

# Implementasi *Convolution-Augmented Transformer* Berbasis Kecerdasan Buatan dalam Analisis Sentimen Teks Hasil Konversi Suara ke Teks

<sup>1</sup>Prabowo Budi Utomo, <sup>2</sup>Dona Wahyudi, <sup>3</sup>Adimas Ketut Nalendra

<sup>1,2,3</sup> Administrasi Server dan Jaringan Komputer, Akademi Komunitas Negeri Putra Sang Fajar Blitar  
E-mail: <sup>1</sup> prabowo86@akb.ac.id, <sup>2</sup> donawahyudi@akb.ac.id, <sup>3</sup> dimas@akb.ac.id

*Corresponden Author:* prabowo86@akb.ac.id

Diterima Redaksi: 13 februari 2024 Revisi Akhir: 09 Maret 2024 Diterbitkan Online: 12 Maret 2024

**Abstrak** – Keterbatasan mendapatkan informasi yang dialami penyandang disabilitas menjadikan mereka kurang update terhadap perkembangan yang terjadi sehingga secara tidak langsung mendorong berbagai upaya untuk mendapatkan informasi tanpa mempedulikan sumber dan konteks informasi yang diperoleh, konteks informasi yang diperoleh dalam hal ini berkaitan dengan emosi yang berusaha disampaikan lewat tulisan atau teks informasi, maka perlu dirancang dan diimplementasikan yang mampu mengekstraksi dan menemukan inteprestasi emosi bermuatan positif, negatif atau netral dari teks hasil konversi teknologi *Speech to Text*, sehingga dapat membantu penderita disabilitas pendengaran dalam memahami konteks dan emosi yang terkandung didalam informasi. Aplikasi *Speech to Text* yang dikombinasikan dengan metode *Conformer* berbasis kecerdasan buatan dapat membantu penyandang disabilitas pendengaran untuk memahami sentimen atau emosi dari teks hasil konversi suara. Dengan menggunakan kecerdasan buatan yang tergabung dalam metode *Conformer* dapat dilakukan klasifikasi sentimen terhadap teks hasil konversi juga dapat dideteksi topik yang disampaikan, sehingga diharapkan dapat dimanfaatkan penyandang disabilitas pendengaran dalam memberikan umpan balik yang tidak menyinggung perasaan dan sesuai topik bahasan.

**Kata Kunci** — Penyandang disabilitas pendengaran, *Speech to Text*, Sentiment Analysis, *Conformer*, Kecerdasan Buatan

**Abstract** – The limitations of obtaining information experienced by people with disabilities make them less updated on developments that occur so that indirectly encourage various efforts to obtain information regardless of the source and context of the information obtained, the context of the information obtained in this case is related to the emotions that are trying to be conveyed through writing or text information, so it is necessary to design and implement an application that is able to extract and find positive, negative or neutral emotional interpretations from the text converted by *Speech to Text* technology, so that it can help people with hearing disabilities in understanding the context and emotions contained in the information. *Speech to Text* applications combined with the artificial intelligence-based *Conformer* method can help people with hearing disabilities to understand sentiment or emotion from voice conversion text. By using artificial intelligence incorporated in the *Conformer* method, the sentiment classification of the converted text can also be detected, so it is hoped that it can be utilized by people with hearing disabilities in providing feedback that is not offensive and according to the topic of discussion.

**Keywords** — People with hearing disabilities, *Speech to Text*, Sentiment Analysis, *Conformer*, Artificial Intelligence



## 1. PENDAHULUAN

Akses terhadap informasi menjadi salah satu hak yang dijamin oleh negara dan dunia sebagaimana diatur dalam Undang-undang Dasar Negara Republik Indonesia Tahun 1945 Pasal 28F yang menyebutkan bahwa setiap Orang berhak untuk berkomunikasi dan memperoleh Informasi untuk mengembangkan pribadi dan lingkungan sosialnya, serta berhak untuk mencari, memperoleh, memiliki, dan menyimpan Informasi dengan menggunakan segala jenis saluran yang tersedia. Dalam perkembangannya, akses informasi cenderung hanya tersedia untuk manusia normal dan belum banyak yang memfasilitasi penyandang disabilitas, padahal sebagaimana yang

