ISSN 2085-4552

U-TAPIS Sal-Tik : Typing Error Detection Using Random Forest Algorithm

Bryan Glennardy[1], Marlinda Vasty Overbeek[2], Niknik Mediyawati[3], Samiaji Bintang Nusantara[4], Rudi Sutomo[5]

[1,2] Informatics Study Program, Universitas Multimedia Nusantara, Tangerang, Indonesia
[1]bryan.glennardi@student.umn.ac.id, [2]marlinda.vasty@umn.ac.id
[3,4] Digital Journalism Study Program, Universitas Multimedia Nusantara, Tangerang, Indonesia
[3]niknik@umn.ac.id, [4]samiaji.bintang@umn.ac.id
[5] Information System Study Program, Universitas Multimedia Nusantara, Tangerang, Indonesia
[5]rudi.sutomo@umn.ac.id

Accepted 27 March 2024
Approved 6 June 2024

***Abstract*—The result of this study indicate that the development of technology in the field of journalism has grown very rapidly. However, there are still frequent deviations in language usage on online news portal, particularly in terms of spelling and word usage. Spelling mistake in news articles can cause the information to be unclear and ambiguous. Based on these issues, a study was conducted to create a model to detects typos in Bahasa Indonesia. This model was conducted to create model to detect typos in Bahasa Indonesia. This model was created using the Random Forest algorithm. The Random Forest algorithm works by constructing several decision decisions tree and then combining the decisions from each tree, taking the majority vote from the predictions of each tree to produce stable and accurate prediction. The result of this study show that the model achieved an accuracy of 100%. However, it should be noted that this 100% result means that the model is able to detect words that are already present in the dataset. Based on the evaluation results, since the detected words were contained in the dataset, the accuracy reported is 100%. The model successfully detects typos in Tribunnews news articles.**

***Index Terms*—detection; news articles; random forest algorithm; type error**

I. INTRODUCTION

Nowadays, technology has developed very rapidly. Technology comes with the aim to help make it easier for humans to do their daily work. One field that has experience significant technological development is journalism. In the past, news media primarily relied on the print media to disseminate news, requiring people to read newspapers or other printed publications. However, with advancements in technology, news can now be spread online, including through websites. This allows people to access news easily and freely, anytime and anywhere. News itself a report about a recent event or the latest information regarding an event. In other words, news consists of interesting facts or important information conveyed to the public through various media channels [1].

In journalism, language is used to convey precise and accurate information to the public through mass media [2]. One of the functions of language is as a tool for communication so that the use of language, especially Indonesian, must use good and correct spelling based on existing rules. Spelling itself is a procedure for using words, sentences, and punctuation both in oral and written form [3].

Technological developments in the field of journalism have developed very rapidly, but there are still frequent deviations from the language on online news portals. Usually, this can be seen from the aspect of using spelling and words that are not in accordance with the established writing rules and this is not uncommon in online news portals [4]. This can occur due to the speed of the news dissemination itself, which usually causes errors when typing the news and also when it is in the editing process. Spelling errors that occur in the news can cause the information contained in the news to be unclear and ambiguous [5]. In research that has been conducted related to the analysis of language errors on online news portals with a case study of the Suara.com online news portal and research related to word writing and punctuation errors in online news, it is concluded that there are still many typing errors in the news [6] [7].

Based on the existing problems, research was conducted to detect type error on online news portals which in this study will use news from the Tribunnews news portal namely U-Tapis Sal-Tik. This is because there are requests from partners regarding the creation of several modules and one of them is a module for detecting type error. There is a reason why Tribunnews wants to make this module, namely because Tribunnews uploads 3000 to 5000 articles a day and each reporter is required to write 20 articles. With such a large amount of production, the possibility of errors in publishing is great. Language errors in news can reduce the credibility and public trust in the information presented by the media. Tribunnews itself is the number one online news portal in Indonesia

managed by PT Tribun Digital Online. Tribunnews.com has a network that has spread throughout Indonesia called Tribun Network [8].

