Analisis Sentimen Komentar Youtube terhadap Kebijakan Menteri Keuangan (*Purbaya Yudhi Sadewa*) Tidak Menaikan Pajak Menggunakan Model Regresi Logistic, Naïve Bayes, Support Vector Machine dan Random Forest

Muhammad Al-farisy
Teknik Informatika S-2, Program Pascasarjana, Universitas Pamulang, Kota Tangerang Selatan, Banten
m.alfarisy797@gmail.com

Abstrak--Kebijakan perpajakan sering memicu beragam respons masyarakat yang terekam dalam media sosial seperti YouTube, sehingga diperlukan pendekatan machine learning untuk menganalisis sentimen secara objektif. Penelitian ini bertujuan menganalisis kecenderungan sentimen masyarakat serta membandingkan kinerja algoritma Regresi Logistik, Naive Bayes, Support Vector Machine (SVM), dan Random Forest dalam mengklasifikasikan komentar YouTube terhadap kebijakan Menteri Keuangan yang tidak menaikkan pajak ke dalam sentimen positif, netral, dan negatif. Data diperoleh dari kolom komentar YouTube CNBC Indonesia dan diproses melalui tahapan data cleansing, pelabelan sentimen menggunakan VADER, preprocessing teks, serta ekstraksi fitur TF-IDF. Hasil analisis menunjukkan adanya ketidakseimbangan kelas dengan dominasi sentimen netral. Evaluasi model menggunakan akurasi, Macro F1-Score, dan Kurva ROC-AUC menunjukkan bahwa Regresi Logistik dan Naive Bayes memiliki akurasi tinggi namun bias terhadap kelas mayoritas, sementara SVM menunjukkan peningkatan performa dalam membedakan kelas sentimen. Random Forest menjadi model paling optimal dengan akurasi tertinggi sebesar 93,95%, Macro F1-Score 0,6181, serta nilai AUC ROC yang tinggi dan seimbang pada seluruh kelas sentimen, sehingga terbukti paling efektif dalam menganalisis sentimen komentar YouTube pada dataset tidak seimbang dan memberikan gambaran objektif mengenai respons publik terhadap kebijakan fiskal.

Kata Kunci : ***Klasifikasi Sentimen, Media Sosial, Ketidakseimbangan Kelas, TF-IDF, Algoritma Machine Learning***

I. PENDAHULUAN

Kebijakan fiskal, khususnya perpajakan, merupakan instrumen pemerintah yang sangat sensitif karena berdampak langsung pada kesejahteraan publik. Keputusan Menteri Keuangan untuk tidak menaikkan pajak memicu beragam opini masyarakat di kolom komentar YouTube. Mengingat volume data yang besar, analisis manual menjadi tidak efektif dan subjektif, sehingga diperlukan pendekatan analisis sentimen berbasis machine learning untuk mengolah opini tersebut secara sistematis.

Penelitian terdahulu oleh Khaidar (2025) membuktikan efektivitas Logistic Regression pada platform Instagram, namun belum memberikan gambaran komparatif terhadap algoritma lain pada platform YouTube yang memiliki karakteristik teks berbeda. Oleh karena itu, rumusan masalah penelitian ini difokuskan pada identifikasi kecenderungan sentimen publik serta penentuan algoritma yang paling optimal di antara Logistic Regression, Naive Bayes, Support Vector Machine (SVM), dan Random Forest. Tujuan penelitian adalah untuk menganalisis opini publik sekaligus membandingkan performa keempat algoritma tersebut guna menemukan model klasifikasi terbaik.

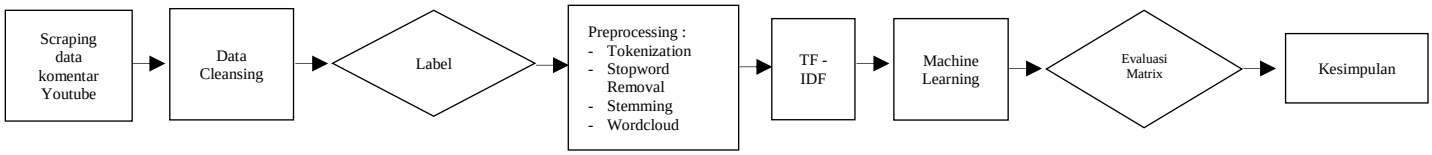
Pemilihan keempat algoritma ini didasarkan pada kemampuan masing-masing dalam menangani teks YouTube yang tinggi akan noise, seperti bahasa tidak baku dan typo. Naive Bayes unggul dalam efisiensi, SVM kuat pada dimensi tinggi, Logistic Regression pada interpretasi data, dan Random Forest dalam mencegah overfitting. Hasil penelitian ini diharapkan memberikan kontribusi akademik pada pengembangan metode NLP dan menjadi referensi berbasis data bagi pembuat kebijakan dalam memahami persepsi masyarakat secara objektif.

II. METODE PENELITIAN

Metode merupakan pendukung penelitian baik dari segi pengumpulan data, tempat penelitian, jenis dan sumber penelitian, tahapan penelitian, metode pengembangan, dan lainnya yang diperlukan dalam metode. Hindari menulis konsep keilmuan yang sudah umum, tinjauan pustaka serta definisi-definisi umum.

A. Tahapan Penelitian

Dalam penelitian ini, penetapan tahapan penelitian memerlukan perencanaan yang matang serta langkah-langkah yang terstruktur guna memastikan kelancaran pelaksanaannya. Tahapan penelitian tersebut disajikan pada Gambar 1.



Gambar 1. *Research Framework Flow Chart*

B. Dataset

Dataset yang digunakan dalam penelitian ini merupakan kumpulan komentar yang diekstraksi dari kanal YouTube CNBC Indonesia, khususnya dari video berjudul 'Menkeu Purbaya Ungkap Alasan Tak Naikkan Pajak'. Proses akuisisi data dilakukan melalui API YouTube, dengan memasukkan ID URL video ke dalam lingkungan Google Colaboratory dan memanfaatkan bahasa pemrograman Python. Setelah data berhasil diperoleh, dilanjutkan dengan tahapan pra-pemrosesan data menggunakan aplikasi Google Colaboratory.

C. Data Cleansing

Proses pembersihan data (data cleansing) merupakan tahapan krusial untuk memastikan kualitas dan konsistensi teks sebelum analisis lebih lanjut. Prosedur ini diimplementasikan menggunakan fungsi clean_text yang secara sistematis melakukan serangkaian operasi sebagai berikut:

1) *Case Folding*
   Tahap ini mengubah seluruh huruf kapital dalam teks menjadi huruf kecil secara konsisten. Tujuannya adalah untuk menyeragamkan representasi kata, menghindari perbedaan makna akibat kapitalisasi, dan mengurangi kompleksitas data untuk analisis.
2) *Removal of HTML Tag*
   Seluruh tag HTML, termasuk elemen <a> (hyperlink) dan tag generik lainnya, dihilangkan dari teks.
3) *Removal of URL Tag*
   Pola Uniform Resource Locator (URL) yang terdeteksi dalam teks dihapus untuk menghilangkan tautan eksternal.
4) *Removal of Repeated Words*
   Kata-kata yang muncul secara berurutan atau berdekatan akan dihapus untuk menghindari redundansi.
5) *Abbreviation Replacement*
   Kata 'de' yang berdiri sendiri diganti dengan 'dengan' untuk standar bahasa yang lebih baku.
6) *Removal of Emojis*
   Karakter emoji dihilangkan dari teks
7) *Removal of Number*
   Karakter numerik dihilangkan dari teks.

8) *Non-Alphanumeric Character Cleaning*
   Hanya karakter alfanumerik dan spasi yang dipertahankan, sementara karakter lain seperti simbol dan tanda baca dihapus.
9) *Removal of HTML Entities*
   Entitas HTML yang mungkin tersisa (misalnya , &, <) juga dibersihkan.
10) *Whitespace Normalization*
   Spasi berlebih dikurangi menjadi satu spasi tunggal, dan spasi di awal atau akhir teks dihilangkan (trimming).
11) *Unicode Normalization*
   Mengubah karakter Unicode khusus ke bentuk dasarnya dan membuang karakter non-ASCII.

12) *Removal of Numbers Attached to Words*
   Menghapus angka yang menempel langsung di belakang huruf.
13) *Custom Character Cleaning*
   Karakter spesifik seperti X1Â¹ yang diidentifikasi sebagai data noise juga dihapus.

D. Label

Proses pelabelan sentimen pada data teks dilakukan dengan memanfaatkan pendekatan leksikon berbasis aturan (rule-based lexicon approach) melalui pustaka *VADER (Valence Aware Dictionary and sEntiment Reasoner)* Sentiment. VADER dipilih karena kemampuannya dalam memahami nuansa sentimen dalam teks media sosial.

Implementasi diawali dengan inisialisasi Sentiment Intensity Analyzer dari VADER. Setiap komentar yang telah dibersihkan (*Dataframe*) kemudian dianalisis untuk menghasilkan skor polaritas. Skor compound VADER, yang merupakan metrik gabungan normalisasi dari polaritas kalimat secara keseluruhan, digunakan sebagai indikator utama.

E. Pre Processing Data

Untuk melakukan analisis sentimen, sebelumnya harus melalui tahap pra-pemrosesan teks (text preprocessing). Ini diperlukan untuk mengoptimalkan hasil dari analisis sentimen. Pra-pemrosesan teks dilakukan melalui 3 (tiga) tahapan utama, yaitu tokenisasi, penghapusan stopword, dan stemming. Tokenisasi memecah teks menjadi unit-unit linguistik diskrit (kata atau token) untuk memudahkan analisis. Setelah itu, penghapusan stopword dilakukan untuk mengeliminasi kata-kata umum yang memiliki kontribusi informatif minimal terhadap sentimen. Kemudian, stemming diterapkan untuk mereduksi setiap kata ke bentuk dasar atau akar katanya, sehingga mengelompokkan variasi kata yang memiliki makna leksikal serupa. Tahapan-tahapan ini secara kumulatif mempersiapkan teks untuk ekstraksi fitur dan analisis lebih lanjut.

1) Tokenization

Tahap ini dimulai dengan proses mengubah seluruh huruf dalam teks menjadi huruf kecil (convert to lowercase) secara konsisten. Langkah ini dilakukan untuk menghindari perbedaan makna yang disebabkan oleh penggunaan huruf kapital, sehingga seluruh kata dianggap seragam oleh sistem. Setelah itu, teks dipecah menjadi unit-unit linguistik diskrit (kata atau token) untuk memudahkan analisis.

2) Stopword Removal

Dilakukan untuk mengeliminasi kata-kata umum yang sering muncul tetapi tidak memiliki pengaruh signifikan terhadap sentimen, sehingga fokus analisis dapat diarahkan pada kata-kata kunci yang lebih bermuatan informasi.

3) Stemming

Merupakan proses mengubah kata ke bentuk dasarnya untuk mengurangi variasi kata dan meningkatkan konsistensi fitur, dengan mengelompokkan bentuk-bentuk infleksi kata ke dalam satu representasi tunggal.

