Brilliance
Research of Artificial Intelligence

E-ISSN : 2807-9035
Volume 5, Number 1, May 2025
https://doi.org/10.47709/brilliance.v5i1.6456

Implementation Of Word Embedding In Book Recommendations Based On Descriptions

Albertus Dwi Andhika Putra Darmansius[1*], Hafiz Irsyad[2]
[1,2]Multi Data Palembang University, Indonesia
[1]dhikaalbertus@mhs.mdp.ac.id, [2]hafizirsyad@mdp.ac.id

open access

***Corresponding Author**

Article History:
Submitted: 10-07-2025
Accepted: 13-07-2025
PUBLISHED: 21-07-2025

Keywords:
Book Recommendation System; CBOW; Knowledge-Based Filtering; Skip-gram; Word2vec.

Brilliance: Research of Artificial Intelligence is licensed under a Creative Commons Attribution-NonCommercial 4.0 International (CC BY-NC 4.0).

ABSTRACT

The rapid development of digital libraries and online bookstores has increased the need for intelligent book recommendation systems that can understand user preferences and provide relevant suggestions. However, many existing systems rely on simple keyword matching or collaborative filtering, which often fail to capture the semantic meaning of complex user descriptions. This study aims to develop and evaluate a content-based book recommendation system that combines Word2Vec word embedding models with Knowledge-Based Filtering to improve the relevance of recommendations based on user-provided descriptions. The proposed system utilizes two Word2Vec architectures, Continuous Bag of Words (CBOW) and Skip-gram, to learn semantic relationships between words in book descriptions and user inputs, while Knowledge-Based Filtering incorporates explicit attributes such as publication year, genre, author, and book length to refine the results. The system was tested using a descriptive query: ***"a good fiction story telling a boy school at great magic school, published on 1995-1999"***. The evaluation, measured by Precision@K and Recall@K at K = 5, 10, and 20, shows that CBOW outperformed Skip-gram, achieving a perfect Precision@5 of 1.00 and balanced precision and recall at higher K values, while Skip-gram exhibited more variability at small K. These results indicate that CBOW is more effective in providing stable and highly relevant recommendations at the top of the list. This research confirms that combining semantic embedding and knowledge-based approaches enhances the accuracy and flexibility of recommendation systems. Further studies can explore diverse datasets and user interfaces to broaden practical applications in digital library and e-commerce platforms.

INTRODUCTION

Information science, including printed and digital literature such as books, is key in bridging the potential of technology with its practical application in various sectors (Lidjali et al., 2024). Books have long been a source of knowledge, entertainment and inspiration for humans as a window to an infinite world. In this development of information technology, accessibility to books is increasing. With the abundance of book choices available, it can cause confusion for readers in choosing books that suit their interests and needs. This phenomenon is known as the "Paradox of Choice", the Paradox of Choice is scientifically defined as a phenomenon in which too many choices actually reduce consumer satisfaction and increase the likelihood of not choosing at all (Mirzaee, 2025).

Advances in information technology have made it easier to access information. The information you are looking for can be obtained easily and quickly and can be used for personal or group purposes (Ritdrix A H, 2018). Based on data obtained from the National Library of Indonesia, the distribution of ISBN numbers continues to increase from 2019 to 2023. From 2019 to October 2023, the number of book titles that have been registered with the ISBN is 675,301 titles (Dharmawan et al., 2023). This shows that the number of books available continues to grow over time. With so many titles available, it is difficult for a person to find a book that suits personal preferences.

Research conducted by Saraswati H. D. explained that there are still many users who still have difficulty searching for books based on keywords from the subject or topic they know because the accuracy achieved does not match the book the reader wants (Saraswati, 2024). Users who enter keywords according to the theme or content of the book they are looking for are often faced with irrelevant search results or even no results at all. In a study conducted by Aman Ahluwalia and colleagues, it was said that keyword search systems often fail to capture the true intentions of users and tend to produce irrelevant results (Ahluwalia et al., 2024). This problem reflects that the system is not yet fully capable of understanding the semantic relationship between keywords and the content of books in the database. In other words, the system still relies heavily on literal word matching, without considering the broader context of meaning of user input.

In previous research conducted by Rahmadhani S. and friends explained that not all keywords entered by users will produce optimal book search recommendations (Rahmadhani et al., 2024). This is related to the analysis of the

This is an Creative Commons License This work is licensed under a Creative Commons Attribution-NonCommercial 4.0 International License.

