Accredited SINTA 2 Ranking
Decree of the Director General of Higher Education, Research, and Technology, No. 158/E/KPT/2021
Validity period from Volume 5 Number 2 of 2021 to Volume 10 Number 1 of 2026

Published online at: **http://jurnal.iaii.or.id**

JURNAL RESTI
(Rekayasa Sistem dan Teknologi Informasi)
Vol. 9 No. 1 (2025) 111 - 117 e-ISSN: 2580-0760

NLP-Based Intent Classification Model for Academic Curriculum Chatbots in Universities Study Programs

Najma Rafifah Putri Syallya[1*], Anindya Apriliyanti Pravitasari[2], Afrida Helen[3]
[1,2]Department of Statistics, Universitas Padjadjaran, Bandung, Indonesia
[3]Department of Computer Science, Universitas Padjadjaran, Bandung, Indonesia
[1]najma20005@mail.unpad.ac.id, [2]anindya.apriliyanti@unpad.ac.id, [3]helen@unpad.ac.id

Abstract

Chatbots are increasingly prevalent in various fields, including academic fields. Universities often rely on lecturers and staff for information access, which can lead to delays, limited availability outside working hours, and the risk of missed questions. This study aims to develop a chatbot model capable of addressing questions about the curriculum through intent classification, reducing reliance on manual responses, and providing a solution that ensures quick, accurate information retrieval. The research focuses on optimizing the IndoBERT model for intent classification and addresses challenges that arose due to imbalanced data, which could have impacted model performance. Data was collected through an open poll on common curriculum-related questions asked by students. To address data imbalance, we tried oversampling techniques, such as SMOTE, B-SMOTE, ADASYN, and Data Augmentation. Data augmentation was chosen and successfully addressed the imbalance problem while maintaining data semantics effectively. We achieved the best model with hyperparameters batch size of 8, learning rate of 0.00001, 15 epochs, and 64 neurons in the hidden layer, resulting in 98.7% accuracy on the test data. Evaluation metrics further demonstrate the model's robustness across multiple intents. This research demonstrates the advantages of the IndoBERT model in intent classification for academic chatbots, achieving excellent performance.

Keywords: *Chatbot, IndoBERT, Intent Classification*

How to Cite: Najma Rafifah Putri Syallya, Anindya Apriliyanti Pravitasari, and Afrida Helen, "NLP-Based Intent Classification Model for Academic Curriculum Chatbots in Universities Study Programs", *J. RESTI (Rekayasa Sist. Teknol. Inf.)* , vol. 9, no. 1, pp. 111 - 117, Feb. 2025.
DOI: https://doi.org/10.29207/resti.v9i1.6276

1. Introduction

Chatbots are rapidly advancing and increasingly prevalent in various fields, including the research and academic fields. Chatbots are an information system that uses artificial intelligence to enable computers to conduct natural language conversations with humans [1]. Chatbots can help students answer academic-related questions, provide additional learning resources and also serve as a means of information dissemination. Chatbot allows students to convey issues or difficulties they face in their studies by asking the university-provided chatbot. Chatbots can lessen the administrative workload, allowing lecturers and staff to focus on academic activities and research [2]. Chatbots are considered effective because they can be used by many users simultaneously, services are always available, and they reduce reliance on manual responses.

Intent classification is an NLP task that groups text data to identify user intentions and is used to create chatbots. At first, intent classification depends on rule-based approaches, where the developer needs to manually code linguistic rules or patterns to match user questions with specific intent categories. Chatbots such as ELIZA [3] were developed using this approach and emulate conversations. The success of the rule-based approach has limitations because the resulting engine cannot handle linguistic variations, cannot capture context dependencies, and is difficult to implement in a new domain. This approach also took a lot of time and effort to develop.

The development of chatbots then utilizes artificial intelligence. Artificial intelligence chatbots can be divided into two, namely, Information Retrieval based models and Generative models. Information Retrieval-based models are designed using algorithms to have the

Received: 02-01-2025 | Accepted: 04-02-2024 | Published Online: 14-02-2024

ability to retrieve the information needed based on user requests or questions. This approach guarantees response quality because it is predefined and suitable for chatbots with specific tasks [4]. The algorithms that have been widely used recently are deep learning because of the need to understand text semantics and complex language patterns. This algorithm replaces algorithms such as Support Vector Machines (SVM) [5] and Naive Bayes [6], which can only learn discriminative features and do not pay attention to context dependencies.

