Indonesian Journal of Multidisciplinary on Social and Technology
Homepage: https://journal.ilmudata.co.id/index.php/ijmst
Vol. 5 No. 1 (2026) pp: 515-522
P-ISSN: 2986-6790, e-ISSN: 2986-6782

Pengembangan Model Transkripsi Ucapan Dasar Bahasa Talaud Berbasis Whisper

Claudia Y. Sarempaa, Sherwin R. U Sompie, Wahyuni F. Zalmi
Informatika, Teknik, Universitas Sam Ratulangi Manado
claudiasarempaa026@student.unsrat.ac.id, aldo@unsrat.ac.id, wahyuni.fithratul.zalmi@unsrat.ac.id

Abstrak

Bahasa Talaud merupakan bahasa daerah di kabupaten kepulauan Talaud, Sulawesi Utar, yang penggunaannya cenderung menurun di kalangan generasi muda. Kondisi ini mendorong perlunya pengembangan teknologi untuk mendukung pelestarian dan digitalisasi bahasa daerah. Penelitian ini bertujuan mengembangkan model transkripsi ucapan dasar bahasa Talaud berbasis Whisper melalui teknik fine-tuning agar dapat mengenali ucapan dalam bahasa Talaud dan mengubahnya menjadi teks secara otomatis. Metode penelitian meliputi pengumpulan data melalui perekaman suara penutur asli di Desa Kiama, Kabupaten Kepulauan Talaud, praprose audio dan teks, pembagian dataset, pelatihan model, serta evaluasi kinerja. Dataset yang digunakan terdiri atas 1.010 rekaman audio dengan total durasi 1 jam 11 menit. Data dibagi secara acak menjadi 808 data latih dan 202 data uji dengan rasio 80:20. Model Whisper-small dari OpenAI dilatih melalui proses fine-tuning selama 10 epoch. Evaluasi dilakukan menggunakan Word Error Rate (WER) dan Character Error Rate (CER). Dari 202 data uji, sebanyak 201 rekaman berhasil diproses dan menghasilkan nilai rata-rata WER sebesar 46,09% dan CER sebesar 24,36%. Sebanyak 114 rekaman atau 56,72% berhasil ditranskripsikan tanpa kesalahan. Hasil penelitian menunjukkan bahwa fine-tuning Whisper-small berpotensi digunakan untuk mengembangkan sistem transkripsi ucapan bahasa Talaud, meskipun tingkat kesalahannya masih perlu dikurangi. Peningkatan jumlah dan variasi data serta penyempurnaan proses pelatihan diperlukan untuk meningkatkan akurasi transkripsi. Penelitian ini diharapkan dapat menjadi dasar pengembangan teknologi pengenalan ucapan yang mendukung pelestarian bahasa Talaud.

Kata kunci: Bahasa Talaud, Transkripsi Ucapan, Whisper, Fine-Tuning, Pengenalan Ucapan Otomatis, Pelestarian Bahasa Daerah.

Abstract

Talaud is a regional language spoken in the Talaud Islands Regency, North Sulawesi, Indonesia, whose use has been declining among younger generations. This situation highlights the need for technological innovations to support the preservation and digitalization of regional languages. This study aims to develop a basic Talaud speech transcription model using Whisper through fine-tuning techniques to automatically recognize Talaud speech and convert it into text. The research methodology includes collecting audio recordings from native speakers in Kiama Village, Talaud Islands Regency, preprocessing audio and text data, splitting the dataset, training the model, and evaluating its performance. The dataset consists of 1,010 audio recordings with a total duration of 1 hour and 11 minutes. The data were randomly divided into 808 training samples and 202 testing samples using an 80:20 ratio. OpenAI's Whisper-small model was fine-tuned for 10 epochs. Model performance was evaluated using Word Error Rate (WER) and Character Error Rate (CER). Of the 202 testing samples, 201 recordings were successfully processed, achieving an average WER of 46.09% and CER of 24.36%. Furthermore, 114 recordings (56.72%) were transcribed without errors. The findings indicate that fine-tuning Whisper-small demonstrates potential for developing Talaud speech transcription systems, although transcription errors remain relatively high. Increasing dataset size and diversity, along with optimizing training procedures, is necessary to improve transcription accuracy and support Talaud language preservation.

