Development Tourism Destination Recommendation Systems Using Collaborative and Content-Based Filtering Optimized with Neural Networks

[1]Diki Fahrizal, [2]Jaja Kustija, [3]Muhammad Aqil Haibatul Akbar
[1,2]Program Study of Electrical Engineering Education, Universitas Pendidikan Indonesia, Bandung, Indonesia
[3]Program Study Magister of Pariwisata, Universitas Pendidikan Indonesia, Bandung, Indonesia
Email: [1]diki15@upi.edu, [2]jaja.kustija@upi.edu, [3]muhammad007aqil@upi.edu

Article Info
Article history:
Received Jan 28th, 2024
Revised Apr 10th, 2024
Accepted Apr 28th, 2024

Keyword:
Collaborative Filtering
Content Based Filtering
Destination Tourism
Neural Network
Recommendation System

ABSTRACT
Tourism, a vital sector in the global economy, benefits significantly from advancements in infrastructure, accessibility, and information availability. However, the vast volume of information can overwhelm travelers, underscoring the need for efficient recommendation systems. This research aims to develop an advanced tourist destination recommendation system by integrating Collaborative Filtering (CF) and Content-Based Filtering (CBF) models with Neural Networks. This approach seeks to enhance recommendation accuracy by closely aligning with user preferences and addressing the challenge of limited data. The study utilizes data from 523 tourist destinations across West Java, along with user preference assessments, encompassing stages of data collection, labeling, pre-processing, pre-training, neural network-based training, model evaluation, and the presentation of recommendation outcomes. The optimization of the recommendation models through neural networks has notably improved the precision of tourist destination suggestions, achieving Root Mean Square Error (RMSE) values below 0.37 for both CF and CBF approaches. This research significantly contributes to increasing the search efficiency and accuracy for tourist destination information, offering a strategic solution to the prevalent issue of information overload in the tourism industry.

Corresponding Author:
Jaja Kustija,
Program Study of Electrical Engineering Education,
Universitas Pendidikan Indonesia, Bandung, Indonesia,
Jl. Dr. Setiabudi No.207, Isola, Kec. Sukasari, Kota Bandung, Jawa Barat 40154.
Email: jaja.kustija@upi.edu
DOI: http://dx.doi.org/10.24014/ijaidm.v7i2.28713

1. INTRODUCTION

Tourism is an important economic sector throughout the world, which plays a major role in increasing income and promoting the culture of each country[1]. With a variety of destinations that offer natural beauty, cultural richness and unique experiences, the tourism sector attracts millions of tourists, both domestic and foreign, every year. This rapid growth in the tourism sector is driven by the development of better tourism infrastructure and increased accessibility. However, in the current digital era, the big challenge faced in the tourism sector is information overload[2]. Tourists are often faced with a large and varied volume of information when searching for and selecting tourist destinations, which can possibly hinder the decision-making process. This difficulty is exacerbated by the speed of information growth which is much faster than humans' ability to process and find relevant information[3], [4]. As a result, the process of searching for tourist destinations has become increasingly complex [5]. This creates a need for a system that can facilitate the search for tourist information, simplify the selection process, and improve the overall tourist experience [6].

In this context, the importance of recommendation systems that can filter, analyze and offer destination suggestions based on user preferences becomes increasingly clear [7], [8]. This recommendation system can not only improve search efficiency but also increase tourist satisfaction by providing more accurate and personalized recommendations to avoid excess information obtained by users[9], [10]. Studies related to recommendation systems have developed significantly, with diverse methods or approaches producing various findings in different contexts. These results depend on many factors, including the data source, selection preferences, data pre-processing techniques, the type of features used, and the type of algorithm model used to recommend[11], [12]. Although many other studies have discussed this topic, there are still several challenges in recommender system research, including cold start problems and data gaps [13].

