

Analisis Sentimen Hasil Transkripsi Audio Berbahasa Indonesia Menggunakan T5 (*Text-to-Text Transfer Transformer*)

Hilman Suhendar^{1*}
Cepy Slamet²
Undang Syaripudin³

^{1,2,3} Teknik Informatika, Fakultas Sains dan Teknologi, UIN Sunan Gunggung Djati Bandung, Jl. A.H. Nasution No. 105 Cibiru Kota Bandung 40614, Indonesia
¹1207050045@student.uinsgd.ac.id, ²cepy_lucky@uinsgd.ac.id, ³undang_if@uinsgd.ac.id

*Penulis Korespondensi:

Hilman Suhendar
1207050045@student.uinsgd.ac.id

Abstrak

Dalam era digital, analisis sentimen menjadi alat penting untuk memahami opini publik, khususnya pada data yang berasal dari media digital seperti video. Namun, analisis sentimen berbasis suara dalam bahasa Indonesia masih jarang dilakukan. Penelitian ini bertujuan untuk mengembangkan model T5 untuk analisis sentimen teks bahasa Indonesia yang dihasilkan dari konversi suara menggunakan teknologi speech-to-text. Keunggulan utama model T5 terletak pada kemampuannya untuk menangani teks panjang, memahami konteks bahasa alami, dan menyesuaikan pelatihan untuk tugas spesifik seperti analisis sentimen. Dataset penelitian diambil dari 20 video YouTube yang dipecah menjadi beberapa segmen berdurasi maksimal 15 detik, menghasilkan total 300 kalimat yang terdiri dari 150 sentimen positif dan 150 sentimen negatif. Data teks yang dihasilkan diproses menggunakan model T5 yang dilatih khusus untuk mendeteksi sentimen positif dan negatif dengan optimalisasi hyperparameter tertentu. Hasil penelitian menunjukkan bahwa model T5 mampu mencapai akurasi sebesar 83%, dengan precision 0.85, recall 0.83, dan f-measure 0.83 pada pengujian menggunakan dataset yang berbeda dari data pelatihan. Penelitian ini menunjukkan bahwa model T5 dapat diadaptasi untuk analisis sentimen berbasis suara dalam bahasa Indonesia dengan hasil yang cukup baik. Temuan ini memberikan kontribusi pada pengembangan teknologi analisis sentimen berbasis suara yang dapat diaplikasikan dalam analisis opini atau ulasan produk. Ke depan, peningkatan tahap pre-processing dan penggunaan dataset yang lebih beragam diharapkan dapat meningkatkan performa model secara keseluruhan.

Kata Kunci: Analisis Sentimen; Pemrosesan Bahasa Alami(NLP); Speech-to-text; T5; Transformer

Abstract

In the digital era, sentiment analysis has become a vital tool for understanding public opinion, particularly from data derived from digital media such as videos. However, voice-based sentiment analysis in the Indonesian language remains uncommon. This research aims to develop the T5 model for sentiment analysis of Indonesian generated from speech using speech-to-text technology. The primary advantages of the T5 model lie in its ability to process lengthy texts, comprehend natural language context, and adapt training for specific tasks such as sentiment analysis. The research dataset was obtained from 20 YouTube videos, segmented into clips of a maximum duration of 15 seconds, resulting in a total of 300 sentences consisting of 150 positive sentiments and 150 negative sentiments. The generated text data was processed using the T5 model, which was specifically trained to detect positive and negative sentiments through the optimization of specific hyperparameters. The results demonstrated that the T5 model achieved an accuracy of 83%, with a precision of 0.85, a recall of 0.83, and an F-measure of 0.83 when tested on datasets different from the training data. This research indicates that the T5 model can be adapted for voice-based sentiment analysis in the Indonesian language with satisfactory results. These findings contribute to the development of voice-based sentiment analysis technology, which can be applied to opinion analysis or product reviews. In the future, improving the pre-processing stage and using more diverse datasets are expected to improve the overall performance of the model.

Keywords: Natural Language Processing (NLP); Sentiment Analysis; Speech-to-text; T5; Transformer

1. Pendahuluan

Perkembangan teknologi informasi, khususnya melalui platform media sosial seperti YouTube, telah mengubah cara masyarakat berinteraksi dan menyampaikan opini terhadap berbagai topik. Sejak pertama kali diluncurkan pada tahun 2005, YouTube telah berkembang menjadi platform berbagi video terbesar di dunia. Pada tahun 2020, YouTube memiliki lebih dari 2 milyar pengguna,

dengan lebih dari 1 milyar jam video yang ditonton setiap hari. Di Indonesia, YouTube menjadi platform media sosial terpopuler, dengan 88% pengguna internet yang mengaksesnya. YouTube memberikan ruang bagi pengguna untuk berbagi pendapat, memberikan ulasan, dan menyampaikan pengalaman mengenai berbagai produk dan layanan yang mereka gunakan[1].

Selain YouTube, platform media sosial lainnya seperti Twitter, Facebook, dan Instagram juga memberikan kesempatan kepada masyarakat untuk mengungkapkan pendapat mereka secara bebas[2]. Data dari Twitter Indonesia pada akhir 2016 menunjukkan bahwa sekitar 77% pengguna Twitter di Indonesia merupakan pengguna aktif yang produktif menulis tweet. Tweet yang dibagikan oleh pengguna ini bisa berupa opini positif, negatif, atau netral terhadap berbagai isu atau produk. Opini yang tersebar di media sosial ini sangat berpengaruh, karena banyak orang bergantung pada ulasan dari pengguna lain sebelum memutuskan untuk menggunakan produk atau layanan tertentu[3], [4].

