Tourism Destination Recommendation System Using Collaborative Filtering and Modified Neural Network

Kurniawan Eka Permana[1](✉), Sri Herawati[2], and Wahyudi Setiawan[2]

[1] Department of Informatics, University of Trunojoyo Madura, Bangkalan, Indonesia
kurniawan@trunojoyo.ac.id

[2] Department of Information System, University of Trunojoyo Madura, Bangkalan, Indonesia
{sriherawati,wsetiawan}@trunojoyo.ac.id

Abstract. Tourism is one of the driving sectors of the national economy. Nowadays, the normal opening of tourist destinations after COVID-19 pandemic, tourist visits are currently increasing rapidly. Indonesia has a unique culture, nature, language, and cuisine. This is certainly a potential that can attract tourists to visit this archipelago country. To increase the attractiveness of tourism, one of the things that can be done is to create a recommendation system. This system needs to be built to provide users with personalized recommendations based on the input that the user has given. This research uses primary data. The data is a tourist place in the island of Madura. The amount of data consists of 160 tourist places. While the rating is done by 120 users. The system built consists of steps: preprocessing, creating, and training the model based on the data split, calculating the error, and showing destination recommendations to a user. Preprocessing converts “user” and “place” into integer indexes. Creating the model is done by embedding between “user” and “place”. The rating is normalized to a number between zero and one using a sigmoid. Furthermore, training data is carried out using Modified Neural Network. The test results show validation_RMSE for each regency (Bangkalan, Sampang, Pamekasan, and Sumenep) is 0.3663, 0.3523, 0.3581, 0.3905. The recommendation system produces seven destinations as recommendations for places that have not been visited by the user.

Keywords: Recommendation System · Tourism Destination · Madura · Modified Neural Network

1 Introduction

Tourism is one of the important sectors of state income. The tourism industry in Indonesia is wide potential. The country has more than 17,000 islands, and a diversity of cultures, languages, and nature. Currently, Bali is still the main tourist destination for foreign and local tourists. Data in April 2022 shows that foreign tourist arrivals have increased significantly by 499%, from 18,540 to 111,057 people. Australia and Singapore became the countries with the largest number of foreign tourists visiting Indonesia [1].

I. H. Agustin (Ed.): ICONNSMAL 2022, AISR 177, pp. 60–70, 2023.
https://doi.org/10.2991/978-94-6463-174-6_7

This massive potential for tourist visits is based on the opening of international airports in Indonesia. The number of arrivals by air was recorded at 87.7%, compared to sea at 11.19% and land at −1.11% [1]. Of course, the tourist must get the best services including facilities in all tourist destinations.

One of the tourist destinations that is quite interesting to explore is the potential of Madura Island. The uniqueness of culture, variety of tourism, cuisine, and history are things that must be explored. Each tour package certainly has a pre-determined itinerary and destination. Not a few tourists do not agree with the specified package. The mismatch of schedule and choice of place is the main factor in the unattractiveness of the tour packages offered. In this way, many travelers choose to travel independently by making travel plans and destinations according to their wishes. However, the number of available places makes it difficult for tourists to determine a place that matches their expectations. Another problem is when a tourist wants to visit a place but does not have time to visit another place because of time constraints. Therefore, tourists need to recommend tourist attractions and explain travel routes to schedule travel information effectively. The recommended system is an application model based on observations of the situation and user desires. The system needs a good recommendation model so that the recommendations fit the user's needs and help him make the right decision when deciding which item to choose.

Research on tourist place recommendations has been carried out, including a tourist recommendation system consisting of several stages, namely input, preprocessing, process and output. Input consists of meteorological data (weather, climate), tourist attractions, user reviews, and ratings. The data is processed in the form of text with preprocessing word embedding Term Frequency-Inverse Document Frequency (TF-IDF). The feature extraction process and recommendations use Convolutional Neural Network (CNN) and Long Short-Term Memory (LSTM). Output in the form of an evaluations index [2].

Another research is on a context-aware-based location recommendation system using LSTM. The system consists of two levels. The first level uses travel history and makes predictions for the location of the next tourist place. At the first level, contextual learning consists of two parts:

- location, user, and environments such as weather, climate, and risk
- popularity location.