diketahui semua makhluk diciptakan dengan kekurangan masing-masing dan berhak mendapatkan informasi yang layak. Keterbatasan mendapatkan informasi yang dialami penyandang disabilitas menjadikan mereka kurang *update* terhadap perkembangan yang terjadi sehingga secara tidak langsung mendorong berbagai upaya untuk mendapatkan informasi tanpa mempedulikan sumber dan konteks informasi yang diperoleh, konteks informasi yang diperoleh dalam hal ini berkaitan dengan emosi yang berusaha disampaikan lewat tulisan atau teks informasi [1]. Hal ini juga mendorong berbagai teknologi yang berupaya mengatasi keterbatasan yang dimiliki para penyandang disabilitas seperti konversi suara ke teks menggunakan teknologi *Speech to Text* (STT) maupun *Automatic Speech Recognition* (ASR), namun keberadaan teknologi belum mampu mengatasi kelemahan penyandang disabilitas dalam memahami konteks informasi, teknologi ini belum mampu menyampaikan atau mengklasifikasi informasi yang disampaikan sesuai konteks atau emosi yang ingin disampaikan. Seperti yang ditunjukkan didalam contoh berikut dimana dalam sebuah berita Gao Fu, kepala Pusat Pengendalian dan Pencegahan Penyakit China, pada hari Sabtu (10/04) mengatakan dalam jumpa pers bahwa vaksin yang ada saat ini "tidak memberi tingkat perlindungan yang sangat tinggi" (<https://www.youtube.com/watch?v=FqKzmJb-Ln0>), pernyataan yang disampaikan didalam berita diatas oleh teknologi *Speech to Text* (STT) akan disampaikan secara datar tanpa ada penekanan/penegasan sehingga tidak dapat menunjukkan emosi negatif dalam berita yang disampaikan yang secara tidak langsung juga dipahami sama oleh penyandang disabilitas, hal ini menyebabkan urgensi dari berita yang disampaikan belum dapat dipahami dengan baik oleh penyandang disabilitas.

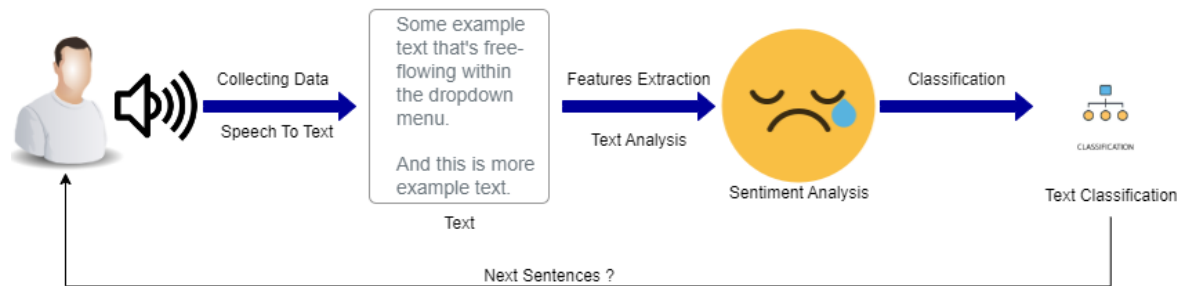
Berkembangnya teknologi penunjang ini telah mendorong terciptanya berbagai aplikasi yang mampu membuat komputer lebih memahami, menganalisa dan mengolah bahasa manusia salah satunya melalui pemanfaatan algoritma *Natural Language Processing* dalam bentuk *Sentiment Analysis*. *Sentiment Analysis* merupakan jenis pengolahan kata berbasis *Natural Language Processing* yang berfokus dalam analisa data berbentuk teks untuk mengidentifikasi pendapat dari teks tersebut. *Sentiment Analysis* dipergunakan untuk mengekstraksi dan menemukan pendapat yang bermuatan positif, negatif atau netral [2], Hasil akhir dari proses *Sentiment Analysis* dapat memvisualisasikan klasifikasi teks sehingga dapat dijadikan pertimbangan untuk mengetahui kondisi emosi pembicara [3]. Dalam perkembangannya, sentimen analysis sudah mampu menemukan interpretasi emosi atau pendapat yang terdapat didalam data suara, dikombinasikan dengan dengan teknologi *Speech to Text* (STT) atau *Automatic Speech Recognition* (ASR) dapat dikembangkan mekanisme sentimen analysis yang dapat menemukan inteprestasi emosi dari data suara sehingga lebih mudah memahami pandangan atau perasaan yang terkandung. Seperti yang dilakukan oleh Elisan & Herny [4] yang memanfaatkan algoritma *Naïve Bayes Classifier* dalam melakukan *sentimen analysis* terhadap data opini *public* program kampus merdeka, namun penggunaan algoritma *Naïve Bayes Classifier* menghasilkan nilai akurasi yang kurang memuaskan sebesar 57% yang disebabkan oleh jumlah data yang sedikit.

Dalam penelitian lain, Febry & Addin [5] melakukan klasifikasi sentimen menggunakan algoritma *Naïve Bayes Multinomial Text* yang mampu menghasilkan klasifikasi pada data berbentuk dokumen dengan nilai akurasi sangat tinggi yaitu 99,62% dalam waktu 0,16 detik, hal ini disebabkan algoritma *Naïve Bayes Multinomial Text* tidak memerlukan pencarian nilai *Evidence* dalam proses klasifikasi. Penggunaan algoritma *Naïve Bayes* dalam *sentiment analysis* memberikan hasil yang bermacam-macam, tidak ada hasil yang konsisten dan komperhensif untuk jenis data tertentu yang secara tidak langsung memerlukan metode pendamping sehingga dapat diperoleh data yang lebih jelas dan lengkap untuk mendapatkan hasil klasifikasi sentimen yang lebih baik. Dalam penelitian yang dilakukan Oscar B. Deho, dkk [6] dipergunakan *Skip-gram* yang merupakan varian dari *Word2Vec* dari teknik ekstraksi *Word Embedding* dalam meningkatkan akurasi klasifikasi sentimen, hal yang hampir sama juga dilakukan oleh Alif Rizal, dll [7] yang menggunakan *Word Embedding Layers* dengan dikombinasikan model klasifikasi *Long Short-Term Memory Network* (LSTM) dalam melakukan klasifikasi sentimen dan mampu menghasilkan nilai *precision* 80%, nilai *recall* 81%, *f1-score* 97% dan *accuracy* 81% dengan jumlah data yang terbatas, sehingga dapat disimpulkan bahwa teknik ekstraksi *Word Embedding* mampu meningkatkan akurasi model klasifikasi sentimen.

