Job matching analysis by latent semantic indexing enhanced on multilingual word meanings

Sukri[1,2], Noor Azah Samsudin[1], Ezak Fadzrin[1], Shamsul Kamal Ahmad Khalid[1], Liza Trisnawati[1,2]
[1]Faculty of Computer Science and Information Technology, Universiti Tun Hussein Onn Malaysia (UTHM), Johor, Malaysia
[2]Department of Informatics Engineering, Universitas Abdurrab, Pekanbaru, Indonesia

Article Info

Article history:

Received Jul 10, 2024
Revised Sep 4, 2024
Accepted Sep 29, 2024

Keywords:

Job matching
Latent semantic indexing
Meanings
Multi-language
Word

ABSTRACT

Job matching is a hiring process that involves a thorough understanding of the context and meaning of words in different languages. The updated and expanded latent semantic indexing (LSI) Framework seeks to improve the precision and relevance of job matching analysis of word meanings in multi-languages. Because they only compare related terms, conventional LSIs are often insufficient to address the complexity of context in job matching. Extending the LSI approach can improve the vector representation of words and help you understand the context and semantic relationships in the text. Improved LSI analyzes context more precisely by using word vector representation. Improved LSI focuses on understanding semantic relationships between words in many languages to produce more accurate and relevant job matches. This paper describes the steps involved in improving LSI, such as data collection, pre-processing, linguistic feature extraction, LSI model training, and evaluation of matching results. The results show that the examined classification model has much better performance in terms of word classification. Conventional LSI has an average prediction value of 79%, once the enhanced LSI can accurately predict about 84% of the entire word, it has a reasonable capacity to recognize the actual words in a natural context.

Corresponding Author:

Noor Azah Samsudin
Faculty of Computer Science and Information Technology, Universiti Tun Hussein Onn Malaysia (UTHM)
Parit Raja, Batu Pahat, Malaysia
Email: azah@uthm.edu.my

1. INTRODUCTION

In today's era of globalization, the challenge of bringing together job candidates with suitable positions is increasingly complex, especially with the diversity of languages that exist in various countries and cultures. Job searches often involve different languages, so sophisticated approaches are needed to ensure an accurate match between a candidate's qualifications and desired job requirements [1]. One approach that has been used is latent semantic indexing (LSI), a technique in the field of natural language processing that makes it possible to extract and analyze the relationships between words in documents [2], [3]. LSI allows for more flexible and relevant searches by recognizing semantic relationships between words [4], [5]. Text indexing and analysis techniques, LSI [6], used to find semantic patterns of documents. When analyzing text, LSI describes terms and documents using vector space. Complex connections or implied meanings between words in a text are beyond the scope of linear representation or LSI [7]. However, in a multi-lingual global context, LSI is not able to read the same meaning of different words.

Trust far, a lot of research has been done to help match jobs by developing analysis applications. In the last 5 years, many researchers have developed LSI techniques as a job matching analysis. However, there

is a huge gap in the inability to analyze contextual in multi-languages globally to get the same meaning in different words. Understanding the contextual, interconnectedness of words, and deeper meanings is essential for job matching models. The number of dimensions (semantic ideas) used to describe a document (which may be difficult to understand), the dimensions of a potential worker's personality traits and cognitive abilities, and sensitivity to changes in a set of documents are often limited by LSI. The majority of contextual analysis of matching uses the LSI technique by comparing words or meanings without analyzing different meanings or words that have the same meaning [8]. An understanding of the context of multi-lingual analysis is essential for the similarity of meaning and intent of matching in a job recruitment agreement [9], [10].
One of the solutions that can be applied for job matching to obtain high accuracy in analyzing multi-language contextuals by improving LSI techniques [11]-[13]. Based on a literature review, few have researched related to the application of contextual analysis LSI techniques [14]-[16]. Moreover, the application of multi-language with contextual analysis of different words with the same meaning [10], [17]. The existing research is only based on comparing the text of words with the same word [18]. A summary of the differences in research and distribution offered can be illustrated in Figure 1. Based on Figure 1, it is hoped that this study will gap the research by developing LSI techniques for multi-language contextual analysis. The improved LSI technique is expected to help to improve the accuracy and ability to analyze multi-language in job-matching.