Seiring berkembangnya media sosial, analisis sentimen menjadi metode yang penting untuk mengevaluasi bagaimana masyarakat merespons suatu topik atau produk. Analisis sentimen adalah teknik untuk mengklasifikasikan opini yang diungkapkan dalam teks, baik itu positif, negatif, maupun netral. Dalam konteks media sosial, analisis sentimen digunakan untuk menilai respons masyarakat terhadap isu-isu tertentu, seperti tren produk atau peristiwa viral[5], [6].

Meskipun analisis sentimen teks telah banyak diterapkan, tantangan muncul seiring dengan meningkatnya penggunaan video suara sebagai sarana untuk menyampaikan pendapat. Pengguna YouTube, misalnya, sering kali memberikan ulasan dalam bentuk video, yang lebih ekspresif dan langsung menunjukkan produk atau layanan yang sedang dibahas.[7], [8] Untuk itu, teknologi speech recognition (SR) diperlukan untuk mengonversi suara dalam video menjadi teks, sehingga bisa dianalisis untuk mendeteksi sentimen[9], [10]. Teknologi ini, meskipun masih dalam tahap awal, memiliki potensi besar untuk memperkaya proses analisis sentimen, terutama dalam konteks media sosial berbasis suara[11].

Model T5 (Text-to-Text Transfer Transformer) memiliki keunggulan dalam menghadapi tantangan ini. Berkat arsitektur berbasis transformer, T5 mampu memahami konteks kalimat secara menyeluruh, memungkinkan model untuk menangkap makna kata-kata ambigu berdasarkan hubungannya dengan kata-kata lain di dalam teks[12]. Fleksibilitas T5 dalam menangani teks panjang dan memahami pola bahasa yang kompleks juga menjadikannya efektif dalam mengatasi struktur kalimat yang fleksibel. Selain itu, kemampuan T5 untuk dilatih secara multitask membuatnya mampu memahami campuran bahasa dan pola penulisan yang khas di media sosial Indonesia[13], [14].

Penelitian ini bertujuan untuk mengimplementasikan model T5 dalam analisis sentimen teks berbahasa Indonesia yang dihasilkan dari konversi suara menggunakan teknologi speech-to-text. Inovasi utama dari penelitian ini terletak pada penerapan model T5 untuk analisis sentimen berbasis suara dalam bahasa Indonesia, suatu bidang yang hingga saat ini masih jarang dieksplorasi. Model T5 dipilih karena fleksibilitasnya dalam menangani berbagai tugas pemrosesan bahasa alami serta kemampuannya untuk menghasilkan akurasi yang kompetitif, seperti yang telah terbukti dalam analisis sentimen berbasis teks[14], [15]. Diharapkan, penelitian ini dapat berkontribusi pada pengembangan teknologi analisis sentimen berbasis suara di Indonesia, yang tidak hanya relevan untuk penelitian akademik, tetapi juga memiliki aplikasi praktis dalam berbagai sektor, seperti analisis opini publik, ulasan produk, dan pelayanan pelanggan.

Penelitian sebelumnya telah mengembangkan berbagai metode analisis sentimen berbasis teks, seperti Naïve Bayes, Support Vector Machine (SVM), dan Convolutional Neural Networks (CNN). Penelitian lainnya telah menggunakan SVM dan Naïve Bayes dalam analisis sentimen Twitter terhadap layanan publik, yang menunjukkan akurasi 84,95% dengan SVM dan 76,43% dengan

Naïve Bayes, mengindikasikan efektivitas kedua metode tersebut dalam analisis sentimen berbasis teks[16].

Selain itu, di penelitian sebelumnya membuktikan bahwa model berbasis Transformer seperti BERT dan XLNet dapat meningkatkan akurasi analisis sentimen berbasis teks, dengan akurasi tertinggi pada model XLNet sebesar 96,2%[17], [18]. Hasil penelitian lain menunjukkan keunggulan BERT dan T5 dalam menganalisis sentimen ulasan produk di Amazon, di mana T5 mencapai akurasi 91%, namun BERT mencapai akurasi hingga 92% dan lebih unggul dalam akurasi, nilai F1, serta efisiensi komputasi, menjadikannya model pilihan untuk analisis sentimen[15].

Ada penelitian yang memberikan kontribusi signifikan dengan memperkenalkan model T5 (Text-to-Text Transfer Transformer), yang menyatukan berbagai tugas pemrosesan bahasa alami dalam format teks-ke-teks. Dalam tugas analisis sentimen berbasis teks, model T5-Base mencatat akurasi 95,2% pada dataset SST-2, yang merupakan bagian dari benchmark GLUE. Hasil ini menunjukkan kemampuan T5 dalam memahami konteks teks panjang serta menangani tugas klasifikasi sentimen dengan sangat baik[14].

