Journal of Theoretical and Applied Information Technology
15th July 2023. Vol.101. No 13

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

UTILIZING TRANSFORMER-BASED DEEP LEARNING FOR INTENT CLASSIFICATION ON TEXT

[1]RICHARD SIMARMATA, [2]JONATHAN KRISTANTO, [2]ANDRY CHOWANDA

[1]Computer Science Department, BINUS Graduate Program - Master of Computer Science, Bina Nusantara University, Jakarta 11480 Indonesia

[2]Computer Science Department, School of Computer Science, Bina Nusantara University, Jakarta 11480 Indonesia

E-mail: Richard.Simarmata@binus.ac.id, jonathan.kristanto@binus.ac.id, achowanda@binus.edu

ABSTRACT

Today, communication is carried out not only between humans but also between machines. Communication between humans and machines is no longer limited to using tools such as buttons or voice commands, but humans can communicate dyadic or two-way communication with computers. This can be done because of the field of Natural Language Processing (NLP). There are several problems exists in building a system that can naturally understand and communicate with human. In communication, there are a number of slang words and local contexts or expressions that might change the meaning of the word. In the communication, there is always an intention of the speaker to the interlocutor. This can improve the machine's understanding of human words and communicate with humans naturally. This research aims to explore and improve intent classification using deep learning technique. In this study, several models of deep learning based on Transformer are proposed that have the best performance in the field of text intention classification, such as Bidirectional Encoder Representations from Transformer and XLNet. There is also another approach for comparison by combining BERT Embedding with the Long-Short Term Memory model. The data used to train the model is out of scope obtained through a Kaggle website with a label intention of 50. By using metrics such as F1-score, recall, precision, accuracy, and top K accuracy, it can be concluded that the BERT model has good results and is the best compared to other models.

Keywords: *Text Classification, Intent Detection, Deep Learning, Transformer Architecture*

1. INTRODUCTION

The importance of communication for human activities in organizational life and social life. By communicating, humans can relate to each other. If there is no communication, Cooperation becomes impossible because people cannot communicate their needs or desires or even convey their feelings to others. Interpersonal communication is communication that occurs between two people who are interacting, and this communication is called dyadic communication [1, 18, 19]. In his book, DeVito defines that dyadic communication as dialogical in nature by involving two people. Like the communication that is done by mother and child, with the intent of the purpose that has been prepared. In the current industrial 4.0 era, many communication applications are integrated with technology such as AI (Artificial Intelligence) based on big data analytics and incorporate the human element into it [2]. Current technologies allow us, human to communicate with machine (e.g., computer, mobile phones or any smart electronic devises).

This application can be seen from the development of how to control IoT, smart homes, chatbots, and several other technologies. Chatbot, which is also one of the applications carried out in the integration of communication between technology and humans, is a tool that is often used as a new technology in the business world [3]. In 2018 a new language representation model called BERT was introduced, derived from the abbreviation Bidirectional Encoder Representations from Transformer [4]. BERT can teach machines to understand more formal and informal human language, called Natural Language Understanding. However, it is not a trivial task to build a model of natural language that can give machine the ability to

communicate with us naturally. There are a number of slang words and local contexts or expressions that might change the meaning of the word, in a dyadic communication. Several techniques [18-21] can be implemented to provide the communication ability to machines, one of them is by recognizing the intention from a utterance. By recognizing the interlocutor's intention, the machine can choose a communication strategy in the interaction process.

From the development of the Eliza Chatbot [5], which only uses a simple algorithm for the NLP and Deep Learning approaches [18-21], this research is carried out to find out which approach is the best among the existing models to get the best Natural Language Understanding model in the classification of sentence intentions. The dataset used in this study came from Kaggle with the name clinc oos-eval. The rest of the paper is arranged as follows: Recent work development are discussed in the next section (Recent Work). The methods and learning algorithms proposed in this research are comprehensively described in The Learning Algorithms section. Moreover, the benchmark to other existing algorithms is presented in the Algorithm Benchmark section. The results then are comprehensively presented in Experiment Results section. Finally, the final section of this research discuss the conclusion of this research.

