

FINAL PROJECT
PENAMBANGAN DATA TEKSTUAL

“Desain dan Analisis Kebutuhan Fungsional dan Non-fungsional Sistem
Perangkat Lunak Analisis Sentimen pada Aplikasi Pajak Kendaraan Online”



Dosen Pengampu Mata Kuliah :
Dr. Ngurah Agus Sanjaya ER, S.Kom., M.Kom.

Disusun Oleh :

I Ketut Oning Pusparama	2008561017
I Wayan Trisna Wahyudi	2008561018
I Gusti Ayu Purnami Pinatih	2008561029
Tristan Bey Kusuma	2008561053
Revi Valen Sumendap	2008561099

PROGRAM STUDI INFORMATIKA
FAKULTAS MATEMATIKA DAN ILMU PENGETAHUAN ALAM
UNIVERSITAS UDAYANA
DENPASAR
2023

DAFTAR ISI

DAFTAR ISI.....	i
BAB I.....	iii
PENDAHULUAN	iii
1.1 Latar Belakang	iii
1.2 Rumusan Masalah	iv
1.3 Tujuan	iv
1.4 Manfaat	v
BAB II.....	1
TINJAUAN PUSTAKA	1
2.1 Landasan Teoritis	1
1. SIGNAL	1
2. Scrapping	2
3. Sentiment Analysis	2
4. SVM.....	3
5. TF-IDF	4
6. Conditional Random Fields	4
7. Metode Evaluasi: F1-score & AUC	5
2.2 Landasan Empiris.....	6
BAB III	8
ANALISIS DAN DESAIN	8
3.1 Tahap Pengumpulan Data	8
1. Scrapping Dataset	8
2. Merapikan Data (Pre Processing).....	9
3.2 Tahap Analisis Kebutuhan Fungsional dan Non-Fungsional.....	10
1. Fungsional.....	10
2. Non-Fungsional.....	11
3.3 Tahap Desain	11
3.4 Tahap desain - desain UI/UX (Mockup)	13
1. Halaman File Input.....	13
2. Halaman Text Input	14
3. Halaman Hasil.....	14

4. Halaman Feedback.....	15
3.5 Tahap Perancangan Eksperimen untuk Evaluasi Model dan Evaluasi Sistem.....	15
1. Pelatihan Model SVM.....	15
2. Evaluasi Model	15
3. Analisis Faktor-Faktor Pengaruh	16
4. Cross-Validation	16
5. Pengukuran Waktu Respon.....	16
6. Evaluasi Kepuasan Pengguna	16
DAFTAR PUSTAKA	18

BAB I

PENDAHULUAN

1.1 Latar Belakang

SIGNAL - Samsat Digital Nasional merupakan sebuah inovasi dalam layanan pembayaran pajak kendaraan bermotor di Indonesia. Aplikasi Signal bertujuan untuk memfasilitasi masyarakat dalam proses pengesahan Surat Tanda Nomor Kendaraan (STNK) tahunan, pembayaran Pajak Kendaraan Bermotor (PKB), dan Sumbangan Wajib Dana Kecelakaan Lalu Lintas Jalan (SWDKLLJ) tanpa perlu datang ke kantor Samsat. Sejak diluncurkan pada tahun 2021 aplikasi ini telah diunduh oleh 36.531 pengguna dan bertambah tiap tahunnya.

SIGNAL telah mengalami beberapa pembaruan. Pembaruan ini mencakup peningkatan keamanan, fungsionalitas baru, dan perbaikan bug untuk meningkatkan pengalaman pengguna secara keseluruhan. Saat ini pada platform Google Play Store aplikasi SIGNAL Korlantas Polri telah diunduh kurang lebih satu juta download dengan jumlah ulasan 101,4 ribu dengan rating 4.6. Berdasarkan ulasan sebagian besar pengguna merasa senang dengan kemungkinan membayar pajak kendaraan secara online tanpa harus datang ke kantor Samsat. Namun terdapat juga pengguna yang menyampaikan ketidakpuasannya terhadap sistem seperti mengalami kendala teknis, seperti kesulitan login atau bug tertentu. Hal tersebut menunjukkan pelayanan yang diberikan belum memenuhi harapan pengguna. Karena banyaknya sentimen dari pengguna aplikasi maka diperlukan lebih banyak waktu untuk mengelompokkan sentimen tersebut termasuk sentimen positif atau sentimen negatif. Oleh karena itu diperlukan sebuah sistem yang dapat secara otomatis mengklasifikasi sentimen positif atau negatif.

Beberapa penelitian terkait sentimen analisis yang pernah dilakukan yaitu analisis sentimen opini masyarakat terhadap vaksinasi Booster COVID-19 dengan membandingkan Metode Naive Bayes, Decision Tree dan SVM. Berdasarkan hasil evaluasi diperoleh hasil bahwa model SVM memiliki akurasi tertinggi dengan akurasi sebesar 83.33%. Selanjutnya model Decision Tree sebesar 79.00% dan Naive Bayes sebesar 70.00% [1]. Penelitian lainnya terkait sentimen analisis dilakukan untuk mengetahui bagaimana sentimen publik terhadap kebijakan yang akan dilakukan pemerintah mengenai kebijakan lockdown ataupun pembatasan sosial berskala besar menggunakan metode Support Vector Machine. Penelitian

tersebut menghasilkan menghasilkan nilai accuracy sebesar 74%, precision sebesar 75%, recall sebesar 92% dan F1-Score sebesar 83% [2]. Berikutnya terdapat penelitian terkait sentimen analisis dilakukan untuk melihat perbandingan metode Klasifikasi Support Vector Machine dan Naïve Bayes untuk Analisis Sentimen pada Ulasan Tekstual di Google Play Store. Hasil pengujian yang didapatkan dari metode 3-folds cross validation menghasilkan bahwa SVM Classifier memiliki nilai yang lebih tinggi jika dibandingkan dengan akurasi dari Naïve Bayes classifier untuk mengklasifikasikan ulasan tekstual berbahasa Indonesia pada Google Play Store, yakni SVM classifier mendapatkan akurasi sebesar 81,46% dan Naïve Bayes classifier sebesar 75,41%. sehingga metode SVM lebih baik untuk dijadikan metode klasifikasi untuk proses Analisis Sentimen ulasan tekstual berbahasa Indonesia pada Google Play Store.[3]

Berdasarkan penelitian-penelitian sebelumnya diperoleh kesimpulan bahwa model SVM menghasilkan hasil terbaik untuk melakukan analisis sentimen. Oleh karena itu penulis akan melakukan analisis sentimen terhadap ulasan aplikasi SIGNAL - Samsat Digital Nasional Korlantas Polri menggunakan metode SVM untuk mengetahui sentimen dari pengguna aplikasi SIGNAL.

