ANALISIS SENTIMEN KOMENTAR YOUTUBE TERKAIT DUGAAN KORUPSI PENGADAAN CHROMEBOOK MENGGUNAKAN MACHINE LEARNING

Sigit Wibisono
Teknik Informatika, Itnstitut Teknologi Budi Utomo,
wsigitwibisono@gmail.com

Abstrak
Pertumbuhan pesat media sosial telah menjadikan YouTube sebagai platform penting untuk mengekspresikan opini publik tentang berbagai isu, termasuk kasus dugaan korupsi yang melibatkan pengadaan Chromebook. Volume komentar yang tidak terstruktur membuat analisis manual menjadi tidak efisien, sehingga muncul kebutuhan akan pendekatan berbasis pembelajaran mesin untuk melakukan analisis sentimen otomatis. Studi ini bertujuan untuk menganalisis komentar YouTube yang terkait dengan kasus dugaan korupsi pengadaan Chromebook dan membandingkan kinerja algoritma Naive Bayes, Support Vector Machine (SVM), dan Regresi Logistik untuk klasifikasi sentimen. Penelitian ini menggunakan pendekatan kuantitatif komparatif menggunakan data yang dikumpulkan melalui YouTube Data API. Dataset menjalani beberapa tahap pra-pemrosesan, termasuk pembersihan, case folding, normalisasi kata slang, tokenisasi, penghapusan stopword, dan stemming. Ekstraksi fitur dilakukan menggunakan metode TF-IDF, diikuti oleh pelabelan dan klasifikasi sentimen menggunakan tiga algoritma pembelajaran mesin. Kinerja model dievaluasi menggunakan metrik akurasi, presisi, recall, dan F1-score. Hasil menunjukkan bahwa di antara 455 komentar YouTube yang dianalisis, sentimen positif dan netral lebih dominan daripada sentimen negatif. Ketiga algoritma mencapai akurasi yang sama yaitu sekitar 51,69%; namun, Regresi Logistik menunjukkan kinerja keseluruhan yang unggul berdasarkan presisi dan skor F1, menjadikannya model paling efektif untuk klasifikasi sentimen dalam penelitian ini. Temuan ini menunjukkan bahwa kualitas fitur dan keseimbangan kelas tetap menjadi faktor penting untuk meningkatkan kinerja analisis sentimen. Di sini Anda diminta untuk menjelaskan hal yang telah dilakukan, hasil utama dan kesimpulan makalah saudara secara jelas dan singkat dalam bahasa Inggris.
Kata Kunci: Analisis Sentimen, YouTube, Machine Learning, Naive Bayes, Support Vector Machine, Regresi Logistik, TF-IDF.

1. PENDAHULUAN

Perkembangan teknologi informasi dan media sosial telah mengubah cara masyarakat dalam menyampaikan opini terhadap berbagai isu publik. Youtube sebagai salah satu platform berbagi video tidak hanya berfungsi sebagai media hiburan, tetapi juga menjadi ruang diskusi terbuka bagi masyarakat. Melalui kolom komentar, pengguna dapat mengekspresikan pendapatnya terhadap suatu peristiwa, termasuk isu dugaan korupsi pengadaan Chromebook. Permasalahan yang muncul adalah komentar komentar tersebut memiliki karakteristik yang kompleks, tidak terstruktur, serta menggunakan bahasa yang beragam seperti bahasa informal, singkatan dan ungkapan sarkasme. Selain itu, jumlah data yang besar serta variasi sentimen yang muncul, baik positif, negatif, maupun netral, menyebabkan proses analisis secara manual menjadi tidak efektif dan berpotensi menimbulkan bias. Di sisi lain, meskipun penelitian analisis sentimen pada media sosial telah banyak dilakukan, masih terdapat keterbatasan dalam studi yang secara komprehensif membandingkan performa algoritma klasifikasi pada konteks komentar Youtube, khususnya yang berkaitan dengan isu dugaan korupsi.

