Khazanah Journal of Religion and Technology
Universitas Islam Negeri Sunan Gunung Djati Bandung, Indonesia
https://journal.uinsgd.ac.id/kjrt
Online ISSN: 2987-6060
Volume 2 Issue 2 (2024)
DOI: https://doi.org/10.15575/kjrt.v2i2.1110

Slang Spelling Detection in Indonesian Religious Content Using Sequence-to-Sequence Algorithm

Nanda Tiara Sabina Hidayatulloh
Reaksi Community
Bandung, Indonesia
nandatiarash@gmail.com

Nur Halizah
Community of Arunika Mengabdi
Bandung, Indonesia
noorhalizah0@gmail.com

Pitriani
Community of Arunika Mengabdi
Bandung, Indonesia
ppitria05@gmail.com

Received: November 20, 2024; Revised: December 20, 2024; Published: January 8, 2025

Abstract— Spelling errors are a common problem in text processing, one of which is in Indonesian. The increasing use of non-standard language, especially in digital text communication, is the background for this research. Spelling errors in sensitive religious content can even cause misunderstandings. This article examines the development of a model for spelling correction with a deep learning-based approach using Sequence-to-Sequence with a GRU-based encoder-decoder architecture and attention mechanism. A dataset containing standard and non-standard text pairs is used to test the model. The experimental results show that the proposed model produces 76.82% accuracy, but is able to recognize and correct spelling errors. However, this research is expected to contribute in the future, so that it can improve improvements in the Indonesian spelling correction system.

***Keywords*- Religious Content, Sequence-to-Sequence, Slang, Spelling Detection.**

I. Introduction

The term "language" comes from Sanskrit, namely "bhāṣā" which means speech or saying. As a means of communication, language has a very important role in various things. Through language, a person can convey opinions, thoughts, and feelings to others [1]. Therefore, good language skills are considered very important so that a person is able to interact well [2]. Because if the interaction is clear, precise, and does not cause double interpretations, it will be easily accepted and understood by others [1]. In addition, language can help someone in learning various things [3].

The many islands in Indonesia have given birth to various regional languages. Before becoming a national language, people in Indonesia communicated using their respective regional languages. On October 28, 1928, through the Youth Pledge, Indonesian was established as the unifying language and national language [3]. The variety of words in Indonesian is divided into two, namely standard words and non-standard words. Standard words are words that are in accordance with the rules or guidelines of the language that have been determined. While non-standard words are words that are not found in the Big Indonesian Dictionary and do not comply with the rules or guidelines of the language that have been determined [4].

In communicating, a person can do it in two ways, namely directly face to face or indirectly with messages, either in text or voice. When someone communicates via text messages, sometimes spelling or writing errors occur. Not only in communicating, but in some conditions, writing errors can occur in documents, scientific papers, or final assignments. This can happen because the position of each letter is close to each other on the keyboard which causes someone to choose the wrong letter, inaccuracy, or even limited knowledge of Indonesian spelling and vocabulary which is influenced by the influence of foreign languages and informal languages, as well as the procedure for writing a word in a good and correct sentence. Spelling or writing errors in a sentence or word will certainly make it difficult to understand the information.

One of the technologies that can be a solution to this problem is Natural Language Processing (NLP) also known as natural language processing. NLP is a science that studies the interaction between computers and human language which aims to design and build software that can analyze and produce human language naturally [5]. One branch of science in NLP is spelling correction, which is the process of detecting and correcting writing or spelling errors in a text. In the process, spelling correction will search for and scan a text then extract the words in it, and then compare each word with the list of words in the main dictionary used as a reference for the spell checker. If the word is not in the dictionary, it will be determined as an incorrect word. As for correcting the error, the model will search for the word that is most similar or close to the incorrect word in the dictionary [6].

To build a model capable of performing the spelling correction process, many algorithms can be used, one of which is the Long Short-Term Memory (LSTM), Gated Recurrent Unit (GRU), and Sequence-to-sequence (seq2seq) algorithms. The LSTM model is a development of the Recurrent Neural Network (RNN) architecture designed to handle the vanishing gradient problem [7]. In addition, the LSTM algorithm is able to model data sequences such as text and has the ability to remember information in the long term [8]. The Gated Recurrent Unit (GRU) is a development of a simpler LSTM cell structure because it has two main gates, namely the update gate which functions to control the amount of information from previous memory that needs to be maintained, and the reset gate which functions to delete information that is not needed [9].

