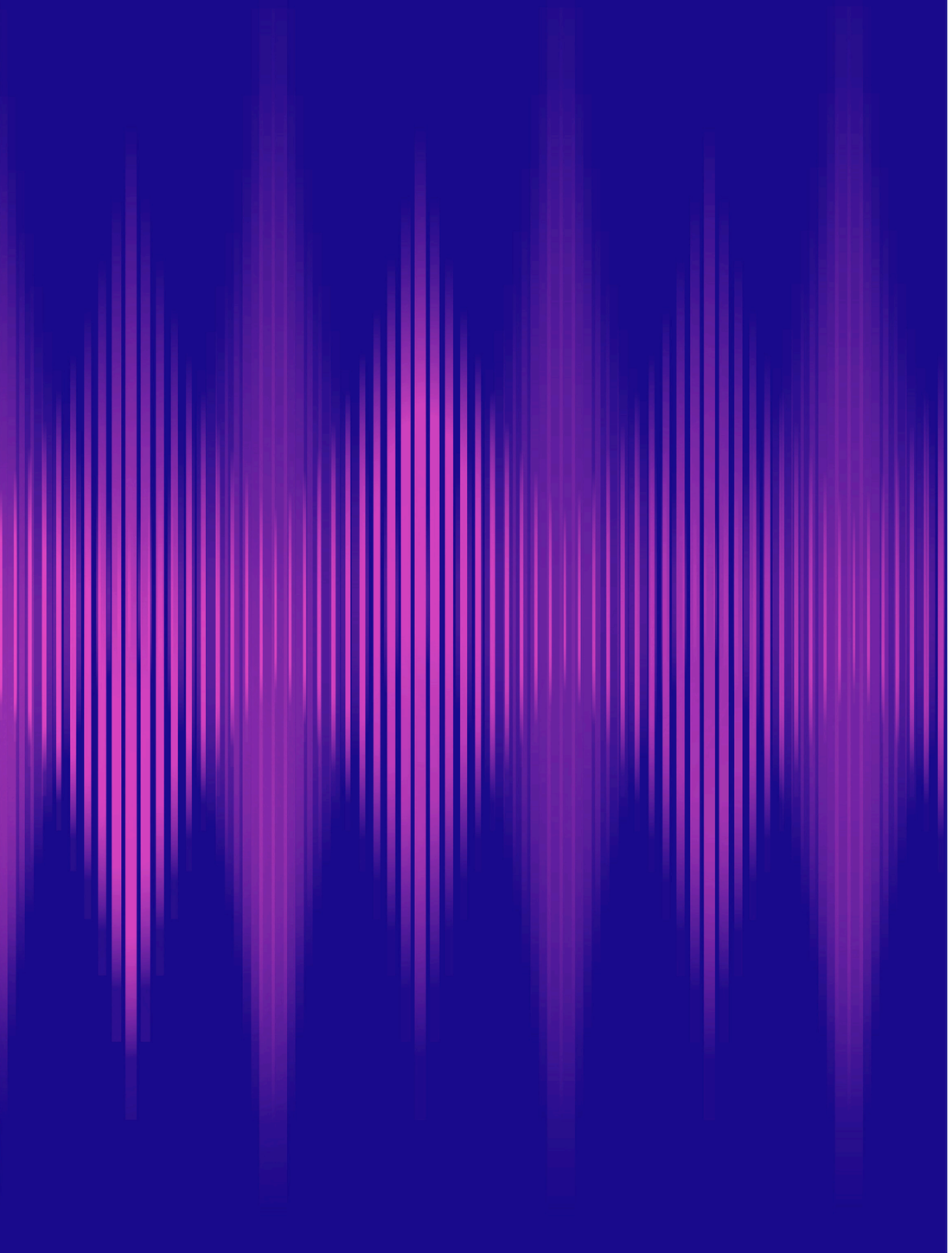
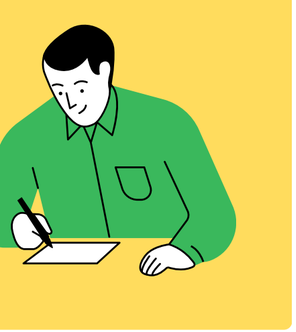