Various studies have been carried out on the topic of recommendation systems for tourist destinations. One of the studies developed a contextual recommendation system using LSTM, which is divided into two stages, namely next location prediction and recommendations with high accuracy reaching 97.2%[14]. Another study developed a system with an input stage involving weather data and tourist reviews, which were processed using methods such as the Term Frequency-Inverse Document Frequency Vectorizer and artificial neural networks to produce recommendations with output in the form of an evaluation index[15]. The third research focuses on comparing hybrid recommendation system algorithms with collaborative content filtering algorithms for tourist destinations. Using samples from 228 tourist locations, it was found that the performance of collaborative content filtering had an accuracy of 93.297% higher than using a hybrid recommendation algorithm which was 87.322% with a statistically significant difference[16]. There is other research, namely a multilevel tourism recommendation system that identifies customer destinations. In addition to identifying travelers with similar interests and recommending destinations, recommendation systems improve system performance by providing multiple similar destinations to users. Data analysis is used to identify key preferences, interests and values. The proposed system improves travel recommendation system services by improving service quality[17]. Finally, a multicriteria tensor algorithm for tourism recommendation systems. The proposed model is first used to analyze places of interest, based on user preferences, reviews, and ratings. Important factors required for a recommendation system are identified through Collaborative Filtering. Experimental results show that the proposed model can make the tourism recommendation system more effective and stable[18].

Researchers highlight gaps in existing research by summarizing key findings from relevant literature. The literature used has certain criteria, such as the year of publication of the article must be within a maximum of 5 years and the research methodology must be relevant to the problem being discussed, be it experimental, observational or qualitative. Thus, the existence of these literature criteria ensures that the selected articles meet the standards of quality and relevance required to support the objectives of this research. Previous research has identified challenges in tourism recommendation systems, especially related to information overload and uncertainty in user preferences. However, little research focuses on integrating Collaborative Filtering and Content-Based Filtering models with Neural Network approaches to improve recommendation accuracy. Therefore, the main contribution of this research is developing a Collaborative Filtering and Content Based Filtering algorithm model that is optimized using a Neural Network for a tourist destination recommendation system. The use of Neural Network optimization in collaborative filtering models to help identify hidden patterns in user interactions or user preferences that may not be visible only with traditional algorithms. Meanwhile, using Neural Network optimization in content-based filtering to improve analysis of item features from tourist destinations, such as city, category, price, or location. There are three types of data sources used in this research, namely tourist destination data, user assessment preference data, and data from the city of the user providing the assessment. Tourist destination data consists of 523 tourist destinations in West Java province which includes the types of tourist destination categories, nature reserves, places of worship, culture and amusement parks which were collected through searching, crawling and scraping processes[19]. Then the preference data along with the user's origin are obtained using the interview method for the purpose of providing an assessment of the tourist destination being assessed.

The aim of this research is to increase the accuracy of the tourist destination recommendation system by paying attention to user preferences and overcoming the problem of lack of data. Analysis was carried out to understand the factors that influence user interest by using tourist destination ratings. Developed using Collaborative filtering and Content based filtering algorithm models which are optimized with Neural Networks to provide accuracy and precision in providing tourist destination recommendations to users. The results of the model that has been developed are then carried out with evaluation metrics to measure the extent to which the model developed is effective and efficient. The evaluation metrics used in this research refer to the Root Square Mean Error (RMSE), precision and recall, F1-Score, and Normalized Discounted Cumulative Gain (NDGC) values[20], [21]. This research explores the potential solutions expected from the problems presented, such as how the integration of Collaborative Filtering and Content-

Based Filtering models optimized with Neural Networks can increase the accuracy of tourist destination recommendations, as well as certain factors that influence the effectiveness of tourist destination recommendation systems, and how to optimize it.

2. RESEARCH METHOD

2.1. Dataset

The data used in this research contains different types of data, namely data on a collection of tourist destinations and user preference data. Data on a collection of tourist destinations was obtained through a process of crawling and scraping data through Indonesian tourist destination websites. This process was carried out to collect data on tourist destinations in various regions in West Java Province. The result was 523 tourist destination datasets in the West Java region with 9 (nine) attributes that had been given labels. Meanwhile, user preference data was obtained through an interview survey of 330 respondents. Respondents randomly assessed 523 tourist destinations in West Java. Ratings are given using a Likert scale from one to five. The results obtained were 10217 preference data from users regarding the assessment of tourist destinations in West Java which had 4 (four) attributes that had been labeled. Apart from that, there is another dataset as information from users who provide ratings on the tourist destination, totaling 330 data with 3 (three) attributes which have been labeled as additional information in the required dataset. The dataset that has been obtained in detail is as in Figure 1, which will be used as material to create a recommendation system algorithm model.