Then the sequence-to-sequence (seq2seq) model is a model consisting of an encoder that functions as an input receiver and a decoder that functions as an output producer. The seq2seq model is usually used for automatic translation systems or text generation. Unfortunately, the seq2seq model is unable to handle or remember information from long sentences. Therefore, to produce more accurate text from even long sentences, an attention mechanism needs to be added to help the model focus more on important parts of the input when producing output [10].

However, along with the many recent developments in the field of NLP, presenting Transformer, a transduction model that relies on self-attention mechanisms in processing and understanding the relationship between input and output. To complete its various tasks, Transformer applies an encoder and decoder-based architecture [11]. In the process, the encoder's task is to process news text into a representation so that the model can easily understand it. Then the decoder, when in the training stage, will receive input [12]. In addition, the transformer will also change one sequence to another using the encoder and decoder [11].

Accurate spelling is critical in religious content, as it preserves the integrity and authenticity of the message being conveyed. Errors or slang usage in such texts can lead to significant misunderstandings, misinterpretations, or even distortions of core teachings and values. For instance, in Islamic content, a slight misspelling of key terms such as "*shalat*" (prayer) or "*zakat*" (charity) could lead to confusion, especially among learners or those new to the faith [13]. Moreover, religious content often serves as a guiding principle for personal and communal life, making clarity and precision in language a priority.

In the digital age, where religious teachings and messages are increasingly disseminated through online platforms, the risk of typographical and slang errors is higher. Automated spelling detection systems using advanced algorithms, such as the Sequence-to-Sequence model, offer a robust solution to address these challenges. By identifying and correcting slang or misspelled words in real time, these systems ensure that the sanctity and accuracy of religious texts are maintained [14], [15]. This is particularly crucial for fostering deeper understanding and engagement with the content, ensuring that the spiritual message resonates accurately with its audience.

Therefore, this study aims to provide a solution to the above problems, namely by developing a model that is able to correct the spelling of non-standard words or Indonesian slang. The algorithm that will be used in building spelling correction is Sequence-to-Sequence with a GRU-based encoder-decoder architecture, which is equipped with an attention mechanism. It is expected that combining the Sequence-to-Sequence and GRU models and the attention mechanism can facilitate and provide significant results in correcting the spelling of non-standard words or Indonesian slang.

II. RELATED WORKS

Relevant research related to spelling correction using the Peter Norvig and N-Gram methods. This study used 55 Indonesian language documents that will be used as the tested dataset. As a result, the application is able to correct documents in the form of text and successfully change incorrect words with an accuracy of 69.09%. However, this system has limitations in handling words with an error rate of more than two letters and the system still depends on the quality of the corpus used [16].

Further research related to spelling correction using the N-Gram and Jaro-Winkler Distance algorithms. The number of sentences used as test data was 180 Indonesian sentences. The results of this study indicate that by combining the N-Gram and Jaro-Winkler Distance methods, the application can find appropriate word correction suggestions for one wrong word in a sentence by achieving an accuracy of 71.348% with a success rate of 98.449% [17].

The next research related to spelling correction combines the Bigram Vector method and Minimum Edit Distance-based Probabilities to identify spelling errors and their types. The dataset used comes from the results of observations in previous research references and a valid word list using the official Indonesian dictionary from the Language Development and Fostering Agency. The results, in detecting errors in the form of vowels, consonants, and diphthongs, the model successfully detected the wrong words with an accuracy of 90.56%. However, the model still gives wrong results in detecting the type of vowel error because it uses probability calculations with a percentage of 26.41% [18].

Then, further research related to spelling correction that applies the Bidirectional Long Short-Term Memory (Bi-LSTM) algorithm and the Multi-Head Attention (MHA) mechanism in detecting and correcting incorrect spelling in formal text documents, such as news or reports in Indonesian. The dataset used in this study is in the form of correct and incorrect sentences. The results of the study show that the Bi-LSTM algorithm provides good performance with an accuracy of 92.26% [19].

The next research is the design of a spelling correction system for Arabic called AraSpell. This system offers a deep learning-based spelling correction method using the seq2seq architecture involving RNN and Transformer. This system utilizes artificial error injection for the automatic labeling mechanism, as well as the Word Error Rate (WER) and Character Error Rate (CER) for evaluation. As a result, this method shows a significant level of effectiveness in reducing spelling errors, making it a relevant reference in applying seq2seq with the attention mechanism in spelling correction tasks [20].