Deep learning models such as Convolutional Neural Networks (CNN) [7], [8], Gated Recurrent Unit (GRU), and models with Attention Mechanism [8] have a great record of improving text classifiers' performance. The attention mechanism is one of the key components of the transformer [9] architecture. This mechanism allows the model to pay attention to the key point by giving different weights to the input, which makes it easier for the model to capture semantics and dependencies between words in sentences. Transformer architecture then emerged as a state-of-the-art approach. Furthermore, transformer-based models were developed, including Text-to-Text Transfer Transformer (T5) [10]. Bidirectional Encoder Representations from Transformer (BERT)[11], Generative Pre-trained Transformer (GPT) [12], etc. Bidirectional Encoder Representations from Transformer (BERT) is a language representation model which was introduced by Google AI. The pre-trained BERT model was later developed into several versions, such as mBERT, MalayBERT, and IndoBERT [13]. Using the same BERT architecture model trained in a variety of different languages to accommodate other language NLP tasks.

In general cases of text classification in Indonesian, fine-tuning the IndoBERT model was recorded as the best performance compared to other models [14]. Indonesian Bidirectional Encoder Representations from Transformer (IndoBERT) is a transformer-based model and is the Indonesian version of the pre-trained BERT model. IndoBERT was trained specifically using data in Indonesian, so it has the ability to understand Indonesian language structure and context compared to other models, such as standard BERT. Several previous studies have developed models for classifying intent. In a study conducted by Larson et al., various classifiers were assessed for their performance in in-scope and out-of-scope intent classification. The classifiers included FastText, SVM, CNN, DialogFlow, Rasa, MLP, and BERT. BERT outperformed the others in detecting in-scope intents, achieving an accuracy of 96%, even when the dataset had limited training examples and faced class imbalance [7] . A study conducted by Hafidz et al. developed an intelligent chatbot model that can understand various requests or questions for JAKI applications. Word embedding tried in the experiment include Word2Vec, GloVe, FastText, and IndoBERT. IndoBERT contextual embedding outperforms another embedding with an F1-score performance of 93% [15].

Despite the increasing adoption of chatbots in many settings, there is a significant research gap regarding the specific application of IndoBERT in the development of educational chatbots. While IndoBERT has demonstrated excellent performance in the general case of Indonesian text classification tasks, its potential in intent recognition within the field of education-focused chatbots remains unexplored.

This study presents an exploration of the applicability and performance of the IndoBERT model in intent classification for educational chatbots. The model understands a broad spectrum of user questions through intent, including questions related to course information, credit details, grading information and other department-specific questions. It is hoped that the concepts applied in this research can produce efficient modelling concepts that can be applied by other universities in Indonesia.

2. Research Methods

This section outlines the methodological framework adopted for developing an intent classification model for a curriculum chatbot using the IndoBERT model. Our approach utilizes IndoBERT for word embeddings to build contextual representations of the text data. The ultimate goal is to support the development of a chatbot system that enhances information access in the Statistics Study Program at Universitas Padjadjaran. The following subsections detail the dataset, data preprocessing, IndoBERT model, fine-tuning setup, and the evaluation procedures employed in this research.

2.1 Dataset

Data collection is conducted through crowdsourcing, aiming to gather ideas and suggestions from students who are potential users. We obtained open poll responses from 233 students in three academic cohorts who participated in the data collection process. We asked, if there was a chatbot that could answer questions about the curriculum, what would you want to ask? Students are allowed to write more than one question. We managed to collect 534 questions, which were then manually grouped/labeled based on 38 intents.

