
UTAPIS Indonesian Word Error Detection Application: Design and Development**Niknik Mediyawati¹, Rudi Sutomo^{✉ 2}, Samiaji Bintang Nusantara³, Marlinda Vasty Overbeek⁴**niknik@umn.ac.id, rudi.sutomo@umn.ac.id, marlinda.vasty@umn.ac.id,samiaji.bintang@umn.ac.id¹Technical Informatics Information System, Universitas Multimedia Nusantara, Tangerang, Indonesia.²Informatics, Universitas Multimedia Nusantara, Tangerang, Indonesia.^{3,4}Digital Journalism, Universitas Multimedia Nusantara, Tangerang, Indonesia

Article Information

Submitted : 6 Feb 2024

Reviewed: 13 Feb 2024

Accepted : 20 Feb 2024

KeywordsSentence Syntax Error
Detection Algorithm,
Cosine Similarity
Algorithm, Jaccard
Similarity, Framework
Flask, Website

Abstract

People misspell and incorrectly write Indonesian words without even recognizing it. Poor and incorrect language proficiency is a common cause of word spelling mistakes. Since PT Tribun Digital Online publishes between 600 and 700 articles a day, the U-Tapis website might be of great assistance to journalists in identifying grammatical problems. Three separate cities—Jakarta, Makassar, and Solo—have participated in the testing of the U-tapis website. @ Tribun Digital Online, PT. This department is in charge of using the Flask framework to create the U-tapis website design. This study develops software or applications that are comparatively brief, succinct, and quick using the RAD (Rapid Application Development) technique. In addition to system development methods. The implementation of this method uses the U-tapis Sentence Syntax Error Detection algorithm which functions to detect an algorithm comparing two pieces of data, calculated using a formula called the Jaccard Index to check spelling errors in words. UAT test results produce all the features according to the wishes of the company owner based on the problems faced.

A. Introduction

Reports about past or currently highly debated events are called news. The word "news" itself is derived from the Sanskrit word "Vrit," which is translated as "Write" in English and signifies "Exists" or "Happens." Some refer to it as "Vritta," which is Sanskrit for "event" or "What Has Happened." Experts have several views of what news is. Paul de Maesenner, for example, defines news as information that is both attention-grabbing and helpful in order to pique the interest of listeners. And finally, according to Mithcell V. Charnley, who said that news is a timely report to find out new facts and opinions that have appeal or are important to the wider community [1][2].

The existence of social media in society also has an impact on the way we communicate with people around us, especially the millennial generation. When communicating on social media, we are not required to use good and correct Indonesian. This can cause the purity of the Indonesian language to fade because Indonesian is used incorrectly and does not comply with standard Indonesian grammatical rules [3].

Errors in correctly spelling Indonesian words are another frequent occurrence that people rarely recognize. Poor reading comprehension or ignorance of proper language usage are major contributors to spelling mistakes. Lack of literacy in reading has become a habit for Indonesian citizens, this can affect their knowledge of the Indonesian language. Apart from a lack of insight, many people learn to spell from their respective regional languages so the way they spell each word can be different. This causes the spelling to change and not be in accordance with the KBBI (Big Indonesian Dictionary). However, in this day and age, good and correct Indonesian language is increasingly fading. Another cause of the fading of good and correct Indonesian is due to a lack of Indonesian language education in an environment. This is because the environment makes Indonesian language mistakes normal, so many people are carried away by these Indonesian language mistakes [4].

Language, in the form of words and sentences, is a weapon for a journalist as well. Journalists also participate in educating the nation through language. Therefore, the language that a journalist must speak must be good and correct so that he can be an example for the Indonesian people in speaking good and correct Indonesian in accordance with the rules of the KBBI (Big Indonesian Dictionary). U- Tapis itself exists as a medium that can help journalists and editors in filtering out errors in the Indonesian language [5].