Further research develops a spelling correction system using two approaches, namely rules and deep learning. The deep learning approach applies an encoder-decoder network using LSTM, a variant of RNN such as GRU. The results show good performance with an accuracy of up to 87% [21]. In contrast to this study, our study uses GRU as a simpler and more computationally efficient alternative but is similar to LSTM in processing sequential data. With the addition of an attention mechanism, our proposed GRU approach is expected to increase the model's focus on relevant text sections for spelling correction tasks.

Further research develops automatic spelling correction for traditional Mandarin corpus on the ASR (Automatic Speech Recognition) system. This study uses a seq2seq neural network model with an attention mechanism [22]. GRU was chosen in this study because it is equivalent to LSTM, this is the inspiration for our research. However, the focus of our research is Indonesian which has unique characteristics, such as more complex morphology and different syntactic structures.

To support this, we adjust the model architecture, including embedding dimensions and other parameters that are adjusted to our dataset. Therefore, our research not only changes the existing approach but also extends its application in the Indonesian language context, providing new contributions to deep learning-based spelling correction.

The next research builds a seq2seq model for spelling correction in Turkish. This research uses an LSTM architecture with an attention mechanism that allows the decoder to focus on certain parts of the input during the decoding process, this helps in overcoming problems when dealing with long input sequences [23]. The study has been shown to show good performance in correcting spelling errors with an accuracy increase of 2.1% compared to the existing Turkish spelling correction system.

Another study related to the use of LSTM in the seq2seq model is a study that examines spelling correction in Turkish with a focus on errors made by humans [24]. The study showed a performance increase of 5% compared to the existing Turkish spelling correction system, so the results are proven to be good. As a new contribution, we replace LSTM with GRU because it is more efficient in terms of computation but still maintains good performance, while still implementing the attention mechanism with adjustments to the model architecture.

III. RESEARCH METHODS

A. Sequence-to-Sequence Algorithm

The Sequence-to-Sequence (Seq2Seq) algorithm is a powerful framework widely utilized in various applications, particularly in natural language processing, speech synthesis, and predictive modeling [25]. This approach typically involves an encoder that processes input sequences and a decoder that generates output sequences, effectively capturing the relationships within the data.

The Seq2Seq model consists of two main components: the encoder, which compresses the input sequence into a fixed-size context vector, and the decoder, which generates the output sequence from this vector [26]. RNN and LSTM: Recurrent Neural Networks (RNNs) and Long Short-Term Memory (LSTM) networks are commonly used due to their ability to handle variable-length sequences and capture temporal dependencies [26].

Seq2Seq models enhance chatbot interactions by predicting future dialogue states, and improving coherence and engagement through reinforcement learning techniques [27]. In voice conversion, Seq2Seq algorithms can maintain prosodic features while converting speech, utilizing adversarial learning for better output quality [28]. In remaining useful life (RUL) predictions, Seq2Seq frameworks can process entire time series data, improving prediction accuracy and efficiency [29]. While Seq2Seq models have shown significant promise, challenges remain, such as the need for large datasets and the complexity of aligning input-output sequences. These issues highlight the ongoing need for advancements in model efficiency and data handling techniques.

B. CRISP-DM

In this study, the CRISP-DM data mining methodology is applied as a general approach in solving problems. The CRISP-DM methodology consists of six stages, namely Business Understanding, Data Understanding, Data Preparation, Modeling, Evaluation, and Deployment [30]. Figure 1 provides six stages of the CRISP-DM process.