Brilliance
Research of Artificial Intelligence

E-ISSN : 2807-9035
Volume 5, **Number** 1, May 2025
https://doi.org/10.47709/brilliance.v5i1.6456

existing system in the book search system in the library of the University of Multi Data Palembang (UMDP). Based on the observation, it was found that the search system still relies on a literal keyword matching method, which is designed based on simple or explicit queries from users. As a result, when the system searches with only partial words such as book titles or phrases representing book contents, the system does not display relevant results, which proves that the keyword book search mechanism still does not handle more specific inputs accurately.

A recommendation system is a software tool and technique that provides suggestions for items that are likely to be of interest to users (Alkaff, Khatimi, & Eriadi, 2020). In this case, the book recommendation system is one of the solutions that can help in finding books that match the books chosen by the reader. The book recommendation system aims to present the most relevant book choices to users where the results of the recommendations given by the system are obtained based on the descriptions provided by users.

Knowledge-Based Filtering is a recommendation system approach that uses explicit knowledge about the domain and user preferences to generate recommendations (Uta et al., 2024). Therefore, the recommendation system model created can accept descriptive text as input. The information obtained will be used as parameters for book recommendations. Text description inputs for searching for book recommendations can include titles, genres, book descriptions, or other story elements. The book recommendation system model is design to work by utilizing the Word Embedding method to match the input book description with the existing book dataset because Word Embedding is a branch of NLP used to convert a word into a word vector, where this word vector represents a word in a vector space (Andira, 2024). The book recommendation system model is also design to work by utilizing the Knowledge-Based Filtering (KBF) and the recommendation system model designed will use methods used in Natural Language Processing (NLP) to understand and process inputted descriptions and convert them into vector representations that can be understood by computers, namely NLP techniques that study how a word can be converted into a vector using the Word Embedding method (Saranya & Amutha, 2024).

Based on the background described above, this study will focus on the use of Word Embedding and Knowledge-Based Filtering (KBF) to develop a book recommendation system model. This study employs Word Embedding and Knowledge-Based Filtering (KBF) for a book recommendation system and evaluates its performance using the Precision@K and Recall@K testing models, utilizing the Kaggle dataset "Goodreads Book Datasets With User Rating 2M." This combination is designed to design and develop a book recommendation system model based on descriptions with good accuracy. The objective of this research is to comprehensively offer better accuracy in recommending books than previous methods, thereby ensuring more accurate book recommendations. Additionally, this research aims to contribute to research on recommendation systems based on user descriptions so that books can be recommended according to user preferences.

LITERATURE REVIEW

Based on the background and problem formulation described above, a strong theoretical understanding is needed as a foundation for developing an effective recommendation system. Therefore, the following section will discuss the literature review related to this research topic.

Previous research conducted by Seegmiller and colleagues showed the fundamental weakness of traditional methods that rely on literal keyword matching, often referred to as the "bag-of-words" approach (Seegmiller, Papanikolaou, & Schmidt, 2022). Their study empirically demonstrated that bag-of-words methods fail to capture semantic similarity, as evidenced by an explained variance (R^2) close to zero when predicting meaningful relationships, while word embeddings achieved an R^2 of approximately 0.42. They also presented examples where two conceptually similar descriptions, such as *"dog"* and *"puppy"*, appeared entirely unrelated under a bag-of-words representation, yet correctly identified as similar using embeddings. These findings illustrate that word embedding-based methods not only outperform bag-of-words quantitatively, but also produce similarity scores more consistent with human judgment. Based on these findings, it is clear that to build recommendation systems that can understand users' narrative descriptions, bag-of-words approaches (e.g., TF-IDF) are inadequate. Therefore, Word Embedding was chosen as the text representation method because it is explicitly designed to understand semantic meaning and overcome the failure of keyword matching, thus enabling the recommendation model to understand the user's description instead of simply matching literal words. As emphasized by Lee and colleagues, accurate recommendation systems in the modern era rely heavily on their ability to process sequential and contextual information, a capability that is inherent in Word Embedding but absent in classical algorithms (Lee, Kim, Lee, & Jun, 2023).