1.2 Rumusan Masalah

Berdasarkan latar belakang tersebut maka diperoleh rumusan permasalahan sebagai berikut :

1. Bagaimana sentimen pengguna terhadap aplikasi SIGNAL?
2. Bagaimana performa dari metode Support Vector Machine (SVM) dalam melakukan analisis sentimen terhadap ulasan pengguna aplikasi SIGNAL?

1.3 Tujuan

Berdasarkan rumusan masalah tersebut adapun tujuan yang ingin dicapai dari penelitian ini adalah sebagai berikut :

1. Mengetahui sentimen pengguna terhadap aplikasi SIGNAL.
2. Mengetahui performa dari metode Support Vector Machine (SVM) dalam melakukan analisis sentimen terhadap ulasan pengguna aplikasi SIGNAL.

1.4 Manfaat

Manfaat dari penelitian ini adalah sebagai berikut :

1. Memberikan informasi kepada pihak terkait mengenai sentimen pengguna terhadap aplikasi SIGNAL.
2. Membantu pihak terkait untuk meningkatkan kualitas layanan aplikasi SIGNAL.

BAB II

TINJAUAN PUSTAKA

2.1 Landasan Teoritis

1. SIGNAL

SIGNAL (Samsat Digital Nasional) merupakan sebuah aplikasi yang mempermudah masyarakat dalam melakukan pengesahan STNK (Surat Tanda Nomor Kendaraan) Tahunan, (PKB) Pembayaran Pajak Kendaraan Bermotor dan (SWDKLLJ) Sumbangan Wajib Dana Kecelakaan Lalu Lintas Jalan. SIGNAL merupakan aplikasi resmi yang berada dibawah naungan Korlantas POLRI - Pembina Samsat Nasional dengan PT Jasa Raharja dan PT. Beta Pasifik Indonesia sebagai pihak pengembang (peveloper). Dengan adanya aplikasi ini, masyarakat tidak perlu lagi datang ke kantor samsat untuk melakukan pengesahan STNK. Cukup dengan mendaftarkan diri dan data kepemilikan kendaraan, maka pengesahan dapat dilakukan dengan beberapa menit saja.

Aplikasi ini merupakan One Stop Service, dimana aplikasi ini dapat melakukan mulai dari pengisian data user, pembayaran hingga pengiriman STNK ke alamat user. Aplikasi ini sementara hanya dapat digunakan untuk pembayaran di 27 Provinsi antara lain, Provinsi DKI Jakarta, Provinsi Banten, Provinsi Jawa Barat, Provinsi Jawa Tengah, Provinsi Jawa Timur, Provinsi Bali, Provinsi Sumatera Barat, Provinsi Riau, Provinsi Kep Riau, Provinsi Jambi, Provinsi Bengkulu, Provinsi Sulawesi Selatan, Provinsi Sulawesi Tenggara, Provinsi Sulawesi Barat, Provinsi NTB, Provinsi Lampung, Provinsi Kalimantan Barat, Provinsi Kalimantan Selatan, Provinsi Sumatera Selatan, Provinsi Daerah Istimewa Yogyakarta, Provinsi Bangka Belitung, Provinsi Kalimantan Tengah, Provinsi Sulawesi Tengah, Provinsi Aceh, Provinsi Sumatera Utara, Provinsi Gorontalo, dan Provinsi Sulawesi Utara. Aplikasi ini juga sementara dapat melakukan pembayaran menggunakan beberapa Bank saja antara lain, Bank Mandiri, BRI, BNI dan BTN serta Bank Pembangunan Daerah yang ada pada 27 Provinsi tersebut.

2. Scrapping

Web scraping atau web crawling mengacu pada prosedur ekstraksi data otomatis dari situs web menggunakan perangkat lunak [4]. Ini adalah proses yang sangat penting dalam bidang seperti Business Intelligence di zaman modern. Web scraping adalah teknologi yang memungkinkan kita untuk mengekstrak data terstruktur dari teks seperti HTML. Web scraping sangat berguna dalam situasi di mana data tidak disediakan dalam format yang dapat dibaca oleh mesin seperti JSON atau XML. Penggunaan web scraping untuk mengumpulkan data memungkinkan kita untuk mengumpulkan harga secara hampir real-time dari situs toko ritel dan memberikan detail lebih lanjut. Telah ditemukan bahwa menggunakan program web scraping akan menghasilkan data yang jauh lebih komprehensif, akurat, dan konsisten daripada entri manual. Berdasarkan hasil ini, disimpulkan bahwa Web scraping adalah alat yang sangat berguna dalam era informasi, dan merupakan salah satu yang penting dalam bidang-bidang modern.

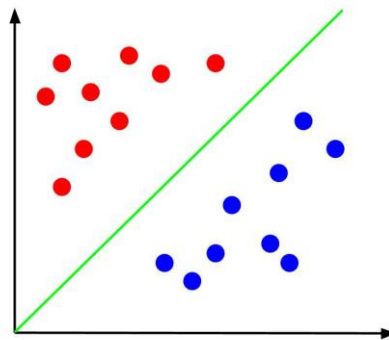
3. Sentiment Analysis

Sentiment analysis merupakan sebuah teknik yang digunakan untuk mengidentifikasi sentimen yang diekspresikan dalam sebuah kalimat [5]. Hal terpenting yang menjadi kebutuhan dasar dalam proses sentiment analysis adalah bagaimana memilih sebuah kata untuk digunakan sebagai fitur. Banyak penelitian yang telah dilakukan dalam proses analisis sentimen yang ada dalam sebuah review berdasarkan domain dictionaries. Meskipun terdapat potensi keunggulan, hasil analisis sentimen dengan menggunakan kamus domain sangat bergantung pada kelengkapan dan akurasi kamus tersebut. Bahasa yang digunakan dalam ulasan online berbeda secara signifikan dari bahasa formal, sehingga tidak hanya mencakup kata-kata khusus domain, tetapi juga banyak frasa yang populer di dunia internet. Hal ini membuat sulit untuk merumuskan kamus domain yang bersifat komprehensif dan akurat, mengingat dinamika linguistik yang terus berkembang dalam konteks lingkungan daring.

Penelitian kami mengatasi kesenjangan ini dengan mengusulkan suatu pendekatan analisis sentimen yang mengintegrasikan *conditional random fields* (CRFs) dan *support vector machines* (SVMs) untuk menganalisis ulasan online. Pendekatan ini pertama-tama mengekstrak fragmen fitur emosional dari ulasan menggunakan algoritma CRF, kemudian mengidentifikasi orientasi sentimen setiap ulasan dalam hal fragmen fitur emosional mereka, yang diklasifikasikan menggunakan SVM.