Furthermore, the second level is a recommendation using LSTM. The results obtained are in the form of location probabilities and the results of top-rank tourist places. Accuracy results show 97.2% using LSTM. In addition, the system also uses a Gated Recurrent Unit (GRU) and a Bidirectional LSTM. Accuracy results are 96.4% and 94.2%, respectively [3].

Another research is making a recommendation system using Question Answer and Knowledge Graph techniques. The case study was conducted at a tourist place in Vietnam. This technique is part of Natural Language Processing (NLP). The stages start from question input, one hot encoding, Seq2seqmodel. Furthermore, the Knowledge Graph Database is processed into a Graph Query. The system is processed using deep learning LSTM. The output produces an NLP answer list for processing through Graph Query and simple NLP answers without using Graph Query. The output results show an F1 measure 83%, while the precision is 87% [4].

In this study, a tourist place recommendation system application was built. The process for recommending tourist places is based on several attributes or parameters including tourist destination, descriptions, categories, cities, ratings, user, user location, and user age. Recommendation system using Collaborative filtering and Modified Neural Network using Recommender_net [5, 6].

2 Research Method

A. *Recommendation System*

A recommendation system is an application that provides and recommends items when making decisions that the user likes [7]–[9]. Implementing recommendations in your system typically anticipates items such as recommendations for movies, music, books, news, and travel. This system works by collecting data directly or indirectly from the user.

The steps taken in the recommendation system include data acquisition from the user which can be done by:

- Ask the user to rate the item.
- Ask users to assign their favorite items to at least one favorite item.
- Give the user many choices and ask them to choose the one that suits them best.
- Ask users to list the items they like or dislike the most.

Data collection is not directly related to the user but is done by observing the items seen by the user on the system. From the data collected, it is processed using a certain algorithm. The result is then returned as a recommended item with parameters from the user.

The recommendation system is also an alternative search engine for the items the user is looking for. When developing a recommender system, there are several ways to solve the problem, such as user-based collaborative filtering, content-based filtering, and hybrids. However, some researchers add knowledge-based behavior methods [10, 11].

A. **Collaborative filtering**

Collaborative filtering uses data from two or more sources. In this study, the data comes from user reviews and user tourism history, to find users with similar tourist interests. It can be assumed that users who visit the same tourist place have the same interests [12, 13].

To get a recommendation, it is necessary to create an embedding that represents the relationship between the user and the tourist place. The result of the first dimension is the matrix, where the user is a row, and a column is a tourist place. Figure 1 is a system description of the embedding process. The results of collaborative filtering will be whether the user in the last line likes or not the Pantai rongkang (rongkang beach) as tourist destination.

We can tell if user 4 doesn't like Rongkang beach. This can be seen in the history of tourist places (Syaichona cholil and Sinjay). If you see users who like Syaichona cholil

	Bukit Jaddih	BanPlaz	Pantai Rongkang	Syaichona cholil	Sinjay
user 1	√		√	√	
user 2		√			√
user 3	√	√	√		
user 4			?	√	√

Fig. 1. Collaborative filtering works

also like rongkang beach. However, some users like Sinjay don't choose Rongkang beach. So, it is not certain if a user who chooses Syaichona cholil and Sinjay also likes Rongkang. No correlation shows this, so the system does not recommend Rongkang as a tourist place. The cosine similarity calculation used to measure the similarity of tourist destinations is shown in Eq. 1.

$$Similarity(i,j) = \frac{\sum_u^U r_{(u,i)} r_{(u,j)}}{\sqrt{\sum_u^U r_{(u,i)}{}^2} \sqrt{\sum_u^U r_{(u,j)}{}^2}} \tag{1}$$

i = main item for which to find other similar items

j = item that is being compared with main item 'i' to find it that are similar.

(u, i) = user 'u' rating for item 'i'.

(u, j) = user 'u' rating for item 'j'.

U = repeat multiplication for all user U.

The Recommendation System in this study was built consisting of seven steps as shown in Fig. 2.

1) *Load Data*

The data is taken from a survey of 300 respondents. The respondents give a rating to 160 tourism destinations in Madura. Rating given with a linkert scale from one to five. The dataset consists of two files with the attributes shown in Table 1.

