*ICIC Express Letters*
Volume **19**, Number **11**, November **2025**
ICIC International ©2025 ISSN 1881-803X
pp. **1183–1190**

# INCORPORATING INDOBERT INTO DIET CLASSIFIER TO ENHANCE INTENT AND ENTITY RECOGNITION IN CULINARY CHATBOT

Nadhifa Sofia, Edi Winarko* and Sigit Priyanta

Faculty of Mathematics and Natural Sciences
Universitas Gadjah Mada
Building C, 4th Floor, North Sekip, Bulaksumur, Yogyakarta 55281, Indonesia
nadhifa.sofia@mail.ugm.ac.id; seagatejogja@ugm.ac.id
*Corresponding author: ewinarko@ugm.ac.id

Received November 2024; accepted February 2025

Abstract. *Dining out, working from cafes, and socializing in restaurants become the modern lifestyles, creating a growing need for accurate tools like culinary chatbots. This paper introduces Restbot, a restaurant-related chatbot. Unlike traditional heuristic or machine learning methods, which often struggle with dynamic data or complex queries, Restbot leverages a customized Dual Intent Entity Transformer (DIET) model integrated with IndoBERT. It improves intent detection and entity recognition, particularly for non-standard and slang Indonesian language inputs common in the culinary domain. Using a dataset of 4,011 sentences across 17 intents and 1,079 menus from 203 restaurants, the DIET+IndoBERT model achieves significant performance compared with DIET: 96% F1-score for entity recognition (4% higher) and 92% F1-score for intent detection (3% improvement). This research contributes a novel approach to enhancing chatbot accuracy in specialized domains, paving the way for more intuitive and user-friendly culinary assistants.*
**Keywords:** Chatbot, Natural language processing, Natural language understanding, System identification

1. **Introduction.** Dining out for lunch, working from cafes, or socializing in restaurant become the lifestyle nowadays. Too many restaurants make people hard to choose. Chatbots will also help them who are unfamiliar with the area. Chatbots are more attractive because they provide short yet precise answers and respond to questions at any time [1].

Researchers developed chatbots with rule-based heuristic method [2]. It processes static data and rules requiring them to keep adding rules and datasets. Machine Learning (ML) overcomes it by computing dynamic data, classifying questions, and finding the answers [3]. ML uses simple statistics; therefore, it is hard to answer various questions and requiring many parameters for finetuning. Deep Learning (DL) tackles it by training more layers and nodes on the system [4,5]. On the other hand, Recurrent Neural Network (RNN) can only understand 86.36% of questions [6], Long Short-Term Memory (LSTM) model produced 87.6% of accuracy, and Gated Recurrent Unit (GRU) produced 87.5% of accuracy [7].

Unlike RNN, DIET considers total loss by accumulating intent and entity loss jointly [8]. DIET enhances intent prediction by utilizing CLS tokens, which capture the meaning of all words in a sentence. Intent detection and entity recognition are essential in understanding chatbots. Without this combined task, the system processes two models separately, reducing performance [9-11]. However, the basic DIET achieved less accuracy than our customized model.

DOI: 10.24507/icicel.19.11.1183

Indonesian chatbots cannot be treated like English chatbots. Indonesians tend to use non-standard and slang words. Those words are updated rapidly in the future, making chatbots challenging to process. There are some Indonesian chatbots with another domains [12,13], but there is no restaurant chatbot. Therefore, we develop Restbot with Indonesian culinary domain. We generated a novel dataset, comprising 4,011 user sentences, 17 intents, and 1,079 unique menu items from 203 restaurants. We incorporated IndoBERT into DIET classifier to significantly improve the intent and entity recognition tasks. By addressing challenges posed by slang and non-standard language, this research provides a practical solution to a growing need in the industry. Additionally, the system demonstrates robustness in handling complex queries, making it a valuable tool for enhancing user experiences in conversational agents.

The remaining of this paper is organized as follows. Section 2 presents the dataset and details the structure of the incorporated DIET classifier. Section 3 explains the experimental setup and interprets the result. Finally, Section 4 provides the conclusions of this research.