2. RECENT WORK

The distance between machines and humans is getting closer with NLP, and this technique allows humans to communicate with machines easily [6]. In recent years more and more technologies have been used NLP, a technology to help humans. These technologies help people's daily lives through voice or text. The research was also carried out to develop NLP technology to find a better model with higher accuracy. One of the latest studies recommends pre-diagnosis medical assistance based on CNN, RNN and a combination of the two [7]. This study uses a crawler or web searcher on a website, namely ask.39.net; each page has 32 questions, and the web searcher took as many as 228,994 questions. This study uses CNN, RNN, CRNN, and RCNN, with the evaluation results CNN getting 86.68% accuracy, while RNN is getting 81.58% accuracy, then RCNN with 87.38% accuracy, after that RCNN with 87.38% accuracy and the last Improved CRNN with 88.63% accuracy. These results show that the Improved CRNN model gets the best performance in the classification compared to the others. Another study also uses NLP to create chatbot applications using Deep Bidirectional Transformers or BERT [8]. This study uses the current state-of-the-art, Transformers as its architecture, intending to create a chatbot financial service. BERT will be compared with other models, namely LSTM, XLNet, Logistic Regression and Xgboost and Naïve Bayes. The feature extraction used is Google Embedding, SharePoint Embeddings, Word2vec, and TF-IDF. The dataset is divided into five intentions: Account maintenance with 9,074 data, Account Permission with 2,961 data, Transfer of Assets with 2,838 data, Banking with 4,788 data, and Tax FAQ with 2,969 data). The results of Performance are BERT small with Sharepoint Embeddings got 94.4% accuracy performance, followed by BERT Small with Google Embeddings with 94.9% accuracy, BERT large and Google Embeddings with 95.4% accuracy, XLNet Large and Google Embeddings 92.7% accuracy, LSTM with attention and Word2Vec 91.3% accuracy, LSTM and Word2Vec was 89.2% accuracy, Logistic Regression and TF-IDF 82% accuracy, Xgboost and TF-IDF 76% accuracy. Finally, Naïve Bayes with TF-IDF got 66.1% accuracy. From this research, BERT with Google Embeddings got the highest accuracy.

Research is also conducted to get the classification of sentiment towards the discovery of topics in Novel Coronavirus or COVID-19 in online discussions using the LSTM Recurrent Neural Network [9]. The model will be compared with the classification of classical machine learning models, namely SVM, Naïve Bayes, Logistic Regression and KNN. The dataset used is comments obtained from Reddit from January 20, 2020, to March 2020, with as many as 563,079 comments. By using Latent Dirichlet Allocation (LDA), the results of this study get SVM accuracy of 77.78%, Naïve Bayes 72.38%, Logistic Regression of 78.72%, KNN 56.18% and LSTM Recurrent Neural Network 81.15%. From these results, the proposed model gets the highest accuracy. Research also focuses on Indonesia NLP using IndoLEM and IndoBERT [10]. The approach in this study is based on the BERT model carried out pre-trained with an Indonesian language dataset. One of the datasets used is POS Tagging. The evaluation results get the accuracy obtained by the indoBERT model of 96.8%.

On the other hand, research has also been developed using BERT as a basic model and conducting pre-training to become an NLP for Indonesia called indoNLU [11]. This study uses 4 billion datasets collected from various places such as social media, blogs, websites and news. The results

of the pre-training resulted in the highest accuracy of 95.0%. Other studies compare the Transformer model with other classification models such as SVM, Bi-LSTM, SVM, and CNN [12]. Various dataset topics are used, one of which is very good to mention is the Health topic with 57,938 data; the F1-score results obtained from the dataset are Logistic Regression 69.7%, SVM 73%, Bi-LSTM 70.0%, CNN 72.4%, mBERT 78.0%, and IndoNLU 80.7%. It can be seen from the evaluation results that IndoNLU has a more significant F Score level compared to other models. The summary of the results of the literature review can be seen in table 1.

Table 1: The summary of the literature review.