A crowdsourcing approach was used to ensure the questions we collect can reflect the natural variability of student input in the context of a curriculum chatbot system. This method allowed us to capture a variety of question formats, representing well-structured and more complex, informally expressed queries. The well-structured questions typically follow proper linguistic conventions, with clear grammar and concise phrasing, such as Question 1 in Table 1: “apa saja mata kuliah di prodi statistika?”. However, the dataset also contains complex inputs, including long, narrative-style questions or informal expressions, as illustrated by

Question 2 in Table 1. These variations were included to ensure that the dataset reflects real-world interactions, where students may express their questions with varying levels of formality and structure. While our data does not contain noise typical of social media (e.g., slang or excessive typos), the inclusion of these diverse formats ensures the model is robust and capable of handling the range of natural expressions found in student input.

Intent labels are formed based on questions grouped by topics in curriculum documents in our domain. The dataset experiences class imbalance, with the class exhibiting the lowest representation comprising a mere 8 data, while the class with the highest representation consisted of 61 data. So, to address the issue, we experimented with several data balancing methods. Including SMOTE [16], B-SMOTE [17], ADASYN and Augmentation. Data augmentation, specifically using EDA [18] and Back Translation was chosen because it can maintain data semantics well when creating data with new variations. We collect additional data using paraphrases as is done in research by Larson et al. In the paraphrasing process, new sentences are constructed by changing the composition and/or changing the vocabulary while retaining the original meaning of the sentence. This process maintains text semantics and enriches the dataset by introducing diverse language variations, which improve model performance. Finally, the dataset consists of 2318 data. Where each intent consists of 61 data. The data is divided into 1872 data in the training set, 214 in the validation set and 232 in the test set. In Table 1, are some examples of data and their intent.

Table 1. Data

No.	Data	Label / Intent
1.	apa saja mata kuliah di prodi statistika?	Penjelasan Semester 5
2.	apa mata kuliah elektif yang ada di prodi statistika unpad?	Penjelasan MK Elective
3.	bolehkah saya mengajukan banding untuk mengubah bobot skor jika hanya berbeda sedikit (misalnya saya ingin mengubah skor b 79 menjadi a 80)?	Penjelasan Penilaian
4.	apa saja yg dipelajari di qc2?	Penjelasan MK QC
5.	bisakah anda memberikan informasi tentang peraturan akademik yang berlaku dalam program studi ini, seperti kebijakan mengenai absensi, keterlambatan, atau kelulusan?	Penjelasan Peraturan Prodi

2.2 Overview of IndoBERT

Bidirectional Encoder Representations from Transformer (BERT) is a language representation model, which was introduced by Google AI. Indonesian Bidirectional Encoder Representations from Transformer (IndoBERT) is a transformer-based model and is the Indonesian version of the pre-trained BERT model. This model utilized an encoder from the transformer architecture and showed outstanding performance in this kind of work for various NLP tasks such as text classification, language translation and sentiment analysis. The bidirectional aspect of this model can understand the relationships between words in a sentence to build context.

Figure 1 is an illustration of the BERT-Base model architecture. In this research, the IndoBERT model used 12-layer encoder transformers, the same as the BERT-Base configuration. The feature extraction process takes place in the IndoBERT encoder to produce a processed sentence representation vector. The resulting vector will be a contextual representation of the semantics in the sentence.

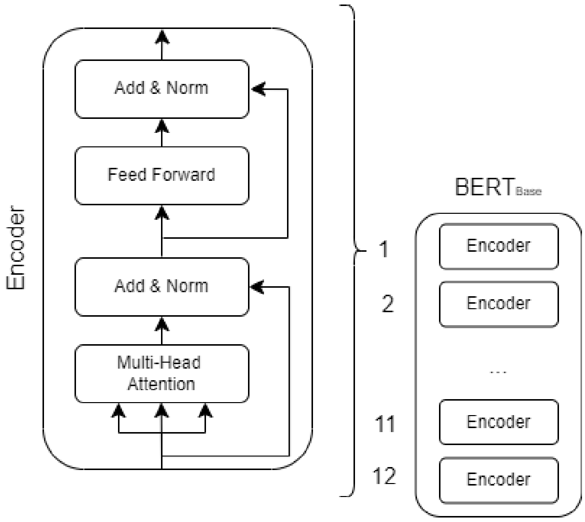

Figure 1. BERT Architecture Illustration

2.3. Data Pre-processing

We did not do much data cleaning such as removing punctuation because our dataset was relatively clean. The questions written will not contain emoji or links such as data from social media. Pre-processing using IndoBERT data does not require a case folding process and removing stop words, because the processing is uncase or words with upper and lower case letters are considered the same. Stop words are also not removed because during the process of creating word embedding the IndoBERT model will try to see the relationship between words in sentences so that the sentences entered into the model must be intact.