4) Wordcloud

Wordcloud adalah representasi visual dari frekuensi atau kepentingan kata-kata dalam suatu teks. Dalam sebuah wordcloud, kata-kata yang muncul lebih sering dalam teks akan ditampilkan dengan ukuran font yang lebih besar dan/atau warna yang lebih menonjol, sementara kata-kata yang kurang sering muncul ditampilkan dengan ukuran yang lebih kecil

F. Model Machine Learning

Tahap pemodelan machine learning dilakukan untuk membangun model yang mampu mempelajari pola dari data teks yang telah melalui proses preprocessing. Pada tahap ini, algoritma machine learning digunakan untuk mengklasifikasikan sentimen berdasarkan fitur-fitur yang diekstraksi dari data. Pemodelan ini bertujuan untuk menghasilkan sistem yang dapat melakukan prediksi sentimen secara otomatis dan konsisten berdasarkan data pelatihan yang tersedia.

G. Evaluasi Kinerja ROC-AUC

Evaluasi kinerja menggunakan metode Receiver Operating Characteristic – Area Under Curve (ROC-AUC) dilakukan untuk mengukur kemampuan model dalam membedakan kelas sentimen secara menyeluruh. ROC-AUC dipilih karena mampu memberikan gambaran performa model yang lebih stabil dan objektif, terutama pada kondisi data yang tidak seimbang (imbalanced data). Nilai AUC yang tinggi menunjukkan bahwa model memiliki kemampuan klasifikasi yang baik dalam memisahkan kelas positif dan negatif.

H. Best Model Machine Learning

Pemilihan best model dilakukan berdasarkan hasil evaluasi kinerja yang telah diperoleh pada tahap sebelumnya. Model dengan nilai ROC-AUC tertinggi dipilih sebagai model terbaik karena dianggap memiliki performa paling optimal dalam melakukan klasifikasi sentimen. Tahap ini penting untuk memastikan bahwa model yang digunakan pada tahap implementasi merupakan model yang paling efektif dan andal.

I. Word Cloud

Tahap word cloud dilakukan untuk memvisualisasikan kata-kata yang paling sering muncul dalam data komentar berdasarkan hasil analisis sentimen. Visualisasi ini bertujuan untuk memberikan pemahaman yang lebih intuitif mengenai topik atau kata dominan yang sering dibahas oleh pengguna. Selain sebagai alat bantu interpretasi hasil, word cloud juga membantu dalam mendukung analisis kualitatif terhadap pola sentimen yang terbentuk.

III. HASIL DAN PEMBAHASAN

A. Dataset

Scraping data menggunakan Google Collabs dan menggunakan Bahasa *Python* serta menggunakan tools tambahan seperti API Youtube, Pandas dan Numpy untuk scraping data dari komentar Youtube.

	Komentar	Penulis	Jumlah Suka	Tanggal Publikasi
0	Avilini banyak bicara	@imah_73	0	2025-12-20T15:05:10Z
1	Avilini cocok jadi wAkil menkeu	@imah_73	0	2025-12-20T15:02:23Z
2	Perasaan Mentri yg cewe kemaren cuma ngomongin...	@SusanLesmana-w5n	0	2025-12-19T20:55:59Z
3	bismillah allhamdulillah baik kami rakyat dise...	@putriputri-f5i	0	2025-12-18T13:40:45Z
4	DARI THN PENJAJAHAN DI SEBUT MERDEKA DI ZAMAN ...	@Rubio-r2h	0	2025-12-18T01:20:38Z
5	T	@Rubio-r2h	0	2025-12-18T00:42:16Z
6	THN 2025 DI KEPEMIMPINAN PRABOWO SUBIYANTO DIS...	@Rubio-r2h	0	2025-12-18T00:41:17Z
7	HATI2 SETIAP JIWA MANUSIA DI BUMI DI KOLONG LA...	@Rubio-r2h	0	2025-12-18T00:33:21Z
8	THN 2025 BANGSA INDONESIA BERSATU WAJIB MAWAS ...	@Rubio-r2h	0	2025-12-18T00:26:21Z
9	THN 2025 BERPIKIR OTAKNYA SONDUL SAMPAI LANGIT...	@Rubio-r2h	0	2025-12-18T00:18:16Z

Gambar 2. Hasil *Scraping Dataset*

B. Data Cleansing

1) Case Folding

Hasil dari proses *Case Folding* pada kolom Komentar Clean secara konsisten mengubah seluruh karakter huruf kapital menjadi huruf kecil (lowercase) untuk menyeragamkan format teks. Implementasi ini terlihat pada transformasi kata-kata di awal kalimat maupun nama tokoh, seperti "Indonesia" menjadi "indonesia" dan "Pak Purbaya" menjadi "pak purbaya". Proses penyeragaman ini bertujuan untuk menghilangkan ambiguitas karakter.

	Komentar	Komentar Clean
7	Cerdas dan bahasa yg mudah dimengerti bg rakyat .. Kereen Pak Menkeu..👍👍👍	cerdas dan bahasa yang mudah dimengerti oleh rakyat keren pak menkeu
10	kecuali anda bayar ke saya SINDIRAN pedas Pak Purbaya😂🤷	kecuali anda bayar ke saya sindiran pedas pak purbaya
12	Pembangunan 2 ruang kelas baru, rehabilitasi 4 ruang kelas lama juga 2 ruang kamar mandi untuk guru juga pagar keliling sekolah di usulkan sekolah ku sejak 3 tahun lalu sampai sekarang belum juga terealisasi Pak, mudahan sekolah ku mendapat kan pembangunan yang diusulkan Karena sekolah ku hanya memiliki 4 ruang kelas sedangkan sekolah ada 6 rombongan kelas 🙏	pembangunan ruang kelas baru rehabilitasi ruang kelas lama juga ruang kamar mandi untuk guru juga pagar keliling sekolah diusulkan sekolahku sejak tahun lalu sampai sekarang belum terealisasi pak mudahan sekolahku mendapat pembangunan yang diusulkan karena sekolahku hanya memiliki ruang kelas sedangkan sekolah ada rombongan kelas
25	memimpin negeri ini teorinya mudah, tapi pelaksanaannya susah karena banyak kepentingan.. negara ini butuh sosok yg berani, dihormati dan disegani, untuk menertibkan tikus2 berbahaya..	memimpin negeri ini teorinya mudah tapi pelaksanaannya susah karena banyak kepentingan negara ini butuh sosok yang berani dihormati dan disegani untuk menertibkan tikus berbahaya
33	InsyaAllah Allah akan melindungi Pak Purbaya dan orang-orang yang berniat Mulia seperti yujarsip atau nabi yusuf ketika dijadikan Allah menjadi pengurus kekayaan Mesir dan membantu penguasa mesir Amenhotep bukankah sejarah berulang mungkin sampai hari Kiamat sesuai yang Allah kehendaki	insya allah akan melindungi pak purbaya dan yang berniat mulia seperti yujarsip atau nabi yusuf ketika dijadikan allah menjadi pengurus kekayaan mesir dan membantu penguasa mesir amenhotep bukankah sejarah berulang mungkin sampai hari kiamat sesuai yang allah kehendaki

Gambar 3. Hasil *Convert Text*

2) Removal of HTML Tags

Removal of HTML Tags pada kolom Komentar Clean berhasil mengeliminasi elemen-elemen sintaksis HTML yang tidak relevan, seperti tag tautan *(<a>)*, penanda baris baru *(
)*, dan tag dekorasi teks *()*. Proses ini secara efektif membersihkan teks dari kebisingan data (noise), sehingga memastikan analisis sentimen terfokus sepenuhnya pada konten tekstual yang bermakna tanpa terganggu oleh kode atau simbol teknis dari platform YouTube. Dengan hilangnya elemen-elemen tersebut, representasi teks menjadi lebih konsisten dan siap untuk diproses lebih lanjut oleh model machine learning.

	Komentar	Komentar Clean
7	Cerdas dan bahasa yg mudah dimengerti bg rakyat . Kereen Pak Menkeu..👍👍👍	cerdas dan bahasa yang mudah dimengerti oleh rakyat keren pak menkeu
10	kecuali anda bayar ke saya SINDIRAN pedas Pak Purbaya😂🤷	kecuali anda bayar ke saya sindiran pedas pak purbaya
12	Pembangunan 2 ruang kelas baru, rehabilitasi 4 ruang kelas lama juga 2 ruang kamar mandi untuk guru juga pagar keliling sekolah di usulkan sekolah ku sejak 3 tahun lalu sampai sekarang belum juga terealisasi Pak, mudahan sekolah ku mendapat kan pembangunan yang diusulkan Karena sekolah ku hanya memiliki 4 ruang kelas sedangkan sekolah ada 6 rombongan kelas 🙏	pembangunan ruang kelas baru rehabilitasi ruang kelas lama juga ruang kamar mandi untuk guru juga pagar keliling sekolah diusulkan sekolahku sejak tahun lalu sampai sekarang belum terealisasi pak mudahan sekolahku mendapat pembangunan yang diusulkan karena sekolahku hanya memiliki ruang kelas sedangkan sekolah ada rombongan kelas
25	memimpin negeri ini teorinya mudah, tapi pelaksanaannya susah karena banyak kepentingan.. negara ini butuh sosok yg berani, dihormati dan disegani, untuk menertibkan tikus2 berbahaya..	memimpin negeri ini teorinya mudah tapi pelaksanaannya susah karena banyak kepentingan negara ini butuh sosok yang berani dihormati dan disegani untuk menertibkan tikus berbahaya
33	InsyaAllah Allah akan melindungi Pak Purbaya dan orang-orang yang berniat Mulia seperti yujarsip atau nabi yusuf ketika dijadikan Allah menjadi pengurus kekayaan Mesir dan membantu penguasa mesir Amenhotep bukankah sejarah berulang mungkin sampai hari Kiamat sesuai yang Allah kehendaki	insya allah akan melindungi pak purbaya dan yang berniat mulia seperti yujarsip atau nabi yusuf ketika dijadikan allah menjadi pengurus kekayaan mesir dan membantu penguasa mesir amenhotep bukankah sejarah berulang mungkin sampai hari kiamat sesuai yang allah kehendaki

Gambar 4. Hasil *Removal of HTML Tags*

3) *Removal of URL Tags*

proses ini mengeliminasi berbagai pola Uniform Resource Locator (URL) seperti protokol HTTP/HTTPS, tautan berbasis 'www', hingga pola domain umum agar tidak mengganggu proses pengolahan mesin. Sebagai contoh, pada data baris ke-2363, elemen tautan video YouTube <a href="https://www.youtube.com/watch?v=Zk7meVzqayM&t=1160"> berhasil dihapus sepenuhnya, menyisakan teks murni yang bermakna. Penghapusan ini sangat penting untuk mengurangi kebisingan data (noise) sehingga analisis sentimen dapat berfokus sepenuhnya pada konten tekstual yang substantif bagi penelitian.