2. **Materials and Methods.** Figure 1 defines the proposed architecture in this research. The first step, data acquisition, is to collect user sentences, restaurant, and menu lists. Secondly, pre-processing is to standardize the data and, next, split data 80% : 20% for training and testing. Subsequently, train the Natural Language Understanding (NLU) model using DIET+IndoBERT. NLU produces total loss and detected intent and entity. The system selects answers by Natural Language Generation (NLG). Lastly, performance

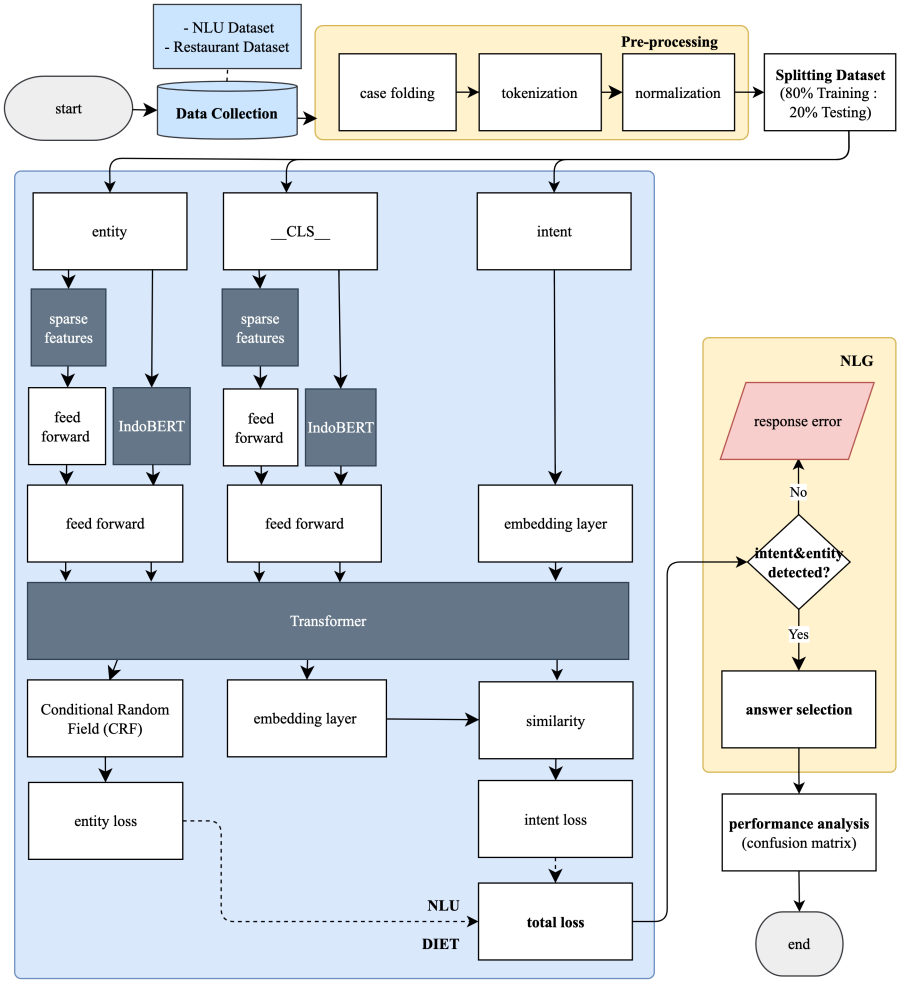



FIGURE 1. Proposed architecture

analysis is to measure the chatbots with a confusion matrix and reduce total loss with Adam Optimizer. The system evaluates the trained model with testing data.

2.1. **Data acquisition.** This research collects primary datasets from real users to achieve higher accuracy. Their user personas are someone who 1) works from the coffee shop, 2) loves exploring restaurants for lunch, 3) goes to a café to hang out with friends, and 4) is new in town. We collected user utterances in the Google Form from 18 participants. There are 4,011 sentences with 17 intents. Table 1 maps the intent with its sentences.