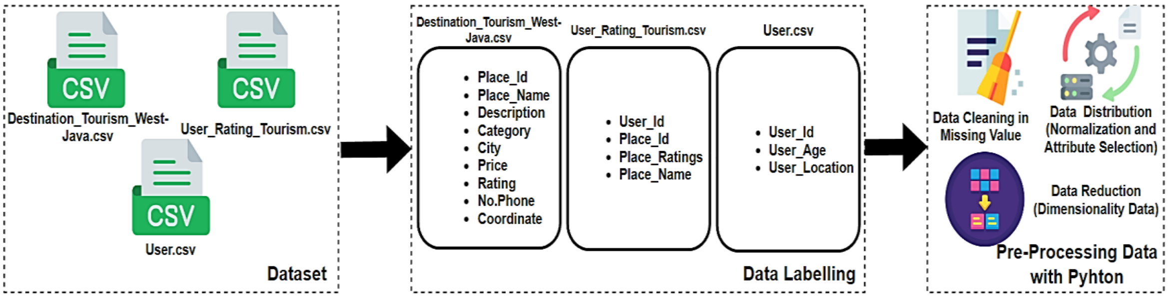

Figure 1. Dataset and Processing

2.2. Overall System Configuration

In this research, we propose an improvement algorithm for Collaborative Filtering (CF) and Content Based Filtering (CBF) that considers user preferences and other related information. The use of one-shot encoding, Term Frequency-Inverse Document Frequency Vectorizer, and Cosine Similarity is accompanied by Neural Network optimization which allows the algorithm to capture differences in user preferences and improve the accuracy of the recommendation system.

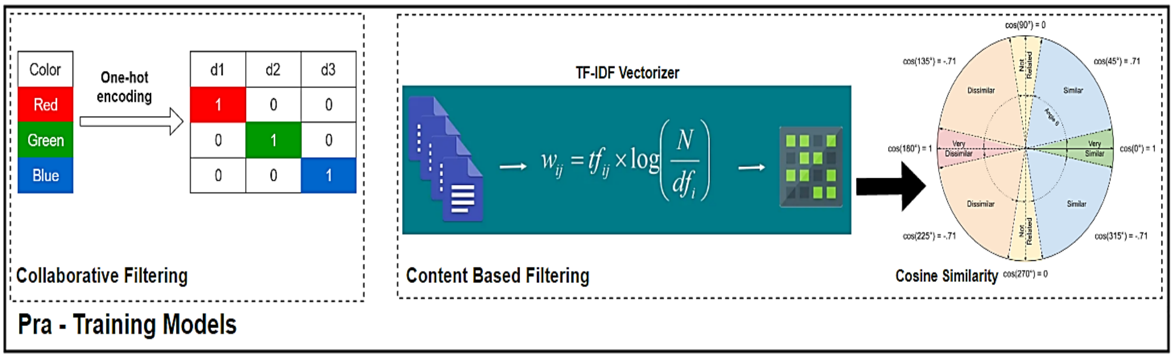

Figure 2. Pra - Training Model Step

Figure 2 shows the pre-training process of the model, where the Collaborative Filtering algorithm model using one-shot encoding is used to convert categorical data into a numerical representation. Meanwhile, Content-Based Filtering uses a combination of Term Frequency-Inverse Document Frequency Vectorizer and Cosine Similarity to analyze and compare text descriptions of tourist destinations. Next, Neural Network optimization is applied to both models to improve recommendation accuracy and precision. This process involves carefully tuning hyperparameters such as the number of epochs, optimizer, dropout rate, and learning rate decay to ensure optimal model convergence.

