Jurnal Media Pratama
Volume 17 No. 1 Edisi: Juni 2023, Halaman: 61 - 73
e-ISSN: 2721-9690

Aplikasi Speech Recognition untuk audio forensik

Speech Recognition application for audio forensics

Mukhlis Prasetyo Aji[1], Ulung Ghozy Aeman[2]
[1,2,]Universitas Muhammadiyah Purwokerto
Jl.Ahmad Dahlan Purwokerto, Indonesia
[1]prasetyo-aji@ump.ac.id, [2]ulungghozy@gmail.com

ABSTRAK

Konten digital berupa audio sebagai alat bukti pengadilan sering dirusak atau dimodifikasi. Alat bukti audio sering dibantah dalam hal penentuan dan penilaian keakuratan suara. Akibatnya, ada persyaratan untuk sistem bantuan keputusan saat mengkonfirmasi forensik audio. Seorang saksi ahli diharapkan dapat memberikan analisis yang mendalam dan relevan secara otentik sebagai penentu nasib baik korban maupun pelakunya, maka setiap bukti suatu kejadian, terutama bukti audio, harus sesuai dengan hukum. Objektif : Tujuan dari penelitian ini untuk Merancang dan membuat aplikasi transkrip suara agar mempermudah tim forensik dalam analisis suara serta Program dapat merubah input suara kebentuk tulisan secara otomatis Metode : Menggabungkan sistem yang berbeda dengan menggunakan kembali bagian yang digunakan sebelumnya. Winston Royce memulai debut model Air Terjun untuk pertama kalinya pada tahun 1970. Model Air Terjun adalah model tradisional langsung yang menggambarkan aliran linier melalui sistem. Keluaran dari satu tahap akan menjadi masukan untuk tahap yang mengikutinya. Model ini, yang diambil dari prosedur rekayasa lainnya, menyajikan pendekatan yang lebih realistis untuk praktik rekayasa perangkat lunak dan memberikan keuntungan bagi penggunanya. Pendekatan ini mencakup tim Software Quality Assurance (SQA) dengan 5 fase, dan setiap tahap selalu divalidasi atau diuji sebelum melanjutkan ke tingkat berikutnya.. Hasil : Aplikasi dengan sebuah tools audio forensik yang dapat bekerja jika fungsi membaca file audio secara binary serta speech recognition beroperasi dengan sample rate 16KHz serta bereksistensi .wav. Kesimpulan : Dalam penelitian ini telah terbentuk sebuah tool audio forensik yang menggunakan auto speech recognition yang menggunakan algoritma machine learning secara open source yang dapat membantu investigator dalam mengangani kasus dengan barang bukti berupa file audio. Media player hanya dapat bekerja jika fungsi membaca file audio secara binary serta Speech Recognition hanya dapat beroperasi pada file audio dengan sampel rate 16 kHz dan dengan eksistensi file .wav. Generate nilai hash tidak dapat dilakukan karena sampel rate file harus dikonversi menjadi 16 kHz dengna eksistensi .wav.

Kata Kunci: Digital forensik audio, speech recognition, bukti digital,aplikasi forensik

ABSTRACT

Digital content in the form of audio as court evidence is often damaged or modified. Audio evidence is often challenged when it comes to determining and assessing the accuracy of sound. As a result, there is a requirement for decision-aid systems when confirming audio forensics. An expert witness is expected to be able to provide in-depth and authentically relevant analysis to determine the fate of both the victim and the perpetrator, so any evidence of an incident, especially audio evidence, must comply with the law. Objective: This research aims to design and create a voice transcription application to make it easier for the forensic team to analyze voice, and the program can automatically change voice input into written form. Method: Combining different systems by reusing previously used parts. Winston Royce debuted the Waterfall model for the first time in 1970. The Waterfall Model is a straightforward traditional model that depicts linear flow through a system. The output from one stage will be the input for the following stage. This model, taken from other engineering procedures, presents a more realistic approach to software engineering practice and benefits its users. This approach includes a Software Quality Assurance (SQA) team with 5 phases, and each phase is always validated or tested before proceeding to the next level. Results: Application with an audio forensic tool that can work if the function of reading audio files in binary and speech recognition is operational with a sample rate 16KHz and has a .wav existence. Conclusion: In this research, an audio forensic tool has been created that

uses auto speech recognition, which uses an open-source machine learning algorithm that can help investigators handle cases with evidence in the form of audio files. The media player can only work if the function reads audio files in binary, and Speech Recognition can only operate on audio files with a sample rate of 16 kHz and with the existence of .wav files. Generating hash values cannot be done because the file sample rate must be converted to 16 kHz with the presence of .wav.

Keywords: Audio forensik digital, pengenalan suara, bukti digital, aplikasi forensik

1. Pendahuluan

Pesatnya perkembangan teknologi telah mempengaruhi banyak factor, terutama di bidang forensic digital. Sebagai hasil dari perkembangan teknologi ini, banyak alat telah ditemukan untuk membantu forensic digital mendapatkan bukti yang lebih akurat. Setelah diadopsi, banyak pihak mengembangkan alat untuk keperluan forensic yang baik berbayar atau open source. Oleh karena itu, penulis berinisiatif untuk mengembangkan aplikasi transkripsi ucapan yang menggunakan auto speech recognition.