Table 1. Example of mapping intents and input sentences

| No | Intent | Sentence samples | |
|---|---|---|---|
| 1 | Greet | *Hi bot* | *Assalamu'alaikum* |
| 2 | Reject | *Hmm kayanya engga deh* | *Ada opsi lain ga bot?* |
| 3 | Bye | *Bye bot!* | *C u soon bot* |
| 4 | Confirm | *Gw bakal nyobain itu* | *Wah, boleh juga saran lo bot* |
| 5 | Restaurant name | *pizza hut delivery tamsis* | *warung kopi klotok* |
| 6 | Thanks | *Makasi ya restbot* | *Thank you banget lho* |
| 7 | Challenge bot | *Kamu bot ya?* | *Bisa jawab ga?* |
| 8 | Negative sample | *Saya tidak suka makanan pedes* | *Lg bokek, gamau auce* |
| 9 | Happy mood | *Luar biasa* | *Enak lho disitu* |
| 10 | Restaurant menu | *Apa saja menu di sushi tei?* | *Share menu di gudeg sagan bot* |
| 11 | Restaurant location | *Dimana ya lokasi bricafe x couvee?* | *Alamat mie sapi banteng* |
| 12 | Restaurant working hours | *Jam berapa fore coffee jakal buka?* | *Pukul berapa golden geisha ramen tutup ya bot?* |
| 13 | Restaurant pricing | *Berapa harga gelato di tempo gelato prawirotaman?* | *bisa informasikan harga menu di giorno's pizzeria?* |
| 14 | Restaurant facility | *Swasana by Temata punya fasilitas apa aja?* | *Apa saja fasilitas di Petit Paris Boulangerie & Bistro?* |
| 15 | Ask 24 hours resto | *Bot, info coffee shop yang buka 24/7 dong.* | *rumah makan yang masih buka jam 11 malem dimana ya?* |
| 16 | Restaurant reservation | *mau bikin reservation di abhaya-giri dong bot* | *mau reserve di wanawatu untuk 10 orang tanggal 3 jam 11.00* |
| 17 | Restaurant discount | *Di mana ya gw bisa dapetin birthday treat, bot?* | *Tau tempat makan yang lagi ada promo imlek nggak, bot?* |

This research also collects some entities so the system can understand the context of the sentences. We collected restaurant type, restaurant name, facility, and location entity. Table 2 represents the example of entities.

Table 2. Example of entities

| Restaurant type | Restaurant name | Facility | Location |
|---|---|---|---|
| Javanese *ngopi* | The House of Raminten BRICafe x Couvee | *WiFi colokan* | *Jakal UGM* |

Some intents require restaurant data for NLG, so we collected 203 restaurant lists with their addresses, categories, working hours, and facilities. We have scraped restaurant and menu lists from Google Maps per May 2024. Table 3 demonstrates restaurant data with the restaurant's address, categories, working hours, and facilities.

Table 3. Sample of restaurant data

| Name | Address | Category | Working hour | Facility |
|---|---|---|---|---|
| *Sate Klathak Pak Pong* | *Jl. Sultan Agung No.18, Jejeran II, Wonokromo, Kec. Pleret, Kabupaten Bantul, Daerah Istimewa Yogyakarta 55791* | Satay, Javanese, Authentic, Top | *Minggu: [09.30-23.00], Senin: [09.30-23.00], Selasa: [09.30-23.00], Rabu: [09.30-23.00], Kamis: [09.30-23.00], Jumat: [09.30-23.00], Sabtu: [09.30-23.00]* | *tempat parkir, mushalla/tempat shalat, ruangan indoor, toilet umum* |

Table 4. Sample of menu lists

| Restaurant name | Menu | Price |
|---|---|---|
| *Sate Klathak Pak Pong* | *sate klathak* | 29000 |
| *Sate Klathak Pak Pong* | *kronyos goreng* | 20000 |
| *Sate Klathak Pak Pong* | *tengkleng* | 35000 |

We differentiated the restaurant and menu table because one restaurant can have more than one menu. We scraped 1,079 menus from 203 restaurants. Table 4 illustrates the menu lists.

2.2. **Data pre-processing.** Input needs to be transformed into the standardized Indonesian words with data pre-processing. We did case folding to make all of the words in lowercase. Next, tokenization is to separate each word with whitespaces. Lastly, normalization is to normalize non-standard words. We have created 1,536 normalized words.

