|  |  |
| --- | --- |
| Improved Recommender System Using Neural Network Collaborative Filtering for E-commerce Cosmetic Product |  |

**Subhan Subhan1\*, Deny Lukman Syarif1, Endah Widhihastuti2, Senda Kartika Rakainsa2, Yahya Nur Ifriza1**

1Computer Science Department, Universitas Negeri Semarang

2Pharmacy Study Program Universitas Negeri Semarang

|  |  |
| --- | --- |
| ***Abstract***  *This study presents an enhanced recommender system tailored for e-commerce platforms specializing in cosmetic products. Traditional recommender systems often face challenges in providing accurate and personalized recommendations due to the complexity and subjectivity inherent in cosmetic preferences. In e-commerce, personalized product recommendations are key to improving user engagement and driving sales. This paper presents an innovative approach to enhance recommendation systems by leveraging neural network collaborative filtering techniques for the cosmetic product domain. The proposed method integrates neural networks into collaborative filtering, namely neural network collaborative filtering with improved preprocessing step. To validate the effectiveness of our proposed system, extensive experiments were conducted using real-world e-commerce cosmetic datasets “eCommerce Event History in Cosmetics Shop”. Additionally, we evaluate the system's performance using historical e-commerce event data in cosmetics stores, demonstrating the system's effectiveness with mean reciprocal ratings (MRR ) and normalized discount cumulative gain (NDCG). Evaluation Metrics of MRR and NDCG in this study got 0.56 and 0.60, respectively, with a split of the data: 70% train data, 15% validation data, and 15% test data. This study obtains better evaluation metrics than the previous study. Furthermore, this model exhibits robustness against data sparsity and cold-start problems commonly encountered in e-commerce platforms. This research advances knowledge of recommendation systems and paves the way for more personalized and efficient online shopping experiences.*  *This is an open access article under the* [*CC BY-SA*](http://creativecommons.org/licenses/by-sa/4.0/) *license* | ***Keywords:***  *Collaborative Filtering;*  *Cosmetics*  *E-commerce;*  *Neural Network;*  *Recommender System;*  ***Article History:***  *Received: May 2, 2019*  *Revised: May 29, 2019*  *Accepted: June 2, 2019*  *Published: June 2, 2019*  ***Corresponding Author:***  *Subhan*  *Computer Science Department, Universitas Negeri Semarang, Indonesia*  *Email:* subhan@mail.unnes.ac.id |

**INTRODUCTION**

Recommender systems play a crucial role in eCommerce, especially in the cosmetics industry, by helping customers make purchase decisions. These systems analyze user interactions, such as clicks, views, and purchases, to understand their preferences and predict the following item(s) for recommendation [1]. However, the ever-growing scale of products and users poses challenges in accurately and efficiently matching products to potential customers. Recommender systems are essential for providing personalized cosmetic product recommendations based on user attributes such as age, skin type, and preferences. These systems leverage ingredient analysis to recommend products that align with the user's desired cosmetic effects and skin type, addressing the challenge of interpreting complex ingredient lists[2], [3], [4], [5].

By utilizing machine learning and AI techniques, recommender systems enhance decision-making by providing accurate and reliable cosmetic recommendations [2], [6]. Recommender systems contribute to user satisfaction and confidence by enabling users to compare and visualize the differences between cosmetic items they own and those they are considering purchasing, thus facilitating informed decision-making [2], [7], [8]. These systems also aid in estimating the efficacy of cosmetics based on their ingredients, catering to the increasing demand for personalized and effective cosmetic recommendations [2].

Recommender systems utilize various techniques, including traditional approaches like sequence similarity and frequent pattern mining, factorization and latent representation methods like matrix factorization, and neural network-based strategies. Additionally, incorporating temporal information and domain knowledge specific to the fashion industry can significantly improve recommendation performance [9]. In cosmetics, a recommendation system based on ingredient analysis and machine learning has been proposed to help users choose suitable products [10].