Salah satu bentuk bukti digital yang bisa didapatkan adalah audio. Kehati-hatian harus dilakukan untuk memastikan bahwa bukti yang sering diajukan di pengadilan tidak dapat dirusak, belum lagi bukti digital (Prayudi, et al., 2014). Menurut Imran dkk. (2017), bukti audio. Hal itu dilakukan agar pengadilan dapat memanfaatkan fakta bahwa beberapa orang terlibat dalam percakapan tersebut. Dalam banyak proses hukum, pengenalan bukti audio seringkali menjadi masalah yang diperdebatkan, terutama dalam menentukan dan mengevaluasi keakuratan rekaman audio. Oleh karena itu, diperlukan sistem pendukung keputusan untuk validasi audio forensik (Renza, et al, 2018). Karena peran sentral yang dimainkan bukti digital dalam proses penetapan kasus, penting untuk memiliki akses ke saksi ahli yang berpengalaman, khususnya di bidang ilmu forensik audio (Handoko, 2017).

Selama negosiasi, analisis otentik, beralasan dan relevan diperlukan dari ahli. Karena menentukan nasib baik korban maupun pelaku, maka semua alat bukti kejahatan, terutama rekaman audio, harus sesuai dengan undang-undang. Standar Operasional Prosedur (SOP) dan sesuai dengan aturan tahap audio forensik DFAT PUSLABFOR harus diikuti selama serangkaian ujian. Pemeriksaan ini harus berdasarkan undang-undang (Putri & Sunarno, 2014). Menurut Maher (2018), Ketersediaan rekaman audio yang dikumpulkan merupakan faktor penting dalam analisis forensik audio. Rekaman wawancara, penyadapan, dan interogasi adalah jenis rekaman audio yang paling umum.

Semua rekaman audio yang ditemukan memiliki ciri khas tersendiri dan tidak ada jaminan bahwa kualitas rekaman yang dikumpulkan akan tinggi dalam hal apapun. Apabila telah terjadi tindak pidana dan diketahui rekaman audio tersangka adalah rekaman audio, maka harus dilakukan analisis untuk mengetahui keaslian rekaman audio yang ditemukan tersebut. identik dengan rekaman penulis. Nada rendah biasanya lebih mudah dikenali daripada nada tinggi. Metode berdasarkan perekam otomatis serta statistik otomatis dapat digunakan dalam proses menganalisis bukti digital yang disajikan dalam bentuk audio.

Berdasarkan latar belakang diatas, peneliti tertarik untuk "Rancang Bangun Aplikasi Audio Untuk Audio Forensik". Diharapkan dalam aplikasi ini dapat membantu pihak kampus atau masyarakat dalam menganalisa kejahatan berbasis pesan suara digital.

2. Tinjauan Pustaka

2.1 Audio

Yang termasuk gambar adalah bagan, grafik, foto, lukisan, iklan, dan sebagainya. Kelengkapan Audio merupakan suara yang dihasilkan oleh getaran suatu benda berupa sinyal analog dengan amplitude yang berubah secara terus agar dapat tertangkap oleh indera pendengaran manusia. Bila

membandingkan gender, gender perempuan memiliki tingkat keberhasilan lebih tinggi dibanding dengan laki-laki. Sehingga, apabila perlu melakukan pengujian adalah suara perempuan, sistem akan mudah mengenalinya sebagai suara dari user yang tidak berhak melakukan akses ke sistem. Sebaliknya, dalam pemakaian sebuah sistem masih memiliki kesulitan dalam membedakan suara yang diuji, apabila suara tersebut merupakan suara dari user gender pria (Imario, dkk., 2017).

Suara dihasilkan melalui dua buah proses yaitu Generation dan Filtering. Pada proses Generation, suara diproduksi dari pita suara yang berada pada laring manusia (vocal cord dan vocal fold) yang berada di larynx untuk menghasilkan bunyi periodik. Bunyi periodik yang sifatnya konstan tersebut kemudian disaring melalui vocal tract yang terdiri dari lidah (tongue), gigi (teeth), bibir (lips), langit-langit (palate) dan lain-lain sehingga bunyi tersebut dapat menjadi bunyi keluaran (output) berupa bunyi vokal (vowel) dan atau bunyi konsonan (consonant) yang membentuk kata-kata yang memiliki arti yang nantinya dapat dianalisis oleh voice recognition (Al-Azhar Nuh, 2011). Rekaman suara berisi gelombang bunyi yang direkam dengan teknologi digital recording (Mansyur, 2017).

Menurut (Al-Azhar Nuh, 2011) rekaman suara dimungkinkan mengandung noise atau disebut gangguan yang "didengar" orang lain, namun dalam istilah telekomunikasi kata noise juga dipakai untuk istilah gangguan yang menimbulkan kebisingan yang dapat didengar suatu sistem. Noise dapat terjadi dengan berbagai macam cara, misalnya rekaman suara barang bukti terdapat bocoran yang terjadi karena pada saat rekaman diperoleh tersangka berada pada lokasi dimana terdapat suara-suara yang saling bercampur. Suara yang demikian perlu adanya peningkatan untuk menaikkan kualitas rekaman, sehingga kosakata pembicaran dapat didengar jelas. Audio dalam bentuk digital :

a. Frekuensi (Hz)
   Frekuensi bunyi adalah jumlah getaran bunyi yang dihasilkan dalam waktu 1 detik. Satuan dari frekuensi bernama Hertz yang disingkat dengan Hz.
b. Intensitas (db / power)
   Intensitas suara mendeskripsikan amplitudo (tinggi) gelombang suara.
c. Sample Rate
   Sampel rate adalah spesifikasi untuk bagaimana komputer membaca file audio. Sederhana nya sebagain resolusi dari suara.

2.2. Audio Forensik

Audio forensik adalah sebagai "penggunaan audio dan penerapan ilmu pengetahuan yang terkait dengannya untuk menyelidiki dan membangun fakta-fakta di persidangan" (Zabri, 2006). Keaslian barang bukti berupa audio dalam menguatkan keputusan hakim dipersidangan perlu dilakukan serangkaian pengamatan dan tes untuk mengevaluasi keaslian rekaman (Mansyur, 2017).