Keywords: Talaud Language, Speech Transcription, Whisper, Fine-Tuning, Automatic Speech Recognition, Regional Language Preservation.

1. Pendahuluan

Penelitian mengenai pengenalan ucapan otomatis telah berkembang melalui pemanfaatan model pralatih dan teknik adaptasi untuk mengatasi keterbatasan data bahasa sasaran. (Radford dkk., 2022) memperkenalkan Whisper, model pengenalan ucapan multibahasa yang dilatih menggunakan data audio dan transkripsi berskala besar. Selanjutnya, (Liu dan Qu., 2024) mengkaji strategi *fine-tuning* Whisper untuk pengenalan ucapan pada bahasa dengan sumber daya terbatas. (Sehar dkk., 2025), juga mengevaluasi kemampuan Whisper pada bahasa Pashto, Punjabi, dan Urdu serta

mengkaji pengaruh fine-tuning terhadap kinerja transkripsi. Sementara itu, (Hasan dan Roy Dipta., 2025) mengembangkan BanglaTalk untuk mendukung pengenalan ucapan pada dialek regional Bengali dengan memanfaatkan model IndicWav2Vec yang disesuaikan terhadap data dialek. Penelitian-penelitian tersebut menunjukkan pentingnya adaptasi model dan evaluasi pada karakteristik bahasa sasaran. Namun, hasil penelitian pada bahasa lain tidak dapat langsung dijadikan bukti bahwa model yang sama akan memiliki kinerja serupa pada Bahasa Talaud. Oleh karena itu, penelitian ini berfokus pada pengembangan model transkripsi ucapan dasar Bahasa Talaud melalui fine-tuning Whisper-small menggunakan rekaman penutur asli serta evaluasi berdasarkan Word Error Rate (WER) dan Character Error Rate (CER).

2. Metode Penelitian

Penelitian ini menggunakan pendekatan eksperimental dengan mengembangkan model transkripsi ucapan dasar Bahasa Talaud melalui teknik *fine-tuning* pada model Whisper-small dari OpenAI. Tahapan penelitian meliputi studi literatur, pengumpulan data, praproses data, pembagian dataset, pelatihan model, pengujian dan evaluasi model. Evaluasi dilakukan menggunakan *Word Error Rate* (WER), *Character Error Rate* (CER), skor akurasi, *confusion matrix*, dan *classification report*.

2.1. Metode Pengumpulan Data

Metode pengumpulan data suara dilakukan langsung oleh penutur asli bahasa talaud yang tinggal di kabupaten kepulauan talaud, khususnya di desa kiama. Dengan mengunjungi rumah ketua adat atau narasumber setempat untuk melakukan sesi wawancara, mengumpulkan dataset berupa rekaman suara sebagai dataset utama. Setelah melakukan beberapa proses metode diatas selanjutnya, dalam pengujian sistem yang dilakukan di lingkungan masyarakat untuk akses internet dengan beberapa device seperti handphone sebagai beckup audio dan mic clipon sebagai penyaring suara yang dipakai juga sangat cukup membantu serta mengsupport, untuk mempermudah proses wawancara dan perekaman audio untuk melatih model whisper ke sistem transkripsi bahasa talaud.

2.2. Praproses Data

Tahap praproses data merupakan konversi format audio pada awal penelitian. Semua data audio bahasa talaud diubah menjadi format wav dengan frekuensi sampling 16 khz, satu saluran mono, dan kedalaman bit 16-bit. Proses konversi ini dilakukan karena model whisper dari OpenAi dilatih menggunakan data audio mono 16 khz, sehingga masukkan yang tidak sesuai dengan spesifikasi ini dapat mengurangi akurasi transkripsi atau menggunakan proses pelatihan. Proses konversi dilakukan dengan memanfaatkan pustaka librosa dan soundfile pada python.