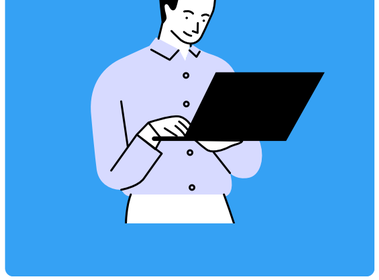
Voice Recognition using Wav2Vec and Whisper Model

NLP PROJECT BY TEAM A

Indonesia AI
AI for Everyone, AI for Indonesia

Spärck





Meet the Team

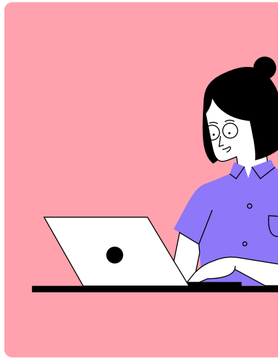
Alfandy Surya

Muhammad Habibullah

Efrad Galio

Anton Pranowo Medianto

Alfian Ali Murtadlo



Spärck

Background &
Problem Statement

Tren e-banking pada era digital saat ini sudah semakin meluas. Tentunya perusahaan perlu memfasilitasi transaksi keuangan nasabah dengan mudah. Namun, interaksi pengguna dengan platform e-banking kadang membutuhkan input teks, menyulitkan beberapa pengguna. Selain itu, keamanan transaksi juga perlu diperkuat.

Dalam konteks ini, penggunaan Automatic Speech Recognition (ASR) dalam e-banking menjadi penting. ASR memungkinkan interaksi melalui suara, mengurangi keterbatasan input teks dan meningkatkan keamanan dengan identifikasi suara pengguna. Sehingga, ASR dapat memperbaiki pengalaman pengguna dan meningkatkan keamanan transaksi e-banking.

Spärck

Objectives & Scope

Objectives:

Melakukan eksperimen dengan membandingkan model-model voice recognition pretrained seperti wave2vec dan whisper.

Scope:

- Menggunakan data minds14
- Menggunakan model wave2vec dan whisper dan hasil fine tune dengan data minds14
- Uji performansi menggunakan rata-rata WER, CER, BLEU, dan waktu inferensi.

Data Information

Dataset used: minds14

Data Fields:

- **path:** path of the audio file
- **audio:** audio object including loaded audio array, sampling rate and path ot audio
- **transcription:** transcription of the audio file
- **english_transcription:** english transcription of the audio file
- **intent_class:** class ID of intent
- **lang_id**: id of language

Test Data:

100 Random data each language

Language:

- English - US
- English - AU

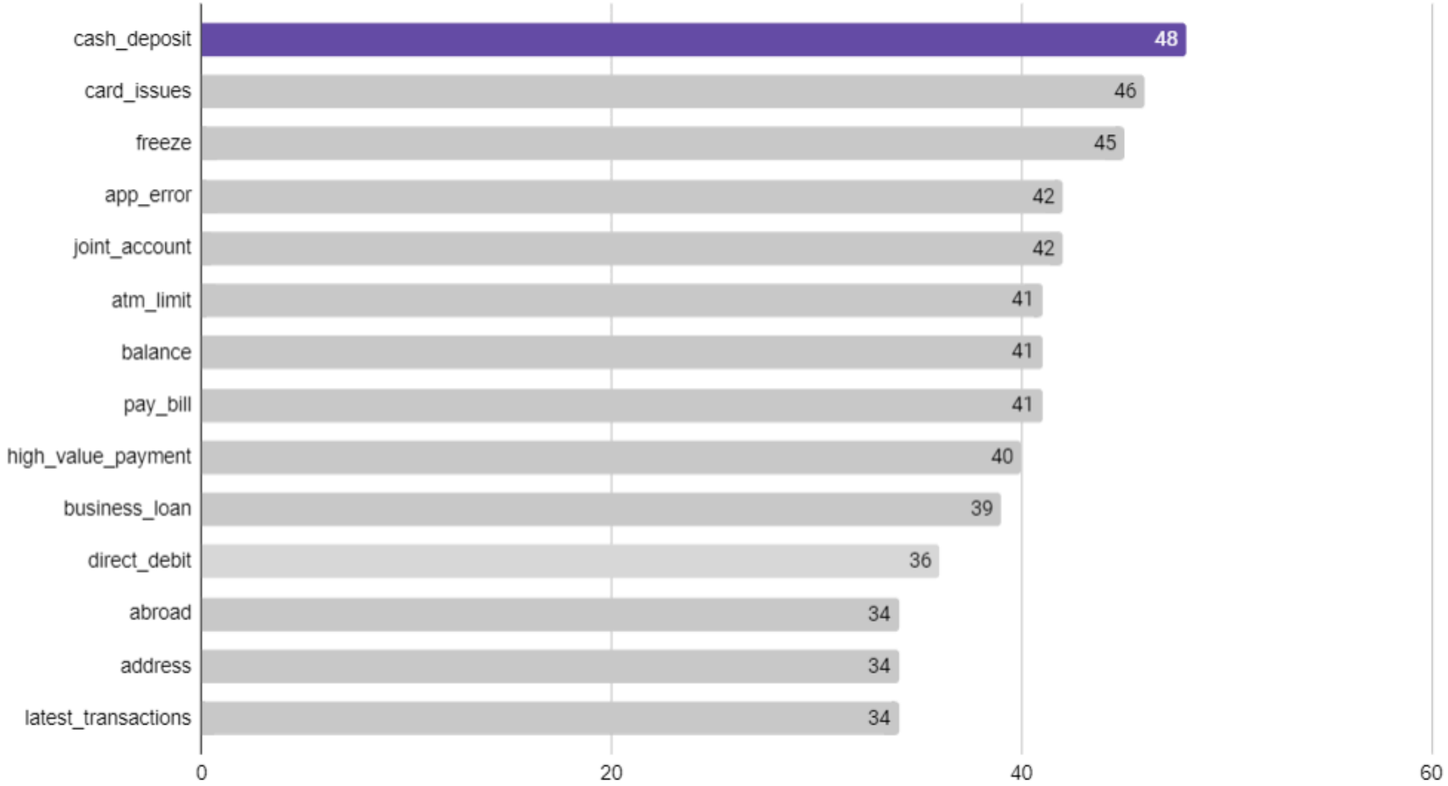
Text Preprocessing for EDA

- Remove parentheses
- Remove punctuation
- Remove extra space
- Case folding (lowercase)
- Remove stopwords

Spärck

Exploratory Data Analysis – 1

Top Intents:



Dua intensi tertinggi pada data yang digunakan adalah terkait **cash_deposit** dan **card_issues.** Artinya kebanyakan isu dari nasabah adalah terkait bagaimana cara melakukan deposit dan permasalahan terkait kartu.

Selanjutnya akan ditunjukkan unigram dan tri-gram dari top 3 intensi yang ada.