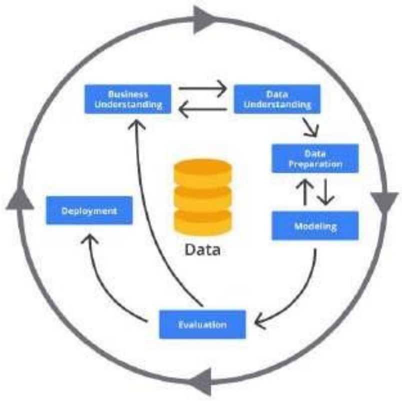

Fig. 1. CRISP-DM Methodology

a. *Business Understanding*, the application of data mining in this study is related to the process of correcting non-standard or slang language in Indonesian language texts. Data obtained through the data crawling process in application X. The purpose of this study is to identify and correct non-standard or slang language errors that often occur and develop effective methods to overcome them.
b. *Data Understanding*, at this stage, an understanding of the data that has been collected from the crawling results on the X application is carried out. The data collected is in the form of tweets with non-standard or slang words as many as 520 data. In addition, there is additional data in the form of a dictionary containing pairs of slang words and standard words that are made manually as a reference for the correction process.
c. *Data Preparation*, in this stage, namely data preparation. This process includes data cleaning, such as removing irrelevant elements, removing symbols, numbers, and inappropriate text. In addition, the process of checking and removing duplicate sentences from the crawled data is carried out. Finally, data transformation is carried out, such as the tokenization process.
d. *Modelling*, at this stage, the model is built using the sequence-to-sequence method with the GRU architecture with the addition of an attention mechanism to handle the problem of spelling correction in Indonesian texts. This process involves training the model using previously prepared data.
e. *Evaluation*, the evaluation stage is carried out to assess the performance of the model that has been built. This process includes accuracy testing to measure how well the model is in correcting sentences containing non-standard or slang words.
f. *Deployment*, this study only focuses on model building without covering implementation or deployment. Therefore, this study produces a model that has been built along with an analysis of model performance using sequence-to-sequence with additional GRU and attention mechanisms.

IV. RESULT AND DISCUSSION

The focus of this research is to build a model that is able to correct the spelling of non-standard words or slang in Indonesian. The dataset used was obtained through crawling techniques from the X application or Twitter, with a total of 535 tweets or data about religious content. After that, the data preprocessing stage was carried out with processes such as changing the text to lowercase, removing URL links, removing punctuation except for spaces, separating words connected with hyphens to make them easier to analyze, removing numbers because they will not be used during the analysis process, normalizing the text by reducing excessive repetition of letters in words, removing emoticons, and tokenizing (Figure 2).

	full_text	preprocess
0	Oktober 7 2024. Satu tahun persis mengedukasi ...	oktober satu h persis mengedukasi diri sendiri...
1	Dari gesture yang bercanda ria video anak-anak...	dari gesture yang bercanda ria video anak a sm...
2	Terima kasih Palestina!! Walaupun kondisi lagi...	terima s palestina walaupun kondisi lagi sulit...
3	https://t.co/tL558Plyiv wst 🇵🇸 Km mau dibantu ...	wst km mau dibantu apa banyak org siap bergera...
4	Selamat datang dan belajar di Perguruan Tinggi...	selamat datang p belajar di perguruan tinggi m...
530	@felixsiauw Makanya orang tuamu jangan banyak ...	makanya orang tuamu jangan banyak tingkah mau ...
531	@felixsiauw Pesantren enggak pernah ngaku usta...	pesantren enggak pernah ngaku ustad cuma karen...
532	@felixsiauw penumpang gelap kenapa ya ngoceh m...	penumpang gelap kenapa ya ngoceh mulu belum te...
533	@felixsiauw Wooiiii...org tua malin kundang ja...	wooiorg tua malin kundang jangan disamakn dg o...
534	@felixsiauw Orang tua seharusnya kasih contoh ...	orang tua seharusnya kasih contoh yang baik da...

Fig. 2. Data Pre-processing Result

After the data preprocessing stage is complete, the next stage is to normalize the text by replacing non-standard or slang words with standard words using a dictionary that has been created by yourself in CSV file format containing pairs of standard words and non-standard or slang words in Indonesian totaling 527 data (example available in Figure 3).

	salah	benar
0	dlm	dalam
1	jg	juga
2	bikin	buat
3	yaa	ya
4	lagiii	lagi

Fig. 3. Slang Dictionary

In the process, non-standard words or slang will be detected and then replaced with standard words. The results of the normalization in Figure 4 will be stored in a new column so that the difference before and after text normalization is visible. In order to see a clearer comparison, the text normalization results are stored in a new column called target_text, while the text before normalization is stored in the input_text column.

	full_text	preprocess	normalized
0	Oktober 7 2024. Satu tahun persis mengedukasi ...	oktober satu h persis mengedukasi diri sendiri...	oktober satu h persis mengedukasi diri sendiri...
1	Dari gesture yang bercanda ria video anak-anak...	dari gesture yang bercanda ria video anak a sm...	dari gesture yang bercanda ria video anak a sm...
2	Terima kasih Palestina!! Walaupun kondisi lagi...	terima s palestina walaupun kondisi lagi sulit...	terima s palestina walaupun kondisi lagi sulit...
3	https://t.co/tL558Plyiv wst 🇵🇸 Km mau dibantu ...	wst km mau dibantu apa banyak org siap bergera...	wst kamu mau dibantu apa banyak orang siap ber...
4	Selamat datang dan belajar di Perguruan Tinggi...	selamat datang p belajar di perguruan tinggi m...	selamat datang p belajar di perguruan tinggi m...

