

Analisis Sentimen Masyarakat terhadap Penggunaan Sepeda Listrik pada Anak-Anak di Media Sosial X Menggunakan Metode SVM

Rohmah Dewi Lestari*, Auliya Rahman Isnain

Fakultas Teknik dan Ilmu Komputer, Sistem Informasi, Universitas Teknokrat Indonesia, Bandar Lampung, Indonesia

Email: ^{1,*}Rohmah_dewi_lestari@teknokrat.ac.id, ²auliyarahman@teknokrat.ac.id

Email Penulis Korespondensi: Rohmah_dewi_lestari@teknokrat.ac.id

Submitted: 12/12/2024; Accepted: 29/12/2024; Published: 30/12/2024

Abstrak—Penggunaan sepeda listrik pada anak menjadi fenomena yang semakin populer di Indonesia. Meskipun praktis dan efisien, penggunaannya menimbulkan kekhawatiran, terutama terkait risiko keselamatan anak di jalan raya. Berbagai pandangan masyarakat mengenai hal ini ramai diperbincangkan di media sosial X (Twitter), sebagian mendukung penggunaannya karena praktis dan ramah lingkungan, sementara yang lain mengusulkan regulasi lebih untuk memastikan keselamatan anak-anak saat berkendara. Penelitian ini bertujuan untuk menganalisis sentimen masyarakat terhadap penggunaan sepeda listrik untuk anak-anak dengan menggunakan metode *Support Vector Machine (SVM)*. Data dikumpulkan melalui proses *crawling* di media sosial X menggunakan *tools Tweet Harvest*, yang menghasilkan total 3.565 data. Data mentah ini kemudian diproses melalui tahap *preprocessing* data, dilanjutkan dengan penerjemahan semua *tweet* ke dalam bahasa Inggris terlebih dahulu karena *TextBlob* lebih efektif dalam menganalisis sentimen pada data berbahasa Inggris. Kemudian, dilakukan pelabelan dengan mengelompokkan data menjadi 1.737 sentimen negatif (64,24%) serta 967 sentimen positif (35,76%), yang selanjutnya dibagi menjadi dua set data dengan 80% untuk pelatihan dan 20% untuk pengujian. Model SVM digunakan dengan pendekatan *kernel linear* untuk membangun klasifikasi. Hasil pengujian model dievaluasi dengan *confusion matrix* yang menunjukkan nilai akurasi 0,84, *precision* 0,84 untuk sentimen negatif dan 0,85 untuk sentimen positif, *recall* 0,92 untuk sentimen negatif dan 0,71 untuk sentimen positif, serta *F1-score* masing-masing 0,88 untuk sentimen negatif dan 0,78 untuk sentimen positif. Berdasarkan hasil tersebut, dapat disimpulkan bahwa mayoritas masyarakat lebih memprioritaskan kekhawatiran terhadap risiko keselamatan anak-anak.

Kata Kunci: Analisis Sentimen; Media Sosial X; Sepeda Listrik Anak; SMOTE; Support Vector Machine; TF-IDF

Abstract—The use of electric bicycles among children is becoming increasingly popular in Indonesia. While offering practicality and mobility efficiency, their usage raises safety concerns, especially for children on the road. Public opinions on this issue are widely discussed on social media platform X (Twitter), with some supporting their use due to practicality and eco-friendliness, while others advocate stricter regulations to ensure children's safety. This study analyzes public sentiment toward the use of electric bicycles for children using the Support Vector Machine (SVM) method. Data was collected through a crawling process on social media X using the Tweet Harvest tool, resulting in 3,565 entries. The data underwent preprocessing and translation into English for sentiment analysis using TextBlob. Sentiments were labeled, identifying 1,737 negative sentiments (64.24%) and 967 positive sentiments (35.76%). The dataset was divided into 80% for training and 20% for testing. An SVM model with a linear kernel was applied for classification. Performance evaluation using a confusion matrix showed 0.84 accuracy, precision scores of 0.84 (negative) and 0.85 (positive), recall scores of 0.92 (negative) and 0.71 (positive), and F1-scores of 0.88 (negative) and 0.78 (positive). The findings reveal that public sentiment predominantly reflects concerns about children's safety risks.

Keywords: Children's Electric Bicycles; Sentiment Analysis; SMOTE; Social Media X; Support Vector Machine; TF-IDF

1. PENDAHULUAN

Penggunaan sepeda listrik di Indonesia dalam beberapa tahun terakhir semakin populer, terutama di kalangan anak-anak. Sepeda listrik menawarkan berbagai keuntungan, seperti kemudahan dan efisiensi dalam beraktivitas di luar rumah [1]. Namun, Peningkatan penggunaan sepeda listrik tersebut, juga memunculkan berbagai permasalahan, seperti tingginya risiko kecelakaan pada anak-anak, mengganggu pengguna jalan yang lain, dan dampak negatif terhadap aktivitas fisik anak [2]. Hal tersebut dikarenakan anak-anak belum lihai dalam mengendarai kendaraan di jalan raya dan belum mengetahui tentang aturan lalu lintas di jalan raya [3]. Berdasarkan data, jumlah kecelakaan yang melibatkan sepeda listrik mencapai 647 kasus di sepanjang Januari-Juni 2024, dengan rata-rata lebih dari 100 kasus per bulan. Salah satu kelompok yang rentan dalam kecelakaan tersebut yaitu anak-anak [4]. Oleh karena itu, penelitian ini sangat penting untuk memahami pandangan masyarakat terhadap penggunaan sepeda listrik, guna mendukung pengambilan kebijakan yang tepat dan meminimalkan risiko yang mungkin terjadi.

Seiring dengan meningkatnya penggunaan sepeda listrik, tanggapan masyarakat terhadap fenomena ini semakin beragam, terutama di platform media sosial X. Beberapa pihak menganggap sepeda listrik sebagai solusi mobilitas yang efisien, sementara pihak lain mengkhawatirkan dampaknya terhadap keselamatan dan gaya hidup anak [5]. Urgensi penelitian ini terletak pada pentingnya menganalisis sentimen masyarakat terhadap penggunaan sepeda listrik untuk anak-anak, terutama di media sosial, yang merupakan sarana utama dalam bertukar informasi dan pendapat. Meskipun terdapat sejumlah penelitian mengenai sentimen terhadap kendaraan listrik secara umum, tidak ada kajian yang secara khusus mengkaji sentimen masyarakat terhadap penggunaan sepeda listrik bagi anak-anak. Penelitian ini bertujuan untuk mengisi kekosongan tersebut dengan memberikan wawasan mengenai persepsi publik terhadap sepeda listrik sebagai alternatif transportasi untuk anak-anak.