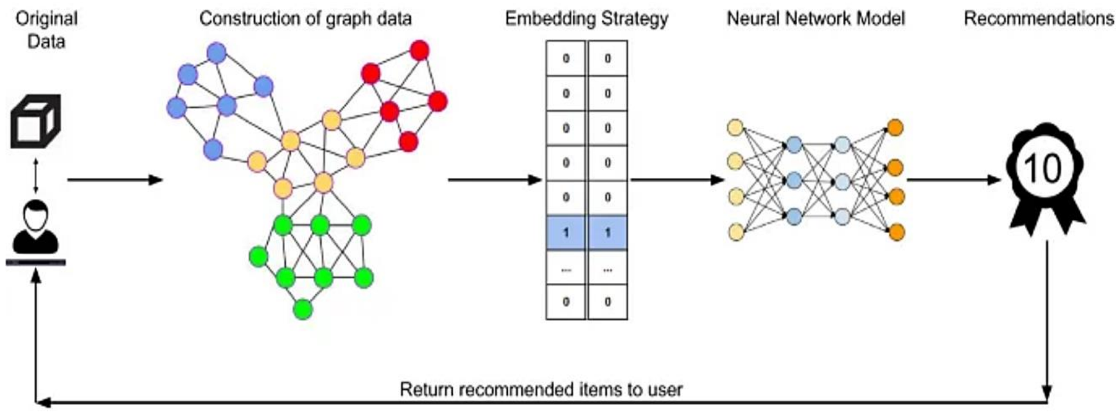

Figure 3. Architecture Overview in Recommendation System

Figure 3 is the recommendation system architecture optimized with Neural Network in this research. In the CF algorithm model, the RecommenderNet approach is used which is then optimized with a Neural Network to get good prediction accuracy[22]. Meanwhile, the CBF algorithm model uses a Sequential approach with several Dense layers to provide tourist destination recommendations based on items or preferences given by previous users[12], [23].

2.3. Collaborative Based Filtering with Neural Network

Collaborative Filtering (CF) is a recommendation approach that focuses on items liked or labeled by users with similar preferences. In Collaborative Filtering, building user profiles that include preferences collected from users with similar interests. This information can be explicit (ratings) or implicit (history of user visits). Collaborative Filtering uses a clustering algorithm to detect similarities between users and groups them into different groups based on these similarities. There are two types of Collaborative Filtering, namely Item – Based and User – Based[24]. The Collaborative Filtering algorithm used for this tourist destination recommendation system considers a collection of dimensional parameter vectors, namely $x^{(0)}, \ldots, x^{(n_m-1)}, \ldots, w^{(0)}, w^{(n_u-1)}$ and $b^{(0)}, \ldots, b^{(n_u-1)}$, where is the rating prediction model for tourist destinations(i) based on the user(j) that is $y^{(i,j)} = w^{(j)} \cdot x^{(i)} + b^{(i)}$. Given a dataset consisting of 10217 rating data generated from 330 users for ratings on 523 tourist destinations in West Java, the model will learn the parameter vector $x^{(0)}, \ldots, x^{(n_m-1)}, \ldots, w^{(0)}, w^{(n_u-1)}$ and $b^{(0)}, \ldots, b^{(n_u-1)}$, to produce the best match (minimizing the squared error). The collaborative filtering cost function for use in this research is given by [25].

$$J\left(x^{(0)}, \ldots, x^{(n_m-1)}, \ldots, w^{(0)}, b^{(0)}, \ldots, w^{(n_u-1)}, b^{(n_u-1)}\right) = \frac{1}{2}\sum_{(i,j):r(i,j)=1}\left(w^{(j)} \cdot x^{(i)} + b^{(i)} - y^{(i,j)}\right)^2 + regularization \quad (1)$$

This regularization is $\frac{\lambda}{2}\sum_{j=0}^{n_u-1}\sum_{k=0}^{n-1}\left(w_k^{(j)}\right)^2 + \frac{\lambda}{2}\sum_{i=0}^{n_m-1}\sum_{k=0}^{n-1}\left(x_k^{(i)}\right)^2$. The first summation in equation (1) is *for all i, j where r(i,j) equals 1*, so therefore.

$$J\left(x^{(0)}, \ldots, x^{(n_m-1)}, \ldots, w^{(0)}, b^{(0)}, \ldots, w^{(n_u-1)}, b^{(n_u-1)}\right) = \frac{1}{2}\sum_{j=0}^{n_u-1}\sum_{i=0}^{n_m-1} r(i,j) * \left(w^{(j)} \cdot x^{(i)} + b^{(i)} - y^{(i,j)}\right)^2 \quad (2)$$

In its implementation, the Collaborative Filtering model in this research uses the RecommenderNet framework approach to produce an architecture that is optimized with a neural network, as in table 1. There are two embedding layers that are used to convert categorical text tokens into dense vectors (*dense vector*) with richer information. In addition, there are two bias layers to increase flexibility and avoid model limitations during the training process.