B. Research Method

The method is described in detail, including the design, population, sample and sampling techniques, how the research works, the parameters observed, and technical analysis. The method is written in narrative form by conveying the importance of the way the research was carried out. Any form of instructions, manuals or technical manuals in research activities that are too detailed should not be included [6].

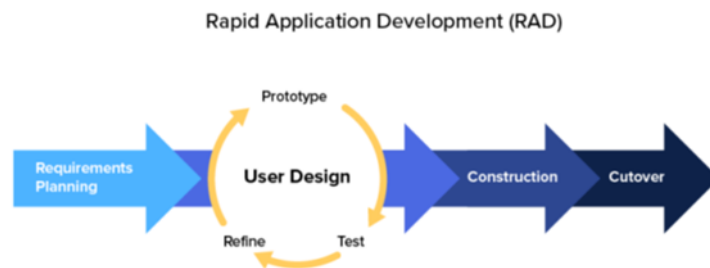


Figure 1. Rapid Application Development Method

RAD (Rapid Application Development) is the method that was selected for this study since it can be implemented quickly and with a limited number of teams. As this research is being conducted by a single individual, it therefore fits this description. It is also possible to fulfill user requests because users and developers communicate constantly. According to [7], the following are the phases of RAD (Rapid Application Development) system development:

Phase 1: Planning Requirements

Several SDLC phases and system planning components are combined in this step. Throughout this stage, the development team will do a number of tasks, including conducting literature reviews, surveying staff members about the requirements for the system that needs to be developed, and observing and interviewing business owners about the issues they are facing and how to resolve them.

Phase 2: User Design

In order to create a prototype or design for the Utapis Application, the user and developer will now completely collaborate. The system to be built's prototype technique makes use of a Figma supporting application; in addition to Flwochart, UX, using draw.io. Furthermore, assessments and alpha testing are used in the testing procedure.

3. Phase 3: Construction

The development team, often known as programmers, are responsible for this phase; their primary objective is to simply concentrate on creating apps that meet user needs in light of the findings from the earlier user design. Using a variety of tools aids in the process of designing an application. For example, the Laravel framework can be implemented using Visual Studio Code, and PHP, JavaScript, and Framework Flask can all be implemented, each of which has advantages in the processes of building a website. The database used for the website's construction is MongoDB, which provides data storage.

Phase 4: Cutover Stage

This cutover step involves testing, moving to a new system, and providing users with application training for the system. During this stage, the system under

construction is tested using the BlackBox Testing model in conjunction with the UAT (User Acceptance Test) approach [8].

This research employs the hashing technique in addition to the system development approach, which makes use of the Rapid Application Development (RAD) development process, to enhance database security in applications.

The preprocessing algorithm processes news data into data per sentence that other algorithms can use. The Conditional Random Field (CRF) algorithm is used to assign tags/labels to each word/symbol in each sentence data. The Context Grammar (CFG) algorithm is used to parse each sentence's data and determine whether the syntax of a sentence is correct (if parsing succeeds) or incorrect (if parsing fails). This algorithm is run on a Virtual Private Server (VPS) owned by Multimedia Nusantara University (UMN) as an API (Application Programming Interface) using the Python Flask Web Framework [9][10][11].

The algorithm applied to U-Tapis for detecting conjunction errors is implemented using Javascript with the Next.JS framework. The algorithm was also deployed using Vercel so that Web U-tapis could detect conjunctions [12][13][14].

The algorithm used to detect spelling errors in the word *teruah* is Jaccard Similarity; this algorithm compares two pieces of data, which are then calculated using a formula called the Jaccard Index. The first data is in the form of articles received from the U- U-Tapis Web [15].