Selain itu, penelitian lainnya menunjukkan penggunaan T5 untuk tugas peringkasan dokumen berbahasa Indonesia. Penelitian ini menemukan bahwa model T5 mampu menghasilkan ringkasan abstraktif yang lebih bermakna dibandingkan metode ekstraktif, dengan skenario terbaik menggunakan stemming tanpa penghapusan stopwords, yang mencapai nilai evaluasi ROUGE-1 sebesar 0,17568. Penelitian ini mempertegas potensi T5 dalam memahami struktur bahasa Indonesia yang kompleks, termasuk menangani fleksibilitas tata bahasa dan kata-kata berimbuhan[12].

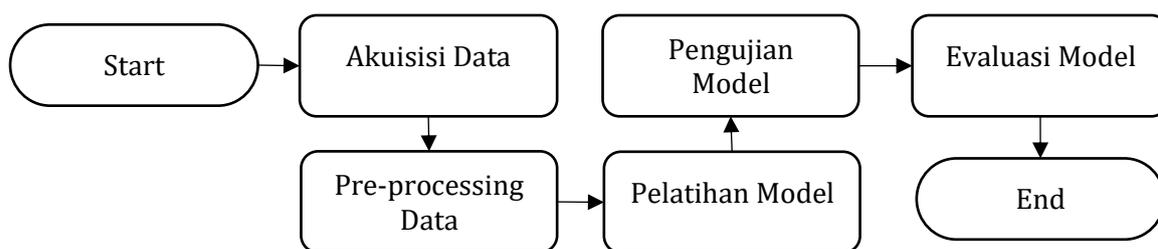
Meskipun demikian, penelitian-penelitian ini terbatas pada data teks tertulis dan belum banyak mengeksplorasi potensi analisis sentimen pada data suara. Sementara itu, beberapa studi lain telah mengimplementasikan teknologi speech-to-text dalam analisis suara. Suatu penelitian juga telah mengembangkan teknologi speech-to-text untuk pencatatan diagnostik pasien dan mencapai akurasi 93,34% dalam mengonversi suara ke teks[19]. Penelitian ini menunjukkan potensi aplikasi speech-to-text dalam berbagai bidang, termasuk deteksi sentimen pada data suara. Penelitian lainnya mendemonstrasikan bagaimana model bahasa terlatih (pre-trained language models) dapat mengurangi kebutuhan anotasi manual hingga 65% dibandingkan dengan metode tradisional, yang memperlihatkan pentingnya pemanfaatan teknologi ini dalam analisis sentimen berbasis suara[20].

Ada suatu penelitian yang memberikan kontribusi pada penerapan teknologi speech recognition dalam analisis sentimen, dengan pendekatan menggunakan metode SVM dengan Sequential Minimal Optimization (SMO) pada video ulasan produk. Penelitian ini menunjukkan peningkatan akurasi, recall, precision, dan f-measure, serta menghasilkan akurasi tertinggi sebesar 89,91%, yang mengindikasikan potensi besar dalam menggabungkan teknologi speech-to-text dengan analisis sentimen berbasis teks[11].

Meskipun begitu, eksplorasi terkait model T5 untuk analisis sentimen berbasis suara dalam bahasa Indonesia masih sangat terbatas. Dengan keunggulan T5 dalam memahami konteks, menangani teks panjang, dan fleksibilitasnya untuk berbagai tugas NLP, penelitian ini bertujuan untuk mengeksplorasi potensi penerapannya dalam analisis sentimen berbasis suara dalam bahasa Indonesia[12], [14], [15]. Pendekatan ini diharapkan dapat memberikan kontribusi penting dalam pengembangan teknologi NLP, khususnya untuk bahasa Indonesia, serta memperluas aplikasi analisis sentimen berbasis suara di berbagai sektor seperti pemasaran, opini publik, dan layanan pelanggan.

2. Metode Penelitian

Pada penelitian ini, data yang dibutuhkan berupa hasil transkripsi teks dari data suara berbahasa Indonesia yang dikonversi menggunakan teknologi speech-to-text. Dataset suara diperoleh dari video ulasan mobil di beberapa channel YouTube yang membahas fitur, performa, dan pengalaman penggunaan mobil. Video dipilih untuk mencakup variasi aksen, gaya berbicara, dan topik pembahasan untuk memastikan keberagaman data. Dataset terdiri dari dua jenis sentimen, yaitu positif dan negatif. Penelitian ini menggunakan Google Colaboratory sebagai alat pemrosesan data dan membangun model T5 dalam bahasa pemrograman Python. Beberapa pustaka yang digunakan antara lain Transformer untuk pengimplementasian model T5, serta pustaka pendukung lainnya untuk pembersihan data dan evaluasi model.



Gambar 1. Tahapan Penelitian

Penelitian ini dilakukan melalui beberapa tahapan, seperti yang ditunjukkan pada Gambar 1. Penjelasan setiap tahapan adalah sebagai berikut. Tahap ini menghasilkan dataset berbahasa Indonesia yang diperoleh dari transkripsi suara 20 video YouTube yang relevan dengan tema ulasan mobil. Data suara dipecah menjadi segmen pendek berdurasi maksimal 15 detik untuk meningkatkan akurasi transkripsi dan analisis sentimen. Setelah dikonversi menggunakan teknologi speech-to-text, dataset akhir berisi 300 kalimat yang terbagi seimbang menjadi 150 kalimat bersentimen positif dan 150 kalimat bersentimen negatif. Dataset ini menjadi representasi awal untuk mendukung tahap pemrosesan selanjutnya.