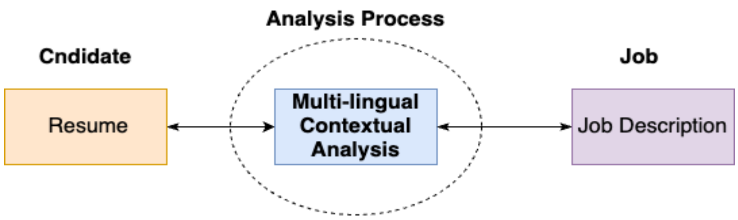

Figure 1. Multi-language analysis process between jobs and candidates

2. METHOD

As stated in the introduction, this study aims to improve the design of LSI Engineering as a multi-language contextual analysis on job matching. However, the study found the main challenge to achieving this goal was that conventional LSI was unable to analyze multi-language contextualities on matching jobs. Therefore, this study is divided into five stages to overcome these challenges to the processing of job matching data, namely Job Matching data collection, word data processing, text data processing, gensim, and the final vocabulary set. To provide a clearer understanding of the approach used in this study, the overall architecture of the job matching data processing method is illustrated in Figure 2.

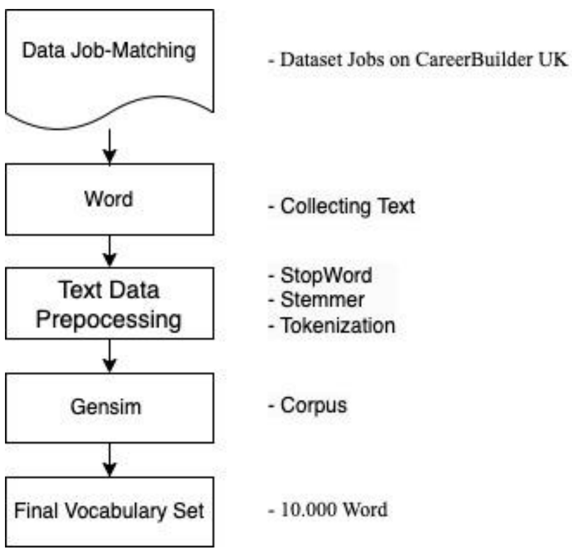

Figure 2. Data collection and preprocessing process

A. Job matching data collection

Job match data collection, which includes detailed information about available jobs, including job descriptions, qualification requirements, and job responsibilities. This data also provides information about potential candidates who are looking for a job, such as their skills, work experience, and educational background [19].

B. Word data processing

Word data processing, is a critical stage where word data is further analyzed. The main purpose of this analysis is to identify keywords in the job description and candidate profiles that are important for job matching. The process also involves the elimination of irrelevant words and the merging of words with the same meaning to simplify further analysis [20], [21].

C. Text data processing

Text data processing, which is a broader step in this process. At this stage, the text data is processed as a whole, including removing special characters, combining words with the same meaning, and eliminating irrelevant words [22], [23]. It aims to clean and prepare the text data for further analysis.

D. Gensim

Gensim is the software used in this process. It helps process text data and generate vector representations of text [24]. This vector representation can match candidates with suitable jobs based on similarities between the candidate's profile and the available job descriptions [8], [24].

E. Final vocabulary set

A collection of terms that have been selected and arranged based on their relevance and frequency of occurrence in the research data [25]. This set consists of words and phrases that cover important aspects of the topic being researched, and have gone through a validation process to ensure their accuracy and consistency. Thus, the overall design of the research framework to obtain an improved LSI on multi-language contextual analysis is shown in Figure 3.

Based on Figure 3, the framework of the word meaning analysis on a process that integrates various important steps to facilitate effective job matching. The first step is to use the Job Dataset as the main data source. This dataset contains information about various job positions, including job descriptions, required qualifications, job location, and other relevant attributes. Using this dataset, we can better begin analysis and modeling to understand the context of the work being done.