C. Result and Discussion

The following are the results and discussions carried out in this study based on the research methodology using the RAD method:

a) Requirement Planning

Utapis Inventor was the target of the interview process that was carried out. The interview results that were acquired from the owner during the procedure are displayed in Table 1 below:

Table 1. Requirement Planning interview

Discussion points	Inventor's opinion
Development of a corpus of Indonesian vocabulary (word mining, text mining, word embedding, identification of entity names)	At this stage, the Proposing Team carries out the process of building a data corpus by inputting millions of basic and standard words in accordance with the Big Indonesian Dictionary (KBBI).
Development of Indonesian syntagmatics and grammatical structure tagging (Part-of-speech tagging)	At this stage, the proposing team adapts the formula for forming affix words that has been agreed upon by language experts. Then, the team also adapted the classification structure/grouping of words, such as based on subject, predicate, object, compound sentences, intra- and inter-sentence conjunctions, active and passive sentences, command sentences, statements and questions, including punctuation (period, comma, alphanumeric, and so on) in Indonesian sentences.
Creation of computational programs for text analysis and language filters using an artificial intelligence-based text classification method approach with Natural	At this stage, the proposing team builds a network system and hardware, such as server installation, which is needed in the process of creating and implementing the computing program to be developed. Then, the team created a computer program based on Natural Language Processing

Language Processing (NLP).	(NLP) by referring to an algorithm that had been determined by an AI engineer. The application of the specified algorithm, the Sympspell algorithm, refers to the level of accuracy and speed of computational results, especially in checking word and sentence errors and providing recommendations for correct words and sentences.
Creating User Experience (UX) for text analysis and Indonesian language filter applications.	At this stage, the proposing team creates a design and creates an interface (user experience) for the U-tapis application which can be accessed on computers/PCs, laptops and devices (mobile phones, tablets). Users will be divided into two categories, U-tapis managers/developers and U-tapis users. U- Tapis developers can modify, add words, language syntagmatic formulas, while U-Tapis users can write and check words and sentences.

b) User Design

A website system's design process begins with the formation or design of the system that will be constructed. Flowchart was used in this planned study to facilitate the system modeling process.

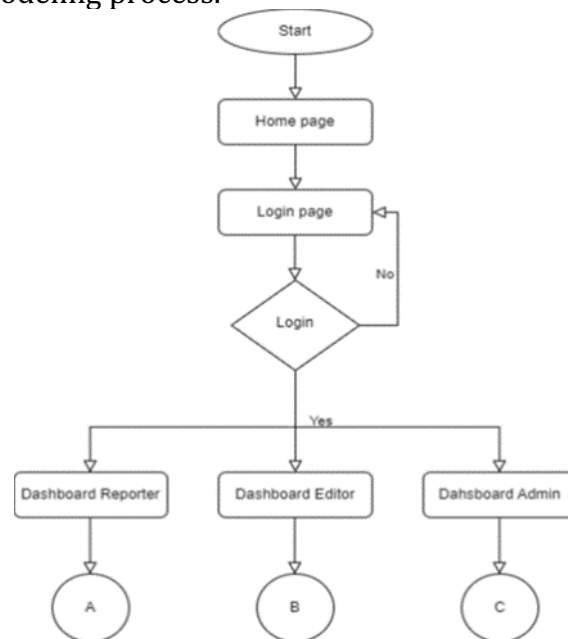


Figure 2. Flowchart of use of the U-tapis website

The U-tapis application flow starts from the main page when the user first enters the U-tapis page. Next, users can go to the login page. If the login fails, the user must repeat the login steps using the email and password that match the registered account. The page matching to the role that is saved in the user's account will then be displayed to the user. The user will be redirected to the admin dashboard page in Figure. 2, if he is an admin, he will be routed to the editor

dashboard page, and if he is a reporter, he will be directed to the reporter dashboard page.



Figure 3. Design of U-Tapis Reporter Web System

In this Flowchart, reporters can enter news titles, news categories, and articles they have written to be checked by reporters using the U-filter system. After filling in the title, category, and news article, the reporter can press the "send" button in the screen's bottom right corner in figure 3.