Mutsaddi's research highlights the fatal flaws of traditional bag-of-words-based text representation approaches such as TF-IDF by comparing TF-IDF and BERT embeddings in detecting plagiarism in Marathi texts (Mutsaddi & Choudhary, 2025). This issue has been empirically proven through document classification experiments, where Mutsaddi compared TF-IDF and BERT embeddings in detecting plagiarism in Marathi texts. The results show that although TF-IDF is capable of capturing statistical information with an accuracy of 58.68%, the TF-IDF model is said to fail to understand semantic context and inter-sentence relationships because TF-IDF only focuses on the presence of the same keywords. In contrast, BERT significantly improved accuracy thanks to its ability to understand meaning and context with an accuracy of 80.64%. This is because BERT Word Embedding can capture the meaning and semantic

This is an Creative Commons License This work is licensed under a Creative Commons Attribution-NonCommercial 4.0 International License.

Brilliance
Research of Artificial Intelligence
E-ISSN : 2807-9035
Volume 5, Number 1, May 2025
https://doi.org/10.47709/brilliance.v5i1.6456

relationships between words and sentences and is more stable anad accurate for semantic classification or implicit meaning detection, making TF-IDF less effective for high-semantic detection tasks such as plagiarism.

One method in book recommendation systems is Collaborative Filtering (CF), which provides recommendations based on user preferences. The cold start problem occurs when the system cannot provide recommendations to new users because it does not have interaction history data (such as ratings) from these users. According to Zhang's research on the cold start problem in recommendation systems, which focuses on new users and items without previous interactions, "The cold start problem is one of the long-standing challenges in recommender systems, focusing on accurately modeling new or interaction-limited users or items to provide better recommendations" (Zhang et al., 2025). The failure of Collaborative Filtering (CF) prompted the selection of an approach that does not rely on historical data, but rather on knowledge explicitly provided by users. This is the justification for choosing the Knowledege-Based Filtering (KBF) framework, which is specifically designed to handle explicit queries without the need for historical data.

METHOD

The research followed a series of well-defined stages, as illustrated in the flowchart in Fig. 1. The process began with a literature review to understand the use of Word2Vec and Knowledge-Based Filtering methods for books recommendation by description.

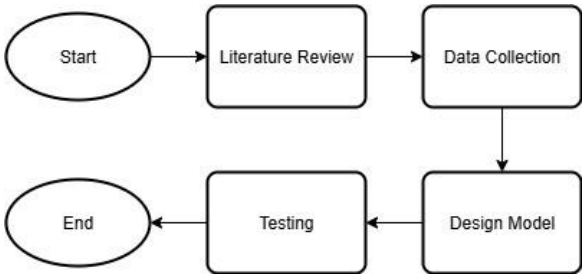


Fig. 1 Research Methodology Flowchart

This study utilized the public "Goodreads Book Datasets With User Rating 2M" dataset from Kaggle, which contains two million book data. As detailed in Table. 1, the dataset has several attributes.

Table 1. Book Dataset Used

Column	DType
RatingDist4	int64
RatingDist1	int64
ISBN	object
Authors	object
Id	int64
pagesNumber	float64
Language	object
RatingDist3	int64
Name	object
PubishYear	int64
CountOfReview	int64
RatingDist5	int64
PublishMonth	int64
RatingDist2	int64
PublishDay	int64
RatingDistTotal	int64
Rating	float64
Publisher	object
Description	object
Count of text reviews	float64
PagesNumber	float64

This dataset was collected from Goodreads, a social catalog website that allows anyone to freely search its database of books, annotations, and reviews. The collection of book information from Goodreads API to analyze the data of books in the world. About 10,000,000 books are available in the site's archive, and this project collects them to query the API using the Goodreads Python library.

The system design involves a book dataset that includes data in the form of book title (Name), description

This is an Creative Commons License This work is licensed under a Creative Commons Attribution-NonCommercial 4.0 International License.

Brilliance
Research of Artificial Intelligence

E-ISSN : 2807-9035
Volume 5, **Number** 1, May 2025
https://doi.org/10.47709/brilliance.v5i1.6456

(Description), year of publication (PublishYear), authors (Authors), and other relevant information. Then text preprocessing is carried out such as tokenization, case folding, stopword removal and stemming.

