 'ri', 'tapi', 'malah', 'sibuk', 'bersihin', 'sampah', 'pasar', 'selain', 'iku', 'apik', 'kabeh', 'terlebih', 'lagulagu', 'emka']	['satunya', 'video', 'dedi', 'mulyadi', 'jelek', 'dedi', 'mulyadi', 'membunuh', 'karakter', 'mahasiswa', 'hmi', 'tupoksi', 'dedi', 'mulyadi', 'dpr', 'ri', 'sibuk', 'bersihin', 'sampah', 'pasar', 'iku', 'apik', 'kabeh', 'lagulagu', 'emka']
['tidak', 'salah', 'ayy', 'coblos', 'kang', 'dedi', 'mulyadi', 'kemarin']	['salah', 'ayy', 'coblos', 'kang', 'dedi', 'mulyadi', 'kemarin']

4.4.6 Stemming

Tahap ini dilakukan setelah tahapan stopwords, tahap ini merupakan tahapan untuk mengubah kata berimbuhan menjadi bentuk dasar. Sebagai contohnya, kata berlari akan diubah menjadi kata lari. Berikut adalah contoh dari tahapan *stemming* Proses stemming dilakukan dengan menggunakan kode program Langkah utama yang dilakukan adalah mengimpor library Sastrawi dan StemmerFactory untuk membentuk objek stemmer. Setelah itu, fungsi stemming diterapkan pada kolom data yang telah melewati tahap penghapusan stopwords. Hasil dari proses stemming adalah sebagai berikut.

Kode Program 4. 6 Stemming

```
factory = StemmerFactory()
stemmer = factory.create_stemmer()
def stem_text(text):
    return [stemmer.stem(word) for word in text]
df['stemming_data'] = df['stopword removal'].apply(lambda x: stem_text(x))
```

Tabel 4. 33 Hasil Stemming Paslon 1

Sebelum Stemming	Setelah Stemming
['signal', 'lup', 'lep', 'dusun', 'patinggen', 'desa', 'karangpawitan', 'kec', 'padaherang', 'kab', 'pangandaran', 'jabar', 'tolong', 'perbaiki']	['signal', 'lup', 'lep', 'dusun', 'patinggen', 'desa', 'karangpawitan', 'kec', 'padaherang', 'kab', 'pangandaran', 'jabar', 'tolong', 'baik']
['acep', 'janjiin', 'listrik', 'gratis', 'rincian', 'biayanya', 'omong', 'doang']	['acep', 'janjiin', 'listrik', 'gratis', 'rincian', 'biaya', 'omong', 'doang']
['tim', 'sukses', 'acepadang', 'isinya', 'keluarga', 'dinasti', 'politik']	['tim', 'sukses', 'acepadang', 'isi', 'keluarga', 'dinasti', 'politik']
['november', 'lupa', 'coblos', 'acep', 'gita', 'ya', 'berharap', 'jabar', 'maju', 'bahagia', 'yuk', 'sukseskan']	['november', 'lupa', 'coblos', 'acep', 'gita', 'ya', 'harap', 'jabar', 'maju', 'bahagia', 'yuk', 'sukses']

Tabel 4. 34 Hasil Stemming Paslon 2

Sebelum Stemming	Setelah Stemming
['calon', 'wakil', 'gubernur', 'jabar', 'dampin', 'jeje', 'wiradinata', 'pose', 'ronal', 'surapradja', 'surat', 'suara', 'curi', 'perhatian']	['calon', 'wakil', 'gubernur', 'jabar', 'dampin', 'jeje', 'wiradinata', 'pose', 'ronal', 'surapradja', 'surat', 'suara', 'curi', 'perhati']
['lagilagi', 'janji', 'manis', 'jeje', 'realisasi', 'nihil', 'capek', 'percaya', 'kenyataan', 'sesuai', 'ucapan']	['lagilagi', 'janji', 'manis', 'jeje', 'realisasi', 'nihil', 'capek', 'percaya', 'nyata', 'sesuai', 'ucap']
['salut', 'jeje', 'turun', 'langsung', 'lapangan', 'duduk', 'kantor', 'pemimpin', 'langka']	['salut', 'jeje', 'turun', 'langsung', 'lapang', 'duduk', 'kantor', 'pimpin', 'langka']
['musim', 'hujan', 'daerah', 'banjir', 'perhatian', 'jeje', 'wiradinata', 'aksi']	['musim', 'hujan', 'daerah', 'banjir', 'perhati', 'jeje', 'wiradinata', 'aksi']

Tabel 4. 35 Hasil Stemming Paslon 3

Sebelum Stemming	Setelah Stemming
['pria', 'lames', 'wakil', 'dki', 'warga', 'jakarta', 'percaya', 'berhianat', 'dibutakan', 'kebodohnya', 'jabar', 'endorse', 'depok', 'kalah', 'dipercaya', 'suswono', 'menjual']	['pria', 'lames', 'wakil', 'dki', 'warga', 'jakarta', 'percaya', 'berhianat', 'buta', 'bodoh', 'jabar', 'endorse', 'depok', 'kalah', 'percaya', 'suswono', 'jual']

['suherman', 'maa', 'syaa', 'allah', 'panutan', 'ustadz', 'contoh', 'kekalahan', 'kontestasi', 'berkontribusi', 'allah', 'siapkan', 'posisi', 'terbaik', 'pasca', 'pilkada']	['suherman', 'maa', 'syaa', 'allah', 'panutan', 'ustadz', 'contoh', 'kalah', 'kontestasi', 'kontribusi', 'allah', 'siap', 'posisi', 'baik', 'pasca', 'pilkada']
['lelah', 'oposisi', 'digaji', 'rakyat', 'lelah', 'rakyat', 'suaranya', 'dijual', 'rezim', 'dinasti', 'oligarki', 'mulyono']	['lelah', 'oposisi', 'gaji', 'rakyat', 'lelah', 'rakyat', 'suara', 'jual', 'rezim', 'dinasti', 'oligarki', 'mulyono']
['legowo', 'mantaps', 'ustd', 'syaikhu', 'cerminan', 'negarawan', 'sejati', 'menang', 'kalah', 'kompetisi', 'terpenting', 'lihat', 'dukung', 'jalannya', 'roda', 'pemerintahan', 'dipimpin', 'dedi', 'selamat', 'bertugas', 'dedi', 'semangat', 'ustd', 'syaikhu', 'tebar', 'kebaikan', 'membangun', 'negeri']	['legowo', 'mantaps', 'ustd', 'syaikhu', 'cermin', 'negarawan', 'sejati', 'menang', 'kalah', 'kompetisi', 'penting', 'lihat', 'dukung', 'jalan', 'roda', 'perintah', 'pimpin', 'dedi', 'selamat', 'tugas', 'dedi', 'semangat', 'ustd', 'syaikhu', 'tebar', 'baik', 'bangun', 'negeri']

Tabel 4. 36 Hasil Stemming Paslon 4

Sebelum Stemming	Setelah Stemming
['dedi', 'mulyadi', 'fokus', 'mensejahterakan', 'rakyat', 'jabar', 'keyakinannya', 'jadikan', 'urusan', 'pribadi', 'diumbang', 'nyembah', 'patung', 'kayak', 'sesajen', 'kayak', 'umat', 'islam', 'unik', 'agamanya', 'disenggol', 'islam', 'ktp', 'saj']	['dedi', 'mulyadi', 'fokus', 'sejahtera', 'rakyat', 'jabar', 'yakin', 'jadi', 'urus', 'pribadi', 'umbar', 'nyembah', 'patung', 'kayak', 'sesajen', 'kayak', 'umat', 'islam', 'unik', 'agama', 'senggol', 'islam', 'ktp', 'saj']
['dedi', 'mulyadi', 'kedatangan', 'keluarga', 'bos', 'persib', 'umuh', 'muchtar', 'subang']	['dedi', 'mulyadi', 'datang', 'keluarga', 'bos', 'persib', 'umuh', 'muchtar', 'subang']
['satunya', 'video', 'dedy', 'mulyadi', 'jelek', 'dedi', 'mulyadi', 'membunuh', 'karakter', 'mahasiswa', 'hmi', 'tupoksi', 'dedy', 'mulyadi', 'dpr', 'ri', 'sibuk', 'bersihin', 'sampah', 'pasar', 'iku', 'apik', 'kابه', 'lagulagu', 'emka']	['satunya', 'video', 'dedy', 'mulyadi', 'jelek', 'dedi', 'mulyadi', 'bunuh', 'karakter', 'mahasiswa', 'hmi', 'tupoksi', 'dedy', 'mulyadi', 'dpr', 'ri', 'sibuk', 'bersihin', 'sampah', 'pasar', 'iku', 'apik', 'kابه', 'lagulagu', 'emka']
['salah', 'ayy', 'coblos', 'kang', 'dedi', 'mulyadi', 'kemarin']	['salah', 'ayy', 'coblos', 'kang', 'dedi', 'mulyadi', 'kemarin']

4.5 Split Data

Dalam analisis sentimen menggunakan metode Support Vector Machine, Data dibagi menjadi dua bagian yaitu data Data Training dan Data Testing dengan perbandingan 80:20 atau 80% untuk data Training dan 20% untuk data Testing. Adapun data Training digunakan untuk melatih model Support Vector Machine, di mana model ini belajar antara fitur teks label dan sentimen. Sementara itu, data testing digunakan untuk menguji kinerja model dengan membandingkan prediksi model dengan label sebenarnya. Pembagian data 80:20 ini memastikan keseimbangan antara melatih model dengan cukup data dan menguji model secara

objektif.

4.6 Pembobotan Kata (TF-IDF)

Setelah data teks dilakukan tahap pre-processing dan split data. Lalu tahap selanjutnya adalah pembobotan kata, jadi untuk setiap data diberi bobot masing-masing dengan metode TF-IDF. Pada pembobotan kata ini ada dua perhitungan yaitu perhitungan TF (term frequency) dan TF-IDF. Metode term frequency merupakan metode yang sering digunakan dengan cara menghitung bobot tiap kata dengan menghitung banyaknya kemunculan kata dalam suatu dokumen. Tahapan pembobotan diawali dengan perhitungan nilai TF (Term frequency), sebagai contoh berikut beberapa dataset uji coba yang berisi dari kata-kata yang telah dilakukan pada tahapan stemming. Terdapat ada sebagian dokumen yang akan dilakukan perhitungan, berikut contoh dokumen yang telah dilakukan stemming.

4.6.1 Perhitungan TF (Term Frequency)

Langkah pertama dalam proses pembobotan kata adalah menentukan nilai frekuensi (TF). Tujuan perhitungan ini adalah untuk mengetahui seberapa sering suatu kata muncul dalam dokumen tertentu, sehingga dapat membantu menemukan kata kunci yang dianggap penting dalam data. Berikut adalah contoh penggunaan perhitungan Term Frequency menggunakan data sampel.

Tabel 4. 37 Data Sample Perhitungan TF-IDF Paslon 1

No	Dokumen	Teks	Label
1	D91	transparan publikasi lapor donasi kampanye	Negatif
2	D92	acep acara tv sindir masyarakat miskin usaha peka banget	Negatif
3	D93	november pasti coblos acep gita ya ayo juang jabar	Positif
4	D94	tanggal november yuk coblos acep gita wajib satu jabar sejahtera	Positif

Setelah melakukan proses pre-processing data selesai dan setiap kata telah ditemukan dalam setiap dokumen. Langkah berikutnya adalah menghitung bobot pada dokumen menggunakan metode TF-IDF, yang merupakan singkatan dari Term Frequency-Inverse Document Frequency. Tujuan perhitungan ini adalah untuk menentukan seberapa penting satu kata dalam dokumen dibandingkan dengan keseluruhan dokumen. Nilai bobot pada TF-IDF ini didasarkan pada representasi numerik dari frekuensi kata yang ditemukan pada setiap dokumen, yang dapat dilihat dibawah ini.

Tabel 4. 38 Hasil Frekuensi Kemunculan Setiap Kata Dalam Dokumen

No	Term	D1	D2	D3	D4
1	Acara	0	1	0	0

2	Acep	0	1	1	1
3	Ayo	0	0	1	0
4	Banget	0	1	0	0
5	Coblos	0	0	1	1
6	Donasi	1	0	0	0
7	Gita	0	0	1	1
8	Jabar	0	0	1	1
9	Juang	0	0	1	0
10	Kampanye	1	0	0	0
11	Lapor	1	0	0	0
12	Masyarakat	0	1	0	0
13	Miskin	0	1	0	0
14	November	0	0	1	1
15	Pasti	0	0	1	0
16	Peka	0	1	0	0
17	Publikasi	1	0	0	0
18	Satu	0	0	0	1
19	Sejahtera	0	0	0	1
20	Sindir	0	1	0	0
21	Tanggal	0	0	0	1
22	Transparan	1	0	0	0
23	Tv	0	1	0	0
24	Usaha	0	1	0	0
25	Wajib	0	0	0	1
26	Ya	0	0	1	0
27	Yuk	0	0	0	1

Nilai Frekuensi Terma untuk setiap kata dalam masing-masing dokumen dihitung dengan menggunakan Rumus Frekuensi Terma yang tercantum pada Perumusan di bawah ini. Hasil perhitungan jumlah kemunculan kata ditampilkan pada di Table 4.38.

$$TF_{(t,d)} = \frac{\text{Jumlah kemunculan kata dalam dokumen}}{\text{Total kata dalam dokumen}}$$

Pada table 4.35 menunjukkan hasil perhitungan pada Term Frequency (TF) untuk setiap kata yang muncul dalam empat dokumen yang dianalisis. Tidak semuaa kata ditemukan di setiap dokumen, sehingga dapat memberikan suatu gambaran mengenai kata-kata yang paling sering digunakan dan memiliki kontribusi penting terhadap dokumen. Selain itu, Panjang dokumen bisa mempengaruhi nilai pada Term frequency (TF). Karena semakin Panjang sebuah dokumen maka semakin kecil proporsi frekuensi sebuah kata jika tidak muncul secara berulang. Misalnya, karena

kata “Acep” sering disebutkan dalam hampir semua dokumen, tetapi pada nilai TF bisa berbeda tergantung pada total jumlah kata dalam setiap masing-masing dokumen. Begitupun juga dengan kata seperti “Donasi” yang hanya muncul pada dokumen tertentu dan tidak muncul pada yang dokumen yang lain, dan menghasilkan nilai TF nol pada sebagian dokumen.

Tabel 4. 39 Hasil Perhitungan TF

No	Term	D1	D2	D3	D4
1	Acara	0	0,111111	0	0
2	Acep	0	0,111111	0,111111	0,1
3	Ayo	0	0	0,111111	0
4	Banget	0	0,111111	0	0
5	Coblos	0	0	0,111111	0,1
6	Donasi	0,2	0	0	0
7	Gita	0	0	0,111111	0,1
8	Jabar	0	0	0,111111	0,1
9	Juang	0	0	0,111111	0
10	Kampanye	0,2	0	0	0
11	Lapor	0,2	0	0	0
12	Masyarakat	0	0,111111	0	0
13	Miskin	0	0,111111	0	0
14	November	0	0	0,111111	0,1
15	Pasti	0	0	0,111111	0
16	Peka	0	0,111111	0	0
17	Publikasi	0,2	0	0	0
18	Satu	0	0	0	0,1
19	Sejahtera	0	0	0	0,1
20	Sindir	0	0,111111	0	0
21	Tanggal	0	0	0	0,1
22	Transparan	0,2	0	0	0
23	Tv	0	0,111111	0	0
24	Usaha	0	0,111111	0	0
25	Wajib	0	0	0	0,1
26	Ya	0	0	0,111111	0
27	Yuk	0	0	0	0,1

4.6.2 Perhitungan IDF (Inverse Document Frequency)

Adalah digunakan untuk mengukur tingkat penyebaran kata di seluruh kumpulan dokumen. Semakin banyak dokumen yang mengandung kata tersebut, maka semakin kecil kontribusi terhadap bobot TF-IDF karena dianggap kurang spesifik. Sebab, DF membantu dalam menemukan kata-kata yang paling sering digunakan di seluruh dataset.

Tabel 4. 40 Hasil Perhitungan DF

No	Term	D1	D2	D3	D4	DF
1	acara	0	1	0	0	1
2	acep	0	1	1	1	3
3	ayo	0	0	1	0	1
4	banget	0	1	0	0	1
5	coblos	0	0	1	1	2
6	donasi	1	0	0	0	1
7	gita	0	0	1	1	2
8	jabar	0	0	1	1	2
9	juang	0	0	1	0	1
10	kampanye	1	0	0	0	1
11	lapor	1	0	0	0	1
12	masyarakat	0	1	0	0	1
13	miskin	0	1	0	0	1
14	november	0	0	1	1	2
15	pasti	0	0	1	0	1
16	peka	0	1	0	0	1
17	publikasi	1	0	0	0	1
18	satu	0	0	0	1	1
19	sejahtera	0	0	0	1	1
20	sindir	0	1	0	0	1
21	tanggal	0	0	0	1	1
22	transparan	1	0	0	0	1
23	tv	0	1	0	0	1
24	usaha	0	1	0	0	1
25	wajib	0	0	0	1	1
26	ya	0	0	1	0	1
27	yuk	0	0	0	1	1

Setelah nilai *document frequency* diketahui bisa dilanjutkan pada tahapan selanjutnya yaitu tahap IDF atau *Invers Document Frequency*. Tahapan IDF digunakan untuk menghitung seberapa jarang suatu kata muncul dalam seluruh dokumen. Semakin jarang suatu kata muncul dalam berbagai dokumen, maka

semakin tinggi nilai IDF. Adapun untuk menghitung nilai IDF menggunakan rumus seperti dibawah ini.

$$IDF_{(t)} = \log \frac{N}{DF_{(t)}}$$

Keterangan :

$IDF_{(t)}$: Nilai Inverse Document Frequency untuk term ke t

N : Jumlah total dokumen dalam koleksi

$DF_{(t)}$: Jumlah dokumen yang mengandung term ke-t

log : Fungsi logaritma (basis ke 10 atau natural log)

Rumus di atas merupakan rumus untuk mencari nilai IDF, dimana rumusnya menggunakan log dari jumlah dokumen dibagi dengan DF_t atau nilai DF dari setiap dokumen yang ada. Berikut adalah hasil dari perhitungan *Invers Document Frequency* Pada table di bawah ini.

Tabel 4. 41 Hasil Perhitungan IDF

No	Term	D1	D2	D3	D4	DF	IDF
1	Acara	0	1	0	0	1	1,386294
2	Acep	0	1	1	1	3	0,287682
3	Ayo	0	0	1	0	1	1,386294
4	Banget	0	1	0	0	1	1,386294
5	Coblos	0	0	1	1	2	0,693147
6	Donasi	1	0	0	0	1	1,386294
7	Gita	0	0	1	1	2	0,693147
8	Jabar	0	0	1	1	2	0,693147
9	Juang	0	0	1	0	1	1,386294
10	Kampanye	1	0	0	0	1	1,386294
11	Lapor	1	0	0	0	1	1,386294
12	Masyarakat	0	1	0	0	1	1,386294
13	Miskin	0	1	0	0	1	1,386294
14	November	0	0	1	1	2	0,693147
15	Pasti	0	0	1	0	1	1,386294
16	Peka	0	1	0	0	1	1,386294
17	Publikasi	1	0	0	0	1	1,386294
18	Satu	0	0	0	1	1	1,386294
19	Sejahtera	0	0	0	1	1	1,386294
20	Sindir	0	1	0	0	1	1,386294
21	Tanggal	0	0	0	1	1	1,386294
22	Transparan	1	0	0	0	1	1,386294
23	Tv	0	1	0	0	1	1,386294
24	Usaha	0	1	0	0	1	1,386294
25	Wajib	0	0	0	1	1	1,386294

26	Ya	0	0	1	0	1	1,386294
27	Yuk	0	0	0	1	1	1,386294

4.6.3 Perhitungan TF-IDF (Term Frequency-Inverse Document Frequency)

Perhitungan TF-IDF dapat digunakan untuk menghitung bobot suatu term yang ada pada dokumen dengan mengalikan hasil dari nilai TF dan IDF. Berikut rumus pembobotannya

$$TFIDF_{(t,d)} = TF \times IDF$$

Tabel 4. 42 Hasil Perhitungan TF-IDF

No	Term	D1	D2	D3	D4
1	Acara	0	0,154032	0	0
2	Acep	0	0,031964	0,031964	0,028768
3	Ayo	0	0	0,154032	0
4	Banget	0	0,154032	0	0
5	Coblos	0	0	0,077016	0,069314
6	Donasi	0,277258	0	0	0
7	Gita	0	0	0,077016	0,069314
8	Jabar	0	0	0,077016	0,069314
9	Juang	0	0	0,154032	0
10	Kampanye	0,277258	0	0	0
11	Lapor	0,277258	0	0	0
12	Masyarakat	0	0,154032	0	0
13	Miskin	0	0,154032	0	0
14	November	0	0	0,077016	0,069314
15	Pasti	0	0	0,154032	0
16	Peka	0	0,154032	0	0
17	Publikasi	0,277258	0	0	0
18	Satu	0	0	0	0,138629
19	Sejahtera	0	0	0	0,138629
20	Sindir	0	0,154032	0	0
21	Tanggal	0	0	0	0,138629
22	Transparan	0,277258	0	0	0
23	Tv	0	0,154032	0	0
24	Usaha	0	0,154032	0	0
25	Wajib	0	0	0	0,138629
26	Ya	0	0	0,154032	0
27	Yuk	0	0	0	0,138629

Tabel 4. 43 Data Sample Perhitungan TF-IDF paslon 2

No	Dokumen	Teks	Label
1	D31	omong pariwisata calon gubernur nomor urut jeje wiradinata bilang	Negatif

		anggur pangandaran tuh kecil sejawa barat iya full detikjabar detikpilkada pilgubjabar	
2	D32	daerah taman kota rapi ramah anak jeje wiradinata peduli ruang publik	Positif
3	D33	jeje sibuk citra urus jalan rusak bertahun-tahun keluhan warga	Negatif
4	D34	salut jeje hadir bencana bantu cepat tanggap	Positif

Tabel 4. 44 Hasil Frekuensi Kemunculan Setiap Kata Dalam Dokumen

No	Term	D1	D2	D3	D4
1	anak	0	1	0	0
2	anggur	1	0	0	0
3	bantu	0	0	0	1
4	barat	1	0	0	0
5	bencana	0	0	0	1
6	bertahun-tahun	0	0	1	0
7	bilang	1	0	0	0
8	calon	1	0	0	0
9	cepat	0	0	0	1
10	citra	0	0	1	0
11	daerah	0	1	0	0
12	detikjabar	1	0	0	0
13	detikpilkada	1	0	0	0
14	full	1	0	0	0
15	gubernur	1	0	0	0
16	hadir	0	0	0	1
17	iya	1	0	0	0
18	jalan	0	0	1	0
19	jeje	1	1	1	1
20	kecil	1	0	0	0
21	keluh	0	0	1	0
22	kota	0	1	0	0
23	nomor	1	0	0	0
24	omong	1	0	0	0
25	pangandaran	1	0	0	0
26	pariwisata	1	0	0	0
27	peduli	0	1	0	0
28	pilgubjabar	1	0	0	0
29	publik	0	1	0	0
30	ramah	0	1	0	0
31	rapi	0	1	0	0
32	ruang	0	1	0	0
33	rusak	0	0	1	0
34	salut	0	0	0	1
35	sejawa	1	0	0	0