Table 2. Tokenization

Data	Adding Special Tokens	Tokenization Results
berapa maksimal jumlah sks yang bisa diambil mahasiswa?	[CLS] berapa maksimal jumlah sks yang bisa diambil mahasiswa ? [SEP] [PAD] [PAD] … [PAD] [PAD]	[‘[CLS]’,‘berapa’, ‘maksimal’, ‘jumlah’, ‘sks’, ‘yang’, ‘bisa’, ‘diambil’, ‘mahasiswa’, ‘?’, ‘[SEP]’, ’[PAD]’, ’[PAD]’, …, ’[PAD]’,’[PAD]’]

IndoBERT model requires a specific format so that the data can be processed, namely by adding a special token consisting of the [CLS] token at the beginning of the sentence, [SEP] as a separator between sentences and [PAD] padding tokens to handle sequence length differences. In this study, the sequence length was equalized to 218. So the amount of padding will be adjusted for each data. Table 2 presents an example of the IndoBERT tokenization process on text data.

2.4 Fine-Tuning IndoBERT

In this research, we focus on applying the pre-trained IndoBERT model to improve model performance in supervised downstream tasks, namely the intent classification task. We use IndoBERT base-p2 and include additional layers. We added a hidden layer with a number of neurons determined based on experiments to be able to adapt to the model capacity and a fully connected output layer to adapt to the multi-class classification task. We use the ReLU activation function on the additional layer neurons and the softmax activation function on the output layer neurons. The ReLU and softmax activation functions are formulated as Formulas 1 and 2.

$$f(x) = \max(0, x) = \begin{Bmatrix} 0 \ if \ x_i < 0 \\ x_i \ if \ x_i \geq 0 \end{Bmatrix} \quad (1)$$

$$\sigma(z_i) = \frac{e^{z_i}}{\sum_{j=1}^{C} e^{z_j}} \, (i = 1,2, \dots, C) \quad (2)$$

x_i is the neuron input. z_i is the logits for the i -th class.

During the process of training the model, the Adam optimizer was used. Adam is an optimization algorithm that estimates first and second-moment gradients to adjust the learning rate for each neural network weight. Adaptive learning rate enables training algorithms to monitor model performance and automatically adjust the learning rate to find the best performance. In this research, the experimental process will be carried out one at a time. We explore various hyperparameters to identify the combination that yields the highest performance. Once we determine the optimal value for one hyperparameter, we move on to the next while retaining the previously established hyperparameters.

The hyperparameters we tuned included the number of neurons, the learning rate, the number of epochs, and the batch size. Epoch is the number of repetitions of the learning process carried out by the model during training. Learning rate is a hyperparameter that is set to control weight changes during neural network parameter optimization. Batch size is the amount of input data when divided into smaller parts to shorten the computing process and simplify the pipeline.

The hyperparameter values tried are based on previous research and we try to expand the value in the experiment. In research by Simanjuntak et al about Fake News Detection, the research achieved an optimum IndoBERT model with a learning rate in the range of 10^{-5}, epoch 10, and batch size of 16 [19]. In research by Prisscilya and Girsang about False News Detection, the research achieved an optimum IndoBERT model with a learning rate in the 10^{-3} - 10^{-5} range [20]. Table 3 shows the hyperparameter values used in the experiment.