Fig. 4. Text Normalization Result

Next, when the text normalization process is complete, the next step is to build a model. The first step is to find unique characters in the input_text and target_text columns so that the data is easy to process. Then, the collected characters are counted and then sorted in the form of a list and stored in input_characters and target_characters. To determine the size of the data to be processed, it is necessary to record the maximum length of the input and output text, which is then stored in max_encoder_seq_length and max_decoder_seq_length. The next step is to map characters to indexes. Then do one hot encoding, which is to convert text into a number or binary vector containing the numbers 0 and 1. Figure 5 provides the numerical form of text data.

```
Number of samples: 534
Number of unique input tokens: 27
Number of unique output tokens: 27
Max sequence length for inputs: 285
Max sequence length for outputs: 348
Encoder input data shape: (534, 285, 27)
Decoder input data shape: (534, 348, 27)
Decoder target data shape: (534, 348, 27)
```

Fig. 5. Representation of Text Data in Numerical Form

The second step is to build a sequence-to-sequence model with a GRU layer and additional attention mechanisms. The first stage sets hyperparameters containing latent_dim with a dimension value of 128, dropout_rate with a value of 0.3, and learning_rate of 0.001. Then prepare and organize the encoder and decoder sections as the core part of the model in processing data by combining GRU. Then add an attention mechanism feature to improve the decoder's capabilities. Furthermore, combine the decoder results with attention which will then be processed through a dense layer. In this process, the model uses the Adam optimizer and categorical cross-entropy. Figure 6 presents the architecture of Sequence-to-sequence model.

Model: "functional"

Layer (type)	Output Shape	Param #	Connected to
encoder_inputs (InputLayer)	(None, None, 27)	0	-
decoder_inputs (InputLayer)	(None, None, 27)	0	-
encoder_gru (GRU)	[(None, None, 128), (None, 128)]	60,288	encoder_inputs[0][0]
decoder_gru (GRU)	[(None, None, 128), (None, 128)]	60,288	decoder_inputs[0][0], encoder_gru[0][1]
attention_layer (Attention)	(None, None, 128)	0	decoder_gru[0][0], encoder_gru[0][0]
decoder_concat_input (Concatenate)	(None, None, 256)	0	decoder_gru[0][0], attention_layer[0][0]
decoder_dense (Dense)	(None, None, 27)	6,939	decoder_concat_input[…

Total params: 127,515 (498.11 KB)
Trainable params: 127,515 (498.11 KB)
Non-trainable params: 0 (0.00 B)