Author	Model	Features	Best Result
[7]	Improved CRNN	CNN	Accuracy 88.63%
[8]	BERT Large	Google Embeddings	Accuracy 95.4%
[9]	LSTM, RNN	LDA	Accuracy 81.15%
[10]	Transformer BERT	IndoBERT	Accuracy 96.8%
[11]	Transformer IndoNLU	BERT Large	Accuracy 95.69%
[12]	IndoNLU	IndoNLU	F1-score 80.7%

3. THE LEARNING ALGORITHMS

3.1 METHODOLOGY

Figure 1 illustrates the proposed methodology in this research. The first step is to explore several promising deep learning architectures and algorithms to model the intent classification from a utterance, from the literature and existing work in this research area. From the exploration, BERT (Bidirectional Encoder Representations from Transformers), LSTM (Long short-term memory) and XLNET architecture are quite promising to model intent classification from a utterance. Which lead to the next step, training step, where those three algorithms and the combination of BERT and LSTM are implemented to model intent classification from a utterance. The detail of the proposed architectures is presented in the next sub-section. The dataset used in this study came from Kaggle with the name clinc oos-eval. Finally the models are evaluate using several performance metrics such as F1-score, recall, precision, accuracy, and top K accuracy.

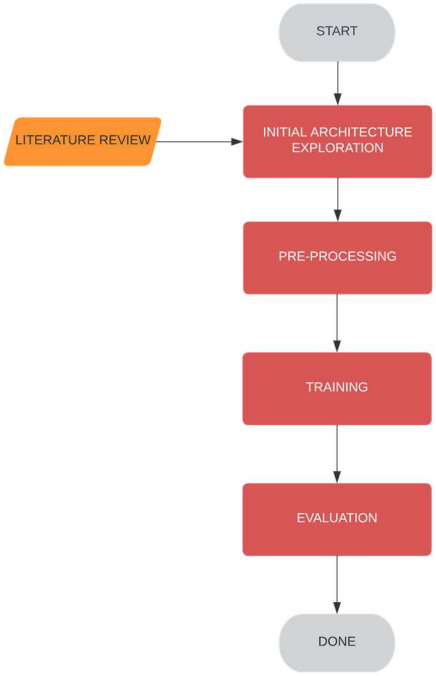

Figure 1: Proposed Methodology

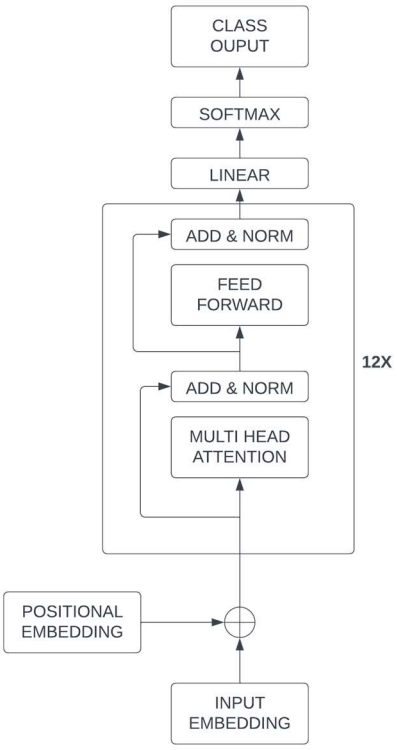

Figure 2: Proposed BERT Architecture

3.2 BERT

In this case, a previously trained BERT model with a large dataset will be used in previous studies. Figure 2 illustrates the proposed general BERT architecture that is implemented in this research. Due to this need, [13] in his research found that fine-tuning can improve the performance of a model. Therefore, the model variant that will be used is bert-base-uncased which has been previously trained using English with 12 layers, 768 hidden, 12 heads, and 110 million parameters. Another variant of BERT is called DistilBERT. This model is also used in this study, where the study [14] found another approach that is smaller, faster, cheaper and lighter, with the difference that DistilBERT only uses 66 million parameters. The last variant of BERT is ALBERT which in his research [15] concluded this variant for self-supervised learning. ALBERT also provides another approach, namely cross-layer parameter sharing, to increase efficiency parameters.

