

Development Of A Chatbot For The Online Application Telegram Chat With An Approach To The Emotion Classification Text Using The Indobert-Lite Method

Khodijah Hullyyah
Department of Informatics
Engineering, Faculty of Science and
Technology
Syarif Hidayatullah State Islamic
University Jakarta
Jakarta, Indonesia
khodijah.hullyyah@uinjkt.ac.id

Faishal Rayyan
Department of Informatics
Engineering, Faculty of Science and
Technology
Syarif Hidayatullah State Islamic
University Jakarta
Jakarta, Indonesia
faishal.rayyan18@mhs.uinjkt.ac.id

Normi Sham Awang Abu Bakar
Computer Science Department,
KICT International
Islamic University Malaysia (IUM)
Gombak, Malaysia
nsham@ium.edu.my

Abstract—The increasing preference for text-based communication on online chat applications has caused the number of social interactions to increase rapidly. However, text-based communication usually results in misunderstandings resulting from the absence of feeling intonation and emotions in the text. This study aims to create a chatbot that can detect emotions text to be entered into online chat applications. This study used a pre-trained model specifically trained from a collection of Indonesian-language datasets, namely IndoBERT-lite. The dataset used to train the model is a collection of Indonesian tweets totaling 4,403 which have been labeled with 5 classes of emotions, namely love, happy, anger, sadness, and fear. The hyperparameters used in this study to train the model were 5 epochs, batch size 16, learning rate 0.000003, and adam optimizer. Based on the test results with the parameters already mentioned, the accuracy, F1 score, recall, and precision values were obtained in the training set of 89%, 89%, 89%, and 90%, while the validation set obtained 70%, 71%, 70%, and 72%.

Keywords— Online chat application, emotion classification, chatbot, pre-trained model, IndoBERT-lite, Telegram.

I. INTRODUCTION

A. Background

The use of online chat applications is commonly used today. Even in Indonesia, based on the results of an APJII (Indonesian Internet Service Providers Association) survey in 2017 of 2,500 respondents who spread throughout Indonesia, the most popular type of internet service application is instant messaging or chatting [1]. There are a lot of online chat applications on the internet. Among the online chat applications commonly used in Indonesia are Whatsapp and Telegram, which have 88.7% and 62.8%, respectively, as of February 2022 [2].

The Most Used Social Media in Indonesia as of
February 2022
(Percentage of Internet Users from Age 16 – 64 Years)

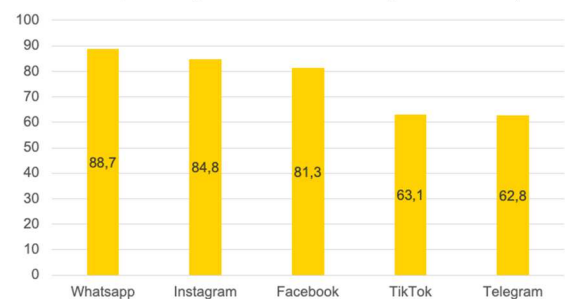


Fig. 1. The Most Popular Social Media in Indonesia

The increasing preference for text-based communication on online chat applications has caused the number of social interactions to increase rapidly, but decreased the quality of the interaction itself which can damage the relationship [3]. Reports from individuals also claim that text-based communication results in many miscommunications compared to face-to-face communication [4] [5]. The combination of a large dependence with text-based communication plus frequent miscommunication can damage relationships because individuals often experience negative feelings when misunderstandings occur [6].

Therefore, in order to reduce conflicts that occur in online chat applications caused by misunderstandings in reading emotions from text messages, it is necessary to create a machine that can classify a person's emotions through text messages. The machine is an artificial intelligence (AI) - based autoresponder message that notifies users of online chat applications of emotions from a text such as a chatbot. Chatbots will be useful as additional functions installed in online chat applications in various fields, so that communication can take place more interactively, relevantly, and avoid misunderstandings.

B. Methods

The AIM etode used in this study is IndoBERT because this metode is the first BERT pre-trained model to be trained using a collection of Indonesian- language datasets and produces a high F 1 score on the task of emotional classification in Indonesian text [7].

C. Hardware

There are 2 types of hardware used in this study, namely hardware based and those that are not. Details of the hardware used include

TABLE I. NON CLOUD HARDWARE SPECIFICATIONS

Device	Lenovo B40
Processor	Intel(R) Core(TM) i3-4005U CPU @ 1.70GHz 1.70 GHz
Hard Disk Drive	238 GB
RAM	4 GB

TABLE II. CLOUD HARDWARE SPECIFICATIONS

Device	Colaboratory: Python 3 Google Compute Engine
Processor	Intel(R) Xeon(R) CPU @2.20GHz
Memory	12.23 GB
GPU	NVIDIA Tesla T4
VRAM	13.62 GB

D. Purpose

This study aims to create an application that can detect IndoBERT-lite-based emotions in the Telegram online chat application as an effort to avoid misunderstandings when communicating using text on Telegram which can cause conflicts and damage relationships.

