CIMB BANK CIMB ISLAMIC

Speech Recognition to Text Analytics

Classifying Phone Banking Customer Sentiment

Group 2

GreatNusa Data Science Training

August 2023

EPICC **Live an EPICC Life**

Meeting agenda

01

Overview

- Background
- Proposed Solution
- Objective
- Potential Benefit

02

Data, Analysis & Results

- Approach & Data Overview
- Speech to Text
- Text Analytics
- Data Prediction Result

03

Conclusion

Overview – Background, Proposed Solution & Potential Benefits

Background

Contact Center is a **key touchpoint** to receive first hand customer experience regarding our Bank

Many **insights** to uncover if we are able to analyze data from contact center

With large volume of calls everyday, how can we **analyze** the **data effectively** & **efficiently?**

Can we use **machine learning** for classification or categorization?

"Selamat malam saya mau masuk ke aplikasi okto kok lama sekali ya?"

Machine learning

Negative Sentiment for OctoMobile

Speech to Text Sentiment Analysis

Capture: Speech Recognition
Translate: *Speech to Text Model*
Evaluate*: Sentiment Analysis*

Proposed Solution

1. **Speech to Text**
   Converting audio recording of customer's call into text for sentiment analysis

2. **Sentiment Analysis**
   Predicting positive / neutral / negative sentiment for a specific banking product / service

Objective

1. Identify & categorize **positive & negative sentiments**

2. *Categorize sentiment* ***based on product***

Potential Benefit

1. ***Provide insight to BU***

2. ***Potential for leads generation***

Approach & Data Overview

Approach

1 Extract train data from Open-Source

Training Dataset obtained from Huggingface

Hugging Face

2 Speech to text Modelling

Using Wav2Vec2 Model & GELUActivation for model evaluation

3 Sentiment Analysis

Sentiment Outcome

Positive | Neutral | Negative

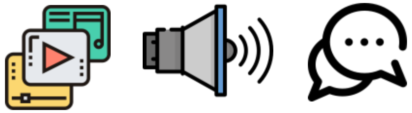
4 Category Prediction

Product Category

Octo Mobile | Tabungan | Credit Card

Data Overview

Extracted Data from Open-Source
5.7 K



1 row audio file = 5 seconds of audio with 5 to 8 words

Each Row use 1 voice source

Sentence Type: Day-to-Day Conversation

Language: Bahasa Indonesia

1 Speech to Text

a. Modeling Framework

1 Data Preprocessing
- Channel Convertion
- Features Extraction
- Tokenization

2 Modeling (Wav2Vec2)
- Parameter Tuning
- Model Optimization

3 Speech to Text Convertion
- Audio Convertion
- Result Evaluation

b. Model Configuration

Wav2Vec2 Model:
•Architecture: Convolutional and self-attention layers.
•Encoder: 24 encoder layers for speech representation.
•Vocabulary Size: 28 tokens (Indonesian).

Wav2Vec2 Processor:
• Mono Channel Convertion.
• Sampling Rate: Resamples audio to 16,000 Hz.

b. Data Splitting

Data Split	Data Count
Data Train	2.1 K
Data Test	1.8 K
Data Validation	1.8 K
Total Dataset	**5.7 K**

b. Model Evaluation

NN Activation Function: GELUActivation

Parameter	Value (%)
CER	0.5
WER	1.9

*** CER = Character Error Rate (%)**
WER = Word Error Rate (%)

2 Text Analytics

a. Data Preprocessing

1. Lower Text
2. Remove Punctuation, Numbers, Slang Language, Whitespace
3. Remove Stop Words
4. Stemming
5. Tokenize
6. Vectorization (TfidfVectorizer)

b. Sentiment & Category Prediction

Sentiment Prediction Model

- Multi Classification Model (Target : Positive, Neutral, Negative)
- Train & Test Split = 0.8 : 0.2
- Total Dataset for Training & Testing = 16,972

Data Training & Testing	Target			Total
	Positive	Neutral	Negative	
Data Train (80%)	5,896	5,896	5,236	13,577
Data Test (20%)	1,474	1,309	612	3,395
Total Dataset (100%)	**7,370**	**6,545**	**3,057**	**16,972**

Best Model = Random Forest

Parameter	Accuracy	Precision	Recall	F-1 Score	Cross-Validation
Value	0.82	0.81	0.82	0.82	0.80

Category Prediction Model

- Multi Classification Model (Target : Octo Mobile, Tabungan, Credit Card, etc.)
- Train & Test Split = 0.8 : 0.2
- Total Dataset for Training & Testing = 36,449