Beberapa Algoritma berbasis machine learning dapat digunakan sebagai pendekatan dari berbagai permasalahan opini yang berkembang, Metode ini memungkinkan sistem untuk

mengklasifikasikan teks secara otomatis berdasarkan pola yang dipelajari dari data. Beberapa algoritma yang umum digunakan dalam klasifikasi teks yaitu diantarana adalah **Naive Bayes, Logistic Regression** dan **Support Vector Machine (SVM)**. Masing-masing algoritma tersebut memiliki karakteristik yang berbeda dalam hal melakukan pendekatan atau menangani data teks. Sehingga diperlukan analisis kuantitatif untuk mengevaluasi dan membandingkan kinerja masing-masing algoritma dalam mengolah data komentar Youtube berdasarkan metrik evaluasi tertentu.

Sebagai cakupan dari penelitian ini untuk melakukan analisis, dibatasi pada opini sentimen terhadap komentar pada Youtub, terkait dugaan korupsi pengadaan Chromebook. Dari hasil perhitungan maka dapat diketahui perbandingan performa algoritma **Naive Bayes, Logistic Regression dan Support Vector Machine** dalam proses klasifikasi sentimen. Hasil penelitian ini diharapkan dapat memberikan gambaran mengenai algoritma yang paling efektif dalam mengklasifikasikan sentimen komentar, serta menjadi referensi bagi penelitian selanjutnya dalam bidang analisis sentimen pada media sosial.

2. METODOLOGI

2.1 Algoritma Naive Bayes

Adalah merupakan metode klasifikasi untuk memprediksi kelas yang ada pada suatu dokumen dengan menghitung probabilitas dari setiap kelas pada kejadian masa lampau. Memiliki asumsi yang kuat (naif) terhadap kejadian-kejadian yang ada merupakan ciri khas dari metode ini. Keuntungan dari penggunaan algoritma Naive Bayes adalah metodenya yang sederhana dan cepat karena memerlukan data latih yang sedikit dalam proses klasifikasi teks (Suryani, Linawati, & Saputra, 2019).

Peneliti memilih Algoritma Naive Bayes karena merupakan suatu metode yang sederhana dan intuitif yang kinerjanya mirip dengan pendekatan lain. Hasil yang diberikan oleh Naive Bayes ini cukup akurat dan dapat bekerja dengan efisien (Murnawan & Sinaga, 2017). Keuntungan dari penggunaan algoritma Naive Bayes adalah metode ini hanya membutuhkan jumlah data latih yang kecil untuk menentukan estimasi parameter yang diperlukan dalam proses pengklasifikasian. Naive Bayes sering bekerja jauh lebih baik dalam kebanyakan situasi dunia nyata yang kompleks dari pada yang diharapkan (Saleh, 2015).

Metode Naive Bayes ini akan digunakan untuk mencari sentimen positif dan negatif pada analisa sentimen Youtube. Adapun rumus Teorema Bayes yang digunakan pada Algoritma Naive Bayes Classifier sebagai berikut:

$$P(H|X) = \frac{P(X|H)P(H)}{P(X)}$$

(Sumber: Introduction to Information Retrieval. Cambridge University Press.(2008)

X = Data yang kelasnya belum diketahui

H = Hipotesis data X

P(H|X) = Probabilitas hipotesis H berdasarkan kondisi X

P(H) = Probabilitas hipotesis H

P(X|H) = Probabilitas hipotesis X berdasarkan kondisi H

P(X) = Probabilitas X

2.2 Algoritma SVM

Support Vector Machine (SVM) adalah algoritma supervised learning yang dapat digunakan untuk klasifikasi dan regresi. SVM bekerja dengan mencari hyperplane terbaik untuk memisahkan data dari berbagai kelas dengan margin maksimal. Dalam analisis sentimen, SVM mencoba menemukan hyperplane yang secara optimal memisahkan teks positif, negatif dan netral secara matematis, SVM menyelesaikan permasalahan optimasi dengan cara berikut:

$$W \cdot X + b = 0$$

(Sumber: The Nature of Statistical Learning Theory. Springer-Verlag. (1995)