Analisis sentimen merupakan proses menilai opini seseorang terhadap suatu objek yang berfungsi untuk membantu dalam pengambilan keputusan dengan memahami apakah seseorang menyukai objek tersebut atau tidak,

dalam hal ini penggunaan sepeda listrik [6]. Pada penelitian kali ini metode yang digunakan yaitu Support Vector Machine (SVM), SVM memungkinkan klasifikasi sentimen opini masyarakat dengan akurasi tinggi dan mampu mencapai hasil signifikan tanpa memerlukan banyak daya komputasi [7]. Selain itu, SVM lebih mudah dilatih dibandingkan algoritma *machine learning* lainnya [8]. Variabel penelitian mencakup sentimen positif, negatif, dan netral yang diungkapkan oleh pengguna media sosial terkait sepeda listrik untuk anak-anak. Permasalahan yang diangkat adalah bagaimana persepsi masyarakat terhadap sepeda listrik, serta faktor-faktor yang mempengaruhi sentimen tersebut.

Penelitian sebelumnya menunjukkan bahwa algoritma SVM memiliki efektivitas yang signifikan dalam menganalisis sentimen yang diungkapkan di media sosial. Dalam penelitian terkait Program Kartu Prakerja, Styawati, Hendrastuty, Isnain, dan Rahmadhani (2021) menganalisis sentimen masyarakat Indonesia terhadap program tersebut dengan menggunakan data Twitter dan metode *Support Vector Machine* (SVM) pada kernel linear dan RBF. Hasil penelitian menunjukkan akurasi sebesar 98,67% untuk kernel linear dan 98,34% untuk kernel RBF, dengan sentimen netral mendominasi sebesar 98,34%. Temuan ini mengonfirmasi bahwa SVM efektif dalam mengklasifikasikan sentimen publik terhadap kebijakan pemerintah selama masa pandemi [9].

Dalam penelitian lain, Melani Alka Syahira dan Rakhmat Kurniawan R. (2024) menggunakan SVM untuk menganalisis sentimen tentang cyberbullying di media sosial X, memperoleh akurasi sebesar 87%, serta presisi, recall, dan F1-score yang tinggi, yang semakin memperkuat klaim bahwa SVM merupakan alat yang andal dalam memprediksi sentimen negatif yang berkembang di dunia maya [6].

Selanjutnya, penelitian oleh Muhammad Daffa Al Fahreza, Ardytha Luthfiarta, Muhammad Rafid, and Michael Indrawan (2024) mengaplikasikan SVM untuk menganalisis sentimen terkait pengaruh jam kerja terhadap kesehatan mental generasi Z di Twitter, dengan akurasi yang sangat baik mencapai 91%. Penelitian ini tidak hanya menunjukkan keefektifan SVM dalam analisis sentimen, tetapi juga menggambarkan bagaimana alat ini dapat digunakan untuk menggali topik-topik terkait kesejahteraan psikologis dan sosial di kalangan kelompok usia tertentu [10].

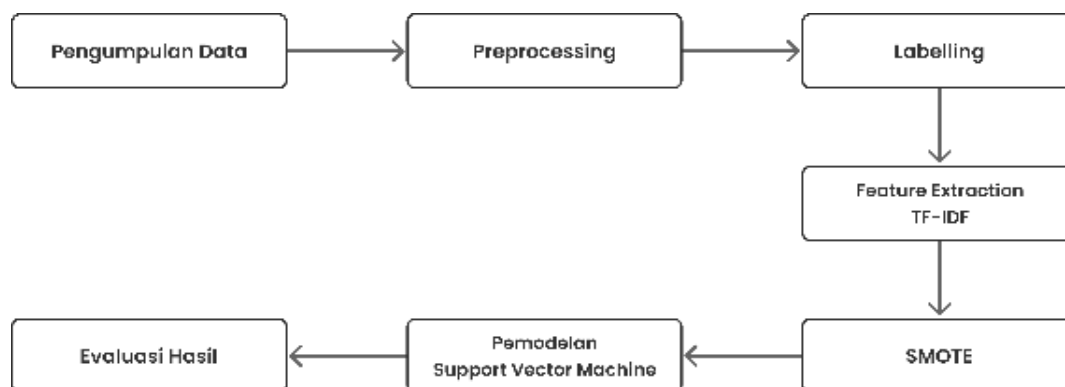
Di sisi lain, Jesica Emarpenta Br Sinulingga dan H. C. K. Sitorus (2024) menggunakan kombinasi SVM dan TF-IDF untuk menganalisis sentimen terhadap film horor Indonesia, memperoleh akurasi 82,51%, yang menunjukkan kemampuan SVM dalam menangani sentimen yang terkait dengan industri hiburan dan pengaruhnya di media sosial [11].

Terakhir, penelitian oleh Dhea Ananda dan Ryan Randy Suryono (2024) mengkaji sentimen publik terhadap kedatangan pengungsi Rohingya di Indonesia, dengan menggunakan dataset Twitter yang berjumlah 3350 tweet. Mereka menemukan bahwa SVM lebih unggul dibandingkan dengan metode *Naive Bayes*, dengan akurasi mencapai 76%. Penelitian ini menunjukkan bahwa SVM memiliki keunggulan dalam menganalisis isu sosial yang kompleks dan kontroversial, seperti pengungsi, dengan lebih efektif dibandingkan dengan algoritma lain yang lebih sederhana [12].

Meskipun penelitian-penelitian ini menunjukkan keunggulan SVM dalam berbagai topik sosial, tidak ada penelitian yang secara spesifik mengkaji sentimen masyarakat terhadap penggunaan sepeda listrik untuk anak-anak. Oleh karena itu, penelitian ini bertujuan untuk mengisi gap ini dengan fokus pada analisis persepsi publik tentang penggunaan sepeda listrik di media sosial X. Dengan mengembangkan model analisis yang lebih mendalam, penelitian ini diharapkan dapat memberikan kontribusi yang signifikan terhadap pemahaman masyarakat mengenai transportasi ramah lingkungan dan manfaat serta risiko yang terkait dengan penggunaan sepeda listrik, terutama bagi anak-anak, yang menjadi target utama dari studi ini.

2. METODOLOGI PENELITIAN

Pada penelitian ini ada beberapa tahapan yang perlu penulis lakukan untuk menyelesaikan proses analisis sentimen, Setiap tahapan memiliki peran penting dalam memastikan kualitas data yang digunakan dan keakuratan model yang dihasilkan. Alur tahapannya secara lengkap dapat dilihat pada Gambar 1.



