Research Article

Digital Mining Algorithm of English Translation Course Information Based on Digital Twin Technology

Juan Yang

School of Foreign Languages, Changzhou Institute of Technology, Changzhou, Jiangsu, China

Correspondence should be addressed to Juan Yang; yangjuan@cit.edu.cn

Received 20 October 2021; Revised 1 December 2021; Accepted 10 December 2021; Published 24 December 2021

Academic Editor: Vinoth Babu Kumaravelu

Copyright © 2021 Juan Yang. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Cross-language communication puts forward higher requirements for information mining in English translation course. Aiming at the problem that the frequent patterns in the current digital mining algorithms produce a large number of patterns and rules, with a long execution time, this paper proposes a digital mining algorithm for English translation course information based on digital twin technology. According to the results of word segmentation and tagging, the feature words of English translation text are extracted, and the cross-language mapping of text is established by using digital twin technology. The estimated probability of text translation is maximized by corresponding relationship. The text information is transformed into text vector, the semantic similarity of text is calculated, and the degree of translation matching is judged. Based on this data dimension, the frequent sequence is constructed by transforming suffix sequence into prefix sequence, and the digital mining algorithm is designed. The results of example analysis show that the execution time of digital mining algorithm based on digital twin technology is significantly shorter than that based on Apriori and Map Reduce, and the mining accuracy rate reached more than 80%, which has good performance in processing massive data.

1. Introduction

With the opening of translation major in colleges and universities, more and more English majors and nonforeign language majors offer translation courses, and translation teaching is gradually being paid attention to. However, there are still many problems in the training of translation talents in colleges and universities in terms of teachers, teaching time, and teaching mode. The information technology has brought about the development space for the reform of translation teaching in colleges and universities and provided unlimited possibilities for the benign and sustainable development of translation teaching. At the same time, under the background of educational information, students' information literacy and computer and smartphone operation ability are constantly improved, and they have the ability of information collection, processing, and application and have been used to the learning methods of "click, search and fragmentation." All of these conditions provide help for students to get rid of the shackles of translation teaching and realize personalized translation learning. Therefore, the in-

depth study of the information of English translation course can promote the development of translation teaching and improve the translation level of students to a certain extent. In the practical of translation, ambiguity and polysemy of word semantics are the key and difficult problems. Therefore, it is necessary to find the translation information accurately, comprehensively, and quickly from a large amount of information data, which requires higher requirements for the information processing technology of translation courses. The key is to measure the semantic similarity of translation words accurately. Through the digital mining of English translation course information, the information matching of bilingual semantic level can be realized, thus providing more accurate and easy to process cross-language communication [1].

Digital twin technology has also become a digital image, which is described by the characteristics of digital technology to form a digital model [2]. In the process of modeling, the consistency of the performance and behavior of the model is ensured, so as to realize intelligent operation and achieve the purpose of optimal management. With the help

of digital twin technology, natural language can communicate with computer. With the support of computer technology, quantitative research on language information can be carried out. At the same time, language description used together between computer and computer can be provided [3]. By constructing the knowledge base of artificial intelligence reasoning, semantic analysis and intelligent reasoning are completed, so as to provide easy access and processing services for English translation course. In the process of digital mining of English translation course information, we need to understand the definition of translation semantic mapping, cross-language implementation, similarity calculation, and the final clustering algorithm of data mining [4]. Therefore, based on digital twin technology, this paper designs a digital mining algorithm for English translation course information. By integrating the uncertainty of translation and retrieval, the gap between cross-language information retrieval and monolingual information retrieval can be narrowed. In order to ensure the semantic integrity and eliminate ambiguity in the process of translation, this paper provides a reference for English translation teaching.

2. Digital Mining Algorithm of English Translation Course Information Based on Digital Twin Technology

2.1. Extracting Feature Words from English Translation Texts. In order to accurately obtain the data information needed in English translation, text feature words are firstly extracted. In order to extract Chinese keywords, word segmentation is the first step. Word segmentation is to divide every sentence in the text into ordered word segments according to the established way [5]. As we all know, Chinese text is composed of words. Therefore, words are the components of Chinese text. So word segmentation is the first step of keyword extraction. The difference between Chinese and English is that in English, there are spaces between words, while in Chinese, there are no spaces. Although Chinese sentences contain punctuation, there is no separation between words. Chinese can be composed of one or more words to express different meanings. The division of words depends on the language environment and the language knowledge accumulated in people's daily life. This is a complex process. Different ways of division and different language environments will lead to different sentence and word meanings. Because of this feature, Chinese word segmentation is more complicated than English word segmentation. This paper uses the thulac word segmentation software to segment and label the text. After word segmentation, the text will contain a large number of stop words. Stop words belong to redundant data in text analysis, which do not have the ability to express the theme of the article, and often have the characteristics of high frequency and meaninglessness [6]. By removing these words, the interference factors of keyword extraction can be reduced. In this process, we need to pay attention to the coding problem and ensure that the content in the TXT text is saved in UTF-8 format. Keywords refer to the words that can reflect