Table 1. Overall Architecture Collaborative Filtering with Neural Networks

Layer (type)	Output Shape	Params
User Embedding	(Num_user, embedding_size)	Dependent on num_users, embedding_size
User Bias	(Num_users, 1)	Dependent on num_users
Places Embedding	(Num_places, embedding_size)	Dependent on num_places & embedding_size
Places Bias	(Num_places, 1)	Dependent on num_places

Layer (type)	Output Shape	Params
Dot Product	(Dependent on user_vector & places_vector)	-
Dropout	Same as input of this layer	Dependent on dropout_rate

2.4. Content Based Filtering with Neural Network

In the context of recommendation systems, Content Based Filtering is an approach that builds profiles for each user and item [26]. In this research, item attributes include information related to the name of the tourist attraction, category, description, price, etc. On the other hand, user attributes consist of descriptions of items that have been previously labeled or assigned a rating value. The comparison process is then carried out between item attributes and user attributes using distance and similarity methods, such as cosine similarity and Term Frequency-Inverse Document Frequency Vectorizer. Based on this, cosine similarity recommends items to target users that best match their user attributes. This approach can overcome the problem of recommending items without prior interaction. Then, the Term Frequency-Inverse Document Frequency Vectorizer is used to convert text or words into a vector representation that can be used for comparison in the Content Based Filtering algorithm[27]. In this research, content based filtering is used to produce attribute vectors for users and tourist destinations. In addition, this additional information is provided to the neural network to generate vectors of users and tourist destinations, as in the image below.

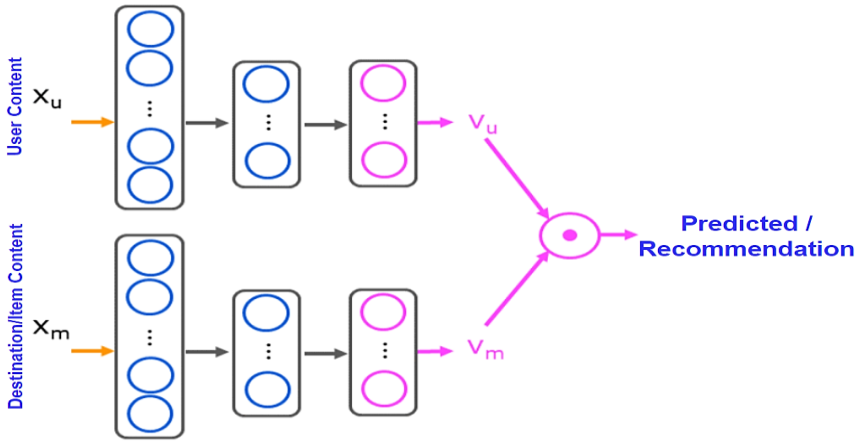

Figure 4. Layer Neural Network with Content Based filtering

In figure 4 above, user content consists only of features that have been generated. The average rating per tourist destination is calculated for each user. Additionally, there are User_Id, Place_Id, Place_Name, and Place_Rating available, but not included in model training or predictions. Nevertheless, this information will be useful in interpreting the data. The training set consists of all ratings provided by users in the dataset. The user and destination vectors are served to the above network as a training set. The user vector remains the same for all destinations rated by users.

n this research, the use of Content Based Filtering uses two types of feature vectors produced by a neural network, namely user feature vectors (v_u) and destination vector (v_m), both vectors have 64 entries in the neural network architecture. However, the similarity between items can be determined through comparing these feature vectors. This information is very useful in providing recommendations to users. For example, if a user gives a high rating to a tourist destination, the system can recommend similar destinations by selecting destinations that have similar destination feature vectors. A measure of the similarity between two feature vectors $\left(v_m^{(k)}\right)$ and destination vector $\left(v_m^{(i)}\right)$. In this study it is determined using the squared distance between the two vectors [28].

$$\left\|\left(v_m^{(k)}\right)-\left(v_m^{(i)}\right)\right\|^2=\sum_{l=1}^{n}\left(\left(v_{ml}^{(k)}\right)-\left(v_{ml}^{(i)}\right)\right)^2 \quad (3)$$

In its implementation, the Content Based Filtering model in this research uses Sequential from TensorFlow to produce an optimized architecture with a neural network consisting of three dense layers to integrate temporal or spatial information and two dropout layers to prevent overfitting of the model during training. The following is the architecture used in the CBF model, as in table 2.