36	sibuk	0	0	1	0
37	taman	0	1	0	0
38	tanggap	0	0	0	1
39	tuh	1	0	0	0
40	urus	0	0	1	0
41	urut	1	0	0	0
42	warga	0	0	1	0
43	wiradinata	1	1	0	0

Tabel 4. 45 Hasil Perhitungan TF

No	Term	D1	D2	D3	D4
1	anak	0	0,090909	0	0
2	anggur	0,05	0	0	0
3	bantu	0	0	0	0,142857
4	barat	0,05	0	0	0
5	bencana	0	0	0	0,142857
6	bertahuntahun	0	0	0,111111	0
7	bilang	0,05	0	0	0
8	calon	0,05	0	0	0
9	cepat	0	0	0	0,142857
10	citra	0	0	0,111111	0
11	daerah	0	0,090909	0	0
12	detikjabar	0,05	0	0	0
13	detikpilkada	0,05	0	0	0
14	full	0,05	0	0	0
15	gubernur	0,05	0	0	0
16	hadir	0	0	0	0,142857
17	iya	0,05	0	0	0
18	jalan	0	0	0,111111	0
19	jeje	0,05	0,090909	0,111111	0,142857
20	kecil	0,05	0	0	0
21	keluh	0	0	0,111111	0
22	kota	0	0,090909	0	0
23	nomor	0,05	0	0	0
24	omong	0,05	0	0	0
25	pangandaran	0,05	0	0	0
26	pariwisata	0,05	0	0	0
27	peduli	0	0,090909	0	0
28	pilgubjabar	0,05	0	0	0
29	publik	0	0,090909	0	0
30	ramah	0	0,090909	0	0
31	rapi	0	0,090909	0	0
32	ruang	0	0,090909	0	0

33	rusak	0	0	0,111111	0
34	salut	0	0	0	0,142857
35	sejawa	0,05	0	0	0
36	sibuk	0	0	0,111111	0
37	taman	0	0,090909	0	0
38	tanggap	0	0	0	0,142857
39	tuh	0,05	0	0	0
40	urus	0	0	0,111111	0
41	urut	0,05	0	0	0
42	warga	0	0	0,111111	0
43	wiradinata	0,05	0,090909	0	0

Tabel 4. 46 Hasil Perhitungan DF

No	Term	D1	D2	D3	D4	DF
1	anak	0	1	0	0	1
2	anggur	1	0	0	0	1
3	bantu	0	0	0	1	1
4	barat	1	0	0	0	1
5	bencana	0	0	0	1	1
6	bertahun	0	0	1	0	1
7	bilang	1	0	0	0	1
8	calon	1	0	0	0	1
9	cepat	0	0	0	1	1
10	citra	0	0	1	0	1
11	daerah	0	1	0	0	1
12	detikjabar	1	0	0	0	1
13	detikpilkada	1	0	0	0	1
14	full	1	0	0	0	1
15	gubernur	1	0	0	0	1
16	hadir	0	0	0	1	1
17	iya	1	0	0	0	1
18	jalan	0	0	1	0	1
19	jeje	1	1	1	1	4
20	kecil	1	0	0	0	1
21	keluh	0	0	1	0	1
22	kota	0	1	0	0	1
23	nomor	1	0	0	0	1
24	omong	1	0	0	0	1
25	pangandaran	1	0	0	0	1
26	pariwisata	1	0	0	0	1
27	peduli	0	1	0	0	1
28	pilgubjabar	1	0	0	0	1

29	publik	0	1	0	0	1
30	ramah	0	1	0	0	1
31	rapi	0	1	0	0	1
32	ruang	0	1	0	0	1
33	rusak	0	0	1	0	1
34	salut	0	0	0	1	1
35	sejawa	1	0	0	0	1
36	sibuk	0	0	1	0	1
37	taman	0	1	0	0	1
38	tanggap	0	0	0	1	1
39	tuh	1	0	0	0	1
40	urus	0	0	1	0	1
41	urut	1	0	0	0	1
42	warga	0	0	1	0	1
43	wiradinata	1	1	0	0	1

Tabel 4. 47 Hasil Perhitungan IDF

No	Term	D1	D2	D3	D4	DF	IDF
1	anak	0	1	0	0	1	1,386294
2	anggur	1	0	0	0	1	1,386294
3	bantu	0	0	0	1	1	1,386294
4	barat	1	0	0	0	1	1,386294
5	bencana	0	0	0	1	1	1,386294
6	bertahun-tahun	0	0	1	0	1	1,386294
7	bilang	1	0	0	0	1	1,386294
8	calon	1	0	0	0	1	1,386294
9	cepat	0	0	0	1	1	1,386294
10	citra	0	0	1	0	1	1,386294
11	daerah	0	1	0	0	1	1,386294
12	detikjabar	1	0	0	0	1	1,386294
13	detikpilkada	1	0	0	0	1	1,386294
14	full	1	0	0	0	1	1,386294
15	gubernur	1	0	0	0	1	1,386294
16	hadir	0	0	0	1	1	1,386294
17	iya	1	0	0	0	1	1,386294
18	jalan	0	0	1	0	1	1,386294
19	jeje	1	1	1	1	4	0
20	kecil	1	0	0	0	1	1,386294
21	keluh	0	0	1	0	1	1,386294
22	kota	0	1	0	0	1	1,386294
23	nomor	1	0	0	0	1	1,386294
24	omong	1	0	0	0	1	1,386294
25	pangandaran	1	0	0	0	1	1,386294

26	pariwisata	1	0	0	0	1	1,386294
27	peduli	0	1	0	0	1	1,386294
28	pilgubjabar	1	0	0	0	1	1,386294
29	publik	0	1	0	0	1	1,386294
30	ramah	0	1	0	0	1	1,386294
31	rapi	0	1	0	0	1	1,386294
32	ruang	0	1	0	0	1	1,386294
33	rusak	0	0	1	0	1	1,386294
34	salut	0	0	0	1	1	1,386294
35	sejawa	1	0	0	0	1	1,386294
36	sibuk	0	0	1	0	1	1,386294
37	taman	0	1	0	0	1	1,386294
38	tanggap	0	0	0	1	1	1,386294
39	tuh	1	0	0	0	1	1,386294
40	urus	0	0	1	0	1	1,386294
41	urut	1	0	0	0	1	1,386294
42	warga	0	0	1	0	1	1,386294
43	wiradinata	1	1	0	0	2	0,693147

Tabel 4. 48 Hasil Perhitungan TF-IDF

No	Term	D1	D2	D3	D4
1	anak	0	0,126026	0	0
2	anggur	0,069315	0	0	0
3	bantu	0	0	0	0,198042
4	barat	0,069314	0	0	0
5	bencana	0	0	0	0,198042
6	bertahun-tahun	0	0	0,154032	0
7	bilang	0,069314	0	0	0
8	calon	0,069314	0	0	0
9	cepat	0	0	0	0,198042
10	citra	0	0	0,154032	0
11	daerah	0	0,126026	0	0
12	detikjabar	0,069314	0	0	0
13	detikpilkada	0,069314	0	0	0
14	full	0,069314	0	0	0
15	gubernur	0,069314	0	0	0
16	hadir	0	0	0	0,198042
17	iya	0,069314	0	0	0
18	jalan	0	0	0,154032	0
19	jeje	0	0	0	0
20	kecil	0,069314	0	0	0
21	keluh	0	0	0,154032	0
22	kota	0	0,126026	0	0

23	nomor	0,069314	0	0	0
24	omong	0,069314	0	0	0
25	pangandaran	0,069314	0	0	0
26	pariwisata	0,069314	0	0	0
27	peduli	0	0,126026	0	0
28	pilgubjabar	0,069314	0	0	0
29	publik	0	0,126026	0	0
30	ramah	0	0,126026	0	0
31	rapi	0	0,126026	0	0
32	ruang	0	0,126026	0	0
33	rusak	0	0	0,154032	0
34	salut	0	0	0	0,198042
35	sejawa	0,069314	0	0	0
36	sibuk	0	0	0,154032	0
37	taman	0	0,126026	0	0
38	tanggap	0	0	0	0,198042
39	tuh	0,069314	0	0	0
40	urus	0	0	0,154032	0
41	urut	0,069314	0	0	0
42	warga	0	0	0,154032	0
43	wiradinata	0,034657	0,063013	0	0

Tabel 4. 49 Data Sample Perhitungan TF-IDF Paslon 3

Dokumen	Dokumen	Teks	Label
1	D15	gentel mundur bukti gagal adik leas stan kecewa langkah praktisnya jaga marwah juang cocok lu tetap	Negatif
2	D16	pks jabar kalah telak ya presiden diturunin jabar kalah ust syaikh prestasi ketua dpw pks kalah sekjend dpw rinso kalah kursi dpri dprd provinsi pks jabar	Negatif
3	D17	haha mumpung angka gue nyoblos pks nyoblos pks anggota dewan mana kader pks dipungkiri bawa manfaat rakyat kemarin gue nyoblos ahmad syaikh ilham jabar gue cocok	Positif
4	D18	pks hatiseemoga orang barokah rahmat allah subhanahuwataala	Positif

Tabel 4. 50 Hasil Frekuensi Kemunculan Setiap Kata Dalam Dokumen

No	Term	D1	D2	D3	D4
1	adik	1	0	0	0
2	ahmad	0	0	1	0
3	allah	0	0	0	1
4	anggota	0	0	1	0
5	barokah	0	0	0	1
6	bawa	0	0	1	0

7	bukti	1	0	0	0
8	cocok	1	0	1	0
9	dewan	0	0	1	0
10	dipungkiri	0	0	1	0
11	diturunin	0	1	0	0
12	dprd	0	1	0	0
13	dprri	0	1	0	0
14	dpw	0	2	0	0
15	gagal	1	0	0	0
16	gentel	1	0	0	0
17	gue	0	0	3	0
18	haha	0	0	1	0
19	hatisemoga	0	0	0	1
20	ilham	0	0	1	0
21	jabar	0	3	1	0
22	jaga	1	0	0	0
23	juang	1	0	0	0
24	kader	0	0	1	0
25	kalah	0	4	0	0
26	kecewa	1	0	0	0
27	kemarin	0	0	1	0
28	ketua	0	1	0	0
29	kursi	0	1	0	0
30	langkah	1	0	0	0
31	leas	1	0	0	0
32	lu	1	0	0	0
33	mana	0	0	1	0
34	manfaat	0	0	1	0
35	marwah	1	0	0	0
36	mumpung	0	0	1	0
37	mundur	1	0	0	0
38	nangka	0	0	1	0
39	nyoblos	0	0	3	0
40	orang	0	0	0	1
41	pks	0	3	3	1
42	praktisnnya	1	0	0	0
43	presiden	0	1	0	0
44	prestasi	0	1	0	0
45	provinsi	0	1	0	0
46	rahmat	0	0	0	1
47	rakyat	0	0	1	0
48	rinso	0	1	0	0
49	sekjend	0	1	0	0
50	stan	1	0	0	0

51	subhanahuwataala	0	0	0	1
52	syaikhu	0	1	1	0
53	telak	0	1	0	0
54	tetap	1	0	0	0
55	ust	0	1	0	0
56	ya	0	1	0	0

Tabel 4. 51 Hasil Perhitungan TF

No	Term	D1	D2	D3	D4
1	adik	0,0625	0	0	0
2	ahmad	0	0	0,038461	0
3	allah	0	0	0	0,142857
4	anggota	0	0	0,038461	0
5	barokah	0	0	0	0,142857
6	bawa	0	0	0,038461	0
7	bukti	0,0625	0	0	0
8	cocok	0,0625	0	0,038461	0
9	dewan	0	0	0,038461	0
10	dipungkiri	0	0	0,038461	0
11	diturunin	0	0,038461	0	0
12	dprd	0	0,038461	0	0
13	dprri	0	0,038461	0	0
14	dpw	0	0,076923	0	0
15	gagal	0,0625	0	0	0
16	gentel	0,0625	0	0	0
17	gue	0	0	0,115384	0
18	haha	0	0	0,038461	0
19	hatisemoga	0	0	0	0,142857
20	ilham	0	0	0,038461	0
21	jabar	0	0,115384	0,038461	0
22	jaga	0,0625	0	0	0
23	juang	0,0625	0	0	0
24	kader	0	0	0,038461	0
25	kalah	0	0,153846	0	0
26	kecewa	0,0625	0	0	0
27	kemarin	0	0	0,038461	0
28	ketua	0	0,038461	0	0
29	kursi	0	0,038461	0	0
30	langkah	0,0625	0	0	0
31	leas	0,0625	0	0	0
32	lu	0,0625	0	0	0
33	mana	0	0	0,038461	0
34	manfaat	0	0	0,038461	0

35	marwah	0,0625	0	0	0
36	mumpung	0	0	0,038461	0
37	mundur	0,0625	0	0	0
38	nangka	0	0	0,038561	0
39	nyoblos	0	0	0,115384	0
40	orang	0	0	0	0,142857
41	pks	0	0,115384	0,115384	0,142857
42	praktisnnya	0,0625	0	0	0
43	presiden	0	0,038461	0	0
44	prestasi	0	0,038461	0	0
45	provinsi	0	0,038461	0	0
46	rahmat	0	0	0	0,142857
47	rakyat	0	0	0,038461	0
48	rinso	0	0,038461	0	0
49	sekjend	0	0,038461	0	0
50	stan	0,0625	0	0	0
51	subhanahuwataala	0	0	0	0,142857
52	syaikhu	0	0,038461	0,38461	0
53	telak	0	0,038461	0	0
54	tetap	0,0625	0	0	0
55	ust	0	0,038461	0	0
56	ya	0	0,038461	0	0

Tabel 4. 52 Hasil Perhitungan DF

No	Term	D1	D2	D3	D4	DF
1	adik	1	0	0	0	1
2	ahmad	0	0	1	0	1
3	allah	0	0	0	1	1
4	anggota	0	0	1	0	1
5	barokah	0	0	0	1	1
6	bawa	0	0	1	0	1
7	bukti	1	0	0	0	1
8	cocok	1	0	1	0	2
9	dewan	0	0	1	0	1
10	dipungkiri	0	0	1	0	1
11	diturunin	0	1	0	0	1
12	dprd	0	1	0	0	1
13	dprri	0	1	0	0	1
14	dpw	0	2	0	0	1
15	gagal	1	0	0	0	1
16	gentel	1	0	0	0	1
17	gue	0	0	3	0	1
18	haha	0	0	1	0	1

19	hatisemoga	0	0	0	1	1
20	ilham	0	0	1	0	1
21	jabar	0	3	1	0	2
22	jaga	1	0	0	0	1
23	juang	1	0	0	0	1
24	kader	0	0	1	0	1
25	kalah	0	4	0	0	1
26	kecewa	1	0	0	0	1
27	kemarin	0	0	1	0	1
28	ketua	0	1	0	0	1
29	kursi	0	1	0	0	1
30	langkah	1	0	0	0	1
31	leas	1	0	0	0	1
32	lu	1	0	0	0	1
33	mana	0	0	1	0	1
34	manfaat	0	0	1	0	1
35	marwah	1	0	0	0	1
36	mumpung	0	0	1	0	1
37	mundur	1	0	0	0	1
38	angka	0	0	1	0	1
39	nyoblos	0	0	3	0	3
40	orang	0	0	0	1	1
41	pks	0	3	3	1	3
42	praktisnnya	1	0	0	0	1
43	presiden	0	1	0	0	1
44	prestasi	0	1	0	0	1
45	provinsi	0	1	0	0	1
46	rahmat	0	0	0	1	1
47	rakyat	0	0	1	0	1
48	rinso	0	1	0	0	1
49	sekjend	0	1	0	0	1
50	stan	1	0	0	0	1
51	subhanahuwataala	0	0	0	1	1
52	syaikhu	0	1	1	0	2
53	telak	0	1	0	0	1
54	tetap	1	0	0	0	1
55	ust	0	1	0	0	1
56	ya	0	1	0	0	1

Tabel 4. 53 Hasil Perhitungan IDF

No	Term	D1	D2	D3	D4	DF	IDF
1	adik	1	0	0	0	1	1,386294
2	ahmad	0	0	1	0	1	1,386294

3	allah	0	0	0	1	1	1,386294
4	anggota	0	0	1	0	1	1,386294
5	barokah	0	0	0	1	1	1,386294
6	bawa	0	0	1	0	1	1,386294
7	bukti	1	0	0	0	1	1,386294
8	cocok	1	0	1	0	2	0,693147
9	dewan	0	0	1	0	1	1,386294
10	dipungkiri	0	0	1	0	1	1,386294
11	diturunin	0	1	0	0	1	1,386294
12	dprd	0	1	0	0	1	1,386294
13	dpri	0	1	0	0	1	1,386294
14	dpuw	0	2	0	0	1	1,386294
15	gagal	1	0	0	0	1	1,386294
16	gentel	1	0	0	0	1	1,386294
17	gue	0	0	3	0	1	1,386294
18	haha	0	0	1	0	1	1,386294
19	hatisemoga	0	0	0	1	1	1,386294
20	ilham	0	0	1	0	1	1,386294
21	jabar	0	3	1	0	2	0,693147
22	jaga	1	0	0	0	1	1,386294
23	juang	1	0	0	0	1	1,386294
24	kader	0	0	1	0	1	1,386294
25	kalah	0	4	0	0	1	1,386294
26	kecewa	1	0	0	0	1	1,386294
27	kemarin	0	0	1	0	1	1,386294
28	ketua	0	1	0	0	1	1,386294
29	kursi	0	1	0	0	1	1,386294
30	langkah	1	0	0	0	1	1,386294
31	leas	1	0	0	0	1	1,386294
32	lu	1	0	0	0	1	1,386294
33	mana	0	0	1	0	1	1,386294
34	manfaat	0	0	1	0	1	1,386294
35	marwah	1	0	0	0	1	1,386294
36	mumpung	0	0	1	0	1	1,386294
37	mundur	1	0	0	0	1	1,386294
38	nangka	0	0	1	0	1	1,386294
39	nyoblos	0	0	3	0	1	1,386294
40	orang	0	0	0	1	1	1,386294
41	pks	0	3	3	1	3	0,287682
42	praktisnnya	1	0	0	0	1	1,386294
43	presiden	0	1	0	0	1	1,386294
44	prestasi	0	1	0	0	1	1,386294
45	provinsi	0	1	0	0	1	1,386294
46	rahmat	0	0	0	1	1	1,386294

47	rakyat	0	0	1	0	1	1,386294
48	rinso	0	1	0	0	1	1,386294
49	sekjend	0	1	0	0	1	1,386294
50	stan	1	0	0	0	1	1,386294
51	subhanahuwataal	0	0	0	1	1	1,386294
52	syaikh	0	1	1	0	2	0,693147
53	telak	0	1	0	0	1	1,386294
54	tetap	1	0	0	0	1	1,386294
55	ust	0	1	0	0	1	1,386294
56	ya	0	1	0	0	1	1,386294

Tabel 4. 54 Hasil Perhitungan TF-IDF

No	Term	D1	D2	D3	D4
1	adik	0,086643	0	0	0
2	ahmad	0	0	0,053319	0
3	allah	0	0	0	0,198042
4	anggota	0	0	0,053319	0
5	barokah	0	0	0	0,198042
6	bawa	0	0	0,053319	0
7	bukti	0,866643	0	0	0
8	cocok	0,043321	0	0,026659	0
9	dewan	0	0	0,053319	0
10	dipungkiri	0	0	0,053319	0
11	diturunin	0	0,053319	0	0
12	dprd	0	0,053319	0	0
13	dprri	0	0,053319	0	0
14	dpw	0	0,016638	0	0
15	gagal	0,086643	0	0	0
16	gentel	0,086643	0	0	0
17	gue	0	0	0,159957	0
18	haha	0	0	0,053319	0
19	hatisemoga	0	0	0	0,198042
20	ilham	0	0	0,053319	0
21	jabar	0	0,079978	0,026659	0
22	jaga	0,086643	0	0	0
23	juang	0,086643	0	0	0
24	kader	0	0	0,053319	0
25	kalah	0	0,213276	0	0
26	kecewa	0,086643	0	0	0
27	kemarin	0	0	0,053319	0
28	ketua	0	0,053319	0	0
29	kursi	0	0,053319	0	0

30	langkah	0,086643	0	0	0
31	leas	0,086643	0	0	0
32	lu	0,086643	0	0	0
33	mana	0	0	0,053319	0
34	manfaat	0	0	0,053319	0
35	marwah	0,086643	0	0	0
36	mumpung	0	0	0,053319	0
37	mundur	0,086643	0	0	0
38	nangka	0	0	0,053319	0
39	nyoblos	0	0	0,159957	0
40	orang	0	0	0	0,198042
41	pks	0	0,033194	0,033194	0,041097
42	praktisnnya	0,086643	0	0	0
43	presiden	0	0,053319	0	0
44	prestasi	0	0,053319	0	0
45	provinsi	0	0,053319	0	0
46	rahmat	0	0	0	0,198042
47	rakyat	0	0	0,053319	0
48	rinso	0	0,053319	0	0
49	sekjend	0	0,053319	0	0
50	stan	0,086643	0	0	0
51	subhanahuwataala	0	0	0	0,198042
52	syaihu	0	0,026659	0,026659	0
53	telak	0	0,053319	0	0
54	tetap	0,086643	0	0	0
55	ust	0	0,053319	0	0
56	ya	0	0,053319	0	0

Tabel 4. 55 Data Sample Perhitungan TF-IDF Paslon 4

No	Dokumen	Teks	Label
1	D15	gembira dedi mulyadi anti baalawi statemennya beliau suka habib habib	Negatif
2	D16	tau suku sunda istilah islam sunda sunda islam saking kental islam masuk budaya sunda kang dedi mulyadi terima jawa barat	Negatif
3	D17	malam tonton podcast dedi mulyadi gue tuju ngakunya era digital sppd numpuk acara hotel	Positif
4	D18	tarik lihat depan sepak terjang dedi mulyadi jawa barat cari solusi intoleransi deradikalisme	Positif