4. SVM

SVM merupakan jenis teknik pembelajaran mesin yang bersifat non-probabilistik yang sering digunakan untuk tugas klasifikasi. SVM dapat mengatasi masalah seperti klasifikasi dan regresi dengan linear maupun non linear. SVM yang menggunakan pendekatan berbasis hyperplane, digunakan untuk menganalisis data dan menentukan batas keputusan. Tujuan utama dari SVM adalah menentukan hyperplane terbaik yang digunakan untuk memisahkan data ke dalam kelas-kelas yang berbeda. Sebagai hasilnya, SVM mencari hyperplane dengan margin yang paling mungkin besar [5].



Gambar 2.1 Hyperline pada SVM

SVM membagi dataset dalam 2 kelas yang dipisahkan oleh hyperline dengan kelas pertama bernilai 1 (positif) dan kelas kedua bernilai -1 (negatif). Dalam perhitungan matematika dapat dibentuk persamaan seperti berikut:

$$X_i \cdot w + b \geq 1 \text{ untuk } Y_i = 1$$

$$X_i \cdot w + b \leq -1 \text{ untuk } Y_i = -1$$

Dimana,

w = nilai weight

b = nilai bias

X = data

Y = kelas data

5. TF-IDF

TF-IDF (Term Frequency-Inverse Document Frequency) adalah metode feature extraction atau ekstraksi fitur yang digunakan untuk memberikan bobot pada kata-kata dalam sebuah dokumen atau korpus berdasarkan seberapa sering kata tersebut muncul dalam dokumen tersebut dan seberapa umum kata tersebut di seluruh korpus. Tujuan utama dari TF-IDF adalah untuk menemukan kata-kata yang paling penting dalam dokumen tertentu. TF-IDF menggabungkan dua konsep untuk perhitungan bobot dalam sebuah dokumen yaitu :

1. Term Frequency (TF) – menghitung frekuensi kemunculan sebuah kata didalam sebuah dokumen tertentu.
2. Inverse Document Frequency (IDF) – menghitung seberapa umum kata tertentu di seluruh korpus dokumen.

Dengan menggabungkan kedua faktor tersebut, TF-IDF memberikan bobot pada kata dalam dokumen tersebut. Kata yang muncul lebih sering dalam dokumen dan kurang umum dalam seluruh korpus mendapatkan bobot yang lebih tinggi, sedangkan kata yang muncul lebih jarang dalam dokumen atau lebih umum dalam seluruh korpus mendapatkan bobot yang lebih rendah. [6]

6. Conditional Random Fields

Conditional Random Fields (CRF) pertama kali diperkenalkan pada tahun 2001 sebagai sebuah model probabilistic yang digunakan untuk segmentasi dan memberikan label pada data [7]. CRF bekerja dengan memodelkan distribusi probabilitas kondisional dari variabel output berdasarkan variabel

input. CRF adalah model probabilitas grafis yang digunakan untuk memprediksi struktur data yang memiliki ketergantungan spasial atau temporal, seperti urutan teks, citra, atau sinyal. CRF dikondisikan pada variabel input, dan digunakan untuk memodelkan distribusi probabilitas kondisional dari variabel output berdasarkan variabel input.

7. Metode Evaluasi: F1-score & AUC

Evaluasi model merupakan langkah penting dalam analisis data terstruktur maupun tidak terstruktur. Evaluasi model digunakan untuk mengetahui seberapa baik atau akurat model yang sudah dibangun dalam memberikan hasil yang tepat. Dalam penelitian ini menggunakan nilai F1-Score dan AUC (Area Under ROC Curve) sebagai metode evaluasi model.

F1-Score merupakan matriks evaluasi yang mencerminkan keseimbangan antara precision dan recall. Dimana precision merupakan matriks evaluasi yang mengukur seberapa baik model dalam memprediksi data yang benar untuk kelas positif dari total prediksi positif yang dilakukan dan recall merupakan matriks evaluasi yang mengukur seberapa baik model dalam memprediksi kelas positif dengan benar. Sehingga nilai F1-Score menjadi sangat penting dalam evaluasi model karena memberikan informasi tentang seberapa efektif model dalam melakukan klasifikasi data.

AUC (Area Under ROC Curve) merupakan metode evaluasi kemampuan model dalam membedakan data positif dan negatif. AUC menghitung nilai total area bawah kurva ROC (Receiver Operating Characteristics). Dimana ROC sendiri menggambarkan TPR (True Positive Rate) atau biasa disebut dengan recall dan FPR (False Positive Rate). Berikut rumus untuk evaluasi model menggunakan F1-Score dan AUC:

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

$$Recall/TPR = \frac{TP}{TP + FN}$$

$$eFPR = \frac{FP}{FP + TN}$$

$$F1 - Score = 2 \times \frac{Precision \times Recall}{Precision + Recall}$$

$$AUC = \frac{1 + TPR - FPR}{2}$$

2.2 Landasan Empiris

Pada penelitian ini, terdapat beberapa penelitian sebelumnya yang dijadikan dasar dalam penyusunan penelitian ini yaitu sebagai berikut ;

1. A survey on sentiment analysis methods, applications, and challenges (Wankhade. et al, 2022)

Pada penelitian ini membahas mengenai penerapan, tantangan, metode, dan pendekatan dalam analisis sentimen. Bagian ini juga menyoroti kesulitan dalam analisis sentimen, termasuk sentimen terstruktur, semi-terstruktur, dan tidak terstruktur, deteksi sarkasme, gaya penulisan informal, dan kesalahan tata bahasa. Dalam penelitian ini dijelaskan SVM adalah algoritma populer untuk analisis sentimen karena kemampuannya menentukan hyperplane yang paling baik memisahkan data ke dalam kelas-kelas berbeda. SVM mencari hyperplane dengan margin tertinggi, sehingga efektif untuk tugas klasifikasi. SVM dikenal memberikan akurasi yang baik untuk kumpulan data yang besar, namun menyempurnakan modelnya bisa jadi sulit dan memakan waktu. Selain itu, SVM memiliki waktu pelatihan yang lama untuk kumpulan data yang besar, yang dapat menjadi kelemahannya.

