A new technology on translating Indonesian spoken language into Indonesian sign language system

Risky Aswi Ramadani, I Ketut Gede Dharma Putra, Made Sudarma, Ida Ayu Dwi Giriantari
Program Doktoral Ilmu Teknik, Universitas Udayana, Denpasar, Indonesia

Article Info

ABSTRACT

People with hearing disabilities are those who are unable to hear, resulted in their disability to communicate using spoken language. The solution offered in this research is by creating a one way translation technology to interpret spoken language to Indonesian sign language system (SIBI). The mechanism applied here is by catching the sentences (audio) spoken by common society to be converted to texts, by using speech recognition. The texts are then processed in text processing to select the input texts. The next stage is stemming the texts into prefixes, basic words, and suffixes. Each words are then being indexed and matched to SIBI. Afterwards, the system will arrange the words into SIBI sentences based on the original sentences, so that the people with hearing disabilities can get the information contained within the spoken language. This technology success rate were tested using Confusion Matrix, which resulted in precision value of 76%, accuracy value of 78%, and recall value of 79%. This technology has been tested in SMP-LB Karya Mulya on the 7th grader students with the total of 9 students. From the test, it is obtained that 86% of students stated that this technology runs very well.

Corresponding Author:

I Ketut Gede Dharma Putra
Program Doktoral Ilmu Teknik (PDIT)
Universitas Udayana
Denpasar, Bali, 80234, Indonesia
Email: ikgdarmaputra@unud.ac.id

1. INTRODUCTION

Deaf people refers to that who has disabilities in hearing function. This condition might be either temporary or permanent. For those living around people with hearing disabilities, a special communication form is highly needed in order to properly deliver the message in a conversation [1-3]. Indonesian sign language system or *sistem isyarat Bahasa Indonesia (SIBI)* is a sign language system specifically developed for people with hearing disabilities. Indonesian sign language system follows Indonesian grammatical structure, and the morphology is also based on *Bahasa Indonesia* [4-6]. The Indonesian sign language system also includes prefixes, basic words, and suffixes [7].

Within the information technology field, especially in Indonesia, there are still very limited researchers developing technology on translating language specific for people with hearing disabilities. Technologies which have been developed currently for people with hearing disabilities are Indonesian vocabularies dictionary, and Indonesian sign language morphology [8, 9]

Creating special facilities for people with hearing disabilities is a heavy task. In 2001, the Indonesian government had just established the dictionary of Indonesian sign language system, of which during that time there were only a few technology developments on the field. Meanwhile, the hearing disabled people were in high needs of communication tools to be used in receiving information in the form of

audio [10]. The solution of this problem is to develop a one-way translation technology from spoken language to Indonesian sign language system.

In 2018, we developed a data set on Indonesian sign language system according to the morphology in *Bahasa Indonesia*. The data set includes prefixes, basic words, and suffixes. The research was entitled *"Database of Indonesian Sign System"* [11]. This research is the continuation from the preceding research in 2018. In this research a technology would be developed to conduct one way translation from spoken language to Indonesian sign language system [12]. Indonesia novelty of this research is being able to make an interpreter of spoken language into Indonesian sign language, by following the morphology of Indonesian.

Input used in this research consists of sentences delivered orally (audio), which then was recorded using microphones in laptops or mobile phones. The voice would subsequently be converted into texts using speech recognition technology [13-15]. The texts would later enter the processing stage. Text processing is used to select input sentences [16, 17]. Text processing consists of several stages such as case folding, tokenizing, and stemming. Case folding is conducted to homogenize the letters to lower case letters. Tokenizing is used to trim the sentences to words. The last stage is stemming which is used to separate word elements to prefixes, basic words, and suffixes [18, 19].

The words which have been identified into a number of prefixes, basic words, and suffixes would then go through indexing process with the prepared database [20]. Finally, the words would be converted to a word in Indonesian sign language system (Images). In order not to change the information contained in Indonesian sign language system, the pictures will be arranged in the formation based on the input sentences. This translation technology has gone through two stages of tests. The first test was using confusion matrix aimed to identify accuracy, precision, and recall [21, 22]. The second test was conducted to understand to what extent the students of SMP-LB Karya Mulya are able to use this technology.