Tabel 4. 56 Hasil Frekuensi Kemunculan Setiap Kata Dalam Dokumen

No	Term	D1	D2	D3	D4
1	acara	0	0	1	0
2	anti	1	0	0	0

3	baalawi	1	0	0	0
4	barat	0	1	0	1
5	beliau	1	0	0	0
6	budaya	0	1	0	0
7	cari	0	0	0	1
8	dedi	1	1	1	1
9	depan	0	0	0	1
10	deradikalisme	0	0	0	1
11	digital	0	0	1	0
12	era	0	0	1	0
13	gembira	1	0	0	0
14	gue	0	0	1	0
15	habib	2	0	0	0
16	hotel	0	0	1	0
17	intoleransi	0	0	0	1
18	islam	0	3	0	0
19	istilah	0	1	0	0
20	jawa	0	1	0	1
21	kang	0	1	0	0
22	kental	0	1	0	0
23	lihat	0	0	0	1
24	malam	0	0	1	0
25	masuk	0	1	0	0
26	mulyadi	1	1	1	1
27	ngakunya	0	0	1	0
28	numpuk	0	0	1	0
29	podcast	0	0	1	0
30	saking	0	1	0	0
31	sepak	0	0	0	1
32	solusi	0	0	0	1
33	sppd	0	0	1	0
34	statemennya	1	0	0	0
35	suka	1	0	0	0
36	suku	0	1	0	0
37	sunda	0	4	0	0
38	tarik	0	0	0	1
39	tau	0	1	0	0
40	terima	0	1	0	0
41	terjang	0	0	0	1
42	tonton	0	0	1	0
43	tuju	0	0	1	0

Tabel 4. 57 Hasil Perhitungan TF

No	Term	D1	D2	D3	D4
1	acara	0	0	0,071428	0
2	anti	0,1	0	0	0
3	baalawi	0,1	0	0	0
4	barat	0	0,05	0	0,076923
5	beliau	0,1	0	0	0
6	budaya	0	0,05	0	0
7	cari	0	0	0	0,076923
8	dedi	0,1	0,05	0,071428	0,076923
9	depan	0	0	0	0,076923
10	deradikalisme	0	0	0	0,076923
11	digital	0	0	0,071428	0
12	era	0	0	0,071428	0
13	gembira	0.1	0	0	0
14	gue	0	0	0,071428	0
15	habib	0.2	0	0	0
16	hotel	0	0	0,071428	0
17	intoleransi	0	0	0	0,076923
18	islam	0	0,15	0	0
19	istilah	0	0,05	0	0
20	jawa	0	0,05	0	0,076923
21	kang	0	0,05	0	0
22	kental	0	0,05	0	0
23	lihat	0	0	0	0,076923
24	malam	0	0	0,071428	0
25	masuk	0	0,05	0	0
26	mulyadi	0,100000	0,05	0,071428	0,076923
27	ngakunya	0	0	0,071428	0
28	numpuk	0	0	0,071428	0
29	podcast	0	0	0,071428	0
30	saking	0	0,05	0	0
31	sepak	0	0	0	0,076923
32	solusi	0	0	0	0,076923
33	sppd	0	0	0,071428	0
34	statemennya	0.1	0	0	0
35	suka	0.1	0	0	0
36	suku	0	0,05	0	0
37	sunda	0	0,2	0	0
38	tarik	0	0	0	0,076923
39	tau	0	0,05	0	0
40	terima	0	0,05	0	0

41	terjang	0	0	0	0,076923
42	tonton	0	0	0,071428	0
43	tuju	0	0	0,071428	0

Tabel 4. 58 Hasil Perhitungan DF

No	Term	D1	D2	D3	D4	DF
1	acara	0	0	1	0	1
2	anti	1	0	0	0	1
3	baalawi	1	0	0	0	1
4	barat	0	1	0	1	2
5	beliau	1	0	0	0	1
6	budaya	0	1	0	0	1
7	cari	0	0	0	1	1
8	dedi	1	1	1	1	4
9	depan	0	0	0	1	1
10	deradikalisme	0	0	0	1	1
11	digital	0	0	1	0	1
12	era	0	0	1	0	1
13	gembira	1	0	0	0	1
14	gue	0	0	1	0	1
15	habib	2	0	0	0	1
16	hotel	0	0	1	0	1
17	intoleransi	0	0	0	1	1
18	islam	0	3	0	0	1
19	istilah	0	1	0	0	1
20	jawa	0	1	0	1	2
21	kang	0	1	0	0	1
22	kental	0	1	0	0	1
23	lihat	0	0	0	1	1
24	malam	0	0	1	0	1
25	masuk	0	1	0	0	1
26	mulyadi	1	1	1	1	4
27	ngakunya	0	0	1	0	1
28	numpuk	0	0	1	0	1
29	podcast	0	0	1	0	1
30	saking	0	1	0	0	1
31	sepak	0	0	0	1	1
32	solusi	0	0	0	1	1
33	sppd	0	0	1	0	1
34	statemennya	1	0	0	0	1
35	suka	1	0	0	0	1
36	suku	0	1	0	0	1
37	sunda	0	4	0	0	1

38	tarik	0	0	0	1	1
39	tau	0	1	0	0	1
40	terima	0	1	0	0	1
41	terjang	0	0	0	1	1
42	tonton	0	0	1	0	1
43	tuju	0	0	1	0	1

Tabel 4. 59 Hasil Perhitungan IDF

No	Term	D1	D2	D3	D4	DF	IDF
1	acara	0	0	1	0	1	1,386294
2	anti	1	0	0	0	1	1,386294
3	baalawi	1	0	0	0	1	1,386294
4	barat	0	1	0	1	2	0,693147
5	beliau	1	0	0	0	1	1,386294
6	budaya	0	1	0	0	1	1,386294
7	cari	0	0	0	1	1	1,386294
8	dedi	1	1	1	1	4	0
9	depan	0	0	0	1	1	1,386294
10	deradikalisme	0	0	0	1	1	1,386294
11	digital	0	0	1	0	1	1,386294
12	era	0	0	1	0	1	1,386294
13	gembira	1	0	0	0	1	1,386294
14	gue	0	0	1	0	1	1,386294
15	habib	2	0	0	0	1	1,386294
16	hotel	0	0	1	0	1	1,386294
17	intoleransi	0	0	0	1	1	1,386294
18	islam	0	3	0	0	1	1,386294
19	istilah	0	1	0	0	1	1,386294
20	jawa	0	1	0	1	2	0,693147
21	kang	0	1	0	0	1	1,386294
22	kental	0	1	0	0	1	1,386294
23	lihat	0	0	0	1	1	1,386294
24	malam	0	0	1	0	1	1,386294
25	masuk	0	1	0	0	1	1,386294
26	mulyadi	1	1	1	1	4	0
27	ngakunya	0	0	1	0	1	1,386294
28	numpuk	0	0	1	0	1	1,386294
29	podcast	0	0	1	0	1	1,386294
30	saking	0	1	0	0	1	1,386294
31	sepak	0	0	0	1	1	1,386294
32	solusi	0	0	0	1	1	1,386294
33	sppd	0	0	1	0	1	1,386294
34	statemennya	1	0	0	0	1	1,386294
35	suka	1	0	0	0	1	1,386294

36	suku	0	1	0	0	1	1,386294
37	sunda	0	4	0	0	1	1,386294
38	tarik	0	0	0	1	1	1,386294
39	tau	0	1	0	0	1	1,386294
40	terima	0	1	0	0	1	1,386294
41	terjang	0	0	0	1	1	1,386294
42	tonton	0	0	1	0	1	1,386294
43	tuju	0	0	1	0	1	1,386294

Tabel 4. 60 Hasil Perhitungan TF-DF

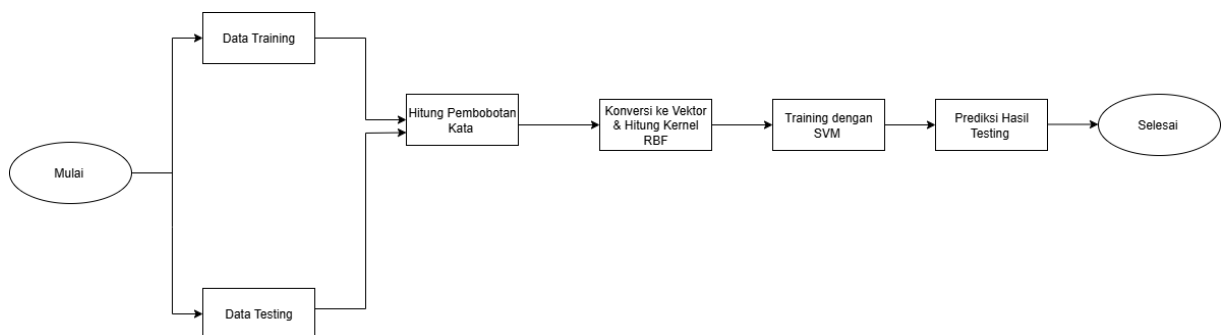
No	Term	D1	D2	D3	D4
1	acara	0	0	0,099021	0
2	anti	0,138629	0	0	0
3	baalawi	0,138629	0	0	0
4	barat	0	0,034657	0	0,053319
5	beliau	0,138629	0	0	0
6	budaya	0	0,069315	0	0
7	cari	0	0	0	0,106638
8	dedi	0	0	0	0
9	depan	0	0	0	0,106638
10	deradikalisme	0	0	0	0,106638
11	digital	0	0	0,099021	0
12	era	0	0	0,099021	0
13	gembira	0,138629	0	0	0
14	gue	0	0	0,099021	0
15	habib	0,277259	0	0	0
16	hotel	0	0	0,099021	0
17	intoleransi	0	0	0	0,106638
18	islam	0	0,207944	0	0
19	istilah	0	0,069315	0	0
20	jawa	0	0,034657	0	0,053319
21	kang	0	0,069315	0	0
22	kental	0	0,069315	0	0
23	lihat	0	0	0	0,106638
24	malam	0	0	0,099021	0
25	masuk	0	0,069315	0	0
26	mulyadi	0	0	0	0
27	ngakunya	0	0	0,099021	0
28	numpuk	0	0	0,099021	0
29	podcast	0	0	0,099021	0
30	saking	0	0,069315	0	0
31	sepak	0	0	0	0,106638
32	solusi	0	0	0	0,106638

33	sppd	0	0	0,099021	0
34	statemennya	0,138629	0	0	0
35	suka	0,138629	0	0	0
36	suku	0	0,069315	0	0
37	sunda	0	0,277259	0	0
38	tarik	0	0	0	0,106638
39	tau	0	0,069315	0	0
40	terima	0	0,069315	0	0
41	terjang	0	0	0	0,106638
42	tonton	0	0	0,099021	0
43	tuju	0	0	0,099021	0

4.7 Klasifikasi Support Vector Machine

Setelah tahap pembobotan kata berhasil dilakukan, langkah selanjutnya yaitu melakukan proses klasifikasi menggunakan *support vector machine* (SVM). Algoritma SVM digunakan untuk mencari *hyperplane* terbaik dari dua kelas, yaitu kelas positif dan kelas negatif [FAJ23]. Nilai input yang digunakan pada tahap ini yaitu nilai vektor dari hasil proses pembobotan atau TF-IDF. Berikut adalah proses perhitungan SVM dari setiap Paslon.

Berikut adalah tahapan dari *Support Vector Machine*.



Gambar 4. 3 Tahapan Support Vector Machine

Tabel 4. 61 Data Vector Paslon 1

Dok	Teks	Label
D72	acep gita nomer bawa jabar bahagia	Positif
D249	program acepadang nyontek paslon kreatif dikit	Negatif
X^0	0,055324 0,078954 0,211304 0,368976 0,040882 0,15143	1
X^1	0,34556 0,877976 1,146216 1,030691 1,030691 1,146216	-1

Nilai vector diperoleh melalui perhitungan menggunakan TF-IDF yang kemudian

nilai tersebut di ubah kedalam bentuk vector. Sehingga label atau nilai y selanjutnya di representasikan dengan nilai -1 jika negatif dan 1 jika nilai tersebut positif. Selanjutnya menghitung nilai $\vec{X}_l$ yang ada pada setiap data, maka dilakukan perhitungan seperti sebagai berikut.

$$\vec{X}_0 = [0,055324 \ 0,078954 \ 0,211304 \ 0,368976 \ 0,040882 \ 0,15143]$$

$$\vec{X}_1 = [0,34556 \ 0,877976 \ 1,146216 \ 1,030691 \ 1,030691 \ 1,146216]$$

Dalam studi kasus klasifikasi data, setiap dokumen atau kalimat direpresentasikan sebagai variabel $\vec{X}_l$, di mana $\vec{X}_l$ menunjukkan dokumen atau kalimat ke-I dalam dataset. Untuk mempermudah penamaan, digunakan notasi x dengan penomoran berurutan dari 0 hingga n, sesuai dengan jumlah total dokumen, dengan indeks I dimulai dari 0. Langkah berikutnya adalah melakukan operasi pengurangan vektor, yang hasilnya akan digunakan dalam fungsi kernel guna mentransformasikan data ke dalam ruang fitur (feature space). Operasi perhitungan vektor tersebut dijelaskan sebagai berikut.

$$\begin{aligned}\vec{X}_0 - \vec{X}_0 &= [0,055324 \ 0,078954 \ 0,211304 \ 0,368976 \ 0,040882 \ 0,15143] - \\ &[0,055324 \ 0,078954 \ 0,211304 \ 0,368976 \ 0,040882 \ 0,15143] \\ &= [0 \ 0 \ 0 \ 0 \ 0 \ 0]\end{aligned}$$

$$\vec{X}_0 - \vec{X}_1 = [-0,290222 \ -0,799022 \ -0,934912 \ -0,661715 \ -0,989809 \ -0,994785]$$

$$\vec{X}_1 - \vec{X}_0 = [0,290222 \ 0,799022 \ 0,934912 \ 0,661715 \ 0,989809 \ 0,994785]$$

$$\vec{X}_1 - \vec{X}_1 = [0 \ 0 \ 0 \ 0 \ 0 \ 0]$$

Setelah berhasil melakukan perhitungan vektor, langkah berikutnya adalah menghitung panjang vektor. Dalam konteks SVM, panjang vektor ini berperan penting untuk menentukan margin maksimum, yaitu jarak terpendek antara hyperplane dan titik-titik data dari dua kelas yang berbeda. Adapun persamaan untuk menghitung panjang vektor dijelaskan sebagai berikut.

$$\|\vec{X}_0 - \vec{X}_0\| = \sqrt{0^2 + 0^2 + 0^2 + 0^2 + 0^2 + 0^2} = 0$$

$$\|\vec{X}_0 - \vec{X}_1\| =$$

$$\begin{aligned}&\sqrt{-0,290222^2 - 0,799022^2 - 0,934912^2 - 0,661715^2 - 0,989809^2 - 0,99478^2} \\ &= 2,001\end{aligned}$$

$$\|\vec{X}_1 - \vec{X}_0\| = \sqrt{0,290222^2 + 0,799022^2 + 0,934912^2 + 0,661715^2 + 0,989809^2 + 0,99478^2}$$

$$= 2,001$$

$$\|\vec{X}_1 - \vec{X}_1\| = \sqrt{0^2 + 0^2 + 0^2 + 0^2 + 0^2 + 0^2} = 0$$

Pada tahap kernel, digunakan fungsi kernel non-linear untuk mentransformasikan data dari ruang input ke dalam ruang fitur (feature space). Transformasi ini bertujuan untuk mengatasi permasalahan ketika data tidak dapat dipisahkan secara linear dalam ruang input. Dengan menerapkan fungsi kernel, setiap titik data dipetakan ke dimensi yang lebih tinggi, sehingga memungkinkan penggunaan hyperplane linear yang lebih efektif dalam ruang fitur. Dalam penelitian ini, digunakan kernel Radial Basis Function (RBF), karena RBF dikenal memiliki performa yang sangat baik dalam menangani klasifikasi data non-linear [HUS21]. Berikut perhitungan fungsi RBF menggunakan nilai gamma (γ) 0,001 karena pada penelitian Margarita saat dilakukan pengujian nilai gamma dengan nilai 0.001, 0.01, 0.1, dan 1 yang memiliki akurasi paling baik adalah nilai gamma 0.001 [MAR24].

$$k\|\vec{X}_0 - \vec{X}_0\| = EXP(-\gamma\|\vec{X}_0 - \vec{X}_0\|^2)$$

$$k\|\vec{X}_0 - \vec{X}_0\| = EXP(-0,001(0^2))$$

$$k\|\vec{X}_0 - \vec{X}_0\| = 1$$

Berikut adalah hasil dari perhitungan fungsi kernel RBF.

$$k\|\vec{X}_0 - \vec{X}_0\| = 1$$

$$k\|\vec{X}_0 - \vec{X}_0\| = 0,996$$

$$k\|\vec{X}_0 - \vec{X}_0\| = 0,996$$

$$k\|\vec{X}_0 - \vec{X}_0\| = 1$$

Hasil dari perhitungan dari fungsi kernel diatas $K(\vec{X}_i - \vec{X}_j)$, Selanjutnya digunakan sebagai variabel dalam proses penentuan nilai bias. Nilai-nilai kernel tersebut disusun dalam bentuk matriks, yang akan digunakan pada tahap perhitungan berikutnya. Berikut ini adalah matriks hasil perhitungannya.

$$K(\vec{X}_i - \vec{X}_j) = \begin{bmatrix} 1 & 0,996 \\ 0,996 & 1 \end{bmatrix}$$

Setelah memperoleh nilai matriks kernel, langkah selanjutnya adalah menghitung nilai bobot. Perhitungan bobot dilakukan dengan menggunakan nilai support vector (α) yang bernilai 0,775. Berikut adalah perhitungan untuk mencari nilai bobot.

$$\vec{w} = \sum_{i=1}^l a_i \cdot y_i \cdot \vec{w}_i$$

$$\begin{aligned}
&= (0,775.(1) \left[\frac{1}{0,996} \right] + (0,775.(-1) \left[\frac{0,996}{1} \right]) \\
&= \left(\frac{0,0031}{-0,0031} \right)
\end{aligned}$$

Saat nilai bobot sudah di dapatkan, selanjutnya yaitu menghitung nilai bias dengan menggunakan perhitungan sebagai berikut.

$$\begin{aligned}
b &= -\frac{1}{2} ((\vec{w} \cdot \vec{x} - 1) + (\vec{w} \cdot \vec{x} + 1)) \\
b &= -\frac{1}{2} \left(\left(\left[\frac{0,0031}{-0,0031} \right] \cdot \left[\frac{0,996}{1} \right] + \left(\left[\frac{0,0031}{-0,0031} \right] \cdot \left[\frac{1}{0,996} \right] \right) \right) \\
b &= -\frac{1}{2} \left(\begin{bmatrix} 0 \\ 0 \end{bmatrix} \right) = 0
\end{aligned}$$

Setelah proses pelatihan menghasilkan nilai bobot dan bias, tahap selanjutnya adalah pengujian model. Pada fase ini, data uji yang telah dilabeli sebagai kelas positif dengan nilai 1 digunakan untuk mengevaluasi akurasi klasifikasi. Evaluasi dilakukan untuk memverifikasi apakah prediksi model sesuai dengan label asli. Seluruh data uji telah melewati tahap transformasi dan pembobotan. Dokumen pengujian disajikan pada bagian berikut.

Tabel 4. 62 Dokumen Testing Paslon 1

Dokumen	jargon	jabar	juara	kota	desadesa	tinggal
TF-IDF	1,030691	0,040882	0,963114	0,847589	0,780012	0,762452

Data di atas digunakan sebagai data uji, adapun bobot data uji tersebut adalah sebagai berikut.

$$\vec{X} = [1,030691 \ 0,040882 \ 0,963114 \ 0,847589 \ 0,780012 \ 0,762452]$$

Langkah selanjutnya yaitu mencari nilai vector dengan perhitungan pengurangan $\vec{x}$ dengan nilai dari data training.

$$\vec{X} - \vec{X}_0 = [1,030691 \ 0,040882 \ 0,963114 \ 0,847589 \ 0,780012 \ 0,762452] - [0,055324 \ 0,078954 \ 0,211304 \ 0,368976 \ 0,040882 \ 0,15143]$$

$$\vec{X} - \vec{X}_0 = [0,975367 \ -0,038072 \ -0,751810 \ 0,478613 \ 0,739130 \ 0,611021]$$

$$\vec{X} - \vec{X}_1 = [0,685145 \ -0,837094 \ -0,183102 \ -0,183102 \ -0,250679 \ -0,383764]$$

Setelah nilai vektor diperoleh, langkah berikutnya adalah menghitung panjang vektor dari data $\vec{x}$. Hasil dari perhitungan ini menghasilkan nilai panjang vektor sebagai berikut.

$$\begin{aligned}
\|\vec{X} - \vec{X}_0\| &= \\
&= \sqrt{0,975367^2 + (-0,038072)^2 + (-0,751810)^2 + 0,478613^2 + 0,739130^2 + 0,611021^2} \\
&= 1,6327
\end{aligned}$$

$$\begin{aligned} \|\vec{X} - \vec{X}_1\| &= \\ \sqrt{0,685145^2 + -0,837094^2 + -0,183102^2 + -0,183102^2 + -0,250679^2 + -0,383764^2} \\ &= 1,2030 \end{aligned}$$

Setelah memperoleh nilai panjang vektor, tahap selanjutnya adalah menghitung nilai fungsi kernel RBF untuk data uji. Penggunaan kernel RBF dipilih karena berdasarkan studi sebelumnya, kernel ini menunjukkan performa yang lebih unggul dibandingkan kernel linear dalam menangani data non-linear (Husada dan Paramita, 2021). Berikut ini disajikan hasil perhitungan dari fungsi kernel RBF yang digunakan pada data uji.

$$\begin{aligned} k\|\vec{X} - \vec{X}_0\| &= EXP(-\gamma\|\vec{X} - \vec{X}_0\|^2) \\ k\|\vec{X} - \vec{X}_0\| &= EXP(-0,001(1,6327^2)) \\ k\|\vec{X} - \vec{X}_0\| &= 0,997337 \end{aligned}$$

$$\begin{aligned} k\|\vec{X} - \vec{X}_1\| &= EXP(-\gamma\|\vec{X} - \vec{X}_1\|^2) \\ k\|\vec{X} - \vec{X}_1\| &= EXP(-0,001(1,2030^2)) \\ k\|\vec{X} - \vec{X}_1\| &= 0,998553 \end{aligned}$$

Setelah fungsi kernel RBF selesai dihitung, langkah selanjutnya adalah melakukan klasifikasi terhadap data uji (testing) menggunakan fungsi hyperplane dalam metode Support Vector Machine (SVM). Dalam tahap ini, nilai kernel yang telah diperoleh digunakan dalam fungsi keputusan untuk menentukan label prediksi dari data uji. Melalui hasil perhitungan ini, dapat diketahui kelas atau kategori dari masing-masing data uji. Perhitungan lengkap disajikan pada bagian berikut.

$$\begin{aligned} f(x) &= \text{sign}(\sum_{i=1, x_i = sv}^l a_i \cdot y_i \cdot K(\vec{x}, \vec{x}_i) + b) \\ f(x) &= \text{sign}(0,775 \cdot 0,997337 + 0) + (0,775 \cdot 0,998553 + 0) \\ f(x) &= 1 \end{aligned}$$

Dalam algoritma Support Vector Machine (SVM), proses klasifikasi ditentukan oleh fungsi keputusan yang membagi dua kelas melalui hyperplane dengan margin maksimum. Nilai keluaran dari fungsi ini menjadi dasar dalam menentukan kelas suatu data. Jika hasil prediksi menghasilkan nilai 1, maka data diklasifikasikan sebagai bagian dari kelas positif. Sebaliknya, jika nilai prediksi adalah -1, maka data tersebut termasuk dalam kelas negatif.