Penggunaan teknologi *Speech to Text* (STT) maupun *Automatic Speech Recognition* (ASR) tidak dapat dipungkiri mampu menjadi solusi bagi penyandang disabilitas khususnya disabilitas pendengaran dalam upaya mendapatkan informasi dari data suara atau video, namun teks hasil konversinya belum mampu menginterpretasikan emosi yang terkandung. Dalam beberapa hal salah satu penyebab belum diperolehnya interpretasi emosi dari teks hasil konversi STT atau ASR adalah kurangnya akurasi teks hasil konversi, dalam upaya meningkatkan akurasi teks hasil konversi maka dipergunakan metode penunjang teknologi *Speech to Text* (STT) bernama *Conformer* [8]. *Conformer* merupakan pengembangan dari metode *Transformer* [9] yang dikombinasikan dengan *Convolution Neural Network* (CNN) dengan ditunjang *Feed Forward Module* [8], *Conformer* dikembangkan dengan tujuan untuk meningkatkan akurasi konversi *Speech to Text*. Didalam penelitian ini akan dirancang dan diimplementasikan teknologi *Speech to Text* (STT) dengan metode *Conformer* berbasis kecerdasan buatan yang mampu mengekstraksi dan menemukan inteprestasi emosi bermuatan positif, negatif atau netral dari teks hasil konversi, dengan harapan dapat membantu penderita disabilitas pendengaran dalam memahami konteks dan emosi yang terkandung didalam sebuah informasi suara.

## 2. METODE PENELITIAN

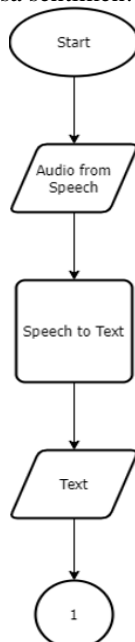
Pemanfaatan teknologi *Speech to Text* (STT) maupun *Automatic Speech Recognition* (ASR) telah mampu menjadi salah satu solusi bagi penyandang disabilitas pendengaran untuk memperoleh informasi dari video atau audio, yang secara tidak langsung dapat juga dimanfaatkan untuk memperoleh interpretasi emosi dari teks hasil konversinya apabila dikombinasikan dengan *sentiment analysis*. Berdasar pada hal tersebut, maka dilakukan penelitian yang berupaya fokus dalam memberikan solusi bagi penyandang disabilitas pendengaran, dimana penelitian yang dilakukan merupakan penelitian eksperimental yang berupaya menemukan sentimen dari teks hasil konversi suara yang dihasilkan dari metode *Speech to Text* yang dikombinasikan dengan metode *Conformer* berbasis kecerdasan buatan. Penelitian ini bersifat deskriptif dimana peneliti akan menggambarkan obyek yang diteliti dan menjabarkan hasil analisa sentimen dan pengujian yang dilakukan. Terkait dengan hal tersebut, terdapat beberapa tahapan penelitian yang dilakukan untuk mendapatkan analisa sentimen sesuai yang diharapkan, secara umum tahapan penelitian yang dilakukan mengacu pada penerapan salah satu model SDLC yaitu model *waterfall* [10], dimana berdasar model tersebut dikembangkan tiga tahapan penelitian yang akan dilaksanakan yaitu pengumpulan data, ekstraksi fitur, dan analisa sentimen yang ditunjukkan sebagaimana gambar 1.



Gambar 1. Tahapan Penelitian

### 2.1 Pengumpulan Data

Mekanisme pengumpulan dataset dilaksanakan dengan memanfaatkan teknologi *Speech to Text* (STT) yang ditujukan untuk menerjemahkan ucapan atau suara manusia ke dalam bentuk teks sesuai dengan bahasa yang dipergunakan. Secara umum penggunaan teknologi *Speech to Text* (STT) akan difokuskan dalam memperoleh data teks hasil konversi dari informasi yang disampaikan lewat audio atau video, dimana data teks yang diperoleh akan dikelola menjadi inputan dalam proses analisa sentimen.



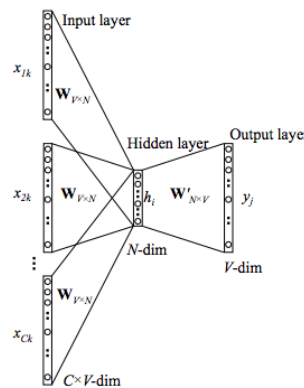
Gambar 2. Flowchart Pengumpulan Dataset

Dalam upaya meningkatkan akurasi teks hasil konversi maka teknologi *Speech to Text* (STT) akan dikombinasikan dengan metode *Conformer* berbasis kecerdasan buatan, dimana metode *Conformer* yang dipergunakan memiliki dua lapisan model yang mengabungkan model pengenalan *Transformer* dengan *Convolution Neural Network* (CNN) yang mampu memodelkan ketergantungan berbasis lokal dan global dalam bentuk urutan audio secara lebih efisien [8], dimana model pengenalan yang dipergunakan telah banyak dipergunakan sebagai model pengenalan dalam teknologi *Speech to Text* (STT) berbasis kecerdasan buatan.