2.3. Pembagian Dataset

Dataset yang telah melalui tahap praproses selanjutnya dibagi menjadi dua bagian, yaitu data latih dan data uji. Pembagian data dilakukan dengan proporsi 80:20, dengan memastikan tidak ada tumpang tindih penutur antara data latih dan data uji agar hasil evaluasi model dapat menghasilkan kemampuan generalisasi model secara objektif.

```
dataset = dataset.train_test_split(
    test_size=0.2,
    seed=42
)
```

Gambar 1. Implementasi Pembagian Dataset Kode

2.4. Pelatihan Model

Model Whisper-small yang telah dirancang selanjutnya dilatih (*fine-tuned*) menggunakan 808 data latih Bahasa Talaud yang telah dipersiapkan pada tahap sebelumnya. Pelatihan dilakukan menggunakan kelas Seq2SeqTrainer dari pustaka Hugging Face Transformers dengan konfigurasi pada Tabel 4.7, yaitu *learning rate* 1e-5, *batch size* 4, dan 10 epoch dengan *optimizer* AdamW. Proses pelatihan dilakukan menggunakan laptop HP tanpa GPU khusus, sehingga

DOI: https://doi.org/10.69693/ijmst.v5i1.14515
Lisensi: Creative Commons Attribution 4.0 International (CC BY 4.0)

pelatihan berlangsung dalam durasi yang relatif lama. Data uji tidak digunakan selama pelatihan. Setelah pelatihan selesai, kinerja model dievaluasi menggunakan data uji.

2.5. Pengujian dan Evaluasi Model

Tahapan pengujian dan evaluasi model Whisper-small hasil *fine-tuning*, yang meliputi tujuan pengujian, skenario pengujian, metode pengujian, metrik evaluasi, dan prosedur pengujian.

3. Hasil dan Pembahasan

Pengumpulan data pada penelitian ini dilakukan di Desa Kiama, Kabupaten Kepulauan Talaud, Sulawesi Utara, selama 2 bulan, mulai dari bulan Januari sampai Maret 2026. Data dikumpulkan dari 5 penutur asli Bahasa Talaud yang terdiri dari 4 orang laki-laki dan 1 orang perempuan, berusia 25-60 tahun. Proses perekaman dilakukan menggunakan *smartphone* dengan bantuan aplikasi ASR untuk menyimpan data audio dalam format Wav, *sample rate* 16 kHz, dalam kondisi lingkungan yang relatif tenang untuk meminimalkan derau latar belakang, dengan format audio wav *sample rate* 16 kHz, dalam kondisi lingkungan yang relatif tenang untuk meminimalkan derau latar belakang. Berhasil dikumpulkan sebanyak 1.010 berkas audio dengan total durasi 1 jam 11 menit. Rincian data yang terkumpul disajikan pada tabel 1.

Tabel 1. Ringkas data audio

Aspek	Jumlah
Jumlah penutur	5 orang
Jumlah file audio	1.010 audio
Total durasi audio	1 jam 11 menit
Format audio	WAV
Rata-rata durasi per audio	5 det

Setiap berkas audio yang terkumpul dipasangkan dengan transkripsi teks yang sesuai dengan ucapan pada rekaman tersebut. Pasangan audio dan transkripsi ini menjadi metadata utama yang digunakan pada tahap pelatihan dan pengujian model. Contoh pasangannya disajikan pada tabel 2.

Tabel 2. Audio dan transkripsi

Nama file	Transkripsi teks
Audio0001.wav	Isai
Audio0002.wav	Suapa
Audio0003.wav	Angngera
Audio0004.wav	Tawe
Audio0005.wav	Ore

Pada Tabel 2, terlihat bahwa sebagian besar data berupa satu kata dalam Bahasa Talaud, seperti kata tanya (*isai*, *suapa*) dan kata ganti atau keterangan (*angngera*, *tawe*, *ore*). Data yang sebagian besar berupa kata pendek ini berpengaruh terhadap hasil evaluasi model pada Sub-bab 4.6, karena nilai *Word Error Rate* (WER) pada satu kata hanya bisa 0% (benar seluruhnya) atau 100% (salah seluruhnya), sehingga metrik *Character Error Rate* (CER) menjadi lebih informatif untuk melihat seberapa dekat hasil transkripsi model dengan jawaban yang benar.

3.1. Hasil Praproses Data

Setelah data audio dan teks terkumpul, dilakukan tahapan praproses untuk mempersiapkan data agar siap digunakan dalam proses *fine-tuning* model. Praproses audio dilakukan penghapusan (*noise*). Duplikat audio (*deduplication*), dan normal volume sedangkan praproses teks meliputi normalisasi ejaan dan meliputi normalisasi ejaan dan penghapusan tanda baca dari total 1.058 data awal, setelah melalui tahapan praproses diperoleh 1.010 data yang layak digunakan, sementara 48 data audio dibuang karena terjadinya *duplication* audio serta buruknya *quality* audio, tabel 3 menunjukkan perbandingan kondisi data sebelum dan sesudah praproses.