Tabel 4. 63 Data Vector Paslon 2

Dok	Teks	Label
D142	jeje wiradinata fokus tingkat sektor sehat	Positif
D188	urus seremoni jeje pas susah cari	Negatif
X^0	0,061213 0,248254 0,749800 0,606433 0,749800 0,500812	1
X^1	0,642029 1,149449 0,061213 0,802875 0,802875 0,825131	-1

Selanjutnya menghitung nilai $\vec{X}_l$ yang ada pada setiap data, maka dilakukan perhitungan seperti sebagai berikut.

$$\vec{X}_0 = [0,061213 \ 0,248254 \ 0,749800 \ 0,606433 \ 0,749800 \ 0,500812]$$

$$\vec{X}_1 = [0,642029 \ 1,149449 \ 0,061213 \ 0,802875 \ 0,802875 \ 0,825131]$$

Operasi perhitungan vektor tersebut dijelaskan sebagai berikut.

$$\begin{aligned}\vec{X}_0 - \vec{X}_0 &= [0,061213 \ 0,248254 \ 0,749800 \ 0,606433 \ 0,749800 \ 0,500812] - [0,061213 \\ 0,248254 \ 0,749800 \ 0,606433 \ 0,749800 \ 0,500812] \\ &= [0 \ 0 \ 0 \ 0 \ 0 \ 0]\end{aligned}$$

$$\vec{X}_0 - \vec{X}_1 = [-0,580816 \ -0,901195 \ 0,688587 \ -0,196442 \ -0,053075 \ -0,324319]$$

$$\vec{X}_1 - \vec{X}_0 = [0,580816 \ 0,901195 \ -0,688587 \ -0,196442 \ -0,053075 \ -0,324319]$$

$$\vec{X}_1 - \vec{X}_1 = [0 \ 0 \ 0 \ 0 \ 0 \ 0]$$

Setelah berhasil melakukan perhitungan vektor, langkah berikutnya adalah menghitung panjang vektor. Dalam konteks SVM, panjang vektor ini berperan penting untuk menentukan margin maksimum, yaitu jarak terpendek antara hyperplane dan titik-titik data dari dua kelas yang berbeda. Adapun persamaan untuk menghitung panjang vektor dijelaskan sebagai berikut.

$$\|\vec{X}_0 - \vec{X}_0\| = \sqrt{0^2 + 0^2 + 0^2 + 0^2 + 0^2 + 0^2} = 0$$

$$\|\vec{X}_0 - \vec{X}_1\| =$$

$$\begin{aligned}&\sqrt{-0,580816^2 + -0,901195^2 + 0,688527^2 + -0,196442^2 + -0,053075^2 + -0,324319^2} \\ &= 1,330\end{aligned}$$

$$\|\vec{X}_1 - \vec{X}_0\| =$$

$$\begin{aligned}&\sqrt{0,580816^2 + 0,901195^2 + -0,688527^2 + 0,196442^2 + 0,053075^2 + 0,324319^2} \\ &= 1,330\end{aligned}$$

$$\|\vec{X}_1 - \vec{X}_1\| = \sqrt{0^2 + 0^2 + 0^2 + 0^2 + 0^2 + 0^2} = 0$$

Berikut perhitungan fungsi RBF menggunakan nilai gamma (γ) 0,001

$$\begin{aligned}k_{||\vec{X}_0 - \vec{X}_0||} &= EXP(-\gamma ||\vec{X}_0 - \vec{X}_0||^2) \\k_{||\vec{X}_0 - \vec{X}_0||} &= EXP(-0,001(0^2)) \\k_{||\vec{X}_0 - \vec{X}_0||} &= 1\end{aligned}$$

Berikut adalah hasil dari perhitungan fungsi kernel RBF.

$$\begin{aligned}k_{||\vec{X}_0 - \vec{X}_0||} &= 1 \\k_{||\vec{X}_0 - \vec{X}_0||} &= 0,998 \\k_{||\vec{X}_0 - \vec{X}_0||} &= 0,998 \\k_{||\vec{X}_0 - \vec{X}_0||} &= 1\end{aligned}$$

Hasil dari perhitungan dari fungsi kernel diatas $K(\vec{X}_i - \vec{X}_j)$, Selanjutnya digunakan sebagai variabel dalam proses penentuan nilai bias. Nilai-nilai kernel tersebut disusun dalam bentuk matriks, yang akan digunakan pada tahap perhitungan berikutnya. Berikut ini adalah matriks hasil perhitungannya.

$$K(\vec{X}_i - \vec{X}_j) = \begin{bmatrix} 1 & 0,998 \\ 0,998 & 1 \end{bmatrix}$$

Setelah memperoleh nilai matriks kernel, langkah selanjutnya adalah menghitung nilai bobot. Perhitungan bobot dilakukan dengan menggunakan nilai support vector (α) yang bernilai 0,775. Berikut adalah perhitungan untuk mencari nilai bobot.

$$\begin{aligned}\vec{w} &= \sum_{i=1}^l a_i \cdot y_i \cdot \vec{w}_i \\&= (0,775 \cdot (1) \begin{bmatrix} 1 \\ 0,998 \end{bmatrix}) + (0,775 \cdot (-1) \begin{bmatrix} 0,998 \\ 1 \end{bmatrix}) \\&= \begin{pmatrix} 0,0016 \\ -0,0016 \end{pmatrix}\end{aligned}$$

selanjutnya yaitu menghitung nilai bias dengan menggunakan perhitungan sebagai berikut.

$$\begin{aligned}b &= -\frac{1}{2}((\vec{w} \cdot \vec{x} - 1) + (\vec{w} \cdot \vec{x} + 1)) \\b &= -\frac{1}{2}((\begin{bmatrix} 0,0016 \\ -0,0016 \end{bmatrix} \cdot \begin{bmatrix} 0,998 \\ 1 \end{bmatrix}) + (\begin{bmatrix} 0,0016 \\ -0,0016 \end{bmatrix} \cdot \begin{bmatrix} 1 \\ 0,998 \end{bmatrix})) \\b &= -\frac{1}{2}(\begin{bmatrix} 0 \\ 0 \end{bmatrix}) = 0\end{aligned}$$

Setelah proses pelatihan menghasilkan nilai bobot dan bias, tahap selanjutnya adalah pengujian model. Pada fase ini, data uji yang telah dilabeli sebagai kelas positif dengan nilai 1 digunakan untuk mengevaluasi akurasi klasifikasi. Evaluasi dilakukan

untuk memverifikasi apakah prediksi model sesuai dengan label asli. Seluruh data uji telah melewati tahap transformasi dan pembobotan. Dokumen pengujian disajikan pada bagian berikut.

Tabel 4. 64 Dokumen Testing Paslon 2

Dokumen	program sosial jeje wiradinata hasil salut
TF-IDF	0,358627 0,642029 0,061213 0,248254 0,577118 0,530520

Data di atas digunakan sebagai data uji, adapun bobot data uji tersebut adalah sebagai berikut.

$$\vec{X} = [0,358627 \ 0,642029 \ 0,061213 \ 0,248254 \ 0,577118 \ 0,530520]$$

Langkah selanjutnya yaitu mencari nilai vector dengan perhitungan pengurangan $\vec{x}$ dengan nilai dari data training.

$$\begin{aligned} \vec{X} - \vec{X}_0 &= [0,358627 \ 0,642029 \ 0,061213 \ 0,248254 \ 0,577118 \\ 0,530520] - [0,061213 \ 0,248254 \ 0,749800 \ 0,606433 \ 0,749800 \\ 0,500812] \end{aligned}$$

$$\vec{X} - \vec{X}_0 = [0,297414 \ 0,393775 \ -0,688587 \ -0,358179 \ -0,172682 \ 0,029708]$$

$$\vec{X} - \vec{X}_1 = [-0,283402 \ -0,50742 \ 0 \ -0,554621 \ -0,225757 \ -0,294611]$$

Setelah nilai vektor diperoleh, langkah berikutnya adalah menghitung panjang vektor dari data $\vec{x}$. Hasil dari perhitungan ini menghasilkan nilai panjang vektor sebagai berikut.

$$\begin{aligned} \|\vec{X} - \vec{X}_0\| &= \sqrt{0,297414^2 + -0,393775^2 + -0,688587^2 + -0,358179^2 + -0,172682^2 + 0,029708^2} \\ &= 0,9363 \end{aligned}$$

$$\begin{aligned} \|\vec{X} - \vec{X}_1\| &= \sqrt{-0,283402^2 + -0,50742^2 + 0^2 + -0,554621^2 + -0,225757^2 + -0,294611^2} \\ &= 0,8849 \end{aligned}$$

Berikut merupakan hasil perhitungan dari fungsi kernel RBF yang digunakan pada data uji.

$$k\|\vec{X} - \vec{X}_0\| = EXP(-\gamma\|\vec{X} - \vec{X}_0\|^2)$$

$$k\|\vec{X} - \vec{X}_0\| = EXP(-0,001(0,9363^2))$$

$$k\|\vec{X} - \vec{X}_0\| = 0,999124$$

$$k\|\vec{X} - \vec{X}_1\| = EXP(-\gamma\|\vec{X} - \vec{X}_1\|^2)$$

$$k\|\vec{X} - \vec{X}_1\| = EXP(-0,001(0,8849^2))$$

$$k\|\vec{X} - \vec{X}_1\| = 0,999217$$

Setelah fungsi kernel RBF selesai dihitung, langkah selanjutnya adalah melakukan klasifikasi terhadap data uji (testing) menggunakan fungsi hyperplane dalam metode Support Vector Machine (SVM). Dalam tahap ini, nilai kernel yang telah diperoleh digunakan dalam fungsi keputusan untuk menentukan label prediksi dari data uji. Melalui hasil perhitungan ini, dapat diketahui kelas atau kategori dari masing-masing data uji. Perhitungan lengkap disajikan pada bagian berikut.

$$f(x) = \text{sign}(\sum_{i=1, x_i = sv}^l a_i \cdot y_i \cdot K(\vec{x}, \vec{x}_i) + b)$$

$$f(x) = \text{sign}(0,775 \cdot 0,999124 + 0) + (0,775 \cdot 0,999217 + 0)$$

$$f(x) = 1$$

Tabel 4. 65 Data Vector Paslon 3

Dok	Teks	Label
D965	gaya beliau tenang sopan debat politik sehat	Positif
D937	Pks cabut dukung anies syaikhu ragu alas	Negatif

Selanjutnya menghitung nilai $\vec{X}_l$ yang ada pada setiap data, maka dilakukan perhitungan seperti sebagai berikut.

$$\vec{X}_0 = [0,756330 \ 0,581505 \ 0,986250 \ 0,887229 \ 0,986250 \ 0,482484 \ 0,448935]$$

$$\vec{X}_1 = [0,210004 \ 0,887229 \ 0,284442 \ 0,406186 \ 0,166739 \ 0,689187 \ 0,829305]$$

Operasi perhitungan vektor tersebut dijelaskan sebagai berikut.

$$\begin{aligned} \vec{X}_0 - \vec{X}_0 &= [0,756330 \ 0,581505 \ 0,986250 \ 0,887229 \ 0,986250 \ 0,482484 \ 0,448935] \\ &- [0,756330 \ 0,581505 \ 0,986250 \ 0,887229 \ 0,986250 \ 0,482484 \ 0,448935] \end{aligned}$$

$$= [0 \ 0 \ 0 \ 0 \ 0 \ 0 \ 0]$$

$$\begin{aligned} \vec{X}_0 - \vec{X}_1 &= [0,546326 \ -0,305724 \ 0,701808 \ 0,481043 \ 0,819511 \ -0,206703 \ - \\ &0,38037] \end{aligned}$$

$$\begin{aligned} \vec{X}_1 - \vec{X}_0 &= [-0,546326 \ 0,305724 \ -0,701808 \ -0,481043 \ -0,819511 \ 0,206703 \\ &0,38037] \end{aligned}$$

$$\vec{X}_1 - \vec{X}_1 = [0 \ 0 \ 0 \ 0 \ 0 \ 0 \ 0]$$

Setelah berhasil melakukan perhitungan vektor, langkah berikutnya adalah menghitung panjang vektor. Dalam konteks SVM, panjang vektor ini berperan penting untuk menentukan margin maksimum, yaitu jarak terpendek antara hyperplane dan titik-titik

data dari dua kelas yang berbeda. Adapun persamaan untuk menghitung panjang vektor dijelaskan sebagai berikut.

$$\|\vec{X}_0 - \vec{X}_0\| = \sqrt{0^2 + 0^2 + 0^2 + 0^2 + 0^2 + 0^2} = 0$$

$$\|\vec{X}_0 - \vec{X}_1\| =$$

$$\sqrt{0,546326^{2+} - 0,305724^{2+} 0,701808^{2+} 0,481043^{2+} 0,819511^{2+} - 0,206703^{2+} - 0,38037^2} \\ = 1,405$$

$$\|\vec{X}_1 - \vec{X}_0\| =$$

$$\sqrt{-0,546326^{2+} 0,305724^{2+} - 0,701808^{2+} - 0,481043^{2+} - 0,819511^{2+} 0,206703^{2+} 0,38037^2} \\ = 1,405$$

$$\|\vec{X}_1 - \vec{X}_1\| = \sqrt{0^2 + 0^2 + 0^2 + 0^2 + 0^2 + 0^2} = 0$$

Berikut perhitungan fungsi RBF menggunakan nilai gamma (γ) 0,001

$$k\|\vec{X}_0 - \vec{X}_0\| = EXP(-\gamma\|\vec{X}_0 - \vec{X}_0\|^2)$$

$$k\|\vec{X}_0 - \vec{X}_0\| = EXP(-0,001(0^2))$$

$$k\|\vec{X}_0 - \vec{X}_0\| = 1$$

Berikut adalah hasil dari perhitungan fungsi kernel RBF.

$$k\|\vec{X}_0 - \vec{X}_0\| = 1$$

$$k\|\vec{X}_0 - \vec{X}_0\| = 0,998$$

$$k\|\vec{X}_0 - \vec{X}_0\| = 0,998$$

$$k\|\vec{X}_0 - \vec{X}_0\| = 1$$

Hasil dari perhitungan dari fungsi kernel diatas $K(\vec{X}_i - \vec{X}_j)$, Selanjutnya digunakan sebagai variabel dalam proses penentuan nilai bias. Nilai-nilai kernel tersebut disusun dalam bentuk matriks, yang akan digunakan pada tahap perhitungan berikutnya. Berikut ini adalah matriks hasil perhitungannya.

$$K(\vec{X}_i - \vec{X}_j) = \begin{bmatrix} 1 & 0,998 \\ 0,998 & 1 \end{bmatrix}$$

Setelah memperoleh nilai matriks kernel, langkah selanjutnya adalah menghitung nilai bobot. Perhitungan bobot dilakukan dengan menggunakan nilai support vector (α) yang bernilai 0,775. Berikut adalah perhitungan untuk mencari nilai bobot.

$$\vec{w} = \sum_{i=1}^l a_i \cdot y_i \cdot \vec{w}_i \\ = (0,775 \cdot (1) \left[\frac{1}{0,998} \right] + (0,775 \cdot (-1) \left[\frac{0,998}{1} \right]) \\ = \left(\frac{0,0016}{-0,0016} \right)$$

selanjutnya yaitu menghitung nilai bias dengan menggunakan perhitungan sebagai berikut.

$$b = -\frac{1}{2}((\vec{w} \cdot \vec{x} - 1) + (\vec{w} \cdot \vec{x} + 1))$$

$$b = -\frac{1}{2} \left(\left(\begin{bmatrix} 0.0016 \\ -0.0016 \end{bmatrix} \cdot \begin{bmatrix} 0.998 \\ 1 \end{bmatrix} \right) + \left(\begin{bmatrix} 0.0016 \\ -0.0016 \end{bmatrix} \cdot \begin{bmatrix} 1 \\ 0.998 \end{bmatrix} \right) \right)$$

$$b = -\frac{1}{2} \left(\begin{bmatrix} 0 \\ 0 \end{bmatrix} \right) = 0$$

Setelah proses pelatihan menghasilkan nilai bobot dan bias, tahap selanjutnya adalah pengujian model. Pada fase ini, data uji yang telah dilabeli sebagai kelas positif dengan nilai 1 digunakan untuk mengevaluasi akurasi klasifikasi. Evaluasi dilakukan untuk memverifikasi apakah prediksi model sesuai dengan label asli. Seluruh data uji telah melewati tahap transformasi dan pembobotan. Dokumen pengujian disajikan pada bagian berikut.

Tabel 4. 66 Dokumen Testing Paslon 3

Dokumen	terus baik masyarakat ahmad syaikh ilham habibie
TF-IDF	0,619828 0,250893 0,385579 0,188607 0,166739 0,205098 0,216382

Data di atas digunakan sebagai data uji, adapun bobot data uji tersebut adalah sebagai berikut.

$$\vec{X} = [0,619828 \ 0,250893 \ 0,385579 \ 0,188607 \ 0,166739 \ 0,205098 \ 0,216382]$$

Langkah selanjutnya yaitu mencari nilai vector dengan perhitungan pengurangan $\vec{x}$ dengan nilai dari data training.

$$\vec{X} - \vec{X}_0 = [0,619828 \ 0,250893 \ 0,385579 \ 0,188607 \ 0,166739 \ 0,205098 \ 0,216382] - [0,756330 \ 0,581505 \ 0,986250 \ 0,887229 \ 0,986250 \ 0,482484 \ 0,448935]$$

$$\vec{X} - \vec{X}_0 = [-0,136502 \ -0,330612 \ -0,600671 \ -0,698622 \ -0,819511 \ -0,277386 \ -0,232553]$$

$$\vec{X} - \vec{X}_1 = [0,409824 \ -0,636336 \ 0,101137 \ -0,217579 \ 0 \ -0,484089 \ -0,612923]$$

Setelah nilai vektor diperoleh, langkah berikutnya adalah menghitung panjang vektor dari data $\vec{x}$.

Hasil dari perhitungan ini menghasilkan nilai panjang vektor sebagai berikut.

$$\|\vec{X} - \vec{X}_0\| =$$

$$\sqrt{-0,136502^2 + -0,330612^2 + -0,600671^2 + -0,698622^2 + -0,819511^2 + -0,277386^2 + -0,232553^2}$$

$$= 1,3339$$

$$\|\vec{X} - \vec{X}_1\| =$$

$$\sqrt{0,409824^2 + -0,636336^2 + 0,101137^2 + -0,217579^2 + 0^2 + -0,484089^2 + -0,612923^2}$$

$$= 1,1137$$

Berikut merupakan hasil perhitungan dari fungsi kernel RBF yang digunakan pada data uji.

$$\begin{aligned}
k||\vec{X} - \vec{X}_0|| &= EXP(-y||\vec{X} - \vec{X}_0||^2) \\
k||\vec{X} - \vec{X}_0|| &= EXP(-0,001(1,3339^2)) \\
k||\vec{X} - \vec{X}_0|| &= 0,998222
\end{aligned}$$

$$\begin{aligned}
k||\vec{X} - \vec{X}_1|| &= EXP(-y||\vec{X} - \vec{X}_1||^2) \\
k||\vec{X} - \vec{X}_1|| &= EXP(-0,001(1,1137^2)) \\
k||\vec{X} - \vec{X}_1|| &= 0,998760
\end{aligned}$$

Setelah fungsi kernel RBF selesai dihitung, langkah selanjutnya adalah melakukan klasifikasi terhadap data uji (testing) menggunakan fungsi hyperplane dalam metode Support Vector Machine (SVM). Dalam tahap ini, nilai kernel yang telah diperoleh digunakan dalam fungsi keputusan untuk menentukan label prediksi dari data uji. Melalui hasil perhitungan ini, dapat diketahui kelas atau kategori dari masing-masing data uji. Perhitungan lengkap disajikan pada bagian berikut.

$$\begin{aligned}
f(x) &= \text{sign}(\sum_{i=1, x_i = sv}^l a_i \cdot y_i \cdot K(\vec{x}, \vec{x}_i) + b) \\
f(x) &= \text{sign}(0,775 \cdot 0,998222 + 0) + (0,775 \cdot 0,998760 + 0) \\
f(x) &= 1
\end{aligned}$$

Tabel 4. 67 Data Vector Paslon 4

Dok	Teks	Label
D281	dedi mulyadi menang tps Prabowo coblos	Positif
D120	dedi mulyadi tunduk mu cihhh harap	Negatif

Selanjutnya menghitung nilai $\vec{X}_l$ yang ada pada setiap data, maka dilakukan perhitungan seperti sebagai berikut.

$$\begin{aligned}
\vec{X}_0 &= [0,020070 \ 0,036106 \ 0,331849 \ 0,710782 \ 0,643204 \ 0,699283] \\
\vec{X}_1 &= [0,020070 \ 0,036106 \ 1,150625 \ 0,919575 \ 1,150625 \ 0,736473]
\end{aligned}$$

Operasi perhitungan vektor tersebut dijelaskan sebagai berikut.

$$\begin{aligned}
\vec{X}_0 - \vec{X}_0 &= [0,020070 \ 0,036106 \ 0,331849 \ 0,710782 \ 0,643204 \ 0,699283] - \\
&[0,020070 \ 0,036106 \ 0,331849 \ 0,710782 \ 0,643204 \ 0,699283] \\
&= [0 \ 0 \ 0 \ 0 \ 0 \ 0] \\
\vec{X}_0 - \vec{X}_1 &= [0 \ 0 \ -0,818776 \ -0,208793 \ -0,507421 \ -0,03719]
\end{aligned}$$

$$\vec{X}_1 - \vec{X}_0 = [0 \ 0 \ 0,818776 \ 0,208793 \ 0,507421 \ 0,03719]$$

$$\vec{X}_1 - \vec{X}_1 = [0 \ 0 \ 0 \ 0 \ 0 \ 0]$$

Setelah berhasil melakukan perhitungan vektor, langkah berikutnya adalah menghitung panjang vektor. Dalam konteks SVM, panjang vektor ini berperan penting untuk menentukan margin maksimum, yaitu jarak terpendek antara hyperplane dan titik-titik data dari dua kelas yang berbeda. Adapun persamaan untuk menghitung panjang vektor dijelaskan sebagai berikut.