Mengidentifikasi rekaman suara perlu dilakukan verikasi dengan suara pembanding yang hampir mirip, sehingga perlu adanya tahapan yang sesuai dengan SOP tentang analisa audio forensics pada DFAT PUSLABFOR yang mengacu pada Good Practice Guide for Computer-Based Electronic Evidence yang diterbitkan oleh ACPO di Inggris, dan Forensic Examination of Digital Evidence: A Guide for Law Enforcement yang diterbitkan oleh National Institute of Justice yang berada di bawah Department of Justice, Amerika Serikat (Al-Azhar Nuh, 2011). Analisa yang harus dilakukan pada audio forensics sebagai berikut:

a. Acquisition
   Proses pengambilan barang bukti yang asli kemudian dicatat menggunakan teknik audio recorder. Hal terpenting dalam rekaman suara sebagai barang bukti adalah mengungkap keidentikan antara suara pada rekaman barang bukti dengan suara yang diduga sebagai pemilik rekaman (Azhar, 2010). Rekaman suara asli seharusnya di backup terlebih dahulu, supaya tetap menjaga nilai keaslian dari barang bukti yang ditemukan. Proses akuisisi audio recorder akan menghasilkan file DD image.
b. Peningkatan Audio (audio enhancement)
   Hasil backup dari rekaman suara asli sebagai barang bukti diputar berkali-kali untuk melihat kualitas rekaman, jika suara yang didengar tidak bagus dikarenakan banyak noise. Suara yang

demikian perlu adanya peningkatan untuk menaikkan kualitas rekaman, sehingga kosakata pembicaran bisa jelas didengar.

c. Decoding
Setelah rekaman suara sudah bisa didengar dengan kosakata yang jelas, Kemudian suara dengan kosakata yang jelas dicatat untuk dijadikan transkrip rekaman. Transkrip rekaman berisi subjek label, waktu pengucapan surasa yang sesuai dengan berjalannya rekaman. Jika terdapat penulisan transkip masih ada suara yang tidak jelas , maka perlu dituliskan pada transkip keterangan tidak jelas atau tidak didengar.

d. Voice Recognition
Proses ini dilakukan untuk memastikan suara pada rekaman identik dengan suara pembanding. Analisis kemiripan atau identifikasi tersebut menggunakan parameter terhadap pitch, formant bandwidth, dan spectrogram. Kosakata yang didapat minimal dua puluh (20) kata yang memiliki kesamaan antara rekaman suara asli dan pembanding, jika kurang dari dua puluh kata maka tidak memenuhi syarat audio forensic.

2.3. Speech Recognition

Speech Recognition disebut sebagai Automatic Speeh Recognition (ASR), yang mana untuk mengenali ucapan (suara) manusia seperti kata, digit, kalimat menggunakan Algoritma yang di eksekusi pada computer. Sistem Automatic Speech Recogntion merubah sinyal suara menjadi bentuk digital, Contohnya Spectrum. (Ambewadikar and Baheti, 2020). Tujuan dari Automatic Speech Recognition (ASR) adalah salinan dari ucapan manusia menjadi kata lisan.

Hal tersebut merupakan tugas yang menantang karena sinyal ucapan manusia sangat bervariasi dikarenakan oleh macam-macam atribut pembicara, perbedaan gaya lisan, suara lingkungan yang tak menentu, dan lain-lain. ASR, dan selebihnya, perlu memetakan variabel panjang dari sinyal suara menjadi variabel panjang urutan kata atau simbol fonetik. Hal tersebut dikenal baik yang Hidden Markov Model (HMM) telah sukses dalam menangani urutan panjang variabel dan juga memodel perilaku sementara dari sinyal suara menggunakan urutan keadaan, masing-masing yang berkaitan dengan distribusi probabilitas tertentu dari observasi.

Model Pencampuran Gaussian (Gaussian Mixture Models (GMMs)). Telah sampai akhirnya dianggap sebagai model yang paling kuat untuk memperkirakan distribusi probabilitas dari sinyal suara terkit dengan masing-masing dari status HMM. (Abdel-Hamid et al., 2014). Auto Speech Recognition (ASR) juga disebut sebagai Speech Recognition, dapat didefinisikan sebagai tampilan grafis dari frekuensi yang dipancarkan sebagai fungsi waktu. Semua teknik pengolahan suara (sintesis dan pemrosesan suara, identifikasi pembicara, Verifikasi pembicara membuat hal tersebut mungkin untuk membuat antarmuka suara (Human Machine Interface) atau melakukan interaksi suara. (Ibrahim and Varol, 2020).

Speech Recognition merupakan disiplin dalam sebuah sub-field dari komputasi linguistik yang membangun teksik dan teknologi yang memfasilitasi pengenalan dan penerjemahan dari kata yang di ucapkan menjadi format teks oleh komputer. Hal tersebut juga dikenal sebagai “Speech to text” (STT). Hal tersebut termasuk ilmu dan penelitian dalam linguistic, ilmu komputer, dan lingkungan teknik kelistrikan. (Y. et al., 2017). Automatic Speech Recognition (ASR) mengijinkan komputer untuk mengidenfitikasi kata yang diucapkan seseorang pada mikrofon atau telefon dan merubahnya menjadi teks tertulis. Sebagai hasilnya dapat berpotensi menjadi mode yang sangan penting pada interaksi antara manusia dan komputer. (Vimala, 2012).