	Komentar	Komentar Clean
2332	Ilmu yg bermanfaat ..semoga berkah dunia akhirat.....bermanfaat bagi org banyak....sehat2 slalu pak Menkeu....	ilmu yang bermanfaat semoga berkah dunia bagi org slalu pak menkeu
2333	Indonesia kaya ya Pa Pur sebagai bendahara ya Kaya....nah...Cara Pa Pur ini tepat karena TRANSPARAN dan Rakyat suka kejujuran tanpa ditutup tutupi dan juga suka ketegasan terhadap selama ini yg SALAH dan harus DIBENAHI dan PATUT DIHUKUM YG KORUP begitukan ya Pa Pur	indonesia kaya ya pa pur sebagai bendahara ya pa pur ini tepat karena transparan dan rakyat suka kejujuran tanpa ditutup tutupi dan juga suka ketegasan terhadap selama ini yang salah dan harus dibenahi dan patut dihukum yang korup begitukan ya pa pur
2335	Skrg pak Menkeu agak langsing...kerja keras fisik mental😂	skrg pak menkeu agak keras fisik mental
2361	Sudah terlalu tinggi.kalau dinaikkan.malah tidak ada yg bayar. Kalau pro rakyat turunin pajak dong	sudah terlalu tidak ada yang pro rakyat turunin pajak dong
2363	Mulyono mana mulyono... 19:20 dengerin.	mulyono mana mulyono dengerin

Gambar 5. Hasil *Removal of URL Tags*

4) *Removal of Repeated Words*

Penelitian ini secara efektif mengeliminasi pengulangan kata berurutan untuk meningkatkan standarisasi teks pada kolom 'Komentar Clean'. Melalui penggunaan ekspresi reguler, kata-kata yang diulang secara redundan seperti "ribut ribut" dikonversi menjadi "bermasalah", serta reduplikasi kata seperti "muda2", "sedikit2", dan "tiba2" disederhanakan menjadi satu kata dasar. Proses ini sangat penting untuk mengurangi variasi leksikal yang tidak perlu, sehingga model machine learning dapat lebih fokus dalam mengenali makna inti dan sentimen dari setiap komentar pengguna.

	Komentar	Komentar Clean
0	Indonesia negara yang ribut ribut masalah korupsi dan fitnah politik kapan indonesia bisa maju.ketinggalan dengan negara taiwan	indonesia negara yang bermasalah dengan korupsi dan fitnah indonesia bisa belajar dari negara taiwan
23	PAK PRABOWO HARUS LINDUNGI PAK PURBAYA, INDONESIA AKAN MENJADI NEGARA MAJU DAN SEJAHTERA, PAK PRABOWO JANGAN HIRAUKAN ANCAMAN DARI ORANG ORANG YG MEGANG KARTU AS ANDA, ORANG YG BAIK BUKAN YG TIDAK PERNAH SALAH...TAPI DIA MAU MEMPERBAIKI DIRI DEMI BANGSA DAN NEGARA INI, GAK USAH GUBRIS ANCAMAN ORANG ORANG ITU, JANGAN GAK ENAKAN SAMA ORANG YG TELAH BERJASA KEPADA ANDA, KALAU MEMANG SALAH YA HARUS DI HUKUM, CONTOH WHOOSH...PROJEK ITU JANGAN DI TERUSKAN DULU, TANGKAP ORANG ORANG YG SUDAH KORUPSI DAN MARK UP BIAYA WHOOSH, PAK PRABOWO...ANDA PRESIDEN TERAKHIR YG AKAN MEMBAWA INDONESIA MAJU SEJAHTERA BILA KEPUTUSAN ANDA BENAR ATAU MENENGGELAMKAN BANGSA DAN NEGARA INI BILA ANDA SALAH LANGKAH...DAN RAKYAT AKAN MENGENANG ITU SEMUA, PAK PRABOWO BERBUATLAH YG BAIK AMANAH DAN TEGAS TERHADAP ORANG ORANG YG SALAH, MAJU DAN TENGGELAMNYA BANGSA DAN NEGARA INI DI TANGAN ANDA	pak prabowo harus lindungi pak purbaya indonesia akan menjadi negara maju dan sejahtera pak prabowo jangan hiraukan ancaman dari yang pegang kartu as anda orang yang baik bukan yang tidak pernah mau memperbaiki diri demi bangsa dan negara ini gak usah gubris ancaman itu jangan gak enakan sama orang yang telah berjasa kepada anda kalau memang salah ya harus dihukum contoh itu jangan diteruskan dulu tangkap yang sudah korupsi dan mark up biaya whoosh pak presiden terakhir yang akan membawa indonesia maju sejahtera bila keputusan anda benar atau menenggelamkan bangsa dan negara ini bila anda salah rakyat akan mengenang itu semua pak prabowo berbuatlah yang baik amanah dan tegas terhadap yang salah maju dan tenggelamnya bangsa dan negara ini di tangan anda
28	Sy sangat apresiasi thdp kebijakan pa Purbaya bhw bendahara negara itu BKN hanya sekedar jurusan bayar tetapi hrs juga mau turun kelapangan dan ada trik trik khusus dlm hal kebijakan penggunaan dan pengeluaran keuangan negara yg bisa dipertanggungjawabkan KPD publik. Lanjutkan pa menkeu	saya sangat apresiasi terhadap kebijakan pak purbaya bahwa bendahara negara itu bukan hanya sekedar urusan bayar tetapi harus juga mau turun ke lapangan dan ada khusus dalam hal kebijakan penggunaan dan pengeluaran keuangan negara yang bisa dipertanggungjawabkan kepada publik lanjutkan pak menkeu
39	Pertama kali keluar di benci sama statment awalnya, tapi makin kesini kesini makin di cintai nih pak purbaya membela rakyat	pertama kali keluar dibenci sama statement awalnya tapi dicintai nih pak purbaya membela rakyat
42	Justru anda itu cuma banyak omong tpi gk ada buat perbaikan ekonomi masyarakat anda tidak amati kehidupan masyarakat ,ngomong aja liku liku .menutupi kebodoaan.	justru anda itu cuma banyak omong tapi gak ada buat perbaikan ekonomi masyarakat anda tidak amati kehidupan masyarakat ngomong saja menutupi kebodohan

Gambar 6. Hasil *Removal of Repeated Words*

5) *Abbreviation Replacement*

Tahapan Abbreviation Replacement pada kolom Komentar Clean secara efektif menormalisasi penggunaan bahasa tidak baku, singkatan, dan slang yang umum ditemukan dalam komentar YouTube menjadi bentuk kata formal yang terstandarisasi. Proses ini mentransformasi singkatan umum seperti "yg" menjadi "yang" dan "gk" menjadi "tidak", serta memperbaiki kata ganti dan kata serapan seperti "lu" menjadi "kamu" dan "statment" menjadi "statement". Dengan menyamakan format penulisan ini, inkonsistensi data dapat diminimalisir sehingga model machine learning mampu menangkap makna semantik dan informasi kunci secara lebih akurat pada tahap analisis selanjutnya.

	Komentar	Komentar Clean
2266	jadi optimisssssss kalo pak pur udah ngomong lanjoot pak	jadi optimis kalo pak pur udah ngomong lanjoot pak

Gambar 7. Hasil *Abbreviation Replacement*

6) *Removal of Emojis*

Removal of Emojis pada kolom Komentar Clean secara efektif mengeliminasi seluruh karakter piktograf dan simbol visual, seperti ekspresi wajah tertawa, simbol jempol, tangan berdoa, hingga ikon hati yang terdapat pada data mentah. Berdasarkan hasil pemrosesan, emoji pada baris 322, 10, 7, dan 1013 berhasil dihapus sepenuhnya untuk memastikan korpus hanya berisi konten teks murni. Penghapusan ini berfungsi untuk mengurangi kebisingan data (noise) karena elemen non-tekstual tersebut tidak memiliki nilai semantik yang dapat diproses langsung oleh algoritma machine learning. Dengan demikian, data menjadi lebih bersih dan terstandarisasi, sehingga model dapat lebih fokus dalam menganalisis informasi tekstual dan sentimen yang disampaikan oleh pengguna.

	Komentar	Komentar Clean
6	Alhamdulillah..terima kasih Pak Menkeu💪	terima kasih pak menkeu
7	Cerdas dan bahasa yg mudah dimengerti bg rakyat . Kereen Pak Menkeu..👍👍👍	cerdas dan bahasa yang mudah dimengerti oleh rakyat keren pak menkeu
9	Mudah2an beliau pajang umur dan jgk mudah2an apa urusan beliau akan di mudah kan sama Allah SWT dan mudah2an beliau selalu sehat beserta semuwa keluwarga beliau tetep di lindungi sama Allah SWT amiiinnn yarobal alamin 🖤🖤🖤	semoga beliau panjang umur dan semoga segala urusan beliau dimudahkan oleh allah swt semoga beliau selalu sehat beserta seluruh keluarga beliau tetap dilindungi oleh allah swt amiin ya rabbal alamin
10	kecuali anda bayar ke saya SINDIRAN pedas Pak Purbaya😂🤠	kecuali anda bayar ke saya sindiran pedas pak purbaya
12	Pembangunan 2 ruang kelas baru, rehabilitasi 4 ruang kelas lama juga 2 ruang kamar mandi untuk guru juga pagar keliling sekolah di usulkan sekolah ku sejak 3 tahun lalu sampai sekarang belum juga terealisasi Pak, mudahan sekolah ku mendapat kan pembangunan yang diusulkan Karena sekolah ku hanya memiliki 4 ruang kelas sedangkan sekolah ada 6 rombongan kelas 🙏	pembangunan ruang kelas baru rehabilitasi ruang kelas lama juga ruang kamar mandi untuk guru juga pagar keliling sekolah diusulkan sekolahku sejak tahun lalu sampai sekarang belum terealisasi pak mudahan sekolahku mendapat pembangunan yang diusulkan karena sekolahku hanya memiliki ruang kelas sedangkan sekolah ada rombongan kelas

Gambar 8. Hasil *Removal of Emojis*

7) Removal of Numbers

Tahapan Removal of Numbers pada kolom Komentar Clean berhasil mengeliminasi berbagai karakter numerik yang tidak memiliki nilai semantik penting dalam analisis teks. Berdasarkan data penelitian, proses ini menghapus angka nominal besar seperti "200T" pada baris 322, indikator jumlah seperti "10 org" dan "11 orang", hingga persentase "6%". Selain itu, angka penanda waktu seperti "19:20" pada baris 2363 serta angka-angka satuan yang muncul di baris 11 dan 12 juga dibersihkan sepenuhnya untuk menyisakan konten tekstual murni. Penghapusan elemen numerik ini sangat krusial untuk menstandarisasi data dan mengurangi dimensi fitur, sehingga model machine learning dapat lebih fokus dalam memproses kata-kata kunci informatif tanpa terganggu oleh karakter angka.