$$\|\vec{X}_0 - \vec{X}_0\| = \sqrt{0^2 + 0^2 + 0^2 + 0^2 + 0^2 + 0^2} = 0$$

$$\begin{aligned} \|\vec{X}_0 - \vec{X}_1\| &= \sqrt{0^2 + 0^2 + (-0,818776)^2 + (-0,208793)^2 + (-0,507421)^2 + (-0,03719)^2} \\ &= 0,986 \end{aligned}$$

$$\begin{aligned} \|\vec{X}_1 - \vec{X}_0\| &= \sqrt{0^2 + 0^2 + 0,818776^2 + 0,208793^2 + 0,507421^2 + 0,03719^2} \\ &= 0,986 \end{aligned}$$

$$\|\vec{X}_1 - \vec{X}_1\| = \sqrt{0^2 + 0^2 + 0^2 + 0^2 + 0^2 + 0^2} = 0$$

Berikut perhitungan fungsi RBF menggunakan nilai gamma (γ) 0,001

$$k\|\vec{X}_0 - \vec{X}_0\| = EXP(-\gamma\|\vec{X}_0 - \vec{X}_0\|^2)$$

$$k\|\vec{X}_0 - \vec{X}_0\| = EXP(-0,001(0^2))$$

$$k\|\vec{X}_0 - \vec{X}_0\| = 1$$

Berikut adalah hasil dari perhitungan fungsi kernel RBF.

$$k\|\vec{X}_0 - \vec{X}_0\| = 1$$

$$k\|\vec{X}_0 - \vec{X}_0\| = 0,999$$

$$k\|\vec{X}_0 - \vec{X}_0\| = 0,999$$

$$k\|\vec{X}_0 - \vec{X}_0\| = 1$$

Hasil dari perhitungan dari fungsi kernel diatas $K(\vec{X}_i - \vec{X}_j)$, Selanjutnya digunakan sebagai variabel dalam proses penentuan nilai bias. Nilai-nilai kernel tersebut disusun dalam bentuk matriks, yang akan digunakan pada tahap perhitungan berikutnya. Berikut ini adalah matriks hasil perhitungannya.

$$K(\vec{X}_i - \vec{X}_j) = \begin{bmatrix} 1 & 0,999 \\ 0,999 & 1 \end{bmatrix}$$

Setelah memperoleh nilai matriks kernel, langkah selanjutnya adalah menghitung nilai bobot. Perhitungan bobot dilakukan dengan menggunakan nilai support vector (α) yang bernilai 0,775. Berikut adalah perhitungan untuk mencari nilai bobot.

$$\begin{aligned}
\vec{w} &= \sum_{i=1}^l a_i \cdot y_i \cdot \vec{w}_i \\
&= (0,775 \cdot (1) \begin{bmatrix} 1 \\ 0,999 \end{bmatrix} + (0,775 \cdot (-1) \begin{bmatrix} 0,999 \\ 1 \end{bmatrix}) \\
&= \begin{pmatrix} 0,0008 \\ -0,0008 \end{pmatrix}
\end{aligned}$$

selanjutnya yaitu menghitung nilai bias dengan menggunakan perhitungan sebagai berikut.

$$\begin{aligned}
b &= -\frac{1}{2} ((\vec{w} \cdot \vec{x} - 1) + (\vec{w} \cdot \vec{x} + 1)) \\
b &= -\frac{1}{2} ((\begin{bmatrix} 0,0008 \\ -0,0008 \end{bmatrix} \cdot \begin{bmatrix} 0,999 \\ 1 \end{bmatrix} + (\begin{bmatrix} -0,0008 \\ 0,0008 \end{bmatrix} \cdot \begin{bmatrix} 1 \\ 0,999 \end{bmatrix})) \\
b &= -\frac{1}{2} (\begin{bmatrix} 0 \\ 0 \end{bmatrix}) = 0
\end{aligned}$$

Setelah proses pelatihan menghasilkan nilai bobot dan bias, tahap selanjutnya adalah pengujian model. Pada fase ini, data uji yang telah dilabeli sebagai kelas positif dengan nilai 1 digunakan untuk mengevaluasi akurasi klasifikasi. Evaluasi dilakukan untuk memverifikasi apakah prediksi model sesuai dengan label asli. Seluruh data uji telah melewati tahap transformasi dan pembobotan. Dokumen pengujian disajikan pada bagian berikut.

Tabel 4. 68 Dokumen Testing Paslon 4

Dokumen	Dedi mulyadi jadi purwakarta asri amp					
TF-IDF	0,020070	0,036106	0,882385	0,628042	1,150625	0,668896

Data di atas digunakan sebagai data uji, adapun bobot data uji tersebut adalah sebagai berikut.

$$\vec{X} = [0,020070 \ 0,036106 \ 0,882385 \ 0,628042 \ 1,150625 \ 0,668896]$$

Langkah selanjutnya yaitu mencari nilai vector dengan perhitungan pengurangan $\vec{x}$ dengan nilai dari data training.

$$\begin{aligned}
\vec{X} - \vec{X}_0 &= [0,020070 \ 0,036106 \ 0,882385 \ 0,628042 \ 1,150625 \ 0,668896] - \\
&[0,020070 \ 0,036106 \ 0,331849 \ 0,710782 \ 0,643204 \ 0,699283]
\end{aligned}$$

$$\vec{X} - \vec{X}_0 = [0 \ 0 \ 0,550536 \ -0,08274 \ 0,507421 \ -0,030387]$$

$$\vec{X} - \vec{X}_1 = [0 \ 0 \ -0,26824 \ -0,291533 \ 0 \ -0,067577]$$

Setelah nilai vektor diperoleh, langkah berikutnya adalah menghitung panjang vektor dari data $\vec{x}$. Hasil dari perhitungan ini menghasilkan nilai panjang vektor sebagai berikut.

$$\|\vec{X} - \vec{X}_0\| = \sqrt{0^{2+} \ 0^{2+} \ 0,550536^{2+} \ -0,08274^{2+} \ 0,507421^{2+} \ -0,030387^{2+}}$$

$$= 0,7538$$

$$\|\vec{X} - \vec{X}_1\| = \sqrt{0^2 + 0^2 + (-0,26824)^2 + (-0,291533)^2 + 0^2 + (-0,067577)^2}$$

$$= 0,4018$$

Berikut merupakan hasil perhitungan dari fungsi kernel RBF yang digunakan pada data uji.

$$k\|\vec{X} - \vec{X}_0\| = EXP(-\gamma\|\vec{X} - \vec{X}_0\|^2)$$

$$k\|\vec{X} - \vec{X}_0\| = EXP(-0,001(0,7538^2))$$

$$k\|\vec{X} - \vec{X}_0\| = 0,999432$$

$$k\|\vec{X} - \vec{X}_1\| = EXP(-\gamma\|\vec{X} - \vec{X}_1\|^2)$$

$$k\|\vec{X} - \vec{X}_1\| = EXP(-0,001(0,4018^2))$$

$$k\|\vec{X} - \vec{X}_1\| = 0,999839$$

Setelah fungsi kernel RBF selesai dihitung, langkah selanjutnya adalah melakukan klasifikasi terhadap data uji (testing) menggunakan fungsi hyperplane dalam metode Support Vector Machine (SVM). Dalam tahap ini, nilai kernel yang telah diperoleh digunakan dalam fungsi keputusan untuk menentukan label prediksi dari data uji. Melalui hasil perhitungan ini, dapat diketahui kelas atau kategori dari masing-masing data uji. Perhitungan lengkap disajikan pada bagian berikut.

$$f(x) = \text{sign}(\sum_{i=1, x_i = sv}^l a_i \cdot y_i \cdot K(\vec{x}, \vec{x}_i) + b)$$

$$f(x) = \text{sign}(0,775 \cdot 0,999432 + 0) + (0,775 \cdot 0,999839 + 0)$$

$$f(x) = 1$$

4.8 Confusion Matrix

Setelah proses klasifikasi dilakukan menggunakan metode *Support Vector Machine*, tahap selanjutnya adalah mengevaluasi kinerja model melalui confusion matrix. Dengan menggunakan confusion matrix ini, peneliti dapat mengukur sejauh mana model mampu mengklasifikasi data dengan benar, serta mengidentifikasi kelemahan model dalam membedakan kategori tertentu. Confusion Matrix ini mencakup perhitungan akurasi, presisi, recall, dan f1-score. Adapun hasil dari Confusion Matrix yaitu berupa nilai TP(*True Positive*), TN(*True Negatif*), FP(*False Positif*), FN(*False Negatif*). Berikut adalah perhitungan menggunakan data 80:20.

Tabel 4. 69 Confusion Matrix Paslon 1

	Prediksi	
	Positif	Negatif
Aktual	Positif 154 Negatif 3	6 31

Dari data tabel di atas dapat disimpulkan bahwa nilai confusion matrixnya yaitu :

$$TP(\text{True Positif}) = 154$$

$$TN(\text{True Negatif}) = 31$$

$$FP(\text{False Positif}) = 3$$

$$FN(\text{False Negatif}) = 6$$

Dengan data di atas maka dapat menghitung hasil dari *accuracy*, *precision*, *recall*, serta *F1 score*. Berikut adalah perhitungannya.

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

$$\text{Accuracy} = \frac{154+31}{154+3+31+6}$$

$$\text{Accuracy} = \frac{185}{194}$$

$$\text{Accuracy} = 0,95$$

$$\text{Precision Positif} = \frac{TP}{TP+FP}$$

$$\text{Precision Positif} = \frac{154}{154+3}$$

$$\text{Precision} = \frac{157}{160}$$

$$= 0,98$$

$$\text{Precision Negatif} = \frac{TN}{TN+FN}$$

$$\text{Precision Negatif} = \frac{31}{31+6}$$

$$\text{Precision} = \frac{31}{37}$$

$$= 0,83$$

$$\text{Recall Positif} = \frac{TP}{TP+FN}$$

$$\text{Recall Positif} = \frac{154}{154+6}$$

$$\text{Recall} = \frac{154}{160}$$

$$= 0,96$$

$$\text{Recall Negatif} = \frac{TN}{TN+FP}$$

$$\text{Recall Negatif} = \frac{31}{31+3}$$

$$\begin{aligned}\text{Recall} &= \frac{31}{34} \\ &= 0,91\end{aligned}$$

$$\text{F1 Score} = 2\left(\frac{\text{Recall} \times \text{Presisi}}{\text{Recall} + \text{Presisi}}\right)$$

$$\text{F1 Score positif} = 2\left(\frac{0,96 \times 0,98}{0,96 + 0,98}\right) = \frac{1,8816}{1,94} = 0,96$$

$$\text{F1 Score negatif} = 2\left(\frac{0,91 \times 0,83}{0,91 + 0,83}\right) = \frac{1,5106}{1,74} = 0,86$$

Tabel 4. 70 Confusion Matrix Paslon 2

		Prediksi	
		Positif	Negatif
Aktual	Positif	91	15
	Negatif	14	78

Dari data tabel di atas dapat disimpulkan bahwa nilai confusion matrixnya yaitu :

$$TP(\text{True Positif}) = 91$$

$$TN(\text{True Negatif}) = 78$$

$$FP(\text{False Positif}) = 14$$

$$FN(\text{False Negatif}) = 15$$

Dengan data di atas maka dapat menghitung hasil dari *accuracy*, *precision*, *recall*, serta *F1 score*. Berikut adalah perhitungannya.

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

$$\text{Accuracy} = \frac{91 + 78}{91 + 14 + 78 + 15}$$

$$\begin{aligned}\text{Accuracy} &= \frac{169}{198} \\ &= 0,85\end{aligned}$$

$$\text{Precision Positif} = \frac{TP}{TP+FP}$$

$$\text{Precision Positif} = \frac{91}{91+14}$$

$$\begin{aligned}\text{Precision} &= \frac{91}{105} \\ &= 0,86\end{aligned}$$

$$\text{Precision Negatif} = \frac{TN}{TN+FN}$$

$$\text{Precision Negatif} = \frac{78}{78+15}$$

$$\begin{aligned}\text{Precision Negatif} &= \frac{78}{93} \\ &= 0,83\end{aligned}$$

$$\text{Recall Positif} = \frac{TP}{TP+FN}$$

$$\text{Recall Positif} = \frac{91}{91+15}$$

$$\begin{aligned}\text{Recall} &= \frac{91}{106} \\ &= 0,85\end{aligned}$$

$$\text{Recall Negatif} = \frac{TN}{TN+FP}$$

$$\text{Recall Negatif} = \frac{78}{78+17}$$

$$\begin{aligned}\text{Recall} &= \frac{78}{92} \\ &= 0,84\end{aligned}$$

$$\text{F1 Score} = 2 \left(\frac{\text{Recall} \times \text{Presisi}}{\text{Recall} + \text{Presisi}} \right)$$

$$\text{F1 Score positif} = 2 \left(\frac{0,85 \times 0,86}{0,85 + 0,86} \right) = \frac{1,462}{1,71} = 0,85$$

$$\text{F1 Score negatif} = 2 \left(\frac{0,84 \times 0,83}{0,84 + 0,83} \right) = \frac{1,3944}{1,67} = 0,83$$

Tabel 4. 71 Confusion Matrix Paslon 3

		Prediksi	
		Positif	Negatif
Aktual	Positif	68	7
	Negatif	6	119

Dari data tabel di atas dapat disimpulkan bahwa nilai confusion matrixnya yaitu :

TP(True Positif) = 68

TN(True Negatif) = 119

FP(False Positif) = 6

FN(False Negatif) = 7

Dengan data di atas maka dapat menghitung hasil dari *accuracy*, *precision*, *recall*, serta

F1 score. Berikut adalah perhitungan nya.

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

$$\text{Accuracy} = \frac{68+119}{68+6+119+7}$$

$$\text{Accuracy} = \frac{187}{200}$$

$$= 0,935$$

$$\text{Precision Positif} = \frac{TP}{TP+FP}$$

$$\text{Precision} = \frac{68}{68+6}$$

$$\text{Precision} = \frac{68}{74}$$

$$= 0,91$$

$$\text{Precision Negatif} = \frac{TN}{TN+FN}$$

$$\text{Precision} = \frac{119}{119+7}$$

$$\text{Precision} = \frac{119}{126}$$

$$= 0,94$$

$$\text{Recall Positif} = \frac{TP}{TP+FN}$$

$$\text{Recall Positif} = \frac{68}{68+7}$$

$$\text{Recall} = \frac{68}{75}$$

$$= 0,90$$

$$\text{Recall Negatif} = \frac{TN}{TN+FP}$$

$$\text{Recall Negatif} = \frac{119}{119+6}$$

$$\text{Recall} = \frac{119}{125}$$

$$= 0,95$$

$$\text{F1 Score} = 2 \left(\frac{\text{Recall} \times \text{Presisi}}{\text{Recall} + \text{Presisi}} \right)$$

$$\text{F1 Score positif} = 2 \left(\frac{0,90 \times 0,91}{0,90 + 0,91} \right) = \frac{1,638}{1,81} = 0,90$$

$$\text{F1 Score negatif} = 2 \left(\frac{0,95 \times 0,94}{0,95 + 0,94} \right) = \frac{1,786}{1,89} = 0,94$$

Tabel 4. 72 Confusion Matrix Paslon 4

		Prediksi	
		Positif	Negatif
Aktual	Positif	114	16
	Negatif	35	35

Dari data tabel di atas dapat disimpulkan bahwa nilai confusion matrixnya yaitu :

$$TP(\text{True Positif}) = 114$$

$$TN(\text{True Negatif}) = 35$$

$$FP(\text{False Positif}) = 35$$

$$FN(\text{False Negatif}) = 16$$

Dengan data di atas maka dapat menghitung hasil dari *accuracy*, *precision*, *recall*, serta *F1 score*. Berikut adalah perhitungannya.

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

$$\text{Accuracy} = \frac{114+35}{114+35+35+16}$$

$$\text{Accuracy} = \frac{149}{200}$$

$$= 0,745$$

$$\text{Precision Positif} = \frac{TP}{TP+FP}$$

$$\text{Precision} = \frac{114}{114+35}$$

$$\text{Precision} = \frac{114}{149}$$

$$= 0,76$$

$$\text{Precision Negatif} = \frac{TN}{TN+FN}$$

$$\text{Precision} = \frac{35}{35+16}$$

$$\text{Precision} = \frac{35}{51}$$

$$= 0,68$$

$$\text{Recall Positif} = \frac{TP}{TP+FN}$$

$$\text{Recall Positif} = \frac{114}{114+16}$$

$$\text{Recall} = \frac{114}{130}$$

$$= 0,87$$

$$\text{Recall Negatif} = \frac{TN}{TN+FP}$$

$$\text{Recall Negatif} = \frac{35}{35+35}$$

$$\begin{aligned}\text{Recall} &= \frac{35}{70} \\ &= 0,5\end{aligned}$$

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

$$\text{F1 Score positif} = 2\left(\frac{0,87 \times 0,76}{0,87 + 0,76}\right) = \frac{1,3224}{1,63} = 0,81$$

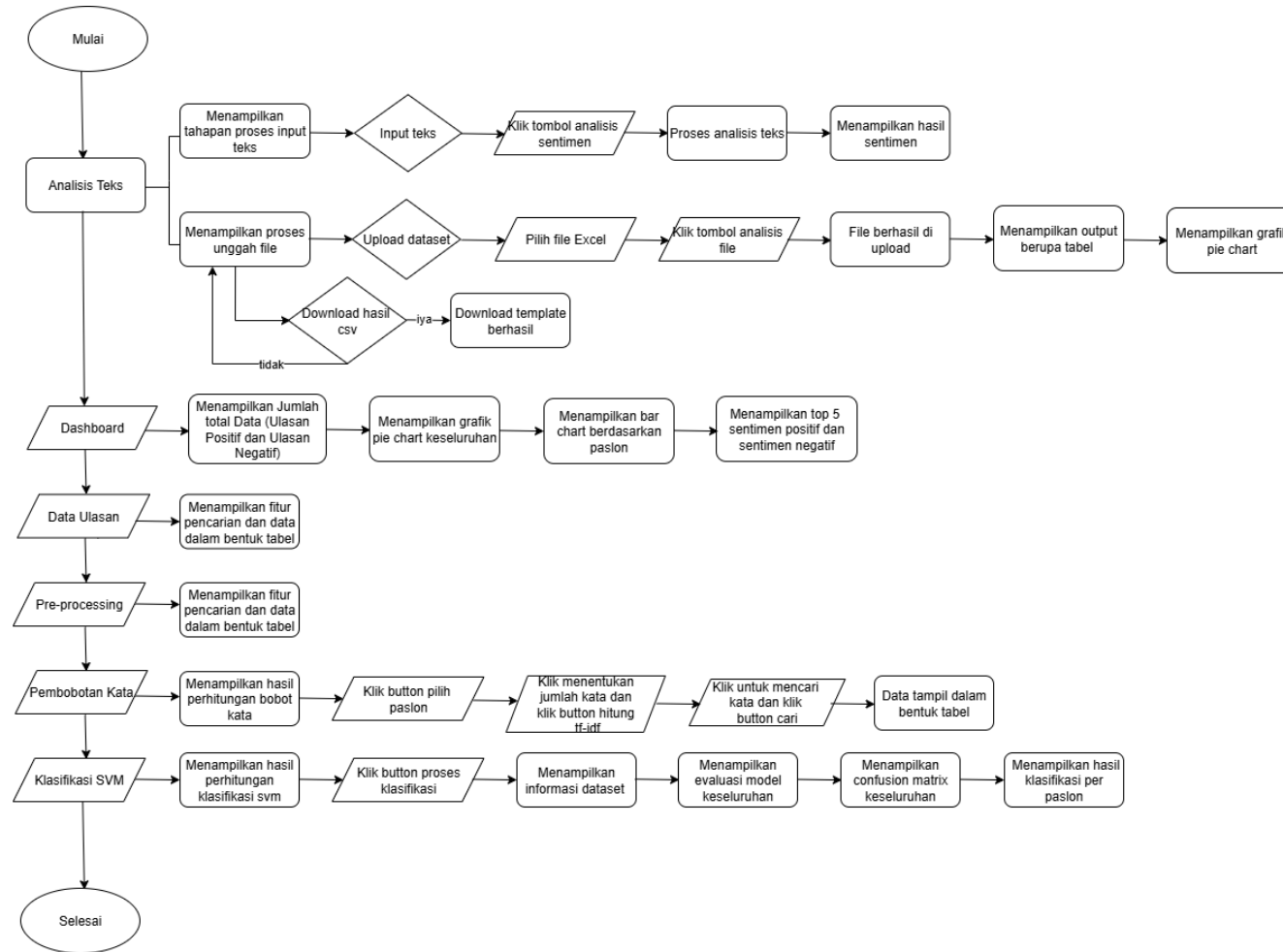
$$\text{F1 Score negatif} = 2\left(\frac{0,5 \times 0,68}{0,5 + 0,68}\right) = \frac{0,68}{1,18} = 0,57$$

4.9 Analisis Data

Analisis data dalam penelitian ini berfokus pada evaluasi hasil klasifikasi menggunakan metode Support Vector Machine (SVM), tingkat akurasi model tersebut, serta faktor-faktor lain yang memengaruhi hasil analisis sentimen terhadap program kampanye calon gubernur Jawa Barat. Selain itu, faktor-faktor seperti kualitas data, ukuran dataset, hasil pelabelan, proses preprocessing, dan pemilihan fitur pembobotan kata (TF-IDF) juga dapat memengaruhi hasil analisis. Evaluasi terhadap faktor-faktor tersebut penting untuk memastikan keandalan dan validitas hasil analisis sentimen yang diperoleh. Dengan demikian, analisis data ini tidak hanya mencakup klasifikasi dan akurasi metode SVM, tetapi juga melibatkan evaluasi terhadap faktor-faktor lain yang dapat memengaruhi hasil analisis sentimen terhadap opini masyarakat di media sosial.

4.10 Perancangan

4.10.1 Flowchart



Gambar 4. 4 Alur Sistem

4.10.2 Penjelasan Skenario

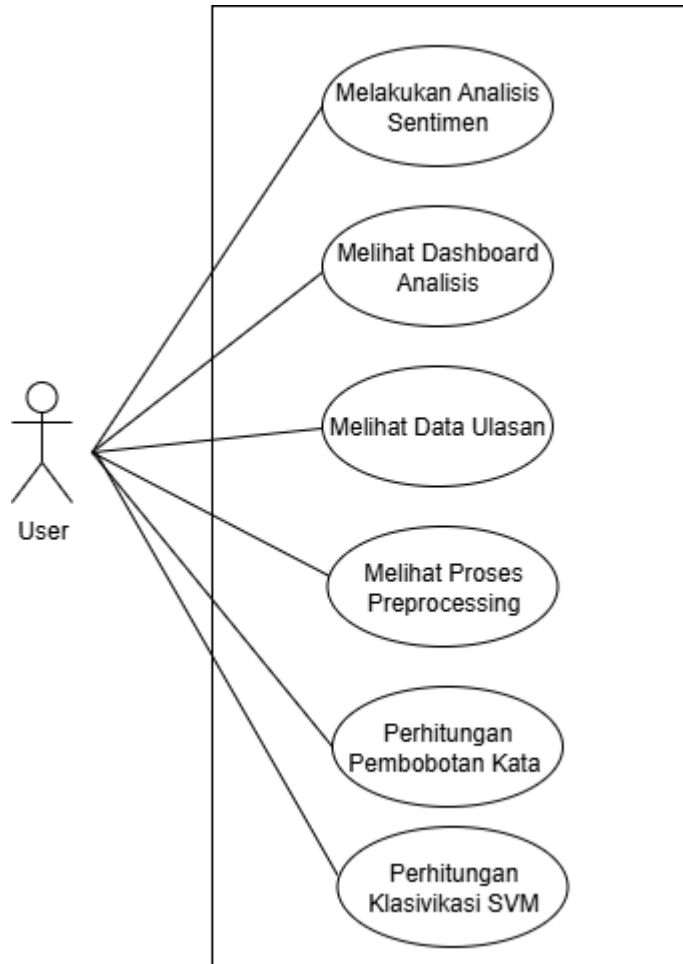
Pada saat user memulai aplikasi, sistem akan menampilkan halaman awal Analisis Teks. Di tahap ini, terdapat dua opsi: input teks langsung atau mengunggah dataset dalam format .csv. Jika memilih input teks, user dapat mengetik kalimat kemudian menekan tombol analisis sentimen untuk mendapatkan hasil sentimen. Jika memilih unggah dataset, user memilih file Excel, lalu sistem memproses dan menampilkan output berupa tabel, grafik pie chart, serta menyediakan opsi untuk mengunduh hasil dalam format .csv.