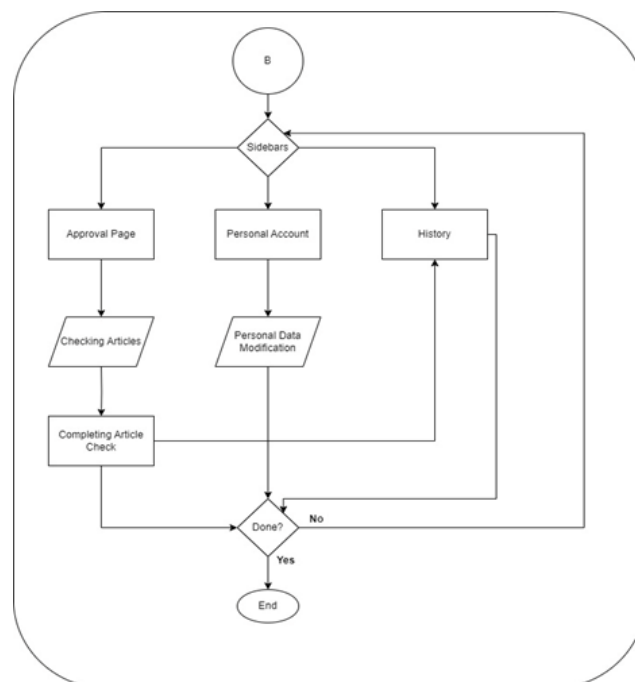


Figure 4. Design of U-Tapis Editor Web System

In this flowchart, editors can see what news reporters have submitted for review. To conduct a review, the editor only needs to click the "view" option on the news to be reviewed. Next, scroll down until you find the 'Delete' button. If the 'filter' button is pressed, the news will be processed using the U-filter system. Currently, U- U-Tapis can detect syntax errors in sentences, conjunction errors in sentences, and word derivation errors in figure 4.

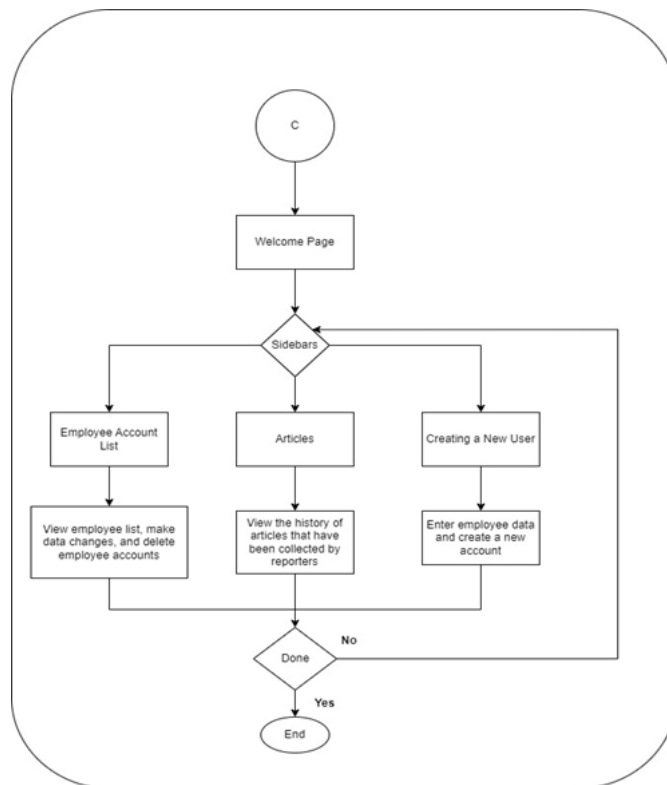


Figure 5. Design of Admin U-Tapis Web System

It contains a summary of the total news articles that have been collected, articles that the editor has approved, articles that are being edited, and a distribution diagram of news categories that have been approved in figure 5.

c) Construction

This subchapter will show an overview of the software interface design that will be built. Figure 6 shows the main page display interface design in this study can be seen in the image below :



Figure 6. Display of the main page of the Admin U-Tapis web system

Figure 7 is a display of the news list page. On this page, the admin can monitor news that has been entered into the U-tapis application, whether the status of the article is still pending or the review process has been completed by the editor. On the news list page, the admin can determine the number of news items that appear on 1 page on the news list page. Admin can choose to display 10, 25, 50, or 100 articles on 1 page in the news list table.

