

INTISARI

MULTICLASS CHAT TOPIC CLASSIFICATION PADA DATASET FALLBACK CHATBOT DALAM DOMAIN TELEKOMUNIKASI MENGGUNAKAN TEMPORAL CONVOLUTIONAL NETWORK (TCN)

Oleh

IVAN LIU NARDO SHIDDIQ

18/427581/PA/18541

Layanan digital berupa *chatbot* dalam domain telekomunikasi masih memiliki kekurangan di mana *chatbot* kerap tidak mengerti pesan yang dikirimkan oleh pelanggan sehingga terjadi *fallback*, yaitu *chatbot* hanya dapat menjawab dengan pesan *default*. Oleh sebab itu, diperlukan suatu pengembangan arsitektur yang dapat mengategorikan pesan-pesan pelanggan yang bersifat teks pendek serta tidak baku yang menyebabkan pesan-pesan tersebut tidak dapat dimengerti oleh *chatbot*.

Dalam penelitian ini, dikembangkan model *Temporal Convolutional Network (TCN)* untuk klasifikasi *multiclass fallback chat* untuk meminimalisir terjadinya *fallback*. *Dataset* pesan-pesan pelanggan yang digunakan pada penelitian ini menggunakan bahasa Indonesia. Pesan-pesan *fallback* tentunya masih belum memiliki label sehingga digunakan metode *Gibbs Sampling Dirichlet Multinomial Mixture* untuk melakukan *topic modelling* dengan tujuan memberikan label kepada setiap pesan. Pesan yang sudah diberikan label kemudian diklasifikasi menggunakan arsitektur *TCN* dengan model *word embedding* berupa *FastText* dan *IndoBERT*.

Tugas *topic modelling* menggunakan metode *Gibbs Sampling Dirichlet Multinomial Mixture* yang digunakan setelah diverifikasi secara manual mampu menghasilkan nilai akurasi sebesar 83,01% dengan 16.602 pesan memiliki label yang benar dan 3.398 pesan memiliki label yang salah. Tugas klasifikasi teks menggunakan arsitektur *TCN – FastText* yang dikembangkan mampu menghasilkan nilai *F1-Score* sebesar 93,49% saat *training* dan 95,51% saat *testing*. Sedangkan arsitektur *TCN – IndoBERT* yang dikembangkan mampu menghasilkan nilai *F1-Score* sebesar 93,1% saat *training* dan 93,44% saat *testing*. Dari sisi waktu *training* model, arsitektur *TCN – FastText* mampu melakukan *training* ± 15 kali lebih cepat dibandingkan arsitektur *TCN – IndoBERT*.

Kata Kunci: *Chatbot Fallback, Gibbs Sampling Dirichlet Multinomial Mixture, Temporal Convolutional Network, FastText, IndoBERT.*

ABSTRACT

MULTICLASS CHAT TOPIC CLASSIFICATION ON CHATBOT'S FALLBACK DATASET IN TELECOMMUNICATION DOMAIN USING TEMPORAL CONVOLUTIONAL NETWORK (TCN)

By

IVAN LIU NARDO SHIDDIQ

18/427581/PA/18541

There is still a shortage of chatbot-style digital services in the telecommunications sector, as chatbots often fail to understand the messages sent by customers, leading to fallbacks where chatbots can only reply with default standard messages. As a result, it is essential to develop an architecture that can classify customer messages that contain short and irregular language, making them difficult for chatbots to understand.

In this research, a Temporal Convolutional Network (TCN) model was developed for the classification of multiclass fallback messages in order to lessen the risk of fallback. The customer message dataset that was utilized in this research is in Indonesian. Obviously, the backup messages still lack labels, thus topic modelling is done using the Gibbs Sampling Dirichlet Multinomial Mixture approach in order to label each message. Once messages have been labelled, they are then categorized using the TCN architecture utilizing FastText and IndoBERT as word embedding models.

The topic modelling task utilizing Gibbs Sampling Dirichlet Multinomial Mixture approach, after being manually validated, was able to produce an accuracy result of 83.01%, with 16,602 messages having the correct label and 3,398 messages having the incorrect label. As for the text classification task, the TCN - FastText architecture was able to obtain F1-Score values of 93.49% during training and 95.51% during testing. The TCN - IndoBERT architecture, on the other hand, was able to generate F1-Score values of 93.1% during training and 93.44% during testing. The TCN - FastText architecture can train models approximately 15 times quicker than the TCN - IndoBERT architecture.

Keywords: Chatbot Fallback, Gibbs Sampling Dirichlet Multinomial Mixture, Temporal Convolutional Network, FastText, IndoBERT.