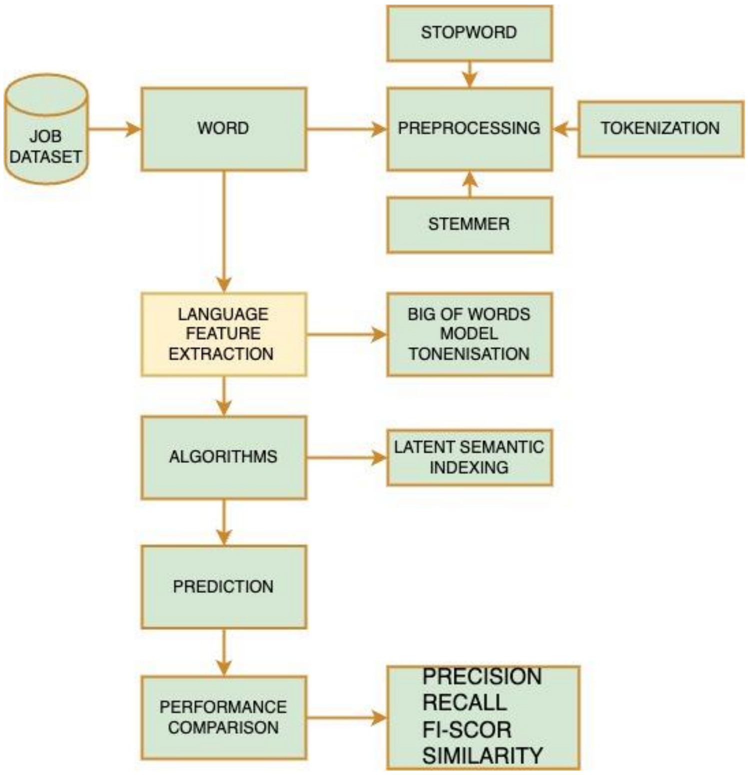

Figure 3. Research outline

Furthermore, word preprocessing is a collection of work data that will undergo a series of text preprocessing processes consisting of three stages (stopwords, Preprocessing, and stemmer). After pre-processing, the next step is the extraction of language features. The processed text is converted into numerical representations using the Bag of Words model [26], [27]. This model records the frequency of each word appearing in a document. The next step at this stage is tokenization, where the text is separated into tokens for further processing.

The resulting bag of words model is fed into the Latent Semantic Indexing (LSI) algorithm. LSI is a dimensionality reduction technique used in text processing to find semantic relationships between words in a document. This allows for the identification of more complex patterns and relationships that may not be directly visible in the text. Using LSI improves our ability to interpret and understand the content of the work. After the model is trained with the LSI algorithm, it then makes predictions or matches the work [28]-[31]. The model predicts a match between an existing job description in the dataset and a job description or a new job seeker. These predictions provide valuable insights to match job seekers with positions that match their skills and experience.

Finally, the prediction results are evaluated by comparing them with the actual data. Performance evaluation metrics such as precision, gain, F1 score, and similarity measure how well the model performs in predicting or matching [32]-[34]. This evaluation provides insight into the accuracy and effectiveness of the model in the context of a job matching application. As such, this framework not only helps in understanding the context of the work and generating relevant predictions but also provides tools to objectively measure the quality and effectiveness of those matches.

Through this series of stages, conventional LSI techniques can be improved and play an important role in providing a systematic and structured approach to improve job matching between job seekers and available job positions. By strengthening each step and providing the importance of performance evaluation, this framework can serve as a strong foundation for the development of more sophisticated and effective job matching systems in the future.

3. RESULTS AND DISCUSSION

The results of the work matching analysis by LSI improved on the proposed multi-language word meaning were determined based on the measurement of similarity, accuracy, precision and recall. Using words in Ireland (IE), Sweden (SE), and United Kingdom (UK), experiments were conducted on conventional LSI techniques for contextual analysis with the same or different meanings. Furthermore, compared to the results using LSI, it is improved.

3.1. Results

3.1.1. LSI conventional

The graph of the results of the conventional LSI trial shown in Figure 4, the test examination of the results of the application of traditional LSI in matching using accuracy produced results of 79%, recall of 79%, precision of 62%, and Fi-Scor of 70%.

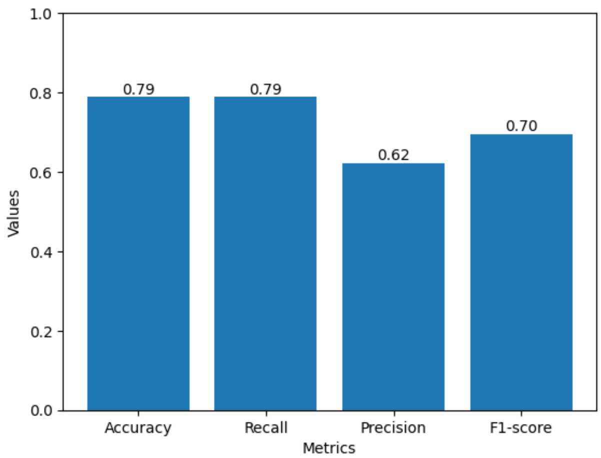

Figure 4. LSI Classification metrics

Based on the graph in Figure 4, the application of traditional LSI in the matching process shows quite good performance with an accuracy of 79%. However, while the recall also reached 79%, the precision level of only 62% indicates that there are a number of less relevant results that are also considered suitable. With an F1-Score of 70%, it can be concluded that while this method is quite effective, there is still room for improvement in terms of precision to ensure more relevant and accurate results.

Furthermore, it can be seen in Figure 5 showing a graph of the increase in LSI results on the similarity score. This graph illustrates a significant improvement in the performance of conventional LSI techniques, which is indicated by a 50% increase in similarity scores. Based on Figure 5 illustrates a significant improvement in the performance of conventional LSI techniques, which is shown by a 50% increase in similarity scores. However, this increase suggests that the similarity of meanings from multi-language contexts is still low, indicating the need for further development to address the complexity and wider variety of languages.