2.3. **Text featurizer.** Text featurizer transforms the pre-processed data into numerical feature vectors by counting the frequency of each word in the dataset. This step is crucial for converting text into a format that the DIET Classifier can process, helping the model learn patterns and relationships between the words and their corresponding intents and entities. We used one-hot encoding as Count Vectorizer to avoid overfitting of long text [8]. The system adds the *_CLS_* token to summarize the utterances and predict the intents. The system combines it with the activation function $\sigma(W_x + b)$, by multiplying the matrix $W_x$ and adding bias. It produces an $array(256)$, and then enters two feed-forward layers before entering the Transformer layer so it can understand the meaning of the words in a sentence.

2.4. **Feed forward layer.** Feed forward layer enables the text featurizer into a higher-dimensional space. This layer applies a set of learned weights to the input vectors, generating output representations with $h = \sigma(W_x + b)$ where $h$ is the hidden layer representation, as $array(256)$. $\sigma$ is the activation function like ReLU multiply by the weighted matrix $W_x$ and bias vector $b$.

2.5. **IndoBERT.** BERT pre-trains deep bidirectional representations and is powerful for question answering, sentiment analysis, and language translation [14,15]. It considers all layers. Standard DIET does not use BERT. We incorporated IndoBERT into DIET to capture richer semantic information from input tokens, especially Indonesian. It finetunes these embeddings within the context of both entity and intent classification tasks. Given an input sequence $X$, the probability of predicting a masked token $x_i$ is shown in Equation (1). $H_i$ is the hidden state of the $i$-th token from the final Transformer layer. $W_h$ is a learnable weight matrix mapping hidden states to vocabulary logics [16].

$$P(x_i|X) = softmax(W_h H_i). \tag{1}$$

2.6. **Transformer.** The Transformer in the DIET Classifier captures relationships between tokens in a sequence by utilizing self-attention mechanisms, enabling the model to process both intents and entities simultaneously. Unlike RNN, which requires learning from the calculations of neighboring nodes, the Transformer only focuses on the nodes without depending on neighboring nodes, so it requires less training time [8]. Equation (2) defines the self-attention mechanism:

$$Attention(Q, K, V) = softmax\left(\frac{QK^T}{\sqrt{d_k}}\right)V, \tag{2}$$

where $Q$, $K$, and $V$ are the query, key, and value matrices derived from the input and $d_k$ is the keys' dimensionality.

2.7. **Total loss.** Entity recognition utilizes the Transformer architecture to process token-level representations $y_{entity}$ with Conditional Random Field (CRF) on Transformer output sequence $a$ to produces entity loss $L_E$. Equation (3) shows the entity loss calculation:

$$L_E = L_{CRF}(a, y_{entity}). \tag{3}$$

Intent detection utilizes dot-product error to optimize the similarity $S_I^+ = h_{CLS}^T h_{intent}^+$ with $y_{intent}^+$ as the target label. It reduces similarities $S_I^- = h_{CLS}^T h_{intent}^-$ with $y_{intent}^-$ as the negative samples shown in Equation (4). It produces intent loss and $_CLS_$ token [8].

$$L_I = -\left\langle S_I^+ - \log\left(e^{S_I^+} + \sum_{\Omega_I^-} e^{S_I^-}\right)\right\rangle. \tag{4}$$

The system calculates total loss with intent loss $L_I$ and entity loss $L_E$, described in Equation (5). The system minimizes the model with Adam Optimizer algorithm. The aim is to get smaller error/loss values over time and to see the effectiveness of a model [8].

$$L_{total} = L_I + L_E. \tag{5}$$

3. **Experiment and Results.** To evaluate the effectiveness of IndoBERT in improving intent classification and entity recognition, we built and compared three different Rasa pipeline configurations: ML, DIET, and DIET+IndoBERT. ML model leverages Standard Vector Machine (SVM) for intent classification (*SklearnIntentClassifier*) and CRF for entity extraction (*CRFEntityExtractor*). The DIET model, a Transformer-based approach, uses the *EmbeddingIntentClassifier*, *CountVectorsFeaturizer*, and *WhitespaceTokenizer*, optimizing loss across both intent and entity recognition. The DIET+IndoBERT model introduces *LanguageModelFeaturizer* with IndoBERT, enhancing contextual embeddings for more accurate language understanding. Additionally, we adjusted *TwoStageFallbackPolicy* for better handling of ambiguous queries by lowering the *nlu_threshold* to 0.3.