DOI: https://doi.org/10.69693/ijmst.v5i1.14515
Lisensi: Creative Commons Attribution 4.0 International (CC BY 4.0)

Tabel 3. Perbandingan data sebelum dan sesudah praproses

Aspek	Sebelum	Sesudah
Jumlah data	1.058	1.010
Sample rate	Suapa	16 kHz
Format audio	Ore	WAV

3.2. Hasil Pembagian Dataset

Data yang telah melalui tahapan praproses sebanyak 1.010 file audio selanjutnya dibagi menjadi 2, data latih dan data uji dengan proporsi 80% : 20%. Pembagian dilakukan secara acak menggunakan *library Hugging Face* datasets dengan *random seed* tetap (seed=42) untuk memastikan hasil pembagian yang konsisten dan dapat direplikasi. Berikut hasil pembagian pada tabel 4.

Tabel 4. Hasil Pembagian Dataset

Jenis data	Jumlah file	Persentase
Data latih training	808	80%
Sample rate	202	20%
Format audio	1.010	100%

3.3. Implementasi Model

Hasil implementasi model whisper untuk transkripsi ucapan dasar bahasa talaud, mencakup lingkungan implemntasi yang digunakan serta konfigurasi arsitektur model yang diterapkan dalam proses fine-tuning.

3.3.1. Lingkungan implementasi

Proses pelatihan (*fine-tuning*) model pada awalnya dilakukan menggunakan laptop HP, sedangkan proses evaluasi, pengujian, dan penyusunan laporan selanjutnya dilakukan menggunakan Laptop Acer Aspire 3 A314-32-C4H6 sebagaimana dijelaskan pada Sub-bab 3.3.1. Adapun perangkat lunak pendukung yang digunakan meliputi bahasa pemrograman Python versi 3.11.0 (64-bit), dengan pustaka (*library*) utama sebagaimana disajikan pada tabel 5.

Tabel 5. Perangkat lunak dan Pustaka yang digunakan

Perangkat lunak	Versi
Python	3.11.0
PyTorch	2.10.0
Transformers HuggingFace	5.13.1
Datasets HuggingFace	5.0.1
Jiwer	4.0.0
Librosa	0.11.0
Soundfile	0.14.0

3.4. Hasil Fine-tuning Model

Hasil fine-tuning model whisper menggunakan dataset bahasa talaud yang telah dipersiapkan, mencakup konfigurasi model, parameter pelatihan yang digunakan, serta hasil yang diperoleh selama proses pelatihan berlangsung.

3.4.1. Parameter Pelatihan

Proses fine-tuning model whisper-small dilakukan menggunakan konfigurasi parameter pelatihan sebagimana disajikan pada tabel 6.

DOI: https://doi.org/10.69693/ijmst.v5i1.14515
Lisensi: Creative Commons Attribution 4.0 International (CC BY 4.0)

Tabel 6. Parameter fine-tuning

Parameter	Nilai
Learning rate	Le-5
Batch size	4
Epoch	10
Optimizer	AdamW

3.4.2. Hasil Training

Proses pelatihan (*fine-tuning*) dilakukan terhadap model Whisper-small menggunakan 808 data latih selama 10 epoch dengan konfigurasi pada Tabel 4.6. Pelatihan dijalankan menggunakan kelas Seq2SeqTrainer dari pustaka Hugging Face Transformers. Pada penelitian ini, kinerja model tidak dianalisis melalui kurva *loss* selama pelatihan, melainkan dievaluasi setelah pelatihan selesai menggunakan data uji yang tidak digunakan dalam pelatihan, dengan metrik WER dan CER.

3.5. Hasil Pengujian Model

Pengujian model dilakukan terhadap 202 data uji, dan 201 di antaranya berhasil diproses (satu berkas audio tidak dapat dibaca). Setiap prediksi transkripsi dibandingkan dengan transkripsi referensi menggunakan metrik WER dan CER. Hasil pengujian menunjukkan rata-rata WER sebesar 46,09% dan rata-rata CER sebesar 24,36%. Sebanyak 114 audio (56,7%) berhasil ditranskripsikan tanpa kesalahan (WER = 0%).