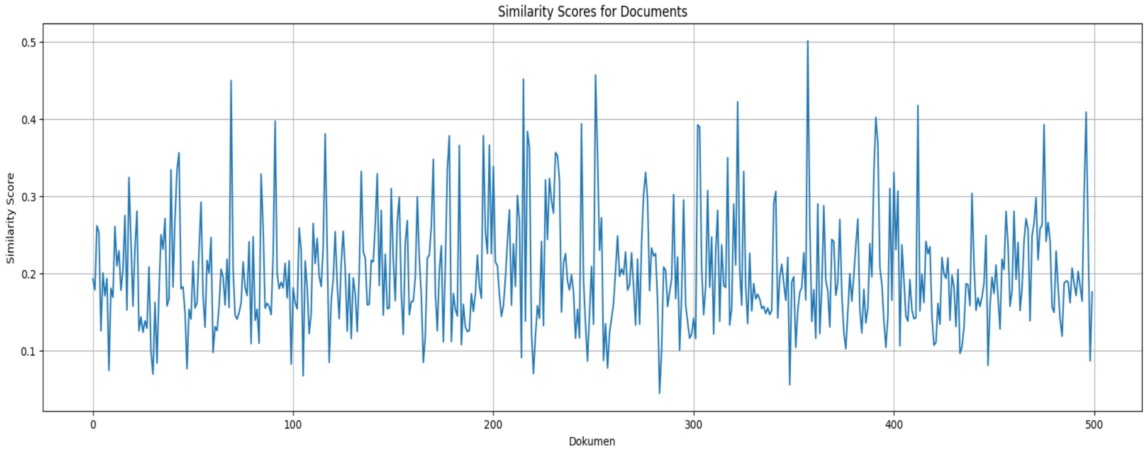

Figure 5. Graph of LSI results improvement on similarity scores

3.1.2. LSI improved

Based on Figure 6, the evaluation of the results of the application of LSI was improved for matching using accuracy, recall, precision, and Fi-Score by having the same value of 84%. This shows that the degree of similarity in contextual reading in multi-languages has the same power and can help analyze a word into the same meaning or the same word but different meanings.

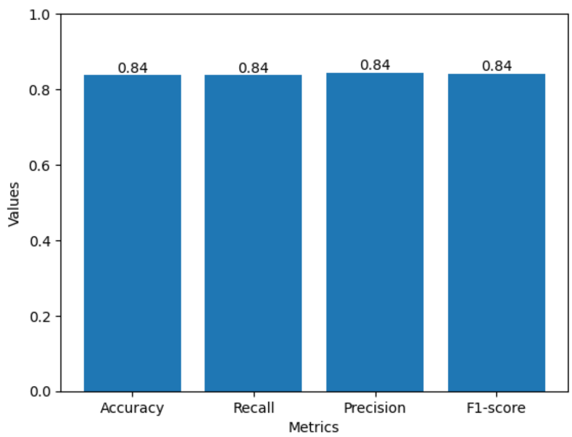

Figure 6. Classification metrics for extended LSI

Furthermore, the discussion in Figure 6, the evaluation of the results of the implementation of the improved LSI shows a significant increase in performance. This increase can be seen from the values of the evaluation metrics used, namely accuracy, recall, precision, and F1-Score, all of which reached 84%. This shows that the improved LSI method is able to provide better and more consistent matching results compared to the previous method.

In Figure 7 it can be seen that d graph has an increase in the result of Extended LSI on the similarity of words to words or to meanings. This graph illustrates a significant improvement in the performance of improved LSI techniques, which is indicated by an increase in similarity scores of 90.06%.