### 2.2 Ekstraksi Fitur

Mengidentifikasi teks berdasarkan suara manusia banyak dipengaruhi oleh variasi dan karakteristik unik dari pembicara, dimana dalam tahapan ini setelah dilakukan konversi suara, teks hasil konversi dipergunakan sebagai *Data Collection* yang akan diekstraksi sebelum dilakukan analisa sentimen. Fungsional dari data teks secara tidak langsung memerlukan mekanisme ekstraksi untuk memperoleh nilai masing-masing fitur yang terdapat didalam data input [11], mekanisme ekstraksi fitur juga berperan penting dalam proses *sentiment analysis* sehingga berpengaruh terhadap hasil klasifikasi [12].

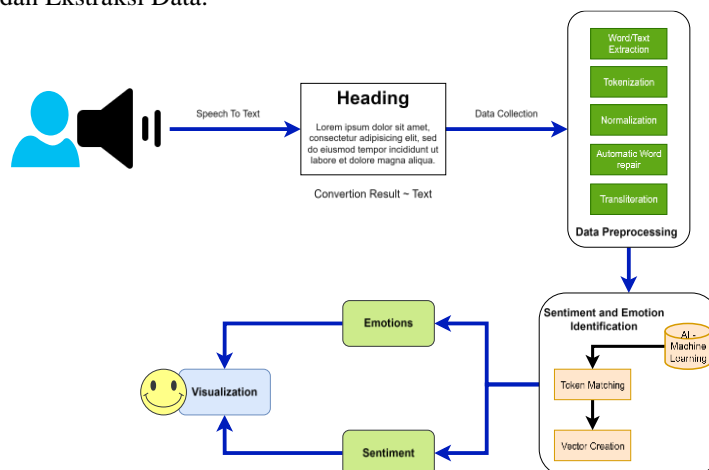
Represitasi teks sebagai salah satu hasil dari ekstraksi fitur cukup mempengaruhi akurasi sentimen, oleh karena itu peneliti menggunakan teknik *Word embedding* dalam ekstraksi fitur pada data teks hasil konversi *Speech to Text*. Metode ekstraksi fitur yang umum digunakan seperti *Bag of Word (BOW)* atau *Lexicon*, dimana BOW sering dipergunakan menghasilkan fitur dalam analisa sentimen yang melibatkan machine learning, sedangkan *Lexicon* merupakan kumpulan kata-kata seperti kamus yang menghasilkan klasifikasi sentimen. Bentuk ekstraksi fitur dengan pendekatan BOW dilakukan dengan mengubah teks mentah ke bentuk BOW yang didahului dengan proses pembersihan dan homogenisasi data, yang diikuti dengan pemberian bobot yang didasarkan pada frekuensi kemunculan data dan jumlah kata, sehingga model BOW akan menghasilkan jumlah fitur yang banyak dan memiliki tingkat dimensi yang tinggi. Dengan mengacu pada kemiripan antar kata dalam satu kalimat, dimana penggunaan *Word embedding* mampu memperoleh representasi *vektor* teks berdasar kemiripan antar teks sebagai invarian [13], maka dalam penelitian ini dimanfaatkan teknik *glove* dari *Word Embedding* yang merupakan gabungan dari model *skip-gram* dan model *CBOW* untuk melakukan ekstraksi fitur dari teks hasil konversi sesuai dengan kemiripan antar kata dalam satu kalimat, sehingga dapat diperoleh fitur dan dimensi yang lebih representatif.



Gambar 3. Multiple Context Word CBOW

### 2.3 Sentiment Analysis

Analisa sentimen merupakan teknologi yang memungkinkan sistem komputer untuk membaca dan memahami emosi manusia dan membuat keputusan berdasar emosi tersebut. Terdapat beberapa proses yang dilakukan dalam analisa sentimen yaitu *Data Collection*, *Data Preprocessing* dan *Sentiment and Emotion Identification*, didalam penelitian ini proses *Data Collection* dan *Data Preprocessing* dilakukan melalui tahap Pengumpulan Dataset dan Ekstraksi Data.