The changing times impact the use of technology in different areas of human life [11]. These developments have a significant influence on trading activities. Activities that used to be done in person can now be done through the online store. Online stores influence consumers' habits when searching for or buying products. Indirectly, this habit affects how buyers make purchasing decisions [12]. This decision can be influenced by several factors, such as product price, product brand, product quality, product reviews from other users, and even the level of confidence in evaluating a product [13]⁠. With these many factors, it is necessary to have a system that can always attract or increase the attractiveness of products in e-commerce to buyers.

One feature that can increase buyers' attractiveness is the recommendation system. The recommendation system displays products to buyers by collecting and analyzing compatibility between products and consumer characteristics. It is one of the features that has a vital role in determining consumer decisions in e-commerce [14]. One business application has implemented a recommendation system, namely automated fashion, to suggest suitable items to consumers [15].

There are generally four types of recommendation systems, namely content-based, collaborative, context-aware, and hybrid filtering [16]. The technique often used is collaborative filtering (CF), which has two methods: memory-based and model-based. In memory-based, there are two techniques, namely user-based and item-based. However, these two techniques are classic in recommendation algorithms that ignore the interaction between the user and the product. However, a new technique has emerged with deep learning that can be integrated with recommendation systems, namely the Neural Network Collaborative Filtering (NNCF). NNCF is a recommendation system technique that combines neural networks with collaborative filtering algorithms[17].

NNCF aims to improve the accuracy of recommendations by modeling the interactions between users and items using neural architectures. NNCF combines the strengths of collaborative filtering algorithms, such as user-based and item-based filtering, with the power of neural networks[17]. This integration allows for more accurate and personalized recommendations[18]. Unlike traditional collaborative filtering algorithms that use dot product operations, NNCF replaces this operation with a multi-layer perceptron, enabling the model to learn complex patterns and capture the collaborative filtering effect. NNCF addresses the cold start problem, where the system lacks sufficient data for new users or items, by integrating item features before the concatenation step in the model[12] on memory-based depending on individual preferences that are available after being registered in the system and have used the system for some time. But in reality, many users are not logged in or using the application for the first time. So, the recommendation system cannot work because there is no individual preference as a reference. This causes a cold-start problem [20].

Natural Language Processing (NLP) methods can be used to build a recommendation system have different roles for each type in the recommendation system, and can improve the recommendation system's performance [21]. The research on film recommendations using a Recurrent Neural Network (RNN) with Embedding-Weight Tying to improve performance [22]. The use of Embedding-Weight Tying on RNN-based is proven to be able to enhance recommendation system performance compared to other baseline methods. Research that used a combination of Deep Neural Network (DNN) and LSTM called DNN-LSTM to create a film recommendation system that can predict ratings on data without ratings to increase efficiency in recommending items to users [23].

Preprocessing is crucial for recommender systems in big data. Existing recommendation systems face scalability and efficiency problems due to the increasing numbers of users and products in the age of big data [24]. Data preprocessing is essential for creating input files with the appropriate format needed by a recommender system, and it helps achieve information-preserving data reduction [25]. Data quality is a prerequisite for the successful application of big data techniques, and data preprocessing is crucial for managing and using big data effectively [26]. Preprocessing plays a significant role in generating valuable data for decision-making and is not solely dependent on advanced algorithms [27]. Preprocessing big data involves dealing with challenges such as concept drift and data streams, which are considered significant challenges in the analysis process [28]. Preprocessing is essential for applying any analytics algorithm to obtain valuable patterns, and it is a key factor in the success of the analysis process in terms of efficiency and output information [29].

Neural networks, conversely, are a powerful subset of artificial intelligence inspired by the human brain's interconnected web of neurons. They excel in learning patterns and understanding complex relationships in data. By integrating neural networks with collaborative filtering, we bridge the gap between traditional recommendation systems and the evolving needs of e-commerce in the cosmetic industry. In this research, we combine NCCF with strengthening preprocessing to improve the accuracy of the recommender system.