U-tapis already conducted before in 2020 until 2023 [9-13]. The research that has been done is the detection of misspelled words using the Jaccard similarity algorithm where this research found an accuracy rate of 93.2% [14]. Then there is also research on detecting the use of conjunctions using the cosine similarity algorithm where this research found an accuracy rate of 92.2% [15].

This type error detection research will use the random forest algorithm. The reason why this research uses the random forest algorithm is because random forest is one of the algorithms in ensemble learning that is used to classify large data sets. In addition, there are several advantages of this algorithm, such as having good accuracy results, relatively strong against outliers and noise, simple and easy to parallelize [16]. In several studies that have been conducted related to text classification also show that the random forest algorithm has a fairly high accuracy rate such as sentiment analysis of the Ruangguru application which has an accuracy rate of 97.16% [17] and fake news detection which has an accuracy rate of 84% [18].

The detection model created can detect by learning from new data that has been trained so that the model can increase its knowledge so that it does not detect by matching words or string matching. In addition, the model does not only detect type error, but it can also provide suggestions for word correction from type error that have been previously detected by the model.

By doing this research, it is hoped that it can detect type error properly and correctly so that it can help news writers in checking type error in the news articles written. Then also with the success of this research, it is hoped that the information contained in the news articles written can be conveyed properly to readers.

II. METHODS

A. Type Error

Type error is an error that occurs during the process of typing text and can change the meaning of a word and even the meaning of a sentence [18]. The occurrence of type error can cause information not to be conveyed properly to the reader and can also cause misunderstanding of the information provided to the reader.

B. Online News Portal

Online News Portal is a site or web page that contains various types of news, such as politics, economics, social, culture to entertainment that is hard news and soft news [19].

C. Text Preprocessing

Text Preprocessing is a process that aims to select text data so that it will be more structured [20]. In text preprocessing, there are several steps that need to be done, including the following.

- Case Folding, is a process that aims to convert all characters into lowercase letters and also eliminate characters that do not include letters [21].
- Tokenizing, is a process for breaking sentences into words [22].
- Filtering, is an advanced stage of tokenizing where this stage is used to select important words from the tokenizing results that have been done previously by removing words that cannot be used or can also be called stopwords [23].
- Stemming, the process of changing a word into its base form [23].

D. Decision Tree

Decision tree is a method for classification using a representation of a tree structure where each node represents the attribute, then the branches represent the value of the attribute and the leaves represent the class [24].

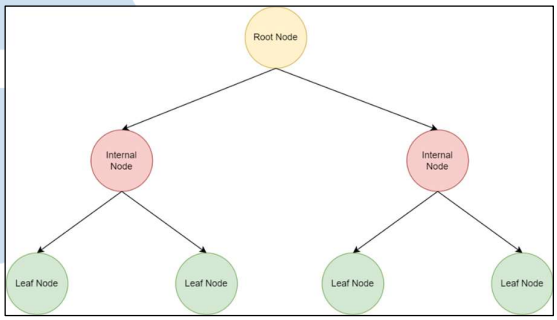

Fig. 1. Decision tree diagram [30]

Fig 1 is a diagram of a decision tree. As can be seen in Figure 1, the nodes in the decision tree are divided into three types, including the following [24].

- Root Node is the node located at the very top where this node has no input and allows it to have no output but it is also possible to have more than one output.
- Internal Node is a branching node. This node has one input and has an output. In the internal node, entropy calculation is performed. Entropy serves to measure the level of uncertainty or impurity of the attribute [25]. After determining the entropy, the calculation of the gain of each attribute is carried out. Gain is a value used to select attributes to generate new nodes. The following is the formula for calculating entropy and gain [25].

$$\boldsymbol{Entropy(S)} = \sum_{i=1}^{n} - \boldsymbol{pi} * \boldsymbol{log_2} * \boldsymbol{pi} \quad (1)$$

Where S is a set of cases, n is a number of S partitions and Pi is a proportion Si to S. And gain formula is

$$\boldsymbol{Gain(S,A) = Entropy(S) - \sum_{i=1}^{n} \frac{|Si|}{|S|} * Entropy(Si)} \quad (2)$$

- Leaf Node or also known as Terminal Node is the final node where this node only has one input and has an output in the form of a decision or final prediction

E. Ensemble Learning

Ensemble Learning is one of the machine learning paradigms where multiple models (base-models) are trained and combined to get better results[26]. In ensemble learning there are three models, including the following[26].