Gambar 4. Model Sentiment Analysis

Dalam melakukan identifikasi emosi, *vektor* teks yang merupakan hasil dari ekstraksi fitur dipergunakan sebagai inputan yang diolah untuk mendapatkan inteprestasi emosi, dimana inteprestasi emosi yang diperoleh

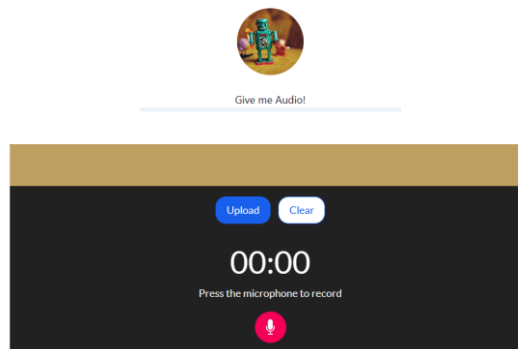
dikelompokkan sesuai kecenderungan opini yang nantinya mengarah pada muatan emosi positif atau emosi negatif atau emosi netral [15]. *Sentimen analysis* dilakukan pada seluruh teks hasil konversi yang telah melalui proses ekstraksi, untuk selanjutnya dilakukan ekstraksi sentimen global dari teks panjang yang mengandung pola lokal, disebabkan banyaknya *noise* maka diperlukan pemahaman terkait struktur kata dan hubungan antar kata dan frasa [16], sedangkan untuk paragraf akan dipecah menjadi beberapa kalimat dimana *polaritas* setiap kalimat akan diidentifikasi [17].

Penggunaan kedua model analisa yaitu *Speech to Text* (STT) yang dikombinasikan dengan metode *Conformer* berbasis kecerdasan buatan yang dikombinasikan dengan model *word embedding* sebagai teknik ekstraksi dalam penelitian ini menjadi kunci yang diharapkan dapat menemukan inteprestasi emosi bermuatan positif, negatif atau netral dari teks hasil konversi. Interpretasi emosi yang diperoleh akan dikelompokkan kedalam beberapa warna yang menjadi penanda tertentu pada jenis emosi tertentu, dalam penelitian ini akan diperggunakan 3 jenis penanda warna yaitu warna merah untuk emosi negatif, warna hijau muda untuk emosi positif dan warna abu-abu muda untuk emosi netral. Semua yang dilakukan dalam penelitian ini diharapkan dapat membantu penderita disabilitas khususnya penderita disabilitas pendengaran dalam memahami konteks dan emosi yang terkandung didalam sebuah informasi suara.

### 3. HASIL DAN PEMBAHASAN

#### 3.1 Implementasi Aplikasi

Pengembangan aplikasi *sentiment analysis* dari teks hasil konversi *Speech to Text* dirancang untuk memperoleh sentimen atau emosi yang terkandung didalam teks, yang diharapkan mampu mempermudah penyandang disabilitas pendengaran untuk lebih merasakan maksud yang ingin disampaikan oleh teks. Pengembangan aplikasi ditujukan untuk mengetahui kekurangan pada metode yang digunakan sekaligus memudahkan pengguna.



Gambar 5. Tampilan aplikasi

Dalam aplikasi terdapat fitur microphone yang dapat diperggunakan dalam proses perekaman suara, yang selanjutnya dapat diketahui teks hasil konversinya menggunakan fitur upload yang telah tersedia. Penggunaan *Speech to Text* (STT) yang dikombinasikan dengan metode *Conformer* berbasis kecerdasan buatan dalam proses klasifikasi sentimen bertujuan untuk meningkatkan akurasi dari teks hasil konversi sehingga dapat diperoleh kandungan sentimen dari teks diharapkan.



Salah satu johnto bahasa pomogrami atala python bahasa pomograman python panya tikunakan undomambat applicasi website applicasi desktop dan banya lucky ini manunjang kapopulan python tikalangan **programmer** tikaranakan kamotahanya.

Gambar 6. Hasil Konversi dengan *Speech to Text*

#### 3.2 Analisa Sentimen

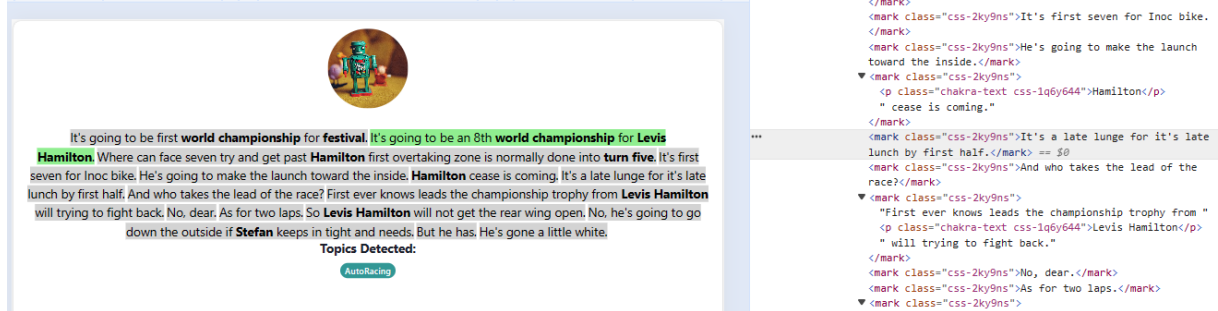
Teks hasil konversi yang diperggunakan sebagai dataset akan diekstraksi sebelum dianalisa sentimennya, Mekanisme ekstraksi fitur dilakukan dengan membagi kalimat ke dalam *box* untuk dipecah, pemecahan kalimat dilakukan dengan mempertimbangkan jumlah kata dan frekuensi kemunculannya sekaligus juga memperhatikan struktur kalimat.

```
text = "I am so Handsome!" --> parts = ["I am", "was handsome"]
entities = [{ text: 'I am', entity_type: 'event' }]

const entityText = entities.map((e) => e.text);
const parts = text.split(new RegExp(`(${entityText.join('|')})`, 'g'))
```

Gambar 7. Salah satu coding untuk ekstraksi kata

Pengecekan dilakukan dengan mempertimbangkan kecocokan antar kata dengan memperhatikan struktur kalimat dan topik bahasan, penggunaan *function RegExp* memungkinkan untuk mencocokkan pola teks dari teks hasil konversi STT untuk kemudian dipecah menjadi *substring*, sebagaimana ditunjukkan dalam gambar berikut.