In the preprocessing stage, stopword removal is applied using the standard English stopword list provided by the Natural Language Toolkit (NLTK), which contains common words such as "*the, a, an, of, in, at*", etc., that are considered to carry little semantic weight. This is followed by stemming to reduce words to their root forms, performed using the Porter stemming algorithm, which is widely used for its simplicity and effectiveness in English text. These steps help reduce noise and standardize the textual data for more effective vectorization.

To address the issue of out-of-vocabulary (OOV) words that is, words in the user query that are not present in the Word2Vec model's vocabulary, the system ignores such words during vector construction. Specifically, only words found in the trained Word2Vec vocabulary are included when averaging vectors to represent the query or the book descriptions, while OOV words are skipped to prevent them from introducing errors into the similarity calculation.

The cleaned and preprocessed dataset is then processed using the Word Embedding technique with the Word2Vec model, train with both the CBOW or Skip-gram architecture to produce a vector representation of the words in the book description. The resulting vectors of the user query and the book descriptions are compared using cosine similarity to identify semantically similar books. After the book vector is available, the system receives input from the user in the form of a description of the book being searched. The input from the user will go through the same preprocessing process.

The processed description is then converted into a vector using the Word2Vec model that has been trained previously. After that, a similarity calculation is performed between the user description input vector and the book description vector in the dataset to identify books with the highest semantic similarity. The resulting similarity value becomes the basis for determining the suitability between the user's wishes and the books in the system. After that, the system applies the Knowledge-Based Filtering method to filter books based on explicit attributes such as year of publication (PublishYear), authors (Authors), or certain keywords that match the domain rules. After that, the recommendation system model displays a list of the most relevant book recommendations that match the user's description as the output of this recommendation model.

The previously created model will perform a system testing stage that aims to evaluate how well and accurately the system provides book recommendations based on user descriptions. The evaluation is carried out by measuring the suitability between the results of the system recommendations and the actual user preferences. The model testing stage will use Precision@K and Recall@K value measurements, which are commonly used in recommendation systems to measure the relevance of results.

Precision@K is used to indicate the ratio of relevant items successfully recommended by the system in the Top-K list of recommendation results, given by equation (1), where, the higher the precision value, the more precise the system is in displaying relevant books from the user's perspective.

$$Precision@K = \frac{|RelevantItem \cap RecommendedItems@K|}{K} \quad (1)$$

Recall@K is used to determine the extent to which the system is able to display all relevant items that should be recommended, given by equation (2). This can help evaluate the system's coverage of the user's overall needs.

$$Recall@K = \frac{|RelevantItem \cap RecommendedItems@K|}{|RelevantItems|} \quad (2)$$

RESULT

This section provides an in-depth analysis of the experimental results aimed at finding relevant book recommendation results for the CBOW and Skip-gram architectures. During the testing phase, the Word2Vec-based book recommendation system model and Knowledge-Based Filtering were evaluated to measure the relevance of the recommendation results to user descriptions. The evaluation was conducted using the Precision@K and Recall@K metrics at values of K = 5, 10, and 20. This step aims to show the ratio of relevant items successfully recommended by the system, as well as to determine the extent to which the system is able to display all relevant items that should be recommended. The test was conducted using the following user description ***"a good fiction story telling a boy school at great magic school, publish on 1995-1999"*** which is a complex yet general description containing semantic keywords about a fictional story involving a boy at a magic school published within a specific time frame.

The complete results of this experiment are presented in Table 2 for CBOW and Table 3 for Skip-gram.

Table 2. Results CBOW Architecture

K	Precision@K	Recall@K
5	1.00	0,217
10	0.90	0,391
20	0.85	0,739

This is an Creative Commons License This work is licensed under a Creative Commons Attribution-NonCommercial 4.0 International License.

Brilliance
Research of Artificial Intelligence

E-ISSN : 2807-9035
Volume 5, **Number** 1, May 2025
https://doi.org/10.47709/brilliance.v5i1.6456

The results of testing at Table 2 present the test using Word2Vec CBOW architecture approach explain that CBOW can predict target words based on context (average surrounding words). The results show that the system performs very well with CBOW architecture, where achieved a Precision@5 of 1.00, demonstrating its ability to recommend only relevant books in the top five results and substantially outperforming traditional keyword matching methods, which have been shown in previous studies to produce sparse and inaccurate similarity scores in descriptive queries. At larger K values, precision decreased slightly while recall improved, reflecting broader coverage of relevant books as more results were considered.