2. RESEARCH METHOD

This research focuses on technology on translating spoken *Bahasa Indonesia* to Indonesian sign language system. To implement this technology, a long process should be conducted. The process spans from inserting input in the form of sentences which are normally used by Indonesian people in regular conversation to produce output in the form of sign language to be received by people with hearing disabilities. This application was built in website base so that it is accessible in all kinds of platforms [23]. More detailed explanation and scheme are depicted in the following flowchart as shown in Figure 1.
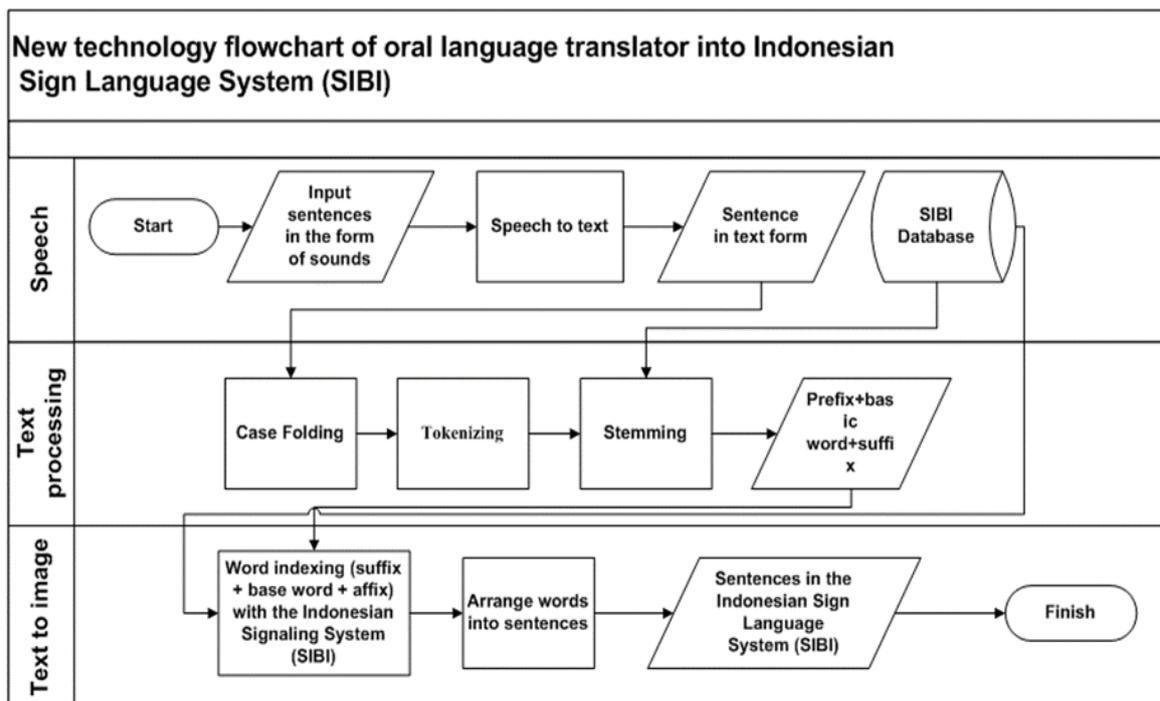

Figure 1. Flowchart of the new technology on translating spoken language to Indonesian sign language system

2.1. Speech recognition

Speech Recognition is the process of converting audio to texts. Audios are sourced from the society who are not familiar with sign language system but are aiming to deliver message towards their hearing-impaired counterparts. Information conveyed by the society is in the form of spoken sentences (audio). Then, the audio input would be converted into texts by using speech recognition [24, 25]. The output produced from speech recognition process are sentences (texts). The Figure 2 is a simplified figure showing the speech recognition process.

Speech recognition is biometric recognition, which is the process of a computer recognizing what someone is saying based on the tone of voice that is converted into digital text. This system is very good if it is used to develop communication applications for the general public with hearing impairers. The process is to convert sound spectrum data into digital form and convert it to text. Speech recognition has several steps that are important for recognizing sounds. The following are the steps that speech recognition uses to recognize sound.