2. A comprehensive review of conditional random fields: variants, hybrids and applications (Yu et al, 2019)

Pada penelitian ini membahas penerapan Conditional Random Fields (CRFs) di berbagai bidang seperti pemrosesan bahasa alami, visi komputer, biomedis, dan banyak lagi. Penelitian ini menyoroti pentingnya dan efektivitas CRF dalam memproses data dalam bentuk teks, suara, gambar, dan video. CRFs memiliki beberapa kelebihan, antara lain kemampuan untuk merepresentasikan ketergantungan jarak jauh dan fitur yang tumpang tindih, serta kemampuan untuk menyelesaikan masalah bias label yang ada pada Maximum Entropy Markov Model (MEMM.). Selain itu, pelatihan CRFs hanya melibatkan convex objective functions yang membantu

mendapatkan estimasi parameter optimal secara global. CRFs juga mampu menyelesaikan tiga masalah dasar dan menunjukkan kinerja yang baik dalam masalah pelabelan urutan.

3. Sentiment analysis for online reviews using conditional random fields and support vector machines (Huosong. et al, 2019)

Pada penelitian ini memberikan tinjauan menyeluruh tentang analisis sentimen opini online, mulai dari konsep dasar hingga aplikasinya. Ini membandingkan kinerja SVM dan CRF untuk analisis sentimen opini ulasan online. SVM mampu mencapai akurasi 93%, sedangkan CRF hanya mencapai akurasi 90%. Hasilnya menunjukkan bahwa SVM dan CRF dapat memberikan kinerja yang baik, tetapi SVM umumnya lebih unggul dalam hal akurasi dan efisiensi komputasi. Namun, dalam penelitian tersebut, kumpulan data kami relatif sederhana sehingga tidak dapat menunjukkan hasil untuk data yang lebih kompleks. Studi ini juga hanya menggunakan metode ekstraksi fitur yang sederhana, yaitu bag-of-words.

BAB III

ANALISIS DAN DESAIN

3.1 Tahap Pengumpulan Data

1. Scrapping Dataset

Di sini, kami menggunakan Google Colab dan bahasa pemrograman Python untuk melakukan proses scrapping data dari Review Google Play Store. Kami menggunakan Library `google-play-scraper` dan Kami juga memanfaatkan library umum seperti `Pandas` dan `NumPy` untuk membantu dalam pengolahan data yang diekstraksi.

Google Colab merupakan platform yang memungkinkan kita untuk menulis dan mengeksekusi kode Python dalam lingkungan cloud. Dengan menggunakan Google Colab, kita dapat mengakses sumber daya komputasi yang kuat dan berbagi kode dengan mudah.

Bahasa pemrograman Python menjadi pilihan kami karena memiliki banyak modul dan pustaka yang mendukung proses scrapping data dengan efisien. Python juga memiliki sintaks yang mudah dipahami dan digunakan oleh banyak pemrogram.

Selain itu, kami memanfaatkan `Pandas` dan `NumPy`, dua pustaka populer dalam ekosistem Python, untuk membantu dalam pemrosesan dan analisis data. `Pandas` menyediakan struktur data yang kuat seperti `DataFrame`, yang sangat berguna dalam mengorganisir dan memanipulasi data yang diperoleh dari scrapping. Sementara itu, `NumPy` memberikan fungsionalitas matematika dan manipulasi array yang efisien, yang dapat digunakan dalam analisis lebih lanjut.

Dengan menggunakan kombinasi Google Colab, Python, `Pandas`, dan `NumPy`, kami dapat mengambil dataset sebanyak 10000 data dari Review Google Play Store tentang Aplikasi `SIGNAL-SAMSAT DIGITAL NASIONAL` dengan fokus pada komentar berbahasa Indonesia dan sort komentar berdasarkan komentar yang "most relevant". Data ini dapat digunakan untuk analisis dan pemahaman lebih lanjut terkait pengalaman pengguna dan ulasan aplikasi di Google Play Store.

1	userName	Sentiment	at	content
2	Ila Laila	Negatif	2023-12-18 11:05:19	parah strnk belum dikirim transaksi udah di anggap selesai. udah 2 minggu lebih belum dikirim kirim. ga lagi deh kedepannya
3	Muhammad Arbi	Negatif	2023-12-18 8:13:37	Aplikasi pembodohan verifikasi data selalu tidak valid katanya perbarui dukcapil., padahal aku baru pindah domisili tentunya
4	Tiara Chairani	Negatif	2023-12-18 6:31:09	Susah dlm pembayaran kode pembayaran selalu salah,pdhl sudah input sesuai kode dr signal,sudah berulang x tetap salah,
5	Edi Sugianto	Negatif	2023-12-18 5:53:58	Saya sudah bayar ongkir yg express tapi tdk di kirim2, akhirnya jemput bola dengan ongkir tdk bisa kembali,pelayanan yg san
6	Dewi Sartika Ghea	Negatif	2023-12-18 3:41:47	Saya Heran kenapa aplikasi tidak layak guna seperti ini bisa di suruh gunakan kepada masyarakat, Saya sudah membayar d
7	Budi Santoso	Negatif	2023-12-18 2:37:23	3x gagal transaksi saat pakai menu kirim paket kerumah sampai dana dikembalikan ke tokopedia. Pas nyoba tanpa dikirim ke
8	Luthfi Heryanto	Negatif	2023-12-18 2:01:43	Tanpa kejelasan berapa hari kerja pengiriman STNK, hanya stuck tidak ada update
9	Kandas Telor	Negatif	2023-12-16 23:01:04	Aplikasi bobrok. pembayaran tidak terverifikasi sampai kadaluarsa...padahal pembayaran melalui BNI sudah berhasil.. di em
10	Amran Hamdani	Negatif	2023-12-16 3:27:51	Mau register susah error terus...padahal koneksi internetnya ngebut
11	Fakhri Maulana Arsyad	Negatif	2023-12-16 3:14:41	mau daftar gabisa bisa error terus