- Bagging is one method of ensemble learning which uses one type of base model by training in parallel and independently on each base model, then combining them to get the best results. Random Forest algorithm is one of the algorithms included in the bagging model.
- Boosting is one of the methods of ensemble learning which uses one type of base model where training is done sequentially and the results of each base model depend on the results of the previous base model. Adaptative boosting algorithm (AdaBoost) is one of the algorithms included in the boosting model.
- Stacking is one of the methods of ensemble learning in which training uses several base models which are then carried out in parallel and independently on each base model and then uses an algorithm derived from other learning to produce output from the combination of each base model. Blending algorithm is one of the algorithms included in the stacking model

F. Random Forest

Random forest a data mining algorithm method used to classify a dataset [27]. The way this random forest algorithm works is to build several decision trees then combine the decisions of each tree that has been built and take the most votes from the predictions of each tree so that later it will produce stable and accurate predictions [28].

G. Confusion Matrix

Confusion matrix is a table used to measure the performance of machine learning. In the confusion matrix table, there are four variables, including True Positive (TP) is data that is positive and predicted to be true positive by the system, False Positive (FP) is data that is negative but predicted to be positive by the system, False Negative (FN) is data that is positive but predicted to be negative by the system, and True Negative (TN) is data that is negative and predicted to be true negative by the system [29]. Table 1 is a form of confusion matrix table.

TABLE I. CONFUSION MATRIX

Actual Value	Predicted Value	
	1	0
1	TP	FN
0	FP	TN

Confusion matrix is used to calculate accuracy, precision, recall and F1 score. The following is the method used in the confusion matrix to calculate these four things [30].

- Accuracy is a description of how accurate the system that has been created in performing the classification correctly. The following is the formula for how to find the accuracy value.

$$\boldsymbol{Accuracy} = \frac{\boldsymbol{TP+TN}}{\boldsymbol{TP+FP+FN+TN}} \quad (3)$$

- Precision is a description of the accuracy of the requested data with the prediction results given by the system. The following is the formula for how to find the precision value.

$$\boldsymbol{Precision} = \frac{\boldsymbol{TP}}{\boldsymbol{TP+FP}} \quad (4)$$

- Recall is the level of success of a system in finding back information. The following is a formula for how to find the recall value.

$$\boldsymbol{Recall} = \frac{\boldsymbol{TP}}{\boldsymbol{TP+FN}} \quad (5)$$

- F1 Score is the average value of precision and recall. Here is the formula for how to find the F1 Score value.

$$F1 = \frac{2*Recall*Precision}{Recall+Precisio} \quad (6)$$

III. RESULT AND DISCUSSIONS

A. Interface Display

The following is the interface of the type error detection model. Figure 2 is the home page of the website.

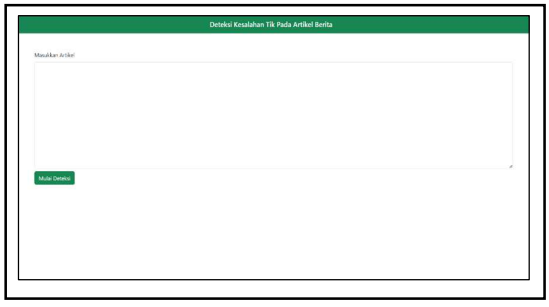

Fig. 2. Home Page Display

Fig 2 is the home page of the website. This page serves as a place for users to enter news articles to be

detected. Users can enter the news article they want to detect through the text area on the home page and then press the *'Mulai Deteksi'* button to start the detection process.