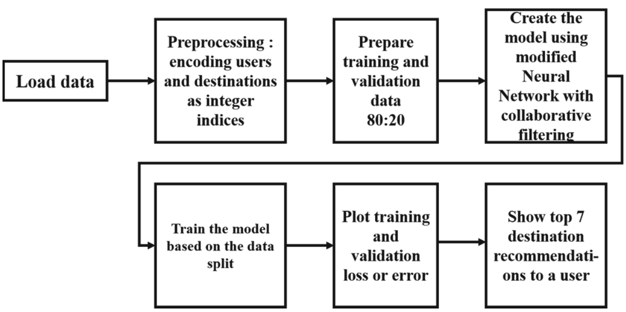


Fig. 2. System Recommendation proposed.

Table 1. Input Data Attributes

File	Attributes
Destinations	Place_id, Place Rating
Users	User_Id, Location, Place Rating

Table 2. Error Result

Region	RMSE	Validation RMSE
Bangkalan	0.3180	0.3663
Sampang	0.3160	0.3523
Pamekasan	0.3140	0.3581
Sumenep	0.3174	0.3905

2) *Preprocessing*

In the dataset, there are two files: "destinations" and "users". The data in the two files is converted into an integer index. So, each "users" and "destinations" has an integer value. This is used to facilitate the collaborative filtering process between users and destinations.

3) *Split Training and Testing Data*

Data needs to be split into the training and testing process. Comparisons are made to find out whether the system made is by the role of the recommendations that are the goal. Split data for training and testing using a percentage of 80:20.

4) *Create the Model Using Modified Neural Network with Collaborative Filtering*

Collaborative filtering process using user and place data. The collaborative process occurs in the model built using a Modified neural network. This model is a library derived from Keras Tensorflow, a library on python [14, 15]. The Modified neural network used has the architecture shown in Fig. 3. The embedding consists of four layers used for the user layer, the user_bias layer, the place layer, and the place_bias layer.

In Fig. 3, the neural network architecture has 19,584 parameters consisting of weight and bias. This study discusses collaborative filtering using the Madura-tourism dataset to recommend tourist places to users. The Madura-tourism rating dataset contains the rating given by the user for the tourist place. The purpose of the system is to be able to predict the rating for tourist places that have not been visited by the user. Tourist places with the highest rating predictions become tourist place recommendations for users.

The stages of the model built are:

- Map "users" to user vector via an embedding matrix
- Map "destinations" to destination vector via an embedding matrix

```
Model: "recommender_net"
_________________________________________________________________
 Layer (type)                Output Shape              Param #
=================================================================
 embedding (Embedding)       multiple                  15000

 embedding_1 (Embedding)     multiple                  300

 embedding_2 (Embedding)     multiple                  4200

 embedding_3 (Embedding)     multiple                  84

=================================================================
Total params: 19,584
Trainable params: 19,584
Non-trainable params: 0
_________________________________________________________________
```

Fig. 3. Architecture of Modified neural network.

- Calculate the multiplication (dot product) between "users-vector and the destinations-vector". The dot product result is the appropriate score between the user and the destination (prediction rating).
- Train embedding via gradient descent using all pairs between user and destination.

5) *Train the Model Based on Split Data*

The next step is to carry out the training process using a modified neural network. The training and validation process uses several parameters, initialization of weights using He_normal [16, 17], sigmoid activation, learning rate 0.0004, Adam optimizer, and epoch 100.

6) *Show the Training and Validation Error*

The training and validation data produces performance in the form of root mean square error, and validation root means square error. A measure of Root Mean Square Error (RMSE) in Eq. (2) [18, 19].

$$RMSE = \sqrt{\frac{1}{N}\sum_{i=1}^{N}\left(y_i - \hat{y}\right)^2} \tag{2}$$

y_i = predicted value of y; $\hat{y}$ = actual value of y; N = amount of data

7) *Recommendations for Tourist Destinations*

In the last section, seven recommendations for the best tourist attractions from each region are presented. The regions of Bangkalan, Sampang, Pamekasan, and Sumenep. Each has seven recommended tourist destinations.

3 Result and Discussion

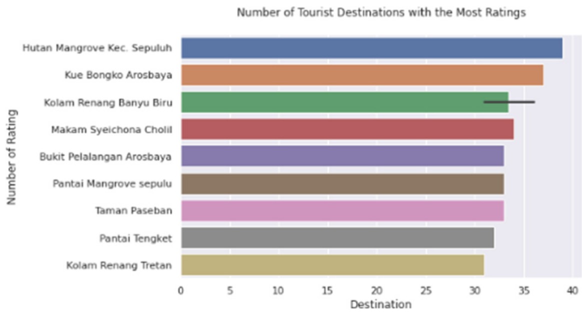
A. *The Testing Environment*

The Recommendation system was carried out using Google Collab with specifications hardware of Core i7–7700 processor, GTX 1060 6 Gb D5 amp, 60 Gb Solid State Drive. It also needs a library of phyton such as TensorFlow, Keras, imageDataGenerator, os, NumPy, pandas, seaborn, and matplotlib.