II. LITERATURE STUDIES

This section will discuss the literature used as well as its differences with this study.

Paper [7] has strengthened the first Indonesian- language natural language understanding benchmark called IndoNLU. In addition, this study also collected Indonesian-language datasets from various sources called Indo4B. Indo4B was used to train the first Indonesian version of the BERT model named IndoBERT. The author used the IndoBERT lite version of this study and compared the results of the F 1score with the F1score from the study conducted by author. In other paper,

In other paper studied about [8] classified emotions from Indonesian song lyrics using the support vector machine method. The accuracy results obtained after conducting training are 62.50%. The research conducted by the author also classifies emotions in Indonesian texts, but the methods and datasets are different.

Furthermore, there is study [9] performed a sentiment analysis task on reviews of English-language films using the BERT method. Meanwhile, the research conducted by the author is the classification of emotions in Indonesian texts using the IndoBERT method, or the Indonesian version of BERT. In addition, Fakhruzzaman et.al[10], created a web-based application to detect clickbait on Indonesian news using the IndoBERT model. The metric used in this study was the average ROC-AUC value obtained at 89%. The research

conducted by the author also uses IndoBERT but to detect emotions in online chat applications and the metrics used are accuracy, F1 score, recall, and precision.

Then, this article [11] talked about the creation of chatbots that can provide psychiatric counseling through dialogue in the form of text in Korean. This study used the GRU (Gated Recurrent Unit) method. Just like in the study, the author will create a chatbot that can classify emotions through Indonesian text .

III. RESEARCH METHODS

A. Implementation Methods

The implementation method that will be carried out consists of three steps: data preparation, modelling, and application development.

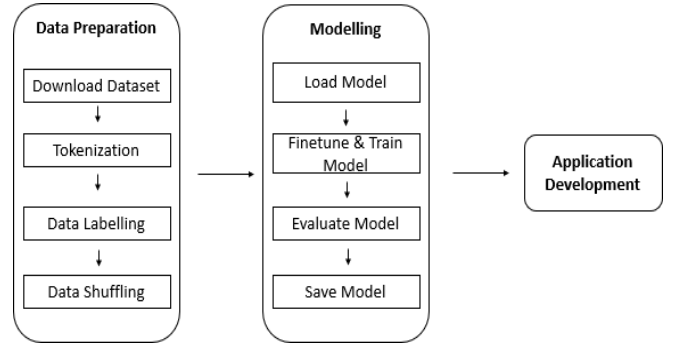


Fig. 2. Implementation Methods

B. User Acceptance Testing

User acceptance testing or abbreviated as UAT is a type of testing carried out by the end user or client to verify or receive a software system before moving a software application to the production stage. There are various types of UAT, however, pthere is this research the type of UAT black box will be used. Black box testing is a test of software functionality by the user. That is, users who run black box testing only know what a piece of software should do and do not need to know the source code of the software .

Theassessment that hany a conducted to test the chatbot in assessing whether the chatbot has the ability to classify statements submitted by the user. Because the chatbot creation project in this study is classified as a small scale (only done by someone, namely the author himself), the minimum number of testers is required i.e. 3 to 5 people [12].

IV. RESULTS AND DISCUSSION

A. Data Preparation

Data preparation is the stage where raw data is cleaned for analysis purposes. There are 4 steps at this stage, namely dataset download, tokenization, data labelling, and data shuffling.

a) *Download Dataset:* This study used a public dataset belonging to Mei Silviana Saputri et al [13]. This dataset has a csv file type (comma-separated values). This dataset consists of 3 sets, namely training, validation, and test. The training set has 3,523 rows of data, the validation set has 440

rows of data, and the test set has 440 rows of data. These data lines are tweets of Indonesian- language netizens taken from Twitter.