	Komentar	Komentar Clean
9	Mudah2an beliau pajang umur dan jgk mudah2an apa urusan beliau akan di mudah kan sama Allah SWT dan mudah2an beliau selalu sehat beserta semuwa keluwarga beliau tetep di lindungi sama Allah SWT amiiinnn yarobal alamin ❤❤❤	semoga beliau panjang umur dan semoga segala urusan beliau dimudahkan oleh allah swt semoga beliau selalu sehat beserta seluruh keluarga beliau tetap dilindungi oleh allah swt amiin ya rabbal alamin
11	Desa SUKADANA kecamatan BUAY BAHUGA kabupaten WAY KANAN propinsi LAMPUNG jembatan nya hancur udah hampir 20 taun tp BELUM di bangun lagi	desa sukadana kecamatan buay bahuga kabupaten way kanan provinsi lampung jembatannya hancur sudah hampir tahun tapi belum dibangun lagi
12	Pembangunan 2 ruang kelas baru, rehabilitasi 4 ruang kelas lama juga 2 ruang kamar mandi untuk guru juga pagar keliling sekolah di usulkan sekolah ku sejak 3 tahun lalu sampai sekarang belum juga terealisasi Pak, mudahan sekolah ku mendapat kan pembangunan yang diusulkan Karena sekolah ku hanya memiliki 4 ruang kelas sedangkan sekolah ada 6 rombongan kelas 🙏	pembangunan ruang kelas baru rehabilitasi ruang kelas lama juga ruang kamar mandi untuk guru juga pagar keliling sekolah diusulkan sekolahku sejak tahun lalu sampai sekarang belum terealisasi pak mudahan sekolahku mendapat pembangunan yang diusulkan karena sekolahku hanya memiliki ruang kelas sedangkan sekolah ada rombongan kelas
13	Jutaan rakyat di blakang purbaya klo byj bukti kroptor2nya di tangkap kropsih klo sri muliani itu malah kropsihnya luar binasa dari luar biasa	jutaan rakyat di belakang purbaya kalau banyak bukti koruptornya ditangkap korpsih kalau sri mulyani itu malah korpsihnya luar biasa dari luar biasa
14	Ujung2nya kerakyat yg hrs bayar	ujungnya kerakyat yang harus bayar

Gambar 9. Hasil *Removal of Numbers*

8) Non-Alphanumeric Character Cleaning

Tahapan *Non-Alphanumeric Character Cleaning* pada kolom Komentar Clean berhasil mengeliminasi berbagai karakter non-standar seperti tanda baca dan simbol khusus yang dapat mengganggu konsistensi data. Berdasarkan hasil pengolahan, berbagai tanda baca seperti koma (,), titik (.), tanda tanya (?), dan tanda seru (!) pada baris 0, 3, 4, dan 7 telah dihapus sepenuhnya untuk menyisakan teks alfabet murni. Selain itu, simbol-simbol khusus seperti tanda kutip dan titik-titik redundan (ellipsis) pada baris 322 dan 2333 juga berhasil dibersihkan. Proses ini sangat krusial dalam menyederhanakan korpus data sehingga model machine learning dapat berfokus pada konten inti teks tanpa adanya gangguan dari karakter non-alfanumerik yang tidak memiliki nilai semantik dalam analisis sentimen.

	Komentar	Komentar Clean
0	Indonesia negara yang ribut ribut masalah korupsi dan fitnah politik.kapan indonesia bisa maju.ketinggalan dengan negara taiwan	indonesia negara yang bermasalah dengan korupsi dan fitnah indonesia bisa belajar dari negara taiwan
3	menteri lama ngapain aja ... ?	menteri lama ngapain aja
4	UMKM harus bisa bergerak bukan mandek.. saya yakin pak purbaya cukup baik.	umkm harus bisa bergerak bukan mandek saya yakin pak purbaya cukup baik
5	Alhamdulillah Menkeu yg baru BPK Purbaya Yudhi ,kami masyarakat awam dpt memahami cara mengelola keuangan Negara,sehingga dpt meningkatkan perekonomian Indonesia seutuhnya amin.	alhamdulillah menkeu yang baru bapak purbaya yudhi kami masyarakat awam dapat memahami cara mengelola keuangan negara sehingga dapat meningkatkan perekonomian indonesia secara menyeluruh amin
6	Alhamdulillah..terima kasih Pak Menkeu 🙏	terima kasih pak menkeu

Gambar 10. Hasil *Non-Alphanumeric Character Cleaning*

9) Removal of HTML Entities

Tahapan Removal of HTML Entities pada kolom Komentar Clean secara efektif berhasil mengidentifikasi dan membersihkan berbagai kode karakter HTML yang muncul akibat proses ekstraksi data mentah dari platform YouTube. Berdasarkan hasil pengolahan pada baris 35, 57, dan

89, entitas " berhasil dihapus sepenuhnya sehingga teks kembali bersih tanpa gangguan simbol kutipan teknis. Selain itu, entitas & pada baris 92 dan entitas karakter khusus ' pada baris 109 juga berhasil dieliminasi untuk menghasilkan kata yang utuh dan normal seperti "assalamu alaikum". Proses pembersihan ini sangat krusial agar model machine learning tidak memproses karakter kode tersebut sebagai fitur data yang tidak bermakna, sehingga integritas teks tetap terjaga untuk tahap analisis sentimen selanjutnya.

	Komentar	Komentar Clean
35	"Bantu saya untuk membantu anda" .. nice pak pur❤	bantu saya untuk membantu anda nice pak pur
57	udah g ada yg bisa d percaya alih" Omon. Omon aj a skat koroptor	sudah tidak ada yang bisa dipercaya saja skat koruptor
89	mindset "diskon" menyelamatkan banyak investor	mindset diskon menyelamatkan banyak investor
92	baru kali ini nungguin update berita menteri di indonesia lebih dari nungguin seri baru drakor. ngefans cm dulu sm bu Susi aja, sekarang ngefans juga sm pak Pur & nungguin update nya. Tuhan, lindunglah orang2 baik, petinggi2 yang PRO rakyat & jauhkan dari siapapun dan apapun yang berniat buruk pd mereka 😢	baru kali ini menunggu update berita menteri di indonesia lebih dari menunggu seri baru drama korea dulu sama ibu susi sekarang ngefans juga sama pak purbaya menunggu update nya tuhan lindungilah orang baik petinggi yang pro rakyat jauhkan dari siapapun dan apapun yang berniat buruk pada mereka
109	Assalamu'alaikum	assalamu alaikum

Gambar 11. Hasil *Removal of* HTML *Entities*

10) Whitespace Normalization

Tahapan *Whitespace Normalization* pada kolom Komentar Clean berhasil merapikan struktur teks dengan menghapus spasi berlebih, spasi di awal atau akhir kalimat, serta karakter baris baru yang tidak diperlukan. Berdasarkan data hasil pemrosesan, spasi ganda atau spasi yang mengikuti tanda baca yang telah dihapus—seperti pada baris 0, 4, dan 2333—disatukan menjadi satu spasi tunggal agar teks lebih rapat dan konsisten. Selain itu, karakter baris baru (\n) yang sebelumnya ditandai oleh tag
 pada baris 7, 12, dan 25 dihilangkan sehingga komentar yang terpisah menjadi satu kesatuan paragraf yang utuh. Proses normalisasi ini sangat penting untuk memastikan bahwa setiap kata dalam korpus hanya dipisahkan oleh satu karakter spasi, sehingga mempermudah akurasi proses tokenisasi pada tahap analisis data selanjutnya.

	Komentar	Komentar Clean
500	Bapak menteri Purbaya yth, mohon maaf seandainya bisa dan tidak mengganggu ke tabilitas negara. Saya mohon dapat ditinjau kembali kepada para pensiunan BUMN yang lama dan sekarang masih hidup banyak sekali yg dapat di bawah 1 juta. Kasihan pak sepertinya tidak cukup untuk 1 bulan dengan keadaan ekonomi saat ini. Trima kasih	bapak menteri purbaya yang terhormat mohon maaf seandainya bisa dan tidak mengganggu stabilitas negara saya mohon dapat ditinjau kembali kepada para pensiunan bumn yang lama dan sekarang masih hidup banyak sekali yang dapat di bawah satu juta kasihan pak sepertinya tidak cukup untuk bulan ini dengan keadaan ekonomi saat ini terima kasih
504	' Anda Bercerita Zaman Sekarang Apa Zaman Presiden Sebelumnya ' Perfect,Halus Tapi Mengenai 😂😂😂	anda bercerita zaman sekarang atau zaman presiden sebelumnya perfect halus tapi mengenai
517	Pak kebijakan duit 50t ini gimana ya orang bawah tdk merasakan aduh apa yang di rasakan saat ini bank .. aja kebijakan aneh aneh.. bank plat merah .. kur aja bilang nya koala abis lah aneh kan	pak kebijakan duit t ini gimana ya orang bawah tidak merasakan aduh apa yang di rasakan saat ini bank aja kebijakan bank plat merah kur aja bilang nya koala abis lah aneh kan
532	Pak dana yang bapak menkeu salurkan buat bank tuh buat apa ya . Saya pelaku usaha kecil mau pinjem ke BANK BRI kok gak di kasih . Aneh ah . Tetep aja jalani usaha yng ada . Jadi Tak bisa melebarkan sayap . Untuk orang kecil seperti saya ... Gebrakan untuk orang2 yang d sana kali ya . 😢😢	pak dana yang bapak menkeu salurkan buat bank tuh buat apa ya saya pelaku usaha kecil mau pinjem ke bank bri kok gak di kasih aneh ah tetep aja jalani usaha yng ada jadi tak bisa melebarkan sayap untuk orang kecil seperti saya gebrakan untuk orang yang d sana kali ya
549	Pak mohon di sidak perusahan berbasis online angkutan barang..sy narik aplikasi lalamove pak pajak ny gila gilaan pak .kasihan kami para supir pak.sudah potongan komisi nya sangat tinggi,hingga mencapai 25% pak,setiap per orderan kami di kenakan PPN untuk nilai ny bervariasi pak sesuai Hargo orderannya.terus kami di kenakan pajak penghasilan bulanan pak sangat tinggi potongan ny pak..tgl 31 Oktober kemaren sy pribadi di kenakan pajak penghasilan senilai Rp 449.000 rupiah pak..sedangkan penghasilan bulanan sy senilai Rp 7.000.000 rupiah pak,itupun penghasilan kotor bukan bersih..mohon di bantu kami para drever online Pak..	pak mohon disidak perusahaan berbasis online angkutan yang menggunakan aplikasi seperti lalamove pak pajaknya sangat tinggi kasihan kami para sopir potongan komisinya sangat tinggi hingga mencapai setiap per pesanan kami dikenakan ppn dengan nilai yang bervariasi sesuai harga kami juga dikenakan pajak penghasilan bulanan yang sangat tinggi potongannya oktober kemarin saya pribadi dikenakan pajak penghasilan senilai rp penghasilan bulanan saya senilai rp pak itu pun penghasilan kotor belum dibantu kami para pengemudi online
558	Pak pur g gmpngan ngluarin uang klau bu sri mulyn gampng d rayyuny ngeluarn uang yg pntg dpt pie amn.. pikrnnya gampang pjk2iin aj rkyt kcl...teryt munth sendr.....	pak purbaya tidak mudah mengeluarkan uang jika bu sri mulyono mudah yang penting memperoleh keuntungan pikiran pak purbaya sederhana pajak ijin saja rakyat menuntut sendiri

Gambar 12. Hasil *Whitespace Normalization*

11) Unicode Normalization

Tahapan Unicode Normalization pada kolom Komentar Clean berhasil menstandarisasi representasi karakter teks dengan mengubah simbol-simbol khusus dan karakter unik ke dalam bentuk yang lebih seragam. Berdasarkan hasil pemrosesan, karakter seperti simbol pangkat pada baris 40 ("X1¹")

berhasil dinormalisasi atau dibersihkan untuk menghindari ambiguitas data. Selain itu, karakter Unicode yang membentuk entitas tertentu seperti tanda kutip miring atau variasi karakter non-standar lainnya pada baris 35, 89, dan 504 diseragamkan menjadi format teks standar. Proses ini sangat penting untuk memastikan bahwa setiap karakter memiliki kode biner yang konsisten, sehingga meningkatkan efisiensi algoritma dalam mengenali kata-kata yang sama namun memiliki variasi pengkodean karakter yang berbeda.