Selanjutnya, user diarahkan ke Dashboard yang menampilkan ringkasan jumlah total data ulasan positif dan negatif. Informasi ini divisualisasikan dalam grafik pie chart keseluruhan, bar chart berdasarkan pasangan calon, serta daftar top 5 sentimen positif dan negatif. Pada menu Data Ulasan, sistem menyajikan seluruh ulasan dalam bentuk tabel dengan fitur pencarian, sehingga memudahkan user untuk menelusuri isi data yang digunakan.

Tahap berikutnya adalah Pre-processing, di mana sistem secara otomatis menjalankan tahapan cleaning, case folding, tokenisasi, normalisasi, stopword removal, dan stemming. Hasil preprocessing ditampilkan kembali dalam bentuk tabel agar pengguna dapat memverifikasi transformasi data. Setelah itu, sistem masuk ke tahap Pembobotan Kata (TF-IDF). Pada tahap ini, teks yang sudah diproses diubah ke dalam bentuk vektor numerik. User dapat melihat hasil bobot kata pada tabel, memilih pasangan calon untuk dianalisis lebih detail, serta melakukan pencarian kata tertentu. Terakhir, pada tahap Klasifikasi SVM, user menekan tombol proses klasifikasi untuk menjalankan model. Sistem kemudian menampilkan hasil berupa informasi dataset, evaluasi model, confusion matrix, dan hasil klasifikasi sentimen per pasangan calon. Dengan demikian, alur aplikasi berjalan end-to-end mulai dari input data hingga visualisasi hasil analisis sentimen.

4.10.3 Use case Diagram

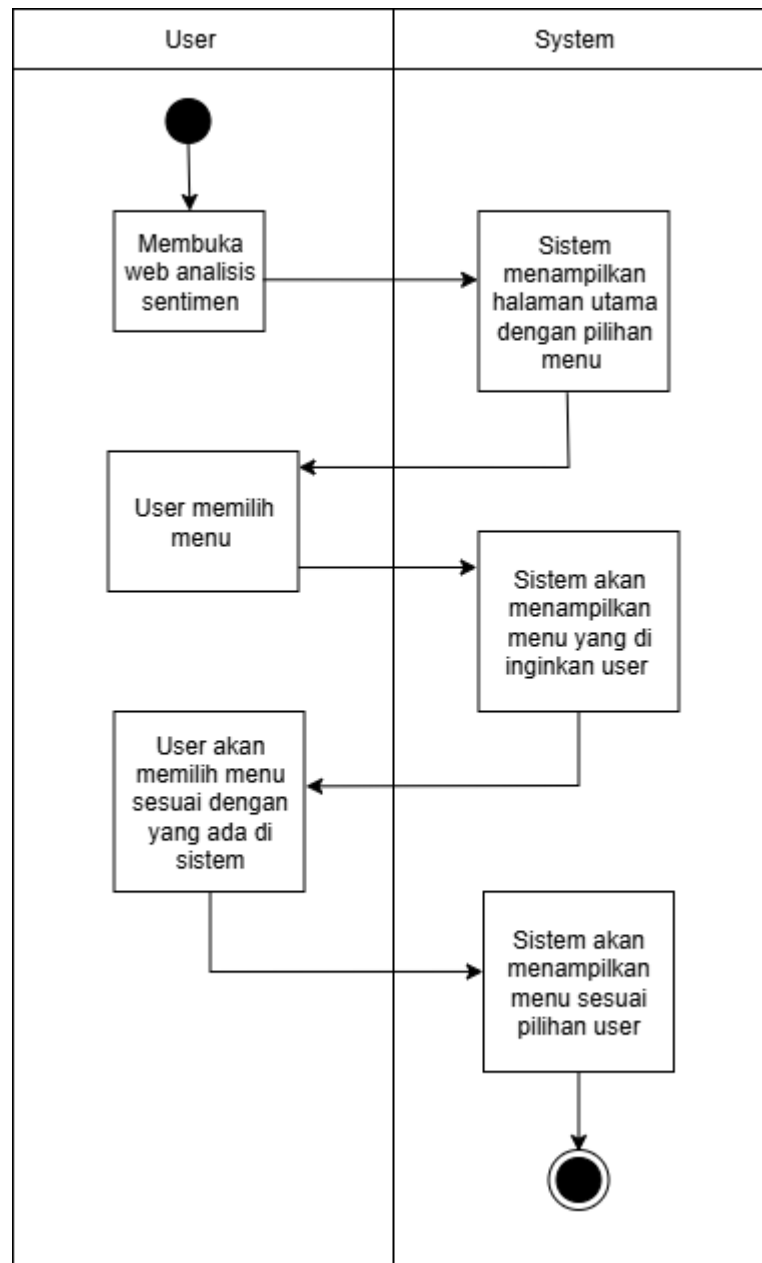
Use case diagram digunakan untuk menggambarkan hubungan antara aktor dengan sistem yang dibangun. Diagram ini memperlihatkan fitur-fitur utama yang dapat diakses oleh user pada sistem analisis sentimen yang dikembangkan. Adapun use case diagram dari sistem ini ditunjukkan pada Gambar 4.4 berikut.



Gambar 4. 5 Usecase Diagram

4.10.4 Activity Diagram

Activity diagram menggambarkan alur aktivitas yang terjadi antara user dengan system pada aplikasi analisis sentimen. Diagram ini memperlihatkan interaksi user mulai dari membuka web hingga sistem menampilkan menu sesuai dengan pilihan yang dipilih user. Dengan adanya activity diagram, alur proses yang dilakukan sistem dapat lebih mudah dipahami, serta menunjukkan peran antara user dan sistem dalam menjalankan proses analisis. Adapun activity diagram dari sistem ini ditunjukan pada gambar 4.5 berikut.

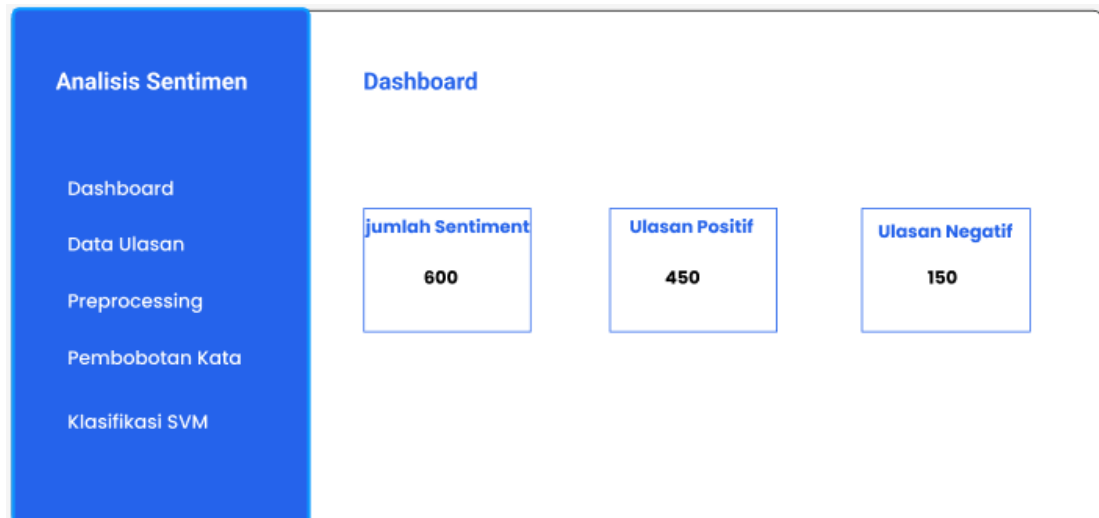


Gambar 4. 6 Activity Diagram

4.10.5 Mock Up

4.10.5.1 Dashboard

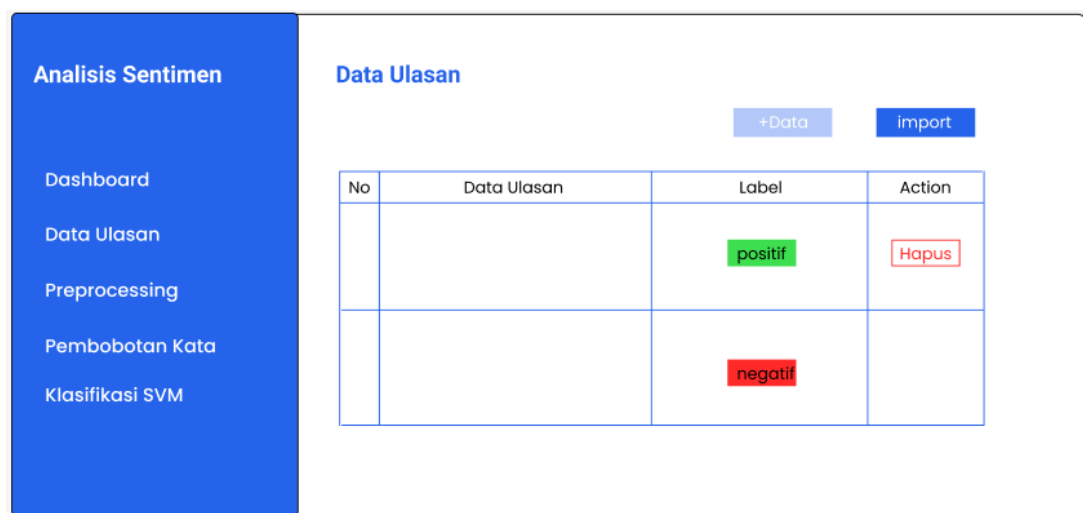
Pada halaman dashboard yaitu menampilkan jumlah data yang digunakan dalam penelitian ini secara keseluruhan. Berikut adalah tampilan antar muka untuk halaman dashboard.



Gambar 4. 7 Mockup Dashboard

4.10.5.2 Data Ulasan

Pada halaman data ulasan adalah data opini yang dari pengguna yang akan digunakan dalam penelitian yang diambil dari X. Dan data ini di import dari tabel, berikut adalah tampilan antar muka untuk data ulasan.

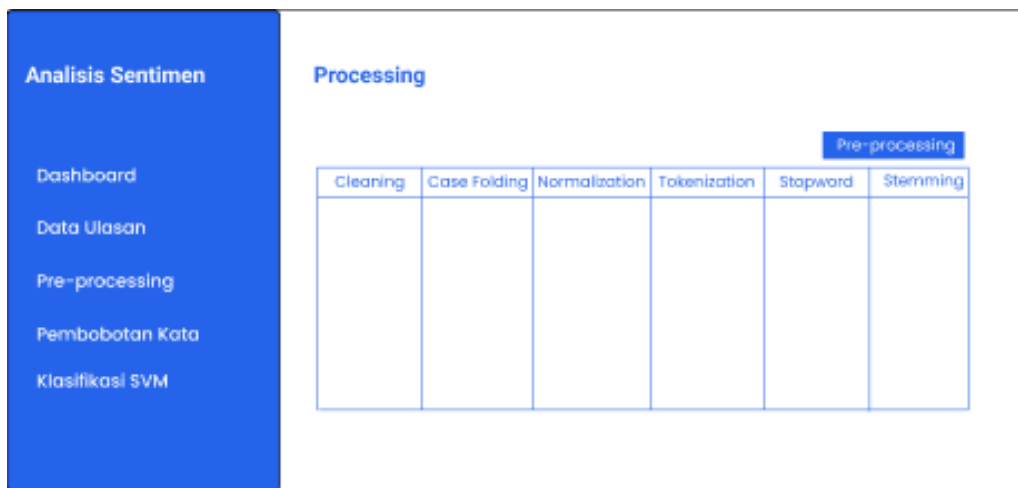


Gambar 4. 8 Mockup Data Ulasan

4.10.5.3 Preprocessing

Pada halaman pre-processing adalah untuk persiapan data sebelum dianalisis lebih lanjut. Dengan tahapan ini akan menampilkan data yang telah diolah dari proses cleaning, case folding, normalization, tokenization, stopword, dan stemming. Berikut

adalah tampilan antar muka untuk pre-processing.



Gambar 4. 9 Mockup Pre-Processing

4.10.5.4 Pembobotan Kata

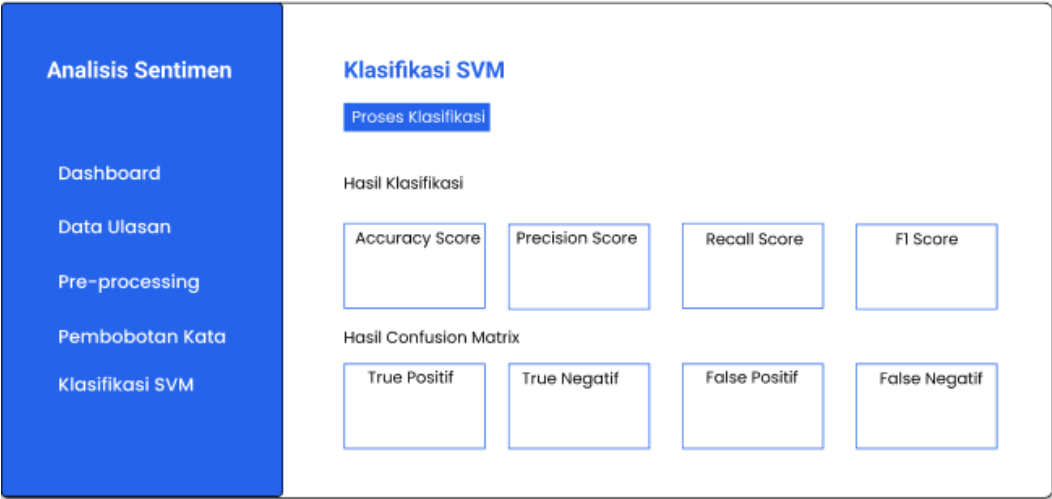
Pada halaman ini ditampilkan proses pembobotan kata menggunakan TF-IDF. Pembobotan dilakukan dengan menghitung nilai TF-IDF, yaitu kombinasi antara frekuensi kemunculan suatu kata dalam dokumen tertentu (TF) dan seberapa jarang kata tersebut muncul dalam seluruh korpus (IDF). Berikut adalah tampilan pada proses pembobotan kata.



Gambar 4. 10 Mockup Pembobotan

4.10.5.5 Klasifikasi SVM

Pada halaman klasifikasi ini, dilakukan pengklasifikasian data ulasan seperti accuracy score, precision score, recall score, dan f1 score. Selain itu ditampilkan hasil dari confusion matrix. Berikut adalah tampilan antar muka dari klasifikasi.



Gambar 4. 11 Mockup Klasifikasi

BAB 5

IMPLEMENTASI

Bab ini akan menjelaskan proses implementasi. Pada tahap ini terdiri dari proses pembangunan dan pengujian pada

5.1 Implementasi Sistem

Tahap implementasi sistem adalah tahap dimana system siap untuk dioperasikan, yang terdiri dari penjelasan mengenai lingkungan implementasi, dan implementasi program.

5.1.1 Kakas dan Kebutuhan perangkat lunak

Dalam melakukan pembangunan perangkat lunak dibutuhkan beberapa kakas dan kebutuhan perangkat lunak. Berikut merupakan kakas dan kebutuhan perangkat lunak yang digunakan :

a. Spesifikasi Perangkat keras

Tabel 5. 1 Spesifikasi Perangkat Keras

No	Nama Perangkat	Sistem Requerment
1	Processor	Intel® Core™ i7-5600 CPU
2	RAM	16384MB
3	Disk	256 GB
4	Operation System	Windows 10 Pro 64-bit

b. Perangkat lunak yang digunakan

Tabel 5. 2 Perangkat Lunak

No	Perangkat Lunak	Keterangan
1	Bahasa Pemrograman	HTML CSS, Java Script, Flask
2	Code Editor	Visual Studio Code

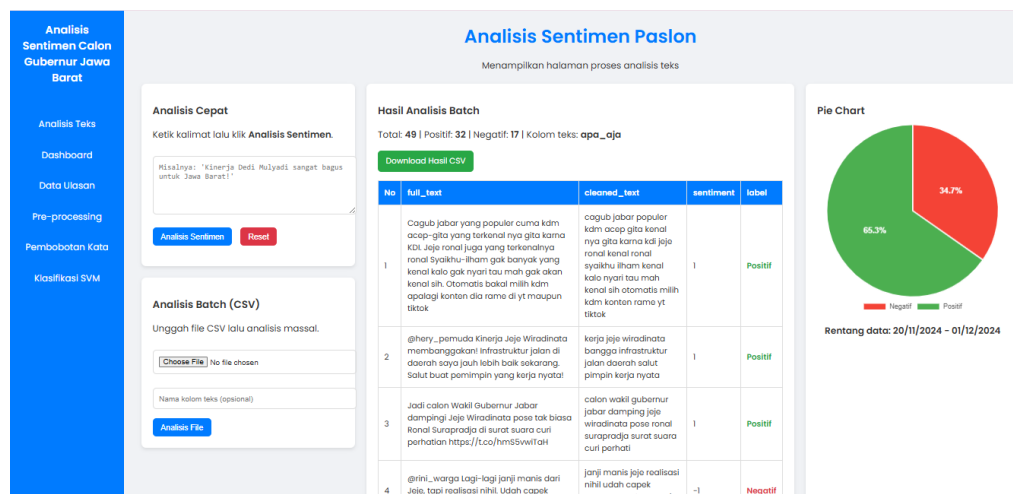
5.1.1.1 Implementasi Tampilan Web

Implementasi antarmuka dilakukan dengan setiap tampilan program yang dibangun dan pengkodeannya dalam bentuk file program.

5.1.1.2 Tampilan Analisis Teks

Pada implementasi fitur analisis teks terdapat beberapa button dan input yang berfungsi untuk memproses analisis. Pada bagian Analisis Cepat, tersedia text area untuk mengetikkan kalimat opini, kemudian button Analisis Sentimen berfungsi menjalankan proses analisis sentimen, sedangkan button Reset digunakan untuk menghapus teks input agar dapat diganti dengan kalimat baru. Pada bagian Analisis Batch (CSV), button Choose File

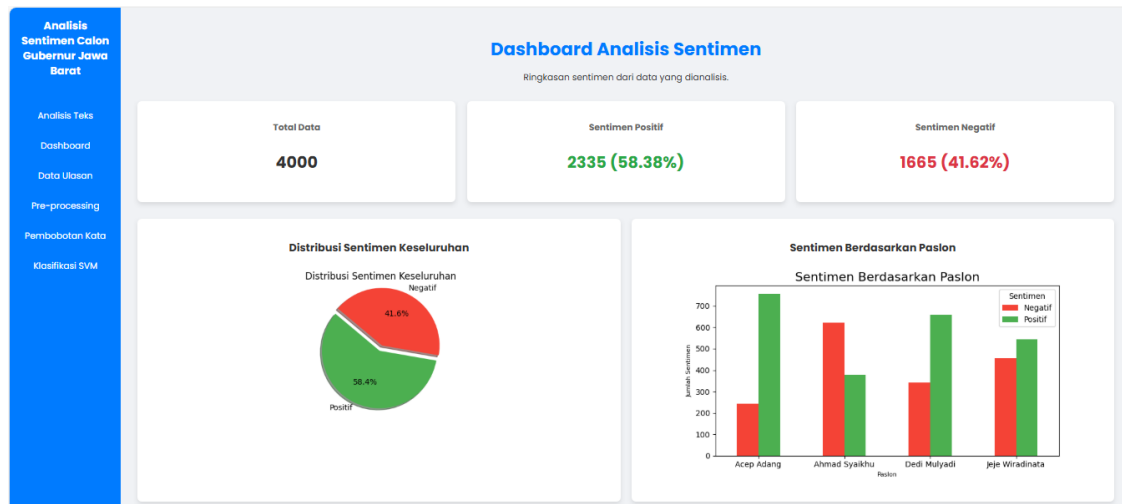
digunakan untuk memilih file dengan format `.csv` dari penyimpanan lokal, kemudian button Analisis File berfungsi memproses seluruh data secara massal, sedangkan field opsional disediakan untuk menuliskan nama kolom teks pada file CSV yang ingin dianalisis. Hasil analisis batch ditampilkan di sisi kanan dalam bentuk tabel berisi nomor, teks asli *full_text*, teks hasil preprocessing stemming, hasil klasifikasi sentimen, serta label yang dihasilkan, dan juga terdapat pie chart yang menampilkan visualisasi distribusi sentimen positif dan negatif agar hasil analisis lebih mudah dipahami.



Gambar 5.1 Website-Analisis Teks

5.1.1.3 Tampilan Dashboard

Pada tampilan dashboard yaitu menampilkan ringkasan hasil analisis sentimen dari keseluruhan data yang digunakan. menampilkan total jumlah data yang dianalisis beserta pembagian sentimen positif dan negatif lengkap dengan persentasenya. Selanjutnya, hasil analisis divisualisasikan dalam bentuk diagram pie yang menunjukkan distribusi sentimen keseluruhan, serta grafik batang yang memperlihatkan jumlah sentimen positif dan negatif berdasarkan masing-masing calon gubernur Jawa Barat. Pada bagian bawah yaitu menampilkan top 5 sentimen positif dan top 5 sentimen negatif untuk memberikan gambaran ulasan. Berikut adalah implementasi dari halaman dashboard.



Gambar 5. 2 Website-Dashboard

5.1.1.4 Tampilan Data Ulasan

Pada tampilan data ulasan, sistem menampilkan daftar ulasan yang sudah melalui proses klasifikasi sentimen dan dikelompokkan berdasarkan masing-masing paslon. Tabel berisi nomor, teks ulasan, dan kategori sentimen positif dan negatif. Serta kolom pencarian dapat digunakan untuk memfilter ulasan berdasarkan kata kunci tertentu sehingga mempermudah dalam menelusuri data untuk setiap pasangan calon.

Analisis Sentimen Calon Gubernur Jawa Barat

Data Ulasan per Paslon
Menampilkan data sentimen dari setiap paslon secara terpisah.