−   Receiving input data
−   Extraction, i.e. input data storage as well as database creation
−   Benchmarking/matching, which is the stage of matching new data with voice data (matching grammar) in the database
−   User identity validation
−   After the data is validated, the user (general public) can submit information that will be conveyed to the deaf through microphone
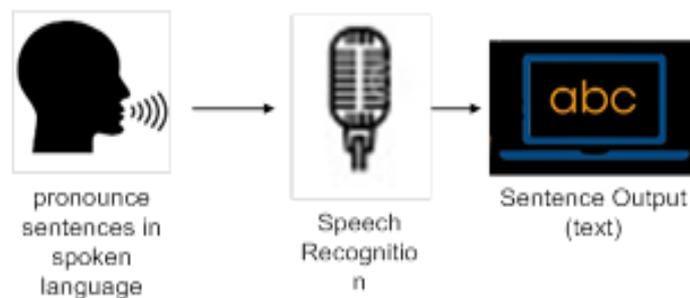−   After the sound is saved, the system will convert the data into text.

Figure 2. Processing audio to texts

2.2. Text processing

The texts produced in the previous stage were not arranged yet, therefore text processing is needed to arrange and to split the texts based on their syllables [26, 27]. The following are stages in text processing.

2.2.1. Case folding

Case folding is the stage where texts were homogenized to lower cases. Figure 3 is the case folding process figure.
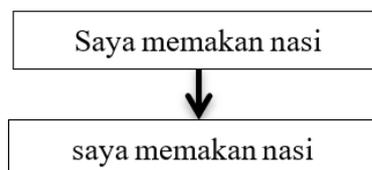

Figure 3. Case folding process

2.2.2. Tokenizing

Tokenizing is splitting the previously homogenized texts into words. The words are recognized by space, enter, and tabulation. The Figure 4 is an illustration of the tokenizing process.
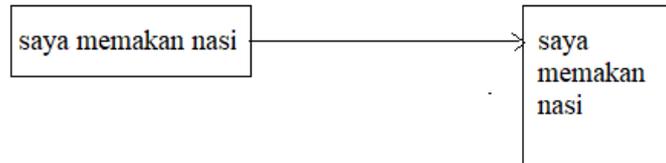
Figure 4. Tokenizing process

2.2.3. Stemming

Furthermore, stemming is the process to separate prefixes, basic words, and suffixes [28, 29] stemming is made from *Bahasa Indonesia* morphology structure, which is depicted as:

$$Prefix + Basic\ Word + Suffix$$

Affixation is the process of adding / affixing or affixing. Affixation consists of:
− prefix (prefix): ber-, be-, PER-, per-, di-, ter-, ke, se
− suffix (suffix): -kan, -an, -i

The use of words in Indonesian usually gets a prefix or suffix, for example the word eats, the word gets a prefix of the word and the word *(makan)* eat basically. The Table 1 is a list of words that get prefix and suffix. Stemming in this research is specifically designed to build translator technology for people with hearing disabilities. The images of sign language system or *sistem bahasa isyarat* (SBI) are sourced from dictionary of sign languages published in 2001 by the Department of Education. The total data of words taken is 3340 words consist of basic words, prefixes, and suffixes. To develop this technology, a data set is needed to store the words/texts. The texts would later be stored in database, and are called as basic words in this research. Texts (SIBI) were stored in repository or specified folder entitled "Image of SIBI". The purpose of separating basic words and images of SIBI is to accelerate the data collection process.

By splitting the prefixes and suffixes, stemming process will identify the basic words. The identified basic words are later traced for its Id. Every word going through this translator technology should pass stemming process, in the Figure 5 show the example of sentence input and the stemming process. The processed basic words were then shown based on their Ids, the Ids were subsequently matched with the images of words (SIBI) stored in the repository.