**METHOD**

The initial stage in the research is to prepare the dataset. The dataset format used is a Comma Separated Value (CSV). The dataset is then preprocessed. The preprocessing stage is an essential step because it can affect model performance. The next step is splitting the data into three parts: training, validation, and testing. Furthermore, the training process uses the NNCF algorithm to make recommendations. After that, validation was carried out using data validation and testing using data testing to calculate the NDCG and MRR values. The final stage of this study was analyzed by comparing accuracy based on previous studies. The flowchart of the methods is shown in Figure 1.

A diagram of a flowchart

Description automatically generated

Figure 1. Flowchart of methods

**Dataset**

The dataset used is the “eCommerce Event History in Cosmetics Shop” dataset, which is public and can be downloaded on Kaggle via https://www.kaggle.com/datasets/mkechinov/ecommerce-events-history-in-cosmetics-shop. The dataset consists of 20,692,840 rows and nine columns containing data for problem-solving trials on recommendation systems. The dataset used is unsupervised, whereas data training does not use data labels. The feature of dataset can be seen in Table 1.

Table 1. Feature on Dataset

|  |  |
| --- | --- |
| Feature | Description |
| event\_time | The time of the user interacted with the product in Coordinated Universal Time (UTC) |
| event\_type | The type of interaction that occurs (view, cart, remove, purchase) |
| product\_id | ID of the product |
| category\_id | ID category of product |
| category\_code | Product categories are in the form of strings |
| Brand | The brand name of the product |
| Price | Product prices are in float form |
| user\_id | ID of user |
| user\_session | Temporary user session ID |

**Preprocessing Data**

The stage after collecting data is preprocessing the data that has been collected. This stage was carried out because not all data will be used in this study. In addition, the reason for preprocessing the data is to save resources, considering that the resources used to conduct research could be more extensive. Some of the steps that will be carried out at this preprocessing stage are data cleaning by checking for missing values, combining data into one file because the data is still separated into five files, data reduction by filtering the required data, converting data into atomic files, data splitting, and evaluation. Data reduction in preprocessing can reduce the amount of irrelevant data and increase accuracy [30]. This stage is carried out to prepare the data for the training model.

**Dataset Loading**

Dataset Loading is the stage for making datasets preprocessing results from atomic files into data in the form of interactions that will be included in the algorithm for training. Interaction is data formed from user interaction with the product.

**Splitting Data**

The splitting data is divided based on the interaction dataset into three parts: training data, validation data, and testing data, 70%, 15%, and 15%, respectively.

**Modeling NNCF**

This stage is to create an NNCF model used in the recommendation system. NNCF is a versatile approach capable of expressing and extending matrix factorization within its framework. To enhance NNCF modeling by incorporating non-linearities, NNCF suggests the utilization of a multi-layer perceptron for learning the function that captures interactions between users and items. Rigorous testing on real-world datasets demonstrates substantial enhancements in the performance of our proposed NNCF framework compared to existing state-of-the-art techniques. Empirical findings support that employing deeper neural network layers leads to superior recommendation performance. The NNCF model is shown in Figure 2 [31]. NNCF is versatile and has the capability to represent and generalize matrix factorization within its structure. To enhance the NNCF model with non-linear features, NNCF suggest incorporating a multi-layer perceptron to capture the user-item interaction function. Practical results indicate that employing deeper neural network layers results in improved recommendation outcomes.

A diagram of a software algorithm

Description automatically generated with medium confidence

Figure 2. Architecture of Neural Network Collaborative Filtering

**Evaluation of Model**

The primary purpose of the recommendation system is to recommend items according to the characteristics and wishes of users, so it is necessary to evaluate the recommendation system that is built [32]⁠. There are several ways to assess the recommendation system, namely by rating prediction metrics, classification accuracy metrics, and ranking metrics. The model used to evaluate this study's model uses the evaluation metric on the ranking metric, namely NDCG and MRR. This calculation aims to find the most relevant recommendation results. The higher the NDCG and MRR values ​​are, the better. Both evaluation matrices have a value range from 0 to 1. A value of 1 on the NDCG indicates that the ranking results are perfect, while a value of 0 indicates no correlation between the ranking results and item relevance. Generally, the NDCG value is considered very good if the value is greater than or equal to 0.9 and harmful if the value is below 0.5. Meanwhile, in the MRR, a value of 1 indicates that the first item in the ranking results is always relevant, while 0 indicates that no suitable item is placed first in the ranking results. In general, the MRR value is considered good when the value is greater than or equal to 0.5 and is considered harmful if the value is below 0.2.