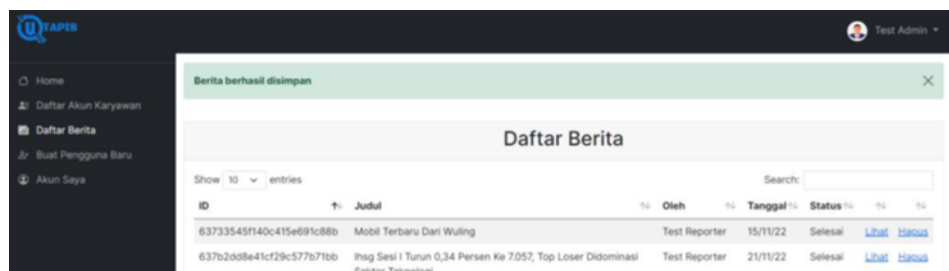


Figure 7. Example of notification/warning "News saved successfully"

Figure 8 is part of the display for checking news like an editor. Admins can modify news data such as news titles, news categories, and also the content of the news. The textbox for news writing uses the tinyMCE javascript text editor extension so that news writing can be neater.

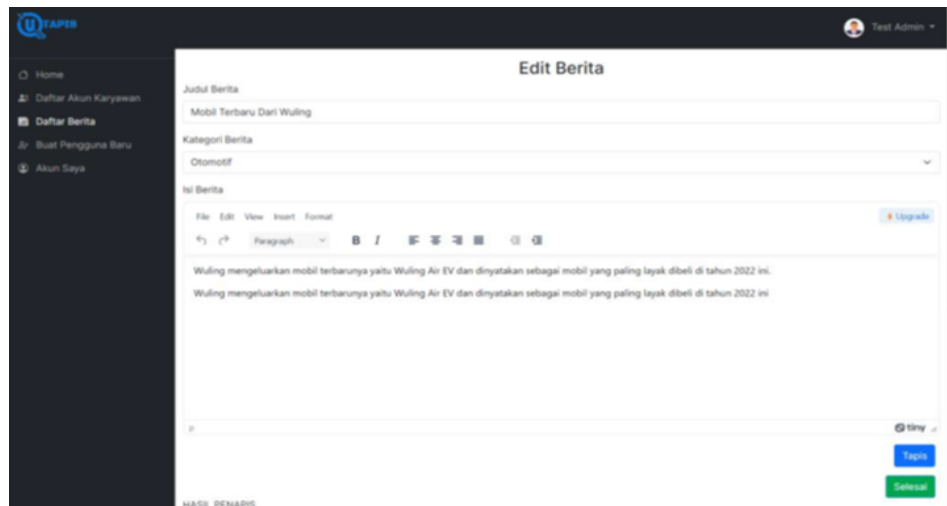


Figure 8. News Edit Page

Figure 9 is a display of the Filter Button which will later produce text as in Figure 9. The results of filtering using machine learning are a display of text on the news editing page that has been tagged through the machine learning process. There are several algorithms that process news text, namely sentence syntax, conjunctions, and word decay. For the sentence syntax algorithm, sentences that are syntactically judged to be incorrect, the font color of the sentence will be marked as red and green.