3. Metode

Menggabungkan sistem yang berbeda dengan menggunakan kembali bagian yang digunakan sebelumnya. Winston Royce memulai debut model Air Terjun untuk pertama kalinya pada tahun 1970. Model Air Terjun adalah model tradisional langsung yang menggambarkan aliran linier melalui sistem. Keluaran dari satu tahap akan menjadi masukan untuk tahap yang mengikutinya. Model ini, yang

diambil dari prosedur rekayasa lainnya, menyajikan pendekatan yang lebih realistis untuk praktik rekayasa perangkat lunak dan memberikan keuntungan bagi penggunanya. Pendekatan ini mencakup tim Software Quality Assurance (SQA) dengan 5 fase, dan setiap tahap selalu divalidasi atau diuji sebelum melanjutkan ke tingkat berikutnya.

3.1. Analisis Kebutuhan Fungsional

Kebutuhan fungsional adalah kebutuhan pada sistem yang merupakan layanan dalam aplikasi yang harus disediakan, serta gambaran proses dari reaksi sistem terhadap masukan sistem dan yang akan dikerjakan oleh sistem diantaranya adalah sebagai berikut:

a. Aplikasi dapat merubah input suara kebentuk tulisan secara otomatis.
b. Aplikasi memberikan visualisasi grafis suara.
c. Aplikasi mengenerate nilai hash dari file yang dibuka.

3.2. Perancangan Sistem

Dalam penelitian ini menggunakan program SpeechRecognition DeepSpeech. Deepspeech merupakan program open source Speech to Text yang menggunakan model yang dilatih menggunakan teknik machine learning berdasarkan pada penelitian Deep Speech oleh Baidu. Projek DeepSpeech menggunakan TensorFlow milik google untuk mempermudah implementasi. Deepspeech merupakan arsitektur jaringann saraf tiruan yang pertama di publikasikan oleh tim peneliti di Baidu. Pada tahun 2017 Mozilla membuat implementasi open source dari lembar penelitian tersebut yang di beri nama Mozilla DeepSpeech. Lembar asli DeepSpeech dari baidu mempopulerkan konsep end - to - end model speech recognition yang berarti model yang mendapat masukan audio dan secara langsung menghasilkan keluaran berupa karakter dan kata. Dibandingkan dengan model speech recognition tradisional seperti yang dibangun menggunakan library open source yang populer seperti Kaldi dan CMU Sphinx, yang memprediksi Fonem dan kemudian merubah Fonem tersebut ke bentuk kata pada proses downstream setelahnya.

Tujuan dari model end – to – end seperti DeepSpeech adalah untuk menyederhanakan elemen pemrosesan data dari speech recognition menjadi satu model. Sebagai tambahan teori yang dikenalkan oleh lembar penelitian Baidu yaitu melatih model deep learning besar pada jumlah data yang besar akan menghasilkan performa lebih bagus daripada model speech recognition model lama. Sekarang library DeepSpeech Mozilla menawarkan model pre-trained (Model yang sudah cerdas) speech recognition yang dapat anda bangun dan juga alat untuk melatih model DeepSpeech anda sendiri.

a. Pre-Trained Model (Model yang sudah cerdas)

Pada transfer learning, kita membutuhkan model yang sudah terlatih, entah itu dilatih oleh orang lain, atau kita melatihnya terlebih dahulu. Model yang sudah terlatih ini disebut dengan Pre-Trained Model. Pre-trained model biasanya sudah dilatih pada dataset yang besar dan merupakan dataset benchmark, sehingga kualitas pre-trained model harusnya sudah sangat baik.

Saat ini sudah banyak pre-trained model yang disediakan untuk beragam kebutuhan. Misalnya :

1. PyTorch menunjukkan pre-trained model yang tersedia secara resmi untuk image classification yang telah dilatih menggunakan dataset ImageNet.
2. Tensorflow juga menunjukkan model-model yang tersedia untuk object detection. Kumpulan pre-trained model ini juga biasa disebut dengan Model Zoo.

Alternatif lain selain menggunakan model resmi dari framework, kita juga bisa browsing-browsing dan pakai pre-trained model yang disediakan peneliti-peneliti lain, misalnya model SaimeseFC. Atau kita juga bisa saja membuat pre-trained model sendiri.

b. Tensorflow

TensorFlow 2.0, dirilis pada Oktober 2019, mengubah kerangka kerja dengan banyak cara berdasarkan masukan pengguna, untuk membuatnya lebih mudah digunakan (misalnya, dengan menggunakan API Keras yang relatif sederhana untuk pelatihan model) dan lebih berkinerja. Pelatihan terdistribusi lebih mudah dijalankan berkat API baru, dan dukungan untuk

TensorFlow Lite memungkinkan penerapan model di berbagai platform yang lebih besar. Namun, kode yang ditulis untuk versi TensorFlow sebelumnya harus ditulis ulang terkadang hanya sedikit, terkadang secara signifikan untuk memanfaatkan fitur TensorFlow 2.0 baru secara maksimal. TensorFlow memungkinkan developer membuat grafik aliran data struktur yang mendeskripsikan bagaimana data bergerak melalui grafik, atau serangkaian node pemrosesan. Setiap node dalam grafik mewakili operasi matematika, dan setiap koneksi atau tepi antar node adalah larik data multidimensi, atau tensor.Tensorflow menggabungkan banyak model dan algoritma machine learning yakni deep learning (neural network). Framework ini disusun menggunakan Python front-end API untuk membuat suatu aplikasi penggunaannya, dan menggunakan C++ yang memiliki kinerja terbaik dalam hal eksekusi. Tensorflow dapat melatih dan menjalankan neural network untuk keperluan mengklasifikasikan tulisan tangan, pengenalan gambar/object, serta menggabungkan suatu kata. Selanjutnya re-current neural network yang merupakan model sequential dapat digunakan untuk Natural Language Processing (NLP). Selain itu, tensorflow digunakan pada skala yang besar untuk produksi dengan menggunakan model yang sama ketika proses training data. Tensorflow ini merupakan library yang sangat populer pada kalangan data enthusiast terutama pelaku machine learning.

c. Akurasi
Untuk mengukur akurasi dari SpeechRecognition menggunakan aplikasi benchmark dari open source project Picovoice speech-to-text benchmark. Di dalam projek open source ini sudah terdapat beberapa engine speech recognition yang siap untuk dites tanpa harus memprogram secara manual, hanya perlu melakukan beberapa konfigurasi spesifik dan program benchmark siap di jalankan.