Fig. 3. Results Page Display

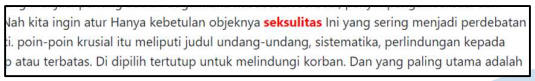

Fig. 4. Display of Detected and Highlighted Type Error Words on the Result Page

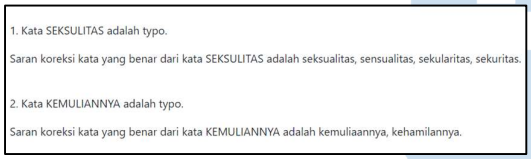

Fig. 5. Display a List of Type Error Words along with Suggestions for Correct Word Correction on the Result Page

Fig 3 is an outline of the results page when the entered news article has been detected. Then in Fig 4 is a display on the results page where the detected type error words will be marked and also in Fig 5 is a display of the list of detected type error words along with suggestions for correct word correction. list of detected mistyped words along with suggestions for correct word correction.

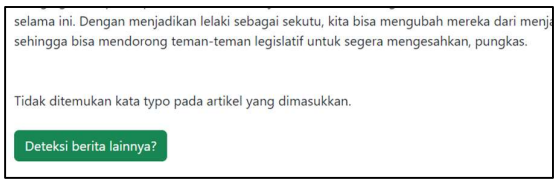

Fig. 6. Display of the Result Page with the Wrong Type Error Word Not Found

Fig 6 is the result page but with different conditions. In Figure 12 no mistyped words are detected so no words are marked and the list of mistyped words is displayed.

B. Testing

At this stage, testing is carried out on the type error detection model that has been created. There are two types of tests carried out, including testing based on the number of news and testing based on the cutoff value.

Testing Based on Number of News

In the first type of test, testing was carried out on the model that had been made with a total of 50 news articles that were entered with different amounts or gradually, namely 20 news to 40 news and to 50 news. Testing by entering the number of news articles in stages aims to see the model's ability to detect tick errors as the number of news increases whether there is a decrease in performance or not as the number of news increases and whether the model can handle larger amounts of data well.

Testing with 20 News

Testing based on the number of news starts with 20 news. Fig 7 is a display of the home page that entered 20 news to be tested.

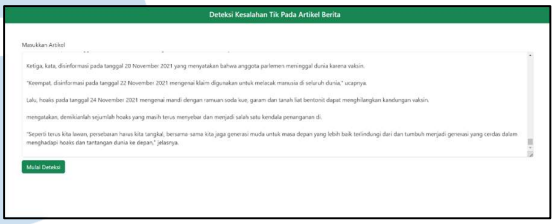

Fig. 7. Home Page Display When 20 News Articles Are Entered

In the 20 news articles entered, there are 3809 words which is the total word result after passing text preprocessing with 29 type error. After detection, 26 type error were detected. Fig 8 is a display of the total type error detected on the results page of the test with 20 news articles.

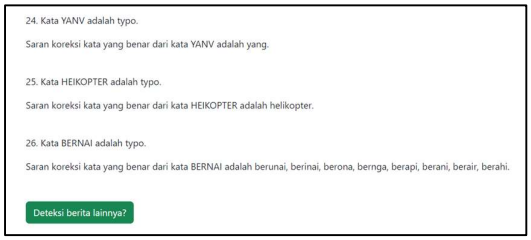

Fig. 8. Display of Test Results With 40 News Articles

Testing with 40 News

The next test was conducted with 40 news articles. Fig 9 is a display of the home page that includes 40 news articles to be tested.

Fig. 9. Home Page Display When 40 News Articles Are Entered

In the 40 news articles entered, there are 6821 words which is the total word result after passing text preprocessing with 61 type error. After detection, there are 48 detected type error. Fig 10 is a display of the total type error detected on the results page of the test with 40 news articles.

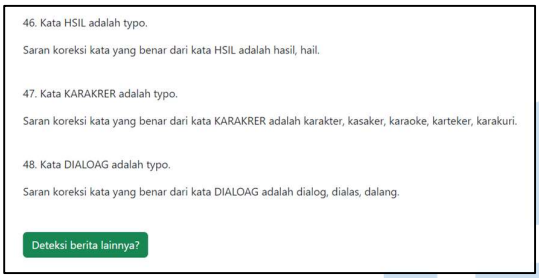

Fig. 10. Display of Test Results With 40 News Articles

Testing with 50 News

The next test was conducted with 50 news articles. Fig 11 is a display of the home page that includes 50 news articles to be tested.