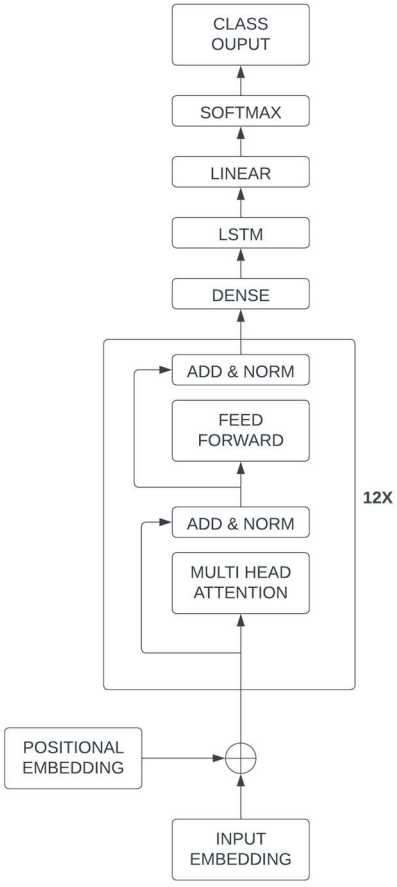

Figure 3: Proposed BERT + LSTM Architecture

3.3 BERT Embedding and LSTM

Figure 3 illustrates the proposed combined BERT and LSTM architecture to improve the modeling performances. By combining pre-trained from BERT and using the LSTM Model. This is an approach to see the difference between transformer-based and non-transformer-based models. Using a pre-trained BERT just like any other BERT model, this model will use a base-uncased BERT. After being pre-trained, the model will be trained with LSTM and will be validated in the same way as the BERT method.

3.4 XLNet

XLNet is a model that combines the concepts of the autoregressive language model and Transformer-XL into pretraining [16]. XLNet introduces a new approach known as Permutation Language Modeling. This approach combines the concept of an autoregressive language model to process information in a directed manner with the ability of an autoencoding language model to process the context of a sentence in a bidirectional manner. XLNet performs permutations that can create a new set of sentences by randomizing the order of words in the sentence, this is different from BERT's Masked Language Modeling (MLM), where the BERT model will cover 15% of a sentence after that prediction is made in the hope of helping the model study context. This permutation allows XLNet to learn the context of a sentence in a bidirectional manner, while maintaining the dependency between tokens that BERT cannot do.

4. ALGORITHM BENCHMARK

In this section, we describe how we benchmark the algorithms in section 3.

4.1 Dataset

In 2019 a collection of research data was created to evaluate a classification intention model [17]. The number of data is 18,025 sentences, with 150 categories of intention. The focus of this dataset is conversations, questions, and statements. In this dataset, several other datasets are also divided to evaluate the model if the sentence is out of context or the categorical intention. One example of a sentence is "how are my sports team doing" while the model being trained does not have a "sport"

category or label, for example, so that the model cannot find its intentions. He will indicate that the sentence is outside the context of the capacity of the 150 predefined intentions. However, because this study only focuses on finding the best model for classification intentions, it was decided to only use data that has labels without using the Out-of-Scope label at all. Table 2 shows for example some of the sentences and some of the intentions contained in this dataset. In this study, several experimental scenarios were tried to determine the best classifier model. The dataset is divided into 3 parts, namely the training set, validation set, and test set.

Table 2: Sample of The Oos-Val Datasets

Domain	Intent	Text
Banking	Transfer	Move 100 dollar from my savings to my checking
Work	PTO Request	Let me know how to make a vacation request
Travel	Travel Suggestion	What site are there to see when in evans
Home	To Do List Update	Nuke all items on my todo list
Utility	Text	Send a text to mom saying I'm on my way
Credit Cards	Rewards	How high are the rewards of my discover card

4.2 Experiment Setup

For training we use Google Colab Pro, with single GPU of NVIDIA P100 / T4 up to 15GB and RAM Up to 25 GB. for pre-trained language model, we use BERT base model, ALBERT, and DistilBERT with Transformer algorithm. and for XLNet we use XLNet base model. After that we use BERT base model with LSTM as the algorithm. We used AdamW Optimizer with the learning rate of 1e-5. The batch size is set to 4, 10 of epoch and max Len is 33 for all datasets and methods to ensure fairness in the experiment.

4.3 Evaluation Matrix

From the results of the output model that has been trained and fine-tuned, performance measurement will be carried out using the Confusion matrix, and will use several metrics including accuracy, precision, recall, F1-score to measure the evaluation performance of all tasks, and Top K Categorical Accuracy from every model.