Fig. 6. Model Architecture

The final step is to train the model by setting the batch size to 16, epochs to 50, and validation split to 20%. As a result, the model achieved an accuracy of 0.7516, with a loss of 0.8476, a val_accuracy of 0.7129, and a val_loss of 0.9526.

```
27/27 ━━━━━━━━━━━━━━━━━━━━ 20s 539ms/step - accuracy: 0.7306 - loss: 0.9000 - val_accuracy: 0.7986 - val_loss: 0.6615
Epoch 40/50
27/27 ━━━━━━━━━━━━━━━━━━━━ 23s 646ms/step - accuracy: 0.7417 - loss: 0.8654 - val_accuracy: 0.8012 - val_loss: 0.6582
Epoch 41/50
27/27 ━━━━━━━━━━━━━━━━━━━━ 20s 631ms/step - accuracy: 0.7334 - loss: 0.8873 - val_accuracy: 0.8016 - val_loss: 0.6569
Epoch 42/50
27/27 ━━━━━━━━━━━━━━━━━━━━ 19s 596ms/step - accuracy: 0.7326 - loss: 0.8880 - val_accuracy: 0.8003 - val_loss: 0.6554
Epoch 43/50
27/27 ━━━━━━━━━━━━━━━━━━━━ 22s 637ms/step - accuracy: 0.7361 - loss: 0.8802 - val_accuracy: 0.8013 - val_loss: 0.6564
Epoch 44/50
27/27 ━━━━━━━━━━━━━━━━━━━━ 22s 688ms/step - accuracy: 0.7258 - loss: 0.9205 - val_accuracy: 0.8023 - val_loss: 0.6537
Epoch 45/50
27/27 ━━━━━━━━━━━━━━━━━━━━ 22s 745ms/step - accuracy: 0.7338 - loss: 0.8948 - val_accuracy: 0.8005 - val_loss: 0.6551
Epoch 46/50
27/27 ━━━━━━━━━━━━━━━━━━━━ 17s 630ms/step - accuracy: 0.7407 - loss: 0.8682 - val_accuracy: 0.8032 - val_loss: 0.6521
Epoch 47/50
27/27 ━━━━━━━━━━━━━━━━━━━━ 20s 608ms/step - accuracy: 0.7363 - loss: 0.8832 - val_accuracy: 0.8040 - val_loss: 0.6516
Epoch 48/50
27/27 ━━━━━━━━━━━━━━━━━━━━ 20s 579ms/step - accuracy: 0.7366 - loss: 0.8798 - val_accuracy: 0.8040 - val_loss: 0.6490
Epoch 49/50
27/27 ━━━━━━━━━━━━━━━━━━━━ 20s 567ms/step - accuracy: 0.7357 - loss: 0.8936 - val_accuracy: 0.8052 - val_loss: 0.6478
Epoch 50/50
27/27 ━━━━━━━━━━━━━━━━━━━━ 20s 539ms/step - accuracy: 0.7276 - loss: 0.9128 - val_accuracy: 0.8036 - val_loss: 0.6480
```

Fig. 7. Training Result

In this study, the model built using sequence-to-sequence with additional GRU and attention mechanism showed consistent accuracy results during the training process, with a large accuracy of 0.7276 at epoch 50. In addition to accuracy, the loss decrease also gradually showed a decrease from 0.9128 to 0.6480. The model evaluation results gave a loss value of 0.7682, which shows that the model is able to maintain good performance on the training data.

The use of the attention mechanism in the sequence-to-sequence method helps the model to focus more on relevant information during data processing, thus increasing the model's performance. In addition, the use of GRU itself also helps the model to be more efficient in capturing relationships between data. While testing the model that has been built previously, it gives good results. As in the following picture.

The test results above show that the model successfully corrects non-standard or slang words in the inputted sentences. Thus, the results of the study show good results with the sequence-to-sequence model with a combination of GRU and attention mechanisms in the task of correcting non-standard or slang words in application X.

V. CONCLUSION

In this study, it can be concluded that the spelling correction model by implementing the sequence-to-sequence algorithm by combining GRU and attention mechanisms is able to detect slang or non-standard words and change them into standard Indonesian words. The model provides good performance with an accuracy of 76.82% with a number of epochs of 50. This research is particularly significant for applications in religious content, where spelling accuracy is crucial to maintaining the integrity and clarity of spiritual messages. By ensuring that non-standard or slang words are corrected, this model can support the dissemination of religious texts and sermons in a more precise and comprehensible manner, reducing the risk of misinterpretation. The suggestions for future research are to add more and varied datasets and to test the model with more complex data.

REFERENCES

[1] M. Qulub, R. Hammad, and P. Irfan, "Improvement of Spelling Correction Accuracy in Indonesian Language through the Application of Hamming Distance Method," 2023. [Online]. Available: http://jurnal.polibatam.ac.id/index.php/JAIC

[2] F. Lubis *et al.*, "Analysis of Methods to Correct Indonesian Language Spelling Errors in Thesis Writing Among Students of State University of Medan," *EDUCTUM: Journal Research*, vol. 2, no. 6, pp. 5–9, Dec. 2023, doi: 10.56495/ejr.v2i6.407.

[3] A. P. S and W. J. Hartono, "Pentingnya Penggunaan Bahasa Indonesia di Perguruan Tinggi," *Jotika Journal in Education*, vol. 2, no. 2, pp. 57–64, Feb. 2023, doi: 10.56445/jje.v2i2.84.