Table 2. Overall Architecture Content Based Filtering with Neural Networks

Layer (type)	Output Shape	Params
dense (Dense)	(None, 128)	425600
dropout (Dropout)	(None, 128)	0
dense_1 (Dense)	(None, 64)	8256
dropout_1 (Dropout)	(None, 64)	0
dense_2 (Dense)	(None, 1)	65

2.5. Term Frequency-Inverse Document Frequency Vectorizer (TF-IDF Vectorizer)

In this research, we develop a method to measure similarities between tourist destinations by focusing on location and tourist categories. This process involves calculating the cosine similarity of the vectors that have been created[29]. The dataset used includes 523 tourist destinations, equipped with place names, categories, descriptions, tourist cities, prices and other important information. The application of TF-IDF Vectorizer makes it possible to convert tourist destination names into vector arrays, as the basis for similarity analysis in this research.

$$TF - IDF = TF(t,d) \times IDF(t) \tag{4}$$

In equation 4, t is the number of times the word or term t appears in document/dataset d. Term Frequency (TF) is used to calculate the frequency of words in a document, using a formula.

$$TF(t,d) = \sum_{x \epsilon d} FR(x,t) \tag{5}$$

$$FR(x,t) = \begin{cases} 1 & if\ x = t \\ 0 & if\ otherwise \end{cases} \tag{6}$$

Then Inverse Document Frequency (IDF), which is used to measure how important a word (tourist destination) appears, uses a formula.

$$IDF(t) = log \frac{|D|}{1+|d:t\epsilon d|} \tag{7}$$

To produce a vector array using TF-IDF, data from this research dataset is entered into the TF-IDF function. This function requires two inputs, namely the dataset which is the main data collection and the column name which determines a particular column in the dataset as the basis for creating a vector array[30].

2.6. Cosine Similarity

This research also applies the cosine similarity method to the Content Based Filtering model which is used to measure the similarity between two non-zero vectors in N-dimensional space. This similarity is inversely proportional to the size of the angle formed between them, the smaller the angle, the greater the similarity[31].

$$S(A,B) = cos\,\theta = \frac{A \cdot B}{|A| \cdot |B|} \tag{8}$$

Cosine similarity measures the cosine orientation between two vectors in multidimensional space, giving a score between -1 and 1. A score close to 1 indicates high similarity, while a score close to -1 indicates a large difference. In this research, the cosine similarity formula is implemented to compare user feature vectors with item feature vectors[32]. For example, in the context of tourist destination recommendations, user profiles containing preferences for certain locations are calculated against tourist destinations to identify destinations that best match the user's preferences. A higher similarity score indicates a closer correlation between user preferences and tourist destination attributes, allowing the system to suggest more relevant options.

2.7. Performance Evaluation

In this research, various metrics are used to evaluate the performance of the recommendation model. RMSE provides a measure of how close the prediction is to the actual value [33], [34]. Precision measures the model's efficiency in correctly classifying items as positive [35]. Recall measures the model's ability to capture all relevant positive samples [36]. F1 Score, which combines precision and recall, is used to assess the accuracy of classification models, especially when the class distribution is unbalanced[37]. NDCG is used to assess the effectiveness of a ranking system based on the order in which relevant items appear[38].

$$RMSE = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(y_i - \hat{y})^2} \quad (9)$$

$$Precision = \frac{Correctly\ recommended\ item}{Total\ recommended\ item} = \frac{TP}{Tp+FP} \quad (10)$$

$$Recall = \frac{Correctly\ recommended\ item}{Total\ recommended\ item} = \frac{TP}{Tp+FN} \quad (11)$$

$$F1\ Score = \frac{2}{\frac{1}{Precision}+\frac{1}{Recall}} = \frac{2\times(Precision\times Recall)}{(Precision+Recall)} \quad (12)$$

$$NDCG = \frac{1}{IDCG} \times \sum_{i=1}^{K} \frac{2^{r_i}-1}{log_2(i+1)} \quad (13)$$

3. RESULTS AND ANALYSIS

This section explains the experimental setup and evaluation methods used in this research. The methodology includes the implementation of an algorithm model using collaborative filtering (CF) and content based filtering (CBF) techniques, with improvements through Neural Network optimization for tourist destination recommendations. Furthermore, this section discusses the use of five evaluation metrics: root mean square error, precision, recall, F1-score, and normalized discounted cumulative gain, which are important for measuring the effectiveness of recommendation systems.