Exploratory Data Analysis - 2

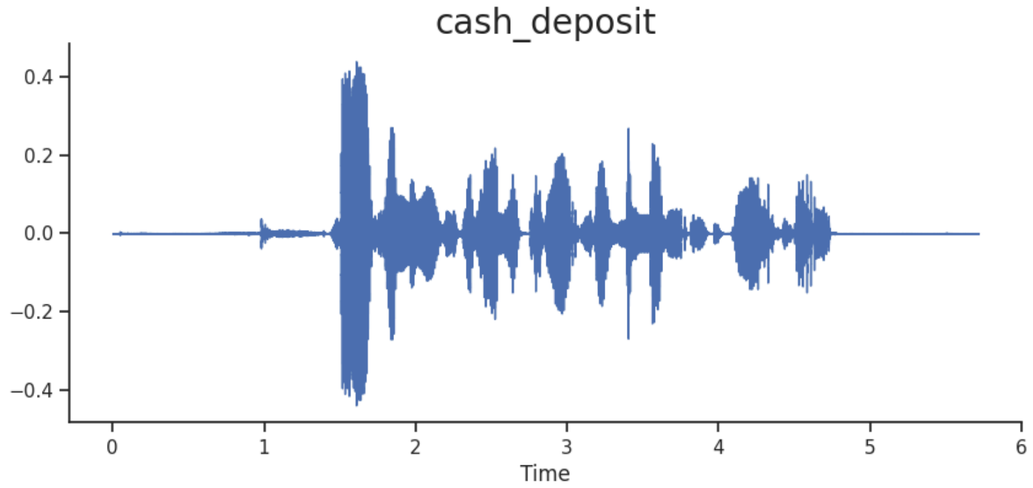
unigram cash deposit

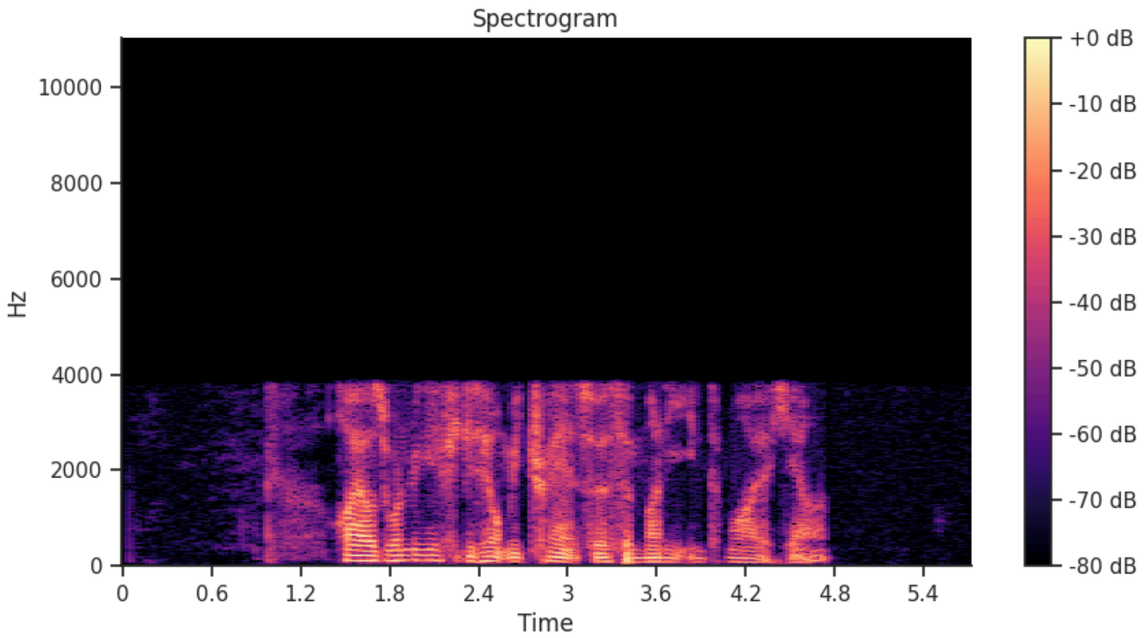
Unigram	Frekuensi
money	44
account	42
deposit	29
transfer	10

tri-gram cash deposit

Trigram	Frekuensi
deposit money account	17
transfer money account	8
money account hi	6
money account deposit	5