The configurations maintained a consistent structure, with 100 epochs for DIET models and batch sizes of 32 and 64. IndoBERT's pre-trained model was integrated to improve feature extraction, leveraging a lower learning rate 0.005 for stable convergence. While machine learning models rely on word-based tokenization and linear SVM, DIET models use Transformers to capture dependencies within sentences efficiently. The DIET+IndoBERT model further enhances performance by utilizing contextual embeddings, leading to better generalization in unseen queries. Notably, entity recognition benefits significantly from the integration of IndoBERT, as it learns richer semantic representations compared to standard DIET and SVM-based approaches.

The system used the confusion matrix to measure the model with several parameters: precision, recall, accuracy, and F1-score, like in [9,10,12,13]. Table 6 presents the evaluation metrics across the three configurations. The DIET+IndoBERT model achieves the highest F1-score, 0.92 for intent recognition and 0.96 for entity recognition, outperforming both Standard DIET (0.88, 0.93) and ML (0.86, 0.90) configurations. Precision and recall

Table 5. Configuration comparison of three models

| Machine learning | Standard DIET | DIET+IndoBERT |
|---|---|---|
| language: id<br>pipeline:<br>- name:<br>CountVectorsFeaturizer<br>analyzer: “word”<br>min_ngram: 1<br>max_ngram: 2<br>- name:<br>SklearnIntentClassifier<br>model: “svm”<br>kernel: “linear”<br>C: 1.0<br>probability: True<br>random_state: 42<br>- name:<br>CRFEntityExtractor<br>- name:<br>EntitySynonymMapper | language: id<br>pipeline:<br>- name:<br>CountVectorsFeaturizer<br>- name:<br>EmbeddingIntentClassifier<br>policies:<br>- name: EmbeddingPolicy<br>max_history: 10<br>epochs: 100<br>batch_size:<br>- 32<br>- 64<br>- name: DIETClassifier<br>epochs: 100<br>learning_rate: 0.001<br>num_transformer_layers:<br>1<br>embedding_dimension: 20<br>weight_sparcity: 0.8<br>hidden_layer_sizes:<br>text: [256]<br>- name:<br>AugmentedMemoization<br>Policy<br>core_threshold: 0.3<br>- name:<br>TwoStageFallbackPolicy<br>nlu_threshold: 0.8<br>- name: FormPolicy<br>- name: MappingPolicy | language: id<br>pipeline:<br>- name: components.Preprocessing_component.Pre-process<br>- name:<br>CountVectorsFeaturizer<br>analyzer: char_wb<br>min_ngram: 1<br>max_ngram: 4<br>- name:<br>LanguageModelFeaturizer<br>model_name: “bert”<br>model_weights: “indobenchmark/indobert-base-p1”<br>- name: DIETClassifier<br>epochs: 20<br>learning_rate: 0.005<br>num_transformer_layers: 2<br>embedding_dimension: 10<br>weight_sparcity: 0.90<br>hidden_layer_sizes:<br>text: [256, 128]<br>- name: FallbackClassifier<br>threshold: 0.3<br>ambiguity_threshold: 0.1<br>tensorboard_log_directory:<br>./.tensorboard/DIET<br>tensorboard_log_level: epoch |

Table 6. Intent and entity evaluation for three pipelines

| Config | Intent evaluation | | | | Entity evaluation | | | |
|---|---|---|---|---|---|---|---|---|
| | Prec. | Recall | Acc. | F1-score | Prec. | Recall | Acc. | F1-score |
| Machine learning | 0.87 | 0.86 | 0.86 | 0.86 | 0.9 | 0.91 | 0.97 | 0.9 |
| DIET | 0.89 | 0.88 | 0.88 | 0.88 | 0.93 | 0.95 | 0.987 | 0.93 |
| DIET+IndoBERT | 0.92 | 0.93 | 0.92 | **0.92** | 0.95 | 0.97 | 0.99 | **0.96** |

improvements are particularly evident in entity recognition, where DIET+IndoBERT attains 0.95 precision and 0.97 recall, demonstrating its superior ability to capture named entities. ML approach, while competitive, lacks contextual awareness, resulting in lower intent classification accuracy 0.86. The DIET+IndoBERT model's overall accuracy 0.92 for intent, 0.99 for entity suggests its effectiveness in handling conversational nuances, making it the most robust configuration for intent classification and entity extraction in Indonesian-language chatbots.