ISSN 2406-9841
e-IISN 2986-2973

Dengan bobot vektor direpresentasikan sebagai W yaitu:

$$W = w1 ,w\ 2,\ldots\ wn$$

(Sumber: *The Nature of Statistical Learning Theory. Springer-Verlag*. (1995)

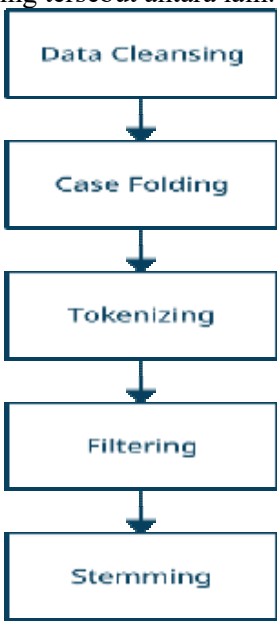
Tuple pelatihan dilambangkan sebagai X, skalar direpresentasikan sebagai b. Hyper plane mengoptimalkan data dengan meminimalkan transformasi masalah ||W|| yang dapat dihitung secara matematis.

2.3 Logistic Regression

Logistic Regression merupakan model statistik yang dalam bentuk dasarnya menggunakan fungsi logistik untuk memodelkan variabel biner (Tolles dan Meurer,2016) Model logistik biner memiliki variabel dependen dengan dua kemungkinan nilai, pada penelitian ini berupa valid / hoaks yang diwakili oleh label '1' dan '0'. 14 Logaritma peluang nilai dengan label '1' merupakan kombinasi linier dari variabel independent yang disebut prediktor. Fungsi logistic mengubah logaritma peluang menjadi probabilitas. Model Logistic Regression hanya memodelkan probabilitas output dan tidak melakukan klasifikasi secara statistik, meski dapat digunakan untuk membuat classifier dengan memilih nilai cutoff dan mengklasifikasikan input dengan probabilitas (Walker dan Duncan, 1967). Logistic Regression dapat diaplikasikan dalam berbagai bidang, termasuk pembelajaran mesin. Dalam pembelajaran mesin, Logistic Regression digunakan untuk menemukan hubungan antara fitur dan probabilitas prediksi (Agrawal, 2017). Model logistik seringkali digunakan sebagai meta learner dalam ensemble learning masalah klasifikasi karena cukup sederhana dan mampu menghasilkan intrepretasi dari prediksi yang dibuat oleh base learners (Brownlee, 2020).

2.4Text Preprocessing

Setelah pengumpulan dan pemilihan data dilakukan, maka akan dilaksanakan text preprocessing dengan tahapan yang terdapat pada **Gambar 2.1** Tahapan text preprocessing tersebut antara lain:



Gambar 1 Tahapan Text Processing
(Sumber: Dokumen Peneitian)

1. Data Cleansing Sebagai tahapan pertama dalam text preprocessing, data cleansing bertujuan untuk membersihkan dokumen dari kata atau karakter yang tidak diperlukan dalam proses analisis. Umumnya, elemen yang dihilangkan meliputi tag HTML, tanda pagar atau hashtag, username, URL, serta emoticon. (Putra, 2018). Penjelasan tentang metode penelitian yang digunakan dalam penelitian.

2. Case Folding Tahap kedua dalam text preprocessing adalah case folding, yaitu proses konversi seluruh huruf dalam data teks menjadi bentuk huruf kecil (lowercase). Proses ini bertujuan untuk menyeragamkan penulisan kata sehingga menghasilkan data teks yang lebih konsisten dan terstruktur (Feldman &Sanger, 2007).

3. Tokenizing Setelah diperoleh data teks yang lebih terstruktur dan konsisten melalui tahap case folding, lalu dilanjutkan dengan tahap tokenizing yang berfungsi untuk memecah teks yang semula dalam bentuk kalimat menjadi unit-unit kata yang disebut

token. Hasil dari proses ini akan digunakan pada tahap pengolahan teks berikutnya (Aggarwal & Zhai, 2012).