5. EXPERIMENT RESULT

Table 3: Score Model in Validation Set

MODEL	ACC	PRE	REC	F1	K5
BERT	94.7	95.0	94.70	94.63	98.44
DistilBERT	93.37	93.38	93.37	93.29	98.66
ALBERT	92.94	92.17	92.08	92.17	98.20
BERT WITH LSTM	94.47	94.87	94.47	94.39	97.10
XLNet	84.73	87.49	84.73	82.78	97.80

Table 3 demonstrates the results of the models in validation set. BERT has the best results in terms of classification intention using the validation data provided. The accuracy achieved by BERT is 94.70%, precision is 95.08%, recall is 94.70%, F1-score is 94.63%, TopK-5 Accuracy is 98.44%. Very thinly followed by a combination of BERT Embedding with LSTM with an accuracy of 94.47%, precision 94.87%, recall 94.47%, F1-score 94.39%, and TopK-5 Accuracy 97.10%. Although other models got unsatisfactory results, DistilBERT got a TopK-5 Accuracy of 98.66% with the highest result. After that, each model was further reviewed using a test dataset in order to better describe the results with data that were not trained.

Table 4: Score Model in Test Set

MODEL	ACC	PRE	REC	F1	K5
BERT	94.29	94.72	94.29	94.24	98.44
DistilBERT	92.58	92.99	92.58	92.45	98.62
ALBERT	91.91	92.46	91.91	92.17	98.28
BERT WITH LSTM	94.13	94.38	94.13	94.07	97.55
XLNet	83.84	85.86	83.84	81.77	97.50

Table 4 demonstrates the results of the models in test set BERT still outperformed other models with 94.29% accuracy, 94.72% precision, 94.29% recall, 94.24% F1-score and 98.44% TopK-5 Accuracy. The best TopK-5 Accuracy is still superior to the DistilBERT model of 0.22% with a TopK-5 Accuracy value of 98.62% compared to the BERT model of 98.44%. Thus, it can be said that BERT is the best model in carrying out classification intentions from the data that has been provided.

6. CONCLUSION

This research was conducted and made to find the best model and method for carrying out classification intentions using the text data that has been provided. BERT has the best results in terms of classification intention using the validation data

provided. The accuracy achieved by BERT is 94.70%, precision is 95.08%, recall is 94.70%, F1-score is 94.63%, TopK-5 Accuracy is 98.44%. On the other hand, the Transformer architecture gets better result that the LSTM architecture in carry out classification intentions with English text-based data. for XLNET Model get lower result than other models, so it is concluded that XLNet does not significantly increase the accuracy of classification intention. More than that, Transformer-based model with DistilBERT method approach has the highest Top-K accuracy rate and low cost.

Considering that there is still not much research done for Indonesian language-based classification intentions using text and taking into account the rapid development of technology and the need for the ability of machines to communicate with humans, the researcher suggests for further research using Indonesian-based data and using the models that have been made in the study. previously, namely IndoBERT and IndoNLU

REFERENCES:

[1] J. A. Devito, "Komunikasi Antarmanusia," *Komunikasi Antarmanusia. Kuliah Dasar*, 2011.

[2] M. Astuti, Rani Natadian; Fatchan, "Perancangan Aplikasi Teknologi Chatbot Untuk INDUSTRI KOMERSIAL 4.0," *Prosiding Seminar Nasional Teknologi dan Sains (SNasTekS)*, vol. 1, no. September, pp. 339–348, 2019.

[3] M. Heo and K. J. Lee, "Chatbot as a New Business Communication Tool: The Case of Naver TalkTalk," *Business Communication Research and Practice*, vol. 1, no. 1, pp. 41–45, 2018, doi: 10.22682/bcrp.2018.1.1.41.

[4] J. Devlin, M. W. Chang, K. Lee, and K. Toutanova, "BERT: Pre-training of deep bidirectional transformers for language understanding," *NAACL HLT 2019 - 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies - Proceedings of the Conference*, vol. 1, no. Mlm, pp. 4171–4186, 2019.

[5] J. Weizenbaum, "ELIZA—A Computer Program For the Study of Natural Language Communication Between Man And Machine," *Commun ACM*, vol. 26, no. 1, pp. 23–28, 1983, doi: 10.1145/357980.357991.