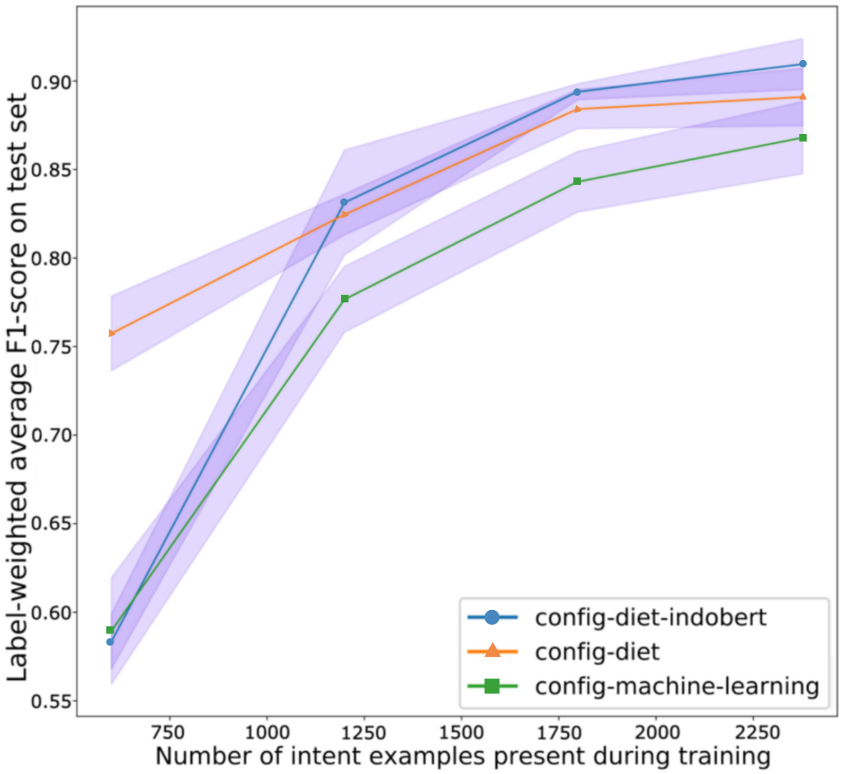



FIGURE 2. F1-score comparison of DIET+IndoBERT with basic DIET and ML model

By integrating IndoBERT into the DIET classifier, the model benefits from the rich contextual embeddings to achieve more accurate intent and entity recognition in overall results. Figure 2 illustrates the F1-score comparison between the ML (config-machine-learning), basic DIET (config-diet), DIET+IndoBERT (config-diet-indobert). As the number of intent examples increases during training, the DIET+IndoBERT model consistently outperforms all models, achieving the highest F1-score across all data points, indicating improved intent detection and entity recognition capabilities. It highlights the effectiveness of incorporating IndoBERT into DIET Classifier in improving accuracy, precision, recall, and F1-score for both tasks.

4. **Conclusions.** Since the culinary industry arose post-pandemic, people tend to go to lunch, work from a café, or hang out with friends. We built the Restbot to ease their search for restaurant information by engaging them with our assistant. We identified that language barriers pose significant challenges in developing a high-accuracy chatbot. Indonesians tend to use non-standard and slang words. They will use it more in the future, making chatbots challenging to process. Our incorporated DIET Classifier significantly improved the performance in overall results, with 92% F1-score for intent detection with 4% increment and 96% F1-score for entity extraction with 3% improvement from basic DIET. In the future, we will incorporate a recommendation system to improve user engagement and make the chatbot a more valuable tool in the culinary industry. We will still allow users to use non-standard words for the Indonesian, so users are still comfortable using it.

**Acknowledgment.** This work is supported by Department of Computer Science and Electronics, Universitas Gadjah Mada. The authors also gratefully acknowledge the helpful comments and suggestions of the reviewers, which have improved the presentation.