Gambar 8. Hasil Ekstraksi Fitur

Hasil ekstraksi fitur mampu memberikan pengelompokan kata sesuai dengan kecocokan antar kata dan kesesuaian topik bahasan yang terbagi dalam box, sehingga dapat mempermudah proses analisa sentimen. Disamping itu dilakukan pula penambahan *highlight* pada beberapa kata melalui penebalan kata untuk menunjukkan obyek dan topik bahasan yang coba disampaikan.

Kemiripan antar kata didasarkan pada kesamaan topik yang dibahas atau pada tata kalimat dari teks tersebut, disamping itu juga dilakukan yang menunjukkan seberapa penting kata tersebut, Hasil dari ekstraksi inilah yang akan dijadikan inputan dalam proses *sentimen analysis*. Hasil ekstraksi kata sebagaimana penjelasan ditunjukkan pada Gambar berikut.



Gambar 9. Ekstraksi Fitur dengan *highlight*

Setelah diperoleh fitur dalam bentuk transkrip, selanjutnya dilakukan analisa sentimen terhadap masing-masing transkripnya. Kecerdasan Buatan dipergunakan sebagai mekanisme analisa yang akan menemukan kandungan sentimen yang terdapat pada masing-masing transkrip, penentuan sentimen dilakukan berdasar interpretasi dari transkrip yang dianalisa.

```
const APIKey = ...

const assemblyApi = axios.create({
  baseURL: "https://api.assemblyai.com/v2/",
  headers: {
    authorization: APIKey,
    "content-type": "application/json",
  },
});

const initialState = { ...
}

function App() {
  const [audioDetails, setAudioDetails] = useState(initialState);
  const [transcript, setTranscript] = useState({ id: '' });
  const [isLoading, setIsLoading] = useState(false);

  const handleAudioUpload = async (audioFile) => {
    setIsLoading(true);

    const { data: uploadResponse } = await assemblyApi.post('/upload', audioFile);

    const { data } = await assemblyApi.post('/transcript', {
      audio_url: uploadResponse.upload_url,
      sentiment_analysis: true,
      entity_detection: true,
      iab_categories: true,
      language_detection: true,
    });
  };
}
```

Gambar 10. Salah satu coding untuk analisa sentiment

Penggunaan kecerdasan buatan dilakukan dengan memanfaatkan fitur API yang terdapat pada assemblyai, yang dipergunakan untuk melakukan analisa sentimen dan deteksi entitas dari transkrip hasil ekstraksi. Kemudian sentimen yang diperoleh akan diklasifikasikan menggunakan warna yang berbeda sesuai dengan jenis sentimennya, untuk sentimen *POSITIVE* akan berwarna *lightgreen*, sentimen *NEGATIVE* akan berwarna *red* dan, sentimen *NEUTRAL* akan berwarna *lightgray*.



Phenomena that happen infrequently and result in damage, injure or death, such as hurricanes, tornadoes, earthquakes, volcano eruption, and firestorms.

**Topics Detected:**

- Disasters

Gambar 10. Hasil analisa sentimen yang bermuatan *NEGATIVE*.

### 3.3 Hasil Pengujian

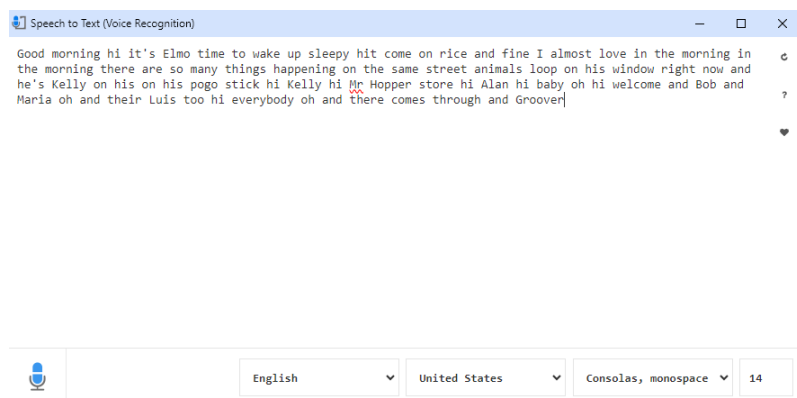
Pengujian dilakukan melalui beberapa mekanisme yang dilaksanakan secara simultan, beberapa mekanisme pengujian yang dilakukan yaitu pengujian kesesuaian teks hasil konversi dengan suara/audio yang disampaikan dan pengujian kesesuaian sentimen yang diperoleh dengan emosi yang terkandung.