the overall content or theme of the article, and they are often representative. In this paper, keywords are used as target feature words, which can effectively improve the effect of feature extraction. By multiplying the word frequency and antidocument frequency, the greater the weight value, the higher the probability of the word as a feature word. The calculation formula of word frequency is as follows:

$$f_1 = \frac{n_1}{w}. \quad (1)$$

In formula (1), f_1 represents the word frequency, that is, the frequency of the word appearing in the document; n_1 is the number of times words appear in the article; w is the total number of words in the article. The frequency of antidocument reflects the classification ability of the word. The calculation formula is as follows:

$$f_2 = \log \left[\frac{n_2}{n_3 + 1} \right]. \quad (2)$$

In formula (2), f_2 is the frequency of antidocument; n_2 is the total number of documents; n_3 is the number of documents containing the word. By analyzing the characteristics of information content, we can see that its title is often very general. If a word appears in the title, it is often more important than other words. The position of words can reflect the importance of words to a certain extent. By analyzing the part of keywords, it is found that the composition of keywords is generally nouns or noun phrases, followed by verbs, and finally numerals, adverbs, and other modifiers. Therefore, considering part of speech features can effectively avoid the defects of traditional linguistic methods. On the basis of word frequency and antidocument frequency, position feature and part of speech feature are introduced into weight calculation to calculate the comprehensive weight of each word [7]. The weight calculation formula is as follows:

$$\omega = \omega_{f_1} + \omega_{f_2} + \omega_1 + \omega_2. \quad (3)$$

In formula (3), ω represents the comprehensive weight; ω_{f_1} is the weight of word frequency; ω_{f_2} represents antidocument frequency weight; ω_1 represents the weight of position factor, which is assigned according to the position of the extracted words in the article; ω_2 represents the weight of the part of speech factor, which is assigned according to the different parts of speech of the extracted words [8]. According to the weight order, the text feature words are determined, which lays the foundation for digital mining.

2.2. Cross-Language Mapping of English Translation Based on Digital Twin Technology. The purpose of sentence processing is to determine the content corresponding to the original text immediately in the process of word conversion translation centered on the establishment of text feature words. In our cognition, a word in a source language is a number of words corresponding to a translation language, and machine translation cannot decide which word to choose as the output, so it will output all possible choices.

Languages and programs do not mix grammar rules with program algorithms. The process of English translation based on digital twin technology is regarded as a process of information transmission, and the channel model is used to explain English translation [9]. The specific method is to take translation as a decoding process and transform the original text into the translation through the model. Therefore, translation can be divided into the following problems: model problem, training problem, and decoding problem. The most important thing is to solve the problem of finding the target language with the highest translation probability for any input source language sentence [10]. The cross-language mapping of English translation based on digital twin technology is shown in Figure 1.

According to the cross-language mapping, three key problems must be solved in English translation: estimating the probability of the language model, estimating the translation probability, and finding a fast and effective search algorithm to maximize the product of the above two probabilities [11]. When translating Chinese into English, good text space mapping can ensure that information is not lost, and it is convenient for calculation. In this paper, we define a triple to represent the cross-language mapping model. The concrete expression is as follows:

$$Y = [C, E, \lambda(C, E)]. \quad (4)$$

In formula (4), Y represents cross-language mapping; C represents Chinese vocabulary set; E represents the vocabulary set of English translation, which has a one to many relationship with Chinese language; $\lambda(C, E)$ is the rule for cross-language mapping. Phrase pairs are extracted from bilinguals, and the probability of phrase pairs is estimated by maximum likelihood estimation. The formula is as follows:

$$p(c|l) = \frac{m(l, c)}{m(l)}. \quad (5)$$

In formula (5), $p(c|l)$ is the translation phrase pair probability; c represents Chinese vocabulary; l represents English translation vocabulary; $m(l, c)$ is the maximum likelihood estimate; $m(l)$ is the maximum likelihood estimator. Through the cross-language mapping model, the phrase is reordered. Since there is a one to many mapping relationship between Chinese and English translation, the corresponding mapping relationship is established with keywords as indexes according to different types of keywords. First, an empty element set is created. By editing the corresponding statements of keywords and entry objects, the key value pairs of all keywords are added to the empty set [12, 13]. Usually, the value of the keyword can be empty, and the corresponding value of the keyword can be added according to the subsequent conditions. When saving the translation information data, the entry data is saved in the text file named by keywords, so it is more convenient to establish the element set. The total elements are stored locally and read directly into the memory, which is convenient for the next step. For text information mining, good document mapping representation is the premise of good text clustering [14].

FIGURE 1: Cross-language mapping of English translation based on digital twin technology.

Based on the establishment of cross-language mapping of English translation by digital twin technology, the text information is further transformed into text vector to evaluate the similarity and category of text vector.

2.3. Calculating Semantic Similarity of Translated Words. In the process of data analysis and data mining, we need to know the difference between information and then evaluate the similarity and category of information. To quantify things with quantitative method, we must use quantitative method to describe the similarity between things. Similarity is an indicator of intimacy between two things. The closer the two things are, the more similar they are. On the contrary, the more distant the meaning of two things is, the less similar they are [15]. At present, the methods of similarity measurement have both diversity and applicability, so they are generally selected according to practical problems. The commonly used similarity measurement methods are as follows: correlation coefficient (proximity between measurement variables) and similarity coefficient (proximity between measurement samples). If the samples give qualitative data, then measure the proximity between samples and use the matching coefficient and consistency of available samples. Vocabulary is described and defined by meaning, and the core of meaning is sememe. Therefore, the similarity of a sense is determined by calculating its sememe similarity. All the sememes of a sense are represented in a hierarchical tree structure according to the context, so the similarity of primitives can be calculated by the relationship between the original nodes in the tree [16]. A single word can contain one or more senses, so the similarity of words can be directly converted into the calculation of the similarity of senses. Firstly, the text vocabulary is vectorized. To some extent, word vector can be used to describe the semantic distance between words [17]. A group of good text vectors can give a better mapping in the text space, so that the computer can calculate more accurate results. In this paper, we use the CBOW model to realize the word vectorization. The idea of CBOW is that the input is the context word vector of a specific word, and the output is the word vector corresponding to a specific word. The objective function is logarithmic likelihood function. The calculation formula is as follows:

$$g = \sum_{s \in V} \log r(s|v(s)). \quad (6)$$

In formula (6), g represents the objective function; s

represents the word vector in the text; v is the text vector; r represents the matrix vector. The input layer of the projection CBOW adopts the method of cumulative summation, which successfully omits the matrix vector calculation originally concentrated between the hidden and output layer and the softmax normalization operation on the output layer, so that the final output layer becomes a Huffman tree and directly outputs the result [18]. In this paper, after vectorization, the semantic similarity of the vector is obtained. There are two ways. One is to organize the concepts of related words in a tree structure through the network or real semantic dictionary. The other is to use the statistical model to solve the problem through the context information. At the same time, we think that word distance and word similarity are different expressions of the same relationship features. Cosine similarity is used to get semantic similarity. The formula is as follows:

$$\cos \alpha = \frac{\sum_1^u (\theta_x \varphi_x)}{\sqrt{\sum_1^u \theta_x^2} \sqrt{\sum_1^u \varphi_x^2}}. \quad (7)$$

In formula (7), $\cos \alpha$ represents the cosine similarity of the lexical vector; u is the total number of word vectors; θ_x and φ_x represent two lexical vectors; x is the total amount of elements in the vector. By calculating the similarity of translated words, the matching degree between translated sentences is determined to ensure the analysis effect of information digital mining.