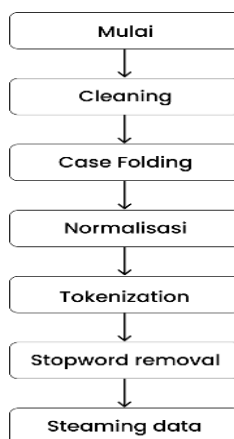
Gambar 1. Tahapan Penelitian

2.1 Pengumpulan Data

Pada tahapan pengumpulan data, dilakukan *crawling* melalui media sosial X menggunakan *tools* Tweet Harvest. Tweet Harvest tersebut digunakan untuk melakukan pengambilan data di media sosial X dengan memanfaatkan *Application Programming Interface (API)* [13]. Kata kunci pencarian yang digunakan dalam pengumpulan data ini adalah “sepeda listrik anak.” Namun, karena kebijakan terbaru dari X yang membatasi jumlah data yang dapat diambil dalam satu waktu, proses *crawling* tidak dapat langsung mengumpulkan data dalam jumlah ribuan seperti sebelumnya. Untuk mengatasi hal ini, pengumpulan data dilakukan secara bertahap, yaitu dengan mengambil data pada interval 1–2 minggu setiap bulannya. Setelah data terkumpul, dilakukan penggabungan secara manual menggunakan Microsoft Excel untuk menghasilkan satu *file* dan akan disimpan dengan format CSV (*Comma Separated Values*), yang kemudian akan diproses lebih lanjut pada tahap berikutnya.

2.2 Preprocessing

Preprocessing adalah tahapan dalam pengolahan data yang bertujuan untuk penyeleksian dan pembersihan data yang tidak diperlukan sebelum di proses ke tahap selanjutnya [14]. *Preprocessing* dilakukan dengan cara eliminasi data yang tidak sesuai dan diubah dalam bentuk yang akan lebih dipahami oleh sistem sehingga proses *mining* nya akan berjalan lebih efektif dan efisien dengan data yang sudah bersih [15]. Berikut merupakan tahapan *preprocessing* yang ditunjukkan pada Gambar 2.



Gambar 2. Tahapan *Preprocessing*

- Cleaning* merupakan tahapan awal dalam preprocessing yang berguna dalam pembersihan data yang tidak diperlukan untuk siap dianalisis. Pembersihan data ini menghilangkan data duplikat, URL, html, emoji, angka, dan tanda baca [12].
- Case folding* merupakan proses menyeragamkan bentuk huruf dan mengubah semua huruf kapital menjadi huruf kecil [16].
- Normalisasi* merupakan proses mengubah kata yang awalnya tidak baku menjadi kata baku yang sesuai dengan KBBI.
- Tokenization* merupakan proses pemisahan teks menjadi bagian yang lebih kecil seperti frasa, kalimat, atau kata [17].
- Stopword Removal* merupakan tahapan untuk menghapus kata-kata tidak relevan untuk proses klasifikasi, seperti ke, dan, atau, dari, siapa, dan lain sebagainya, guna meningkatkan efisiensi analisis [18].
- Steaming* adalah Proses untuk menyaring kata-kata dengan menghapus elemen seperti kata sambung, kata ganti, atau kata depan, serta mengubah kata menjadi bentuk dasarnya dengan menghilangkan imbuhan di awal atau akhir [19].

2.3 Labeling

Pelabelan merupakan tahapan untuk memberikan kategori positif, negatif maupun netral pada dataset, yang bertujuan memudahkan dalam analisis dan pengolahan data selanjutnya [6]. pelabelan ini menggunakan *textblob* dalam menganalisis sentimen dan NLP, karena kemampuannya mengevaluasi polaritas (positif, netral, negatif) serta subjektivitas teks melalui kamus bawaan, sambil mendukung berbagai fungsi tambahan seperti tokenisasi, analisis frasa, dan penerjemahan bahasa.

2.4 TF-IDF

Setelah melakukan tahapan *preprocessing*, tahapan selanjutnya yaitu pembobotan menggunakan *Term Frequency Inverse Document Frequency (TF-IDF)* dengan penggabungan 2 cara perhitungan bobot kata. TF-IDF merupakan suatu metode pembobotan yang memberikan bobot di setiap kata pada dokumen menurut *frekuensi term (TF)* dan *invers frekuensi document (IDF)* [6]. TF-IDF terdiri dari dua komponen utama, yaitu TF yang menunjukkan data yang

lebih dominan pada dokumen, dan IDF yang merepresentasikan kebalikan dari frekuensi dokumen tempat kata tersebut ditemukan [20]. Tujuan dari pembobotan ini untuk memberikan nilai frekuensi kata sebagai bobot yang akan di proses ke tahap selanjutnya.

$$W_{tf_{t,d}} = \begin{cases} 1 + \log_{10} tf_{t,d}, & \text{if } tf_{t,d} > 0 \\ 0, & \text{if } tf_{t,d} = 0 \end{cases} \quad (1)$$

$$idf_t = \log_{10} \frac{N}{df_t} \quad (2)$$

$$W_{t,d} = W_{tf_{t,d}} \cdot idf_t \quad (3)$$

Bobot kata di setiap dokumen dilambangkan dengan $W_{tf_{t,d}}$ di mana $tf_{t,d}$ merepresentasikan jumlah kemunculan suatu term dalam dokumen. Simbol N menunjukkan jumlah keseluruhan dokumen, sedangkan D merupakan jumlah dokumen yang mengandung term tertentu. Nilai idf menggambarkan bobot inverse dalam nilai df dan $W_{t,d}$ digunakan untuk pembobotan TF-IDF.

2.5 SMOTE

SMOTE (*Synthetic Minority Over-sampling Technique*) merupakan metode untuk mengatasi ketidakseimbangan kelas dan *dataset* [12]. Metode ini bekerja untuk menambahkan sampel sintetik untuk kelas minoritas agar jumlah data seimbang dengan kelas mayoritas.

2.6 Support Vector Machine (SVM)

SVM adalah algoritma yang berguna untuk memprediksi kategori data dengan melakukan pemberian label berdasarkan kategori yang termasuk di dalamnya [8]. Hasil tingkat akurasi model yang diperoleh melalui penerapan SVM sangat dipengaruhi oleh jenis fungsi kernel serta parameter yang diterapkan, di mana SVM berdasarkan karakteristiknya terbagi menjadi dua jenis, yaitu *linear* dan *non-linear*, *SVM linear* memanfaatkan *hyperplane* dengan *soft margin* untuk memisahkan dua kelas secara langsung dalam ruang data *linier* [21]. Penerapan fungsi *hyperline* pada data digunakan untuk pemisahan setiap kelas dengan tingkat dan akurasi yang besar menggunakan bahasa pemrograman Python di google colab. *Hyperplane* berfungsi sebagai batas matematis yang membagi data ke dalam ruang fitur yang mewakili masing-masing kelas, sehingga memudahkan klasifikasi data yang kompleks [22].