To confirm the system's ability to handle the cold start problem, where a new user has no prior interaction history, the test relied solely on the descriptive query without any historical user data. The system was able to recommend books that matched both the semantic meaning and explicit constraints (publication year 1995–1999), validating its effectiveness in overcoming the cold start issue.

At K = 5, the perfect Precision@5 (1.00) indicates that the system successfully presents completely relevant recommendations at the top. Recall@5 is at 0.217, indicating that only a small portion of the total relevant items were captured in the top five.

When K is increased to 10, Precision@10 decreases slightly to 0.90, while Recall@10 increases to 0.391. At K = 20, Precision@20 remains fairly high at 0.85, while Recall@20 reaches its highest value, 0.739. This indicates that the system is able to cover up to 73.9% of relevant items when given more space (K = 20).

Table 3. Results Skip-gram Architecture

K	Precision@K	Recall@K
5	0.80	0,174
10	0.90	0,391
20	0.85	0,739

The results of testing at Table 3 present the test using the Word2Vec Skip-gram architecture approach explain that Skip-gram is known to be more effective in capturing associations between words that rarely appear, because this model learns to predict the context of the target word.

Regarding the claim that Skip-gram is more sensitive to rare words, an analysis of the recommended books revealed that Skip-gram indeed tended to select titles with less frequent and more specific terms, such as *"wizardry"*, *"headmaster"*, *"enchantment"*, and *"sorcery"*. In contrast, CBOW prioritized more common and general terms like "magic" and "school". This observation empirically confirms that Skip-gram better captures associations involving rare words, although it exhibits less stability in top-ranked results, as reflected by its lower Precision@5 of 0.80 compared to CBOW's perfect score.

A Precision@5 of 0.80 indicates that 80% of the top 5 recommendations are truly relevant to user needs. A Recall@5 of 0.174 indicates that only about 17.4% of all relevant books are captured when only the first 5 recommendations are considered. This means that the top list of Skip-gram is already quite good, but not as precise as CBOW, which achieves 1.00. Precision@10 increases to 0.90, while Recall@10 rises to 0.391. This shows that as the number of recommendations increases, the Skip-gram system can cover more relevant books while maintaining a high level of relevance. Precision@20 is at 0.85, while Recall@20 reaches 0.739. At this point, the system has successfully identified approximately 73.9% of all relevant books, demonstrating broad coverage of all relevant books in the dataset.

In this case, the user description contains specific terms such as "magic school," "boy," "fiction," and "publish on 1995-1999," some of which are rare combinations. Therefore, Skip-gram can recognize less common but relevant semantic patterns, although it sometimes sacrifices stability in the top recommendations (small K). This explains why at K=5, its precision is slightly lower than CBOW, but at larger K values, Skip-gram remains stable and its recall is sufficiently high.

These both results between Word2Vec architecture CBOW and Skip-gram confirm that are approach is effective in understanding the semantic meaning of user descriptions, including complex and lengthy ones. The addition of Knowledge-Based Filtering based on explicit attributes (year, author, title, book length, etc.) strengthens the alignment of results with users' explicit preferences. The combination of both methods enables the system to find recommendations that are not only semantically relevant but also meet explicit constraints, such as genre or book length.

Overall, combining both methods allows the system to find recommendations that are not only semantically relevant but also meet explicit constraints, such as genre or book length, and allows the system to findings demonstrate that the system effectively addresses both the semantic relevance and cold start challenges, while also empirically validating the distinct strengths of CBOW and Skip-gram in generating book recommendations from descriptive queries. The system also demonstrates good stability against variations in K length, with precision remaining high even as recall increases significantly at larger K values. This makes the system flexible for users who want short recommendations with high relevance, as well as longer lists with broader coverage.

This is an Creative Commons License This work is licensed under a Creative Commons Attribution-NonCommercial 4.0 International License.

Brilliance
Research of Artificial Intelligence

E-ISSN : 2807-9035
Volume 5, **Number** 1, May 2025
https://doi.org/10.47709/brilliance.v5i1.6456

DISCUSSION

The findings from the experimental evaluation confirm that the proposed book recommendation system effectively addresses the limitations of traditional keyword matching while also mitigating the cold start problem. The CBOW architecture demonstrated superior performance at small K, achieving a perfect Precision@5 of 1.00, which underscores its ability to deliver highly relevant recommendations at the top of the list. This result validates the system's effectiveness in overcoming the documented shortcomings of literal keyword matching, which fails to capture context and synonyms, often resulting in sparse and misleading similarity scores in descriptive queries.