Table 1. List of sentences to test

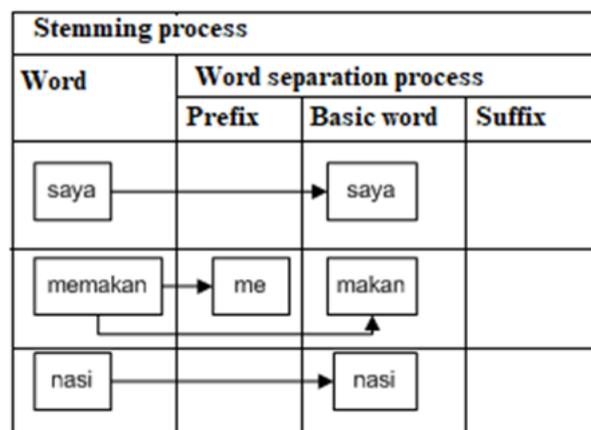| Number	Word	Prefix	Basic Word	Suffix
1.	Memakan (eat)	Me-	Makan (eat)	
2.	Membaca (read)	Mem-	Baca (read)	
3.	Bersihkan (cleaning)		Bersih (cleaning)	-kan
4.	Menulis (write)	Me-	Tulis (write)	
5.	Mandikan (bathe)		Mandi (bathe)	-kan

Figure 5. Stemming process

2.2.4. Text to image

The process of converting text to images (SIBI) uses an Id as a link between text and image text (SIBI). Text images (SIBI) are stored in a specific repository, to allow faster access and lighten system performance. Figure 6 is a simple image showing the process of converting text to text image (SIBI). In this research, Indexing is conducted to match the texts with images (SIBI). The following chart illustrates the matching process as shown in Figure 7.
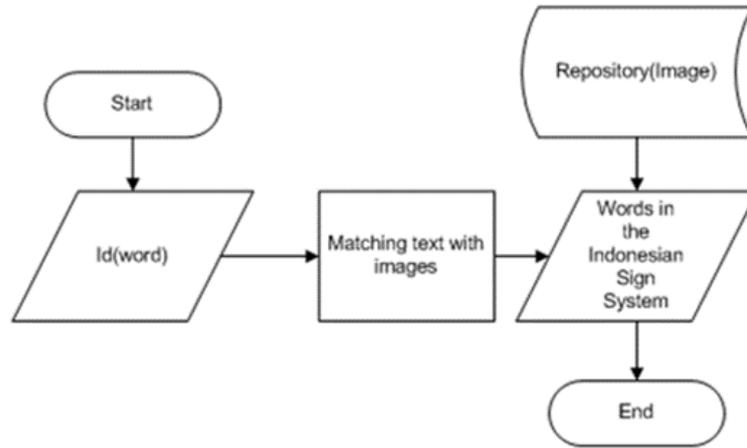
Figure 6. Flowchart of indexing texts with images

Figure 7. Illustration on matching text Id to image Id

2.2.5. Sentences in sign language

Images of words (SIBI) which were previously separated then being arranged in a sentence (SIBI). Images of sentences (SIBI) can be easily seen by the people with hearing disabilities. The display in this output is still in black and white based on its sources in the dictionary which were obtained through image scanning. The Figure 8 is the image of sentence (SIBI) output.

Figure 8. Sentences in the form of Indonesian sign language system (SIBI)

3.   RESULTS AND DISCUSSION
3.1.   Precision and recall accuracy analysis

Model testing is an essential step, considering the application will be directly related to the people with hearing disability [30-32]. To analyze accuracy, precision, and recall test of this application, confusion matrix is used. Data is used to test 100 sentences (audio) to be converted to images of texts (SIBI). Data in Table 2 is data used in system testing. By using confusion matrix, researcher can assess system success rate, by calculating accuracy, precision, and recall as success parameter. The Figure 9 shows accuracy, precision and recall calculation.