4. Filtering Kata (token) yang dihasilkan dari tahap tokenizing kemudian disaring melalui proses filtering. Tahap ini bertujuan untuk menghilangkan kata-kata umum (stopwords) yang tidak memiliki makna penting dalam proses analisis, seperti kata “dan”, “di”, “yang”, “ke”, serta kata-kata umum lainnya. Dengan demikian, hanya kata-kata yang dianggap penting dan relevan yang akan digunakan pada tahap selanjutnya (Putra, 2018).

5. Stemming Tahap stemming dilakukan untuk mengubah kata yang memiliki imbuhan menjadi bentuk kata dasarnya. Proses ini bertujuan untuk mengurangi variasi bentuk kata sehingga kata-kata yang memiliki makna sama dapat dipresentasikan dalam satu bentuk dasar. Sebagai contoh, kata “membeli”, “dibeli” dan “pembelian” akan diubah menjadi kata dasar “beli” (Nazief & Adriani, 1996).

3. ANALISIS DAN PEMBAHASAN

3.1 Web Scrapping Komentar Video Youtube

Web Scraping merupakan teknik pengumpulan data secara otomatis dari sebuah situs web menggunakan program komputer. Dalam penelitian ini, web scraping digunakan untuk mengambil data komentar dari video Youtube yang berkaitan dengan topik penelitian. Data komentar tersebut kemudian digunakan sebagai dataset untuk proses analisis sentimen. Proses pengambilan komentar dilakukan dengan memanfaatkan Youtube Data API atau pustaka pemrograman tertentu yang memungkinkan sistem untuk mengakses dan mengekstraksi informasi dari halaman Youtube. Informasi yang diambil biasanya meliputi isi komentar, waktu komentar diposting, serta informasi tambahan lainnya yang relevan. Komentar yang diperoleh dari proses scraping kemudian disimpan dalam bentuk dataset, misalnya dalam format CSV atau DataFrame. Dataset ini selanjutnya akan digunakan dalam tahap pengolahan data, mulai dari text preprocessing, ekstraksi fitur menggunakan TF-IDF, algoritma machine learning. hingga proses klasifikasi menggunakan Tahap web scraping sangat penting karena menyediakan raw data atau data mentah, yang berasal langsung dari pengguna Youtube. Data ini mencerminkan opini atau reaksi masyarakat terhadap suatu topik, sehingga ada relevansinya untuk digunakan dalam penelitian analisis sentimen.

3.2 Pendaftaran API Key Youtube

Cara Membuat API Key Youtube (Youtube Data API): API Key digunakan untuk mengakses layanan Youtube Data API agar aplikasi dapat mengambil data seperti video, channel, komentar, dan informasi lainnya dari platform Youtube.

1. Masuk ke Google Cloud Console Buka halaman Google Cloud Console kemudian login menggunakan akun Google.

2. Membuat Project Baru a. Klik menu Select Project pada bagian atas halaman. b. Pilih New Project. c. Masukkan nama project, misalnya Youtube API Project. d. Klik tombol Create.

3. Mengaktifkan Youtube Data API a. Masuk ke menu APIs & Services. 48 b. Pilih Library. c. Cari Youtube Data API v3. d. Klik API tersebut, kemudian tekan tombol Enable untuk mengaktifkan layanan API.

4. Membuat API Key a. Masuk ke menu APIs & Services → Credentials. b. Klik tombol + Create Credentials. c. Pilih API Key. d. Sistem akan secara otomatis menghasilkan API Key baru.

5. Menyimpan API Key API Key yang berhasil dibuat akan ditampilkan pada layar. Selanjutnya: a. Klik ikon Copy untuk menyalin API Key. b. Simpan API Key di tempat yang aman agar tidak disalahgunakan oleh pihak lain

.