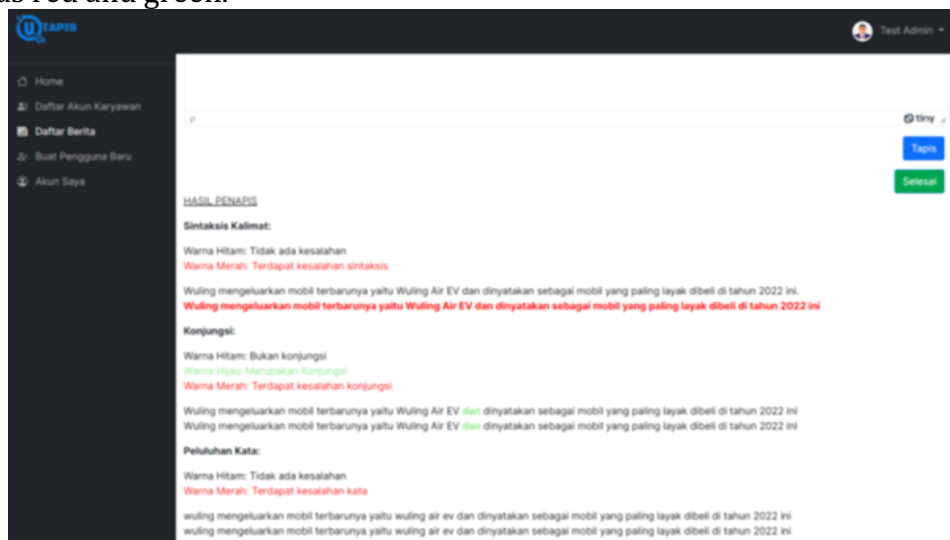


Figure 9. Sentence Syntax Filter Results

D. Cutover Phase

The User Acceptance Test, or UAT, is used in the technique stage of this research to verify that the website is constructed in accordance with user requirements. The Black Box Testing method is used to test the website for the owner, who is both the website user and owner of the company. In this way, all of the features for the admin, reporter, and editor sections have been tested and deemed successful. Table 2 indicates that all of the features pertaining to the prior firm or needs have been successful and operating as intended. Every position—

reporter, administrator, and editor—listed in is tested in this way. A super admin job on the system can also access all menu.

These findings indicate that every test conducted on the admin end and in the Utapis area was successful, indicating that these features meet the requirements that were asked for.

Table 2. User acceptance test results

No	Test case	Results
1	Login and update password	Successful
2	Upload news	Successful
3	View the news list	Successful
4	View news content	Successful
5	Do a news check	Successful
6	Show the results of checking the news	Successful
7	Sending information on the results of checking news	Successful
8	Revise the results of checking the news	Successful
9	Confirm receipt of news	Successful
10	View the list of confirmation of news receipt	Successful

D. Conclusion

One firm in the media space, specifically in the news production division, is PT Tribun Digital Online. As to The Tribune, reporters and journalists receive thousands of news items every day. Consequently, many of them lack the time to edit the reports beforehand, resulting in disorganized phrasing. U-tapis were therefore developed in partnership with PT Tribun Digital Online thanks to the initiative of Multimedia Nusantara University. Web-based program U-tapis was developed with appropriate and effective usage of Indonesian as its foundation. U-tapis' primary purpose is to assist editors and reporters in decreasing the amount of linguistic blunders they make, particularly when producing news. Building and implementing a website, particularly on the back-end with Python Flask, integrating a website with a MongoDB database, and integrating a website with three algorithms have all been done properly.

E. Acknowledgment

We express our gratitude to Universitas Multimedia Nusantara for their crucial help. Their assistance was very important to the accomplishment of our research. We would like to express our profound appreciation to our Information Systems Department colleagues, whose knowledgeable advice and deep experience were invaluable to us over the entire research process. Their significant input improved the overall quality of the study and helped determine its course..