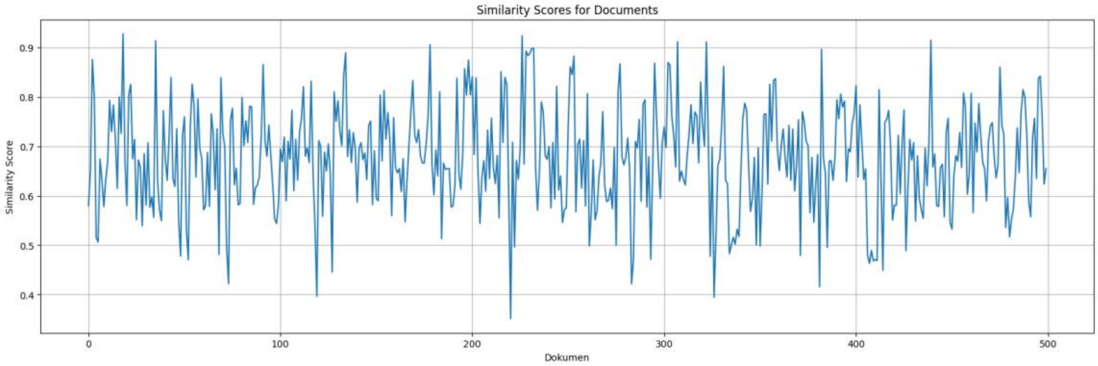

Figure 7. Improved LSI yield improvement graph on similarity scores

3.2. Discussion

The end result of vectorizing a document is a numerical vector representation with meaning and contextual information. These vectors can be used for information retrieval, analysis of word meanings, and topic modeling, among other things, to understand and classify documents according to their vector space similarities. Comparison of the results that have been studied in the previous paper [30], [31], in Table 1 have an average value of 0.788 while the results of this paper in Table 2 there is an average value of 0.571 and increased to 0.629.

Table 1. Performance comparison results

Word	Accuracy	Precision	Recall	F1-Score
1	0,7908	0,8347	0,7908	0,7003
2	0,7890	0,8335	0,7890	0,695
3	0,7899	0,7642	0,7899	0,6999
4	0,7890	0,7633	0,7890	0,6986
5	0,7881	0,7274	0,788 1	0,6964
6	0,7881	0,8330	0,7881	0,6947
7	0,7881	0,7274	0,7881	0,6964
8	0,7890	0,7633	0,7890	0,6986
9	0,7871	0,6920	0,7871	0,6959
10	0,7890	0,8335	0,7890	0,6969
Berarti	0,7888	0,7772	0,7888	0,6974
SD	0,0010	0,0505	0,0010	0,0017

Table 2. Word similarity search results

Word	The Standar LSI	The Proposed Method
1	0,844	0,906
2	0,520	0,559
3	0,309	0,473
4	0,839	0,853
5	0,349	0,358
Rata-rata	0,571	0,629

Table 1 shows an average accuracy value of 0.788848 with a standard deviation of 0.001001. Accuracy measures how many predictions are correct out of the total predictions made. In this context, accuracy measures the percentage of correctly predicted words out of the total number of words observed. Mathematically, accuracy is calculated as:

$$\text{Accuracy} = \frac{\text{Number of correct predictions}}{\text{Total Number of Predictions}}$$

From the average value of accuracy given (0.788848), it can be interpreted that the classification model correctly predicts about 78.88% of all words. About 78.88% of words have the expected label according to the actual label. While the remaining 21.12% are words whose predictions are not accurate. A low standard deviation (SD) (0.001001) indicates that the accuracy values of the tested models are close together, not too far from the average. This shows the consistency of the model's prediction performance against accuracy. In Table 3, the average value of the precision is around 0.777276. In this case, precision represents the percentage of words that are predicted accurately and correctly, or how many positive predictions the model is 100% accurate.

$$\mathrm{P}recision = \frac{\text{True Positives}}{\text{True Positives+False Positives}}$$

True positives (TP) is the number of words predicted correctly and accurately, while false positives (FP) is the number of words predicted correctly but incorrectly. So, an average precision of about 0.777276 shows that about 77.73% of the total words expected by the model to be the same are correct. In other words, about 77.73% of all the words correctly predicted by the model are also correct in the actual context. This shows how accurate the model is in identifying words that are in line with expectations. In table 3, the average recall value is 0.788848. Recall measures how much of the overall relevant item the model predicts, in this context how many words the model predicts are the same compared to the overall words that are actually the same. Mathematically, a recall is calculated as:

$$\text{Recall} = \frac{\text{True Positives}}{\text{True Positives+False Negatives}}$$

True Positives (TP) is the number of words predicted correctly and accurately, while False Negatives (FN) is the number of words predicted correctly but incorrectly predicted. So, the average recall of about 0.788848 shows that about 78.88% of the exact words are also expected to be the same by the model. In other words, about 78.88% of all the exact words were identified and predicted the same by the model. This explains how well the model can find all the correct cases in the class. A metric called F1-Score seeks to measure the alignment between gains, harmonic averages, and precision in a single number. The following formula displays the F1-Score calculation:

$$\text{F1 Score} = 2 \times \frac{\text{Precision}\times\text{Recall}}{\text{Precision+Recall}}$$

An understanding of how effectively the model balances the correct positive classification (precision) and the classification that ignores the positive (recall) is obtained by comparing precision and recall using F1-Score. From Table 3, the average F1-Score is around 0.69. This shows that on average around 69.74% of the ratio of correctly predicted words and the same weight. This is a valuable measure in evaluating the overall performance of the model, as it provides an idea of how well the model can find true positive cases and avoid false positives. The model balance between acquisition and precision is better the higher the F1-Score.