Dataset hasil transkripsi suara kemudian di proses dengan mengubah semua teks menjadi huruf kecil (lowercase) dan menyesuaikan format data agar sesuai dengan kebutuhan model T5. Format data terdiri dari kolom input_text, yang berisi teks dengan prompt seperti "sentiment detection:", serta kolom target_text yang berisi label sentimen (positif atau negatif). Dataset kemudian dibagi menjadi data pelatihan dan data pengujian menggunakan metode train-test split, serta dikonversi ke format Hugging Face Dataset untuk kompatibilitas dengan pustaka pemrosesan bahasa alami. Proses tokenisasi dilakukan untuk mengubah teks menjadi representasi numerik yang sesuai dengan karakteristik bahasa Indonesia. Setelah tahap ini selesai, dataset siap digunakan dalam pelatihan model analisis sentimen.

Proses pelatihan model dilakukan dengan mengonfigurasi parameter menggunakan objek TrainingArguments dari Hugging Face Transformer. Parameter yang diatur meliputi laju pembelajaran, ukuran batch, jumlah epoch, strategi logging, dan penyimpanan model. Metode early stopping diterapkan untuk menghentikan pelatihan jika tidak ada peningkatan pada nilai eval_loss berturut-turut, guna mencegah overfitting dan menghemat waktu komputasi.

Model diuji menggunakan 30 data uji yang telah dipisahkan dari data pelatihan. Data ini tidak pernah digunakan selama pelatihan, sehingga hasil pengujian mencerminkan kemampuan model dalam mengenali pola sentimen pada data baru. Model dievaluasi menggunakan beberapa metrik kinerja: precision, recall, F1-score, dan akurasi. Precision mengukur ketepatan model dalam mengidentifikasi sentimen positif dan negatif, sementara recall mengukur kemampuan model dalam mendeteksi semua contoh relevan di setiap kategori sentimen. F1-score, sebagai rata-rata

harmonis dari precision dan recall, digunakan untuk menilai keseimbangan antara keduanya, terutama dalam kasus data yang tidak seimbang. Hasil evaluasi model menunjukkan akurasi sebesar 83%, dengan precision, recall, dan F1-score yang bervariasi antara sentimen positif dan negatif.

3. Hasil

Tahap akuisisi data pada penelitian ini dimulai dengan pengumpulan data suara dari 20 video YouTube yang relevan dengan tema penelitian, khususnya ulasan mobil yang membahas fitur, performa, dan pengalaman pengguna. Pemilihan video dilakukan secara cermat untuk memastikan keberagaman data, mencakup variasi aksen, gaya berbicara, serta topik pembahasan. Keberagaman ini bertujuan untuk meningkatkan generalisasi model terhadap data baru di luar dataset pelatihan.

Setiap video yang terkumpul kemudian dipecah menjadi beberapa segmen pendek dengan durasi maksimal 15 detik. Segmentasi dilakukan untuk memastikan potongan suara lebih mudah diolah oleh teknologi speech-to-text sekaligus meminimalkan kesalahan transkripsi. Proses ini menghasilkan potongan suara yang singkat dan berfokus pada satu konteks pembicaraan tertentu, sehingga mempermudah analisis sentimen di tahap selanjutnya.

Potongan suara yang telah disegmentasi kemudian dikonversi menjadi teks menggunakan teknologi speech-to-text yang mendukung bahasa Indonesia. Proses konversi ini menghasilkan teks yang kemudian diberi label sentimen positif atau negatif berdasarkan konteks pembicaraan. Dataset akhir terdiri dari 300 kalimat teks berbahasa Indonesia, yang masing-masing terbagi secara seimbang menjadi 150 kalimat dengan sentimen positif dan 150 kalimat dengan sentimen negatif. Dataset ini dirancang untuk menjadi representasi awal dari ulasan mobil berbasis suara, sekaligus menjadi acuan untuk tahap-tahap pemrosesan selanjutnya.

Pada tahap pre-processing data, teks hasil transkripsi diproses lebih lanjut untuk memastikan kualitas dan konsistensi data sebelum digunakan dalam pelatihan model. Semua huruf dalam teks diubah menjadi huruf kecil (lowercase) guna menjaga konsistensi dalam pengolahan teks. Selanjutnya, dilakukan penyesuaian format data agar sesuai dengan kebutuhan model T5. Proses ini melibatkan penambahan kolom `input_text`, yang berisi teks dengan prompt seperti "sentiment detection:" untuk mengarahkan tugas analisis sentimen, serta kolom `target_text` yang menyimpan label sentimen (positif atau negatif).

Dataset yang telah diproses ini kemudian dibagi menjadi dua bagian, yaitu data pelatihan (training set) dan data pengujian (test set), menggunakan metode train-test split. Pembagian ini dilakukan untuk memastikan bahwa model dapat diuji pada data yang belum pernah dilihat selama pelatihan, sehingga mengurangi risiko overfitting. Selanjutnya, dataset yang telah terbagi ini dikonversi ke format Hugging Face Dataset, yang kompatibel dengan pustaka pemrosesan bahasa alami modern.

Tahapan berikutnya adalah tokenisasi, di mana teks diubah menjadi representasi numerik yang dapat dipahami oleh model T5. Tokenisasi dilakukan dengan mempertimbangkan panjang maksimal teks dan karakteristik bahasa Indonesia. Setelah proses ini selesai, dataset siap digunakan dalam pelatihan model untuk mendeteksi sentimen berdasarkan data teks yang dihasilkan dari suara.