	Komentar	Komentar Clean
40	X1¹	NaN
322	Gaskan pak bekingan bapak RAKYAT.. yg empunya daulat dinegeri ini... 😎😎😎 Pada intinya Menkeu udah suplay 200T ke bankir.. lu yg muda² sekarang kudu mikir usahanya apa.. ajukan kredit krna bunga nya sudah dijelaskan turun.. kalo persyaratan rumit atau diperhambat lu bisa komplen ke satgas yg dibentuk pak Mentri.... Gimana guys??.. pak Mentri udah buka peluang nih.. kalo 1 org jadi pemodal untuk pekerjakan 10 org penganggur, 11 orang ini rutin bayar pajak dari penghasilan, yakin aku janji pertumbuhan ekonomi 6% bakal terwujud😎😎	gaskan pak bekingan bapak rakyat yang empunya daulat di negeri ini pada intinya menkeu sudah suplai ke bankir lu yang muda sekarang kudu mikir usahanya apa ajukan kredit karena bunganya sudah dijelaskan turun kalau persyaratan rumit atau diperhambat lu bisa komplain ke satgas yang dibentuk pak menteri gimana guys pak menteri sudah buka peluang nih kalau orang jadi pemodal untuk pekerjakan orang penganggur orang ini rutin bayar pajak dari penghasilan yakin aku janji pertumbuhan ekonomi bakal terwujud
950	Ekonom sejati y kyk gini memberikan solusi yg baik dan tdk merugikan rakyat trutama rakyat kecil, gak kyk mentri wonderwomen kemarin sedikit² naikin pajak doang nyekik rakyat kecil tdk memberikan solusi yg baik yg tdk merugikan masyarakat kecil. Klu solusinya naikin pajak trus y anak SMA bsa jd mentri g harus cri profesor.	ekonom sejati y kyk gini memberikan solusi yang baik dan tdk merugikan rakyat trutama rakyat kecil gak kyk mentri wonderwomen kemarin sedikit naikin pajak doang nyekik rakyat kecil tdk memberikan solusi yang baik yang tdk merugikan masyarakat kecil klu solusinya naikin pajak trus y anak sma bsa jd mentri g harus cri profesor
991	Harusnya yg pertama Wapresnya copot ganti gk rela negara dipimpin org dongo yg tiba² entah siapa yg milih. Ke 2 perbaiki DPR dan selidiki dan harus transparasi mengenai anggaran ini itu sampai tingkat rt rw hrs jelas	harusnya yang pertama wapresnya copot ganti gk rela negara dipimpin org dongo yang tiba entah siapa yang milih ke perbaiki dpr dan selidiki dan harus transparasi mengenai anggaran ini itu sampai tingkat rt rw harus jelas
1013	Entah kenapa kalau liat pak Purbaya mata saya berkaca² bangga sekali punya mentri yg pintar seperti Pak Purbaya ♥	entah kenapa kalau liat pak purbaya mata saya berkaca bangga sekali punya mentri yang pintar seperti pak purbaya

Gambar 13. Hasil Unicode Normalization

12) *Removal of Numbers Attached to Words*

Tahapan *Removal of Numbers Attached to Words* pada kolom Komentar Clean secara efektif menormalisasi kata-kata yang mengandung angka sebagai bentuk reduplikasi atau penjamak bahasa tidak baku menjadi bentuk kata tunggal yang terstandarisasi. Berdasarkan hasil pengolahan data, penggunaan angka di akhir kata seperti pada kata "muda2" (baris 322), "sedikit2" (baris 950), "tiba2" (baris 991), "berkaca2" (baris 1013), serta istilah "tikus2" (baris 25) dan "kroptor2nya" (baris 13) berhasil dibersihkan dari elemen numeriknya. Proses ini sangat penting untuk mengurangi variasi penulisan yang redundan, sehingga memudahkan model machine learning dalam mengenali akar kata yang sama dan meningkatkan akurasi ekstraksi informasi pada tahap analisis selanjutnya.

	Komentar	Komentar Clean
9	Mudah2an beliau pajang umur dan jgk mudah2an apa urusan beliau akan di mudah kan sama Allah SWT dan mudah2an beliau selalu sehat beserta semuwa keluwarga beliau tetep di lindungi sama Allah SWT amiiinnn yarobal alamin ♥♥♥	semoga beliau panjang umur dan semoga segala urusan beliau dimudahkan oleh allah swt semoga beliau selalu sehat beserta seluruh keluarga beliau tetap dilindungi oleh allah swt amiin ya rabbal alamin
13	Jutaan rakyat di blakang purbaya klo byj bukti kroptor2nya di tangkap kropsih klo sri muliani itu malah kropsihnya luar binasa dari luar biasa	jutaan rakyat di belakang purbaya kalau banyak bukti koruptornya ditangkap korpsih kalau sri mulyani itu malah korpsihnya luar biasa dari luar biasa
14	Ujung2nya kerakyat yg hrs bayar	ujungnya kerakyat yang harus bayar
20	PGI pa gmna kbr PPA PGI in semoga Allah slalu beri kesehatan SMA PPA oiya pa Diana punya kisah inspirasi nh buat PPA ad seorang Mentri yg jujur yg dermawan yg amanah dan dia TDK lupa untk mengeluarkan sedekahnya dan dia menjln kn tugasnya sbgai Mentri bnr2 mnjga dengan baik dan Mentri tersebut sangat dermawan dia slalu ringan tangan kepada siapa btuh pertolongannya ya udh pa gtu aj semoga kisah inspirasi in PPA pkai Diana berhrp menjadi Mentri ekonomi yg amanah yg jujur dan jga dermawan semoga itu harapan diana	pagi pak bagaimana kabar ppa pagi ini semoga allah selalu beri kesehatan sama ppa oiya pak diana punya kisah inspirasi nih buat ppa ada seorang menteri yang jujur dermawan dan amanah dia tidak lupa untuk mengeluarkan sedekahnya dan menjalankan tugasnya sebagai menteri menjaga dengan baik menteri tersebut sangat dermawan dia selalu ringan tangan kepada siapa yang butuh pertolongannya ya sudah pak gitu aja semoga kisah inspirasi ini ppa pakai diana berharap menjadi menteri ekonomi yang amanah jujur dan juga dermawan semoga itu harapan diana
25	memimpin negeri ini teorinya mudah, tapi pelaksanaannya susah karena banyak kepentingan.. negara ini butuh sosok yg berani, dihormati dan disegani, untuk menertibkan tikus2 berbahaya..	memimpin negeri ini teorinya mudah tapi pelaksanaannya susah karena banyak kepentingan negara ini butuh sosok yang berani dihormati dan disegani untuk menertibkan tikus berbahaya

Gambar 14. Hasil Removal of Numbers Attached to Words

C. Label

Dalam tahapan pelabelan sentimen, VADER (Valence Aware Dictionary and sEntiment Reasoner) dimanfaatkan sebagai model berbasis leksikon dan aturan untuk mengklasifikasikan setiap komentar YouTube ke dalam kategori sentimen positif, netral, dan negatif. Pendekatan ini secara spesifik

menggunakan compound score yang dihasilkan oleh VADER, sebuah metrik terstandardisasi yang berkisar antara -1 (sentimen paling negatif) hingga +1 (sentimen paling positif).

	Komentar	Komentar Clean	sentiment_label	sentiment_score
0	Indonesia negara yang ribut ribut masalah korupsi dan fitnah politik kapan indonesia bisa maju.ketinggalan dengan negara taiwan	indonesia negara yang bermasalah dengan korupsi dan fitnah indonesia bisa belajar dari negara taiwan	neutral	0.0
1	Hahka aku sayang Kamu Yolamikooka miko	hahka aku sayang kamu yolamikooka miko	neutral	0.0
2	Aku sayang Kamu Yolamikooka	aku sayang kamu yolamikooka	neutral	0.0
3	menteri lama ngapain aja ... ?	menteri lama ngapain aja	neutral	0.0
4	UMKM harus bisa bergerak bukan mandek.. saya yakin pak purbaya cukup baik.	umkm harus bisa bergerak bukan mandek saya yakin pak purbaya cukup baik	neutral	0.0

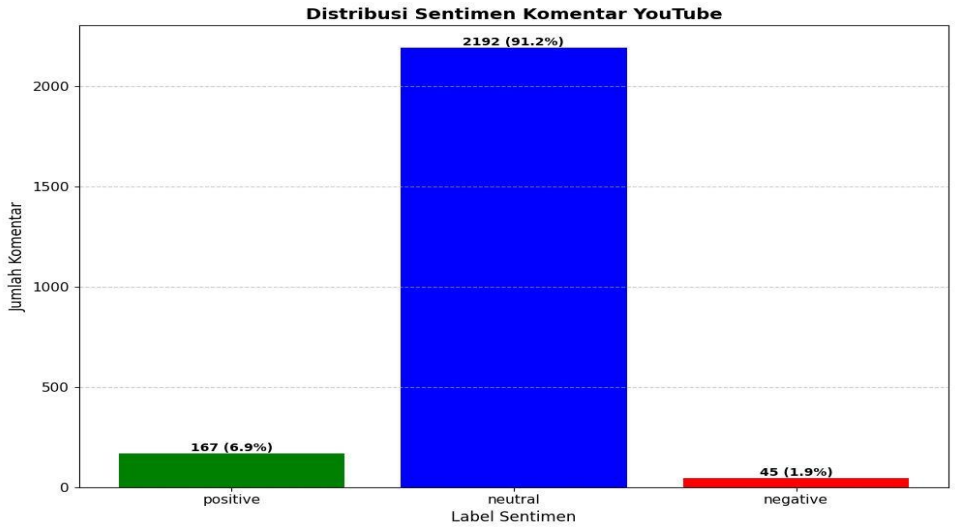
Gambar 16. Hasil Label

Grafik tersebut berjudul "Distribusi Sentimen Komentar YouTube", yang secara visual merepresentasikan hasil analisis sentimen dari sejumlah komentar YouTube. Sumbu horizontal (x-axis) menunjukkan "Label Sentimen" yang terbagi menjadi tiga kategori: "positive" (positif), "neutral" (netral), dan "negative" (negatif). Sementara itu, sumbu vertikal (y-axis) menunjukkan "Jumlah Komentar", dengan skala yang berkisar dari 0 hingga lebih dari 2000.