2.7 Evaluasi

Evaluasi adalah tahapan terakhir dalam proses menganalisis data untuk mengukur seberapa baik model dapat memprediksi sentimen pada teks dengan menggunakan *confusion matrix* yang akan menghasilkan akurasi, *recall*, *precision* dan *F1-Score/F-Measure* [20]. *Confusion matrix* merupakan tabel yang berguna untuk mengukur hasil dari prediksi apakah hasil tersebut termasuk ke dalam *true positive*, *true negative*, *false positive* dan *false negative* [21]. *Confusion matrix* dan rumusnya dapat dilihat pada Tabel 1 berikut ini:

Tabel 1. *Confusion matrix*

		Prediksi	
		Positif	Negatif
Aktual	Positif	True Positive (TP)	False Negative (FN)
	Negatif	False Positive (FP)	True Negative (TN)

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN} \times 100\% \quad (4)$$

$$Precision = \frac{TP}{TP+FP} \times 100\% \quad (5)$$

$$Recall = \frac{TP}{TP+FN} \times 100\% \quad (6)$$

$$F1-Score = 2 \times \frac{(precision \times recall)}{precision + recall} \times 100\% \quad (7)$$

3. HASIL DAN PEMBAHASAN

3.1 Pengumpulan Data

Data dalam penelitian ini diperoleh melalui metode *crawling* menggunakan Python di platform X, yang dilakukan dalam rentang waktu Januari 2024 hingga selesai. Dari total 3.565 data yang dikumpulkan, hanya kolom *full text* yang



diambil, karena kolom ini berisi ulasan dan komentar yang paling penting untuk analisis sentimen. Data mencakup pengalaman masyarakat tentang penggunaan sepeda listrik untuk anak-anak, termasuk persepsi mereka mengenai aspek seperti performa, desain, kenyamanan, dan keamanan. Berikut contoh *dataset* dari proses pengumpulan data dapat di lihat pada Tabel 2:

Tabel 2. Hasil Pengumpulan Data

No	Hasil Pengumpulan Data
1	Tanyarl Di sini ada yang ahli sepeda listrik? Bantu jelasin dong. Kira-kira dari spesifikasi 2 sepeda ini mana yang lebih bagus. : https://t.co/2EIkY1YyC3
2	ada yg pernah long trip bawa sepeda listrik? https://t.co/MU02ISoIqC
...	...
3564	ANGGOTA SATLANTAS POLRES ACEH TIMUR HIMBAU PENGGUNA SEPEDA LISTRIK Masyarakat diharapkan menggunakan sepeda listrik dengan bijak sesuai ketentuan yang berlaku. @DivHumas_Polri @HumasPoldaAceh https://t.co/KtmgVmks7N https://t.co/HJdN7BardZ
3565	Apakah Sepeda Listrik Boleh Digunakan di Jalan Raya? Begini Aturannya #TempoOtomotif

3.2 Preprocessing

a. *Cleaning*

Pada proses ini, data yang masih mentah dilakukan verifikasi untuk menghapus data yang tidak relevan. Dari total 3.565 data yang dikumpulkan, hanya ditemukan satu data duplikat, yang kemudian dihapus. Selain itu, elemen-elemen yang dianggap tidak penting seperti URL, HTML, emoji, angka, dan tanda baca juga dihapus untuk memastikan kebersihan data. Semua elemen yang telah dibersihkan disimpan dalam kolom *cleaning*. Hasil data sebelum dan sesudah proses *cleaning* dapat dilihat dalam Tabel 3:

Tabel 3. Hasil *Cleaning*

<i>Tweet</i>	<i>Cleaning</i>
Akhir2 ini sering liat bocil2 pake sepeda listrik di jalan raya tanpa pengawasan orang tua. Kadang beriringan sama teman2nya. Ngebut pula. Sepeda lipat saya kebalap. Bagi saya alasannya demikian: 1. Orang tuanya tau tapi membiarkan. Daripada tantrum di rumah.	Akhir ini sering liat bocil pake sepeda listrik di jalan raya tanpa pengawasan orang tua Kadang beriringan sama temannya Ngebut pula Sepeda lipat saya kebalap Bagi saya alasannya demikian Orang tuanya tau tapi membiarkan Daripada tantrum di rumah

b. *Case Folding*

Case Folding dilakukan untuk menyeragamkan semua huruf dalam data menjadi huruf kecil. Langkah ini memastikan konsistensi dalam analisis sentimen, sehingga pengolahan data dapat lebih akurat. Hasil data sebelum dan sesudah proses *case folding* dapat dilihat dalam Tabel 4 :

Tabel 4. Hasil *Case Folding*

<i>Cleaning</i>	<i>Case Folding</i>
Akhir ini sering liat bocil pake sepeda listrik di jalan raya tanpa pengawasan orang tua Kadang beriringan sama temannya Ngebut pula Sepeda lipat saya kebalap Bagi saya alasannya demikian Orang tuanya tau tapi membiarkan Daripada tantrum di rumah	akhir ini sering liat bocil pake sepeda listrik di jalan raya tanpa pengawasan orang tua kadang beriringan sama temannya ngebut pula sepeda lipat saya kebalap bagi saya alasannya demikian orang tuanya tau tapi membiarkan daripada tantrum di rumah

c. Normalisasi

Normalisasi merupakan proses untuk mengubah kata tidak baku dalam teks menjadi bentuk kata baku. Proses ini dilakukan dengan menggunakan kamus baku yang diperoleh dari sumber internet, yang kemudian disesuaikan berdasarkan data awal yang dikumpulkan. Kamus baku yang digunakan berisi sekitar 15.186 entri, yang membantu memastikan konversi kata menjadi bentuk baku sesuai standar bahasa. Langkah ini meningkatkan akurasi analisis sentimen dengan menjaga konsistensi penggunaan bahasa dalam *dataset*. Hasil data sebelum dan sesudah proses Normalisasi dapat dilihat dalam Tabel 5:

Tabel 5. Hasil Normalisasi

<i>Case Folding</i>	<i>Normalisasi</i>
akhir ini sering liat bocil pake sepeda listrik di jalan raya tanpa pengawasan orang tua kadang beriringan sama temannya ngebut pula sepeda lipat saya kebalap bagi saya alasannya demikian orang tuanya tau tapi membiarkan daripada tantrum di rumah	akhir ini sering lihat anak pakai sepeda listrik di jalan raya tanpa pengawasan orang tua kadang beriringan sama temannya laju pula sepeda lipat saya kebalap bagi saya alasannya demikian orang tuanya tau tapi membiarkan daripada tantrum di rumah

d. *Tokenization*

Tokenization dilakukan setelah proses normalisasi untuk memecah teks yang sudah diubah menjadi bentuk baku menjadi unit kata individu. Langkah berfungsi untuk mempersiapkan data agar dapat dianalisis dalam analisis sentimen. Hasil data sebelum dan sesudah proses *Tokenization* dapat dilihat dalam Tabel 6 :

Tabel 6. Hasil *Tokenization*

Normalisasi	Tokenizatio
akhir ini sering lihat anak pakai sepeda listrik di jalan raya tanpa pengawasan orang tua kadang beriringan sama temannya laju pula sepeda lipat saya kebalap bagi saya alasannya demikian orang tuanya tau tapi membiarkan daripada tantrum di rumah	['akhir', 'ini', 'sering', 'lihat', 'anak', 'pakai', 'sepeda', 'listrik', 'di', 'jalan', 'raya', 'tanpa', 'pengawasan', 'orang', 'tua', 'kadang', 'beriringan', 'sama', 'temannya', 'laju', 'pula', 'sepeda', 'lipat', 'saya', 'kebalap', 'bagi', 'saya', 'alasannya', 'demikian', 'orang', 'tuanya', 'tau', 'tapi', 'membiarkan', 'daripada', 'tantrum', 'di', 'rumah']

e. *Stopword Removal*

Stopword Removal dilakukan setelah proses *Tokenization* untuk menghapus kata-kata umum yang tidak memiliki makna yang signifikan dalam analisis sentimen, seperti “dan,” “di,” “ke,” “adalah,” dan sebagainya. Hasil data sebelum dan sesudah proses *Stopword Removal* dapat dilihat dalam Tabel 7:

Tabel 7. Hasil *Stopword Removal*

Tokenization	Stopword Removal
['akhir', 'ini', 'sering', 'lihat', 'anak', 'pakai', 'sepeda', 'listrik', 'di', 'jalan', 'raya', 'tanpa', 'pengawasan', 'orang', 'tua', 'kadang', 'beriringan', 'sama', 'temannya', 'laju', 'pula', 'sepeda', 'lipat', 'saya', 'kebalap', 'bagi', 'saya', 'alasannya', 'demikian', 'orang', 'tuanya', 'tau', 'tapi', 'membiarkan', 'daripada', 'tantrum', 'di', 'rumah']	['lihat', 'anak', 'pakai', 'sepeda', 'listrik', 'jalan', 'raya', 'pengawasan', 'orang', 'tua', 'kadang', 'beriringan', 'temannya', 'laju', 'sepeda', 'lipat', 'kebalap', 'alasannya', 'orang', 'tuanya', 'tau', 'membiarkan', 'tantrum', 'rumah']

f. *Stemming*

Stemming merupakan proses yang bertujuan untuk mengubah kata berimbuhan menjadi kata dasar. Proses ini dilakukan untuk mengurangi variasi kata dalam analisis sentimen. Hasil stemming dapat dilihat dalam Tabel 8:

Tabel 8. Hasil *Stemming*

Stopword Removal	Stemming
['lihat', 'anak', 'pakai', 'sepeda', 'listrik', 'jalan', 'raya', 'pengawasan', 'orang', 'tua', 'kadang', 'beriringan', 'temannya', 'laju', 'sepeda', 'lipat', 'kebalap', 'alasannya', 'orang', 'tuanya', 'tau', 'membiarkan', 'tantrum', 'rumah']	lihat anak pakai sepeda listrik jalan raya awas orang tua kadang iring teman laju sepeda lipat balap alas orang tua tau biar tantrum rumah

3.3 *Labeling*

Setelah melalui proses *preprocessing*, data masuk ke tahap *labeling*, yang bertujuan untuk menentukan jenis sentimen dalam ulasan yang dianalisis. Data yang digunakan berasal dari kolom *Stemming*, yang merupakan hasil akhir dari proses *preprocessing* dan berisi teks ulasan yang telah dibersihkan dari berbagai elemen tidak relevan. Sebelum proses *labeling* dilakukan, semua ulasan diterjemahkan ke dalam bahasa Inggris agar dapat dianalisis menggunakan *library TextBlob*, yang efektif dalam analisis sentimen bahasa Inggris. Hasil dari penerjemahan ini dapat dilihat pada Tabel 9:

Tabel 9. Hasil Penerjemahan Teks

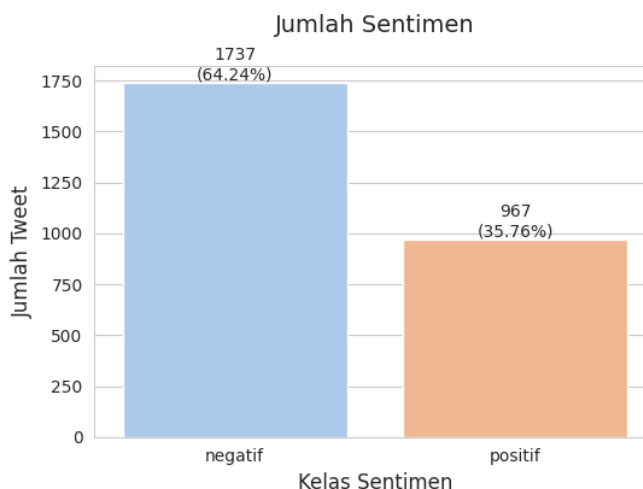
Stemming	Translate_tweet
guna sepeda listrik banget regulasi cepat anak pakai sepeda listrik jalan raya seram	Using electric bicycles is really fast regulations for children using electric bicycles on scary roads
untuk orang tua negara cinta tolong deh anak kasih sepeda listrik ngundang celaka mental anak stabil kendara tolol	For parents, country of love, please give your child an electric bicycle, it's inviting mental harm to the child's stable, stupid ride
ayo toyy sepeda listrik bareng haha semangat nyupirin orang achristyjkt	Let's play electric bikes together haha, let's get excited about inspiring the people of Achristyjkt
ahli sepeda listrik bantu jelasin spesifikasi sepeda bagus	Electric bicycle experts help explain the specifications of a good bicycle

Dalam tahap *labeling*, analisis sentimen dilakukan untuk mengklasifikasikan ulasan menjadi dua kategori: positif dan negatif. Proses ini dilakukan berdasarkan nilai polaritas yang dihitung oleh *library TextBlob*. Jika nilai polaritas positif (lebih besar dari 0), maka teks diklasifikasikan sebagai sentimen positif. Sebaliknya, jika nilai polaritas kurang dari atau sama dengan 0, maka teks dikategorikan sebagai sentimen negatif. Tabel 10 menunjukkan distribusi hasil labelling data setelah analisis dilakukan.

Tabel 10. Hasil Labeling Data

<i>translate_tweet</i>	<i>sentiment_polarity</i>	<i>sentiment</i>
<i>Using electric bicycles is really fast regulations for children using electric bicycles on scary roads</i>	-0.15	Negatif
<i>For parents, country of love, please give your child an electric bicycle, it's inviting mental harm to the child's stable, stupid ride</i>	-0.1333333333333333	Negatif
<i>hello friends, then the Permenhub article numbers electric bicycles for special lanes for highway driving areas, yes</i>	0.3583333333333333	Positif
<i>look for recommendations for electric bicycles</i>	0.7	Positif

Setelah dilakukannya proses *labelling*, proses selanjutnya adalah melakukan distribusi sentimen. Sentimen yang akan didistribusikan ditampilkan dalam bentuk visualisasi *sns plot*, sehingga dapat dilihat dengan jelas perbedaan jumlah data berdasarkan sentimen yang ada. Hasil visualisasi ini dapat dilihat pada Gambar 3 :



Gambar 3. Hasil Visualisasi *Labeling*

Didalam visualisasi Gambar 3 diatas menunjukkan terdapat sentimen yang paling unggul adalah negatif sebanyak 1737 (64,24%), dan sentimen positif sebanyak 967 (35,76%).