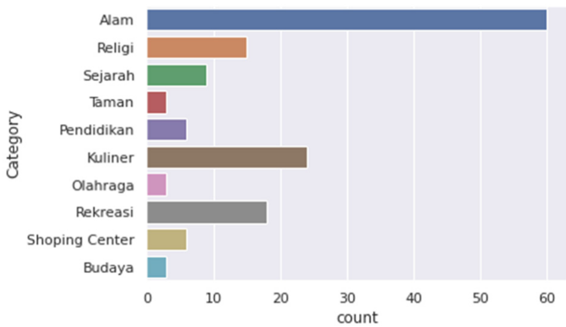
B. *Exploratory Data Analysis (EDA)*

EDA is a data exploration that is used for the initial testing process which aims to identify patterns, find outliers (if any), and validate assumptions [20]. Data exploration consists of tourist places with the highest rating by users in each region. Figures 4 (a) and (b) are the results of tourism data exploration.

Figure 4 (a) shows a tourist place that has been given a rating by the user. The most ratings tourist destination given by users is Hutan Mangrove Sepulu. Furthermore, Fig. 4 (b) shows the tourism categories in Bangkalan in the form of culture, park/theme park, nature conservation, nautical, shopping center, and worship place. In addition, data



(a)



(b)

Fig. 4. (a) Rating for tourist places in Bangkalan; (b) Category of tourism in Bangkalan

exploration shows the number of users, and cities of origin that give ratings to tourist attractions.

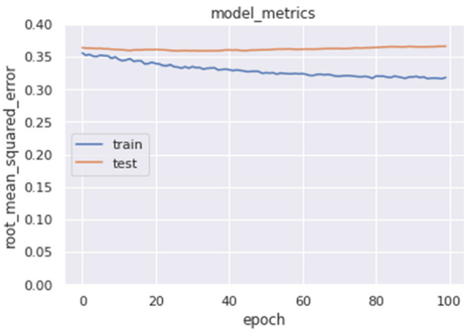
C. *Recommendation Result*

The test results consist of two: the output of tourism recommendations from each regency and the performance of RMSE. Figure 5 shows the recommendations for the top seven tourist destinations in the regency of Bangkalan. The performance of recommendation system with collaborative filtering uses RMSE because the system performs a regression task [21, 22]. RMSE shows the model_metric generated when making a tourism recommendation system in each region. By using the research steps that have been described in the research method section. Figure 6 shows the RMSE during training and validation in four regions. The complete results in RMSE can be seen in Table 2. The test results show that RMSE is above 0.25.

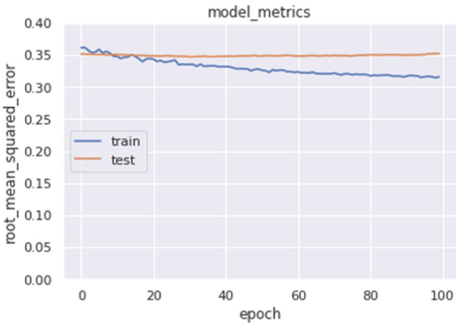
The recommendation system that has been made is a recommendation system that involves two data: user and place. The system architecture is made simple by only involving four embedding layers. The system needs to improve the RMSE. First by modifying the layer and component parameters (learning rate, epoch, optimizer) that are contained in the model. In addition, it is also necessary to involve other recommended attributes such as user age, tourism category, and ticket.

```
------------------------------------------------------------
Tempat dengan rating wisata paling tinggi dari user
------------------------------------------------------------
Bukit Kapur Arosbaya : Alam
Kolam Renang Banyu Biru : Rekreasi
Air Terjun Kokop : Alam
Kapal jodoh : Rekreasi

------------------------------------------------------------
Top 7 place recommendation
------------------------------------------------------------
1 . Pantai Rongkang
     Alam , Harga Tiket Masuk  20000 , Rating Wisata  4.6

2 . Pantai Siring Kemuning
     Alam , Harga Tiket Masuk  20000 , Rating Wisata  4.4

3 . Taman Rekreasi Kota
     Taman , Harga Tiket Masuk  50000 , Rating Wisata  4.5

4 . Bukit Pelalangan Arosbaya
     Alam , Harga Tiket Masuk  15000 , Rating Wisata  4.4

5 . Pantai Telangoh
     Alam , Harga Tiket Masuk  10000 , Rating Wisata  4.7

6 . Pemandian Sumber Pocong
     Alam , Harga Tiket Masuk  5000 , Rating Wisata  4.0

7 . Masjid Agung Sultan Abdul Kadirun
     Religi , Harga Tiket Masuk  0 , Rating Wisata  4.7

=============================================
```