[4] R. Devianty, "Penggunaan Kata Baku dan Tidak Baku dalam Bahasa Indonesia," *Jurnal Pendidikan Bahasa Indonesia)*, vol. 1, no. 2, pp. 121–132, 2021, [Online]. Available: http://jurnaltarbiyah.uinsu.ac.id/index.php/eunoia/index

[5] A. M. B. Ledjap, F. P. Rochmawati, D. A. E. Marsanda, and A. P. Sari, "Pemanfaatan Natural Language Processing Untuk Pengecekan Ejaan

Sesuai KBBI," *Jurnal Mahasiswa Teknik Informatika*, vol. 3, no. 2, pp. 46–56, Oct. 2024, doi: 10.35473/jamastika.v3i2.3255.
[6] M. H. Ferdiansyah and I. K. D. Nuryana, "Analisis Perbandingan Metode Burkhard Keller Tree dan SymSpell dalam Spell Correction Bahasa Indonesia," *Journal of Informatics and Computer Science (JINACS)*, pp. 305–313, Jan. 2023, doi: 10.26740/jinacs.v4n03.p305-313.
[7] K. S. Nugroho, I. Akbar, A. N. Suksmawati, and Istiadi, "Deteksi Depresi dan Kecemasan Pengguna Twitter Menggunakan Bidirectional LSTM," Jan. 2023.
[8] M. Khadapi and V. M. Pakpahan, "Analisis Sentimen Berbasis Jaringan LSTM dan BERT terhadap Diskusi Twitter tentang Pemilu 2024," *JUKI: Jurnal Komputer dan Informatika*, vol. 6, no. 2, pp. 130–137, 2024.
[9] M. A. Syifa and D. R. S. Saputro, "Stance Detection Dengan Algoritme Gated Recurrent Unit (GRU)," in *Prosiding Seminar Nasional Matematika dan Statistika*, 2023, pp. 267–275.
[10] A. N. Khasanah and M. Hayaty, "Abstractive-based Automatic Text Summarization on Indonesian News Using GPT-2," *JURTEKSI (Jurnal Teknologi dan Sistem Informasi)*, vol. 10, no. 1, pp. 9–18, Dec. 2023, doi: 10.33330/jurteksi.v10i1.2492.
[11] M. H. Mori Hovipah, E. Hearani, J. Jasril, and F. Syafria, "Klasifikasi Clickbait Menggunakan Transformers," *Jurnal CoSciTech (Computer Science and Information Technology)*, vol. 4, no. 1, pp. 172–181, Apr. 2023, doi: 10.37859/coscitech.v4i1.4713.
[12] A. Bahari and K. E. Dewi, "Peringkasan Teks Otomatis Abstraktif Menggunakan Transformer Pada Teks Bahasa Indonesia," *Komputa : Jurnal Ilmiah Komputer dan Informatika*, vol. 13, no. 1, pp. 83–91, Apr. 2024, doi: 10.34010/komputa.v13i1.11197.
[13] M. I. Yahya, Arini, V. Amrizal, I. M. M. Matin, and D. Khairani, "Spelling Correction Using the Levenshtein Distance and Nazief and Adriani Algorithm for Keyword Search Process Indonesian Qur'an Translation," in *2022 Seventh International Conference on Informatics and Computing (ICIC)*, IEEE, Dec. 2022, pp. 01–06. doi: 10.1109/ICIC56845.2022.10006994.
[14] A. Musyafa, Y. Gao, A. Solyman, C. Wu, and S. Khan, "Automatic Correction of Indonesian Grammatical Errors Based on Transformer," *Applied Sciences*, vol. 12, no. 20, p. 10380, Oct. 2022, doi: 10.3390/app122010380.
[15] Muhammad zaky ramadhan and Kemas Muslim Lhaksmana, "Improving Document Retrieval with Spelling Correction for Weak and Fabricated Indonesian-Translated Hadith," *Jurnal RESTI (Rekayasa Sistem dan Teknologi Informasi)*, vol. 4, no. 3, pp. 551–557, Jun. 2020, doi: 10.29207/resti.v4i3.1913.
[16] R. Martin, D. S. Naga, and V. C. Mawardi, "Penggunaan Spelling Correction Dengan Metode Peter Norvig dan N-Gram," *Jurnal Ilmu Komputer dan Sistem Informasi*, vol. 9, no. 1, p. 175, Jan. 2021, doi: 10.24912/jiksi.v9i1.11591.