Fig. 11. Home Page Display When 50 News Articles Are Entered

In the 50 news articles entered, there are 8535 words which is the total result of words after passing text preprocessing with 79 mistyped words. After detection, there are 60 detected type error. Fig 12 is a display of the total type error detected on the results page of the test with 50 news articles.

From the test results based on the number of news stories that have been carried out, it can be concluded that the model has successfully performed type error detection. However, there are still some type error words that are not detected. Fig 13 is a comparison graph of testing based on the number of news.

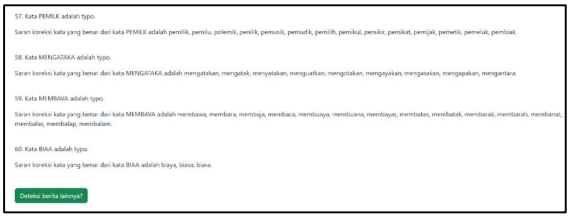

Fig. 12. Display of Test Results With 50 News Articles

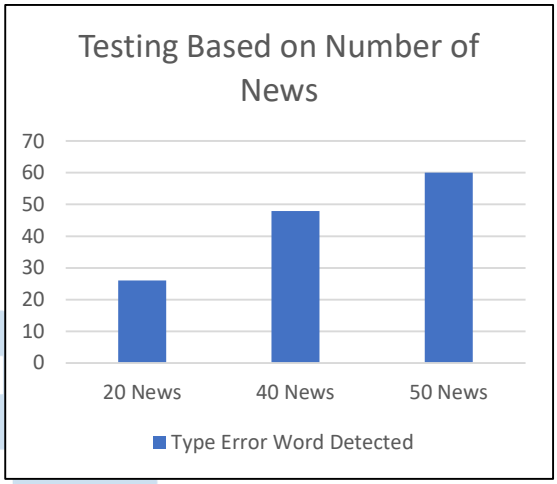

Fig. 13. Comparison Chart of The Test Based on The Number of News

Testing Based on Cutoff Value

In the second type of test, tests were conducted using different cutoff values. This aims to see the word correction generated from mistyped words with different cutoff values.

Testing with Cutoff Value 0.85

The test was conducted by giving a cutoff value of 0.85. Fig 14 is the display when a news article is entered to test with a cutoff of 0.85.

Fig. 14. Home Page View When Tested With a Cutoff Value of 0.85

After testing, in the tested news there is one type error word, namely '*kpeada*' which should be '*kepada*'. However, the result issued is that the word suggestion is not found. Fig 15 is a list of detected mistyped words on the results page with testing using a cutoff value of 0.85.

Fig. 15. Display Page Test Results With a Cutoff Value of 0.85

Testing with Cutoff Value 0.75

Further testing is done by giving a cutoff value of 0.75. The news used is the same as the news used in Fig 15. After testing using a cutoff value of 0.75, the results of word correction suggestions from the type error word '*kpeada*' were successfully obtained. Fig 16 is a list of type error words detected on the results page by testing using a cutoff value of 0.75.

Fig. 16. Display Page Test Results With a Cutoff Value of 0.75

Testing with Cutoff Value 0.65

Furthermore, testing is done by giving a cutoff value of 0.65. Fig17 is the display when the news article is entered for testing with a cutoff value of 0.65. testing with a cutoff value of 0.65.

Fig. 17. Home Page View When Tested With a Cutoff Value of 0.65

After testing, the tested news contained four mistyped words, namely '*seksulitas*', '*kemuliannya*', '*menberi*', and '*kemanusian*' which should be '*seksualitas*', ' *kemuliaannya*', ' *memberi'*, and ' *kemanusiaan*'. These four misspelled words have successfully obtained word correction suggestions. Fig 18 is a list of mistyped words detected on the results page by testing using a cutoff value of 0.65.

Then testing using the same news with a cutoff value of 0.75 can be used as a comparison of the resulting word correction. Fig 19 is a list of mistyped words detected on the results page by testing using a cutoff value of 0.75.