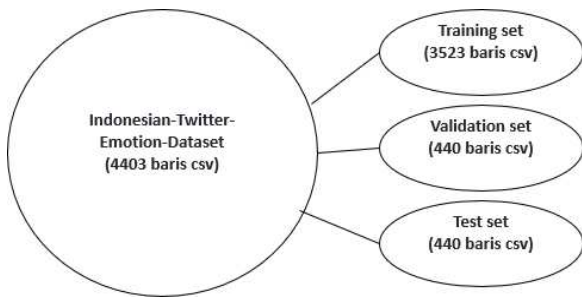


Fig. 3. Visualization of Dataset Division

b) *Tokenization*: Tokenization is the process of breaking text into pieces of words or tokens. This study used a special pre-trained tokenizer belonging to the indobert- liteLARGE model of the first phase [7]. This is done because pretrained-models such as IndoBERT, will only work well if given the same token inputs as the tokens used for tokenization of the pre-trained model.

c) *Data Labelling*: In this study, the dataset to be used has been labeled to categorize text into classes. These classes are in the form of emotions that will be understood by the model, namely love, joy, anger, sadness, and fear. Later the text with sad emotions will be given a value of 0, angry emotions will be given a value of 1, love emotions will be given a value of 2, fear emotions will be given a value of 3, and emotions of fear will be given a value of 3, and emotions of anger will be given a value of 3 glad to be rated 4.

d) *Data Shuffling*: Data Shuffling is commonly done before training machine learning models. This is done to minimize variance and ensure that the model will learn the data representation well. In this study, the authors will do data shuffling on the validation set, and test set so that the model can generalize the data well and reduce overfitting on the model.

B. Modelling

There are 4 steps at this stage, namely dataset download, tokenization, labelling data, and data shuffling

a) *Load Model*: This study used the first phase of the IndoBERT-liteLARGE model that has been trained [7]. The reason is because the model does not have a file size that is too large, which is 67.5 MB, so it will succeed when loaded at the deployment model stage.

b) *Finetune & Train Model*: IndoBERT is a pre-trained model, meaning that the model has been created and trained. However, the model is not completely accurate and ready-to-use for the purposes of this study, so a finetune & train model process is needed.

The fine-tuning carried out is to determine the ADAM optimizer, the learning rate is 0.000003, the batch size is 16, and the epoch is 5. After that, the next step is to train the model using the GPU in the Colaboratory.

c) *Evaluate Model*: This research uses the values in the confusion matrix, namely accuracy, precision, recall, and F1. Here are the results of the finetune & train model stage process

TABLE III. TRAINING RESULTS

Training Set				Validation Set			
Accura cy	F1- scor e	Reca ll	Preci sion	Accura cy	F1- scor e	Reca ll	Preci sion
89%	89 %	89%	90%	70%	71 %	70%	72%

d) *Save Model*: At this stage, the model is saved to go into the deployment model stage. The part that is saved at the save stage of this model is the weight and bias on the model.

C. Application Development

After going through the modelling process, the next stage is application development. Here's a picture that explains the process.

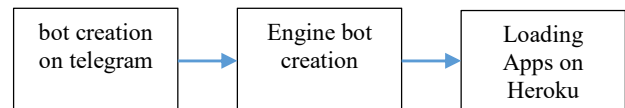


Fig. 4. Application Development

This study uses a smartphone that will run one of the features of Telegram, namely a bot. Later the bot will respond and classify emotions automatically based on input from the user. The user sends a text message to the account belonging to the bot through the Telegram server which will later be forwarded to the bot server. Text messages will be processed on the bot server to be classified and the response resulting from the classification, i.e. the type of emotion, will be given which is interpreted from the text. A response from the bot will be sent to the user via the Telegram server.

a) *Bot Creation on Telegram*: The creation of bots in Telegram is done through a bot called BotFather which is available on Telegram. In BotFather, the following steps are required to create the bot

- Type the command "/start" to start the interaction to BotFather.
- Type the command "/newbot" to create a new bot.
- Type a name for the bot to be created.
- Type a username for the bot.

b) *Bot Engine Creation*: Creation of a bot engine to connect the Telegram server with the bot server using a Python library called Flask. Flask functions such as responding to the results of the classification of emotions, calling the bot server, and classifying emotions in the text using the model created at this stage.

c) *Loading Apps on Heroku*: Loading the application on Heroku is the last stage after creating a bot on Telegram and creating an engine for the bot. This is done so that the application for this research can get a cloud-based server and other features so that the application can be run easily in real time.

The process to load the application on Heroku is not difficult, it only takes a few steps, namely

- Create a file containing the libraries needed for the application to be loaded in Heroku.
- Create a Profile file to tell Heroku what commands to run first in order for the application to run.
- Copy the Heroku domain and pass it to the
- bot engine variable.

- Copy the token to access the Telegram HTTP API and pass it into the bot engine variable.

d) *User Acceptance Testing*: The tests that have been carried out are distributing test questionnaires to respondents who use the Telegram application. In the questionnaire, the respondents were asked to enter 5 sentences (1 sentence of sad emotions, 1 sentence of angry emotions, 1 sentence of love emotions, 1 sentence of fear emotions, 1 sentence of happy emotions) that they had tested on the Telegram chatbot application. In addition, the respondents also got a question whether the sentences they had tested were successfully classified by chatbots or not.