The simplest way to measure recommendation results is with Cumulative Gain (CG), which can be calculated by equation 1.

(1)

(2)

Where the value of k is the maximum ranking, i is the position of the item in the ranking, is the relevance value of the item in position i, and is the order of the data item based on its relevance value in equation (1), CG calculates the total number of item values ​​regardless of the position or order of the items in the results. So, CG is developed into Discounted Cumulative Gain (DCG). The DCG equation can be seen in equation (2). The DCG aims to provide a proportion of values ​​according to the position of the items in the recommendation results. The higher the DCG value, the more items are in a higher position. However, DCG still has drawbacks, namely, it is difficult to evaluate the results of different recommendations. Therefore, equation (4) Normalized Discounted Cumulative Gain (NDCG) exists. In this method, the Ideal DCG (IDCG) is used as a constant, which will be used as a divider to evaluate the results of one recommendation with another. IDCG is a virtual list; positions result in descending order according to their DCG values.

(3)

The MRR calculation can be seen in equation (5).

(5)

Where i is the ith query, is the number of queries performed, and the ranking position of the first relevant item from the ith query.

**RESULTS AND DISCUSSION**

**Result**

The dataset that has been collected is preprocessed. Some data has the possibility that there are missing values ​​which can cause the data to become very large but actually the data is incomplete, so it needs to be checked. Missing value checking can be done in several ways. In this study, we used functions from the pandas library, namely isna() and sum() to find out how many missing values ​​were found. After knowing each number of missing values ​​in the category code column every month the next step is to retrieve data that does not contain a missing value in the category\_code column. After each month's data is filtered based on the category\_code column without a missing value, we can find out the results of the amount of each data each month as shown in Table 2.

(4)

Table 2. Amount of Data for Each Category Code Without Missing Values

|  |  |
| --- | --- |
| **Months** | **Count** |
| October | 67.477 |
| November | 75.748 |
| December | 58.465 |
| January | 74.719 |
| February | 77.185 |

After the monthly data without missing values ​​in the category\_code column is obtained, the next step is to combine all the separated data into one file. The next stage is reducing the amount of data based on the value in a particular column. In this study, the reduction was done by removing data based on the event\_type column with the value "remove\_from\_cart".

The next stage after the reduction process is carried out is to change the data in the form of atomic files. This process is necessary because the model creation process will use pipelines from the recbole library from python. The first stage is to create a file related to the interaction between the user and the product into a file format with the .inter extension. The result of .inter file is shown in Figure 2. The second stage is to make files related to the product into a file format with the .item extension. The result of .inter file is shown in Figure 3.

A screenshot of a computer

Description automatically generated

Figure 2. Result of .inter file

**A screenshot of a computer

Description automatically generated**

Figure 3. Result of .item file

Then the atomic file data is converted into interaction data which will be included in the algorithm for training. Interaction is data formed from user interaction with the product. The results of the loading dataset produce a total of 223,791 interactions grouped by user and sorted by timestamp. The number of interactions generated at this stage is different from the previous amount of data in the atomic files which totaled 315,136, because they have not been grouped by user. The result of the interaction dataset is shown in Figure 4.

A screenshot of a computer

Description automatically generated

Figure 4. The result of the interaction dataset

Table 3 shows the amount of each data split into training, validation, and testing data. The amount of validation data and testing data does not have the same number, even though the percentage of validation data and testing data is the same, which is 0.15%. This happens because the data is divided based on the timestamp so that data in one session is still in the same series so that it cannot be split so as not to lose context.