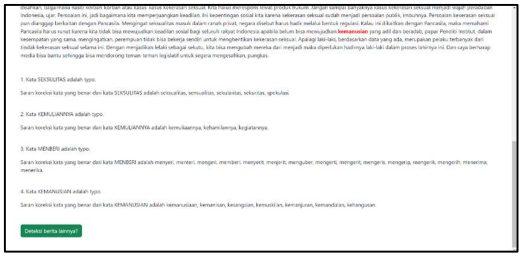

Fig. 18. Display Page Test Results With a Cutoff Value of 0.65

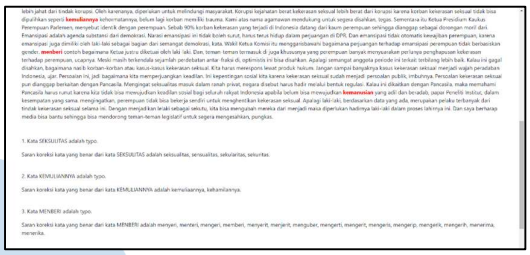

Fig. 19. Display Page The Second News Article Test Result With a Cutoff Value of 0.75

From the results of the tests conducted, the word correction suggestions from the type error words '*seksualitas*', ' *kemuliaannya*' displayed with a cutoff value of 0.75 resulted in a smaller number of correction suggestions than using a cutoff value of 0.65. This shows that the smaller the cutoff value, the more and further the similarity level of the word correction suggestions generated from the detected type error words

C. Evaluation

After conducting the testing stage, the next step is evaluation. In this evaluation stage, calculations are carried out using the confusion matrix. This aims to determine the accuracy, precision, recall, and F1 score of the model. In making the evaluation, the training data is 80% and the test data is 20% of the training data and the accuracy, precision, recall, and F1 score results are 100%. Fig 20 is a classification report from the results of the calculations that have been done.

```
Classification Report:
              precision    recall  f1-score

     correct       1.00      1.00      1.00
   incorrect       1.00      1.00      1.00

    accuracy                           1.00
   macro avg       1.00      1.00      1.00
weighted avg       1.00      1.00      1.00
```

Fig. 20. Classification Report Display of Calculations Performed

IV. CONCLUSION

Based on the research that has been done, it can be concluded that the typing error detection model using the random forest algorithm has been successfully built. From the results of the confusion matrix calculation carried out using 80% of the dataset as training data and 20% of the training data as test data at the evaluation stage, the results of accuracy, precision, recall, and F1 score are 100%. This shows that the model built has been able to detect type error, especially in the 50 Tribunnews news articles entered when testing. But keep in mind that this 100% result is that the model is able to detect words that are already contained in the dataset. Based on the evaluation results that have been carried out, because the detected word is contained in the dataset, the accuracy issued is 100%. However, if it is not contained in the dataset, such as one of the letters is mixed up or deleted, it will allow the word to be recognized by the model which is done by majority vote and will provide correction of the correct word from the word by looking at the level of similarity performed by the get close matches function from the difflib library. The reason for using 20% of the training data as data test data is because when calculating the evaluation, the model becomes very dependent on the dataset so that if in the evaluation calculation there is a word that is not in the dataset, the model will not be able to evaluate it. that is not in the dataset, it will cause the model to be unable to perform the evaluation calculation. perform evaluation calculations.

ACKNOWLEDGMENT

The authors would like to thank Universitas Multimedia Nusantara for supporting this research as well as Tribunews as a partner for this research.

REFERENCES

[1] D. N. Chandra, G. Indrawan and I. N. Sukajaya, "Klasifikasi Berita Lokal Radar Malang Menggunakan Metode Naïve Bayes Dengan Fitur N-Gram," *Jurnal Ilmiah Teknologi dan Informasia ASIA (JITIKA),* vol. 10, no. 1, pp. 11-19, 2016.

[2] Waridah, "Ragam Bahasa Jurnalistik," *Jurnal Simbolika: Research and Learning in Communication Study,* vol. 4, no. 2, pp. 121-129, 2018.