[17] H. Sujaini, H. Muhardi, and J. H. Simanjuntak, "Aplikasi Pengoreksi Ejaan (Spelling Correction) pada Naskah Jurnal Bidang Informatika dengan N-Gram dan Jaro-Winkler Distance," *Jurnal Edukasi dan Penelitian Informatika (JEPIN)*, vol. 8, no. 2, p. 235, Aug. 2022, doi: 10.26418/jp.v8i2.48092.
[18] E. Erwina, T. Tommy, and M. Mayasari, "Indonesian Spelling Error Detection and Type Identification Using Bigram Vector and Minimum Edit Distance Based Probabilities," *SinkrOn*, vol. 6, no. 1, pp. 183–190, Nov. 2021, doi: 10.33395/sinkron.v6i1.11224.
[19] Y. Yanfi, H. Soeparno, R. Setiawan, and W. Budiharto, "Multi-Head Attention Based Bidirectional LSTM for Spelling Error Detection in the Indonesian Language," *IEEE Access*, vol. 12, pp. 188560–188571, 2024, doi: 10.1109/ACCESS.2024.3422318.
[20] M. Salhab and F. Abu-Khzam, "AraSpell: A Deep Learning Approach for Arabic Spelling Correction," Jun. 02, 2023. doi: 10.21203/rs.3.rs-2974359/v1.
[21] Sa. Kasmaiee, Si. Kasmaiee, and M. Homayounpour, "Correcting spelling mistakes in Persian texts with rules and deep learning methods," *Sci Rep*, vol. 13, no. 1, p. 19945, Nov. 2023, doi: 10.1038/s41598-023-47295-2.
[22] Y.-C. Chao and C.-H. Chang, "Automatic Spelling Correction for ASR Corpus in Traditional Chinese Language using Seq2Seq Models," in *2020 International Computer Symposium (ICS)*, IEEE, Dec. 2020, pp. 553–558. doi: 10.1109/ICS51289.2020.00113.
[23] O. Büyük, "Context-Dependent Sequence-to-Sequence Turkish Spelling Correction," *ACM Transactions on Asian and Low-Resource Language Information Processing*, vol. 19, no. 4, pp. 1–16, Jul. 2020, doi: 10.1145/3383200.
[24] O. Büyük and L. M. Arslan, "Learning from mistakes: Improving spelling correction performance with automatic generation of realistic misspellings," *Expert Syst*, vol. 38, no. 5, Aug. 2021, doi: 10.1111/exsy.12692.
[25] E. Egonmwan and Y. Chali, "Transformer and seq2seq model for Paraphrase Generation," in *Proceedings of the 3rd Workshop on Neural Generation and Translation*, Stroudsburg, PA, USA: Association for Computational Linguistics, 2019, pp. 249–255. doi: 10.18653/v1/D19-5627.
[26] Y. Zhang, "Encoder-decoder models in sequence-to-sequence learning: A survey of RNN and LSTM approaches," *Applied and Computational Engineering*, vol. 22, no. 1, pp. 218–226, Oct. 2023, doi: 10.54254/2755-2721/22/20231220.
[27] J. Ismail, A. Ahmed, and E. ouaazizi Aziza, "Improving a Sequence-to-sequence NLP Model using a Reinforcement Learning Policy Algorithm," in *Artificial Intelligence, Soft Computing and Applications*, Academy and Industry Research Collaboration Center (AIRCC), Dec. 2022, pp. 221–231. doi: 10.5121/csit.2022.122317.
[28] F. Bous, L. Benaroya, N. Obin, and A. Roebel, "Voice Reenactment with F0 and timing constraints and adversarial learning of conversions," Oct. 2021.
[29] J. Li, R. Chen, and X. Huang, "A sequence-to-sequence remaining useful life prediction method combining unsupervised LSTM encoding-decoding and temporal convolutional network," *Meas Sci Technol*, vol. 33, no. 8, p. 085013, Aug. 2022, doi: 10.1088/1361-6501/ac632d.
[30] M. A. Hasanah, S. Soim, and A. S. Handayani, "Implementasi CRISP-DM Model Menggunakan Metode Decision Tree dengan Algoritma CART untuk Prediksi Curah Hujan Berpotensi Banjir," *Journal of Applied Informatics and Computing*, vol. 5, no. 2, pp. 103–108, Oct. 2021, doi: 10.30871/jaic.v5i2.3200.