3.4 TF-IDF

Tahapan TF-IDF ini dilakukan karena data hasil dari *labeling* menghasilkan *imbalance* data, maka perlu dilakukan *handling* data agar distribusinya lebih seimbang. Data teks yang berupa *string* perlu diubah menjadi numerik menggunakan metode TF-IDF, proses TF-IDF itu sendiri memiliki proses yaitu, mengubah data yang awalnya berupa teks kemudian diubah menjadi vektor dengan cara menghitung frekuensi kata dan mengurangi bobot kata yang sering muncul di seluruh dokumen.

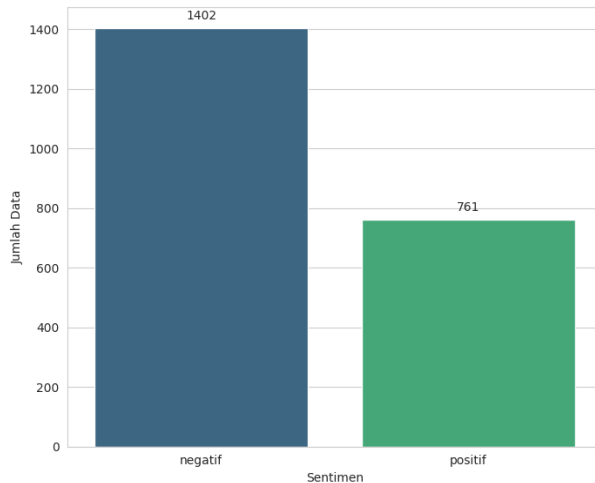
Tabel 11. Hasil TF-IDF

Kata	TF	IDF	TF-IDF
Bicycle	0.054836	2.067673	0.113382
The	0.059257	1.887145	0.111826
Bike	0.039623	2.446088	0.096921
An	0.042243	2.222003	0.093864
Bicycles	0.039337	2.323606	0.091403
...
Installation	0.000049	8.209710	0.000404
Welcome	0.000049	8.209710	0.000404

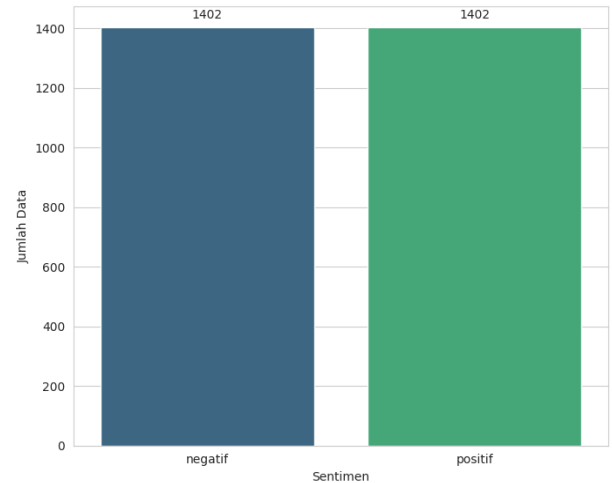
Pada Tabel 11 di atas menampilkan hasil pembobotan kata menggunakan TF-IDF, di mana beberapa kata dengan nilai TF-IDF tinggi seperti "bicycle" dan "bike" dianggap lebih relevan dan berperan sebagai kata kunci dalam analisis teks. Sebaliknya, kata dengan nilai TF-IDF rendah, seperti "installation" dan "sewing," cenderung memiliki signifikansi yang lebih kecil.

3.5 SMOTE

Pada tahap ini, karena distribusi data pelabelan menunjukkan adanya ketidakseimbangan data, perlu dilakukan teknik *oversampling* untuk menghindari performa model yang tidak optimal. Grafik pertama menunjukkan jumlah data sebelum dilakukan SMOTE, di mana kelas negatif memiliki jumlah yang jauh lebih tinggi (1402 data) dibandingkan kelas positif (761 data). Ketidakseimbangan ini dapat mengakibatkan model cenderung bias terhadap kelas mayoritas.



Gambar 4a. Distribusi Kelas Sebelum SMOTE



Gambar 5b. Distribusi Kelas Sebelum SMOTE

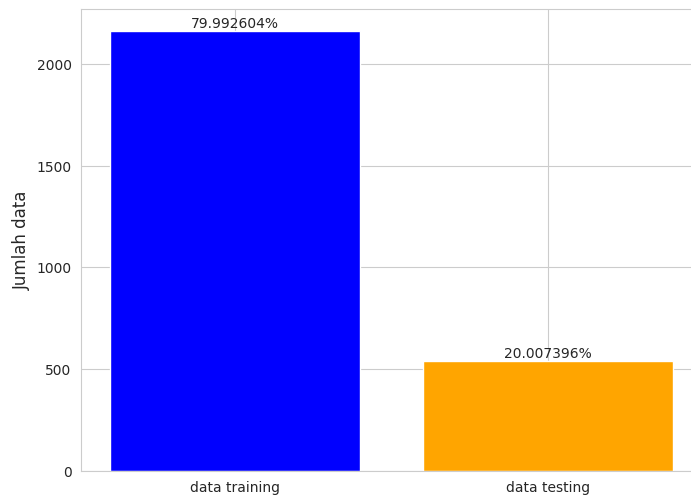
Pada Gambar 4a, menunjukkan grafik dataset yang tidak seimbang, dengan kelas negatif yang lebih tinggi sebesar 1402 data dan mendominasi kelas positif yang hanya sebesar 761 data. Untuk mengatasi hal ini, dilakukan *oversampling* menggunakan metode SMOTE, yang bertujuan meningkatkan jumlah data pada kelas positif hingga mendekati jumlah data pada kelas negatif. Grafik kedua memperlihatkan hasil setelah proses SMOTE, di mana jumlah data pada kedua kelas kini seimbang (masing-masing 1402 data). Proses ini dilakukan dengan membuat sampel sintetis pada kelas positif sehingga jumlahnya menjadi sama dengan kelas negatif.

3.6 WordCloud

Data penelitian saat ini sudah dikelompokkan ke dalam label positif dan negatif, dan akan melalui proses pembuatan *wordcloud* untuk menghasilkan representasi visual dari teks yang terdapat dalam data. *Wordcloud* adalah sebuah alat yang digunakan untuk memvisualisasikan teks, di mana kata-kata dalam teks ditampilkan dalam bentuk grafis dengan ukuran huruf yang mencerminkan frekuensi kemunculannya. Dengan kata lain, kata-kata yang muncul lebih sering dalam data akan memiliki ukuran huruf yang lebih besar dalam *wordcloud*. Alat ini berfungsi untuk mempermudah pengamatan dan representasi kata-kata dari setiap segmen, sehingga kata-kata yang dominan dapat diidentifikasi dengan lebih jelas. Berikut adalah hasil *wordcloud* berdasarkan sentimen data penelitian dapat di lihat pada Gambar 5:



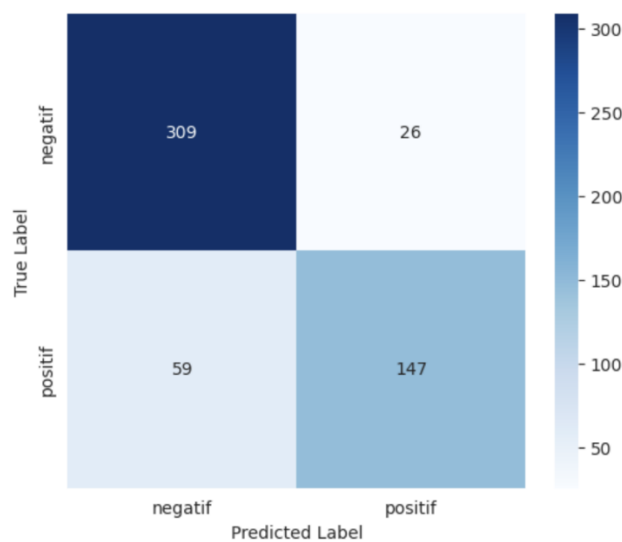
Gambar 5. Wordcloud Sentimen Negatif



Gambar 8. Jumlah Data Training dan Testing

3.8 Evaluasi Kinerja Model

Setelah melalui tahapan preprocessing, pelabelan data, dan penerapan TF-IDF, dataset kemudian diproses dengan menggunakan model SVM dengan pendekatan kernel linear. Proses klasifikasi ini bertujuan untuk menganalisis sentimen terkait penggunaan sepeda listrik pada anak-anak di media sosial, khususnya pada platform X. *Confusion matrix* digunakan untuk mengevaluasi performa model dalam mengukur nilai *accuracy*, *precision*, *recall*, dan *f1-score*.



Gambar 9. SVM Confusion Matrix

Berdasarkan Gambar 9 dapat disimpulkan bahwa sentimen masyarakat terhadap penggunaan sepeda listrik untuk anak cenderung negatif. Selanjutnya, kita dapat menjelaskan hasil prediksi pada masing-masing label sebagai berikut:

- Pada label negatif, sebanyak 309 data terklasifikasi sebagai negatif secara benar, sementara ada 26 data sebagai positif yang salah diklasifikasikan.
- Pada label positif, sebanyak 147 data terklasifikasi sebagai positif secara benar, namun ada 59 data sebagai negatif yang salah diklasifikasikan.

```

Classification Report:
              precision    recall  f1-score   support

   Negative      0.84      0.92      0.88       335
   Positive      0.85      0.71      0.78       206

 accuracy              0.84       541
 macro avg              0.84       541
 weighted avg          0.84       541
    
```

Gambar 10. Clasification Report

Berdasarkan Gambar 10 yang menampilkan *classification report*, performa model SVM pada dataset ini dapat dijelaskan sebagai berikut:

- Accuracy* model mencapai 0,84 yang menunjukkan bahwa model secara keseluruhan dapat mengklasifikasikan data dengan cukup baik.
- Precision* untuk label *negative* adalah 0,84 yang berarti bahwa dari semua data yang diprediksi sebagai *negative*, 0,84 di antaranya benar-benar *negative*. Sementara pada label *positive*, *precision* tercatat 0,85, menunjukkan model cukup tepat dalam mengidentifikasi data yang benar-benar *positive*.
- Recall* untuk label *negative* tercatat 0,92, yang berarti bahwa model berhasil mengidentifikasi 0,92 dari seluruh data yang seharusnya *negative*. Namun, *recall* untuk label *positive* lebih rendah, yaitu 0,71, yang menunjukkan bahwa ada sejumlah data *positive* yang tidak teridentifikasi dengan baik oleh model.
- F1-score*, yang menggabungkan *precision* dan *recall*, untuk label *negative* adalah 0,88, menunjukkan performa yang cukup baik dalam mengklasifikasikan sentimen *negative*. Sedangkan untuk label *positive*, *f1-score* tercatat 0,78, yang mengindikasikan adanya potensi peningkatan pada klasifikasi label *positive*.

Secara keseluruhan, model SVM ini menunjukkan performa yang solid dengan nilai *macro average* untuk *precision* 0,84, *recall* 0,82, dan *f1-score* 0,83. Hasil ini membuktikan bahwa model lebih efektif dalam mengklasifikasikan sentimen negatif dibandingkan positif, terutama pada aspek *recall* untuk label positif yang sedikit lebih rendah.

4. KESIMPULAN

Hasil analisis menggunakan metode SVM pada opini masyarakat terkait penggunaan sepeda listrik untuk anak-anak di media sosial X menunjukkan bahwa sentimen negatif mendominasi dengan persentase 64,24%, sementara sentimen positif mencapai 35,76%. Proses analisis melibatkan tahap pembersihan dan penyesuaian data, pelabelan sentimen, serta pembobotan menggunakan TF-IDF dan penanganan ketidakseimbangan data dengan SMOTE. Evaluasi model menunjukkan nilai *recall* tertinggi sebesar 0,92 untuk sentimen negatif, sedangkan *precision* dan *F1-score* untuk sentimen positif masing-masing 0,85 dan 0,78. Hasil ini mengindikasikan bahwa SVM efektif dalam mengklasifikasikan opini masyarakat terhadap penggunaan sepeda listrik untuk anak-anak. Namun, penelitian ini memiliki beberapa keterbatasan, seperti penggunaan data hanya dari media sosial X yang mungkin tidak mewakili opini seluruh masyarakat, serta pengklasifikasian sentimen yang terbatas pada dua kategori (positif dan negatif) tanpa mempertimbangkan sentimen netral atau lebih kompleks. Selain itu, terdapat kemungkinan bias dalam pemilihan fitur dan proses pelabelan data. Penelitian lanjutan diharapkan dapat mengatasi keterbatasan ini dengan memperluas cakupan dataset dari berbagai platform media sosial, menggunakan metode lain, dan mengklasifikasikan sentimen yang lebih variatif. Dengan demikian, hasil penelitian ini dapat memberikan kontribusi dalam pengembangan model analisis sentimen yang lebih akurat dan mendukung pengambilan keputusan yang lebih efektif terkait penggunaan sepeda listrik di Indonesia.