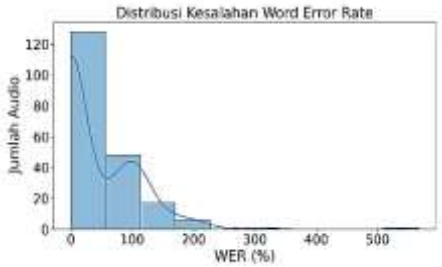


Gambar 2. Perbandingan rata-rata WER dan CER

Gambar 2. Menunjukkan bahwa nilai rata-rata WER lebih tinggi dibandingkan CER. Hal ini terjadi karena WER mengukur kesalahan pada tingkat kata, sehingga satu kesalahan huruf pada sebuah kata tetap dihitung sebagai satu kesalahan kata penuh, sedangkan CER mengukur kesalahan pada tingkat karakter sehingga lebih sensitif terhadap kesalahan kecil.

3.5.1. Evaluasi WER

Evaluasi *Word Error Rate* (WER) dilakukan dengan membandingkan hasil transkripsi model terhadap transkrip referensi (*ground truth*) pada seluruh data uji. Dari 201 data uji yang berhasil diproses, sebanyak 114 audio (56,7%) berhasil ditranskripsikan dengan sempurna (WER = 0%), dan 115 audio (57,2%) memiliki WER di bawah 20%. Nilai WER tertinggi yang tercatat adalah 566,7%. Distribusi nilai WER disajikan pada Gambar 3.



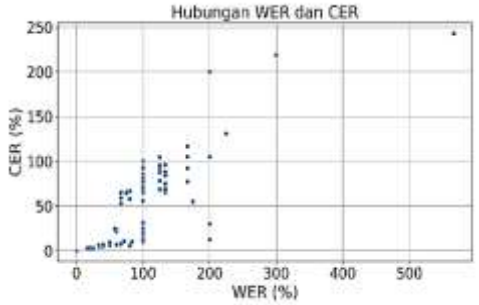
Gambar 3. Distribusi Nilai WER pada Data Uji

DOI: https://doi.org/10.69693/ijmst.v5i1.14515
Lisensi: Creative Commons Attribution 4.0 International (CC BY 4.0)

Gambar 3 menunjukkan bahwa distribusi nilai WER didominasi oleh rentang 0-20%, dengan frekuensi tertinggi berada pada kelompok WER mendekati 0%. Hal ini menunjukkan model cukup mampu mengenali sebagian besar pola ucapan Bahasa Talaud pada data uji, meskipun terdapat sejumlah kecil kasus dengan nilai WER tinggi yang mengindikasikan kegagalan transkripsi total pada beberapa audio.

3.5.2. Evaluasi CER

Evaluasi *Character Error Rate* (CER) dilakukan untuk melihat tingkat kesalahan model pada satuan karakter. Rata-rata CER tercatat sebesar sebesar 24,36%, lebih rendah dibandingkan rata-rata WER sebesar 46,09%, karena sebagian besar data berupa satu kata sehingga satu kesalahan huruf sudah dihitung sebagai kesalahan kata penuh dalam WER. Berikut disajikan pada gambar 4.



Gambar 4. Hubungan WER dan CER

3.5.3. Akurasi Skor

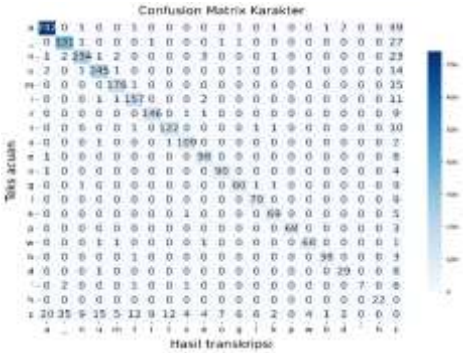
Akurasi skor dihitung dari hasil transkripsi yang sesuai dengan teks referensi, dari 201 data uji yang berhasil diproses, sebanyak 114 audio (56,72%) menghasilkan transkripsi tanpa kesalahan. Akurasi pada tingkat kata dan karakter masing-masing sebesar 53,91% dan 75,64%, berdasarkan nilai WER sebesar 46,09% dan CER sebesar 24,36%. Perbedaan ini menunjukkan bahwa model lebih baik dalam mengenali karakter dibandingkan menyusun kata secara tepat. Hasil tersebut cukup menjanjikan dengan durasi data latih 1 jam 11 menit, tetapi masih perlu ditingkatkan melalui penambahan jumlah dan variasi data.