Acep Adang

Cari ulasan untuk Acep Adang...

No.	Teks	Sentimen
1	@jabar...peduli Debat tadi malam Acep Adang malah salah sebut lawan debat, bikin nggak fokus nontonnya	Negatif
2	@Telkomsel Signal ga bener di patinggen 1. Desa Karangpawitan kec Padaherang Pangandaran Jabar.	Negatif
3	@Telkomsel Signal lup lep ga bener. Di dusun patinggen 1. Desa Karangpawitan. kec Padaherang kab Pangandaran Jabar. Tolong perbaiki	Negatif
4	@ekonomi...oke Acep janjiin listrik gratis, tapi nggak ada rincian biayanya. Jangan cuma omong doang!	Positif
5	@anti...nepot Tim sukses Acep-Adang isinya keluarga semua. Emang mau jadi dinasti politik?	Negatif
6	@pemilih...muda Pas debat, Adang malah ngelantur bahas masa lalu terus. Topik sekarang kok nggak dibahas?	Negatif
7	Katanya anti korupsi, tapi dapet dana dari partai yang bosnya koruptor. Hipokrit banget siht	Negatif
8	Acep janjiin listrik gratis tapi pas ditanya sumber dananya malah beralasan 'hanti dibahas'. Klasik	Negatif
9	27 November nanti jangan lupa coblos Acep - Gita ya! Kita semua berharap Jabar bisa menjadi lebih maju dan lebih bahagia. Yuk kita sukseskan bersama! https://t.co/BpK5AhK3	Positif
10	27 November ayo kita pilih Acep - Gita! Mari kita bersatu membangun Jabar yang sejahtera dan BAHAGIA! Cek info lanjut di [link] https://t.co/8u7XbpKISF	Positif
11	27 November coblos Acep - Gita! Avok kita ramalkan Jabar vana makmur dan penuh kebahagiaan! Mari kita beraandanaan tanaan membanaun Jabar vana lebih	Positif

Gambar 5. 3 Website-Data Ulasan

5.1.1.5 Tampilan Pre-processing

Pada tampilan pre-processing yaitu menampilkan hasil pengolahan teks ulasan sesuai tahapan yang telah ditentukan. Data asli dari ulasan ditampilkan pada kolom Full_text, kemudian secara bertahap diolah melalui proses cleaning, case folding, normalization, tokenization, stopword,. Hasil dari setiap tahap ditampilkan pada tabel sehingga dapat dilihat perbedaan teks sebelum dan sesudah

preprocessing. Halaman ini berfungsi untuk memastikan bahwa teks ulasan sudah siap digunakan pada tahap pembobotan kata dan klasifikasi.

Analisis Sentimen Calon Gubernur Jawa Barat

Analisis Teks

Dashboard

Data Ulasan

Pre-processing

Pembobotan Kata

Klasifikasi SVM

Hasil Preprocessing per Paslon

Menampilkan hasil preprocessing untuk setiap paslon.

Acep Adang

Cari di tabel Acep Adang...

No.	Full Text	Sentimen	Cleaning	Case Folding	Normalized	Tokenization	Stopword Removal	Stemming
1	@jabar_peduli Debat tadi malam Acep Adang malah salah sebut lawan debat, bikin nggak fokus nontonnya	Negatif	Debat tadi malam Acep Adang malah salah sebut lawan debat, bikin nggak fokus nontonnya	debat tadi malam acep adang malah salah sebut lawan debat, bikin nggak fokus nontonnya	[debat, tadi, malam, acep, adang, malah, salah, sebut, lawan, debat, bikin, tidak, fokus, nontonnya]	[debat, tadi, malam, acep, adang, malah, salah, sebut, lawan, debat, bikin, tidak, fokus, nontonnya]	[debat, malam, acep, adang, salah, lawan, debat, bikin, fokus, nontonnya]	debat malam acep adang salah lawan debat bikin fokus nontonnya
2	@telkomsel Signal ga bener di patinggen 1 Desa Karangpawitan kec Padaherang Pangandaran Jabar.	Negatif	Signal ga bener di patinggen 1 Desa Karangpawitan kec Padaherang Pangandaran Jabar.	signal ga bener di patinggen 1 desa karangpawitan kec padaherang pangandaran jabar.	[signal, tidak, benar, di, patinggen, desa, karangpawitan, kec, padaherang, pangandaran, jabar]	[signal, tidak, benar, di, patinggen, desa, karangpawitan, kec, padaherang, pangandaran, jabar]	[signal, patinggen, desa, karangpawitan, kec, padaherang, pangandaran, jabar]	signal patinggen desa karangpawitan kec padaherang pangandaran jabar
3	@telkomsel Signal lup lep ga bener. Di dusun patinggen 1 Desa Karangpawitan, kec Padaherang kab Pangandaran Jabar. Tolong perbaiki	Negatif	Signal lup lep ga bener. Di dusun patinggen 1 Desa Karangpawitan, kec Padaherang kab Pangandaran Jabar. Tolong perbaiki	signal lup lep ga bener. di dusun patinggen 1 desa karangpawitan, kec padaherang kab pangandaran jabar. tolong perbaiki	[signal, lup, lep, tidak, benar, di, dusun, patinggen, desa, karangpawitan, kec, padaherang, kab, pangandaran, jabar, tolong, perbaiki]	[signal, lup, lep, tidak, benar, di, dusun, patinggen, desa, karangpawitan, kec, padaherang, kab, pangandaran, jabar, tolong, perbaiki]	[signal, lup, lep, dusun, patinggen, desa, karangpawitan, kec, padaherang, kab, pangandaran, jabar, tolong, perbaiki]	signal lup lep dusun patinggen desa karangpawitan kec padaherang kab pangandaran jabar tolong baik
4	@ekonomi_oke Acep janjin listrik gratis, tapi nggak ada rincian biayanya. Jangan cuma omong doang!	Positif	Acep janjin listrik gratis, tapi nggak ada rincian biayanya. Jangan cuma omong doang!	acep janjin listrik gratis, tapi nggak ada rincian biayanya. jangan cuma omong doang!	[acep, janjin, listrik, gratis, tapi, tidak, ada, rincian, biayanya, jangan, cuma, omong, doang]	[acep, janjin, listrik, gratis, tapi, tidak, ada, rincian, biayanya, jangan, cuma, omong, doang]	[acep, janjin, listrik, gratis, rincian, biayanya, omong, doang]	acep janjin listrik gratis rincian biaya omong doang

Gambar 5. 4 Website-Pre-processing

5.1.1.6 Tampilan Pembobotan Kata

Pada tampilan Pembobotan Kata TF-IDF yaitu menghitung dan menampilkan bobot kata menggunakan metode TF-IDF dari dataset yang telah melalui proses preprocessing. Pengguna dapat memilih paslon yang akan dianalisis serta menekan tombol *Hitung TF-IDF* untuk menampilkan hasil pembobotan dalam tabel yang berisi daftar kata beserta nilai bobotnya. Selain itu, tersedia kolom pencarian yang memungkinkan pengguna menelusuri bobot dari kata tertentu, misalnya kata “jabar”, sehingga memudahkan dalam menemukan kata spesifik. Dropdown digunakan untuk menampilkan data dengan jumlah tertentu, misal menampilkan 100 atau 200 data.

Analisis Sentimen Calon Gubernur Jawa Barat

Analisis Teks

Dashboard

Data Ulasan

Pre-processing

Pembobotan Kata

Klasifikasi SVM

Pembobotan Kata (TF-IDF)

Pilih paslon dan hitung pembobotan kata menggunakan metode TF-IDF dari dataset hasil preprocessing.

Pilih Paslon: dedi_mulyadi

Jumlah Kata: 100

Hitung TF-IDF

Cari kata:

Cari

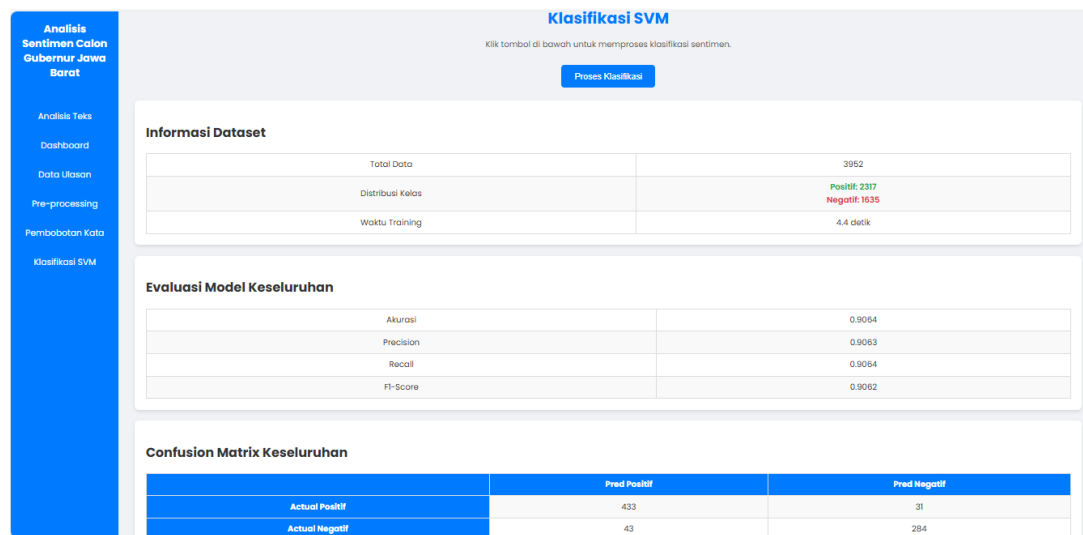
Hasil Pembobotan (Top 100 Kata)

Kata	Skor TF-IDF
dedi	167.7229
mulyadi	166.4497
jabar	81.4562
kang	55.4931
menang	55.4665
pilih	47.7272
kdm	43.4961
jawa	42.2531
gubernur	41.8967
barat	40.8783
pimpin	39.7297
ya	37.8473
masyarakat	36.2971
arena	33.7583

Gambar 5. 5 Website-Pembobotan Kata

5.1.1.7 Tampilan Klasifikasi SVM

Pada tampilan klasifikasi svm, user dapat melakukan proses klasifikasi sentimen menggunakan metode svm dengan menekan tombol proses klasifikasi. Setelah di proses, sistem akan menampilkan informasi dataset berupa jumlah data ulasan, distribusi kelas sentimen positif dan sentimen negatif, serta lama waktu training model. Selain itu menampilkan hasil evaluasi keseluruhan medel meliputi nilai akurasi, precision, recall, dan F1-score. Hasil ini dilengkapi dengan confusion matrix untuk menunjukkan jumlah prediksi benar maupun salah pada setiap kelas sentimen. User juga dapat melihat hasil klasifikasi berdasarkan masing-masing paslon melalui dropdown pilihan. Setelah paslon dipilih, sistem menampilkan evaluasi performa klasifikasi khusus untuk paslon tersebut, termasuk total data, metrik evaluasi, distribusi prediksi sentimen positif dan negatif, serta confusion matrix yang lebih detail.



Gambar 5. 6 Website-Klasifikasi SVM

5.2 Tahap Pengujian

Pengujian merupakan tahapan yang sangat penting dalam pembangunan sebuah web. Pengujian dilakukan untuk memastikan aplikasi web yang telah dibuat dapat berjalan dengan baik sesuai fungsinya. Dengan pengujian ini juga diharapkan agar web yang dibuat dapat diketahui kekurangan dan kelebihanannya.

Rencana pengujian web analisis sentimen cagub jabar menggunakan data uji berupa data masukan, yaitu sebagai berikut :

Tabel 5. 3 Rencana Pengujian

Item	Deskripsi	Pengujian
Halaman Analisis Teks	Pengujian fungsi tombol pada analisis teks	Black Box
Halaman Dashboard	Menampilkan ringkasan data hasil analisis	Black Box
Halaman Data Ulasan	Menampilkan halaman data ulasan	Black Box
Halaman Preprocessing	Menampilkan halaman preprocessing	
Halaman Pembobotan Kata	Menampilkan hasil pembobotan kata dengan metode TF-IDF	Black Box
Halaman Klasifikasi SVM	Menampilkan hasil klasifikasi sentimen menggunakan metode SVM	Black Box

Tabel 5. 4 Kasus dan Hasil Uji

Data Masukan	Kasus dan Hasil Uji		
	Data yang di harapkan	Hasil Uji	Kesimpulan
Halaman Analisis Teks	Pengujian fungsi tombol pada analisis teks	Semuaa tombol berfungsi dengan baik, yaitu : • Analisis menampilkan hasil analisis • Reset mengosongkan input teks • Analisis file menampilkan tabel serta grafik pie chart dari data CSV	Berhasil
Halaman Dashboard	Menampilkan ringkasan data hasil analisis	menampilkan ringkasan data ulasan dan grafik pie chart	Berhasil
Halaman Data Ulasan	Menampilkan halaman data ulasan	Tabel data ulasan berhasil di tampilkan	Berhasil
Halaman Preprocessing	Menampilkan halaman preprocessing	Semuaa hasil preprocessing berhasil di tampilkan sesuai alur	Berhasil
Halaman Pembobotan Kata	Menampilkan hasil pembobotan kata dengan metode TF-IDF	Tabel bobot kata berhasil di tampilkan	Berhasil
Halaman Klasifikasi SVM	Menampilkan hasil klasifikasi sentimen menggunakan metode SVM	Hasil klasifikasi sentimen berhasil di tampilkan	Berhasil

5.3 Evaluasi

Tahap evaluasi merupakan bagian krusial dalam penyelesaian Tugas Akhir. Pada tahap ini, seluruh hasil penelitian, analisis data, serta kesimpulan yang telah diperoleh akan dinilai secara komprehensif. Tujuan utama evaluasi adalah untuk memastikan bahwa Tugas Akhir memenuhi standar akademik yang berlaku serta memberikan kontribusi yang berarti dalam bidang analisis sentimen.

Hasil dari evaluasi ini diperoleh setelah melakukan pengujian sistem web Analisis Sentimen Program Kampanye Calon Gubernur Jawa Barat menggunakan metode SVM. Dari hasil uji coba, pengguna dapat memasukkan opini atau komentar terkait paslon, kemudian sistem secara otomatis mengklasifikasikan opini tersebut menjadi sentimen positif maupun negatif. Uji coba juga menunjukkan bahwa sistem mampu memberikan hasil klasifikasi dengan tingkat akurasi yang cukup baik sesuai data latih yang digunakan.

Selain itu, dilakukan evaluasi terhadap pengalaman pengguna saat mencoba fitur web. Sebagian besar pengguna dapat dengan mudah memahami cara kerja sistem, baik pada fitur analisis cepat (input teks langsung) maupun pada analisis batch menggunakan file CSV.

BAB 6

KESIMPULAN

Bab ini menjelaskan mengenai kesimpulan yang didapatkan dari hasil penelitian yang dilakukan serta saran yang diberikan untuk penelitian selanjutnya.

6.1 Kesimpulan

Berdasarkan penelitian yang telah dilakukan oleh penulis, yaitu penelitian tentang analisis sentimen program kampanye calon gubernur Jawa Barat menggunakan metode *Support Vector Machine*, maka dapat disimpulkan bahwa:

1. Penelitian ini berhasil memenuhi tujuan untuk memprediksi popularitas calon gubernur berdasarkan opini publik melalui klasifikasi sentimen berbasis machine learning. Hal ini ditunjukkan dari hasil pengambilan data menggunakan metode crawling dengan kata kunci nama calon gubernur, yang menghasilkan 4000 tweet (3952 data setelah pembersihan), terdiri dari 2317 sentimen positif dan 1635 sentimen negatif.
2. Penelitian ini juga berhasil merancang hyperparameter serta preprocessing yang sesuai untuk dataset sentimen media sosial. Proses preprocessing dilakukan mulai dari pembersihan data, normalisasi teks, hingga pembobotan TF-IDF, sementara pemilihan kernel RBF pada metode SVM mampu meningkatkan kinerja klasifikasi.
3. Implementasi Support Vector Machine dalam analisis sentimen program kampanye calon gubernur Jawa Barat berhasil dilakukan melalui serangkaian tahapan yang telah dirancang. Hasil pengujian menunjukkan bahwa metode SVM mampu mengklasifikasikan opini masyarakat ke dalam kategori positif dan negatif dengan performa yang baik, ditunjukkan melalui metrik evaluasi berupa Akurasi, Presisi, Recall, dan F1-Score. Dengan demikian, tujuan untuk mengetahui dan menganalisis opini masyarakat serta menghitung akurasi hasil klasifikasi dapat tercapai.
4. Penelitian ini berhasil membangun aplikasi berbasis web yang dapat digunakan untuk melakukan analisis sentimen secara langsung, sehingga memudahkan pemetaan opini masyarakat terhadap program kampanye calon gubernur Jawa Barat.
5. Pola kata yang dikenali oleh mesin dalam proses klasifikasi menunjukkan adanya keterbatasan. Mesin cenderung lebih mudah mengenali kata-kata yang secara eksplisit bernuansa positif atau negatif (misalnya *bagus*, *terbaik*, *buruk*, *gagal*), namun masih kesulitan dalam memahami konteks kalimat yang bersifat ambigu, mengandung sarkasme, atau menggunakan bahasa campuran. Hal ini menjadi keterbatasan yang perlu diperhatikan untuk pengembangan penelitian lebih lanjut.

6.2 Saran

Berdasarkan hasil penelitian yang telah dilakukan, diharapkan penelitian dapat menjadi acuan untuk penelitian lebih lanjut. Oleh karena itu, penulis memberikan usulan pengembangan sebagai berikut:

1. Perluasan dataset penelitian ini masih menggunakan jumlah data yang terbatas. Untuk memperoleh lebih representative dan akurat, disarankan menggunakan dataset yang lebih besar dan bervariasi.
2. Penggunaan metode klasifikasi lain selain *Support Vector Machine (SVM)*, metode lain seperti Naïve Bayes, Random Forest, atau model berbasis deep learning seperti (LSTM) dapat digunakan untuk dibandingkan performanya dalam klasifikasi sentimen.
3. Menggunakan kamus kata baku yang lebih lengkap dan relevan untuk preprocessing pada tahap normalisasi data untuk meningkatkan kualitas data sebelum di klasifikasi.
4. Pengembangan aplikasi lebih lanjut aplikasi berbasis web yang telah dikembangkan masih bersifat sederhana. Pengembangan lebih lanjut seperti integrasi API untuk pengambilan data secara real-time dari media sosial atau platform ulasan dapat membuat aplikasi lebih dinamis dan bermanfaat.

DAFTAR PUSTAKA

- [ARV21] G. N. Arviana. Sentiment Analysis, Teknik untuk Pahami Maksud di Balik Opini Pelanggan. 1 Februari 2021.
- [AZH24] Nur Fadhilah Az-haari, Didi Juardi, Asep Jamaludin. Analisis Sentimen Terhadap Boikot Brand Pro-Israel Pada Twitter Menggunakan Metode Naïve Bayes. *JATI (Jurnal Mahasiswa Teknik Informatika)*, 8(3), 4256–4261. 2024.
- [DOS22] Doloksaribu, H. P., & Samuel, Y. T. Komparasi Algoritma Data Mining untuk Analisis Sentimen Aplikasi PeduliLindungi. *Jurnal Teknologi Informasi: Jurnal Keilmuan dan Aplikasi Bidang Teknik Informatika*, 2022.
- [FAJ23] Muhamad Fajri, Aji Primajaya. Komparasi Teknik Hyperparameter Optimization Pada SVM Untuk Permasalahan Klasifikasi Dengan Menggunakan Grid Search Dan Random Search. *Journal of Applied Informatics and Computing*, 7(1), 14–19. 2023.
- [FEB24] Febby Wilyani, Qonaah Nuryan Arif, Fitri Aslimar. Pengenalan Dasar Pemrograman Python Dengan Google Colaboratory. *Jurnal Pelayanan Dan Pengabdian Masyarakat Indonesia*, 3(1), 08–14. 2024.
- [GIF22] Okta Ihza Gifari, Muh. Adha, Fernandito Freddy, Fernandito Freddy Setlight Durrand. Analisis Sentimen Review Film Menggunakan TF-IDF Dan Support Vector Machine. *Journal of Information Technology*, 2(1), 36–40. 2022.
- [HER20] Nuraeni Herlinawati, Yuri Yuliani, Siti Faizah, Windu Gata, Samudi Samudi. Analisis Sentimen Zoom Cloud Meetings Di Play Store Menggunakan Naïve Bayes Dan Support Vector Machine. *CESS (Journal of Computer Engineering, System and Science)*, 5(2), 293. 2020.
- [HUS21] Hendry Cipta Husada, Adi Suryaputra Paramita. Analisis Sentimen Pada Maskapai Penerbangan Di Platform Twitter Menggunakan Algoritma Support Vector Machine (SVM). *Teknika*, 10(1), 18–26. 2021.
- [KAN23] KANTINIT. Confusion Matrix: Pengertian, Cara Kerja Dan Contoh Soal. *KantinIT*, 2023.
- [MAR24] Devina Margarita, Hendra Maulana, Eka Prakarsa Mandyartha. Klasifikasi Penyakit Daun Padi Menggunakan Support Vector Machine Berdasarkan Fitur Mendalam (Deep Feature). *JUPI (Jurnal Ilmiah Penelitian Dan Pembelajaran Informatika)*, 9(4), 2256–70. 2024.
- [NAT20] Amar P. Natasuwarna. Seleksi Fitur Support Vector Machine Pada Analisis Sentimen Keberlanjutan Pembelajaran Daring. *Techno.Com*, 19(4), 437–48. 2020.
- [FRI24] Fried Sinlae, Laudza K Kalmamy, Ruly Setiaji, dan M Syahrul. *Menjelajahi Dunia Web: Panduan Pemula Untuk Pemrograman Web*. *Jurnal Siber Multi Disiplin*, 2(2), 107–118. 2024. <https://doi.org/10.38035/jsmd.v2i2.170>