3.3. Model Perancangan Sistem atau Skenario Program

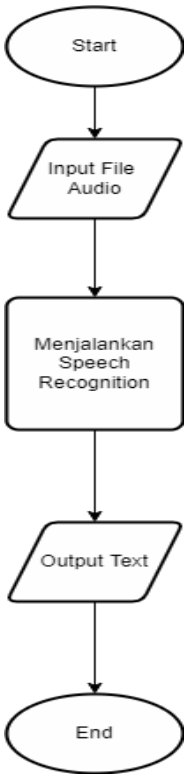
Aplikasi Auto transkrip ini bertujuan untuk dapat membantu DFC dalam mengangani kasus kejahatan digital dengan alat bukti berupa rekaman suara. Dengan aplikasi ini, DFC di mudahkan

analisa kasus berdasarkan alat bukti yang ada. Aplikasi ini bersifat open source, tidak ada batasan bagi semua orang jika ingin mengembangkan nya.

Penelitian Dalam pembuatan tugas akhir ini meliputi beberapa langkah:

1. Mencari Referensi
   Hal pertama yang penulis lakukan adalah mencari referensi menggunakan internet dan perpustakaan, setelah memiliki referensi yang cukup, penulis mengumpulkan data penting.
2. Perencanaan Program
   Selain itu, penulis bertanggung jawab atas desain dan pembuatan program; setelah desain selesai, penulis melanjutkan ke pembuatan program.
3. Pengujian
   Setelah pengembangan program selesai, penulis akan mengujinya, dan jika pengujian berhasil, penulis akan memeriksa hasil pengujian.
4. Penyusunan laporan

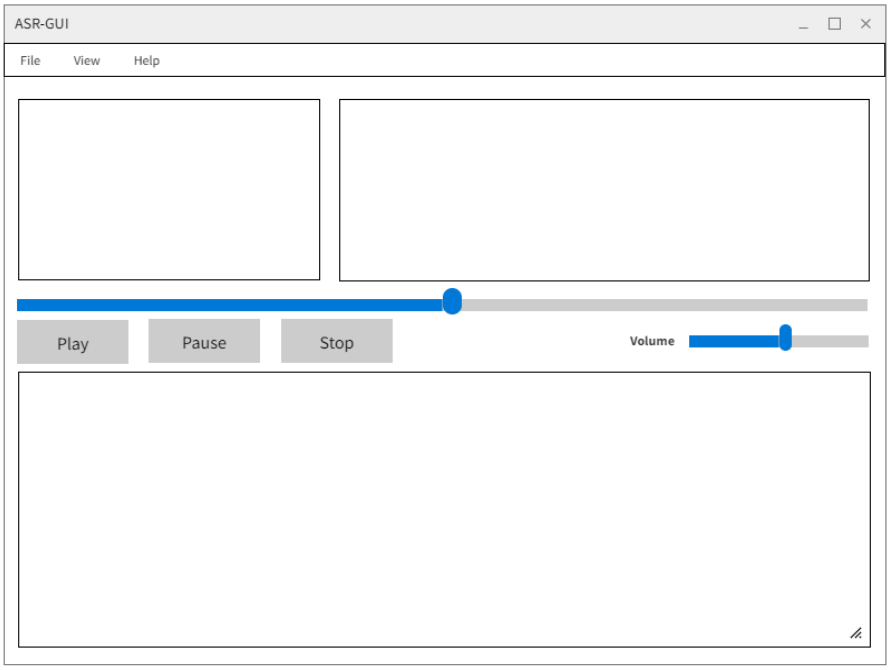
Flowchar Program :



Gambar 1. Flowchart

3.4. Rancangan Antarmuka

Rancangan antarmuka aplikasi merupakan tampilan yang menampilkan seperti apa nantinya aplikasi tersebut, dalam tampilan ini terdapat view untuk menampilkan lokasi file yang di muat, view yang menampilkan spectro analyzer untuk memonitor sinyal audio secara real-time, dan kolom text yang akan secara otomatis mentranskrip audio ketika tombol play ditekan.