Analisis sentimen pada grafik ini menunjukkan distribusi sebagai berikut:

1) Sentimen Netral: Kategori ini mendominasi dengan jumlah komentar terbanyak, yaitu 2192 komentar, yang merepresentasikan 91.2% dari total seluruh komentar yang dianalisis. Bar untuk sentimen netral ditandai dengan warna biru.
2) Sentimen Positif: Kategori sentimen positif berada di urutan kedua dengan jumlah 167 komentar, yang setara dengan 6.9% dari total komentar. Bar untuk sentimen positif ditandai dengan warna hijau.
3) Sentimen Negatif: Kategori ini memiliki jumlah komentar paling sedikit, yaitu 45 komentar, yang hanya menyumbang 1.9% dari total keseluruhan. Bar untuk sentimen negatif ditandai dengan warna merah.

Secara keseluruhan, grafik ini dengan jelas menunjukkan bahwa sebagian besar komentar di YouTube yang dianalisis cenderung bersifat netral terhadap topik yang dibahas, dengan proporsi sentimen positif dan negatif yang jauh lebih kecil.



Gambar 17. Distribusi Sentimen Komentar YouTube

D. Pre Processing Data

1) Tokenization

Tahapan Tokenizing pada kolom tokens secara efektif memecah teks dari kolom Komentar Clean menjadi unit-unit kata tunggal (token) dalam bentuk struktur data daftar (list) untuk memudahkan analisis komputasional. Berdasarkan hasil pengolahan pada baris 0 hingga 4, kalimat yang telah dibersihkan didekomposisi menjadi potongan kata yang berdiri sendiri, seperti pada baris 0 di mana kalimat tentang masalah korupsi dipecah menjadi token [indonesia, negara, yang, bermasalah, ...]. Proses ini memastikan bahwa setiap elemen bahasa, termasuk kata benda, kata kerja, dan kata depan, terisolasi secara konsisten tanpa gangguan tanda baca. Dengan mengubah teks menjadi sekumpulan token ini, data siap digunakan untuk tahap pemrosesan lebih lanjut seperti penghitungan frekuensi kata atau pembobotan fitur dalam model machine learning.

	Komentar	Komentar Clean	tokens
0	Indonesia negara yang ribut ribut masalah korupsi dan fitnah politik.kapan indonesia bisa maju.ketinggalan dengan negara taiwan	indonesia negara yang bermasalah dengan korupsi dan fitnah indonesia bisa belajar dari negara taiwan	[indonesia, negara, yang, bermasalah, dengan, korupsi, dan, fitnah, indonesia, bisa, belajar, dari, negara, taiwan]
1	Hahka aku sayang Kamu Yolamikooka miko	hahka aku sayang kamu yolamikooka miko	[hahka, aku, sayang, kamu, yolamikooka, miko]
2	Aku sayang Kamu Yolamikooka	aku sayang kamu yolamikooka	[aku, sayang, kamu, yolamikooka]
3	menteri lama ngapain aja ... ?	menteri lama ngapain aja	[menteri, lama, ngapain, aja]
4	UMKM harus bisa bergerak bukan mandek.. saya yakin pak purbaya cukup baik.	umkm harus bisa bergerak bukan mandek saya yakin pak purbaya cukup baik	[umkm, harus, bisa, bergerak, bukan, mandek, saya, yakin, pak, purbaya, cukup, baik]

Gambar 18. Hasil *Tokenization*

2) Stopword Removal

Penelitian ini secara spesifik menerapkan penghapusan stopword, sebuah langkah krusial yang bertujuan untuk mengeliminasi kata-kata umum dan tidak informatif dari teks (*seperti 'yang', 'dan', 'aku', 'kamu', 'bisa', 'dengan', 'saya', 'harus'*). Proses ini, yang mengubah representasi teks dari kolom Komentar_Clean menjadi stopword_removal, secara signifikan mengurangi noise dan kompleksitas data, sembari menyoroti kata-kata kunci yang lebih substantif dan relevan secara semantik (*contohnya: 'korupsi', 'fitnah', 'umkm', 'bergerak', 'mandek', 'purbaya'*). Dengan demikian, penghapusan stopword memungkinkan analisis selanjutnya untuk lebih fokus pada inti pesan pengguna, meningkatkan efisiensi dan akurasi dalam identifikasi topik dan sentiment.

	Komentar	Komentar Clean	stopword_removal
0	Indonesia negara yang ribut ribut masalah korupsi dan fitnah politik.kapan indonesia bisa maju.ketinggalan dengan negara taiwan	indonesia negara yang bermasalah dengan korupsi dan fitnah indonesia bisa belajar dari negara taiwan	[indonesia, negara, bermasalah, korupsi, fitnah, indonesia, belajar, negara, taiwan]
1	Hahka aku sayang Kamu Yolamikooka miko	hahka aku sayang kamu yolamikooka miko	[hahka, sayang, yolamikooka, miko]
2	Aku sayang Kamu Yolamikooka	aku sayang kamu yolamikooka	[sayang, yolamikooka]
3	menteri lama ngapain aja ... ?	menteri lama ngapain aja	[menteri, ngapain, aja]
4	UMKM harus bisa bergerak bukan mandek.. saya yakin pak purbaya cukup baik.	umkm harus bisa bergerak bukan mandek saya yakin pak purbaya cukup baik	[umkm, bergerak, mandek, purbaya]

Gambar 19. Hasil *Stopword Removal*

3) Stemming

Secara keseluruhan, kolom ini menyajikan inti informasi dari komentar pengguna dalam bentuk kata kunci, seperti keresahan mengenai korupsi dan perbandingan negara pada baris pertama, ungkapan personal pada baris kedua dan ketiga, serta opini terkait kinerja menteri dan

sektor UMKM pada baris terakhir. Hasil ekstraksi ini menghilangkan kata hubung dan imbuhan, sehingga hanya menyisakan kata-kata esensial yang siap digunakan untuk tahap analisis data.

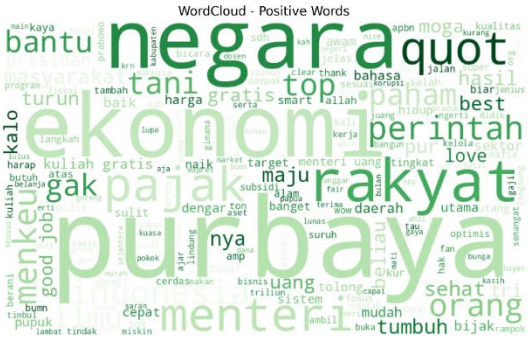
	Komentar	Komentar Clean	text_stemmed
0	Indonesia negara yang ribut ribut masalah korupsi dan fitnah politik.kapan indonesia bisa maju.ketinggalan dengan negara taiwan	indonesia negara yang bermasalah dengan korupsi dan fitnah indonesia bisa belajar dari negara taiwan	[indonesia, negara, masalah, korupsi, fitnah, indonesia, ajar, negara, taiwan]
1	Hahka aku sayang Kamu Yolamikooka miko	hahka aku sayang kamu yolamikooka miko	[hahka, sayang, yolamikooka, miko]
2	Aku sayang Kamu Yolamikooka	aku sayang kamu yolamikooka	[sayang, yolamikooka]
3	menteri lama ngapain aja ... ?	menteri lama ngapain aja	[menteri, ngapain, aja]
4	UMKM harus bisa bergerak bukan mandek.. saya yakin pak purbaya cukup baik.	umkm harus bisa bergerak bukan mandek saya yakin pak purbaya cukup baik	[umkm, gerak, mandek, purbaya]

Gambar 19. Hasil *Stemming*

4) Word Cloud

a) Word Cloud Sentimen Positif

Visualisasi ini menonjolkan sentimen positif yang sangat kuat terhadap figur **Purbaya** dalam konteks pembangunan **ekonomi** dan kesejahteraan **negara**. Kata-kata yang mendominasi seperti **rakyat**, **menteri**, dan **pajak** menunjukkan bahwa diskusi positif ini berkaitan erat dengan kebijakan publik dan peran pemerintah dalam membantu masyarakat, terutama di sektor seperti per**tani**an. Penggunaan istilah pendukung seperti **maju**, **paham**, **cerdas**, dan **tumbuh** mencerminkan adanya apresiasi serta optimisme publik terhadap kinerja yang dianggap efektif dan membawa dampak nyata bagi kondisi finansial serta kemajuan bangsa Indonesia



Gambar 20. Hasil *Wordcloud* Sentimen Positif

b) Word Cloud Sentimen Netral

Visualisasi ini menunjukkan bahwa diskusi publik, baik dalam kategori sentimen positif maupun netral, berpusat pada tokoh **Purbaya** serta isu-isu strategis terkait **ekonomi**, **rakyat**, dan tata kelola **negara**. Pada kategori positif, muncul apresiasi terhadap kinerja pemerintah dengan kata kunci pendukung seperti **menteri**, **tani**, **maju**, **cerdas**, dan **paham**, yang mengindikasikan kepuasan masyarakat terhadap kebijakan yang dianggap memberi dampak nyata bagi kesejahteraan. Sementara itu, pada kategori netral, diskusi lebih bersifat informatif dan objektif dengan kemunculan kata-kata seperti **pajak**, **uang**, **indonesia**, dan **Sri Mulyani**, yang mencerminkan perhatian publik yang tinggi terhadap pengelolaan keuangan negara serta interaksi antarpejabat publik dalam menjalankan fungsinya.

Gambar 21. Hasil *Wordcloud* Sentimen Netral

c) Word Cloud Sentimen Negatif

Visualisasi ini menunjukkan pada tokoh **purbaya**, **rakyat**, dan sektor **ekonomi**, namun kini disertai dengan istilah-istilah yang menunjukkan kekhawatiran atau kritik. Kata-kata seperti **belanja**, **uang**, **pajak**, dan **hutang** muncul secara signifikan, mengindikasikan adanya keresahan publik terkait pengelolaan keuangan negara. Selain itu, terdapat kemunculan kata-kata bernada negatif atau waspada seperti **mafia**, **defisit**, **susah**, **marah**, dan **lambat** yang menyertai pembahasan mengenai institusi seperti **pemerintah**, **presiden**, dan **menteri**. Secara keseluruhan, awan kata ini mencerminkan sentimen negatif masyarakat yang berfokus pada hambatan kerja, masalah anggaran, dan beban ekonomi yang dirasakan oleh rakyat.

.

Gambar 22. Hasil *Word Cloud* Sentimen Negatif

E. TF-IDF

Secara keseluruhan, hasil TF-IDF ini secara efektif mengidentifikasi kata-kata kunci dan frasa (n-gram seperti "indonesia negara", "purbaya optimis", "sehat purbaya") yang paling representatif untuk setiap komentar, menyoroti tema-tema berulang seperti isu negara, ekonomi, korupsi, peran menteri, UMKM, dan figur Purbaya, serta sentimen yang bervariasi dari harapan, kritik, hingga dukungan. Meskipun sebagian besar sel dalam matriks TF-IDF tetap nol karena sifat sparse data teks, nilai-nilai bukan nol yang terekam dengan jelas menunjukkan fokus tematik dan informasi penting dari masing-masing komentar.