ISSN 2406-9841
e-IISN 2986-2973

Gambar 2 Pembuatan API Key pada Google Cloud Console Platform (Sumber: Dokumentasi Penelitian)

3.3 Pembuatan Fungsi Web Scrapping

Pada program ini merupakan implementasi bagaimana cara pengambilan data komentar pada Youtube menggunakan Youtube Data API v3 yang dirancang untuk memperoleh dataset lengkap berupa komentar utama dan balasannya dari suatu video. Dengan memanfaatkan pustaka googleAPIclient.discovery, program akan melakukan autentikasi menggunakan API key dan mengirimkan permintaan bertingkat yang mencakup parameter snippet dan replies untuk mengekstrak informasi penting seperti waktu publikasi, nama pengguna, teks komentar dan jumlah like. Mekanisme paginasi dengan nextPageToken diterapkan untuk menjamin tidak ada data yang terlewat, sehingga menghasilkan kumpulan data yang representatif. Hasil pengambilan data, yang disimpan dalam bentuk daftar, dapat langsung dikonversi ke dalam DataFrame pandas untuk memudahkan analisis lanjutan seperti analisis sentimen atau pemodelan topik, sehingga program ini menjadi fondasi penting dalam proses pengumpulan data untuk penelitian berbasis opini publik di platform Youtube.

3.4 Memanggil Fungsi Web Scrapping

Setelah fungsi video comments didefinisikan, tahap selanjutnya adalah mengeksekusinya dengan memberikan video_id dan API_key yang telah ditentukan. Hasil eksekusi berupa daftar mentah yang berisi seluruh komentar dan balasan kemudian dikonversi menjadi struktur DataFrame menggunakan pustaka pandas. Proses konversi ini memberikan format tabular dengan kolom publishedAt, authorDisplayName, textDisplay dan likeCount, yang memudahkan proses penelitian untuk melakukan eksplorasi data, penyaringan, serta analisis lebih lanjut seperti statistik deskriptif atau visualisasi. Dengan demikian, data komentar Youtube yang semula tidak terstruktur berhasil diorganisasikan menjadi dataset yang siap digunakan dalam tahapan penelitian berikutnya.

1. Deskripsi Dataset Data yang berhasil dikumpulkan memiliki 463 baris dan 4 kolom dengan struktur sebagai berikut:

Tabel 3.1 Struktur Dataset Hasil Pengambilan Data Komentar YouTube (Sumber: Dokumentasi Penelitian)

2. Karakteristik Data a. Rentang waktu: komentar dikumpulkan dari Juli 2025 hingga Maret 2026. b. Distribusi like: mayoritas komentar memiliki 0 like, namun terdapat beberapa komentar dengan jumlah like cukup tinggi atau sekitar 43. c. Keberagaman konten: teks komentar mencakup berbagai topik terkait video, termasuk opini, kritik dan diskusi antar pengguna. Data dalam bentuk DataFrame ini siap digunakan untuk analisis lanjutan seperti pembersihan teks, analisis sentimen, pemodelan topik atau visualisasi tren opini publik.

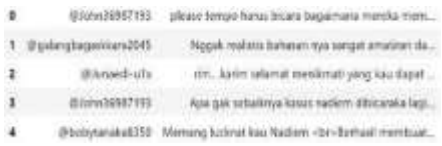
3.5 Pemilihan Kolom yang Digunakan

Setelah data berhasil dikonversi ke dalam DataFrame, langkah selanjutnya adalah menyederhanakan struktur data dengan hanya mengambil kolom yang relevan untuk analisis lebih lanjut. Pada tahap ini, kolom publishedAt dan likeCount tidak digunakan karena fokus analisis adalah pada konten teks

komentar dan identitas pengguna. Oleh karena itu, DataFrame dipilih hanya dua kolom, yaitu:

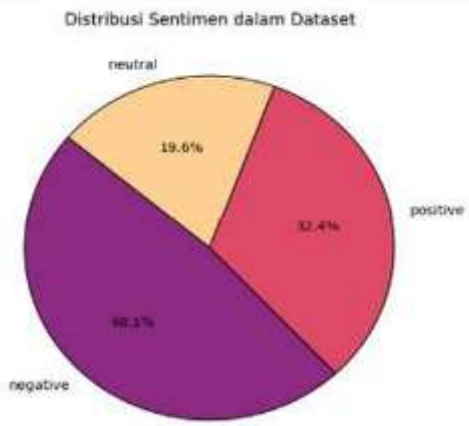
1. authorDisplayName: nama pengguna yang memberikan komentar.
2. textDisplay: isi teks komentar (masih dalam bentuk asli, termasuk tag HTML jika ada).