3.1. Dataset Analysis

The dataset used in this research consists of synthetic data that is useful for training recommendation system models, user demographic info, and user rating preferences. This data was obtained from the process of searching, crawling and scraping data to obtain data on tourist destinations throughout West Java Province, as well as conducting interviews with users to provide ratings of tourist destinations in West Java. The data that has been obtained and processed is then subjected to Exploratory Data Analysis (EDA) which is used for initial tests with the aim of identifying patterns, looking for unusual values, and testing the correctness of assumptions. This involves analyzing the tourist destinations with the highest ratings from users, comparing the number of tourist categories in West Java, the age distribution of users, and the number of districts or cities from previous user preferences.

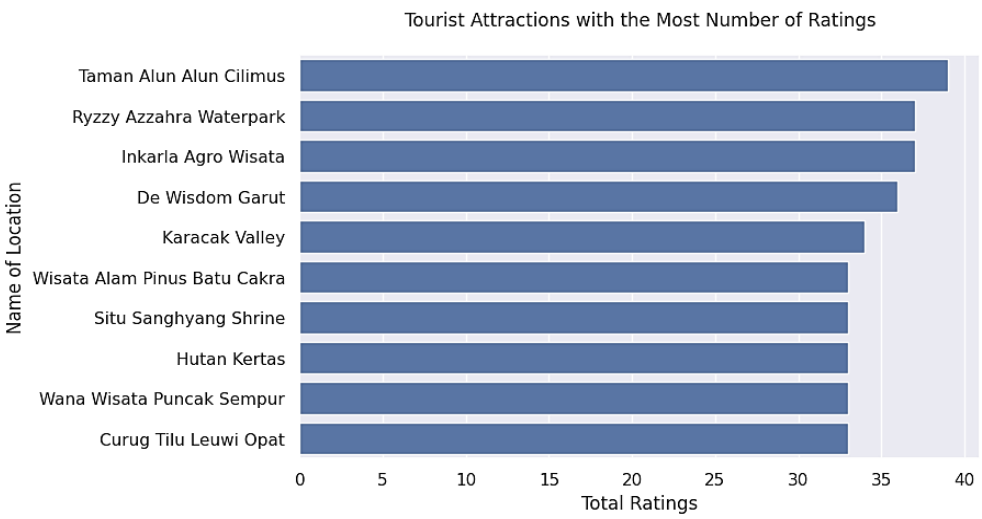

Figure 5. Tourist Attractions with the Most Number of Ratings

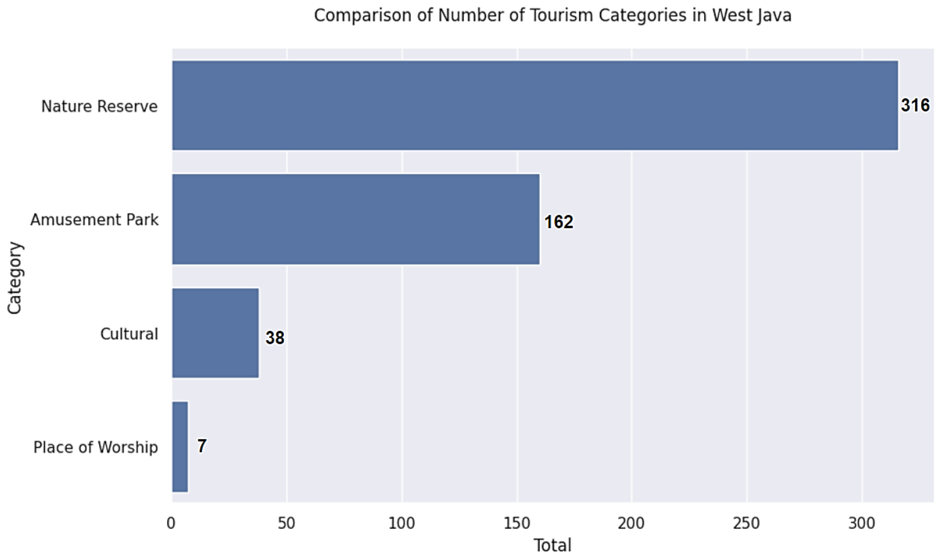

Figure 6. Comparison Number of Tourism Categories in West Java

Exploratory data analysis of tourist destination data in various regions of West Java province which is used as training data objects for modeling recommendation systems. Figure 5 shows the results of the analysis of the entire dataset to see which tourist destinations have the highest rating factors from users. Meanwhile, Figure 6 shows that the distribution of tourist destination categories from the data that has been collected is 60.4% in the “Cagar Alam” category; 30.9% for “Taman Hiburan”; 7.3% for “Budaya”; and 1.4% for “Tempat Ibadah”. This information can be especially useful in designing algorithm modeling using a content based filtering approach.