2.4. Design the Digital Mining Algorithm of English Translation Course Information. After the definition and processing of the information related data of English translation course, the data grouping conditions and related dimensions are defined. Then, the digital mining algorithm is designed to extract valuable and meaningful information. Association rules have two important properties, which can make association rules, which are support and confidence [19]. Support is used to determine the frequency of a given vocabulary dataset. Confidence is used to determine the frequency of the occurrence of a lexical item set in a set containing another. Firstly, the minimum support and the minimum reliability are given to determine the association rules. Lexical item sets have an important property, that is, the number of transactions containing a particular set of items, known as support count. Mathematically, the support count of the lexical item set can be expressed by the following formula:

$$\varepsilon(A) = |\{b|A \subseteq b, b \in I\}|. \quad (8)$$

In formula (8), $\varepsilon(A)$ represents the support of lexical item sets; A represents lexical item set; b represents the antecedent set of rules; I represents the subsets of the rule. Because English translation information contains complex data, only item set is not enough to express, so it is necessary to define sequence. A set of sequences is a collection of tuples with one ID. Items can appear at most once in the item set, but they can appear times in different sequence. If sequence x is a subset of sequence y , then the support of x is the number of

x tuples, which can be marked when context is clear. Given a positive integer as the minimum support threshold, when the support is greater than or equal to the threshold, it is considered frequent, which is called sequence pattern [20]. Because frequent pattern mining will produce a large number of patterns and rules, which hinder the mining work, this paper designs the sequential pattern mining algorithm to improve the efficiency of the algorithm. In this paper, we use suffix sequence to prefix sequence to construct frequent sequence, which is explained in detail below. First, scan the database to get the frequent sequence patterns with length of 1, and form the corresponding prefix subset group. Such subsets can be obtained by constructing the corresponding projection database and mining the subsets of each sequential pattern recursively [21, 22]. Taking the sequence pattern prefix $\langle p \rangle$ as an example, the projection database collects all the subsequences with $\langle p \rangle$ prefix. For example, in the sequence $\langle p \text{ (pqr) (rw)} \rangle$, only $\langle \text{(pqr)(rw)} \rangle$ is calculated, where $\langle _s \rangle$ in $\langle _s \text{r(pt)} \rangle$ denotes the prefix p but matches in the item set. The projection database of $\langle p \rangle$ is $\langle \text{(pqr)(rw)} \rangle$, $\langle _s \text{r(pt)} \rangle$, and $\langle _w \text{q} \rangle$. Similarly, by scanning the projection database of $\langle p \rangle$, all sequence patterns with prefix $\langle p \rangle$ of length N will be found, which is $\langle \text{pr} \rangle: N$. Similarly, the sequence patterns prefixed with $\langle \text{pr} \rangle$ are $\langle _w \rangle$ and $\langle \text{pf} \rangle$, and the same operation is performed on other frequent sequence patterns of length 1. In this way, we continue to explore recursively and finally find all frequent sequence patterns, as shown in Table 1.

In this setting, the frequent sequence can be guaranteed to be monotonic. The most special pattern sequence is the basis of retrieving all pattern sequences. In this way, the problem of mining sequential patterns is reduced to only finding the most special pattern sequence, which can significantly reduce the complexity of mining. So far, the digital mining algorithm of English translation course information based on digital twin technology has been designed.

3. Example Analysis

3.1. Data Preparation. This case study takes English translation text as the research object for digital mining analysis. The primary task for these data is to realize the normalization and standardization of data information and apply it to Hadoop platform for processing. First of all, we need to preprocess the complex dataset and get standardized and clear data information through data filtering and cleaning, so as to get hidden and meaningful correlation information from a large number of course information. In this case study, we collected bilingual news text data in Chinese and English from Hong Kong news website through crawler, including headlines and text content. The original Web page information obtained by the theme crawler is saved in the local disk. By observing the content and composition of its Web data, content extraction is carried out according to the following tags. Regular expressions are used to extract text information under specific tags. The Chinese English news text dataset is used as experimental data. Besides 285 single documents without comparison, a total of 4082 bilingual documents were obtained, including finance, urban life,

TABLE 1: Sequence pattern.

Prefix	Suffix	Sequential pattern
<p>	<p (pqr) (rw)> <(_s)r(pt)> <(_w)q>	<p> <pp> <pr> <pq>
<q>	<(_r)(rw)>	<q>
<r>	<(_w)> <pt> <q>	<r>
<s>	<(_w)qr> <w(pw)q>	<t>
<t>	<w(pw)q> <q>	<qt>
<w>	<qr> <q>	<w>

and environment. The vocabulary of the local Chinese vocabulary database is relatively rich. The original web page is analyzed, and the method of extracting content based on specific tags is formulated to extract and integrate the English vocabulary, so as to make the visualization results more neat. The results of the whole process provide a good database for the digital mining of text vocabulary. The test data is provided by the original dataset and the expanded dataset. The original dataset contains 10-20 item sets. Each phase set consists of random numbers between 1000 and 15000. The expanded datasets are 2-5 times of the original datasets. The original dataset and the expanded dataset are marked as U0-U4, respectively. Frequent patterns in digital mining algorithms become increasingly complex as the amount of data grows. The more complex the frequent patterns, the longer the execution time, so the amount of data is a key factor affecting the processing time. The research goal of this paper is to reduce the execution time of the digital mining algorithm, so in order to test whether this method achieves the research goal, set different size datasets to compare the execution time of this method and other algorithms.