--- Bobot TF-IDF untuk 5 Komentar Teratas ---

	ajar	indonesia	indonesia negara	korupsi	negara	sayang	aja	menteri	ngapain	gerak	purbaya	umkm
0	0.377713	0.467285	0.497092	0.350153	0.518912	0	0	0	0	0	0	0
1	0	0	0	0	0	1	0	0	0	0	0	0
2	0	0	0	0	0	1	0	0	0	0	0	0
3	0	0	0	0	0	0	0.521494	0.369273	0.769208	0	0	0
4	0	0	0	0	0	0	0	0	0	0.685969	0.242677	0.685969

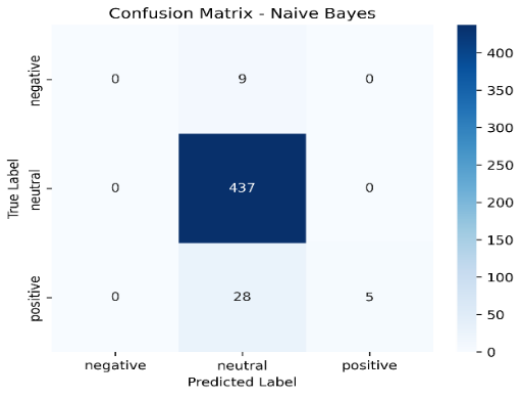
--- Bobot TF-IDF untuk 5 Komentar Terbawah ---

	lanjut	negara	optimis	program	purbaya	purbaya optimis	sehat	sehat purbaya	visi	aamiin	adem	denger	ekonomi
2389	0.373803	0.226007	0.324345	0.356852	0.142088	0.433006	0.246255	0.347606	0.433006	0	0	0	0
2390	0	0	0	0	0.10779	0	0	0	0	0.283573	0.322711	0.297673	0.143422
2391	0	0	0	0	0	0	0	0	0	0	0	0	0
2392	0	0	0	0	0	0	0	0	0.40886	0	0	0	0
2393	0.611852	0	0	0	0.232573	0	0	0	0	0	0	0	0

Gambar 23. Hasil TF-IDF

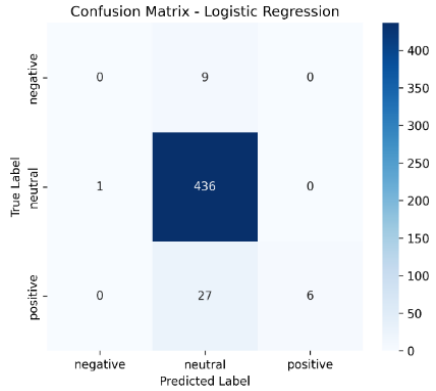
F. Model Machine Learning

Model Naive Bayes ini memiliki akurasi tinggi sebesar 92,28% karena sangat akurat dalam mengenali sentimen "netral" sebagai kelas mayoritas. Namun, nilai Macro F1-Score sebesar 0,4075 menunjukkan adanya tantangan besar dalam mendeteksi kelas minoritas, di mana sentimen "positif" hanya terdeteksi sebagian kecil (recall 0,15) dan sentimen "negatif" belum berhasil teridentifikasi sama sekali. Fenomena ini umum terjadi akibat ketidakseimbangan jumlah data, sehingga model cenderung mengarahkan prediksinya ke kelas dengan sampel terbanyak untuk menjaga performa akurasi secara keseluruhan..



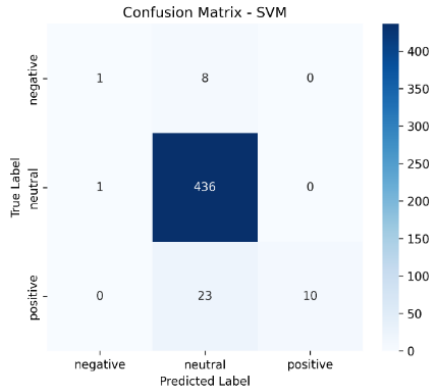
Gambar 24. Hasil Machine Learning Naïve Bayes

Model Regresi Logistik ini mencapai performa yang sangat unggul dalam mengenali sentimen "**neutral**" dengan keberhasilan mengidentifikasi 436 dari 437 data. Namun, terdapat tantangan dalam mendeteksi kelas minoritas, di mana sentimen "**positive**" baru teridentifikasi sebanyak 6 dari 33 instansi dan sentimen "**negative**" belum berhasil terdeteksi sama sekali. Hasil ini menunjukkan adanya kecenderungan model untuk mengarahkan prediksi ke kelas mayoritas akibat ketidakseimbangan data, sehingga diperlukan langkah optimasi lebih lanjut agar model lebih sensitif dalam mengenali kategori sentimen yang jumlah sampelnya terbatas.



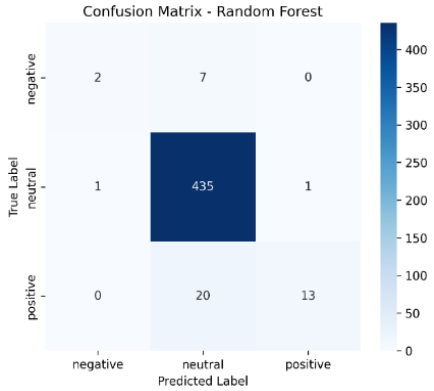
Gambar 25. Hasil *Machine Learning Regresi Logistic*

Model Support Vector Machine (SVM) menunjukkan peningkatan kinerja yang signifikan dengan akurasi 93,32% dan kenaikan drastis pada *Macro F1-Score* menjadi 0,5372, yang mengindikasikan kemampuan klasifikasi yang lebih seimbang di tengah ketidakseimbangan data. Dibandingkan model-model sebelumnya, SVM terbukti lebih efektif dalam menangani kelas minoritas dengan berhasil mendeteksi sebagian sentimen negatif dan meningkatkan *recall* sentimen positif hingga 0,30 sambil tetap mempertahankan presisi sempurna. Meskipun masih terdapat ruang untuk pengembangan pada kategori negatif, SVM sejauh ini menjadi model yang paling andal karena tetap menjaga performa kuat pada sentimen netral sekaligus menunjukkan sensitivitas yang lebih baik terhadap kategori sentimen lainnya



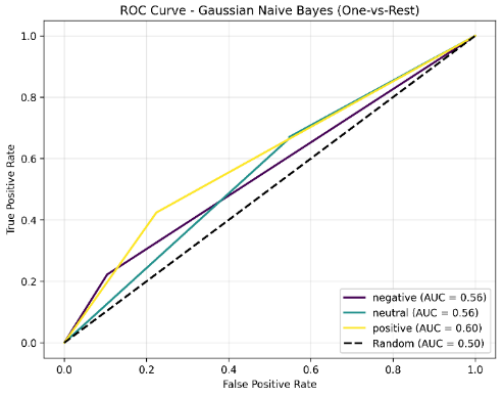
Gambar 26. Hasil *Machine Learning* SVM

Model Random Forest menunjukkan performa yang menjanjikan dengan mencapai akurasi tertinggi sebesar 93,95% dibandingkan model-model sebelumnya. Peningkatan signifikan pada Macro F1-Score menjadi 0,6181 mengindikasikan kemampuan model yang lebih baik dalam menyeimbangkan kinerja di seluruh kelas sentimen, termasuk pada kategori negatif dan positif yang memiliki jumlah data terbatas. Dengan hasil ini, Random Forest terbukti menjadi model yang paling efektif dalam menangani kompleksitas serta ketidakseimbangan kelas dalam tugas analisis sentimen ini



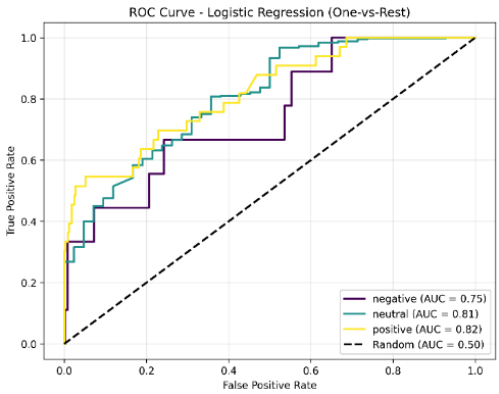
Gambar 27. Hasil *Machine Learning Random Forest*

Kurva ROC (Receiver Operating Characteristic) untuk model Gaussian Naive Bayes dengan pendekatan One-vs-Rest menunjukkan bahwa kemampuan model dalam membedakan kelas sentimen negatif, netral, dan positif sangat terbatas, dengan nilai AUC masing-masing 0,56 untuk kelas negatif dan netral serta 0,60 untuk kelas positif. Nilai-nilai AUC yang hanya sedikit lebih tinggi dari 0,50 nilai kinerja acak menandakan bahwa model ini hampir tidak mampu melakukan diskriminasi yang baik antara kelas-kelas tersebut. Hal ini terlihat dari kurva yang berada sangat dekat dengan garis acak dan jauh dari sudut kiri atas grafik, sehingga dapat disimpulkan bahwa model Gaussian Naive Bayes memiliki performa yang lemah dalam klasifikasi sentimen pada dataset ini.



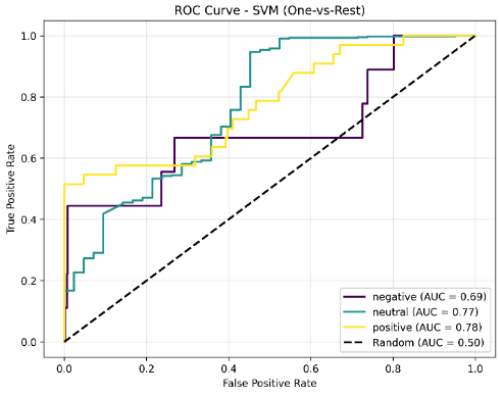
Gambar 28. Hasil ROC Curve Naïve Bayes

Kurva ROC untuk model Logistic Regression dengan pendekatan One-vs-Rest menunjukkan peningkatan performa yang signifikan dibandingkan Gaussian Naive Bayes, dengan nilai AUC sebesar 0,75 untuk kelas negatif, 0,81 untuk kelas netral, dan 0,82 untuk kelas positif. Nilai AUC yang mendekati atau melebihi 0,8 menandakan kemampuan model yang baik dalam membedakan masing-masing kelas sentimen dari kelas lainnya. Kurva yang semakin menjauh dari garis acak dan mendekati sudut kiri atas grafik mengindikasikan tingkat sensitivitas tinggi dengan tingkat kesalahan rendah, sehingga model Logistic Regression dapat diandalkan untuk klasifikasi sentimen dalam dataset ini dengan performa yang lebih unggul.



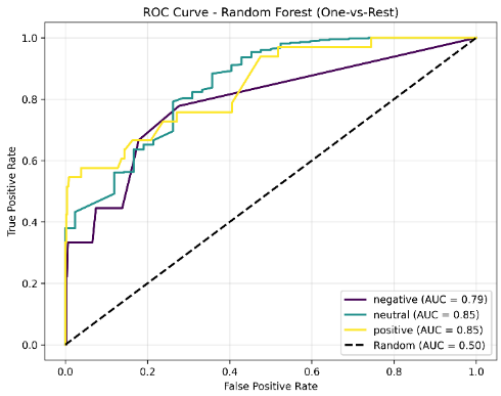
Gambar 29. Hasil ROC Curve Logistic Regresion

Kurva ROC untuk model Support Vector Machine (SVM) dengan pendekatan One-vs-Rest menunjukkan performa yang cukup baik dalam klasifikasi sentimen, dengan nilai AUC sebesar 0,69 untuk kelas negatif, 0,77 untuk kelas netral, dan 0,78 untuk kelas positif. Nilai AUC ini menunjukkan kemampuan model yang lebih baik dibanding Gaussian Naive Bayes, meskipun sedikit lebih rendah dibandingkan Logistic Regression, terutama pada kelas negatif. Kurva yang menjauh dari garis acak dan mendekati sudut kiri atas mengindikasikan model SVM mampu membedakan kelas sentimen dengan tingkat sensitivitas dan spesifisitas yang cukup tinggi, menjadikan SVM sebagai alternatif yang efektif untuk tugas klasifikasi sentimen pada dataset ini.