Table 2. List of sentences to test

Number	Sentence
1	*saya membaca buku setiap sore*
2	*stop, apa yang kamu lakukan*
3	*ayo silahkan airnya diminum*

N=100		Prediction Value	
		False	True
True Value	True	True Negative(TN)= 40 Correction of no results	False Positive(FP)=12 unexpected results
	False	False Negative(FN)=10 missing results	True Positive(TP)=38 Correct results

Figure 9. Research result analysis using confusion matrix

The data is obtained from the application testing results. From the data, researcher is allowed to acquire accuracy, precision, and recall values of the application. The detailed calculation is shown in (1)-(3). The calculation has resulted in the values of precision is 76%, accuracy is 78%, and recall is 79%. It can be concluded from the values that this technology run well.

$$Accuracy = \frac{(TP+TN)}{TP+TN+FP+FN} \tag{1}$$
$$= \frac{(38+40)}{38+40+12+10}$$
$$= 0,78$$

$$Precision = \frac{TP}{TP+FP} \tag{2}$$
$$= \frac{38}{38+12}$$
$$= 0,76$$

$$Recall = \frac{TP}{TP+FN} \tag{3}$$
$$= \frac{38}{38+10}$$
$$= 0,79$$

3.2.   Testing of translator technology in students of SMP-LB Karya Mulia Surabaya

Testing of translator technology in Surabaya Mulya Junior High School-LB students is needed, considering that this translator technology will be used by deaf people. Translator technology was tested on 1st grade students of SMP-LB Surabaya, these students were aged 12-13 years. The system was tested on 9 students. The Figure 10 is a photo of the trial activities carried out at SMP-LB Karya Mulya Surabaya. Instrument is needed to measure the success of translator technology, in the pilot phase a questionnaire is used. To analyze the number of students from SMP-LB Karya Mulya Surabaya who rate this application it is good to use (4).

$$\text{Index formula } \% = \frac{\text{total score}}{Y} * 100 \tag{4}$$

From (4), it is found that the percentage of students who gave excellent grades, can be seen more clearly in the Table 3.

Figure 10. Trial translator system translating Indonesian spoken language into Indonesian sign language system (SIBI)

Table 3. Results of the assessment by questionnaire

Number	Assessment parameters	The results that show are very good
1.	Interface	87 %
2.	Navigation menu	71 %
3.	Translation results	98 %
4.	User Guide	73 %
5.	Informative	98 %
6.	Recording sensitivity	87 %
Average		86 %

From the results in Figure 3 it can be concluded that 86% of respondents rated translator technology as very good. This research also gets some advice from people with hearing impairment such as improving accuracy, increasing vocabulary, and improving appearance. And needs to be made an Android-based application.

4.   CONCLUSION

This new technology on translating spoken language (audio) to Indonesian sign language system is capable to translate spoken language to SIBI language. The stages in conducting translation are speech recognition, text processing, basic words, images of SIBI, stemming, indexing, sentences in sign language, and video in SIBI. Video is used to allow the people with hearing disabilities to receive output easily. This translating technology has been tested through converting spoken sentences. The total number of spoken sentences is 100 sentences. From the test, it was obtained the data such as: precision value of 76% and accuracy value of 78%. The recall value in this system is already good in the number of 79%. From that, it can be concluded that the system has been able to translate spoken sentences to pictured sentences in the form of SIBI images. Whereas for interface assessment, navigation menu, translation results, user guide, informative, and recording sensitivity. Value by respondents is very good, with an average value of 86%.

REFERENCES

[1]   Toba, Y., Horiuchi, H., Matsumoto, S., Saiki, S., Nakamura, M., Uchino, T *et al.*, "Considering multi-modal speech visualization for deaf and hard of hearing people," *2015 10th Asia-Pacific Symposium on Information and Telecommunication Technologies (APSITT)*, Colombo, 2015, pp. 1–3, doi: 10.1109/APSITT.2015.7217102.

[2]   V. E Newton and S. R. Shah, "Improving communication with patients with a hearing impairment," *Community Eye Health*, vol. 26, pp. 6–7, 2013.

[3]   U. Lemke and S. Scherpiet, "Oral communication in individuals with hearing impairment-considerations regarding attentional," *Front Psychol.*, vol. 17, pp. 8–15, 2015.

[4]   S. D Larasati, V. Kuboň, and D. Zeman, "Indonesian Morphology Tool (MorphInd): Towards an Indonesian Corpus," *International Workshop on Systems and Frameworks for Computational Morphology*, 2011, vol. 4, pp. 119–192.