Pengujian kesesuaian teks hasil konversi dilakukan melalui perbandingan teks hasil konversi dari aplikasi sentimen yang dikembangkan dengan teks hasil konversi yang diperoleh dari aplikasi lain seperti *Google Voice Recognition* untuk diketahui kemungkinan kesalahan didapatkan. Hasil dari perbandingan yang dilakukan dapat dilihat pada gambar 11 dan gambar 12 berikut.



Good morning. Hi, it's **Elmo**. Time to wake up, sleepy heap. Come on, rise and sign. **Elmo** loved in the morning in the morning there are so many things happening on the same street. **Elmo** looked out his windows right now and he see **Kelly** of his pogosti. Hi, **Kelly**. Hi, Mr. **Hooverstar**. Hi, **Alan**. Hi, **Gabby**. Hi. Oh, and there **Gordon** and **Bob** and **Maria**. Oh, and there **Lois**, too. Hi, everybody. Oh, and here comes **Julie** and **Groover**.

Gambar 11. Hasil *Speech to Text* menggunakan Bahasa Inggris



Gambar 12. Hasil *Speech to Text* menggunakan *Google Voice Recognition*

Hasil *Speech to Text* yang ditunjukkan dapat dilihat pada gambar 11 merupakan hasil konversi dari aplikasi yang dikembangkan dimana menunjukkan teks yang utuh sesuai dengan data suaranya, aplikasi dapat melakukan konversi dengan baik namun untuk jenis bahasa tertentu seperti bahasa Inggris, dikarenakan hasil yang berbeda akan diperoleh saat konversi menggunakan bahasa yang lain. Hasil konversi pada gambar 11 juga langsung diekstraksi per kata sesuai dengan kesesuaian antar kata juga dilakukan *highlight*/penekanan pada beberapa kata seperti kata yang mengandung nama, sedangkan dalam hasil konversi pada gambar 12 yang menggunakan *Google Voice Recognition* menghasilkan teks yang juga utuh sesuai dengan data suara yang diberikan, namun belum ada ekstraksi fitur dan *highlight*.

Dalam pengujian kesuaian sentimen yang diperoleh, hal ini didasarkan pada konsep analisa sentimen yang diperoleh berdasar inteprestasi dari transkrip hasil ekstraksi, sehingga cukup rawan untuk mengalami ketidaksesuaian dengan sentimen berdasar persepsi manusia. Maka dalam pengujian ini dilakukan perbandingan antara sentimen kecerdasar buatan per transkrip dengan sentimen persepsi manusia per transkrip.

Tabel 1 Perbandingan Sentimen

Transkrip	Terjemah	Kecerdasan Buatan	Manusia	Keterangan
More than 1.1 million people have died as a result of Covid-19 in the US and many more have been hospitalized or lost loved ones	Lebih dari 1,1 juta orang telah meninggal akibat Covid-19 di AS dan lebih banyak lagi yang dirawat di rumah sakit atau kehilangan orang yang dicintai	NEGATIVE	SAD	Sesuai
Faris Jr. Says, knowing that Levy wanted a participate but could not fill him with sadness	Faris Jr. mengatakan, mengetahui bahwa Levy ingin berpartisipasi tetapi tidak bisa mengisinya dengan kesedihan	NEGATIVE	SADNESS	Sesuai
It's my passion because I get to build bonds with the children. I truly love every minute of it.	Ini adalah gairah saya karena saya dapat membangun ikatan dengan anak-anak, saya benar-benar mencintai setiap menitnya.	POSITIVE	HAPPY	Sesuai
Is it going to be an eight world championship for Lewis Hamilton ?	Apakah ini akan menjadi kejuaraan dunia kedelapan bagi Lewis Hamilton?	NEUTRAL	TEGANG	Tidak Sesuai

Pengujian yang dilakukan melalui perbandingan hasil konversi ini secara tidak langsung menunjukkan kedua aplikasi dapat memberikan teks hasil konversi yang utuh sesuai dengan data suara, sehingga secara tidak langsung aplikasi sentimen yang dikembangkan memiliki performa konversi yang minimal sama dengan aplikasi *Google Voice Recognition* meskipun ada beberapa fitur yang tidak sama. Sedangkan dalam pengujian melalui kesuaian sentimen diperoleh hasil yang cukup memuaskan dengan 75% ditemukan kesesuaian sentimen.

Penggunaan *sentimen analysis* berbasis kecerdasan utama mampu memberikan sentimen yang cukup sesuai pada transkrip teks sehingga diharapkan dapat mempermudah penyandang disabilitas pendengaran untuk lebih memahami konteks atau emosi yang terkandung didalam teks. Hasil penelitian ini juga menunjukkan bahwa topik yang ingin disampaikan melalui teks dapat dideteksi melalui ekstraksi fitur dan analisa sentimen yang telah dilakukan, sehingga akan mempermudah penyandang disabilitas pendengaran untuk memberikan umpan balik yang sesuai topik bahasan.