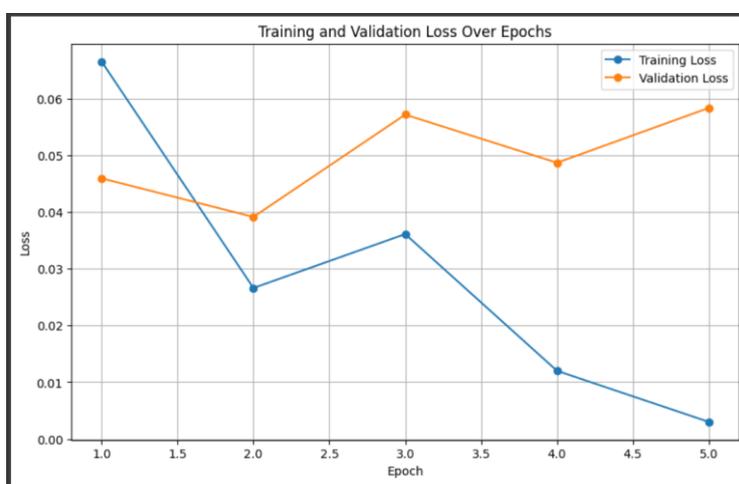
4. Pembahasan

Pada tahap pelatihan model, parameter pelatihan dikonfigurasi menggunakan objek `TrainingArguments` dari Hugging Face Transformer, mencakup laju pembelajaran, ukuran batch, jumlah *epoch*, strategi *logging*, dan penyimpanan model. Metode *early stopping* diterapkan untuk menghentikan pelatihan jika tidak ada peningkatan pada nilai evaluasi *loss* selama tiga *epoch* berturut-turut, guna menghindari *overfitting* dan menghemat waktu komputasi. Model kemudian

dilatih dengan objek *Trainer*, yang menggabungkan model, dataset pelatihan, dan evaluasi. Setiap *epoch* dievaluasi berdasarkan nilai *eval_loss*, dan pelatihan berlanjut hingga mencapai *epoch* maksimal atau kondisi *early stopping* terpenuhi, dengan model terbaik dimuat di akhir pelatihan. Tahap ini bertujuan untuk menghasilkan model T5 yang optimal dalam mendeteksi sentimen dari teks berbahasa Indonesia hasil konversi speech-to-text.

Tabel 1. Hasil Pelatihan Model

<i>Epoch</i>	<i>Training Loss</i>	<i>Validation Loss</i>
1	0.066500	0.045922
2	0.026600	0.039128
3	0.036100	0.057161
4	0.012000	0.048678
5	0.003000	0.058335



Gambar 2. Grafik Pelatihan Model

Tabel 1. Hasil Pelatihan Model dan Gambar 2. Grafik Pelatihan Model menunjukkan nilai training loss dan validation loss pada setiap epoch selama proses pelatihan. Nilai training loss mencerminkan seberapa baik model belajar dari data pelatihan, sedangkan validation loss mengindikasikan seberapa baik model mampu generalisasi terhadap data yang tidak dilihat selama pelatihan.

Pada epoch pertama, nilai training loss sebesar 0.066500 dan validation loss sebesar 0.045922, menunjukkan bahwa model mulai belajar pola dari data pelatihan namun masih belum optimal. Seiring bertambahnya epoch, training loss terus menurun secara signifikan, mencapai 0.003000 pada epoch kelima, menunjukkan bahwa model semakin mampu mempelajari pola data pelatihan dengan lebih baik.

Namun, nilai validation loss tidak menunjukkan pola penurunan yang konsisten. Pada epoch kedua, nilai validation loss turun menjadi 0.039128, yang merupakan nilai terendah selama pelatihan. Setelah itu, validation loss mulai meningkat kembali pada epoch ketiga hingga kelima, dengan nilai terakhir sebesar 0.058335. Pola ini mengindikasikan bahwa model mulai mengalami overfitting, di mana model terlalu fokus pada data pelatihan dan kurang mampu menangani data baru.

Berdasarkan hasil ini, model terbaik kemungkinan diperoleh pada epoch kedua, ketika validation loss mencapai nilai terendah sebesar 0.039128. Penggunaan metode early stopping membantu

mencegah pelatihan berlanjut terlalu lama dan menghindari overfitting, dengan memuat model terbaik dari epoch tersebut untuk digunakan dalam pengujian lebih lanjut.

Model dievaluasi menggunakan 30 data uji yang dipisahkan dari data pelatihan. Data ini belum pernah dilihat selama pelatihan, sehingga hasil pengujian mencerminkan kemampuan model dalam mengenali pola sentimen pada data baru. Proses ini bertujuan untuk mengukur seberapa baik model memprediksi sentimen berdasarkan teks hasil konversi suara. Hasil prediksi kemudian dibandingkan dengan label asli untuk mengevaluasi akurasi dan performa model, memastikan bahwa model tidak hanya efektif pada data pelatihan, tetapi juga konsisten pada data baru yang lebih mirip dengan penggunaan nyata.

```

Kalimat: speedomernya keren terdapat animasi saat mobilnya dinyalakan
Sentimen Asli: Positive
Prediksi Sentimen: Negative
-----
Kalimat: head unitnya memiliki resolusi yang tinggi dan juga responsif
Sentimen Asli: Positive
Prediksi Sentimen: Positive
-----
Kalimat: fitur-fiturnya kurang banyak
Sentimen Asli: Negative
Prediksi Sentimen: Negative
-----
Kalimat: mobil ini juga belum memiliki Cruise control
Sentimen Asli: Negative
Prediksi Sentimen: Positive
    
```