Tabel 7. Akurasi Skor Hasil Pengujian

Level	Perhitungan	Akurasi
Kalimat exact match	114 ÷ 201	56,72%
Kata	100%-WER (46,09%)	53,91%
Karakter	100%-CER (24,36%)	75,64%

3.5.4. Confusion Matrix

Confusion matrix tidak hanya menunjukkan jumlah prediksi yang benar dan salah, tetapi juga memberikan gambaran mengenai karakter yang paling sering mengalami kesalahan, pada proses transkripsi bahasa talaud.



Gambar 5. Confusion Matrix Karakter

DOI: https://doi.org/10.69693/ijmst.v5i1.14515
Lisensi: Creative Commons Attribution 4.0 International (CC BY 4.0)

Gambar 5. Menunjukkan *confusion matrix* pada level karakter dengan baris sebagai teks acuan dan kolom sebagai hasil transkripsi. Nilai diagonal yang dominan menunjukkan bahwa sebagian besar karakter berhasil ditranskripsi dengan benar, misalnya karakter a sebanyak 742 kali dan spasi sebanyak 331 kali. Simbol ε menunjukkan karakter yang hilang (*deletion*) pada kolom dan karakter tambahan (*insertion*) pada baris. Rincian kesalahan terbanyak disajikan pada Tabel 8.

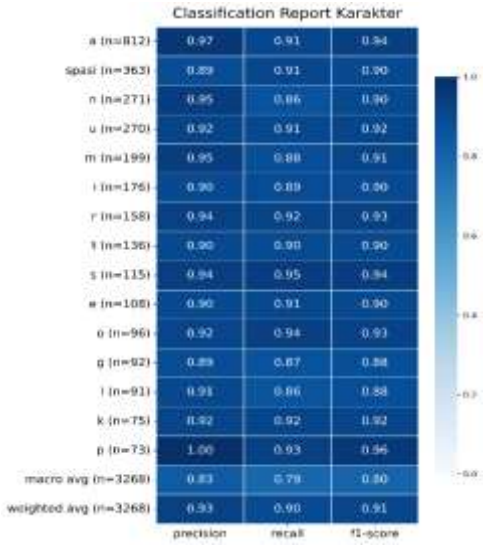
Tabel 8. Analisis 10 Kesalahan Karakter pada Hasil Transkripsi

Karakter Acuan	Hasil Transkripsi	Jumlah Kesalahan
a	ε	49
ε	_	35
_	ε	27
n	ε	23
ε	a	20
m	ε	15
ε	u	15
u	ε	14
ε	t	12
ε	i	12

Kesalahan karakter yang paling banyak terjadi adalah pasangan a → ε sebanyak 49 kali, selanjutnya, pasangan ε → _ sebanyak 35 kali dan _ → ε sebanyak 27 kali. Kesalahan lainnya adalah n → ε sebanyak 23 kali dan ε → a sebanyak 20 kali. Selain itu, terdapat kesalahan pada pasangan m → ε dan ε → u yang masing-masing terjadi sebanyak 15 kali, serta u → ε sebanyak 14 kali. Pasangan ε → t dan ε → i masing-masing ditemukan sebanyak 12 kali. Hasil tersebut menunjukkan bahwa kesalahan transkripsi tidak hanya berupa pergantian karakter, tetapi juga terdapat karakter yang tidak dikenali atau karakter tambahan pada hasil transkripsi. Nilai pada diagonal utama confusion matrix menunjukkan bahwa model mampu mengenali sebagian besar karakter dengan benar. Namun, masih terdapat beberapa kesalahan karakter yang ditunjukkan oleh nilai di luar diagonal. Hasil *confusion matrix* ini menjadi analisis pendukung terhadap nilai CER, sehingga dapat diketahui karakter-karakter yang paling sering mengalami kesalahan pada proses transkripsi Bahasa Talaud.