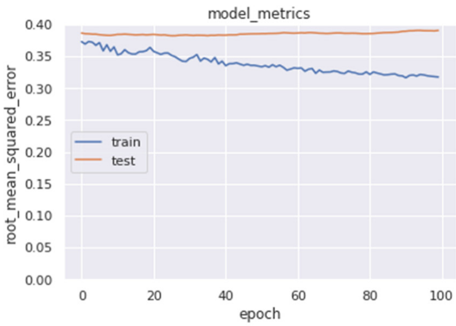
Fig. 5. Top seven tourist destinations in Bangkalan



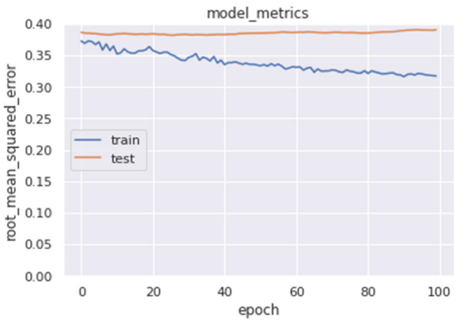
(a)



(b)



(c)



(d)

Fig. 6. RMSE during training and validation (a) Bangkalan (b) Sampang (c) Pamekasan (d) Sumenep

4 Conclusion

In this article, a recommendation system using collaborative filtering with Recommender_Net has been carried out. Collaborative filtering is done between "users" and "destinations" to produce the best recommendations based on ratings. Then, the training and testing process is carried out using Recommender_Net which consists of four embedding layers. The parameters used in Recommender_Net are sigmoid activation, learning rate 0.0004, Adam optimizer, and epoch 100. The percentage of training and testing data is 80:20. The test results show that the RMSE and RMSE validation are still more than 0.30. This shows that the system still needs improvement. For further experiments, it is necessary to involve all attribute elements from tourism data such as user age, tourism category, and ticket. In addition, it is required an alternative recommendation model with another architecture and hyperparameter tuning to find out a better model.

Acknowledgment. This research is one of the outputs of Penelitian Grup Riset, Penelitian Mandiri, University of Trunojoyo Madura No 229/UN46.4.1/PT01.03/2022.

References

1. Kemenparekraf, "Perkembangan Wisatawan Mancanegara," Kemenparekraf RI, 2022.
2. H. woo An and N. Moon, "Design of recommendation system for tourist spot using sentiment analysis based on CNN-LSTM," J. Ambient Intell. Humaniz. Comput., vol. 13, no. 3, pp. 1653–1663, 2022, doi: https://doi.org/10.1007/s12652-019-01521-w.
3. W. Shafqat and Y. C. Byun, "A context-aware location recommendation system for tourists using hierarchical LSTM model," Sustain., vol. 12, no. 10, 2020, doi: https://doi.org/10.3390/su12104107.
4. P. Do, T. H. V. Phan, and B. B. Gupta, "Developing a Vietnamese Tourism Question Answering System Using Knowledge Graph and Deep Learning," ACM Trans. Asian Low-Resource Lang. Inf. Process., vol. 20, no. 5, pp. 1–18, 2021, doi: https://doi.org/10.1145/3453651.
5. D. S. Pradana and G. P. Hartawan, "Perbandingan Algoritma Content-Based Filtering dan Collaborative Filtering dalam Rekomendasi Kegiatan Ekstrakurikuler Siswa," Progresif, vol. 18, no. 2, pp. 151–160, 2014.
6. O. G. Yalçın, Applied Neural Networks with TensorFlow 2. 2021.
7. P. Yochum, L. Chang, T. Gu, and M. Zhu, "Linked Open Data in Location-Based Recommendation System on Tourism Domain: a Survey," IEEE Access, vol. 8, pp. 16409–16439, 2020, doi: https://doi.org/10.1109/aCCESS.2020.2967120.
8. R. A. Hamid et al., "How smart is e-tourism? A systematic review of smart tourism recommendation system applying data management," Comput. Sci. Rev., vol. 39, p. 100337, 2021, doi: https://doi.org/10.1016/j.cosrev.2020.100337.
9. C.-S. Jeong, K.-H. Ryu, J.-Y. Lee, and K.-D. Jung, "Deep Learning-based Tourism Recommendation System using Social Network Analysis," Int. J. Internet, Broadcast. Commun., vol. 12, no. 2, pp. 113–119, 2020, [Online]. Available: http://dx.doi.org/https://doi.org/10.7236/IJIBC.2020.12.2.113.
10. A. A. Fakhri, Z. K. A. Baizal, and E. B. Setiawan, "Restaurant Recommender System Using User-Based Collaborative Filtering Approach: A Case Study at Bandung Raya Region," J. Phys. Conf. Ser., vol. 1192, no. 1, 2019, doi: https://doi.org/10.1088/1742-6596/1192/1/012023.