Gambar 2. Rancang Aplikasi Audio Forensik

4. Hasil dan Pembahasan

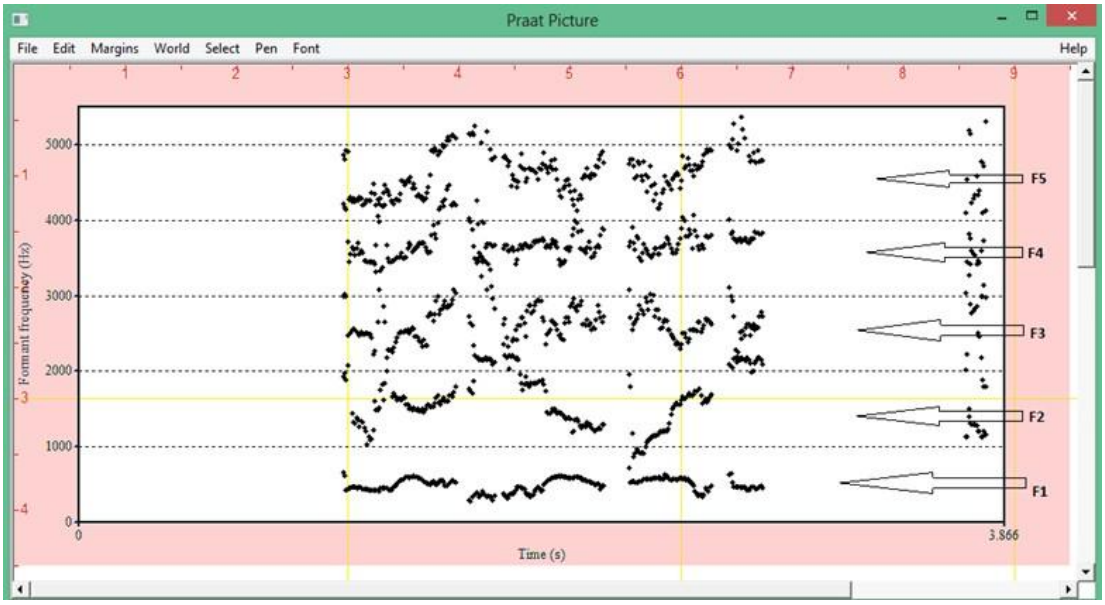
4.1. Hasil Analisis Metode

Dalam perjalanan skenario kasus eksperimental, rekaman suara dikumpulkan sebagai bukti untuk diperiksa. Setelah melalui proses decoding, transkrip lengkap dialog yang terekam dalam rekaman suara tersebut diambil sebagai tersangka. Selain itu, pembahasan juga tersegmentasi berdasarkan hasil analisis transkrip rekaman tersebut, yang dapat dilihat pada Tabel 1:

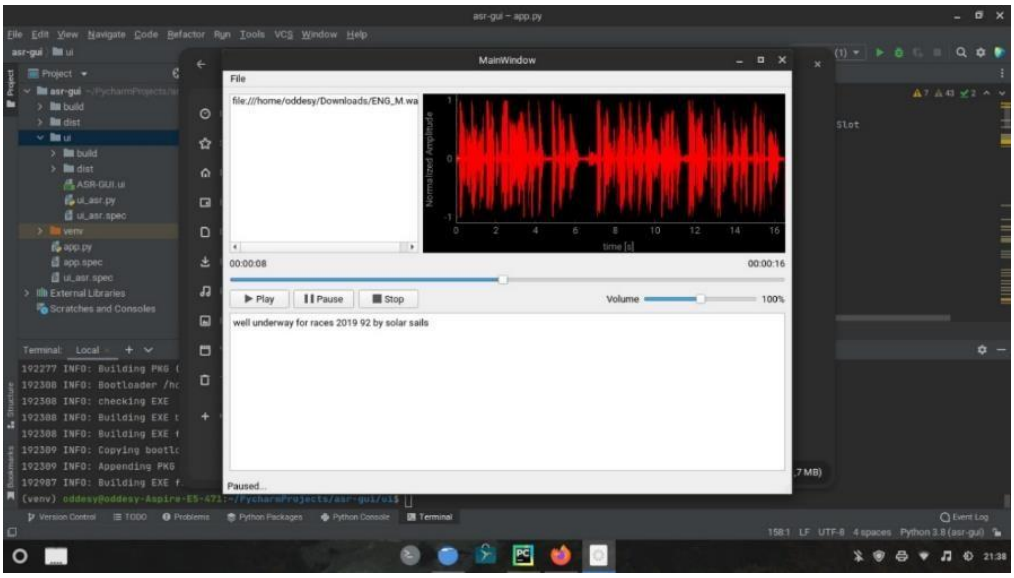
TABEL 1. Rekaman Penentuan Data Analisis

Subyek Rekaman	Kalimat	Jumlah
Rekaman suara *unknown* (A_n)	"Nak Naura, karena Anda dipilih oleh pak Rektor untuk mengikuti lomba ke Singapura. Diharapkan Anda untuk mentransfer uang sebesar 1 juta rupiah hari ini untuk keperluan tiket penerbangan dan dokumentasi.	28 kata / subyek
Rekaman suara *known* (P_n)		

Setelah memilih panjang frasa yang akan dijadikan dasar perbandingan, langkah selanjutnya adalah mencari rekaman suara dari sejumlah contoh yang berbeda. Di area di mana asumsi penerapan metode yang digunakan perlu diperkuat, penelitian ini menggunakan rekaman suara yang diketahui dan rekaman suara yang tidak diketahui. Dengan menggunakan program Matlab 2015b, berikut adalah langkah-langkah yang harus dilakukan untuk menerapkan pendekatan selama tahap pengenalan suara:



Gambar 3. Audio Forensik Hasil Rekaman



Gambar 4.. Spektogram Rekaman

Plot Lowpass Filter dari spektogram yang direkam ditampilkan pada Gambar 4.3. Plot ini menunjukkan perbedaan dalam pengurangan frekuensi ke 0 dari rentang waktu 1 dan seterusnya. Selain itu, evaluasi kualitas ucapan dilakukan berdasarkan prediksi linier antara rekaman suara (A1) dan rekaman suara komparatif (P1), yang keduanya dinilai melalui penggunaan fungsi LPC. Untuk mencapai tingkat akurasi 76,218 persen menggunakan 100000 frame dari total jumlah frame, baik perekaman suara A1 maupun perekaman suara P1 digunakan.