4. CONCLUSION

The rated classification model has satisfactory performance when classifying words. With an average accuracy value of about 78.88%, the model can accurately predict about 78.88% of all words. In addition, an average precision value of around 77.73% indicates that most of the words expected by the model are also completely correct in the actual context. The model's ability to recognize the same word is also quite good with an average recall value of around 78.88%. In addition, an F1 score of 69.74% indicates that the model can achieve a decent balance between precision and acquisition. As such, the classification model demonstrates adequate ability to identify and classify words, however further evaluation may be required depending on the specific needs of the problem at hand.

ACKNOWLEDGEMENTS

The author would like to thank you for the support provided. The author congratulates Abduurrab University under the leadership of Abdurrab Foundation Pekanbaru for supporting this paper so that it has obtained satisfactory results, as well as the Faculty of Computer Science and Information Technology, Universiti Tun Hussein Onn Malaysia (UTHM), especially all tutors involved in the completion of this article.

REFERENCES

[1] L. Arendt, E. Gałecka-Burdziak, F. Núñez, R. Pater, and C. Usabiaga, "Skills requirements across task-content groups in Poland: What online job offers tell us," *Technological Forecasting and Social Change*, vol. 187, p. 122245, Feb. 2023, doi: 10.1016/j.techfore.2022.122245.

[2] D. Mhamdi, R. Moulouki, M. Y. El Ghoumari, M. Azzouazi, and L. Moussaid, "Job recommendation based on job profile clustering and job seeker behavior," *Procedia Computer Science*, vol. 175, pp. 695–699, 2020, doi: 10.1016/j.procs.2020.07.102.

[3] F. G. Çalli and G. K. Orman, "Job seeker recommendation for employers: a graph-based recommendation approach using node embedding," *Procedia Computer Science*, vol. 225, pp. 3660–3669, 2023, doi: 10.1016/j.procs.2023.10.361.

[4] A. D. Margulescu, "LSI predicts durable scarring and absence of conduction gaps following pulmonary vein isolation procedures," *Journal of Atrial Fibrillation and Electrophysiology*, vol. 15, no. 4, Jul. 2022, doi: 10.4022/jafib.20200560.

[5] K. Venkatesh Prasad *et al.*, "Lesion index–guided workflow for the treatment of paroxysmal atrial fibrillation is safe and effective – Final results from the LSI workflow study," *Heart Rhythm O2*, vol. 3, no. 5, pp. 526–535, Oct. 2022, doi: 10.1016/j.hroo.2022.06.004.

[6] S. Aouichaty, Y. Maleh, A. Hajami, and H. Allali, "Automated decision classification model for tax appeals commission in Morocco using latent dirichlet allocation," *Indonesian Journal of Electrical Engineering and Computer Science*, vol. 31, no. 3, pp. 1811–1820, Sep. 2023, doi: 10.11591/ijeecs.v31.i3.pp1811-1820.

[7] M. Xu and S. Liu, "Semantic-enhanced and context-aware hybrid collaborative filtering for event recommendation in event-based social networks," *IEEE Access*, vol. 7, pp. 17493–17502, 2019, doi: 10.1109/ACCESS.2019.2895824.

[8] M. M. Haider, M. A. Hossin, H. R. Mahi, and H. Arif, "Automatic text summarization using gensim Word2Vec and K-means clustering algorithm," in *2020 IEEE Region 10 Symposium, TENSYMP 2020*, 2020, vol. 2, pp. 283–286, doi: 10.1109/TENSYMP50017.2020.9230670.

[9] Y. Kumar, K. Kaur, and S. Kaur, "Study of automatic text summarization approaches in different languages," *Artificial Intelligence Review*, vol. 54, no. 8, pp. 5897–5929, Feb. 2021, doi: 10.1007/s10462-021-09964-4.

[10] K. Watanabe, "Latent semantic scaling: a semisupervised text analysis technique for new domains and languages," *Communication Methods and Measures*, vol. 15, no. 2, pp. 81–102, Nov. 2021, doi: 10.1080/19312458.2020.1832976.

[11] R. E. Hall and S. Schulhofer-Wohl, "Measuring job-finding rates and matching efficiency with heterogeneous job-seekers," *American Economic Journal: Macroeconomics*, vol. 10, no. 1, pp. 1–32, Jan. 2018, doi: 10.1257/mac.20170061.

[12] K. Song, H. Kim, J. Cha, and T. Lee, "Matching and mismatching of green jobs: A big data analysis of job recruiting and searching," *Sustainability (Switzerland)*, vol. 13, no. 7, p. 4074, Apr. 2021, doi: 10.3390/su13074074.

[13] Y. Kino, H. Kuroki, T. Machida, N. Furuya, and K. Takano, "Text analysis for job matching quality improvement," *Procedia Computer Science*, vol. 112, pp. 1523–1530, 2017, doi: 10.1016/j.procs.2017.08.054.