Gambar 3. Hasil Pengujian Model

Penelitian ini mengukur kinerja model dalam menganalisis sentimen dengan menggunakan beberapa metrik: *precision*, *recall*, *F1-score*, dan akurasi. *Precision* menunjukkan seberapa tepat model dalam mengidentifikasi sentimen positif dan negatif tanpa memasukkan kesalahan, sedangkan *recall* mengukur seberapa baik model mendeteksi semua contoh yang relevan dalam setiap kategori sentimen. *F1-score*, yang merupakan rata-rata harmonis dari *precision* dan *recall*, digunakan untuk mengevaluasi keseimbangan antara kedua metrik tersebut, terutama dalam kasus data yang mungkin tidak seimbang.

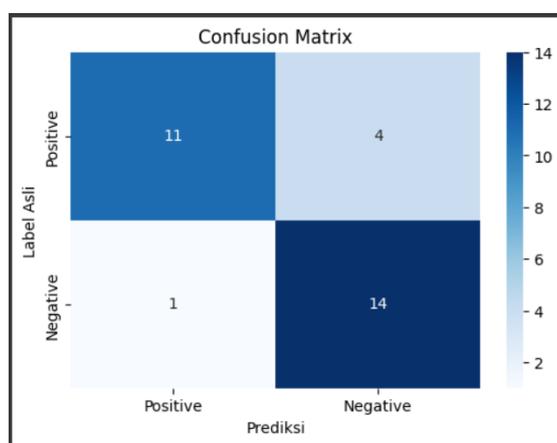
Tabel 2. Hasil Evaluasi Model

	<i>Precision</i>	<i>Recall</i>	<i>F1-score</i>	<i>Support</i>
<i>Positive</i>	0.78	0.93	0.85	15
<i>Negative</i>	0.92	0.73	0.81	15
<i>Accuracy</i>			0.83	30
Macro Avg	0.85	0.83	0.83	30
Weighted Avg	0.85	0.83	0.83	30

Dari data hasil evaluasi model di atas, dapat disimpulkan bahwa kinerja model dalam mengklasifikasikan sentimen positif dan negatif menunjukkan beberapa kekuatan dan kelemahan. Untuk sentimen positif, model memiliki *precision* sebesar 0.78, yang menunjukkan bahwa 78% dari seluruh prediksi yang dibuat oleh model untuk sentimen positif adalah benar. Sementara itu, *recall* sebesar 0.93 menunjukkan bahwa model mampu mendeteksi 93% dari total data yang sebenarnya bersentimen positif. Nilai *F1-score* sebesar 0.85 mencerminkan keseimbangan yang cukup baik antara *precision* dan *recall* untuk sentimen positif. Di sisi lain, untuk sentimen negatif, model menunjukkan *precision* yang lebih tinggi, yaitu sebesar 0.92, yang berarti 92% dari seluruh prediksi negatif adalah benar. Namun, *recall* untuk sentimen negatif hanya sebesar 0.73, yang menunjukkan bahwa model hanya mampu mendeteksi 73% dari total data bersentimen negatif.

Meskipun F1-score untuk sentimen negatif sebesar 0.81, ini menunjukkan bahwa kinerja model untuk sentimen negatif sedikit lebih rendah dibandingkan dengan sentimen positif.

Secara keseluruhan, model mencapai akurasi sebesar 0.83, atau 83%, yang berarti bahwa 83% dari seluruh prediksi yang dilakukan oleh model sesuai dengan label yang benar (baik positif maupun negatif). Rata-rata makro, yang menghitung nilai rata-rata untuk setiap metrik (precision, recall, dan F1-score) tanpa memperhatikan distribusi jumlah data pada setiap kelas, menunjukkan nilai masing-masing sebesar 0.85 untuk precision, 0.83 untuk recall, dan 0.83 untuk F1-score. Ini menunjukkan kinerja model secara keseluruhan pada kedua kelas (positif dan negatif). Selain itu, rata-rata tertimbang (weighted average), yang mempertimbangkan jumlah data pada masing-masing kelas (dalam hal ini, kelas positif dan negatif masing-masing terdiri dari 15 data), memberikan nilai yang sama dengan rata-rata makro. Hal ini menunjukkan bahwa distribusi data yang seimbang antara kelas positif dan negatif tidak mempengaruhi hasil evaluasi model.