Berdasarkan Tabel 4.3, dapat ditarik kesimpulan bahwa rekaman dapat diperiksa jika karakter memiliki suara dan kosa kata yang baik, jika pengucapan kosa kata cepat tetapi ambigu, dan jika ada noise. Pendekatan konvensional dapat diterapkan pada analisis rekaman suara yang berkualitas tinggi dan mengandung terminologi yang dapat dipahami dengan baik. Perbandingan satu rekaman suara wanita yang tidak diketahui dan satu yang diketahui dilakukan dalam penelitian sebelumnya oleh korban dan pelaku, dan temuannya disajikan dalam bentuk persentase untuk menunjukkan akurasi.

4.2. Hasil Pengujian Sistem

Black-box testing yaitu pengujian yang dilakukan hanya mengamati hasil eksekusi melalui data uji dan memeriksa fungsional dari perangkat lunak (Astuti, 2018).

Metode BlackboxTesting merupakan salah satu metode yang mudah digunakan karena hanya memerlukan batas bawah dan batas atas dari data yang di harapkan,Estimasi banyaknya data uji dapat dihitung melalui banyaknya field data entri yang akan diuji, aturan entri yang harus dipenuhi serta kasus batas atas dan batas bawah yang memenuhi. Dan dengan metode ini dapat diketahui jika fungsionalitas masih dapat menerima masukan data yang tidak diharapkan maka menyebabkan data yang disimpan kurang valid (Cholifah, Yulianingsih and Sagita, 2018).

Tabel 2. Pengujian Sistem

Modul yang diuji	Prosedur Pengujian	Masukan	Keluaran yang diharapkan	Hasil yang didapat	Kesimpulan
New	-Klik File		-	O	n
			Klik New	p e	File

-Klik File	-Klik Open File -Pilih File Audio	Menampilkan jendela baru Menampilkan keterangan file audio yang dibuka	Menampilkan jendela baru Menampilkan keterangan file audio yang dibuka	
Select Models				
Play	-Klik Open File -Pilih File Audio -Klik Tombol Play	File audio yang dibuka dapat diputar	File audio yang dibuka dapat diputar	Fungsi hanya dapat bekerja jika file audio dibaca secara binary/raw.
Pause	-Klik Open File -Pilih File Audio -Klik Tombol Play -Klik Tombol Pause	File audio yang sedang diputar dapat di jeda	File audio yang sedang diputar dapat di jeda	
Stop	-Klik Open File -Pilih File Audio -Klik Tombol Play -Klik Tombol Stop	File audio yang yang sedang diputar dapat diberhentikan	File audio yang yang sedang diputar dapat diberhentikan	
Volume Bar	-Klik Play -Menggeser Volume bar	Volume suara berubah sesuai volume bar	Volume suara berubah sesuai volume bar	
Speech to text	-Klik Open File -Pilih File Audio -Klik Tombol Play	Menampilkan text yang ditrasncribe	Menampilkan text yang ditrasncribe	
Spectogram	-Klik Open File -Pilih File Audio	Menampilkan bentuk visual gelombang suara	Menampilkan bentuk visual gelombang suara	Fungsi hanya dapat bekerja jika file audio dibaca secara binary/raw.

4.3. Pengujian DeepSpeech

Pengujian model DeepSpeech dilakukan menggukan aplikasi dari PicoVoice untuk menguji seberapa ampuh akurasi DeepSpeech dalam mentranscribe masukan audio menjadi bentuk text.

Adapun Dataset yang digunakan dalam pengujian ini yaitu :

1. LibriSpeech
   Merupakan koleksi tulisan kata dengan panjang sekitar 1000 jam dengan frekuensi 16kHz dibaca dalam ucapan bahasa Inggris. Data berasal dari berbagai audiobook dari projek LibriVox yang sudah disegmentasi dan di urutkan dengan hati – hati.
2. TED-LIUM
   Merupakan koleksi tulisan kata untuk melatih speech recognition berbahasa inggris berasal dari TED talks, yang dibuat oleh Laboratoire d‟Informatique de l‟Université du Maine (LIUM). Data berisi ucapan audio dengan panjang sekitar 118 jam dengan frekuensi 16kHz.
3. Common Voice
   Merupakan projek yang dilakukan oleh Mozilla untuk membuat database gratis dari software speech recognition. Projek ini didukung oleh relawan yang merekam sampel kalimat dengan microphone dan rekaman review dari pengguna lain . Kalimat transcribe akan dikumpulkan pada database suara yang tersedia secara publik. Tujuan dari projek ini yaitu untuk menyediakan beraneka ragam sampel suara. Berdasarkan Katharina Borchert, banyak projek yang ada menggunakan dataset dari radio umum atau memiliki dataset yang kurang mempresentasikan perempuan dan orang dengan gaya ucapan/logat.

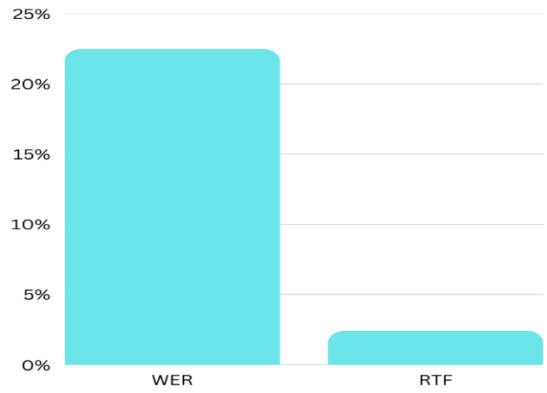
Metric dari hasil yang diperoleh data dari pengujian ini yaitu :

1. Word Error Rate
   Word error rate (WER) adalah rasio jarak edit antara kata yang ada referensi transcript dan kata yang dihasilkan oleh keluaran dari mesin speech to text pada angka dari kata pada referensi transcript.
2. Real Time Factor
   Real-time factor (RTF) adalah rasio dari waktu CPU(pemrosesan) ke panjang dari masukan file ucapan. Mesin speech-to-text dengan RTF rendah lebih efisien secara kompulasi.