Table 3. Hyperparameter List

Hyperparameter	Value
Number of Neuron	64, 128, 512
Epoch	10, 15
Learning Rate	10^{-3}, 10^{-4}, 10^{-5}.
Batch Size	8, 16

We calculate accuracy, precision, recall and f1-score for comparative evaluation and analysis during the fine-tuning and testing stages. This metric is calculated as Formulas 3 – 6.

$$\text{Precision} = \frac{TP}{TP+FP} \times 100\% \quad (3)$$

$$\text{Recall} = \frac{TP}{TP+FN} \times 100\% \quad (4)$$

$$\text{Accuracy} = \frac{TP+TN}{TP+FP+FN+TN} \times 100\% \quad (5)$$

$$F1 - Score = 2 \times \left(\frac{precision \times recall}{precision + recall}\right) \quad (6)$$

3. Results and Discussions

3.1 Results

During the tuning process, the validation set accuracy and loss are used as a reference. Meanwhile, the training set accuracy and loss are also considered to diagnose signs of overfitting and underfitting. Table 4 presents the results of the experiment on the number of neurons used. In the experiment on the number of neurons, other hyperparameters were kept constant, including batch size 8, learning rate 0.00001, and 10 epoch. The model with 64 and 512 neurons has the highest validation set accuracy 99.52% with a very small difference. The number of neurons 64 is used for the next experiment.

Table 4. Number of Neuron Experiments

Neuron	Training Accuracy	Training Loss	Validation Accuracy	Validation Loss
64	0.9979	0.0146	0.9952	0.0153
128	0.9968	0.0204	0.9900	0.0136
512	0.9968	0.0168	0.9952	0.0275

Table 5. Learning Rate Experiment

Learning Rate	Training Accuracy	Training Loss	Validation Accuracy	Validation Loss
10^{-3}	0.0208	3.6384	0.0337	3.6383
10^{-4}	0.0267	3.6816	0.0192	3.6372
10^{-5}	0.9979	0.0146	0.9952	0.0153

Table 5 presents the results of the experiment with different learning rates. In the experiment on the learning rate, other hyperparameters were kept constant, including a number of neurons 64, batch size 8, and 10 epochs. The model with a learning rate of 0.00001 has the highest validation accuracy 99.52% which is significantly different from another model with different learning rates. Based on the graph in the training process, the learning process is stable and minimizes the loss value until it obtains 0.0146 training loss and 0.0153 validation loss which is much better than other learning rates.

Table 6 presents the results of the experiment on the epochs used. The number of neurons is set to 64, the learning rate is 0.00001, and the batch size 8 is set constant when we try different epochs. The model with 15 epochs has the highest accuracy 99.95% slightly higher than the model with 10 epochs. Furthermore, this model successfully minimizes the loss obtaining 0.0076 training loss and 0.0023 validation loss.

Table 6. Epoch Experiment

Epoch	Training Accuracy	Training Loss	Validation Accuracy	Validation Loss
10	0,9979	0,0146	0,9952	0,0153
15	0,9995	0,0076	1	0,0023

Table 7 presents the results of the experiment with different batch sizes. In the batch size experiment, we did not see any significant changes in either model performance or speeding up the computational process. The model training time with batch sizes 8 and 16 showed no difference, ranging between 18-25 minutes.

Table 7. Batch Size Experiment

Batch Size	Training Accuracy	Training Loss	Validation Accuracy	Validation Loss
8	0,9995	0,0076	1	0,0023
16	0,9989	0,0123	1	0,0041

High accuracy values and low loss indicate that there is no underfitting in the model. Meanwhile, the difference in accuracy and loss values between the training data and validation data is not significant, indicating that there is no overfitting and the model can predict new data on the validation data well.

Figures 2 and 3 are the loss and accuracy curves during model training. The optimization process runs stably.