Tabel 1 Tampilan DataFrame Hasil Pemilihan Kolom
(Sumber: Dokumentasi Penelitian)



(Sumber: Dokumentasi Penelitian)

3.6 Visualisasi Data



Gambar 3. Pie Chart Distribusi Sentimen
(Sumber: Dokumentasi Penelitian)

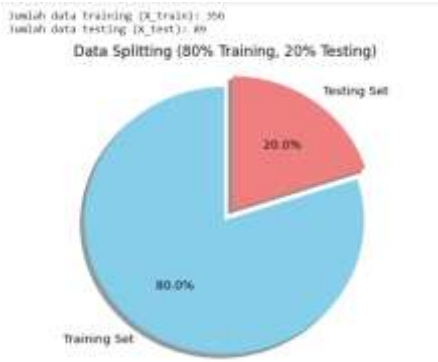
Dari Gambar 3.2 diagram tersebut dapat diamati bahwa sentimen negatif mendominasi dengan hampir setengah dari total komentar (48,1%), disusul sentimen positif sebesar 32,4% dan sentimen netral hanya 19,6%. Distribusi yang timpang ini mengindikasikan bahwa mayoritas komentar cenderung bersifat kritis atau tidak puas terhadap topik yang dibahas dalam video. Kondisi ini perlu menjadi perhatian dalam pemodelan klasifikasi karena adanya ketidakseimbangan kelas (imbalanced class) yang dapat menyebabkan model bias terhadap kelas negatif jika tidak ditangani dengan teknik seperti resampling atau penggunaan metrik evaluasi yang tepat (misalnya F1-score). Dengan mengetahui distribusi ini, peneliti dapat merancang strategi pra-pemrosesan dan pemodelan yang lebih sesuai untuk menghasilkan analisis sentimen yang akurat.

3.7 Pembagian Data Training dan Data Testing

Tahap awal dalam pemodelan klasifikasi adalah mempersiapkan data teks yang telah melalui proses stemming menjadi representasi numerik yang dapat diproses oleh algoritma machine learning. Pada langkah ini, digunakan TF-IDF Vectorizer untuk mengubah kolom steming_data yang berisi teks hasil stemming menjadi matriks fitur numerik (X). Proses fit-transform dilakukan pada seluruh data untuk menghasilkan vektor fitur yang merepresentasikan pentingnya setiap kata terhadap dokumen secara keseluruhan. Target klasifikasi ditetapkan sebagai kolom polarity (y) yang berisi label sentimen (misalnya positif, negatif, netral). Selanjutnya, data dibagi menjadi dua bagian menggunakan train_test_split dengan proporsi 80% data latih (training) dan 20% data uji (testing) serta nilai random_state=42 untuk memastikan reproduksibilitas. Hasil pembagian menunjukkan: 1. Jumlah data training: X_train.shape[0] sampel 2. Jumlah data testing: X_test.shape[0] sampel Visualisasi menggunakan diagram lingkaran (pie chart) memperjelas distribusi tersebut, di mana bagian terbesar (80%) digunakan untuk melatih model, sedangkan 20% sisanya disisihkan untuk menguji kinerja model pada data yang belum pernah dilihat sebelumnya. Pembagian ini penting untuk menghindari overfitting dan memastikan model mampu menggeneralisasi pola dari data baru. Dengan demikian, data siap digunakan untuk membangun dan mengevaluasi model klasifikasi seperti Logistic

ISSN 2406-9841
e-IISN 2986-2973

Regression, SVM atau Naive Bayes pada tahap berikutnya.