[3] R. Tussolekha, "Kesalahan Penggunaan Ejaan Bahasa Indonesia," *AKSARA Jurnal Bahasa dan Sastra,* vol. 20, no. 1, pp. 35-43, 2019.

[4] N. Faizah and I. S. Ramadhani, "Analisis Kesalahan Berbahasa Pada Penulisan Berita OnlineLiputan6 Edisi 18 Juli 2022," *Jurnal Pendidikan dan Konseling,* vol. 5, no. 1, pp. 850-854, 2023.

[5] F. Achsani, "KESALAHAN BERBAHASA PADA PENULISAN," *SIROK BASTRA,* vol. 8, no. 2, pp. 246-255, 2020.

[6] A. Andriani, D. I. Sari, E. M. Choirunisa, N. P. Ariska and C. Ulya, "ANALISIS KESALAHAN BERBAHASA DALAM TATARAN MORFOLOGI PADA PORTAL BERITA ONLINE SUARA.COM," *Nivedana : Jurnal Komunikasi & Bahasa,* vol. 2, no. 2, pp. 128-139, 2021.

[7] R. R. Apriliana, A. Firdaus and F. Suparman, "KESALAHAN PENULISAN KATA DAN TANDA BACA PADA ONLINE NEWS," *BAHASTRA: Jurnal Pendidikan Bahasa dan Sastra Indonesia,* vol. 5, no. 1, pp. 13-19, 2020.

[8] Marsuki, Rasmila, R. Avindo and D. Safitri, "ANALISIS WEBSITE TRIBUNNEWS MENGGUNAKAN SUS (SYSTEM USABILITY SCALE)," in *Seminar Hasil Penelitian Vokasi (SEMHAVOK)*, Palembang, 2021.

[9] Mediyawati N, Bintang S. Platform Kecerdasan Buatan Sebagai Media Inovatif Untuk Meningkatkan Keterampilan Berkomunikasi: U-Tapis. Pros Semin Nas Pendidik Progr Pascasarj Univ PGRI Palembang 21 Augustus 2021. 2021;69–79

[10] Niknik M, Sutomo R, Nusantara, Samiaji Bintang Overbeek MV. U-Tapis Web: An Automated Indonesian News Text Error Detection System. J Syst Manag Sci. 2024;13(1):1–20

[11] Saputra AKB, Overbeek MV. Harnessing Long Short-Term Memory Algorithm for Enhanced Di-Di Word Error Detection and Correction. In: 1ST INTERNATIONAL CONFERENCE TRACK ON ARTIFICIAL INTELLIGENCE HORIZONS & SOCIETY. 2024

[12] Dwitya NR, Overbeek MV. Development of Detection and Correction of Errors in Spelling and Compound Words Using Long Short-Term Memory. In: 1ST INTERNATIONAL CONFERENCE TRACK ON ARTIFICIAL INTELLIGENCE HORIZONS & SOCIETY. 2024

[13] Siswanto VGA, Overbeek MV. Development of " Kata Terikat " Detection and Writing Errors Correction Using Rabin-Karp and Random Forest Algorithm. In: 1ST INTERNATIONAL CONFERENCE TRACK ON ARTIFICIAL INTELLIGENCE HORIZONS & SOCIETY. 2024

[14] N. Evan, "Deteksi kesalahan eja kata luluh pada berita dengan algoritma jaccard similarity (studi kasus : Tribunnews)," 2022.

[15] J. Olwen, "Deteksi penggunaan kata konjungsi pada portal berita dengan algoritma cosine similarity (studi kasus: Tribun news)," 2022.

[16] E. Fitri, Y. Yuliani, S. Rosyida and W. Gata, "Analisis Sentimen Terhadap Aplikasi Ruangguru Menggunakan Algoritma Naive Bayes, Random Forest Dan Support Vector Machine," *Jurnal Transformatika,* vol. 18, no. 1, pp. 71-80, 2020.