The results of the test conducted by 6 respondents with each giving 5 sentences, obtained as many as 23 sentences classified precisely according to the class of emotions, and 7 sentences were not classified appropriately.

TABLE IV. TEST RESULTS

Total Sentences: 30	Percentage
Properly classified	76,7%
Not Properly Classified	23,3%

V. CONCLUSIONS AND SUGGESTIONS

Based on the results of the analysis and design that have been made, the following conclusions were obtained

1. A chatbot application for the classification of emotions on text was successfully created on the Telegram messaging app platform.
2. Based on the test results with a total number of epochs of 5, batch size 16, learning rate 0.000003, ADAM optimizer, and using indobert-liteLARGE first phase, obtained accuracy, F1 score, recall, and precision values on training sets were 89%, 89%, 89%, and 90%, respectively, while in the validation set, it was 70%, 71%, 70%, and 72%.
3. The test results by 6 respondents with a total of 30sentences showed the model had an accuracy of 76.7%

The suggestions that can be considered for future research are

1. Using a better dataset with a larger amount of data to increase the accuracy and accuracy of the model in classifying emotions.
2. Using the second phase indobertlarge model to obtain better accuracy [7].
3. Rent a cloud server so that it can load larger models such as IndoBERTLARGE.

REFERENCES

- [1] Association of Indonesian Internet Service Providers. (2018, February 21). Chat, the Most Popular Application of The Indonesian People.Databoks. <https://databoks.katadata.co.id/datapublish/2018/02/21/chatting-aplikasi-paling-digemari-masyarakat-indonesia>
- [2] Kemp, S. (2022, February 15). Digital 2022: Indonesia.DataReportal.<https://datareportal.com/reports/digital-2022-Indonesia>
- [3] Mulqueen, M. (2019, Desember 07). Texting really is ruining personal relationships. NBC News. <https://www.nbcnews.com/think/opinion/texting-really-ruining-personal-relationships-ncna1097461>
- [4] Johnson, A. J., Bostwick, E., & Anderson, C. (2016). How do computer-mediated channels negatively impact existing interpersonal relationships? In E. S. Gilchrist-Petty & S. D. Long (Eds.), Contexts of the dark side of communication (pp. 241–252). New York, NY: Peter Lang.
- [5] Kelly, L., Keaten, J. A., Becker, B., Cole, J., Littleford, L., Rothe, B. (2012). “It’s the American Lifestyle!”: An Investigation of Text Messaging by College Students. Qualitative Research Reports in Communication, 13(1), 1–9. doi:10.1080/17459435.2012.719203
- [6] Dun, T. (2005, Mei). Troublesome differences: A typology of lay views of misunderstanding. Paper presented at the meeting of the International Communication Association Conference, New York, NY.
- [7] Wilie, B., Vincentio, K., Winata, G. I., Cahyawijaya, S., Li, X., Lim, Z. Y., Soleman, S., Mahendra, R., Fung, P., Bahar, S., & Purwarianti, (2020). IndoNLU: Benchmark and Resources for Evaluating Indonesian Natural Language Understanding. <http://arxiv.org/abs/2009.05387>
- [8] Abirawa, I. D., Osmond, A. B., & Setianingsih C. (2018). Classification of Emotions in Song Lyrics Using Support Vector Machine. 5(3), 6203-6209.
- [9] Princess, C. A. (2020). Sentiment Analysis of English-Language Film Reviews With a Bidirectional Encoder Representations from Transformers Approach. JATISI (Journal of Informatics And Information Systems Engineering), 6(2),181–193. <https://doi.org/10.35957/jatisi.v6i2.206>
- [10] Fakhruzzaman, M. N., & Gunawan, S. W. (2021). Web-based Application for Detecting Indonesian Clickbait Headlines using IndoBERT. <https://arxiv.org/abs/2102.10601>
- [11] Oh, K., Lee, D., Ko, B., & Choi, H. (2017). A Chatbot for Psychiatric Counseling in Mental Healthcare Service Based on Emotional Dialogue Analysis and Sentence Generation. 2017 18th IEEE International Conference on Mobile Data Management (MDM), pp. 371-375, doi: 10.1109/MDM.2017.64.
- [12] Rachlin, R. (n.d.). How to rightsize your mobile app beta program. TechBeacon. techbeacon.com/app-dev-testing/how-rightsized-your-mobile-app-beta-program#:~:text=Generally%2C%20you%20only%20need%20about,more%20testers%20for%20ga ming%20apps.
- [13] Saputri, M. S., Mahendra, R., & Adriani, M. (2018). Emotion Classification on Indonesian Twitter Dataset. 2018 International Conference on Asian Language Processing (IALP), pp. 90-95, doi: 10.1109/IALP.2018.8629262.