[5]   F. Pisceldo, R. Mahendra, R. Manurung, and I. W. Arka, "A Two-Level Morphological Analyser for the Indonesian Language," *Association for Computational Linguistics*, vol. 5, pp. 1–6, 2014.

[6]   H. Alwi, S. Ardjowidjojo, and H. Lapoliwa, "Indonesian Language Standard Grammar," (in Indonesia), Pusat dan Balai Pustaka, Jakarta, 2003.

[7]   Departermen Pendidikan Nasional Indonesia, "Indonesian Sign System," (in Indonesia), Pusat dan Balai Pustaka, Jakarta, 2002.

[8]   A. W. Yanuardi, S. Prasetio, and P. P. J. Adi, "Indonesian Sign Language Computer Application for the Deaf," *International Conference on Education Technology and Computer*, Shanghai, 2010, vol. 2, pp. 57–65.

[9]   M. N. Baehaqi, M. Irzal, and F. H. Indiyah, "Morphological Analysis of Speech Translation into Indonesian Sign Language System (SIBI) on Android Platform," *International Conference on Advanced Computer Science and information Systems (ICACSIS)*, Bali, 2015, pp. 205–210.

[10] W. K. Mprah, "Sources and Use of Sexual and Reproductive Health Information Among Deaf People in Ghana," *Indonesian Journal of Disability Studies*, vol. 1, pp. 1–11, 2014.

[11] R. A. Ramadhani, I. D. Putra, M. Sudarma, and I. A. Giriantari, "Database of Indonesian Sign System," in *International Conference on Smart Green Technology in Electrical and Information Systems*, Bali, Indonesia, 2018, vol. 3, pp. 225–228.

[12] R. M. McGuire, J. Hernandez-Rebollar, T. Starner, V. Henderson, H. Brashear, and D. S. Ross, "Towards a one-way American sign language translator," *International Conference on Automatic Face and Gesture Recognition*, Seoul, 2004, vol. 6, pp. 620–625.

[13] T. Shimada, R. Nishimura, M. Tanaka, H. Kawahara, and T. Irino, "Developing a method to build Japanese speech recognition system," *Conference Anthology*, China, Jan. 2013, pp. 120–125.

[14] A. N. de Andrade, C. Maria, and Daniela G., "Speech recognition in individuals with sensorineural hearing loss," *Brazilian Journal of Otorhinolaryngology*, vol. 28, pp. 1–5, 2016.

[15] C. Brookes, "Speech-to-text systems for deaf, deafened and hard-of-hearing people," in *Seminar on Speech and Language Processing for Disabled and Elderly People*, London, 2000, vol. 4, pp. 5/1–5/4.

[16] X. Lin, J. Yang, and J. Zhao, "The text analysis and processing of Thai language text to speech conversion system," in *International Symposium on Chinese Spoken Language Processing*, Singapore, 2014, vol. 9, pp. 436–436.

[17] R. Forgáč and R. Krakovský, "Text processing by using projective ART neural networks," in *New Trends in Signal Processing (NTSP)*, Demanovska Dolina, vol. 6, pp. 1–5, 2016.

[18] A. Ghazvini and M. J. Ab Aziz, "Stemming algorithm for different tenses to improve Persian dictionary," in *Symposium on Industrial Electronics and Applications*, Bandung, 2012, pp. 50–53.

[19] F. S. Alotaibi and V. Gupta, "A cognitive inspired unsupervised language-independent text stemmer for Information retrieval," *Cognitive Systems Research,* vol. 58, pp. 291–300, 2000.

[20] T. Neugebauer, "Image indexing," *The Indexer*, vol. 28, pp. 98–103, 2010.

[21] S. Ruuska, W. Hämäläinen, S. Kajava, M. Mughal, P. Matilainen, J. Mononen, and M. Joliat, "A Evaluation of the confusion matrix method in the validation of an automated system for measuring feeding behaviour of cattle," *Behavioural Processes*, vol. 148, pp. 56–62, 2018.