Table 3. Amount of Each Splitting Data

|  |  |
| --- | --- |
| **Data Split** | **Amount of Interaction** |
| Training data | 154,636 |
| Validation data | 30,367 |
| Testing data | 38,788 |

The training and evaluation process is carried out based on Top-K, and Evaluation Model, where 𝐾 is the number of top candidates. In the training process, the data used are training data and validation data. While in the data evaluation process used is testing data. Figure 5 and Figure 6 evaluate model-based MRR and NDGC. Based on Figure 5, it's evident that the MRR values show a steady rise with successive iterations, stabilizing around the 15th iteration. Similarly, Figure 6 illustrates that the NDGC values also exhibit an incremental growth with each iteration, stabilizing approximately by the 20th iteration.

A graph with a line

Description automatically generated

Figure 5. MRR Evaluation model based TOP-K

A graph with a line

Description automatically generated

Figure 6. NDGC Evaluation model based TOP-K

**Discussion**

Furthermore, the research obtained better results than previous studies. Results comparing this study with previous studies are shown in Table 4.

Table 4. Accuracy Comparison with Previous Research

|  |  |  |  |
| --- | --- | --- | --- |
| **Researcher** | **Research Methods** | **Object** | **Accuracy** |
| [33] | Deep Neural Network (DNN) with Optimizer Adam | Dataset Kechinov  ‘eCommerce Events History in Cosmetics Shop’ | Recall 0.31 |
| [34] | Hierarchical Sequence Probability (HSP) and Recurrent Iem Co-occurrence (RIC) with Graph Modelling | Dataset Leather, Diginetica, YooChoose 1/64, and electronics | Average MRR 0.26 |
| [35] | BERT4REC for sequential recommendation | Dataset Amazon Beauty, Steam, Movielens-1m, and Movielens-2m | Average MRR 0.31 and Avertage NDCG 0.32 |
| Proposed Method | NNCF with Data Reduction | Dataset Kechinov  ‘eCommerce Events History in Cosmetics Shop’ | MRR 0.56, NDCG 0.60 |

The results of the evaluation metrics in this study using MRR and NDCG were 0.56 and 0.60 respectively, which were higher than those of [35] who obtained an average MRR of 0.31 and NDCG 0.32 using the same dataset. The method used is NNCF with Data Reduction. In his research, he received suggestions to use other features besides product\_id and user\_id to get better results. Therefore, this research utilized the product's features and performed data reduction in the preprocessing section.

This study also obtained better results than the research conducted by Delianidi et al [34], who used the HSP and RIC models obtained an average MRR of 0.26. In his study, the model was trained by predicting the next item based on the sequence, while in this study, the model was introduced by masking the NNCF algorithm in the sequence. The dataset used in the research by Delianidi et al uses a dataset with the same object, namely e-commerce [34].

The results of this study also obtained better results than those of research conducted by Szabo [33] with the same dataset. Although the research focuses on measuring the value of MRR and NDCG, this study also calculates Recall. In Szabó and Genge's research [33], they obtained a Recall value of 0.31, while in this study, they obtained a Recall of 0.67. In this study, the same preprocessing was applied to the research conducted by Szabo and Genge [33], namely data reduction and the conversion method of user interaction event types with products into a single UX value. In contrast, the algorithm used in this study differs from his research, namely NNCF.

**CONCLUSION**

This study focuses on implementing NNCF in e-commerce to overcome the cold start problem so that we can continue to provide product recommendations to users without referring to user preferences. The results of the metric evaluation using NDCG and MRR in this study were better than those of other studies, namely 0.60 and 0.54. In future research, this can be done by using more data and trying different distribution percentages to improve accuracy.

**ACKNOWLEDGMENT**

The authors thank The LPPM Universitas Negeri Semarang through the DIPA program Fund 2023 No 023.17.2.690645/2023.10.