The evaluation results show some important insights, that the Word2Vec-based book recommendation system with CBOW architecture and Knowledge-Based Filtering provides overall better results compared to Skip-gram. For the user description ***"a good fiction story telling a boy school at great magic school, published on 1995-1999"***, CBOW's embeddings, which are derived from predicting target words based on the average of their surrounding context, tend to produce stable and general representations that prioritize the most semantically relevant recommendations. CBOW yields Precision@5 of 1.00, Precision@10 of 0.90, and Precision@20 of 0.85, with Recall@5 of 0.217, Recall@10 of 0.391, and Recall@20 of 0.739.

In contrast, Skip-gram, although achieving comparable performance to CBOW at larger K values (K=10 and K=20), exhibited more variability at K=5, reflected in its lower Precision@5 of 0.80. This aligns with the nature of Skip-gram, which is more sensitive to rare words and specific associations, as confirmed by the analysis of recommended titles that included uncommon terms like *"wizardry"*, *"headmaster"*, and *"enchantment"*. While this makes Skip-gram advantageous in scenarios where rare or niche terms are important, it also compromises the stability and precision of the top-ranked recommendations.

Compared to the Skip-gram architecture, the CBOW architecture shows a significant advantage at small K (especially at K=5), where all recommendations provided are fully relevant to the user's description. This indicates that the embedding generated by CBOW is more stable and tends to represent the general meaning well, making it suitable for providing precise results at the top position.

Meanwhile, Skip-gram, despite having equivalent performance at K=10 and K=20, is unable to match the precision of CBOW at K=5. This is in line with Skip-gram's nature of focusing more on rare words and more variable embedding, which may cause the results at the top to be less stable.

The system also demonstrated its ability to handle the cold start problem effectively. By relying solely on the user's descriptive input without requiring any prior interaction data, the system successfully generated recommendations that matched both the semantic meaning and explicit constraints stated by the user, such as publication year. This capability is critical in real-world applications where new users or infrequent users may not have sufficient historical data.

This finding reinforces the view that CBOW is superior for recommendations that prioritize high precision at the top of the list. By combining Word2Vec CBOW and Knowledge-Based Filtering, the system is able to provide relevant, accurate results and effectively fulfill the user's explicit preferences.

CONCLUSION

This study demonstrated the effectiveness of a book recommendation system that integrates Word2Vec word embeddings with Knowledge-Based Filtering in addressing the limitations of traditional keyword matching and the cold start problem. The system successfully interprets complex, descriptive user queries by leveraging semantic embeddings to capture meaning beyond literal keywords, while knowledge-based filtering ensures that explicit user preferences—such as year, genre, author, and book length are respected.

Among the two Word2Vec architectures tested, CBOW consistently showed superior performance at small K values, achieving a perfect Precision@5 of 1.00, which is crucial for ensuring that the most relevant recommendations appear at the top of the list. At larger K values (K=10 and K=20), CBOW maintained stable performance comparable to Skip-gram, with a good balance between precision and recall. Although Skip-gram demonstrated its strength in identifying rare and specific terms, it was less stable than CBOW in the highest-ranked recommendations.

These findings confirm that the integration of CBOW and Knowledge-Based Filtering is an effective and reliable approach for book recommendation systems by description, particularly when relevance and alignment with explicit user input are prioritized. Overall, this research contributes by providing a practical and efficient solution for book recommendation systems based on diverse and descriptive user inputs. Future work could explore the application of this model to larger and more varied datasets and enhance its usability through improved user interfaces and real-time interaction to support adoption in library and digital bookstore environments.

REFERENCES

Ahluwalia, A., Sutradhar, B., Ghost, K., Yadav, I., Sheetal, A., & Patil, P. (2024). Hybrid Semantic Search: Unveiling User Intent Beyond Keywords. *ArXiv*.

Alkaff, M., Khatimi, H., & Eriadi, A. (2020). Sistem Rekomendasi Buku pada Perpustakaan Daerah Provinsi

This is an Creative Commons License This work is licensed under a Creative Commons Attribution-NonCommercial 4.0 International License.