REFERENCES

- P. Maramis, R. A. Maramis, and R. Dientje, "Penegakan Hukum Penggunaan Sepeda Listrik oleh Anak di Bawah Umur di Jalan Raya Sesuai Hukum Positif di Indonesia," *Univ. Sam Ratulangi_Fakultas Huk.*, vol. 2, no. 8, 2023.
- F. Nurfadlilah, A. Sofi Malika, A. Rifqi Naufal, and R. Wikansari, "Peran Sepeda Listrik dalam Mewujudkan Mobilitas Berkelanjutan," *Humanit. J. Humaniora, Sos. dan Bisnis*, vol. 2, no. 1, pp. 136–141, 2024, [Online]. Available: <https://humanisa.my.id/index.php/hms/article/view/91>
- H. Lesmana and A. Fithry, "Pengaturan dan Perlindungan Hukum terhadap Penggunaan Sepeda Listrik di Jalan Raya Indonesia," *Pros. SNAPP Sos. Humaniora, Pertanian, Kesehat. dan Teknol.*, vol. 2, no. 1, pp. 109–113, 2024, doi: 10.24929/snapp.v2i1.3180.
- Y. D. R. PUSPARISA, "Sepeda Listrik Banyak Dikendarai Anak di Jalan Raya, Bagaimana Aturannya?," *kompas.id*. Accessed: Nov. 20, 2024. [Online]. Available: <https://www.kompas.id/baca/ekonomi/2024/07/21/sepeda-listrik-banyak-dikendarai-anak-anak-bagaimana-ideal-penggunaannya>
- A. Fithry, Sjaifurrachman, and R. W. S. Sakinah, "Edukasi Pendampingan Bahaya Penggunaan Sepeda Listrik pada Anak," *Pros. SNAPP Sos. Humaniora, Pertanian, Kesehat. dan Teknol.*, pp. 433–438, 2023.
- M. A. Syahira and R. Kurniawan, "Analisis Sentimen Cyberbullying pada Media Sosial X Menggunakan Metode Support Vector Machine," *J. Media Inform. Budidarma*, vol. 8, no. 3, p. 1724, 2024, doi: 10.30865/mib.v8i3.7926.
- S. W. Iriananda, R. W. Budiawan, A. Y. Rahman, and I. Istiadi, "Optimasi Klasifikasi Sentimen Komentar Pengguna Game Bergerak Menggunakan Svm, Grid Search Dan Kombinasi N-Gram," *J. Teknol. Inf. dan Ilmu Komput.*, vol. 11, no. 4, pp. 743–752, 2024, doi: 10.25126/jtiik.1148244.
- A. Syahri, A. Angraini, and F. Muttakin, "Sentimen Analisis pada Ulasan Aplikasi Ajaib di Google Play Store dengan Algoritma Support Vector Machine," *INOVTEK Polbeng - Seri Inform.*, vol. 9, no. 1, pp. 257–271, 2024, doi: 10.35314/isi.v9i1.4047.
- N. Hendrastuty, A. Rahman Isnain, and A. Yanti Rahmadhani, "Analisis Sentimen Masyarakat terhadap Program Kartu Prakerja pada Twitter dengan Metode Support Vector Machine," *J. Inform. J. Pengemb. IT*, vol. 6, no. 3, pp. 150–155, 2021.
- M. D. Al Fahreza, A. Luthfiarta, M. Rafid, M. Indrawan, and A. Nugraha, "Analisis Sentimen: Pengaruh Jam Kerja terhadap Kesehatan Mental Generasi Z," *J. Appl. Comput. Sci. Technol.*, vol. 5, no. 1, pp. 16–25, 2024, doi: 10.52158/jacost.v5i1.715.
- J. E. Br Sinulingga and H. C. K. Sitorus, "Analisis Sentimen Opini Masyarakat terhadap Film Horor Indonesia Menggunakan



- Metode SVM dan TF-IDF,” *J. Manaj. Inform.*, vol. 14, no. 1, pp. 42–53, 2024, doi: 10.34010/jamika.v14i1.11946.
- [12] D. Ananda and R. R. Suryono, “Analisis Sentimen Publik terhadap Pengungsi Rohingya di Indonesia dengan Metode Support Vector Machine dan Naïve Bayes,” *J. Media Inform. Budidarma*, vol. 8, no. 2, p. 748, 2024, doi: 10.30865/mib.v8i2.7517.
- [13] S. A. Putra and A. Wijaya, “Analisis Sentimen Artificial Intelligence (AI) pada Media Sosial Twitter Menggunakan Metode Lexicon-Based,” *JuSiTik J. Sist. dan Teknol. Inf. Komun.*, vol. 7, no. 1, pp. 21–28, 2023, doi: 10.32524/jusitik.v7i1.1042.
- [14] F. Amaliah and I. K. Dwi Nuryana, “Perbandingan Akurasi Metode Lexicon-Based dan Naive Bayes Classifier pada Analisis Sentimen Pendapat Masyarakat Terhadap Aplikasi Investasi pada Media Twitter,” *J. Informatics Comput. Sci.*, vol. 3, no. 03, pp. 384–393, 2022, doi: 10.26740/jinacs.v3n03.p384-393.
- [15] R. N. R. dan A. S. K. Suropto, “Teknik pre-processing dan classification dalam data science,” Binus University. [Online]. Available: <https://mie.binus.ac.id/2022/08/26/teknik-pre-processing-dan-classification-dalam-data-science/>
- [16] D. Darwis, E. S. Pratiwi, and A. F. O. Pasaribu, “Penerapan Algoritma SVM untuk Analisis Sentimen pada Data Twitter Komisi Pemberantasan Korupsi Republik Indonesia,” *EduTic - Sci. J. Informatics Educ.*, vol. 7, no. 1, pp. 1–11, 2020, doi: 10.21107/edutic.v7i1.8779.
- [17] S. T. M. Yessi Asri, M. K. Dr.Dra. Dwina Kuswardani, M. kom. Listra Frigia missianes Horhoruw, S.Kom., and S. Siti Aisyah ramadhana, *MACHINE LEARNING & DEEP LEARNING Analisis Sentimen Menggunakan Ulasan Pengguna Aplikasi*. Jawa timur: Uwais Inspirasi Indonesia, 2019.
- [18] A. Reandito, I. Sumantri, M. Fatchan, and T. N. Wiyatno, “Analisis Sentimen Produk Makanan Jepang di Indonesia pada Twitter Menggunakan Naïve Bayes,” *J. Ilm. Tek. Inform. dan Sist. Inf.*, 2024.
- [19] J. Supriyanto, D. Alita, and A. R. Isnain, “Penerapan Algoritma K-Nearest Neighbor (K-NN) untuk Analisis Sentimen Publik terhadap Pembelajaran Daring,” *J. Inform. dan Rekayasa Perangkat Lunak*, vol. 4, no. 1, pp. 74–80, 2023, doi: 10.33365/jatika.v4i1.2468.
- [20] C. Zai and A. R. Isnain, “Komparasi Algoritma Naïve Bayes dan Support Vector Machine (SVM) pada Analisis Sentimen Capcut,” *J. Process.*, vol. 18, no. 2, pp. 272–285, 2023, doi: 10.33998/processor.2023.18.2.897.
- [21] 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, p. 31, 2021, doi: 10.33365/jdmsi.v2i1.1021.
- [22] Y. Laia, S. Sandino Berutu, el Pieter Sumihar, and H. Budiati, “Implementasi Library Textblob dan Metode Support Vector Machine Pada Analisis Sentimen Pelanggan Terhadap Jasa Transportasi Online,” *Technol. Sci.*, vol. 6, no. 1, pp. 1–10, 2024, doi: 10.47065/bits.v6i1.5090.