sample waveplot cash deposit





Exploratory Data Analysis - 3

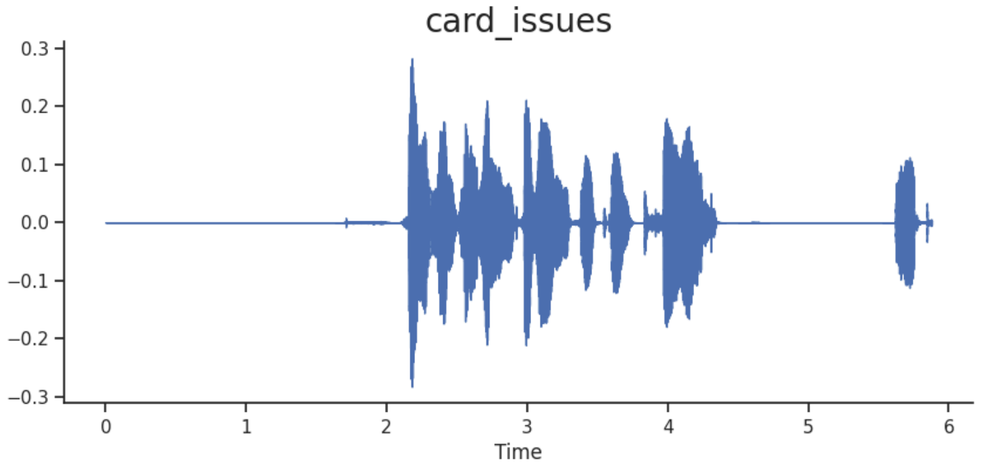
unigram card issues

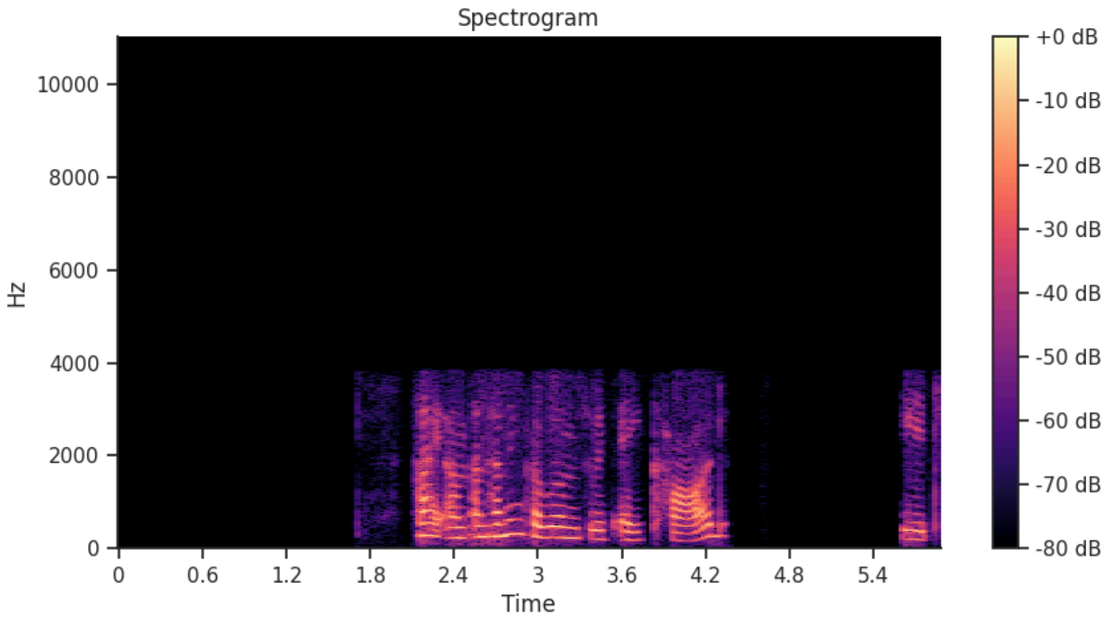
Unigram	Frekuensi
card	44
(not) working	42
help	29
payment	10

tri-gram card issues

Trigram	Frekuensi
please help card	3
wondering could help	2
card payment declined	2
help card (not) working	2