**REFERENCES**

[1] M. Nasir and C. I. Ezeife, “A Survey and Taxonomy of Sequential Recommender Systems for E-commerce Product Recommendation,” *SN Comput Sci*, vol. 4, no. 6, p. 708, Sep. 2023, doi: 10.1007/s42979-023-02166-5.

[2] J. Lee, H. Yoon, S. Kim, C. Lee, J. Lee, and S. Yoo, “Deep learning-based skin care product recommendation: A focus on cosmetic ingredient analysis and facial skin conditions,” *J Cosmet Dermatol*, 2024, doi: 10.1111/jocd.16218.

[3] R. Iwabuchi *et al.*, “Proposal of recommender system based on user evaluation and cosmetic ingredients,” in *2017 International Conference on Advanced Informatics, Concepts, Theory, and Applications (ICAICTA)*, IEEE, Aug. 2017, pp. 1–6. doi: 10.1109/ICAICTA.2017.8090967.

[4] R. S, H. S, K. Jayasakthi, S. Devi. A, K. Latha, and N. Gopinath, “Cosmetic Product Selection Using Machine Learning,” in *2022 International Conference on Communication, Computing and Internet of Things (IC3IoT)*, IEEE, Mar. 2022, pp. 1–6. doi: 10.1109/IC3IOT53935.2022.9767972.

[5] Y. Nakajima, H. Honma, H. Aoshima, T. Akiba, and S. Masuyama, “Recommender System Based on User Evaluations and Cosmetic Ingredients,” in *2019 4th International Conference on Information Technology (InCIT)*, IEEE, Oct. 2019, pp. 22–27. doi: 10.1109/INCIT.2019.8912051.

[6] S. A. Abu-Shanab, S. Alzu’Bi, and A. Zraiqat, “A Novel Virtual Cosmetics Recommender System Based on Pre-Trained Computer Vision Models,” in *2023 International Conference on Information Technology: Cybersecurity Challenges for Sustainable Cities, ICIT 2023 - Proceeding*, 2023, pp. 765–770. doi: 10.1109/ICIT58056.2023.10225835.

[7] M. Ueda, S. Yabe, D. Li, and S. Nakajima, “A Cosmetic Differences Visualization System for Beauty Recommendation using the Scores of Various Evaluation Items,” in *CEUR Workshop Proceedings*, 2022, pp. 137–150.

[8] S. Yabe, M. Ueda, and S. Nakajima, “A Comparative Method Based on the Visualization of Cosmetic Items Using Their Various Aspects,” in *Digest of Technical Papers - IEEE International Conference on Consumer Electronics*, 2021. doi: 10.1109/ICCE50685.2021.9427774.

[9] A. Singha Roy, E. D’Amico, A. Lawlor, and N. Hurley, “Addressing Fast Changing Fashion Trends in Multi-Stage Recommender Systems,” *The International FLAIRS Conference Proceedings*, vol. 36, May 2023, doi: 10.32473/flairs.36.133307.

[10] R. S, H. S, K. Jayasakthi, S. Devi. A, K. Latha, and N. Gopinath, “Cosmetic Product Selection Using Machine Learning,” in *2022 International Conference on Communication, Computing and Internet of Things (IC3IoT)*, IEEE, Mar. 2022, pp. 1–6. doi: 10.1109/IC3IOT53935.2022.9767972.

[11] D. Pradana and E. Sugiharti, “Implementation Data Mining with Naive Bayes Classifier Method and Laplace Smoothing to Predict Students Learning Results,” *Recursive Journal of Informatics*, vol. 1, no. 1, pp. 1–8, 2023, doi: 10.15294/rji.v1i1.63964.

[12] B. Galhotra, “Evolution of E-commerce in India: A Review and Its Future Scope,” in *Proceedings of the International Conference on Machine Learning, Big Data, Cloud and Parallel Computing: Trends, Prespectives and Prospects, COMITCon 2019*, IEEE, 2019, pp. 226–231. doi: 10.1109/COMITCon.2019.8862252.