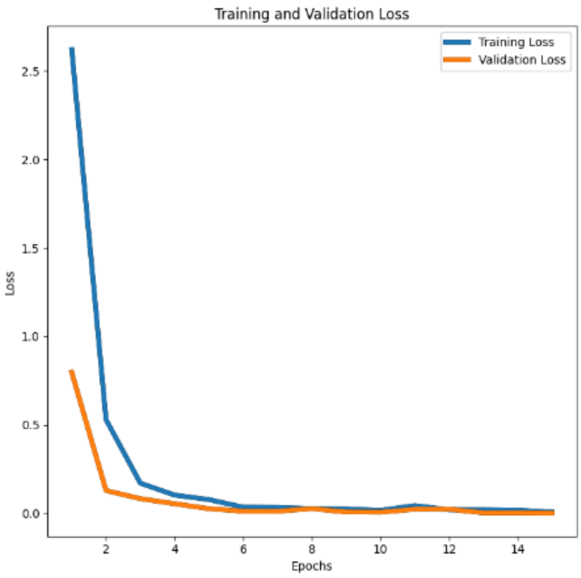

Figure 2. Loss Curve

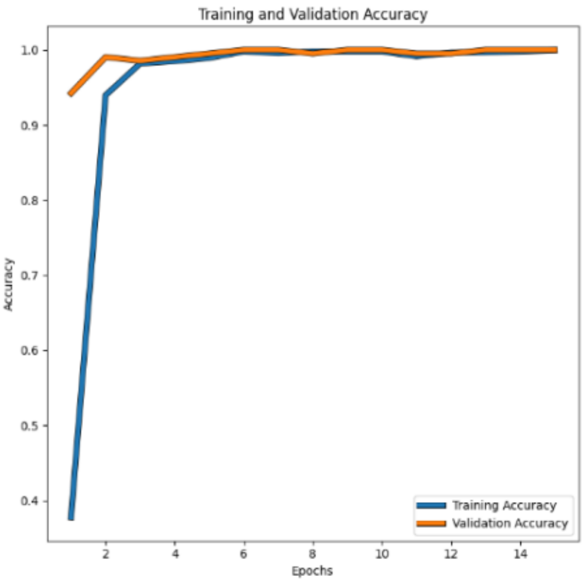

Figure 3. Accuracy Curve

3.2 Model Evaluation

The best-performing model is a model with a number of neurons 64, a learning rate of 0.00001, 15 epochs, and a batch size of 8. Then we evaluate the model with a test set. The following is a table of metric results for each intent class. The best-performing model achieved an accuracy of 98.7% on the test data. Additionally, the precision and recall scores show excellent scores to ensure the model performs well across 38 intents. With an average precision of 98%, recall of 98%, and f1-score of 98%. We succeeded in utilizing data balancing because the resulting model was not biased towards the majority class. Table 8 presents the evaluation results of the model's performance across the different intents.

There are several questions that are not classified correctly. Our investigation revealed that the data is multi-label, meaning it can belong to multiple classes due to covering two or more intent topics. This poses a limitation for our research, as the model assumes each question belongs to only one intent label. In this case, the model forces the intent prediction into one class and relationships between labels are unfortunately ignored. Multi-label prediction information is valuable information so that the chatbot can provide the best response possible. Table 9 contains examples of data that were misclassified by the model.

The model is limited to detecting intent within the specific curriculum scope, similar to the data. It currently cannot identify requests that fall outside the service's scope, known as out-of-scope cases. The questions in the dataset only involve questions about the curriculum in the Statistics Study Program at Universitas Padjadjaran. Due to the dataset's domain-specific nature, any out-of-scope questions will be classified by the model into the most probable relevant category. Detecting out-of-scope cases is crucial because it can lead to confusion and chatbot can't provide relevant responses.