3.5.4. Classification Report

Classification report merupakan laporan evaluasi yang digunakan untuk mengukur kinerja model dalam mengenali setiap karakter berdasarkan nilai *precision*, *recall*, dan *F1-score*. Pada penelitian ini, *classification report* digunakan untuk mengetahui kemampuan model Whisper dalam mengenali karakter pada hasil transkripsi ucapan Bahasa Talaud dibandingkan dengan teks referensi. Hasil evaluasi ditunjukkan pada Gambar 6.



Gambar 6. Classification Report Karakter

DOI: https://doi.org/10.69693/ijmst.v5i1.14515
Lisensi: Creative Commons Attribution 4.0 International (CC BY 4.0)

4. Kesimpulan

Disimpulkan bahwa model transkripsi ucapan dasar Bahasa Talaud berhasil dikembangkan melalui teknik fine-tuning pada model Whisper-small menggunakan dataset rekaman audio penutur asli yang telah melalui tahap pengumpulan data, praproses, serta pembagian data latih dan data uji. Hasil penelitian menunjukkan bahwa pendekatan transfer learning melalui Whisper dapat diadaptasi untuk pengembangan sistem transkripsi bahasa daerah dengan sumber daya terbatas (low-resource language), seperti Bahasa Talaud. Model yang dikembangkan menghasilkan rata-rata Word Error Rate (WER) sebesar 46,09% dan Character Error Rate (CER) sebesar 24,36%, dengan 114 rekaman audio atau 56,72% berhasil ditranskripsikan tanpa kesalahan. Hasil tersebut menunjukkan bahwa model telah mampu mentranskripsikan sebagian ucapan dasar Bahasa Talaud, tetapi tingkat kesalahan yang dihasilkan masih menunjukkan perlunya peningkatan kinerja, terutama dalam menghadapi variasi kosakata dan pengucapan. Dengan demikian, penelitian ini dapat menjadi dasar awal bagi pengembangan teknologi transkripsi otomatis untuk mendukung dokumentasi, digitalisasi, dan pelestarian Bahasa Talaud.