sample waveplot card issues





Exploratory Data Analysis - 4

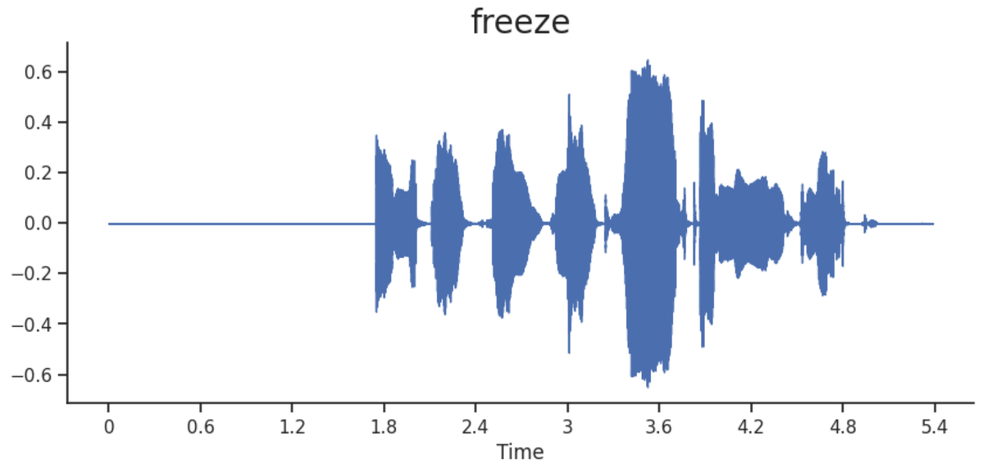
unigram freeze

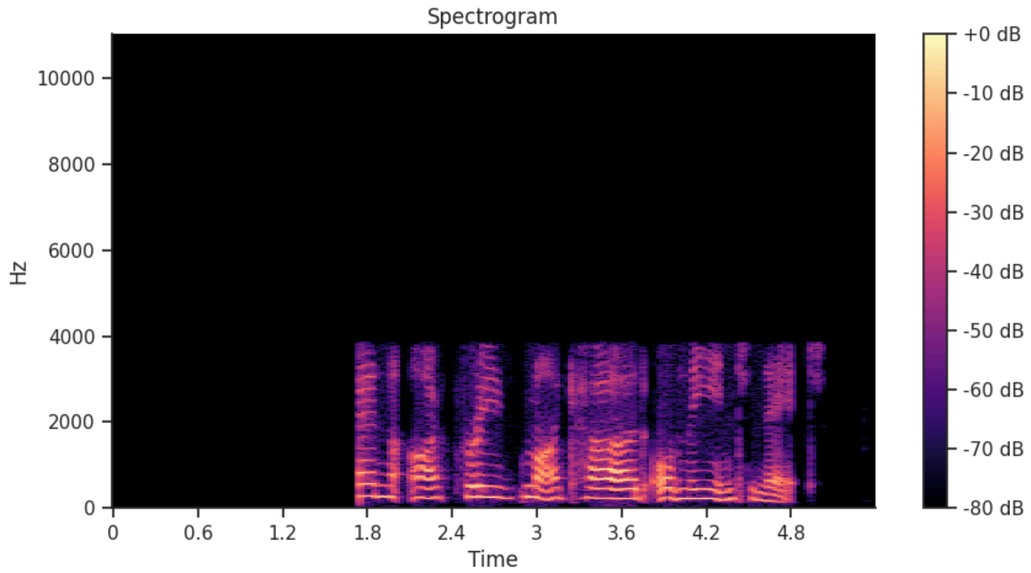
Unigram	Frekuensi
card	42
freeze	27
please	19
like	15

tri-gram freeze

Trigram	Frekuensi
please freeze card	4
id like freeze	4
stop transactions card	3
like freeze card	3