Table 8. Evaluation Metric

Intent	Acc	Prec	Recall	F1-Score
List Mata Kuliah	0.987	1	1	1
Penjelasan Minor &Profesi		1	1	1
Penjelasan Administrasi		1	1	1
Penjelasan Bidang Aktuaria		1	1	1
Penjelasan Buku Kuliah		1	0.83	0.91
Penjelasan Cumlaude		1	1	1
Penjelasan Dosen MK		1	1	1
Penjelasan Fast Track		1	1	1
Kampus Merdeka		1	0.83	0.91
Penjelasan Ketentuan SKS		0.86	1	0.92
Penjelasan Konversi SKS		1	1	1
Penjelasan KRS		1	1	1
Penjelasan Kurikulum		1	1	1
Penjelasan Magang		1	1	1
Penjelasan MK ADM		1	1	1
Penjelasan MK Ekonometrika		1	1	1
Penjelasan MK Elective		1	1	1
Penjelasan MK Kalkulus		1	1	1
Penjelasan MK MSNP		1	1	1
Penjelasan MK Prasyarat		1	1	1
Penjelasan MK QC		1	1	1
Penjelasan MK Spatial		1	1	1
Penjelasan MK Teos		1	1	1
Penjelasan MK Wajib		0.86	1	0.92
Penjelasan Penilaian		1	1	1
Penjelasan Peraturan		1	1	1
Penjelasan Semester 1		1	1	1
Penjelasan Semester 2		1	1	1
Penjelasan Semester 3		1	1	1
Penjelasan Semester 4		1	1	1
Penjelasan Semester 5		1	1	1
Penjelasan Semester 6		0.86	1	0.92
Penjelasan Semester 7		1	1	1
Penjelasan Semester 8		1	1	1
Penjelasan Skripsi		1	1	1
Penjelasan Syarat Lulus		1	0.83	0.91
Penjelasan UTS UAS		1	1	1
Rincian Jadwal		1	1	1
Avg.		0.98	0.98	0.98

Table 9. Examples of Misclassified Questions

No.	Question	Actual Intent	Predicted
1.	berapa batas minimum dan maksimum total kredit dalam gelar statistik?	Penjelasan Syarat Kelulusan	Penjelasan Ketentuan SKS
2.	bagaimana ketentuan untuk mengikuti msib di semester 6?	Penjelasan Kampus Merdeka	Penjelasan Semester 6

4. Conclusions

Based on intent classification for curriculum chatbots model, fine-tuned IndoBERT, achieving optimal hyperparameters with a batch size of 8, a learning rate of 0.00001, and 64 hidden layer neurons. The model was trained for 15 epochs and reached 98.7% of accuracy on the test data. Its precision and recall scores indicate excellent performance across 38 intents, showcasing its superior ability to understand the context of students' questions. The model's ability to understand the context of students questions is superior and the training process for intent classification tasks is relatively short. The training process for intent classification took less than 25 minutes. This research demonstrates the advantages of using the IndoBERT model for intent classification in an academic chatbot, achieving high accuracy performance. To generalize this approach and adapt it to other domains, you can fine-tune IndoBERT on intent data specific to your domain. This enables the model to accurately handle domain-specific queries. Future work could focus on several key areas to address current limitations and enhance the chatbot's performance. Expanding the model to effectively handle out-of-scope cases would ensure it can manage unexpected input. Furthermore, incorporating multi-label classification models could improve the chatbot's reliability by allowing it to identify and address multiple intents within a single input. These advancements would significantly contribute to the chatbot's ability to provide accurate and relevant answers.

Acknowledgements

This work was supported by Universitas Padjadjaran Research Grant No. 1698/UN6.3.1/PT.00/2024 and the Research Center for Artificial Intelligence and Big Data Universitas Padjadjaran.

References

[1] C. W. Okonkwo and A. Ade-Ibijola, “Chatbots applications in education: A systematic review,” *Computers and Education: Artificial Intelligence*, vol. 2, p. 100033, 2021, doi: 10.1016/j.caeai.2021.100033.

[2] S. Mohamad Suhaili, N. Salim, and M. N. Jambli, “Service chatbots: A systematic review,” *Expert Systems with Applications*, vol. 184, p. 115461, Dec. 2021, doi: 10.1016/j.eswa.2021.115461.

[3] J. Weizenbaum, “ELIZA—a computer program for the study of natural language communication between man and machine,” *Commun. ACM*, vol. 9, no. 1, pp. 36–45, Jan. 1966, doi: 10.1145/365153.365168.

[4] G. Caldarini, S. Jaf, and K. McGarry, “A Literature Survey of Recent Advances in Chatbots,” 2022, doi: 10.48550/ARXIV.2201.06657.

[5] J. Schuurmans and F. Frasincar, “Intent Classification for Dialogue Utterances,” *IEEE Intell. Syst.*, vol. 35, no. 1, pp. 82–88, Jan. 2020, doi: 10.1109/MIS.2019.2954966.