Gambar 4. Confusion Matrix

Berdasarkan confusion matrix, kita dapat mengamati hasil prediksi model yang lebih mendetail. Dalam hal ini, dari total 15 data yang sebenarnya positif, model berhasil memprediksi 11 data dengan benar sebagai positif (True Positives), namun ada 4 data yang seharusnya positif, tetapi salah diprediksi sebagai negatif (False Negatives). Untuk sentimen negatif, dari 15 data yang sebenarnya negatif, model memprediksi 14 dengan benar sebagai negatif (True Negatives), namun 1 data yang seharusnya negatif salah diprediksi sebagai positif (False Positive). Hal ini menjelaskan bahwa model lebih baik dalam menghindari kesalahan saat memprediksi sentimen negatif dengan precision yang cukup tinggi (0.92), namun model sedikit kesulitan dalam mendeteksi seluruh data negatif, seperti yang terlihat pada recall yang hanya 0.73. Sebaliknya, meskipun model sangat andal dalam mendeteksi data positif (recall 0.93), terdapat risiko lebih tinggi untuk membuat kesalahan saat memprediksi sentimen positif, yang tercermin dari precision yang lebih rendah (0.78). Secara keseluruhan, meskipun model menunjukkan performa yang baik dengan akurasi 83%, terdapat ketidakseimbangan dalam precision dan recall antara sentimen positif dan negatif. Masalah ini dapat diatasi dengan mengoptimalkan hyperparameter model, menyeimbangkan dataset, atau menyesuaikan metode pre-processing guna mengurangi bias terhadap salah satu kelas.

5. Penutup

Penelitian ini bertujuan untuk mengimplementasikan model T5 dalam analisis sentimen berbasis suara, khususnya untuk teks berbahasa Indonesia yang dihasilkan dari konversi suara menggunakan teknologi speech-to-text. Inovasi utama dari penelitian ini terletak pada penerapan model T5 untuk analisis sentimen berbasis suara dalam bahasa Indonesia, suatu bidang yang hingga saat ini masih jarang dieksplorasi. Hasil pelatihan model menunjukkan bahwa T5 mampu

mengklasifikasikan sentimen dengan akurasi 83%, dengan keseimbangan yang cukup baik antara precision dan recall pada kedua kategori sentimen (positif dan negatif). Meskipun model ini menunjukkan kinerja yang baik, terdapat beberapa area yang dapat diperbaiki, seperti meningkatkan recall pada sentimen negatif.

Secara keseluruhan, penelitian ini telah menunjukkan bahwa T5 merupakan model yang potensial untuk analisis sentimen berbasis suara dalam bahasa Indonesia, meskipun masih terdapat tantangan dalam meningkatkan deteksi sentimen negatif. Temuan ini membuka peluang bagi pengembangan lebih lanjut dalam pengolahan data suara untuk analisis sentimen, yang dapat diterapkan dalam berbagai sektor, seperti riset pasar, layanan pelanggan, dan analisis opini publik. Melalui penelitian ini, diharapkan dapat memberikan kontribusi pada pengembangan teknologi speech-to-text dalam konteks analisis sentimen di Indonesia, serta mendorong penggunaan model-model bahasa alami yang lebih canggih dalam pemrosesan data suara.

Salah satu tantangan utama dalam penelitian ini adalah kesulitan mendapatkan video yang mengandung sentimen positif dan negatif dalam satu video secara eksplisit. Banyak video ulasan yang cenderung berfokus pada sentimen yang lebih dominan, baik itu positif atau negatif, tanpa memberikan keseimbangan yang jelas antara kedua sentimen tersebut. Hal ini menyulitkan dalam pembuatan dataset yang seimbang antara sentimen positif dan negatif, karena umumnya video ulasan cenderung lebih mengutamakan satu sisi dari sentimen tersebut. Tantangan ini dapat menyebabkan ketidakseimbangan dalam distribusi data, yang pada gilirannya mempengaruhi kualitas pelatihan model dan performa analisis sentimen.

Untuk penelitian selanjutnya, disarankan agar dataset yang digunakan diperbesar, baik dalam hal jumlah kalimat maupun keberagaman topik. Penambahan dataset yang lebih besar dapat membantu model dalam mengidentifikasi pola sentimen yang lebih beragam, sehingga meningkatkan akurasi dan kemampuan generalisasi model terhadap data yang belum terlihat sebelumnya. Selain itu, menggunakan data dari berbagai sumber (misalnya, ulasan produk dari berbagai industri selain otomotif) dan melibatkan variasi aksen dan gaya bicara yang lebih banyak dapat memperkaya model dalam mendeteksi sentimen secara lebih luas. Penelitian lebih lanjut juga dapat mempertimbangkan penggunaan teknik augmentasi data untuk meningkatkan jumlah data pelatihan tanpa harus mengumpulkan data baru secara manual.

Referensi

- [1] S. Thomas, T. Informasi, and I. Shanti Bhuana, "Studi Analisis Metode Analisis Sentimen pada YouTube," *JIFOTECH (JOURNAL OF INFORMATION TECHNOLOGY)*, vol. 1, no. 1, 2021.
- [2] I. A. Safra and E. Zuliarso, "ANALISA SENTIMEN PERSEPSI MASYARAKAT TERHADAP PEMINDAHAN IBUKOTA BARU DI KALIMANTAN TIMUR PADA MEDIA SOSIAL TWITTER," *Proceeding SENDIU*, pp. 214–219, 2020.
- [3] M. Yunus, M. Husni, and M. M. Mufadhdhal, "Klasifikasi Sentimen Terhadap Badan Penyelenggara Jaminan Sosial (BPJS) Pada Media Sosial Twitter Menggunakan Naive Bayes," *SMATIKA JURNAL*, vol. 11, no. 02, pp. 81–91, Dec. 2021, doi: 10.32664/smatika.v11i02.577.
- [4] A. A. Permana *et al.*, "SENTIMEN ANALISIS OPINI MASYARAKAT PADA MEDIA SOSIAL TWITTER TERHADAP VAKSIN BERBAYAR MENGGUNAKAN METODE NAÏVE BAYES CLASSIFIER (NBC)," *Jurnal Teknik FT UMT*, vol. 10, no. 2, pp. 84–92, 2021, [Online]. Available: <http://jurnal.umt.ac.id/index.php/jt/index>
- [5] K. Nugroho and F. N. Hasan, "Analisis Sentimen Masyarakat Mengenai RUU Perampasan Aset Di Twitter Menggunakan Metode Naïve Bayes," *SMATIKA JURNAL*, vol. 13, no. 02, pp. 273–283, Dec. 2023, doi: 10.32664/smatika.v13i02.899.