Gambar 3.1 Contoh Hasil Scrapping Dataset Sentiment Negatif

1	userName	sentiment	at	content
2	Shony Apriliant	Positif	2023-12-18 12:08:50	Pelayanan cepat bagus dan murah tanpa biaya apapun.
3	Koce Diesel	Positif	2023-12-18 11:12:21	bayar pajak mudah cukup sambil tiduran di rumah sambil ngopi, untuk kedepan nya signal mudah'n bisa lebih mudah lagi dalam proses pembayaran nya
4	Waluyo	Positif	2023-12-18 11:01:02	sangat membantu sekali bagi yg tidak memiliki banyak waktu karena cukup dengan proses lewat online tanpa mengantri STNK langsung d kirim k alamat
5	Toni Kurniawan	Positif	2023-12-18 6:38:59	biar Rutin bayar pajak, pembayaran pajak ga usah dipersulit. susah klo beli kendaraan second ga ada KTP pemilik lama,
6	Urai Ikhsan Fadhillah	Positif	2023-12-18 6:14:38	Saya sudah bayar Melalui dana Namun status di aplikasi signal masih belum terbayarkan, sedangkan kita diberi waktu yang terus berjalan bagaimana jika
7	Safar Azmi	Positif	2023-12-18 5:30:17	mantap proses cepet,nga perlu antri kayak mau ambil sembako. good job E samsat
8	Naa nia	Positif	2023-12-18 3:19:47	bayar pajak sangat mudah. mesko pun kita di luar propinsi. terimakasih SIGNAL
9	Copral ariyanto	Positif	2023-12-18 1:55:01	pengisian data yg mudah proses verifikasi nya pun cepat dan metode pembayaran nya pun banyak pilihan, pertahankan pelayanan seperti ini, terima kasih
10	Dicky Ardiansyah	Positif	2023-12-18 0:15:43	sangat membantu sekali untuk perpanjang strnk lebih cepat
11	edwin eko	Positif	2023-12-17 15:40:32	Dengan adanya signal sangat membantu dan mudah utk bayar perpanjang STNK tahunan. terima kasih SIGNAL

Gambar 3.2 Contoh Hasil Scrapping Dataset Sentiment Positif

2. Merapikan Data (Pre Processing)

Setelah mengambil data dari Review Google Play Store, kami akan menjalani serangkaian tahapan pemrosesan teks untuk membersihkan dan mempersiapkannya untuk analisis lebih lanjut. Berikut adalah tahapan yang akan kami lakukan:

- Convert Lowercase:** Mengonversi semua teks review ke huruf kecil. Hal ini dilakukan agar tidak ada perbedaan dalam pengolahan kata yang memiliki huruf besar atau kecil.
- Removal of Punctuations:** Menghapus tanda baca dari teks review, seperti titik, koma, tanda tanya, dan sebagainya. Hal ini bertujuan untuk memfokuskan pada kata-kata penting dalam teks dan menghindari pemrosesan yang tidak relevan dengan tanda baca.
- Removal of Stopwords:** Menghapus kata-kata umum yang tidak memberikan kontribusi signifikan dalam pemahaman konten, seperti "the", "is", "and", "a", dan lain-lain. Penghapusan stopwords dapat membantu mengurangi ukuran dataset dan memfokuskan pada kata-kata kunci yang lebih informatif.
- Removal of Special Characters:** Menghapus karakter khusus yang tidak relevan atau tidak diinginkan dalam teks review, seperti simbol, emotikon, URL, atau karakter khusus lainnya. Hal ini

dilakukan untuk mempertahankan hanya karakter yang penting dalam analisis.

- e. Lemmatization: Menerapkan proses lemmatisasi pada kata-kata dalam teks review. Lemmatisasi adalah proses mengubah kata-kata ke bentuk dasarnya (kata dasar). Misalnya, kata "menjalankan" akan diubah menjadi "jalan", sehingga kata-kata dengan akar yang sama dapat dianggap sebagai entitas yang sama.

Dengan menjalani kelima tahap ini, kami akan memiliki dataset yang lebih bersih, terstruktur, dan siap untuk diolah lebih lanjut. Pemrosesan ini membantu mengurangi noise dalam data dan memfokuskan pada kata-kata kunci yang memberikan informasi berharga dalam analisis ulasan di Google Play Store.

3.2 Tahap Analisis Kebutuhan Fungsional dan Non-Fungsional

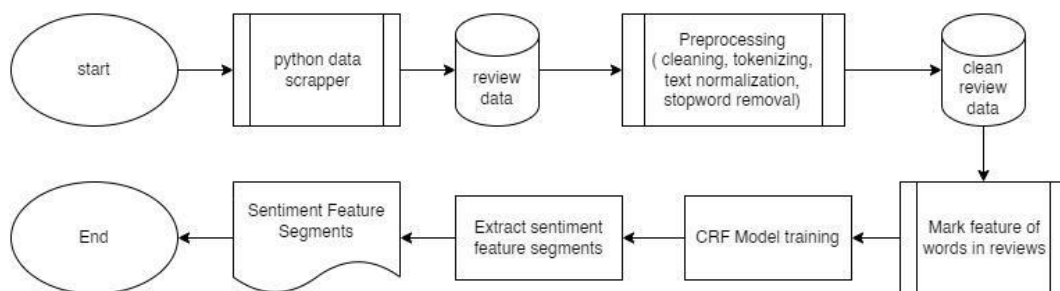
1. Fungsional

- a. Pengumpulan Data Sentimen: Situs web harus mampu mengambil data sentimen dari inputan user yang dapat berupa sebuah kalimat atau file dengan format csv.
- b. Analisis Sentimen: Kemampuan untuk menganalisis sentimen dalam teks yang dikumpulkan, yang mencakup pengenalan sentimen positif dan negatif.
- c. Manajemen Data: Kemampuan untuk menyimpan dan mengelola data sentimen yang dikumpulkan.
- d. Visualisasi Data: Menampilkan hasil analisis sentimen dalam bentuk grafik atau laporan yang mudah dimengerti.
- e. Interaksi Pengguna: Sistem harus memungkinkan pengguna untuk memasukkan teks atau sumber data yang akan dianalisis, dan untuk melihat hasil analisis dengan jelas.
- f. Kemampuan Beradaptasi: Memungkinkan penambahan atau perubahan aspek yang akan dianalisis dalam masa depan.

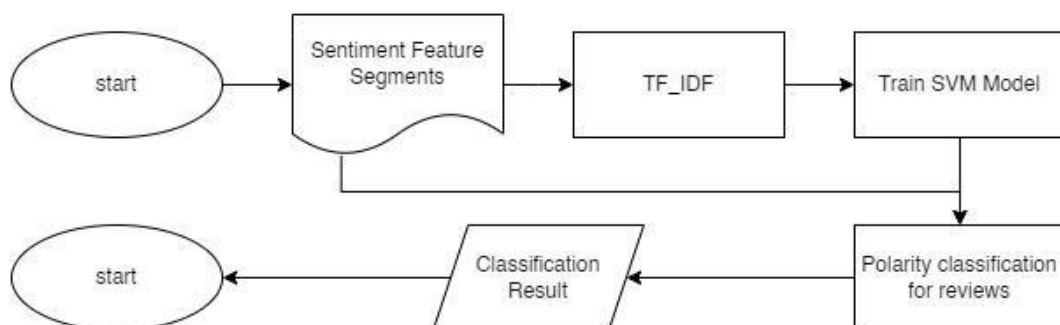
2. Non-Functional

- Kinerja: Website harus mampu mengatasi volume besar data yang akan dianalisis dalam waktu yang wajar.
- Keamanan: Perlindungan data dan privasi pengguna harus diutamakan, dengan enkripsi yang kuat dan perlindungan terhadap ancaman siber.
- Skalabilitas: Kemampuan untuk berkembang sesuai kebutuhan, termasuk peningkatan kapasitas untuk menangani lebih banyak data.
- Kemudahan Penggunaan: Antarmuka pengguna harus ramah pengguna dan mudah dinavigasi.
- Ketersediaan: Website harus tersedia sepanjang waktu atau memiliki waktu jeda yang minimal untuk pemeliharaan.