[14] Q. Cheng *et al.*, "Bert-based latent semantic analysis (Bert-lsa): A case study on geospatial data technology and application trend analysis," *Applied Sciences (Switzerland)*, vol. 11, no. 24, p. 11897, Dec. 2021, doi: 10.3390/app112411897.

[15] F. Liang and X. Wan, "Jobmatching analysis based on text mining and multicriteria decision-making," *Mathematical Problems in Engineering*, vol. 2022, pp. 1–10, Sep. 2022, doi: 10.1155/2022/9245876.

[16] F. zahra El-Alami, S. Ouatik El Alaoui, and N. En Nahnahi, "Contextual semantic embeddings based on fine-tuned AraBERT model for Arabic text multi-class categorization," *Journal of King Saud University - Computer and Information Sciences*, vol. 34, no. 10, pp. 8422–8428, Nov. 2022, doi: 10.1016/j.jksuci.2021.02.005.

[17] S. Sukri, N. A. Samsudin, E. Fadzrin, S. K. A. Khalid, and L. Trisnawati, "Word2vec-based latent semantic indexing (Word2Vec-LSI) for contextual analysis in job-matching application," *International Journal of Advanced Computer Science and Applications*, vol. 15, no. 3, pp. 699–707, 2024, doi: 10.14569/IJACSA.2024.0150371.

[18] S. H. Mohammed and S. Al-Augby, "LSA & LDA topic modeling classification: Comparison study on E-books," *Indonesian Journal of Electrical Engineering and Computer Science*, vol. 19, no. 1, pp. 353–362, Jul. 2020, doi: 10.11591/ijeecs.v19.i1.pp353-362.

[19] A. Newman, Y. L. Bavik, M. Mount, and B. Shao, "Data collection via online platforms: challenges and recommendations for future research," *Applied Psychology*, vol. 70, no. 3, pp. 1380–1402, Jan. 2021, doi: 10.1111/apps.12302.

[20] N. Garg and K. Sharma, "Text pre-processing of multilingual for sentiment analysis based on social network data," *International Journal of Electrical and Computer Engineering*, vol. 12, no. 1, pp. 776–784, Feb. 2022, doi: 10.11591/ijece.v12i1.pp776-784.

[21] W. Wang, K. Zhang, H. Ren, D. Wei, Y. Gao, and J. Liu, "UULPN: An ultra-lightweight network for human pose estimation based on unbiased data processing," *Neurocomputing*, vol. 480, pp. 220–233, Apr. 2022, doi: 10.1016/j.neucom.2021.12.083.

[22] S. Kim, H. Park, and J. Lee, "Word2vec-based latent semantic analysis (W2V-LSA) for topic modeling: A study on blockchain technology trend analysis," *Expert Systems with Applications*, vol. 152, p. 113401, 2020, doi: 10.1016/j.eswa.2020.113401.

[23] A. Sharma and S. Kumar, "Ontology-based semantic retrieval of documents using Word2vec model," *Data and Knowledge Engineering*, vol. 144, p. 102110, Mar. 2023, doi: 10.1016/j.datak.2022.102110.

[24] A. Qostal, A. Moumen, and Y. Lakhrissi, "CVs classification using neural network approaches combined with BERT and gensim: CVs of Moroccan engineering students," *Data*, vol. 9, no. 6, p. 74, May 2024, doi: 10.3390/data9060074.

[25] J. Choi and S. W. Lee, "Improving FastText with inverse document frequency of subwords," *Pattern Recognition Letters*, vol. 133, pp. 165–172, May 2020, doi: 10.1016/j.patrec.2020.03.003.

[26] M. Mujahid *et al.*, "Data oversampling and imbalanced datasets: an investigation of performance for machine learning and feature engineering," *Journal of Big Data*, vol. 11, no. 1, Jun. 2024, doi: 10.1186/s40537-024-00943-4.

[27] K. M. Kuo, Y. L. Lin, C. S. Chang, and T. J. Kuo, "An ensemble model for predicting dispositions of emergency department patients," *BMC Medical Informatics and Decision Making*, vol. 24, no. 1, Apr. 2024, doi: 10.1186/s12911-024-02503-5.

[28] P. Liu *et al.*, "LinkSAGE: Optimizing job matching using graph neural networks," *arXiv preprint arXiv:2402.13430*, 2024.

[29] F. Sami, "Optimize electric automation control using artificial intelligence (AI)," *Optik*, vol. 271, p. 170085, Dec. 2022, doi: 10.1016/j.ijleo.2022.170085.

[30] J. Priambodo, "Automatic text summarization using gensim Word2Vec and K-Means clustering algorithm," *Jurnal Informatika Universitas Pamulang*, vol. 3, no. 1, pp. 39–45, 2018.