- [PRA22] Prasetyo, Sofyan Mufti, Muhammad Ivan Prayo Nugroho, Riris Lima Putri, dan Opa Fauzi. *Pembahasan Mengenai Front-End Web Developer Dalam Ruang Lingkup Web Development*. Jurnal Multidisiplin Ilmu, 1(6), 1015–1020. 2022. <https://journal.mediapublikasi.id/index.php/bullet>
- [NOR21] Dwi Normawati, Surya Allit Prayogi. Implementasi Naïve Bayes Classifier Dan Confusion Matrix Pada Analisis Sentimen Berbasis Teks Pada Twitter. Jurnal Sains Komputer & Informatika (J-SAKTI), 5(2), 697–711. 2021.
- [NUG11] Anto Satriyo Nugroho, Arief Budi Witarto, Dwi Handoko. Support Vector Machine. *Proceedings of the 2011 Chinese Control and Decision Conference, CCDC*. 2011, 842–847.
- [PRA19] Kartika Prameswari, Erwin Budi Setiawan. Analisis Kepribadian Melalui Twitter Menggunakan Metode Logistic Regression Dengan Pembobotan TF-IDF Dan AHP. *E-Proceeding of Engineering*, 6(2), 9667–82. 2019.
- [REG23] Regina Lo, Alfred Edbert Yunanto, Rasya Nuhaifa Movia, Lambert Aditama Soehardjianto, Ferdinand Wangsa, Natalie Asdyo Lidjaja, Rahmi Yulia Ningsih. Penggunaan Bahasa Pemrograman Python Dalam Menganalisis Hubungan Kualitas Kopi Dengan Lokasi Pertanian Kopi. *Jurnal Publikasi Teknik Informatika*, 2(2), 100–109. 2023.
- [RIZ21] Muhammad Fitra Rizki, Karina Auliasari, Renaldi Primaswara Prasetya. Analisis Sentiment Cyberbullying Pada Sosial Media Twitter Menggunakan Metode Support Vector Machine. JATI (Jurnal Mahasiswa Teknik Informatika), 5(2), 548–56. 2021.
- [SAP22] Irwansyah Saputra, D. A. K. Machine Learning untuk Pemula. Informatika Bandung, 2022.
- [SHA21] Jafar Shadiq, Ahmad Safei, Rayhan Wahyudin Ratu Loly. Pengujian Aplikasi Peminjaman Kendaraan Operasional Kantor Menggunakan BlackBox Testing. *Information Management for Educators and Professionals*, 5(2), 97–110. 2021.
- [SIT20] Christ Memory Sitorus, Adhi Rizal, Mohamad Jajuli. Prediksi Risiko Perjalanan Transportasi Online Dari Data Telematik Menggunakan Algoritma Support Vector Machine. Jurnal Teknik Informatika Dan Sistem Informasi, 6(2), 254–65. 2020.
- [SUS20] Imelya Susianti, Sri Soerya Ningsih, M. Al Haris, Tiani Wahyu Utami. *Analisis Sentimen Pada Twitter Terkait New Normal Dengan Metode Naïve Bayes Classifier*. *Prosiding Seminar Edusainstech FMIPA UNIMUS*, 354–63. 2020.
- [SUR21] Suryadi. Analisis Sentimen Review Hotel Menggunakan Algoritma Naïve Bayes Classifier. TECHSI - Jurnal Teknik Informatika, 13(2), 95. 2021.
- [WID21] Widi Widayat. Analisis Sentimen Movie Review Menggunakan Word2Vec Dan

- Metode LSTM Deep Learning. *Jurnal Media Informatika Budidarma*, 5(3), 1018. 2021.
- [WIL23] Nicholas Wilyanto, Januar Firnando, Billy Franko, Surya Pratama Tanzil, Handy Christiano Tan, Ery Hartati. Pembuatan Website Menggunakan Visual Studio Code Di SMA Xaverius 3 Palembang. *Fordicate*, 3(1), 1–8. 2023.
- [WIN23] Wilson Winarto, Izmy Alwiah Musdar, Hasniati Hasniati. Analisis Sentimen Bakal Calon Presiden 2024 Menggunakan Algoritma Support Vector Machine Pada Twitter. *KHARISMA Tech*, 19(1), 86–98. 2023.
- [GFD22] Gifari, O. I., Muh. Adha, Fernandito Freddy, dan Fernandito Freddy Setlight Durrand. *Analisis Sentimen Review Film Menggunakan TF-IDF dan Support Vector Machine*. *Journal of Information Technology*, 2(1), 36–40, 2022. DOI: <https://doi.org/10.46229/jifotech.v2i1.330>

LAMPIRAN

Lampiran 1 Dataset Paslon 1

No	full_text	sentiment
1	@jabar_peduli Debat tadi malam Acep Adang malah salah sebut lawan debat, bikin nggak fokus nontonnya	-1
2	@Telkomsel Signal ga bener di patinggen 1. Desa Karangpawitan kec Padaherang Pangandaran jabar.	-1
3	@Telkomsel Signal lup lep ga bener. Di dusun patinggen 1. Desa Karangpawitan. kec Padaherang kab Pangandaran Jabar. Tolong perbaiki	-1
...	...	...
1000	@prabowo @LawrenceWongST Endorse cagub skrg pak? Kirain beda ternyata lebih parah.	-1

Lampiran 2 Dataset Paslon 2

No	full_text	sentiment
1	Cagub jabar yang populer cuma kdm acep-gita yang terkenal nya gita karna KDI. Jeje ronal juga yang terkenalnya ronal Syaikhul- ilham gak banyak yang kenal kalo gak nyari tau mah gak akan kenal sih. Otomatis bakal milih kdm apalagi konten dia rame di yt maupun tiktok	-1
2	@hery_pemuda Kinerja Jeje Wiradinata membanggakan! Infrastruktur jalan di daerah saya jauh lebih baik sekarang. Salut buat pemimpin yang kerja nyata!	1
3	Jadi calon Wakil Gubernur Jabar dampingi Jeje Wiradinata pose tak biasa Ronal Surapradja di surat suara curi perhatian https://t.co/hmS5vwiTaH	1
...	...	...
1000	@maya Banyak spanduk kampanye di fasilitas umum tak dicopot meski sudah lewat masa pemilu. Mengganggu dan jorok	-1

Lampiran 3 Dataset Paslon 3

No	full_text	sentiment
1	@AjiTeguh04 @aindraku @choliqnur @dapitnih @DenaKubang81 @frans_surya @syaikhu_ahmad Uas di Sumut mengendors Edy hasilnya????? Semoga sadar diri	-1
2	@Mata_Netizen62 @nongandah @GunRomli @PaltiWest2024 @ahmad_bellamy @Anak__Ogi @AleenaDi09 @AntoniusCDN @Cintada16 @denni_sauya @Dharma_tc saya melihatnya sbg teguran utk partai relijeus saat dititipi amanah utk capres A kok kmdn bersekutu dg lawannya hany	-1
3	@syaikhu_ahmad Akhirnya menjadi satu kolam juga dengan psi dalam genggamannya raja Jawa.	1
...	...	...
1000	@diskusimilenial Tapi ia tetap menunjukkan konsistensi nilai kejujuran dan kesederhanaan	1

Lampiran 4 Dataset Paslon 4

No	full_text	sentiment
1	Trimakasih dari WARGA JABAR ..setelah terpilihnya KANG DEDI MULYADI maka kebudayaan SUNDA turut bangkit di ALAM JAWA BARAT . Kita dukung KDM 2 Periode #TikTok https://t.co/FgvKYr2tdC	1
2	udh kebiasaan begitu dr dulu sjk era Jokowi anggaran begituan dikurang2i mknya banyak pejabat yg gak suka sm Jkw karna anggaran begituan gak sebesar dulu memperbaiki mental korup pejabat ternyata gak cukup 10 tahun bro gak cukup sumpah mi apapun cek Dedi Mulyadi soal beginian	1
3	@mhuseinali Kagak. Dedi Mulyadi beragama islam kak. DM sangat peduli ke budaya sunda jd kadang muncul anggapan macam-macam. Banyak artikel berita yang beritain dia sholat dan umroh. https://t.co/HRLLKWyye4	1
...	...	...
1000	#PREDIKSI PEMENANG PILKADA GUBERNUR 2024 1. DK Jakarta (3) #Pramono_Anung Rano Karno 2. Jabar (4) #Dedi_Mulyadi Erwan 3.	1

	Banten (1) #Airin_Rachmi_D Ade S 4. Jateng (1) #Andika_Perkasa Hendrar P 5. Jatim (2) #Khoififah_Indar_P Emil Dardak 6. Sumut (2) #Edy_Rahmayadi Hasan B f??<?&	
--	---	--

```

import re
import string
import nltk

# Fungsi untuk menghapus URL
def remove_URL(tweet):
    if tweet is not None and isinstance(tweet, str):
        url = re.compile(r'https?://\S+|www\.\S+')
        return url.sub(r'', tweet)
    else:
        return tweet

# Fungsi untuk menghapus HTML
def remove_html(tweet):
    if tweet is not None and isinstance(tweet, str):
        html = re.compile(r'<.*?>')
        return html.sub(r'', tweet)
    else:
        return tweet

# Fungsi untuk menghapus emoji (opsional, bisa diaktifkan)
def remove_emoji(tweet):
    if tweet is not None and isinstance(tweet, str):
        emoji_pattern = re.compile("[
            u\"\\U0001F600-\\U0001F64F"
            u\"\\U0001F300-\\U0001F5FF"
            u\"\\U0001F680-\\U0001F6FF"
            u\"\\U0001F700-\\U0001F77F"
            u\"\\U0001F780-\\U0001F7FF"
            u\"\\U0001F800-\\U0001F8FF"
            u\"\\U0001F900-\\U0001F9FF"
            u\"\\U0001FA00-\\U0001FA6F"
            u\"\\U0001FA70-\\U0001FAFF"
            u\"\\U0001F004-\\U0001F0CF"
            u\"\\U0001F1E0-\\U0001F1FF"
            "]" +, flags=re.UNICODE)
        return emoji_pattern.sub(r'', tweet)
    else:
        return tweet

# Fungsi untuk menghapus simbol
def remove_symbols(tweet):
    tweet = re.sub(r'^a-zA-Z0-9\s', '', tweet)

```

```

        return tweet

# Fungsi untuk menghapus angka
def remove_numbers(tweet):
    tweet = re.sub(r'\d+', '', tweet)
    return tweet

# Fungsi untuk menghapus username
def remove_username(text):
    return re.sub(r'@\w+\s*', '', text).strip()

# ===== Preprocessing =====
df['cleaning'] = df['full_text'].apply(lambda x: remove_URL(x))
df['cleaning'] = df['cleaning'].apply(lambda x: remove_html(x))
df['cleaning'] = df['cleaning'].apply(lambda x: remove_numbers(x))
df['cleaning'] = df['cleaning'].apply(lambda x: remove_username(x))
df['cleaning'] = df['cleaning'].apply(lambda x: remove_symbols(x))

# ===== Menampilkan hasil =====
df.head(500)

```

Lampiran 5 Kode Proses Cleansing

```

# Fungsi untuk melakukan case folding
def case_folding(text):
    if isinstance(text, str):
        lowercase_text = text.lower()
        return lowercase_text
    else:
        return text

# Fungsi case folding pada kolom cleaning
df['case_folding'] = df['cleaning'].apply(case_folding)

# ===== Menampilkan hasil =====
df.head(500)

```

Lampiran 6 Kode Proses Case Folding

```

def replace_taboo_words(text, kamus_tidak_baku):
    if isinstance(text, str):

        # Pisahkan kalimat menjadi kata per kata
        words = text.split()

        # Menyimpan kalimat hasil penggantian
        hasil_kalimat = []

```

```

# Menyimpan kata baku yang dipakai
kata_baku = []

# Menyimpan kata tidak baku yang ditemukan
kata_tidak_baku = []

# Menyimpan hash unik dari kata tidak baku
kata_tidak_baku_hash = []

for word in words:

    # Jika kata ada di kamus kata tidak baku
    if word in kamus_tidak_baku:

        # Ambil kata baku dari kamus
        baku_word = kamus_tidak_baku[word]
        if isinstance(baku_word, str) and all(char.isalpha() or
char.isspace() for char in baku_word):

            # Masukkan kata baku ke hasil kalimat
            hasil_kalimat.append(baku_word)

            # Simpan daftar kata baku
            kata_baku.append(baku_word)

            # Simpan daftar kata tidak baku
            kata_tidak_baku.append(word)

            # Simpan hash kata tidak baku
            kata_tidak_baku_hash.append(hash(word))
        else:
            # jika kata baku tidak valid, tetap gunakan kata asli
            hasil_kalimat.append(word)
    else:
        hasil_kalimat.append(word)

    replaced_text = " ".join(hasil_kalimat)
    return replaced_text, kata_baku, kata_tidak_baku,
kata_tidak_baku_hash
else:
    # Jika input bukan string, kembalikan nilai kosong
    return "", [], [], []

# Baca dataset
data = pd.DataFrame(df[['full_text ', 'sentiment', 'sentiment_polarity',
'cleaning', 'case_folding']])

# ===== Menampilkan hasil =====

```

```

data.head(500)

#Baca kamus tidak baku
kamus_data = pd.read_excel("kamuskatabaku.xlsx")
kamus_tidak_baku =dict(zip(kamus_data['tidak_baku'],
kamus_data['kata_baku']))

#Terapkan fungsi penggantian kata tidak baku
data['hasil_normalisasi'], data['Kata_Baku'], data['Kata_Tidak_Baku'],
data['Kata_Tidak_Baku_Hash'] = zip(*data['case_folding'].apply(lambda x:
replace_taboo_words(x, kamus_tidak_baku)))

# Membuat dataframe
df = pd.DataFrame(data[['full_text ', 'sentiment', 'sentiment_polarity',
'cleaning', 'case_folding', 'hasil_normalisasi']])

# ===== Menampilkan hasil =====
df.head(500)

```

Lampiran 7 Kode Proses Normalization

```

# Fungsi untuk memecah teks menjadi token (berdasarkan spasi)
def tokenize(text):

# Memisahkan teks menjadi list kata
tokens = text.split()
    return tokens

# Menerapkan fungsi tokenize pada kolom 'hasil_normalisasi'
df['tokenize'] = df['hasil_normalisasi'].apply(tokenize)

# ===== Menampilkan hasil =====
df.head(500)

```

Lampiran 8 Kode Proses Tokenization

```

# Import daftar stopwords dari NLTK
from nltk.corpus import stopwords
nltk.download('stopwords')

# Mengambil stopwords bahasa Indonesia
stop_words = stopwords.words('indonesian')

# Fungsi untuk menghapus stopwords dari tek
def remove_stopwords(text):
    return [word for word in text if word not in stop_words]

```

```
# Menerapkan fungsi remove_stopwords pada kolom 'tokenize'
# Hasilnya disimpan pada kolom baru 'stopword removal'
df['stopword removal'] = df['tokenize'].apply(lambda x:
remove_stopwords(x))

# ===== Menampilkan hasil =====
df.head(500)
```

Lampiran 9 Kode Proses Stopword

```
# Install library Sastrawi untuk stemming Bahasa Indonesia
!pip install Sastrawi

# Import library stemming dari Sastrawi dan NLTK
from Sastrawi.Stemmer.StemmerFactory import StemmerFactory
from nltk.stem import PorterStemmer
from nltk.stem.snowball import SnowballStemmer

# Membuat objek stemmer menggunakan Sastrawi
factory = StemmerFactory()
stemmer = factory.create_stemmer()

# Fungsi untuk melakukan stemming pada teks
def stem_text(text):
    return [stemmer.stem(word) for word in text]

# Menerapkan stemming pada kolom 'stopword removal'
# Hasil stemming disimpan pada kolom baru 'stemming_data'
df['stemming_data'] = df['stopword removal'].apply(lambda x: stem_text(x))

# ===== Menampilkan hasil =====
df.head(1000)
```

Lampiran 10 Kode Proses Stemming

```
import pandas as pd
from sklearn.model_selection import train_test_split
from sklearn.feature_extraction.text import TfidfVectorizer
from sklearn.svm import LinearSVC
from sklearn.metrics import classification_report, accuracy_score,
confusion_matrix
import seaborn as sns
import matplotlib.pyplot as plt

# Load dan bersihkan data
df = pd.read_excel('hasil_dedi.xlsx')
```

```

df = df.dropna(subset=['steming_data ', 'sentiment '])
df['sentiment '] = df['sentiment '].str.strip().str.lower()

# =====
# 1. SPLIT DATA (DULUAN)
# =====
X = df['steming_data ']
y = df['sentiment ']

# Split data menjadi train dan test
X_train_text, X_test_text, y_train, y_test = train_test_split(
    X, y, test_size=0.2, random_state=42
)

# Tampilkan info hasil split
print("Jumlah data total:", len(df))
print("Jumlah data training:", len(X_train_text))
print("Jumlah data testing :", len(X_test_text))

# print(y_train.head())

```

Lampiran 11 Kode Proses Split Data

```

import pandas as pd
import math
from collections import Counter

# ===== 1. Kalimat Uji =====
uji_texts = [
    "gembira dedi mulyadi anti baalawi statemennya beliau suka habib habib",
    "tau suku sunda istilah islam sunda sunda islam saking kental islam masuk budaya sunda kang dedi mulyadi terima jawa barat",
    "malam tonton podcast dedi mulyadi gue tuju ngakunya era digital sppd numpuk acara hotel",
    "tarik lihat depan sepak terjang dedi mulyadi jawa barat cari solusi intoleransi deradikalisme",
]

# ===== 2. Tokenisasi =====
tokenized_docs = [text.lower().split() for text in uji_texts]
N = len(tokenized_docs)

# ===== 3. TF Relatif =====
tf_list = []
for doc in tokenized_docs:
    counts = Counter(doc)
    total = len(doc)
    tf = {term: count / total for term, count in counts.items()}
    tf_list.append(tf)

```

```

# ===== 4. DF (Document Frequency) =====
df_counts = {}
for doc in tokenized_docs:
    for term in set(doc):
        df_counts[term] = df_counts.get(term, 0) + 1

# ===== 5. IDF =====
# Handle potential division by zero if a term doesn't appear in any
# document (though set(doc) should prevent this)
idf = {term: math.log(N / (df_counts[term] if df_counts[term] > 0 else
1)) for term in df_counts}

# ===== 6. TF-IDF =====
tfidf_list = []
for tf in tf_list:
    tfidf = {term: tf[term] * idf.get(term, 0) for term in tf} # Use get
    # with default 0 for terms not in idf
    tfidf_list.append(tfidf)

# ===== 7. TF Count (Jumlah Absolut) =====
tf_count_list = []
for doc in tokenized_docs:
    tf_count_list.append(Counter(doc))

# ===== 8. Buat DataFrame =====
all_terms = sorted(df_counts.keys())

# A. TF Relatif
tf_df = pd.DataFrame([{'term': tf.get(term, 0) for term in all_terms} for
tf in tf_list])
tf_df.insert(0, 'Dokumen', [f'Uji{i+1}' for i in range(N)])

# B. IDF
idf_df = pd.DataFrame({
    'Term': all_terms,
    'DF': [df_counts[t] for t in all_terms],
    'IDF': [idf[t] for t in all_terms]
})

# C. TF-IDF
tfidf_df = pd.DataFrame([{'term': tfidf.get(term, 0) for term in all_terms}
for tfidf in tfidf_list])
tfidf_df.insert(0, 'Dokumen', [f'Uji{i+1}' for i in range(N)])

# D. TF Count
tfc_df = pd.DataFrame([{'term': tfc.get(term, 0) for term in all_terms}
for tfc in tf_count_list])

```

```

tfcount_df.insert(0, 'Dokumen', [f'Uji{i+1}' for i in range(N)])

# ===== 9. Tampilkan Hasil di Terminal =====
# Print only a sample or summary as the full matrices are too large
print("\n=== Term Frequency Relatif (TF) (Sample) ===")
print(tf_df.to_string(index=False))

print("\n=== Document Frequency & IDF ===")
print(idf_df.to_string(index=False))

print("\n=== TF-IDF (Sample) ===")
print(tfidf_df.to_string(index=False))

print("\n=== Term Frequency Count (Jumlah Kemunculan Kata) (Sample) ===")
print(tfc_df.to_string(index=False))

# ===== 10. Simpan ke Excel =====
with pd.ExcelWriter('tfidf_dedi.xlsx', engine='openpyxl') as writer:
    tf_df.to_excel(writer, sheet_name='TF', index=False)
    idf_df.to_excel(writer, sheet_name='DF_IDF', index=False)
    tfidf_df.to_excel(writer, sheet_name='TF-IDF', index=False)
    tfcount_df.to_excel(writer, sheet_name='TF_Count', index=False)

print("\n✅ File 'tfidf_dedi.xlsx' berhasil disimpan.")

```

Lampiran 12 Kode Proses TF-IDF

```

from sklearn.svm import SVC
from sklearn.metrics import classification_report, confusion_matrix,
accuracy_score
import seaborn as sns
import matplotlib.pyplot as plt

# Inisialisasi TF-IDF Vectorizer
vectorizer = TfidfVectorizer()

# Fit dan transform data training (mengubah teks → vektor numerik)
X_train_vectorized = vectorizer.fit_transform(X_train_text)

# Transform data testing (pakai model yang sama dari training)
X_test_vectorized = vectorizer.transform(X_test_text)

# Tampilkan matriks hasil vektorisasi untuk data training
print("Matriks Vektorisasi untuk Data Training:")
print(X_train_vectorized.toarray())

# =====
# # TRAINING MODEL SVM DENGAN KERNEL RBF

```

```

# =====
svm = SVC(kernel="rbf", decision_function_shape='ovo')
svm.fit(X_train_vectorized.toarray(), y_train.values)

# Prediksi label sentimen pada data testing
y_pred = svm.predict(X_test_vectorized.toarray())

from sklearn.metrics import classification_report, accuracy_score
from sklearn.model_selection import train_test_split
from sklearn.feature_extraction.text import TfidfVectorizer
from sklearn.naive_bayes import MultinomialNB
import pandas as pd

# =====
# (Alternatif) Vektorisasi TF-IDF untuk Naive Bayes (jika ingin uji
banding)
# =====
vectorizer = TfidfVectorizer()
X_tfidf = vectorizer.fit_transform(X)

# Split dataset menjadi training dan testing
X_train, X_test, y_train, y_test = train_test_split(X_tfidf, y,
test_size=0.2, random_state=42)

# Hitung akurasi prediksi SVM
accuracy_SVM = accuracy_score(y_test.values, y_pred)
print("SVM Accuracy:", accuracy_svm*100)
print("")

# Tampilkan classification report (precision, recall, f1-score)
print("SVM Clasification Report:")
print(classification_report(y_test, y_pred, labels=['positif',
'negatif']))

from sklearn.metrics import confusion_matrix

# Tentukan urutan label yang digunakan dalam confusion matrix
labels = ['positif', 'negatif']

# Hitung confusion matrix berdasarkan hasil prediksi (y_pred) dan label
asli (y_test)
cm = confusion_matrix(y_test, y_pred, labels=labels)

# Tampilkan confusion matrix dalam bentuk array
print("SVM Confusion Matrix:")
print(cm)

from sklearn.metrics import confusion_matrix
import matplotlib.pyplot as plt

```

```

import seaborn as sns

# =====
# CONFUSION MATRIX
# =====
labels = ['positif', 'negatif']
cm = confusion_matrix(y_test, y_pred, labels=labels)

# =====
# VISUALISASI CONFUSION MATRIX DENGAN HEATMAP
# =====
plt.figure(figsize=(6, 5)) # Ukuran proporsional
sns.heatmap(cm, annot=True, fmt='d', cmap='Blues', cbar=True,
            xticklabels=labels, yticklabels=labels, square=True)

plt.title("Confusion Matrix")
plt.xlabel("Predicted label")
plt.ylabel("True label")
plt.tight_layout() # Supaya tidak terpotong
plt.show()

```