Gambar 4 Pie Chart Pembagian Data Testing dan Data Training
(Sumber: Dokumentasi Penelitian)

Diagram lingkaran di atas menunjukkan proporsi pembagian data yang digunakan dalam proses pelatihan dan pengujian model klasifikasi sentimen. Berdasarkan visualisasi, data dibagi menjadi dua bagian:

Tabel 2 Klasifikasi Sentimen
(Sumber: Dokumentasi Penelitian)

Kategori	Persentase
Training Set	80,0%
Testing Set	20,0%

Dari diagram ini dapat diamati bahwa 80% dari total data dialokasikan sebagai data latih (training set), sementara 20% digunakan sebagai data uji (testing set). Pembagian ini mengikuti praktik umum dalam machine learning di mana proporsi 80:20 digunakan untuk memastikan model memiliki cukup data untuk belajar pola (training) serta data yang representatif untuk mengevaluasi kinerjanya pada data yang belum pernah dilihat sebelumnya (testing). Dengan pemisahan ini, dapat mengukur akurasi, presisi, recall dan F1-score model secara objektif, serta menghindari overfitting karena evaluasi dilakukan pada data yang tidak dilibatkan dalam proses pelatihan.

3.8 World Cloud

Word Cloud adalah memvisualisasikan kata-kata yang paling sering muncul pada setiap kategori sentimen menggunakan word cloud. Teknik ini bertujuan untuk memberikan gambaran intuitif mengenai topik atau istilah khas yang mendominasi komentar bersentimen positif, negatif maupun netral. Proses diawali dengan memisahkan teks berdasarkan label sentimen (polarity), kemudian menggabungkan seluruh kata pada kolom stopword removal menjadi satu string utuh untuk setiap kelas. Fungsi plot_wordcloud dibuat dengan parameter lebar, tinggi dan skema warna yang seragam, serta dilengkapi pemeriksaan untuk menghindari kesalahan jika tidak ada data pada suatu sentimen. Hasil visualisasi berupa tiga word cloud terpisah: Positif, Negatif dan Netral, yang masing-masing menampilkan 100 kata teratas. Word cloud sentimen positif cenderung menonjolkan kata-kata bernada apresiatif atau dukungan, sentimen negatif memperlihatkan kritik atau kekecewaan, sedangkan sentimen netral memuat kata-kata faktual atau pertanyaan. Dengan demikian, teknik ini tidak hanya memperkaya analisis eksploratif, tetapi juga membantu peneliti memahami secara visual perbedaan karakteristik linguistik antar sentimen.

Gambar 5 Hasil Word Cloud Sentimen Positif
(Sumber: Dokumentasi Penelitian)

ISSN 2406-9841
e-IISN 2986-2973

4. KESIMPULAN

Berdasarkan hasil penelitian yang telah dilakukan mengenai Analisis Sentimen Komentar Youtube terkait Dugaan Korupsi Pengadaan Chromebook menggunakan Machine Learning, maka dapat ditarik kesimpulan sebagai berikut:

1. Karakter Sentimen Komentar Berdasarkan hasil analisis terhadap 455 data komentar Youtube, diketahui bahwa sentimen yang muncul terdiri dari tiga kategori, yaitu positif, negatif dan netral. Distribusi sentimen menunjukkan bahwa sentimen positif dan netral lebih dominan dibandingkan sentimen negatif. Hal ini mengindikasikan bahwa opini masyarakat terhadap isu dugaan korupsi pengadaan Chromebook tidak hanya didominasi oleh kritik, tetapi juga mencerminkan adanya dukungan serta komentar yang bersifat informatif.

2. Performa Algoritma Klasifikasi Hasil evaluasi menunjukkan bahwa algoritma Naive Bayes, Support Vector Machine (SVM) dan Logistic Regression memiliki performa yang relatif serupa dengan nilai akurasi sebesar 51,69%. Namun demikian, berdasarkan matrix precision, recall dan F1-score, performa ketiga model masih tergolong rendah, khususnya dalam mengklasifikasikan sentimen netral. Hal ini menunjukkan bahwa model belum mampu secara optimal membedakan antar kelas sentimen.

3. Algoritma Terbaik Berdasarkan hasil perbandingan metrik evaluasi, algoritma Logistic Regression merupakan metode yang memberikan performa terbaik dibandingkan dengan Naive Bayes dan SVM. Hal ini ditunjukkan oleh nilai precision dan F1-score yang lebih tinggi, sehingga Logistic Regression dinilai lebih efektif dalam melakukan klasifikasi sentimen pada penelitian ini.