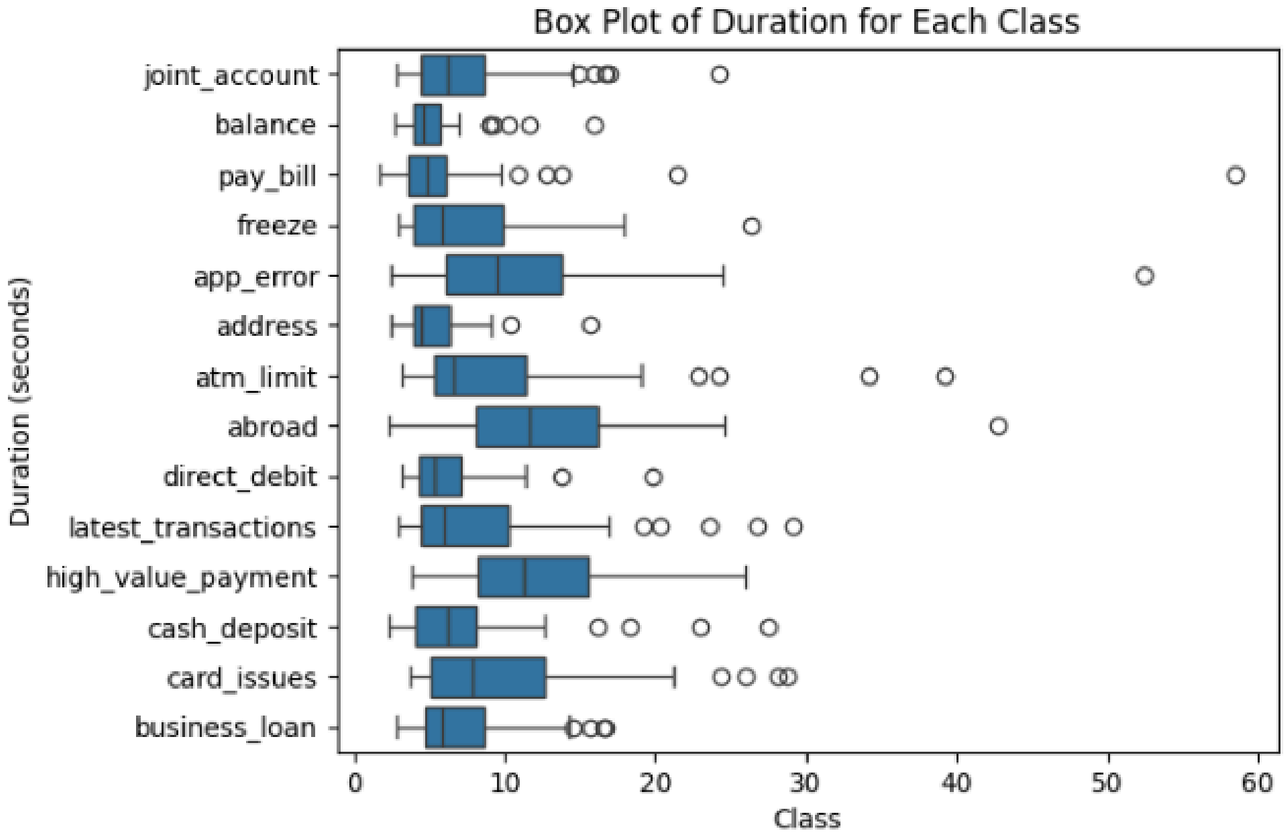
sample waveplot freeze



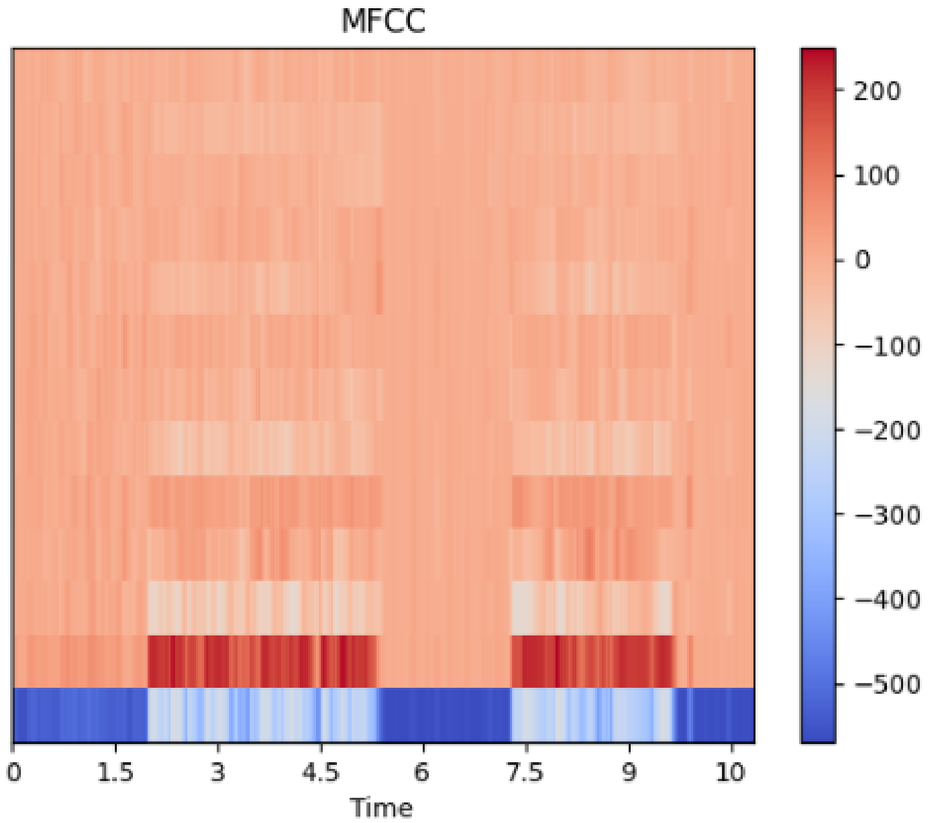


Exploratory Data Analysis – 5

Duration Box Plot



sample MFCC joint_account



Inference Experimentation

Data Usage:

Language	Rows	Used Rows
English - US (EN-US)	563	563
English - AU (EN-AU)	654	100

Performance Metrics:

- **Word Error Rate (WER)**: Mengukur kesalahan dalam perbandingan teks sistem dengan teks referensi, diukur dalam persentase.
- **Character Error Rate (CER)**: Mirip dengan WER, tetapi fokus pada kesalahan karakter, diukur dalam persentase.
- **BLEU (Bilingual Evaluation Understudy)**: Mengukur kesamaan antara teks hasil prediksi dengan teks referensi, menggunakan penghitungan n-gram overlaping precision, memberikan nilai skor antara 0 dan 1.
  - Poor Quality: BLEU < 0.4
  - Reasonable Quality: BLEU 0.4 - 0.5
  - Good Quality: BLEU 0.5 - 0.6
  - High Quality: BLEU > 0.6