   Saat pengujian dilakukan file audio harus dikonversi menjadi 16kHz dengan ekstensi wav karena DeepSpeech hanya mandukung spesifikasi file audio tersebut untuk beroperasi. Anda dapat mengkonfersi file audio menggunakan paket ffmpeg jika sudah tersedia di system anda, atau anda dapat menginstalnya dengan perintah “sudo apt install ffmpeg” pada terminal untuk sistem operasi linux, atau jika menggunakan windows and dapat menjalankan perintah perintah “pip install ffmpeg” pada command prompt.

4.4. Hasil Pengujian Analisis

Berikut merupakan data dari pengujian akurasi speech recognition yang telah dirata rata secara keseluruhan :



Gambar 5. WER Dan RTF Lower Is Better

TABEL 3.Hasil Pengujian Analisis

	LibriSpeech-test Clean	**LibriSpeech-test Other**	**TED-LIUM**	**CommonVoice**	**Rata - Rata**
WER	7.16%	20.26%	18.90%	43.51%	22.46%
RTF	2.46%	2.23%	1.82%	3.14%	2.41%

5. Kesimpulan

Dalam penelitian ini telah terbentuk sebuah tool audio forensik yang menggunakan auto speech recognition yang menggunakan algoritma machine learning secara open source yang dapat membantu investigator dalam mengangani kasus dengan barang bukti berupa file audio. Media player hanya dapat bekerja jika fungsi membaca file audio secara binary serta Speech Recognition hanya dapat beroperasi pada file audio dengan sampel rate 16 kHz dan dengan eksistensi file .wav. Generate nilai hash tidak dapat dilakukan karena sampel rate file harus dikonversi menjadi 16 kHz dengna eksistensi .wav.

DAFTAR PUSTAKA

ADDIN Mendeley Bibliography CSL_BIBLIOGRAPHYAbdel-Hamid, O. et al. (2014) „Convolutional neural networks for speech recognition", IEEE Transactions on Audio, Speech and Language Processing, 22(10), pp. 1533–1545. doi: 10.1109/TASLP.2014.2339736.

Ambewadikar, M. A. and Baheti, M. R. (2020).Review on Speech Recognition System for Disabled People Using Automatic Speech Recognition (ASR)", in Proceedings of the 2020 International Conference on Smart Innovations in Design, Environment, Management, Planning and Computing, ICSIDEMPC 2020. Institute of Electrical and Electronics Engineers Inc., pp. 31–34. doi: 10.1109/ICSIDEMPC49020.2020.9299615.

Azhar, M.N.(2010). Audio Forensic: Theory and Analysis. Pusat Laboratorium Forensik Polri Bidang Fisika Dan Komputer Forensik.

Ibrahim, H. and Varol, A. (2020).A Study on Automatic Speech Recognition Systems", 8th International Symposium on Digital Forensics and Security, ISDFS 2020, (November). doi: 10.1109/ISDFS49300.2020.9116286.

Mohd. Ehmer, K. and Farmeena, K. (2012).A Comparative Study of White Box , Black Box and Grey Box Testing Techniques", International Journal of Advanced Computer Science and Applications, 3(6), pp. 12– 15.

Vimala, C. (2012).A Review on Speech Recognition Challenges and Approaches", 2(1), pp. 1–7.

Y., A. et al. (2017).Survey paper on Different Speech Recognition Algorithm: Challenges and Techniques", International Journal of Computer Applications, 175(1), pp. 31–36. doi: 10.5120/ijca2017915472.

Zabri, M.A.(2006). Introduction to Audio Forensics. Presented at the 8th Mycert Sig. 24 Juny.

Abdel-Hamid, O. et al. (2014) „Convolutional neural networks for speech recognition", IEEE Transactions on Audio, Speech and Language Processing, 22(10), pp. 1533–1545. doi: 10.1109/TASLP.2014.2339736.

Ambewadikar, M. A. and Baheti, M. R. (2020) „Review on Speech Recognition System for Disabled

People Using Automatic Speech Recognition (ASR)‟, in Proceedings of the 2020 International Conference on Smart Innovations in Design, Environment, Management, Planning and Computing, ICSIDEMPC 2020. Institute of Electrical and Electronics Engineers Inc., pp. 31–34. doi: 10.1109/ICSIDEMPC49020.2020.9299615.

Astuti, P. (2018) „PENGGUNAAN METODE BLACK BOX TESTING (BOUNDARY VALUE ANALYSIS) PADA SISTEM AKADEMIK (SMA/SMK)‟, Faktor Exacta, 11(2), p. 186. doi: 10.30998/faktorexacta.v11i2.2510.

Cholifah, W. N., Yulianingsih, Y. and Sagita, S. M. (2018) „Pengujian Black Box Testing pada Aplikasi Action & Strategy Berbasis Android dengan Teknologi Phonegap‟, STRING (Satuan Tulisan Riset dan Inovasi Teknologi), 3(2), p. 206. doi: 10.30998/string.v3i2.3048.

Ibrahim, H. and Varol, A. (2020) „A Study on Automatic Speech Recognition Systems‟, 8th International Symposium on Digital Forensics and Security, ISDFS 2020, (November). doi: 10.1109/ISDFS49300.2020.9116286.

Vimala, C. (2012) „A Review on Speech Recognition Challenges and Approaches‟, 2(1), pp. 1–7.

Y., A. et al. (2017) „Survey paper on Different Speech Recognition Algorithm: Challenges and Techniques‟, International Journal of Computer Applications, 175(1), pp. 31–36. doi: 10.5120/ijca2017915472.