[6] M. Y. Helmi Setyawan, R. M. Awangga, and S. R. Efendi, “Comparison Of Multinomial Naive Bayes Algorithm And Logistic Regression For Intent Classification In Chatbot,” in *2018 International Conference on Applied Engineering (ICAE)*, Batam: IEEE, Oct. 2018, pp. 1–5. doi: 10.1109/INCAE.2018.8579372.

[7] S. Larson *et al.*, “An Evaluation Dataset for Intent Classification and Out-of-Scope Prediction,” in *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP)*, Hong Kong, China: Association for Computational Linguistics, 2019, pp. 1311–1316. doi: 10.18653/v1/D19-1131.

[8] A. Cohan, W. Ammar, M. Van Zuylen, and F. Cady, “Structural Scaffolds for Citation Intent Classification in Scientific Publications,” in *Proceedings of the 2019 Conference of the North*, Minneapolis, Minnesota: Association for Computational Linguistics, 2019, pp. 3586–3596. doi: 10.18653/v1/N19-1361.

[9] A. Vaswani *et al.*, "Attention Is All You Need," Aug. 01, 2023, *arXiv*: arXiv:1706.03762. Accessed: Jul. 14, 2024. [Online]. Available: http://arxiv.org/abs/1706.03762

[10] C. Raffel *et al.*, "Exploring the Limits of Transfer Learning with a Unified Text-to-Text Transformer," Sep. 19, 2023, *arXiv*: arXiv:1910.10683. Accessed: May 02, 2024. [Online]. Available: http://arxiv.org/abs/1910.10683

[11] J. Devlin, M.-W. Chang, K. Lee, and K. Toutanova, "BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding," May 24, 2019, *arXiv*: arXiv:1810.04805. Accessed: May 02, 2024. [Online]. Available: http://arxiv.org/abs/1810.04805

[12] T. B. Brown *et al.*, "Language Models are Few-Shot Learners," Jul. 22, 2020, *arXiv*: arXiv:2005.14165. Accessed: May 02, 2024. [Online]. Available: http://arxiv.org/abs/2005.14165

[13] F. Koto, "IndoLEM and IndoBERT: A Benchmark Dataset and Pre-trained Language Model for Indonesian NLP," *COLING 2020 - 28th International Conference on Computational Linguistics, Proceedings of the Conference*. pp. 757–770, 2020.

[14] K. S. Nugroho, A. Y. Sukmadewa, H. W. DW, F. A. Bachtiar, and N. Yudistira, "BERT Fine-Tuning for Sentiment Analysis on Indonesian Mobile Apps Reviews," in *6th International Conference on Sustainable Information Engineering and Technology 2021*, Sep. 2021, pp. 258–264. doi: 10.1145/3479645.3479679.

[15] A. Fadhlurohman, "Development of Indonesian Language Intelligent Chatbot for Public Services in JAKI Application," *Proceedings of 2023 IEEE International Smart Cities Conference, ISC2 2023*. 2023. doi: 10.1109/ISC257844.2023.10293514.

[16] N. V. Chawla, K. W. Bowyer, L. O. Hall, and W. P. Kegelmeyer, "SMOTE: Synthetic Minority Over-sampling Technique," 2011, doi: 10.48550/ARXIV.1106.1813.

[17] H. Han, W. Wang, and B. Mao, "Borderline-SMOTE: A New Over-Sampling Method in Imbalanced Data Sets Learning," in *International Conference on Intelligent Computing*, 2005. [Online]. Available: https://api.semanticscholar.org/CorpusID:12126950

[18] J. Wei and K. Zou, "EDA: Easy Data Augmentation Techniques for Boosting Performance on Text Classification Tasks," 2019, doi: 10.48550/ARXIV.1901.11196.

[19] A. Simanjuntak *et al.*, "Studi dan Analisis Hyperparameter Tuning IndoBERT Dalam Pendeteksian Berita Palsu."

[20] V. Prisscilya and A. S. Girsang, "Classification of Indonesia False News Detection Using Bertopic and Indobert," *jist*, vol. 5, no. 8, pp. 3061–3079, Aug. 2024, doi: 10.59141/jist.v5i8.1310.