5. DAFTAR PUSTAKA

1. Absharina, E. D., & Nurqolbiah, F. (2025). Sentiment analysis of YouTube video comments review of Starlink Indonesia by Gadgetin using VADER and TextBlob. *The Indonesian Journal of Computer Science, 14*(3). https://doi.org/10.33022/ijcs.v14i3.4686
2. Aiswarya, A. S., & Rajeev, H. (2024). YouTube comment sentimental analysis. *Indian Journal of Data Mining, 4*(1), 5–8. https://doi.org/10.54105/ijdm.A1633.04010524
3. Bird, S., Klein, E., & Loper, E. (2009). *Natural language processing with Python*. O'Reilly Media.
4. Cox, R. (2007). Regular expression matching can be simple and fast. Google Research. https://swtch.com/~rsc/regexp/regexp1.html
5. Feldman, R., & Sanger, J. (2007). *The text mining handbook: Advanced approaches in analyzing unstructured data*. Cambridge University Press.
6. Hidayat, R., Arifiyanti, A., & Kartika, D. S. (2024). Analisis sentimen komentar YouTube konferensi tingkat tinggi G20 menggunakan metode Naive Bayes. *Jurnal Teknologi dan Sistem Informasi Bisnis, 6*(2), 282–285. https://doi.org/10.47233/jteksis.v6i2.1263
7. Ismardani, T., & Fatah, Z. (2025). Sentiment analysis of YouTube user comments on government policies using the Naïve Bayes method. *Journal of Data Insights, 3*(2), 113–123. https://doi.org/10.26714/jodi.v3i2.876
8. Khomsah, S. (2024). Sentiment analysis on YouTube comments using Word2Vec and Random Forest. *Jurnal Telematika dan Teknologi Informasi, 18*(1). https://doi.org/10.31315/telematika.v18i1.4493
9. Manning, C. D., Raghavan, P., & Schütze, H. (2008). *Introduction to information retrieval*. Cambridge University Press.

ISSN 2406-9841
e-IISN 2986-2973

10. Mahfudza, N., & Ihksan, M. (2025). Sentiment analysis of YouTube comments on Indonesian presidential candidates. *Jurnal Riset Komputer (JURIKOM), 12*(2), 1–9. https://doi.org/10.30865/jurikom.v12i2.8538
11. Nazief, B., & Adriani, M. (1996). Confix-stripping approach to stemming algorithm for Indonesian language. *Proceedings of the International Conference on Computational Linguistics*.
12. Pasaribu, R., & Putra, I. A. G. S. (2025). Analisis sentimen ulasan aplikasi dengan Multinomial Naïve Bayes, Logistic Regression, dan SVM. *Jurnal Nasional Teknologi Informasi dan Aplikasinya, 4*(1), 38–50. https://doi.org/10.24843/JNATIA.2025.v04.i01.p08
13. Prastyo, D., Irawan, D., & Mursyidin, I. H. (2024). Klasifikasi sentimen komentar YouTube dengan NLP pada debat Pilkada Banten 2024. *Bit-Tech Journal, 7*(2), 1833. https://doi.org/10.32877/bt.v7i2.1833
14. Sihombing, K. E. V., & Pakereng, M. A. I. (2025). Analisis sentimen komentar YouTube terhadap wawancara Presiden Prabowo menggunakan machine learning dan Orange Data Mining. *STORAGE: Jurnal Ilmiah Teknik dan Ilmu Komputer, 4*(4).
15. Siregar, S. C., Adek, R. T., & Fitri, Z. (2025). The sentiment analysis of comments on YouTube channel beauty vlogger in Indonesian language using Support Vector Machine method. *Proceedings of the International Conference on Multidisciplinary Engineering (ICOMDEN).*
16. The pandas development team. (2025). *pandas documentation*. https://pandas.pydata.org/docs/
17. Witten, I. H., Frank, E., & Hall, M. A. (2011). *Data mining: Practical machine learning tools and techniques* (3rd ed.). Morgan Kaufmann.