[6] M. C. Surabhi, "Natural language processing future," *2013 International Conference on Optical Imaging Sensor and Security, ICOSS 2013*, pp. 3–5, 2013, doi: 10.1109/ICOISS.2013.6678407.

[7] X. Zhou, Y. Li, and W. Liang, "CNN-RNN Based Intelligent Recommendation for Online Medical Pre-Diagnosis Support," *IEEE/ACM Trans Comput Biol Bioinform*, pp. 1–1, 2020, doi: 10.1109/tcbb.2020.2994780.

[8] S. Yu, Y. Chen, and H. Zaidi, "AVA: A Financial Service Chatbot based on Deep Bidirectional Transformers," *ArXiv*, 2020.

[9] H. Jelodar, Y. Wang, R. Orji, and H. Huang, "Deep sentiment classification and topic discovery on novel coronavirus or COVID-19 online discussions: NLP using LSTM recurrent neural network approach," *ArXiv*, vol. 24, no. 10, pp. 2733–2742, 2020.

[10] F. Koto, A. Rahimi, J. H. Lau, and T. Baldwin, "IndoLEM and IndoBERT: A Benchmark Dataset and Pre-trained Language Model for Indonesian NLP," pp. 757–770, 2021, doi: 10.18653/v1/2020.coling-main.66.

[11] B. Wilie *et al.*, "IndoNLU: Benchmark and resources for evaluating indonesian natural language understanding," *ArXiv*, 2020.

[12] M. N. Nityasya, H. A. Wibowo, R. E. Prasojo, and A. F. Aji, "Costs to Consider in Adopting NLP for Your Business," pp. 1–12, 2020, [Online]. Available: http://arxiv.org/abs/2012.08958

[13] J. Howard and S. Ruder, "Universal language model fine-tuning for text classification," *ACL 2018 - 56th Annual Meeting of the Association for Computational Linguistics, Proceedings of the Conference (Long Papers)*, vol. 1, pp. 328–339, 2018, doi: 10.18653/v1/p18-1031.

[14] V. Sanh, L. Debut, J. Chaumond, and T. Wolf, "DistilBERT, a distilled version of BERT: smaller, faster, cheaper and lighter," pp. 2–6, 2019, [Online]. Available: http://arxiv.org/abs/1910.01108

[15] Z. Lan, M. Chen, S. Goodman, K. Gimpel, P. Sharma, and R. Soricut, "ALBERT: A Lite BERT for Self-supervised Learning of Language Representations," pp. 1–17, 2019, [Online]. Available: http://arxiv.org/abs/1909.11942

[16] Z. Yang, Z. Dai, Y. Yang, J. Carbonell, R. Salakhutdinov, and Q. v. Le, "XLNet: Generalized autoregressive pretraining for

language understanding," *Adv Neural Inf Process Syst*, vol. 32, no. NeurIPS, pp. 1–18, 2019.

[17] S. Larson *et al.*, "An evaluation dataset for intent classification and out-of-scope prediction," *EMNLP-IJCNLP 2019 - 2019 Conference on Empirical Methods in Natural Language Processing and 9th International Joint Conference on Natural Language Processing, Proceedings of the Conference*, pp. 1311–1316, 2020, doi: 10.18653/v1/d19-1131.

[18] Chowanda, A., & Chowanda, A. D. (2017). Recurrent neural network to deep learn conversation in indonesian. Procedia computer science, 116, 579-586.

[19] Sutoyo, R., Chowanda, A., Kurniati, A., & Wongso, R. (2019). Designing an emotionally realistic chatbot framework to enhance its believability with AIML and information states. Procedia Computer Science, 157, 621-628.

[20] Chowanda, A., Sutoyo, R., & Tanachutiwat, S. (2021). Exploring text-based emotions recognition machine learning techniques on social media conversation. Procedia Computer Science, 179, 821-828.

[21] Zhu, W., Chowanda, A., & Valstar, M. (2016). Topic switch models for dialogue management in virtual humans. In Intelligent Virtual Agents: 16th International Conference, IVA 2016, Los Angeles, CA, USA, September 20–23, 2016, Proceedings 16 (pp. 407-411). Springer International Publishing.