Reference

1. Fathony, I. A. N., Mareta, A., Adiana, B. E., Wardhani, O., & Halim, D. A. (2025). Optimisasi Whisper Speech-to-Text Bahasa Indonesia dengan Hybrid Cloud dan Multi Engine. *Elkom: Jurnal Elektronika dan Komputer*, *18*(1), 60-72.
2. Pertiwi, A. (2024). MEMBUKA DUNIA BAGI PENYANDANG TUNARUNGU: REVOLUSI SISTEM TRANSKRIPSI SUARA. *Martabe: Jurnal Pengabdian Masyarakat*, *7*(1).
3. Suhendar, H., Slamet, C., & Syaripudin, U. (2025). Analisis Sentimen Hasil Transkripsi Audio Berbahasa Indonesia Menggunakan T5 (Text-to-Text Transfer Transformer). *SMATIKA JURNAL*, *15*(01), 115-125.
4. Novela, M., & Basaruddin, T. (2021). Dataset suara dan teks berbahasa Indonesia pada rekaman podcast dan talk show. *Jurnal Fasilkom*, *11*(2), 61-66.
5. NOERCHOLIS, A., DWIANDINI, T., & MUKTI, F. S. (2024). Optimasi Teknologi WAV2Vec 2.0 menggunakan Spectral Masking untuk meningkatkan Kualitas Transkripsi Teks Video bagi Tuna Rungu. *ELKOMIKA: Jurnal Teknik Energi Elektrik, Teknik Telekomunikasi, & Teknik Elektronika*, *12*(4), 877.
6. Rifaldi, R. (2025). *Pengembangan Fitur Speech Recognition Pada Aplikasi Notulensi Menggunakan Whisper AI* (Doctoral dissertation, Universitas Islam Indonesia).
7. Syaifullah, M. I. (2024). *Penerapan Teknologi Automatic Speech Recognition (Asr) Sebagai Alat Bantu Untuk Mendeteksi Kesalahan Dalam Membaca Al-Qur'an Menggunakan Model Wav2vec2. 0 Berbasis Mobile* (Doctoral dissertation, Universitas Islam Sultan Agung Semarang).
8. Quinton, F. Y., Pranata, Y. W., Palenjawi, P., Febriansyah, R., Putry, S. A., Pratiwi, L., & Nainggolan, E. R. (2026). Perancangan Aplikasi Percakapan AI Berbasis Web Menggunakan Speech-to-Text, Text-to-Speech, dan Voicevox untuk Melatih Percakapan dalam Bahasa Jepang. *Jurnal Ilmu Komputer*, *1*(1), 1-9.
9. Ferdiansyah, D., & Aditya, C. S. K. (2024). Implementasi automatic speech recognition bacaan Al-Qur'an menggunakan metode Wav2Vec 2.0 dan OpenAI-Whisper. *Jurnal Teknik Elektro Dan Komputer TRIAC*, *11*(1), 11-16.
10. Putri, A. I. (2025). Implementasi Aplikasi Transkripsi Instan dalam Pembelajaran Inklusi di SMK Marsudirini Marganingsih Surakarta. *JIMU: Jurnal Ilmiah Multidisipliner*, *3*(03), 1798-1803.
11. Putri, A. A., Anwar, M., & Anggrellanggi, A. (2024). Pengaruh Transkripsi Instan Terhadap Pemahaman Mahasiswa Tunarungu Universitas Sebelas Maret. *Jurnal Inovasi Pendidikan*, *2*(3), 175-183.
12. Wilyanti, L. S., Larlen, L., & Wulandari, S. (2022). Transkripsi Sastra lisan melayu Jambi sebagai alternatif bahan ajar sastra di perguruan tinggi. *Jurnal Ilmiah Universitas Batanghari Jambi*, *22*(1), 247-252.
13. SIEMWIJAYA, J. A. ANALISIS KESALAHAN PELAFALAN NADA 2 DAN NADA 3 SERTA DAMPAKNYA OLEH MAHASISWA PRODI PENDIDIKAN BAHASA MANDARIN DI UNIVERSITAS MA CHUNG.
14. Pratama, R. S. A., & Amrullah, A. (2024). Analysis of whisper automatic speech recognition performance on low resource language. *Jurnal Pilar Nusa Mandiri*, *20*(1), 1-8.
15. Ariyati, D. A., & Rostina, R. (2024). Efektivitas Game English Whisper dalam Meningkatkan Keterampilan Berbahasa Inggris pada Siswa SMP 19 Makassar. *Jurnal Pengabdian Masyarakat Nusantara*, *6*(4), 222-236.
16. Rumandjo, N., Dyantari, P., & Putra, G. N. (2023). DEVELOPMENT OF EARLY CHILDREN'S LANGUAGE ABILITY THROUGH CHAIN WHISPER PLAY METHOD AT KASIH BANGSA PALU KINDERGARTEN. *JARITA: Journal of Children's Education and Learning Research*, *1*(2), 67-72.
17. Ahsanurrijal, A., Akbar, F., Ahmad, A., Abdullah, A. R., & Ramli, A. (2025). Educational Chain Whisper Game for Improving Listening and Memory Based on Local Wisdom. *Jurnal Sipakatau: Inovasi Pengabdian Masyarakat*, *2*(6), 268-274.
18. Putra, R. H., Alam, R., Indriasari, M., & Kom, M. (2025). PERANCANGAN WEBSITE NOTULENSI RAPAT OTOMATIS BERBASIS KECERDASAN BUATAN DENGAN FITUR TINDAK LANJUT. *Technopex 2025*, *9*(1), 495-500.
19. Anggara, F. (2021). *Implementasi Speech To Text Pada Pencarian Buku Berbasis Android* (Doctoral dissertation, Universitas Islam Riau).
20. Haq, M. D., & Prasetiyo, N. T. (2025). Deep learning sebagai pendekatan transformasional dalam pendidikan: Sebuah tinjauan literatur. *Jurnal Studi Guru Dan Pembelajaran*, *8*(3), 1826-1842.
21. Nzenwata, U., & Ogbuigwe, D. U. (2024). Automatic speech recognition for the Ika language. arXiv preprint arXiv:2410.00940.
22. Pranida, S. Z., Airlangga, M. C., Genadi, R. A., & Shehata, S. (2025). ASR under noise: Exploring robustness for Sundanese and Javanese. arXiv preprint arXiv:2509.25878.
23. Pratama, R. S. A., & Amrullah, A. (2024). Analysis of Whisper automatic speech recognition performance on low resource language. Jurnal Pilar Nusa Mandiri, 20(1), 1–8

DOI: https://doi.org/10.69693/ijmst.v5i1.14515
Lisensi: Creative Commons Attribution 4.0 International (CC BY 4.0)