[17] N. G. Ramadhan, F. D. Adhinata, A. J. T. Segara and D. P. Rakhmadani, "Deteksi Berita Palsu Menggunakan Metode Random Forest dan Logistic," *JURIKOM (Jurnal Riset Komputer),* vol. 9, no. 2, p. 251−256, 2022.

[18] A. I. Fahma, I. Cholissodin and R. S. Perdana, "Identifikasi Kesalahan Penulisan Kata (Typographical Error) pada Dokumen Berbahasa Indonesia Menggunakan Metode N-gram dan Levenshtein Distance," *Jurnal Pengembangan Teknologi Informasi dan Ilmu Komputer,* vol. 2, no. 1, pp. 53-62, 2018.

[19] W. H. Kencana, I. V. O. Situmeang, Meisyanti, K. J. Rahmawati and H. Nugroho, "Penggunaan Media Sosial dalam Portal Berita Online," *Jurnal IKRAITH-HUMANIORA,* vol. 6, no. 2, pp. 136-145, 2022.

[20] A. N. Rohman, A. N. Rohman and S. Raharjo, "Deteksi Emosi Media Sosial Menggunakan Pendekatan Leksikon dan Natural Language Processing," *JURNAL EKSPLORA INFORMATIKA,* vol. 9, no. 1, pp. 70-76, 2019.

[21] N. A. Purwitasari and M. Soleh, "Implementasi Algoritma Artificial Neural Network Dalam Pembuatan Chatbot Menggunakan Pendekatan Natural Language Processing," *Jurnal IPTEK,* vol. 6, no. 1, pp. 14-21, 2022.

[22] F. S. Jumeilah, "Penerapan Support Vector Machine (SVM) untuk Pengkategorian Penelitian," *Jurnal RESTI (Rekayasa*

Sistem dan Teknologi Informasi, vol. 1, no. 1, pp. 19-25, 2017.

[23] L. Ardiani, H. Sujaini and Tursina, "Implementasi Sentiment Analysis Tanggapan Masyarakat Terhadap Pembangunan di Kota Pontianak," *JUSTIN (Jurnal Sistem dan Teknologi Informasi),* vol. 8, no. 2, pp. 183-190, 2020.

[24] A. Andriani, "SISTEM PENDUKUNG KEPUTUSAN BERBASIS DECISION TREE DALAM PEMBERIAN BEASISWA STUDI KASUS: AMIK "BSI YOGYAKARTA"," in *Seminar Nasional Teknologi Informasi dan Komunikasi 2013 (SENTIKA 2013)*, Yogyakarta, 2013.

[25] L. S. Muchlis, "PROSES DECISION TREE PADA DATAMINING DENGAN ALGORITMA ID3," *Jurnal Saintek,* vol. 2, no. 1, pp. 87-93, 2010.

[26] L. M. Cendani and A. Wibowo, "Perbandingan Metode Ensemble Learning pada Klasifikasi Penyakit Diabetes," *Jurnal Masyarakat Informatika,* vol. 13, no. 1, pp. 33-43, 2022.

[27] F. Y. Pamuji and V. P. Ramadhan, "Komparasi Algoritma Random Forest Dan Decision Tree Untuk Memprediksi Keberhasilan Immunotheraphy," *Jurnal Teknologi dan Manajemen Informatika,* vol. 7, no. 1, pp. 46-50, 2021.

[28] R. Supriyadi, W. Gata, N. Maulidah and A. Fauzi, "Penerapan Algoritma Random ForestUntuk Menentukan Kualitas Anggur Merah," *JURNAL ILMIAH EKONOMI DAN BISNIS,* vol. 13, no. 2, pp. 67-75, 2020.

[29] R. Siringoringo, "KLASIFIKASI DATA TIDAK SEIMBANG MENGGUNAKAN ALGORITMA SMOTE DAN k-NEAREST NEIGHBOR," *Jurnal ISD,* vol. 3, no. 1, pp. 44-49, 2018.

[30] A. M. Argina, "Penerapan Metode Klasifikasi K-Nearest Neigbor pada Dataset Penderita Penyakit Diabetes," *Indonesian Journal of Data and Science,* vol. 1, no. 2, pp. 29-33, 2020.