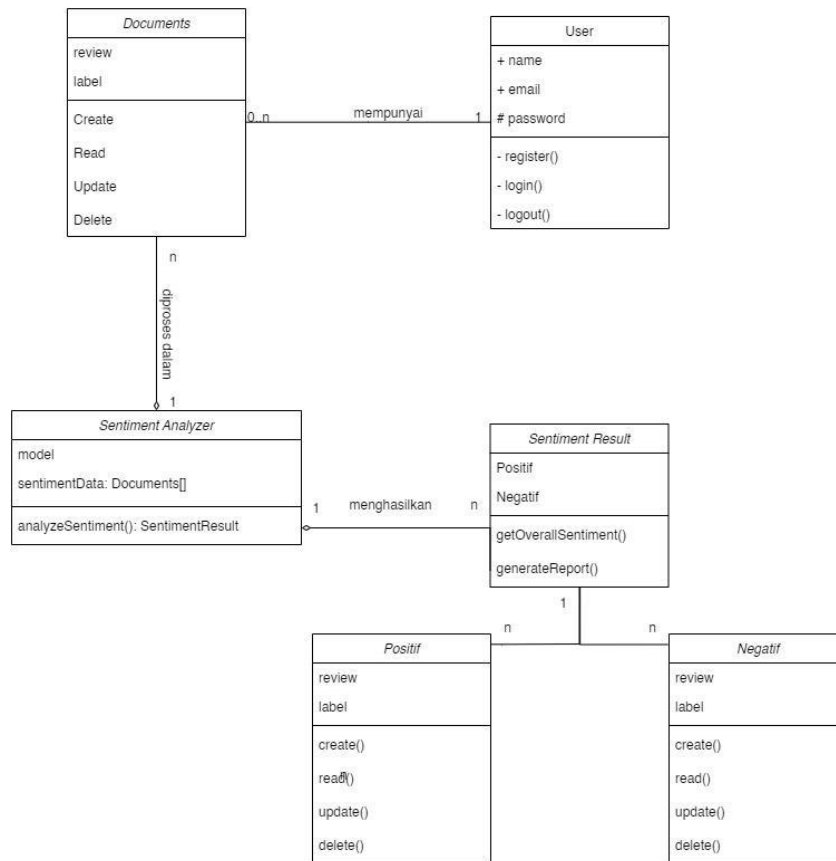
3.3 Tahap Desain



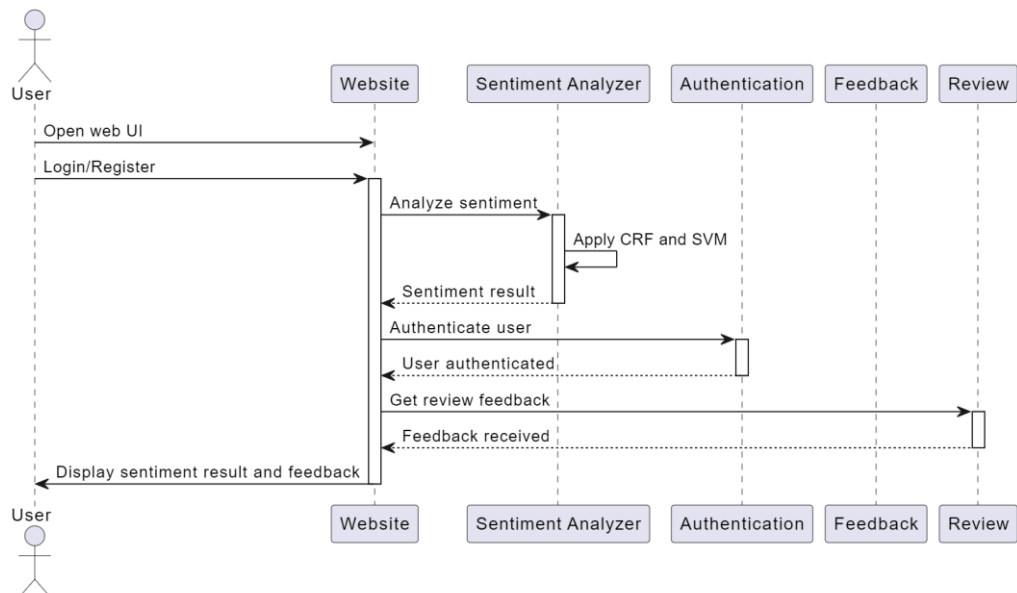
Gambar 3.3 Ekstraksi fragmen fitur sentimen dengan conditional random field (CRF)



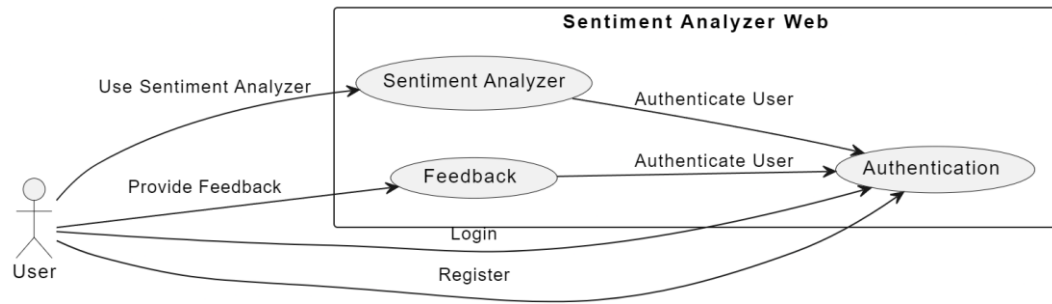
Gambar 3.4 Klasifikasi ulasan dengan support vector machine (SVM)