11. N. Nassar, A. Jafar, and Y. Rahhal, "A novel deep multi-criteria collaborative filtering model for recommendation system," Knowledge-Based Syst., vol. 187, p. 104811, 2020, doi: https://doi.org/10.1016/j.knosys.2019.06.019.
12. V. Subramaniyaswamy, R. Logesh, M. Chandrashekhar, A. Challa, and V. Vijayakumar, "A personalised movie recommendation system based on collaborative filtering," Int. J. High Perform. Comput. Netw., vol. 10, no. 1–2, pp. 54–63, 2017, doi: https://doi.org/10.1504/IJHPCN.2017.083199.
13. M. D. Ekstrand, J. T. Riedl, and J. A. Konstan, "Collaborative filtering recommender systems," Found. Trends Human-Computer Interact., vol. 4, no. 2, pp. 81–173, 2010, doi: https://doi.org/10.1561/1100000009.
14. Y. Tay, "Deep Learning based Recommendation," 2018, [Online]. Available: https://github.com/cheungdaven/DeepRec.
15. R. E. Nakhli, H. Moradi, and M. A. Sadeghi, "Movie Recommender System Based on Percentage of View," 2019 IEEE 5th Conf. Knowl. Based Eng. Innov. KBEI 2019, pp. 656–660, 2019, doi: https://doi.org/10.1109/KBEI.2019.8734976.
16. K. He, X. Zhang, S. Ren, and J. Sun, "Delving Deep into Rectifiers : Surpassing Human-Level Performance on ImageNet Classification," 2015.
17. S. K. Kumar, "On weight initialization in deep neural networks," pp. 1–9, 2017, [Online]. Available: http://arxiv.org/abs/1704.08863.
18. R. M. Sallam, M. Hussein, and H. M. Mousa, "An Enhanced Collaborative Filtering-based Approach for Recommender Systems," Int. J. Comput. Appl., vol. 176, no. 41, pp. 9–15, 2020, doi: https://doi.org/10.5120/ijca2020920531.
19. R. Ahuja, A. Solanki, and A. Nayyar, "Movie recommender system using k-means clustering and k-nearest neighbor," Proc. 9th Int. Conf. Cloud Comput. Data Sci. Eng. Conflu. 2019, pp. 263–268, 2019, doi: https://doi.org/10.1109/CONFLUENCE.2019.8776969.
20. M. Komorowski, D. C. Marshall, J. D. Salciccioli, and Y. Crutain, "Secondary Analysis of Electronic Health Records," Second. Anal. Electron. Heal. Rec., no. September, pp. 1–427, 2016, doi: https://doi.org/10.1007/978-3-319-43742-2.
21. F. O. Isinkaye, Y. O. Folajimi, and B. A. Ojokoh, "Recommendation systems: Principles, methods and evaluation," Egypt. Informatics J., vol. 16, no. 3, pp. 261–273, 2015, doi: https://doi.org/10.1016/j.eij.2015.06.005.
22. M. Saadati, S. Shihab, and M. S. Rahman, "Movie recommender systems: Implementation and performace evaluation," arXiv, 2019.