3.2. Result Analysis. This paper compares the digital mining algorithm based on digital twin technology with the digital mining algorithm based on Apriori and Map Reduce and proves the application effect of the algorithm through data. U0-U4 datasets of different sizes are selected as test datasets. In different datasets, the above three algorithms are executed in turn. And in the main function of the algorithm, we call the currentTimeMillis method to get the system time before and after the execution of the algorithm and record the two time differences as the total execution time of the algorithm. The test results are shown in Figure 2.

According to the results of Figure 2, when the test set is the original dataset U0, the data scale is small, and the execution time of the three algorithms is close. When the test set is the extended dataset U1, the data is expanded to twice the original data size, and the execution time of the three algorithms begins to lag. When the test set is the extended dataset U2-U4, the data is expanded to 3-5 times the original data scale. With the increase of the dataset size, the execution time gap of the three algorithms increases. The execution time of digital mining algorithm based on digital twin technology is obviously shorter than that based on Apriori and Map Reduce. This is because the algorithm designed in this paper is optimized in frequent pattern mining of item sets. By constructing frequent sequence of item sets, the

FIGURE 2: Algorithm execution time.

query speed is improved and execution time is shortened. Therefore, the performance of the algorithm is better than the two comparison algorithms when processing the same size of datasets. Therefore, the algorithm designed can reduce the time complexity of the algorithm and give full play to the advantages of digital mining.

The data of the news text extension test set is compared with the data mined by the three algorithms to test the accuracy of the three mining algorithms. The test results are shown in Figure 3.

According to the results of Figure 3, starting with the original dataset U0 with the smallest data size, the accuracy of the other two methods reaches above 90%. As the data expansion scale increased, the execution accuracy of all three algorithms decreased, but the execution accuracy of digital mining algorithms based on digital twin technology remained basically stable, with little decline, while the accuracy of the other two algorithms decreased significantly. Finally, the digital mining algorithm based on digital twin technology is designed with above 80% accuracy, significantly greater than that based on Apriori and Map Reduce. This is because the algorithm uses digital twin technology to establish a cross-language mapping model of text, calculates the semantic similarity of text, judges the degree of translation matching, ensures the maximum text translation correspondence, and narrows the gap between cross-

FIGURE 3: Algorithm execution accuracy.

language information retrieval and monolingual information retrieval; semantic description eliminates the ambiguity as far as possible, so the mining data is more accurate.

4. Conclusion

This paper designs a digital mining algorithm for English translation course information based on digital twin technology. Example analysis shows that the proposed algorithm can effectively reduce the time complexity and execution time, and the mining accuracy remains above 80%, which has certain advantages in processing massive data. In terms of feature word extraction and weight calculation, this paper only uses the reverse frequency method of words to calculate the weight of text feature words and does not consider the semantic association between words. In addition, although this method significantly reduces the execution time of the digital mining algorithm, there is still room for improving the mining accuracy. Future studies can use analytical methods to fully extract and obtain the weights of feature words and use bilingual semantic clustering methods to obtain more information to improve the accuracy of the algorithm.

Data Availability

The datasets used and/or analyzed during the current study are available from the corresponding author on reasonable request.

Conflicts of Interest

It is declared by the author that this article is free of conflict of interest.