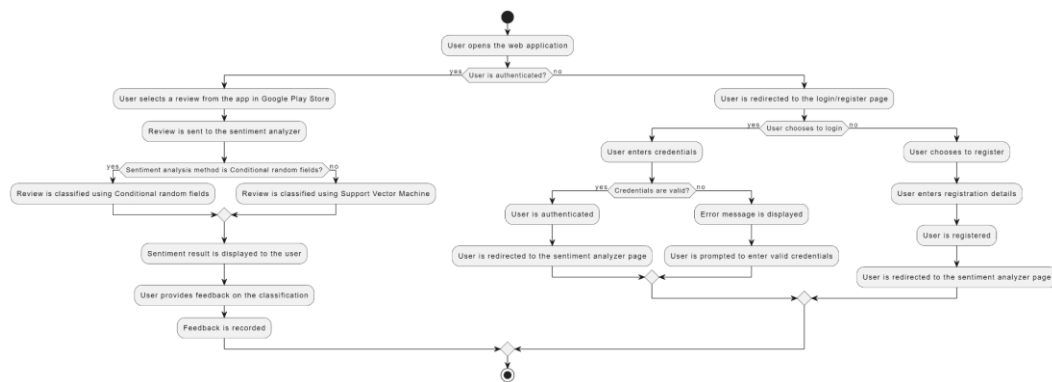
Gambar 3.5 Class Diagram



Gambar 3.6 Sequence Diagram



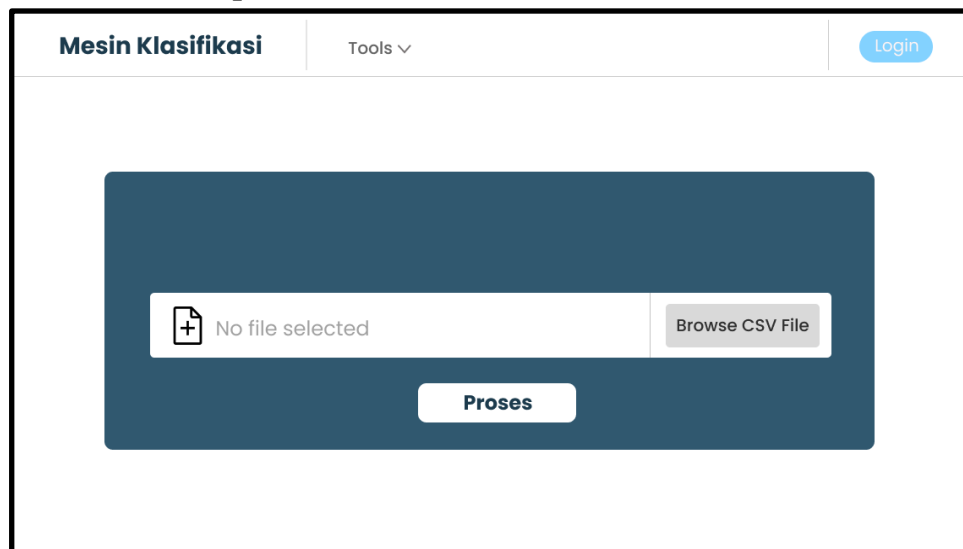
Gambar 3.7 Use Case Diagram



Gambar 3.8 Activity Diagram

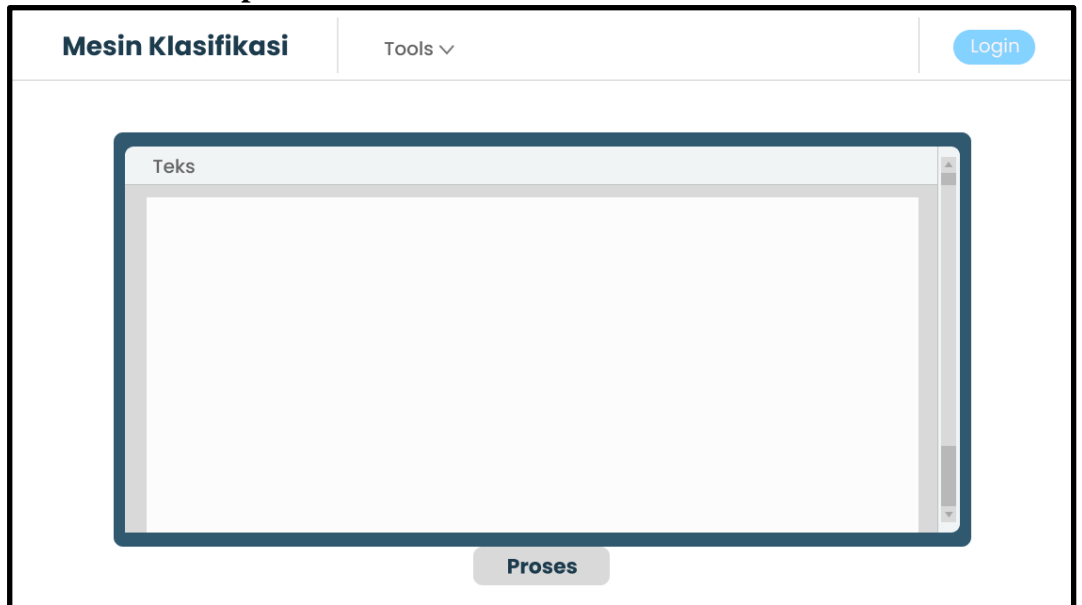
3.4 Tahap desain - desain UI/UX (Mockup)

1. Halaman File Input



Gambar 3.9 Desain File Input

2. Halaman Text Input



The screenshot shows a web interface for a classification machine. The header includes the title 'Mesin Klasifikasi', a 'Tools' dropdown menu, and a 'Login' button. The main content area features a large text input field with the placeholder text 'Teks' and a 'Proses' button located below the input field.