[13] S. Dinesh, R. Nalini, N. Shobhana, R. Amudha, R. Alamelu, and V. Renaarajan, “Impact of Digital Touchpoints towards Consumer Decision Journey with reference to Delta Districts of Tamilnadu,” in *Proceedings of IEEE International Conference on Disruptive Technologies for Multi-Disciplinary Research and Applications, CENTCON 2021*, Institute of Electrical and Electronics Engineers Inc., 2021, pp. 188–193. doi: 10.1109/CENTCON52345.2021.9687926.

[14] Y. Lyu, “Recommender Systems in e-Commerce,” in *Proceedings - 2021 International Conference on Intelligent Computing, Automation and Applications, ICAA 2021*, Institute of Electrical and Electronics Engineers Inc., 2021, pp. 209–212. doi: 10.1109/ICAA53760.2021.00044.

[15] P. Jariha and S. K. Jain, “A state-of-the-art Recommender Systems: An overview on Concepts, Methodology and Challenges,” in *Proceedings of the International Conference on Inventive Communication and Computational Technologies, ICICCT 2018*, Institute of Electrical and Electronics Engineers Inc., Sep. 2018, pp. 1769–1774. doi: 10.1109/ICICCT.2018.8473275.

[16] M. Sadeghi, S. A. Asghari, and M. M. Pedram, “An Improved Method Multi-View Group Recommender System (IMVGRS),” in *2020 8th Iranian Joint Congress on Fuzzy and intelligent Systems (CFIS)*, IEEE, Sep. 2020, pp. 127–132. doi: 10.1109/CFIS49607.2020.9238688.

[17] H. Liu, “Implementation and Effectiveness Evaluation of Four Common Algorithms of Recommendation Systems - User Collaboration Filter, Item-based Collaborative Filtering, Matrix Factorization and Neural Collaborative Filtering,” in *2022 International Conference on Cloud Computing, Big Data Applications and Software Engineering (CBASE)*, IEEE, Sep. 2022, pp. 224–227. doi: 10.1109/CBASE57816.2022.00049.

[18] Z. Wang and D. Wang, “Research on recommendation algorithm based on neural network fusing user behavior sequence,” in *Third International Conference on Artificial Intelligence and Computer Engineering (ICAICE 2022)*, X. Li, Ed., SPIE, Apr. 2023, p. 142. doi: 10.1117/12.2671253.

[19] A. C. Hansel and A. Wibowo, “Using Movie Genres in Neural Network Based Collaborative Filtering Movie Recommendation System to Reduce Cold Start Problem,” *International Journal of Emerging Technology and Advanced Engineering*, vol. 12, no. 3, pp. 63–73, Mar. 2022, doi: 10.46338/ijetae0322\_08.

[20] S. M. Choi, K. Jang, T. D. Lee, A. Khreishah, and W. Noh, “Alleviating Item-Side Cold-Start Problems in Recommender Systems Using Weak Supervision,” *IEEE Access*, vol. 8, pp. 167747–167756, 2020, doi: 10.1109/ACCESS.2020.3019464.

[21] K. Binabdullah and N. Tongtep, “Comparative Study on Natural Language Processing for Tourism Suggestion System,” in *2021 36th International Technical Conference on Circuits/Systems, Computers and Communications (ITC-CSCC)*, IEEE, 2021, pp. 1–4. doi: 10.1109/ITC-CSCC52171.2021.9501422.

[22] M. H. Kwon, D. S. Chang, and Y. S. Choi, “Improving RNN Based Recommendation by Embedding-Weight Tying,” in *Proceedings - 2018 IEEE International Conference on Systems, Man, and Cybernetics, SMC 2018*, Institute of Electrical and Electronics Engineers Inc., Jan. 2019, pp. 4017–4022. doi: 10.1109/SMC.2018.00681.