## REFERENCES


[1] D. Jannach, A. Manzoor, W. Cai and L. Chen, A survey on conversational recommender systems, *ACM Comput. Surv.*, vol.54, no.5, DOI: 10.1145/3453154, 2021.

[2] J. Lee, T. An, H. E. Chu, H. G. Hong and S. N. Martin, Improving science conceptual understanding and attitudes in elementary science classes through the development and application of a rule-based AI chatbot, *Asia-Pacific Sci. Educ.*, vol.13, no.2, pp.365-412, DOI: 10.1163/23641177-bja10070, 2023.

[3] S. N. Shirsat, T. J. Khan, P. A. Wable, D. B. Mandalik and P. M. T. Jagtap, Healthcare chatbot system using multinomial Naïve Bayes algorithm, *IJARIIE*, vol.8, no.2, pp.1826-1831, 2022.

[4] Nurjayanti, Adiwijaya and S. A. Faraby, Learning contextual meaning for question retrieval using Siamese LSTM on Islamic question answering system, *ICIC Express Letters*, vol.16, no.9, pp.1011-1017, DOI: 10.24507/icicel.16.09.1011, 2022.

[5] R. Kusumaningrum, K. Kadarisman, S. N. Endah, P. S. Sasongko, K. Khadijah, S. Sutikno, R. Rismiyati and A. Afriani, Automated essay scoring using convolutional neural network long short-term memory with mean of question-answer encoding, *ICIC Express Letters*, vol.18, no.8, pp.785-792, DOI: 10.24507/icicel.18.08.785, 2024.

[6] P. Muangkammuen, Automated Thai-FAQ chatbot using RNN-LSTM, *2018 22nd Int. Comput. Sci. Eng. Conf.*, pp.1-4, 2018.

[7] P. Florian and W. Louis, *Design and Implementation of Chatbot in the Context of Customer Support*, Master Thesis, Faculty of Applied Sciences, University of Liège, 2018.

[8] T. Bunk, D. Varshneya, V. Vlasov and A. Nichol, DIET: Lightweight language understanding for dialogue systems, *arXiv Preprint*, arXiv: 2004.09936, 2020.

[9] P. Lorenc, Joint model for intent and entity recognition, *International Journal of Hybrid Intelligent System*, pp.1-5, 2019.

[10] W. A. Abro, G. Qi, M. Aamir and Z. Ali, Joint intent detection and slot filling using weighted finite state transducer and BERT, *Appl. Intell.*, DOI: 10.1007/s10489-022-03295-9, 2022.

[11] H. Weld, X. Huang, S. Long, J. Poon and S. C. Han, A survey of joint intent detection and slot filling models in natural language understanding, *ACM Comput. Surv.*, vol.37, no.4, DOI: 10.1145/3547138, 2022.

[12] L. W. Narendra and E. R. Setyaningsih, Designing a transactional smart assistant in Indonesian using Rasa framework, *The 7th Int. Conf. Electr. Electron. Inf. Eng. Technol. Breakthr. Gt. New Life (ICEEIE 2021)*, DOI: 10.1109/ICEEIE52663.2021.9616946, 2021.

[13] Y. Windiatmoko, R. Rahmadi and A. F. Hidayatullah, Developing Facebook chatbot based on deep learning using RASA framework for university enquiries, *IOP Conf. Ser. Mater. Sci. Eng.*, vol.1077, no.1, 012060, DOI: 10.1088/1757-899x/1077/1/012060, 2021.

[14] A. Chowanda, Emotions classification from conversation using BERT, *ICIC Express Letters*, vol.16, no.7, pp.787-794, DOI: 10.24507/icicel.16.07.787, 2022.

[15] E. Maekawa and H. Murao, Measuring sentence similarity for text simplification: Identifying effective metrics, *ICIC Express Letters*, vol.18, no.9, pp.927-933, DOI: 10.24507/icicel.18.09.927, 2024.

[16] M. C. Kenton, L. Kristina and J. Devlin, BERT: Pre-training of deep bidirectional Transformers for language understanding, *arXiv Preprint*, arXiv: 1810.04805, 2019.