Gambar 30. Hasil ROC Curve SVM

Kurva ROC pada model Random Forest menunjukkan performa klasifikasi yang cukup baik dan stabil untuk ketiga kategori sentimen (negatif, netral, dan positif) menggunakan strategi One-vs-Rest. Kategori netral dan positif menunjukkan kinerja tertinggi yang identik dengan nilai AUC sebesar 0,85, sedangkan kategori negatif memiliki performa sedikit di bawahnya dengan nilai AUC 0,79. Secara keseluruhan, karena semua nilai AUC berada jauh di atas garis acuan diagonal (AUC = 0,50), dapat disimpulkan bahwa model memiliki kemampuan diskriminasi yang efektif dalam membedakan antar kelas, meskipun model tampak lebih optimal dalam mengenali sentimen positif dan netral dibandingkan sentimen negatif.



Gambar 31. Hasil ROC *Curve Random Forest*

IV. KESIMPULAN

Penelitian analisis sentimen komentar YouTube terhadap kebijakan Menteri Keuangan (Purbaya Yudhi Sadewa) Tidak Menaikkan Pajak menggunakan model Regresi Logistik, Naive Bayes, Support Vector Machine, dan Random Forest menghasilkan beberapa kesimpulan penting:

1. Hasil analisis sentimen terhadap komentar YouTube mengenai kebijakan Menteri Keuangan yang tidak menaikkan pajak, dapat disimpulkan bahwa kecenderungan sentimen masyarakat terbagi ke dalam tiga kategori utama, yaitu sentimen negatif, netral, dan positif. Keberadaan ketiga sentimen tersebut menunjukkan bahwa respons publik terhadap kebijakan tersebut bersifat beragam, mencerminkan adanya perbedaan persepsi, sikap, dan tingkat penerimaan masyarakat terhadap kebijakan fiskal yang diterapkan.
2. Hasil perbandingan performa algoritma klasifikasi Logistic Regression, Naive Bayes, Support Vector Machine (SVM), dan Random Forest, diperoleh bahwa algoritma Random Forest memiliki tingkat akurasi dan kinerja terbaik dalam mengklasifikasikan sentimen komentar YouTube. Hal ini dibuktikan melalui evaluasi ROC Curve dengan nilai Area Under Curve (AUC) yang tinggi pada seluruh kelas sentimen, yaitu 0,88 untuk sentimen negatif, 0,94 untuk sentimen netral, dan 0,93 untuk sentimen positif, yang menunjukkan kemampuan pemisahan kelas yang sangat baik dan stabil dibandingkan algoritma lainnya.

DAFTAR PUSTAKA

[1] M. Quraisy, “Analisis sentimen ulasan aplikasi MyUnpam di Google Play Store menggunakan metode Naive Bayes,” in *Prosiding Seminar Kecerdasan Artifisial, Sains Data, dan Pendidikan Masa Depan (PROKASDADIK)*, vol. 2, Sep. 2024.
[2] S. A. Nugraha, “Penerapan lexicon based untuk analisis sentimen masyarakat Indonesia terhadap Danantara,” *JATI (Jurnal Mahasiswa Teknik Informatika)*, vol. 9, no. 3, Jun. 2025.
[3] I. R. Ainunnisa and Sulastri, “Analisis sentimen aplikasi TikTok dengan metode support vector machine (SVM), logistic regression, dan Naïve Bayes,” *J. Teknol. Sist. Inf. Apl.*, vol. 6, no. 3, pp. 423–430, Jul. 2023, doi: 10.32493/jtsi.v6i3.31076.
[4] Z. Fatah and R. A. Ningsih, “Analisis sentimen komentar YouTube terhadap tragedi demo 25 Agustus menggunakan pendekatan lexicon-based,” makalah tidak dipublikasikan, Universitas Ibrahimy, Situbondo, 2024.
[5] A. Khaidar, “Analisis sentimen di Instagram terhadap Menteri Keuangan Purbaya Yudhi Sadewa menggunakan metode logistic regression,” *JITET (Jurnal Informatika dan Teknik Elektro Terapan)*, vol. 13, no. 3S1, 2024, doi: 10.23960/jitet.v13i3S1.8002.
[6] N. Fauziah, M. Alkautsar, Y. Suryaman, and F. F. Roji, “Pelabelan VADER dalam menganalisis persepsi masyarakat terhadap kenaikan tarif PPN di Indonesia,” makalah tidak dipublikasikan, Universitas Garut, 2025.
[7] A. Aziz, A. B. Susanto, and S. Wiharjo, “Analisis sentimen pelayanan pelanggan mini market Alfamart pada media sosial Twitter dengan Naïve Bayes classifier,” Program Pascasarjana, Universitas Pamulang, Banten, 2025.
[8] A. Andhini, F. N. Handayani, I. Diasih, and N. Nurmalitasari, “Analisis sentimen opini publik pada channel YouTube Mata Najwa menggunakan metode SVM,” *J. Tekn. Inf. dan Teknol. Inf.*, vol. 5, no. 2, pp. 139–154, Aug. 2025, doi: 10.55606/jutiti.v5i2.5426.
[9] S. F. Huwaida, R. Kusumawati, and B. Isnaini, “Analisis sentimen komentar YouTube terhadap pemindahan ibu kota negara menggunakan metode Naïve Bayes,” makalah tidak dipublikasikan, 2024.
[10] A. Danil, “Analisis sentimen masyarakat terhadap pemilihan Bupati Cirebon 2024 berdasarkan komentar pada video debat di YouTube dengan metode Naïve Bayes,” *J. Inform. Tek. Elektro Terap.*, vol. 13, no. 1, 2025.

[11] S. Syafrizal, M. Afdal, and R. Novita, “Analisis sentimen ulasan aplikasi PLN Mobile menggunakan algoritma Naïve Bayes classifier dan K-nearest neighbor,” *MALCOM: Indones. J. Mach. Learn. Comput. Sci.*, vol. 4, no. 1, pp. 10–19, 2024.
[12] A. Zakira, M. Arhami, M. I. Abdi, and S. Safriadi, “Text-based emotion sentiment analysis on social media using NLP and lexicon approach (case study: Gaza conflict),” *J. Inform. Eng. Softw. Appl.*, vol. 1, no. 1, pp. 139–149, 2025.
[13] L. D. Putra, *Analisa konten media sosial Instagram @folkative dalam membentuk opini publik*, Doctoral dissertation, Universitas Buddhi Dharma, 2024.
[14] “Purbaya Yudhi Sadewa,” Wikipedia bahasa Indonesia, Sep. 2025. [Online]. Available: https://id.wikipedia.org/wiki/Purbaya_Yudhi_Sadewa. [Accessed: Sep. 30, 2025].
[15] Kementerian Keuangan Republik Indonesia, “Profil pejabat—Menteri Keuangan: Purbaya Yudhi Sadewa,” 2025. [Online]. Available: https://www.kemenkeu.go.id/profile/profile-pejabat/Menteri-Keuangan. [Accessed: Sep. 30, 2025].
[16] MUC Consulting, “Purbaya Yudhi Sadewa gantikan Sri Mulyani sebagai Menteri Keuangan,” Sep. 8, 2025. [Online]. Available: https://muc.co.id/id/article/purbaya-yudhi-sadewa-gantikan-sri-mulyani-sebagai-menteri-keuangan. [Accessed: Sep. 30, 2025].
[17] “Profil Sri Mulyani, dua dekade urus keuangan negara kini diganti Purbaya,” *Bisnis.com*, Jakarta, Sep. 8, 2025. [Online]. Available: https://ekonomi.bisnis.com/read/20250908/9/1909320/profil-sri-mulyani-2-dekade-urus-keuangan-negara-kini-diganti-purbaya. [Accessed: Sep. 30, 2025].
[18] P. Y. D. Abigail, “Profil Purbaya Yudhi Sadewa, Menteri Keuangan pengganti Sri Mulyani,” *Bisnis.com*, Jakarta, Sep. 8, 2025. [Online]. Available: https://ekonomi.bisnis.com/read/20250908/9/1909315/profil-purbaya-yudhi-sadewa-menteri-keuangan-pengganti-sri-mulyani. [Accessed: Sep. 30, 2025].
[19] A. S. D. P. Sinaga and A. S. Aji, “Analisis sentimen publik terhadap Mayor Teddy Indra Wijaya dengan pendekatan logistic regression,” *MALCOM: Indones. J. Mach. Learn. Comput. Sci.*, vol. 5, no. 1, pp. 222–231, 2025.
[20] D. W. P. Lestari, R. S. Perdana, and P. P. Adikara, “Klasifikasi video clickbait pada YouTube berdasarkan analisis sentiment komentar menggunakan Learning Vector Quantization (LVQ) dan Lexicon-Based Features,” *J. Pengemb. Teknol. Inf. dan Ilmu Komput.*, vol. 3, no. 2, pp. 1184–1189, 2019. [Online]. Available: http://j-ptiik.ub.ac.id
[21] A. R. Ismail and R. B. F. Hakim, “Implementasi Lexicon Based untuk analisis sentiment dalam mengetahui trend wisata Pantai di DI Yogyakarta berdasarkan data twiter,” *Emerging Statistic and Data Science J.*, vol. 1, no. 1, pp. 37–46, 2023.
[22] L. Wirakarsa, A. Angdresey, and J. D. Kapantow, “Implementasi metode Navie Bayes dan Lexicon-Based approach untuk mengkalsifikasi sentiment netizen pada tweet berbahasa Indonesia,” *J. Ilm. Realtech*, vol. 18, no. 1, pp. 15–23, 2022.
[23] I. G. N. A. Wiswasta, I. M. Sukamerta, D. M. Wedagama, and I. G. A. A. Agung, *Metode Penelitian dan Analisis Statistik Kuantitatif Deskriptif*, UNMAS Press, 2017.
[24] D. Wahyuni, N. Fadhillah, and W. W. Ariestya, “Metode Long Short-Term Memory dan Lexicon Based untuk analisis sentiment ulasan aplikasi TikTok,” *J. Ilm. KOMPUTASI*, vol. 23, no. 2, pp. 173–189, 2024, doi: 10.32409/jikstik.23.2.3579.
[25] R. A. S. Nurillah, M. Imrona, and A. Alamsyah, “Prediksi pola penyebaran penyakit DBD di Kota Pagar Alam menggunakan Long Short-Term Memory (LSTM),” *eProceedings of Eng.*, vol. 8, no. 1, pp. 867–882, 2021. [Online]. Available: http://openlibrary.telkomuniversity.ac.id.