- [6] A. Fauzi, M. Faittullah Akbar, Y. Ferdi, and A. Asmawan, "Sentimen Analisis Berinternet Pada Media Sosial dengan Menggunakan Algoritma Bayes," *JURNAL INFORMATIKA*, vol. 6, no. 1, pp. 77–83, 2019, [Online]. Available: <http://ejournal.bsi.ac.id/ejurnal/index.php/ji/article/view/5437>
- [7] A. A. Ningtyas, A. Solichin, and R. Pradana, "ANALISIS SENTIMEN KOMENTAR YOUTUBE TENTANG PREDIKSI RESESI EKONOMI TAHUN 2023 MENGGUNAKAN ALGORITME NAÏVE BAYES," *Bit (Fakultas Teknologi Informasi Universitas Budi Luhur)*, vol. 20, no. 1, pp. 9–16, 2023.
- [8] A. Karimah and G. Dwilestari, "ANALISIS SENTIMEN KOMENTAR VIDEO MOBIL LISTRIK DI PLATFORM YOUTUBE DENGAN METODE NAIVE BAYES," 2024. [Online]. Available: <https://www.kaggle.com/datasets/billycemerson/anali>
- [9] G. Hakobyan, "From Voice to Text: The Evolution of Speech-to-Text APIs," *krisp.ai*. Accessed: Nov. 15, 2024. [Online]. Available: <https://krisp.ai/blog/speech-to-text-api-evolution/>
- [10] O. Koalu and S. Reinaldo Unsratdianto Aldo Sompie, "Rancang Bangun Aplikasi Pengenalan Bahasa Tountemboan Menggunakan Speech Recognition," *Jurnal Teknik Informatika*, vol. 14, no. 2, 2019.
- [11] M. A. Subarkah *et al.*, "Analisis Sentimen Terhadap Video Ulasan Produk Menggunakan Metode Support Vector Machine Dengan Sequential Minimal Optimization," *Jurnal Riset Inovasi Bidang Informatika Dan Pendidikan Informatika (KERNEL)*, vol. 3, no. 2, 2022.
- [12] I. N. Purnama and N. N. Widya Utami, "IMPLEMENTASI PERINGKAS DOKUMEN BERBAHASA INDONESIA MENGGUNAKAN METODE TEXT TO TEXT TRANSFER TRANSFORMER (T5)," *Jurnal Teknologi Informasi dan Komputer*, vol. 9, p. 4, 2023.
- [13] A. Mastropaolo *et al.*, "Studying the usage of text-to-text transfer transformer to support code-related tasks," in *Proceedings - International Conference on Software Engineering*, IEEE Computer Society, May 2021, pp. 336–347. doi: 10.1109/ICSE43902.2021.00041.
- [14] C. Raffel *et al.*, "Exploring the Limits of Transfer Learning with a Unified Text-to-Text Transformer," Oct. 2019, [Online]. Available: <http://arxiv.org/abs/1910.10683>
- [15] M. K. Shaik Vadla, M. A. Suresh, and V. K. Viswanathan, "Enhancing Product Design through AI-Driven Sentiment Analysis of Amazon Reviews Using BERT," *Algorithms*, vol. 17, no. 2, Feb. 2024, doi: 10.3390/a17020059.
- [16] K. F. Ramdhanian, D. F. Hidayat, and R. Salkiawati, "Implementasi Metode Naïve Bayes dan Support Vector Machine (SVM)," *JMPM: Jurnal Matematika dan Pendidikan Matematika*, 2024.
- [17] K. Pipalia, R. Bhadja, and M. Shukla, "Comparative analysis of different transformer based architectures used in sentiment analysis," in *Proceedings of the 2020 9th International Conference on System Modeling and Advancement in Research Trends, SMART 2020*, Institute of Electrical and Electronics Engineers Inc., Dec. 2020, pp. 411–415. doi: 10.1109/SMART50582.2020.9337081.
- [18] S. Tabinda Kokab, S. Asghar, and S. Naz, "Transformer-based deep learning models for the sentiment analysis of social media data," *Array*, vol. 14, Jul. 2022, doi: 10.1016/j.array.2022.100157.
- [19] R. F. Fahturohman, N. P. Putra, P. Santri, and A. P. Sari, "Implementasi Speech to Text untuk Mempermudah Catatan Praktik Diagnosis Pasien Dengan Metode NLP," *Journal of*

Multidisciplinary Inquiry in Science Technology and Educational Research, vol. 1, no. 3c, pp. 1456–1463, 2024, doi: 10.32672/mister.v1i3c.1935.

- [20] S. Shon, P. Brusco, J. Pan, K. J. Han, and S. Watanabe, “Leveraging Pre-trained Language Model for Speech Sentiment Analysis.”