#### 4. SIMPULAN

Pengembangan aplikasi *Speech to Text* yang dikombinasikan dengan metode *Conformer* berbasis kecerdasan buatan dapat membantu penyandang disabilitas pendengaran untuk memahami sentimen atau emosi dari teks hasil konversi suara. Dengan menggunakan kecerdasan buatan yang tergabung dalam metode *Conformer* dapat dilakukan klasifikasi sentimen terhadap teks hasil konversi juga dapat dideteksi topik yang disampaikan, sehingga diharapkan dapat dimanfaatkan penyandang disabilitas pendengaran dalam memberikan umpan balik yang tidak menyinggung perasaan dan sesuai topik bahasan.

#### DAFTAR PUSTAKA

- [1] R. Husnaini, "HADIS MENGENDALIKAN AMARAH DALAM PERSPEKTIF PSIKOLOGI."
- [2] RVS College of Engineering & Technology and Institute of Electrical and Electronics Engineers, *Proceedings of the International Conference on Inventive Research in Computing Applications (ICIRCA 2018) : date: July 11-12, 2018.*
- [3] H. Utama and A. Masruro, "Analisis Sentimen pada Twitter menggunakan Word Embedding dengan Pendekatan Word2Vec," 2022.
- [4] E. Febriyani and H. Februariyanti, "Analisis Sentimen Terhadap Program Kampus Merdeka Menggunakan Algoritma Naive Bayes Classifier Di Twitter," vol. 17, no. 1.
- [5] F. Eka Purwiantono and A. Aditya, "KLASIFIKASI SENTIMEN SARA, HOAKS DAN RADIKAL PADA POSTINGAN MEDIA SOSIAL MENGGUNAKAN ALGORITMA NAIVE BAYES MULTINOMIAL TEXT," *Jurnal TEKNOKOMPAK*, vol. 14, no. 2, pp. 68–73, 2020.
- [6] O. B. Deho, W. A. Agangiba, F. L. Aryeh, and J. A. Ansah, "Sentiment Analysis with Word Embedding."
- [7] A. Rizal Maulana, S. Hadi Wijoyo, and Y. T. Mursityo, "ANALISIS SENTIMEN KEBIJAKAN PENERAPAN KURIKULUM MERDEKA SEKOLAH DASAR DAN SEKOLAH MENENGAH PADA MEDIA SOSIAL TWITTER DENGAN MENGGUNAKAN METODE WORD EMBEDDING DAN LONG SHORT-TERM MEMORY NETWORKS (LSTM)," vol. 10, no. 3, pp. 523–530, 2023, doi: 10.25126/jtiik.2023106977.
- [8] A. Gulati *et al.*, "Conformer: Convolution-augmented Transformer for Speech Recognition," May 2020, [Online]. Available: <http://arxiv.org/abs/2005.08100>
- [9] *ICASSP 2020 - 2020 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. IEEE, 2020.
- [10] G. F. Avisyah, I. J. Putra, and S. S. Hidayat, "Open Artificial Intelligence Analysis using ChatGPT Integrated with Telegram Bot," *Jurnal ELTIKOM*, vol. 7, no. 1, pp. 60–66, Jun. 2023, doi: 10.31961/eltikom.v7i1.724.
- [11] T. S. Sabrila, V. R. Sari, and A. E. Minarno, "Analisis Sentimen Pada Tweet Tentang Penanganan Covid-19 Menggunakan Word Embedding Pada Algoritma Support Vector Machine Dan K-Nearest Neighbor," *Fountain of Informatics Journal*, vol. 6, no. 2, p. 69, Jul. 2021, doi: 10.21111/fij.v6i2.5536.
- [12] X. Chen, Y. Xue, H. Zhao, X. Lu, X. Hu, and Z. Ma, "A novel feature extraction methodology for sentiment analysis of product reviews," *Neural Comput Appl*, vol. 31, no. 10, pp. 6625–6642, Oct. 2019, doi: 10.1007/s00521-018-3477-2.
- [13] IEEE Staff, *2017 International Conference on Intelligent Communication and Computational Techniques (ICCT)*. IEEE, 2017.
- [14] Universitas Atma Jaya Yogyakarta, Institute of Electrical and Electronics Engineers. Indonesia Section. Computer Society Chapter., Institute of Electrical and Electronics Engineers. Indonesia Section., and Institute of Electrical and Electronics Engineers, *2019 International Conference of Artificial Intelligence and Information Technology (ICAIIIT 2019) : 13-15 March 2019, Platinum Adisucipto Yogyakarta Hotel & Conference Center.*
- [15] M. Z. Rahman, Y. A. Sari, and N. Yudistira, "Analisis Sentimen Tweet COVID-19 menggunakan Word Embedding dan Metode Long Short-Term Memory (LSTM)," 2021. [Online]. Available: <http://j-ptiik.ub.ac.id>
- [16] R. Arulmurugan, K. R. Sabarmathi, and H. Anandakumar, "Classification of sentence level sentiment analysis using cloud machine learning techniques," *Cluster Computing*, vol. 22. Springer Science and Business Media, LLC, pp. 1199–1209, Jan. 16, 2019. doi: 10.1007/s10586-017-1200-1.
- [17] P. Nandwani and R. Verma, "A review on sentiment analysis and emotion detection from text," *Social Network Analysis and Mining*, vol. 11, no. 1. Springer, Dec. 01, 2021. doi: 10.1007/s13278-021-00776-6.