F. References

- [1] L. F. Kusmanto and H. Subagio, "Analisa Perbandingan Faktor-Faktor Penentu Pemilihan Konsumen Dalam Penggunaan Koran Cetak Atau E-Paper Pada Produk Harian Kompas Di Surabaya Jawa Timur," *J. Strateg. Pemasar.*, vol. 3, no. 2, 2016.
- [2] Y. Febrianti, D. Cristiana Victoria, and J. Priyanto, "Analisis Kesalahan Berbahasa Pada Bidang Ejaan dan Sintaksis dalam Karangan Berbahasa

- Indonesia," *Primaria Educ. J. /*, vol. 2, no. 2, pp. 199–209, 2019.
- [3] Z. Ardi and S. A. Putri, "The analysis of the social media impact on the millennial generation behavior and social interactions," *Southeast Asian J. Technol. Sci.*, vol. 1, no. 2, pp. 70–77, 2020.
- [4] T. Journal and N. Industrialization, "张小华 1 , 季永林 2 (1," vol. 10, no. 8, 2020.
- [5] N. Mediyawati and S. Bintang, "Platform Kecerdasan Buatan Sebagai Media Inovatif Untuk Meningkatkan Keterampilan Berkomunikasi: U-Tapis," *Pros. Semin. Nas. Pendidik. Progr. Pascasarj. Univ. PGRI Palembang 21 Agustus 2021*, pp. 69–79, 2021.
- [6] M. M. Dr. H. Darmadi, S.Ag., "Cakrawala Pendidikan Akhlak," vol. 41, no. Terbaru, p. 157, 2018, doi: 10.21831/cp.v40i2.40812.
- [7] A. Lestari, N. Nugrahaningsih, and D. Septiana, "Perancangan Aplikasi Kamus Digital Bahasa Lawangan – Bahasa Indonesia," *J. Teknol. Inf. J. Keilmuan dan Apl. Bid. Tek. Inform.*, vol. 7, no. 2, pp. 205–211, 2023, doi: 10.47111/jti.v7i2.9095.
- [8] D. Harmilasari and C. Munggaran, "Evaluasi Kepuasan Pengguna Portal Berita Menggunakan Usability Metric," *J. Ilm. Komputasi*, vol. 19, no. 3, pp. 293–300, 2020, doi: 10.32409/jikstik.19.3.23.
- [9] D. V. Kornienko, S. V. Mishina, S. V. Shcherbatykh, and M. O. Melnikov, "Principles of securing RESTful API web services developed with python frameworks," *J. Phys. Conf. Ser.*, vol. 2094, no. 3, 2021, doi: 10.1088/1742-6596/2094/3/032016.
- [10] R. Yusuf Azhari, "Web Service Framework: Flask Dan Fastapi," *Technol. Informatics Insight J.*, vol. 1, no. 1, pp. 58–65, 2022, doi: 10.32639/tij.v1i1.54.
- [11] M. S. Bonney *et al.*, "Development of a digital twin operational platform using Python Flask," *Data-Centric Eng.*, vol. 3, no. 1, pp. 1–14, 2022, doi: 10.1017/dce.2022.1.
- [12] R. H. Mondy and A. Wijayanto, "Recommendation System With Content-Based Filtering Method for Culinary Tourism in Mangan Application," *ITSMART J. Ilm. Teknol. dan Inf.*, vol. 8, no. 2, pp. 65–72, 2019.
- [13] N. B. Ahmat Baseri, J. A. Bakar, A. Ahmad, H. Jafferi, and M. F. Zamri, "SMVS: A Web-based Application for Graphical Visualization of Malay Text Corpus," *ISCAIE 2020 - IEEE 10th Symp. Comput. Appl. Ind. Electron.*, pp. 30–35, 2020, doi: 10.1109/ISCAIE47305.2020.9108705.
- [14] T. M. P. Aulia, A. Jamaludin, and ..., "Extractive Text Summerization Pada Berita Berbahasa Indonesia Menggunakan Algoritma Support Vector Machine," *J-SAKTI (Jurnal Sains ...)*, vol. 5, no. September, pp. 727–735, 2021.
- [15] J. Ahmed and M. Ahmed, "Online News Classification Using Machine Learning Techniques," *IIUM Eng. J.*, vol. 22, no. 2, pp. 210–225, 2021, doi: 10.31436/iiumej.v22i2.1662.