Data Training & Testing	Total
Data Train (80%)	29,159
Data Test (20%)	7,290
Total Dataset (100%)	**36,449**

Best Model = Random Forest

Parameter	Accuracy	Precision	Recall	F-1 Score	Cross-Validation
Value	0.73	0.73	0.73	0.72	0.69

3 Data Prediction Result

Illustration

Customer Phone Call with 14041

Contact Center

Audio

Speech to Text Conversion

Customer

"Mau konfirmasi untuk kartu kredit saya sudah bisa digunakan ya prosesnya cepat ya"

Output: Customer Sentiment & Banking Product in question

Sentiment: Positive

Product: Credit Card

Sample:

Text	Sentiment	Category
"selamat malam saya mau buka tabungan di sini bagaimana caranya"	neutral	Tabungan
"selamat malam saya mau masuk ke aplikasi okto kok lama sekali ya"	negative	Octo Mobile
"selamat malam terima kasih untuk kredit kar nya ya banyak promusi"	positive	Credit Card
"apa ada tawaran kartu husus untuk pengguna dibawah umur dua puluh lima tahun"	neutral	Credit Card
"halo terdapat kekurangan saldo ditabungan saya harap di perbaiki"	neutral	Tabungan

Key Highlight:

Model is able to convert audio to text and predict what kind of sentiment and product category being discussed in the audio file

Summary & What's Next

Conclusion

1. **Wav2Vec2 model** is used using convolutional and self attention layers, 24 encoders for speech representation and 28 tokens of vocabulary size using Bahasa Indonesia

2. Using **GELUActivation** for model evaluation resulted in **CER** values of 0.5 and **WER** values of 1.9

3. **Random forest** is the best model for both Sentiment Prediction & Category Prediction Model, with **accuracy** of **0.82 & 0.73** and **cross-validation** of **0.80 & 0.69** respectively

4. **Model is ready** to be used, however **audio quality** from Contact Center should be **clear** for better result

Recommendation & Next Action Plan

Optimize Resource used

Using this framework from end-to-end requires large amount of resource to execute

Differentiating multiple voice source

Currently the model only use 1 voice source as basis

Apply Real-Time Sentiment Analysis from Voice Recognition

Real-time sentiment Analysis model has been developed, but there is issue regarding noise reduction

Thank You

Appendix

Data	Target			Total
	Positive	Neutral	Negative	
Data Train (80%)	5,896	5,236	2,445	13,577
Data Test (20%)	1,474	1,309	612	3,395
Total Dataset (100%)	7,370	6,545	3,057	16,972

Year	Weight	Probability of Positive Sentiment	Scoring per Year
Year 0 & -1 (2023 – 2022)	45%	Total Positive Sentiment / Total Review in respectively each year's group	0.45 * Probability of positive sentiment
Year -2 & -3 (2021 – 2020)	35%		0.35 * Probability of positive sentiment
<= Year -4 (<=2019)	20%		0.2 * Probability of positive sentiment
Total Scoring			Total of Scoring each Year's Group

Parameter	Accuracy	Precision	Recall	F-1 Score	Cross-Validation
Value	0.73	0.73	0.73	0.72	0.69

Sample Messages

Sample Messages	Sentiment	MCC
Rotinya beda dari yang lain, lembut banget, banyak rasa, murah dan meriahðŸ' ðŸ' ðŸ'	Positive	Restaurant
Pesan antar	Neutral	Restaurant
Penjualnya sangat arogan. Mereka melayani saya dengan sikap yang sangat buruk, bahkan salah satu dari mereka tidak menjawab pertanyaan saya tentang promo dengan cara yang tepat. Ketika saya mengambil pesanan saya, mereka sangat kasar untuk memberitahu saya untuk menyimpan tanda terima ketika saya mengatakan kepadanya bahwa saya telah membuangnya ke tempat sampah. Pengalaman yang sangat tidak menyenangkan.	Negative	Restaurant
Good place, barangnya juga lumayan bagus, gak nyesel belanja disini ðŸ'	Positive	Clothing Store
Tiba-tiba ingatanku kembali ke tahun 1992-1995... Itu sudah menjadi bagian dari waktu hidupku yang berharga...	Neutral	Clothing Store
Pelayanannya buruk setiap kali ke sana kasirnya tidak sopan atau mungkin malas melayani...kalau tidak mau jadi kasir stop saja	Negative	Clothing Store