Experimentation:

whisper model

- openai/whisper-tiny (Pretrained)
- openai/whisper-large (Pretrained)
- openai/whisper-medium (Pretrained)
- NemesisAlm/whisper-tiny-en (Fine Tuned)

wave2vec model

facebook/wav2vec2-base-960h (Pretrained)

Spärck

Inference Performance Result
(EN-US)

Spärck

Model	Type	WER	CER	BLEU	Avg. Inf. Time
openai/whisper-tiny	Pretrained	51.17%	34.86%	49.85%	1.32 Sec/Pred
openai/whisper-medium	Pretrained	47.47%	32.22%	53.05%	1.10 Sec/Pred
openai/whisper-large	Pretrained	48.37%	32.59%	5215%	1.68 Sec/Pred
NemesisAlm/whisper-tiny-en	Pretrained + Fine Tuned	31.39%	25.64%	71.52%	1.23 Sec/Pred
facebook/wav2vec2-base-960h	Pretrained	52.45%	35.81%	48.67%	0.14 Sec/Pred

Untuk dataset EN-US Model whisper tiny yang difine tune menghasilkan performansi yang jauh lebih baik dibandingkan whisper large dari openai yang dibuktikan dari adanya penurunan WER sekitar 7% dibandingkan whisper large. Namun jika meninjau dari segi efisiensi, whisper medium merupakan model dengan rata-rata waktu inferensi paling rendah dibandingkan model lainnya.

Inference Performance Result (EN-AU)

Spärck

Model	Type	WER	CER	BLEU	Avg. Inf. Time
openai/whisper-tiny	Pretrained	42.70%	23.32%	45.14%	1.58 Sec/Pred
openai/whisper-medium	Pretrained	36.19%	19.60%	53.10%	1.02 Sec/Pred
openai/whisper-large	Pretrained	35.67%	19.07%	53.49%	1.63 Sec/Pred
NemesisAlm/whisper-tiny-en	Pretrained + Fine Tuned	25.85%	18.09%	63.34%	1.54 Sec/Pred
facebook/wav2vec2-base-960h en-AU	Pretrained	42.65%	22.98%	41.71%	0.85 sec/ Pred

Untuk dataset EN-AU Model whisper tiny yang difine tune menghasilkan performansi yang jauh lebih baik dibandingkan whisper large dari openai yang dibuktikan dari adanya penurunan WER sekitar 10% dibandingkan whisper large. Namun jika meninjau dari segi efisiensi, whisper medium merupakan model dengan rata-rata waktu inferensi paling rendah dibandingkan model lainnya.

Inference Sample
using NemesisAlm/whisper-tiny-en (EN-US)

Spärck

Generated Transcription	Reference Transcription
I am trying to use the banking app on my phone and currently my checking and savings account balance is not refreshing	hi I'm trying to use the banking app on my phone and currently my checking and savings account balance is not refreshing
Even possible to deposit the money	is it possible to deposit the money
I want to place a freeze in my card	I want to see freezing my card
I'm trying to make a pretty large payment and I'm getting a message that I will get a text message it says it's an SMS I don't really know what that is or what I'm supposed to do with it	hi I'm trying to make a pretty large payment and I'm getting a message that I will get a text message it says it's an SMS I don't really know what that isn't what I'm supposed to do
payment I have to pay my bill	I have to pay my bill

Summary

- Model fine-tuned whisper **NemesisAlm/whisper-tiny-en** adalah model dengan performa terbaik untuk EN-US dan EN-AU yang dibuktikan dari skor BLEU di atas 60% dengan rata-rata waktu inferensi diatas 1.2 detik/prediksi.
- Model **facebook/wav2vec2-base-960h** adalah model dengan efisiensi terbaik karena proses inferensi untuk EN-US dan EN-AU rata-rata hanya membutuhkan waktu dibawah 1 detik/prediksi.

Future Improvements

- Melakukan proses training dan atau fine-tuning sendiri untuk model wav2vec2 dan whisper
- Melakukan deployment dari model automatic voice recognition
- Melakukan error analysis.
- Menggunakan model automatic voice recognition yang *up-to-date* (SOTA)

Thank you, any question?