[23] S. K. Rachamadugu, J. R. Dwaram, and K. R. Patike, “Recommender System based on Deep Neural Network and Long Short Term Memory,” in *Proceedings - 2021 International Conference on Computing, Networking, Telecommunications and Engineering Sciences Applications, CoNTESA 2021*, Institute of Electrical and Electronics Engineers Inc., 2021, pp. 50–55. doi: 10.1109/CoNTESA52813.2021.9657131.

[24] S. Dongre and J. Agrawal, *Comprehensive Assessment of Big Data in Recommendation Systems*, vol. 528. 2023. doi: 10.1007/978-981-19-5845-8\_11.

[25] V. Stergiopoulos, T. Tsianaka, and E. Tousidou, “AMiner Citation-Data Preprocessing for Recommender Systems on Scientific Publications,” in *ACM International Conference Proceeding Series*, 2021, pp. 23–27. doi: 10.1145/3503823.3503828.

[26] P. Zhang, F. Xiong, J. Gao, and J. Wang, “Data quality in big data processing: Issues, solutions and open problems,” in *2017 IEEE SmartWorld Ubiquitous Intelligence and Computing, Advanced and Trusted Computed, Scalable Computing and Communications, Cloud and Big Data Computing, Internet of People and Smart City Innovation, SmartWorld/SCALCOM/UIC/ATC/CBDCom/IOP/SCI 2017 -* , 2018, pp. 1–7. doi: 10.1109/UIC-ATC.2017.8397554.

[27] B. Kotiyal and H. Pathak, *Big Data Preprocessing Phase in Engendering Quality Data*, vol. 768. 2022. doi: 10.1007/978-981-16-2354-7\_7.

[28] S. Dalal and V. Dahiya, “Big data preprocessing: Needs and methods,” *International Journal of Engineering Trends and Technology*, vol. 68, no. 10, pp. 100–104, 2020, doi: 10.14445/22315381/IJETT-V68I10P217.

[29] M. J. Reena, “Preprocessing Big Data using Partitioning Method for Efficient Analysis,” in *Proceedings of IEEE InC4 2023 - 2023 IEEE International Conference on Contemporary Computing and Communications*, 2023. doi: 10.1109/InC457730.2023.10262924.

[30] A. Solikhatun and E. Sugiharti, “Application of the Naïve Bayes Classifier Algorithm using N- Gram and Information Gain to Improve the Accuracy of Restaurant Review Sentiment Analysis,” *Journal of Advances in Information Systems and Technology*, vol. 2, no. 2, pp. 1–12, 2020.

[31] X. He, L. Liao, H. Zhang, L. Nie, X. Hu, and T.-S. Chua, “Neural Collaborative Filtering,” in *Proceedings of the 26th International Conference on World Wide Web*, Republic and Canton of Geneva, Switzerland: International World Wide Web Conferences Steering Committee, Apr. 2017, pp. 173–182. doi: 10.1145/3038912.3052569.

[32] A. Da’u and N. Salim, “Recommendation System Based on Deep Learning Methods: a Systematic Review and New Directions,” *Artif Intell Rev*, vol. 53, no. 4, pp. 2709–2748, 2020, doi: 10.1007/s10462-019-09744-1.

[33] P. Szabo and B. Genge, “Efficient Conversion Prediction in E-Commerce Applications with Unsupervised Learning,” in *2020 28th International Conference on Software, Telecommunications and Computer Networks, SoftCOM 2020*, IEEE, Sep. 2020. doi: 10.23919/SoftCOM50211.2020.9238344.

[34] M. Delianidi, K. Diamantaras, D. Tektonidis, and M. Salampasis, “Session-Based Recommendations for e-Commerce with Graph-Based Data Modeling,” *Applied Sciences (Switzerland)*, vol. 13, no. 1, pp. 1–16, 2023, doi: 10.3390/app13010394.

[35] F. Sun *et al.*, “BERT4Rec,” in *International Conference on Information and Knowledge Management, Proceedings*, New York, NY, USA: ACM, Nov. 2019, pp. 1441–1450. doi: 10.1145/3357384.3357895.