[22] P. Trajdos and M. Kurzynski, "Weighting scheme for a pairwise multi-label classifier based on the fuzzy confusion matrix," *Pattern Recognition Letters*, vol. 103, pp. 60–67, 2018.

[23] K. P. Dipin, J. Bose, and S. Pal, "User Friendly Tab Management in Web Browsers," in *Annual Consumer Communications & Networking Conference (CCNC)*, Las Vegas, vol. 16, pp. 1–4, 2019.

[24] V. Mitra, H. Franco, M. Graciarena, and D. Vergyri, "Medium-duration modulation cepstral feature for robust speech recognition," in *International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 2014, pp. 1749–1753.

[25] T. Athanaselis, S. Bakamidis, G. Giannopoulos, I. Dologlou, and E. Fotinea, "Robust speech recognition in the presence of noise using medical data," *International Workshop on Imaging Systems and Techniques*, Crete, 2008, pp. 349–352.

[26] X. Lin, J. Yang, and J. Zhao, "The text analysis and processing of Thai language text to speech conversion system," in *International Symposium on Chinese Spoken Language Processing*, Singapore, vol. 9, pp. 436–436, 2014, doi: 10.1109/ISCSLP.2014.6936630.

[27] L. Skovajsová, "Long short-term memory description and its application in text processing," *Communication and Information Technologies*, Vysoke Tatry, pp. 1–4, 2017.

[28] G. Septian, A. Susanto, and G. F. Shidik, "Indonesian news classification based on NaBaNA," in *International Seminar on Application for Technology of Information and Communication*, Semarang, 2017, pp. 175–180, doi: 10.1109/ISEMANTIC.2017.8251865.

[29] D. S. Maylawati, W. B. Zulfikar, C. Slamet, M. A. Ramdhani, and Y. A. Gerhana, "An Improved of Stemming Algorithm for Mining Indonesian Text with Slang on Social Media," *2018 6th International Conference on Cyber and IT Service Management (CITSM)*, Parapat, 2018, vol. 6, pp. 1–6.

[30] L. Chen and H. L. Tang, "Improved computation of beliefs based on confusion matrix for combining multiple classifiers," in *Electronics Letters*, vol. 40, pp. 238–239, 2004.

[31] F. J. Ariza-Lopez, J. Rodriguez-Avi, and M. V. Alba-Fernandez, "Complete Control of an Observed Confusion Matrix," *International Geoscience and Remote Sensing Symposium*, Valencia, 2018, pp. 1222–1225, doi: 10.1109/IGARSS.2018.8517540.

[32] D. E. Zomahoun, "A Semantic Collaborative Clustering Approach Based on Confusion Matrix," *International Conference on Signal-Image Technology & Internet-Based Systems*, 2019, pp. 688–692, doi: 10.1109/SITIS.2019.00112.

BIOGRAPHIES OF AUTHORS

Risky Aswi Ramadhani, Born April 8, 1990, S1 at Nusantara University PGRI Kediri, S2 at the Yogyakarta Academy of Computer Science Management, is currently a PhD student in the Engineering Study Program of Doctoral Engineering at Udayana University, Bali, the field of interest is natural langunge processing.

I Ketut Gede Darma Putra, Born Mengwitani, April 24, 1974. Earning a Doctorate degree at Gajah Mada University, in the Electrical Engineering Study Program, the area of interest is Image Processing/ Biometrics. Currently teaching at the Doctoral Program in Engineering, Udayana University.

Made Sudarma, Completed Undergraduate Study In 1992 at the Ten November Institute of Technology Surabaya, Electrical Engineering study, Completed Masters in 2000 at the School of Information Technology and Engineering, Otawa University, Canada, S3 completed in 2012 in the Post-Graduate Program, Udayana University, Field which are interested in data mining, HCI and Smart Systems.

Ida Ayu Dwi Giriantari, Chair of the Core (Center of Excellence Community Based Renewable Energy), Teaching in the Doctoral Studies Program in Engineering, Udayana University, in 2005 earned a Ph. D at the University of New South Wales, Sydney, Australia.