E-ISSN : 2807-9035
Volume 5, **Number** 1, May 2025
https://doi.org/10.47709/brilliance.v5i1.6456

Kalimantan Selatan Menggunakan Metode Content-Based Filtering. *MATRIK : Jurnal Manajemen, Teknik Informatika Dan Rekayasa Komputer*, *20*(1), 193–202. https://doi.org/10.30812/matrik.v20i1.617

Andira, A. (2024). *ANALISIS PERBANDINGAN KINERJA METODE WORD EMBEDDING GLOVE DAN WORD2VEC DALAM ANALISIS SENTIMEN SKRIPSI*.

Dharmawan, H., Shofiah Hilabi, S., Kaniawulan, I., Informasi Fakultas Ilmu Komputer Universitas Buana Perjuangan Karawang Jl Ronggo Waluyo Sirnabaya, S., Timur, T., & Barat, J. (2023). Sistem Rekomendasi Buku Dengan Metode K-Nearest Neighbor Pada Gramedia SISTEM REKOMENDASI BUKU DENGAN METODE K-NEAREST NEIGHBOR (K-NN) PADA GRAMEDIA. In *Sistem Informasi* (Vol. 5).

Lee, T. H., Kim, S., Lee, J., & Jun, C. H. (2023). Word2Vec-based efficient privacy-preserving shared representation learning for federated recommendation system in a cross-device setting. *Information Sciences*, *651*. https://doi.org/10.1016/j.ins.2023.119728

Lidjali, N., Pongoh, D. S., Tongkotou, A., Sondak, N., Mewaikere, S., Baduang, P., … Utara, S. (2024). The Role of Information Science in the Development of Science and Technology. *Jurnal Syntax Admiration*, *5, No. 12*(12).

Mirzaee, M. (2025). *Operations Management under Paradox of Choice*.

Mutsaddi, A., & Choudhary, A. (2025). *Enhancing Plagiarism Detection in Marathi with a Weighted Ensemble of TF-IDF and BERT Embeddings for Low-Resource Language Processing*. Retrieved from https://github.com/aditya-choudhary599/

Rahmadhani, S., Hakim, L., Hendra Wibowo, G., Purnama Kristanto, S., Mistiko Rini, E., Studi Teknologi Rekayasa Perangkat Lunak, P., … Negeri Banyuwangi, P. (2024). JIP (Jurnal Informatika Polinema) SISTEM REKOMENDASI PENELUSURAN BUKU BERBASIS CONTENT-BASED FILTERING DENGAN PEMBOBOTAN TF-RF. *JIP (Jurnal Informatika Polinema)*, *10*(4).

Ritdrix A H. (2018). *SISTEM REKOMENDASI BUKU MENGGUNAKAN METODE ITEM-BASED COLLABORATIVE FILTERING*.

Saranya, M., & Amutha, B. (2024, May). *A Survey of Machine Learning Technique for Topic Modeling and Word Embedding*. 1–6. https://doi.org/10.1109/ICACCS60874.2024.10716820

Saraswati, H. D. (2024). *Pentingnya Penambahan Subjek dengan Pemanfaatan Tesaurus pada Sistem Pengolahan Bahan Perpustakaan*.

Seegmiller, B., Papanikolaou, D., & Schmidt, L. D. W. (2022). Measuring Document Similarity With Weighted Averages of Word Embeddings. *ScienceDirect*. https://doi.org/https://doi.org/10.1016/j.eeh.2022.101494

Uta, M., Felfernig, A., Le, V. M., Tran, T. N. T., Garber, D., Lubos, S., & Burgstaller, T. (2024). Knowledge-based recommender systems: overview and research directions. *Frontiers in Big Data*, Vol. 7. Frontiers Media SA. https://doi.org/10.3389/fdata.2024.1304439

Zhang, W., Bei, Y., Yang, L., Zou, H. P., Zhou, P., Liu, A., … Yu, P. S. (2025). *Cold-Start Recommendation towards the Era of Large Language Models (LLMs): A Comprehensive Survey and Roadmap*. Retrieved from http://arxiv.org/abs/2501.01945

This is an Creative Commons License This work is licensed under a Creative Commons Attribution-NonCommercial 4.0 International License.