[31] I. N. Khasanah, "Sentiment classification using fasttext embedding and deep learning model," *Procedia CIRP*, vol. 189, pp. 343–350, 2021, doi: 10.1016/j.procs.2021.05.103.

[32] P. A. Hamza *et al.*, "Recruitment and selection: the relationship between recruitment and selection with organizational performance," *International Journal of Engineering, Business and Management*, vol. 5, no. 3, pp. 1–13, 2021, doi: 10.22161/ijebm.5.3.1.

[33] D. E. Cahyani and I. Patasik, "Performance comparison of tf-idf and word2vec models for emotion text classification," *Bulletin of Electrical Engineering and Informatics*, vol. 10, no. 5, pp. 2780–2788, Oct. 2021, doi: 10.11591/eei.v10i5.3157.

[34] T. W. Chang and Y. Kim, "Performance analysis of promotion programs of the smart factory using propensity score matching," *Procedia Computer Science*, vol. 232, pp. 1909–1917, 2024, doi: 10.1016/j.procs.2024.02.013.

BIOGRAPHIES OF AUTHORS

Sukri, S.T., M.Kom is currently lecturers at Abdurrab University, Indonesia. She obtained her Master in Information Technology from the Indonesian Islamic University, Indonesia. He is currently pursuing his Ph.D. in Doctor of Philosophy of Information Technology at Tun Hussein Universiti Onn Malaysia (UTHM). His research interests include data mining and machine learning. He can be contacted at email: sukri@univrab.ac.id.

Assoc. Prof. Dr. Noor Azah Binti Samsudin is Associate Professor at the Faculty of Computer Science and Information Technology, Universiti Tun Hussein Onn Malaysia (UTHM). She holds a Ph.D. degree from the University of Queensland. Bachelor Of Computer Science, Universiti Kebangsaan Malaysia, Bachelor of Computer Science, University of Missouri, Columbia, US. She is Deputy Director of the Center for Academic Excellence and Development and Principal Investigator of the Faculty of Computer Science and Information Technology, Soft Computing, and Data Mining Center (SMC). Her areas of expertise are information, computer and communications technology (ICT), artificial intelligence, and machine learning. Her research area interests were then classification and feature selection. She can be contacted at email: azah@uthm.edu.my.

Dr. Ezak Fadzrin Bin Ahmad Shaubari is a lecturer at the Faculty of Computer Science and Information Technology, Universiti Tun Hussein Onn Malaysia (UTHM). His higher studies are in Doctor of Multimedia Computing Philosophy at Manchester Metropolitan University, UK. The M.Sc. degree and the B.Sc. degree are in Bachelor of Medical Engineering at the University of Bradford, UK. His areas of expertise are information, computer and communications technology (ICT), multimedia, and digital image processing systems. His research areas of interest were image processing, computer vision, medical imaging, human physiology, and medical instruments. His innovation fields of interest are biomechanics, artificial intelligence and automotive racing technology. He is the Principal Investigator, Faculty of Computer Science and Information Technology, Media and Signal Processing (MSP), Head of Department International Office, and Allied Investigator, Faculty of Mechanical Engineering and Manufacturing, Research Center for Unmanned Vehicles (RECUV). He can be contacted at email: ezak@uthm.edu.my.

Dr. Shamsul Kamal Bin Ahmad Khalid is lecturer at the Faculty of Computer Science and Information Technology, Universiti Tun Hussein Onn Malaysia (UTHM). The B.Sc. degree in Bachelor of Computer Science from New York University, US, the M.Sc. degree in Bachelor of Science (Computer), Universiti Kebangsaan Malaysia, and the Doctor of Philosophy in Information Technology from Tun Hussein Universiti Onn Malaysia (UTHM), Malaysia. He is the Principal Investigator, Faculty of Computer Science and Information Technology, Center of Information Security Research (CISR), Universiti Tun Hussein Onn Malaysia (UTHM). His areas of expertise are Information, Computer and Communications Technology (ICT), security systems, and security services (including digital forensics, steganography, network security, public key infrastructure, and biometrics). His research area of interest was watermarking, and his innovation fields of interest are GPS-based applications and the Internet of Things. He can be contacted at email: shamsulk@uthm.edu.my.

Liza Trisnawati, S.T., M.Kom is a lecturer at Abdurrab University, Indonesia. She obtained her Master in Information Technology from Putera Indonesia University (YPTK, Indonesia). He is currently pursuing his Ph.D. in Doctor of Philosophy of Information Technology at Tun Hussein Universiti Onn Malaysia (UTHM). Her research interests include data mining and machine learning. She can be contacted at email: liza.trisnawati@univrab.ac.id.