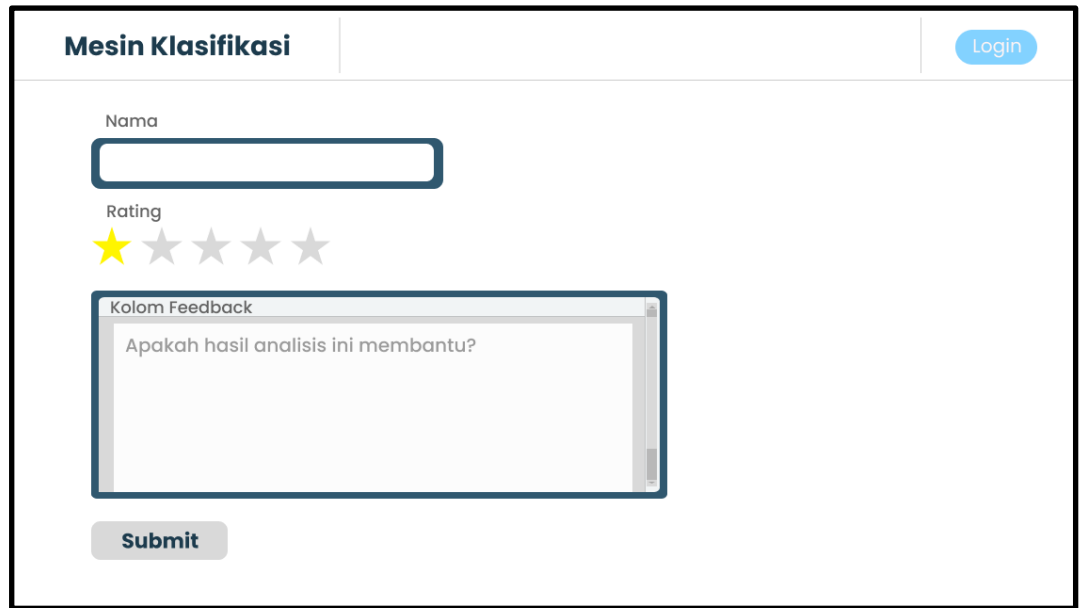
Gambar 3.10 Desain Text Input

3. Halaman Hasil



Gambar 3.11 Desain Hasil Analisis Sentimen

4. Halaman Feedback



Gambar 3.12 Desain Feedback Model

3.5 Tahap Perancangan Eksperimen untuk Evaluasi Model dan Evaluasi Sistem

Perancangan eksperimen untuk sentimen analisis teks dari opini Twitter harus mempertimbangkan beberapa aspek penting. Berikut adalah langkah-langkah eksperimen yang kami pertimbangkan:

1. Pelatihan Model SVM

Gunakan bagian dataset pelatihan untuk melatih model SVM. Disini kami menyesuaikan hyperparameter SVM, seperti kernel, gamma dan C, melalui validasi silang untuk meningkatkan kinerja.

2. Evaluasi Model

Gunakan dataset pengujian yang terpisah untuk mengevaluasi kinerja model. Akan diukur metrik evaluasi seperti akurasi, presisi, recall, F1 score, dan matriks confusion.

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

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

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

$$F1-score = \frac{2 \times Precision \times Recall}{Precision + Recall}$$

Keterangan :

TP (True Positive) : Jumlah data yang diprediksi benar. Misalnya, dalam analisis sentimen opini, sebuah tweet diprediksi sebagai sentimen negatif dan ternyata memang benar sentimennya negatif.

TN (True Negative) : Data yang diprediksi tidak dan ternyata memang tidak. Misalnya, sebuah tweet diprediksi sebagai sentimen positif dan ternyata memang benar sentimennya positif.

FP (False Positive) : Data yang diprediksi salah positif. Misalnya, sebuah tweet diprediksi sebagai sentimen negatif padahal sentimennya sebenarnya positif.

FN (False Negative) : Data yang diprediksi salah negatif. Misalnya, sebuah tweet diprediksi sebagai sentimen positif padahal sentimennya sebenarnya negatif.

3. Analisis Faktor-Faktor Pengaruh

Identifikasi faktor-faktor yang dapat memengaruhi kinerja model, seperti jumlah data, kualitas pra-pemrosesan, dan parameter SVM. Kemudian melakukan analisis sensitivitas terhadap parameter untuk memahami dampaknya terhadap hasil.

4. Cross-Validation

Menerapkan validasi silang pada dataset pelatihan untuk memastikan keandalan model dan mengurangi risiko overfitting.

5. Pengukuran Waktu Respon

Setelah itu, diukur waktu yang diperlukan untuk memproses analisis sentimen setiap opini media sosial. Ini adalah evaluasi kinerja sistem dari segi respons waktu.

6. Evaluasi Kepuasan Pengguna

Kami mengimplementasikan sebuah fitur feedback dimana user dapat menilai apakah hasil analisis sentimen dianggap membantu atau tidak. Melalui skor rating dari 1 hingga 5 maka kami dapat menilai kepuasan pengguna secara kuantitatif.

Rencana eksperimen ini melibatkan evaluasi model analisis sentimen dan sistem aplikasi. Tentukan matrik evaluasi seperti akurasi, presisi, recall, dan F1

score. Spesifikasi pengaturan eksperimen termasuk dataset untuk pelatihan dan pengujian, parameter untuk model SVM, dan kriteria keberhasilan. Evaluasi sistem aplikasi melibatkan faktor-faktor seperti kepuasan pengguna, waktu respon, dan kinerja keseluruhan.

DAFTAR PUSTAKA

- [1] Rima Aldisa and Pandu Maulana, "Analisis Sentimen Opini Masyarakat Terhadap Vaksinasi Booster COVID-19 Dengan Perbandingan Metode Naive Bayes, Decision Tree dan SVM," *Build. Inform. Technol. Sci. BITS*, vol. 4, no. 1, Jun. 2022, doi: 10.47065/bits.v4i1.1581.
- [2] A. R. Isnain, A. I. Sakti, D. Alita, and N. S. Marga, "Sentimen Analisis Publik Terhadap Kebijakan Lockdown Pemerintah Jakarta Menggunakan Algoritma Svm," *J. Data Min. Dan Sist. Inf.*, vol. 2, no. 1, pp. 31–37, 2021.
- [3] L. B. Ilmawan and M. A. Mude, "Perbandingan metode klasifikasi Support Vector Machine dan Naïve Bayes untuk analisis sentimen pada ulasan tekstual di Google Play Store," *Ilk J Ilm*, vol. 12, no. 2, pp. 154–161, 2020.
- [4] M. A. Khder, "Web Scraping or Web Crawling: State of Art, Techniques, Approaches and Application.," *Int. J. Adv. Soft Comput. Its Appl.*, vol. 13, no. 3, 2021.
- [5] M. Wankhade, A. C. S. Rao, and C. Kulkarni, "A survey on sentiment analysis methods, applications, and challenges," *Artif. Intell. Rev.*, vol. 55, no. 7, pp. 5731–5780, Oct. 2022, doi: 10.1007/s10462-022-10144-1.
- [6] S. Qaiser and R. Ali, "Text mining: use of TF-IDF to examine the relevance of words to documents," *Int. J. Comput. Appl.*, vol. 181, no. 1, pp. 25–29, 2018.
- [7] B. Yu and Z. Fan, "A comprehensive review of conditional random fields: variants, hybrids and applications," *Artif. Intell. Rev.*, vol. 53, no. 6, pp. 4289–4333, Aug. 2020, doi: 10.1007/s10462-019-09793-6.