References

- [1] Q. X. Yang, G. N. Wang, and T. Wang, "Simulation of multi-source log security data mining based on time series," *Computer Simulation*, vol. 36, no. 2, pp. 297–301, 2019.
- [2] Y. F. Yang, "On English translation skills from cross cultural perspective based on bidirectional association rules," *Journal of Hubei University of Education*, vol. 38, no. 9, p. 6, 2021.
- [3] X. Y. H and M. J. Li, "A knowledge domain analysis of research subjects and trends in current international translation studies," *Foreign Languages and Their Teaching*, vol. 6, p. 13, 2020.
- [4] N. Mansouri, M. M. Javidi, and B. Zade, "Using data mining techniques to improve replica management in cloud environment," *Soft Computing*, vol. 24, no. 10, pp. 7335–7360, 2020.
- [5] N. Chatterjee and S. Gupta, "Efficient phrase table pruning for Hindi to English machine translation through syntactic and marker-based filtering and hybrid similarity measurement," *Natural Language Engineering*, vol. 25, Part 1, pp. 171–210, 2019.
- [6] N. S. Khan, A. Abid, and K. Abid, "A novel natural language processing (NLP)-based machine translation model for English to Pakistan sign language translation," *Cognitive Computation*, vol. 12, no. 2, pp. 1–18, 2020.
- [7] M. X. Huang, "Vietnamese-English cross language query expansion based on weighted association patterns mining," *Journal of the China Society for Scientific and Technical Information*, vol. 36, no. 3, p. 12, 2019.
- [8] J. B. Jin, L. Ge, and X. L. Xu, "Development and application of computer data mining technology," *Journal of Zhejiang University of Water Resources and Electric Power*, vol. 31, no. 6, p. 5, 2019.
- [9] L. Grealish, S. Myers, C. Scott, M. Krug, and J. A. Todd, "Simulation as a research translation technique," *Clinical Simulation in Nursing*, vol. 31, pp. 17–20, 2019.
- [10] A. H. Nasution, Y. Murakami, and T. Ishida, "A generalized constraint approach to bilingual dictionary induction for low-resource language families," *Bilingual*, vol. 10, 2020.
- [11] G. X, Q. I. Le, and H. Guo, "On cross language writing theory: explanation, differentiation and reflection," *Journal of Guangdong University of Foreign Studies*, 2019.
- [12] S. Z. Li, L. J. Zhou, F. Zhou, and J. A. Zhang, "Laos-Chinese bilingual sentence similarity calculation method based on mutual translation feature word pair matching," *Modern Electronics Technique*, vol. 42, no. 4, pp. 79–83, 2019.
- [13] J. Dong, Y. Cong, G. Sun, Z. Fang, and Z. Ding, "Where and how to transfer: knowledge aggregation-induced transferability perception for unsupervised domain adaptation," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 1, 2021.
- [14] Z. Adak and A. Demiriz, "Hybridization of population-based ant colony optimization via data mining," *Intelligent Data Analysis*, vol. 24, no. 2, pp. 291–307, 2020.
- [15] I. Rm and R. N. Rosa, "Translation techniques used by English department students of Universitas Negeri Padang in translating a narrative text," *English Language and Literature*, vol. 9, no. 1, 2020.
- [16] M. Salihoun, "State of art of data mining and learning analytics tools in higher education," *International Journal of Emerging Technologies in Learning (iJET)*, vol. 15, no. 21, p. 58, 2020.
- [17] A. Valeev, I. Gibadullin, A. Khusainova, and A. Khan, *Application of Low-Resource Machine Translation Techniques to*

Russian-Tatar Language Pair 2019, <https://arxiv.org/abs/1910.00368>.

- [18] S. S. Shao, L. F. Wang, and Z. Y. Liu, "Sentiment analysis of Chinese and English version in Chinese and international standards based on big data, China Standardization," *China Standardization*, vol. 17, p. 6, 2019.
- [19] T. V. Nguyen, L. Zhou, A. Chong, B. Li, and X. Pu, "Predicting customer demand for remanufactured products: a data-mining approach," *European Journal of Operational Research*, vol. 281, no. 3, pp. 543–558, 2020.
- [20] H. Mallik, L. K. Arjuna, and V. Reddy, "An adaptive correlation based video data mining using machine learning," *International Journal of Knowledge-Based and Intelligent Engineering Systems*, vol. 24, no. 1, pp. 1–9, 2020.
- [21] Y. P. Singh and P. Gautam, "Development of data mining algorithm for giving the loan in banks introduction," *High Technology Letters*, vol. 27, no. 1, pp. 284–295, 2021.
- [22] Y. Sinambela, S. Herman, A. Takwim, and S. R. Widiyanto, "A study of comparing conceptual and performance of k-means and fuzzy c means algorithms (clustering method of data mining) of consumer segmentation," *Jurnal Riset Informatika*, vol. 2, no. 2, pp. 49–54, 2020.