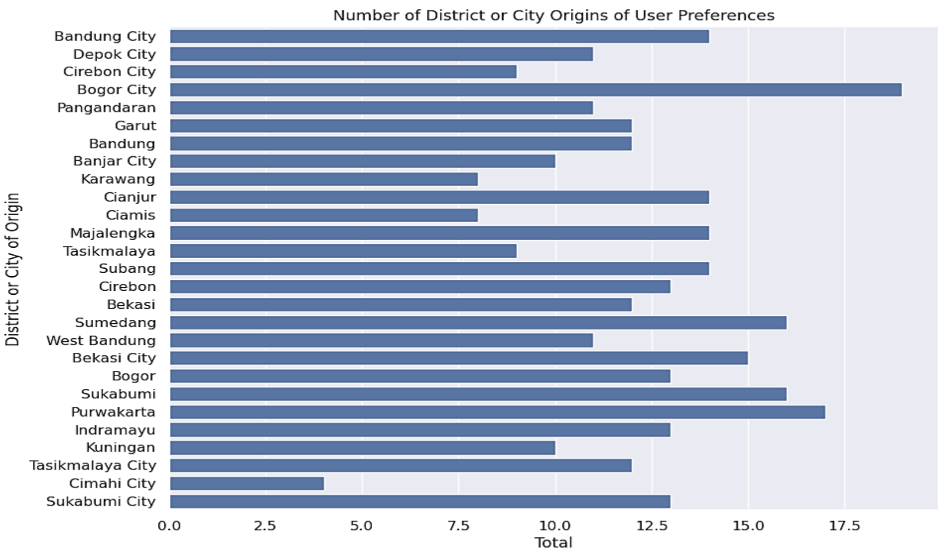

Figure 7. Comparison District or CIty Origins of User Preference

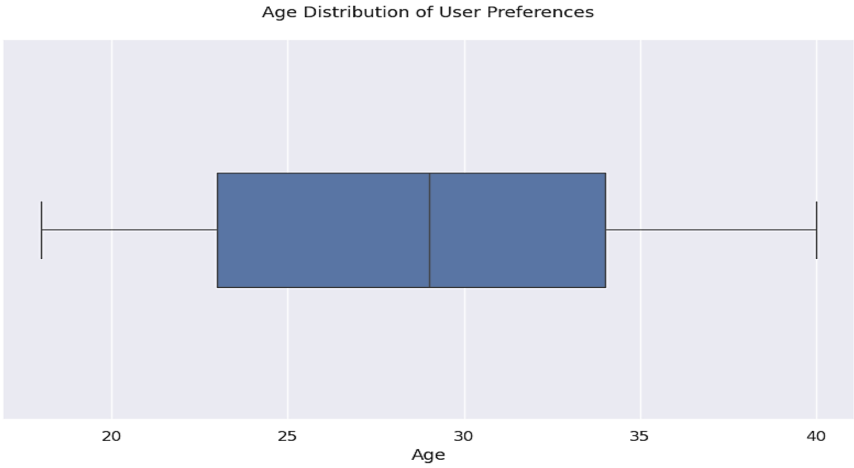

Figure 8. Age Distribution of User Preference

Exploratory data analysis of user preference data that provides an assessment of tourist destinations in West Java. Figure 7 shows the distribution of the number of cities or districts of users who rated tourist destinations. Meanwhile, Figure 8 shows the distribution of the age range of users who provide ratings on tourist destinations. These two pieces of information are really needed for collaborative filtering modeling with a user based approach or content based filtering modeling to provide recommendations for tourist destinations based on certain items.

3.2. Recommendation System Result

In this model training process, there are differences in the pre-model training process methods. The Collaborative Filtering (CF) model uses Integer Encoding before training the model, this aims to convert categorical data to numeric, as well as create a unique mapping between each categorical value and an integer. This can help recommend system algorithms that require a numeric input. The Content Based Filtering (CBF) model uses TF-IDF Vectorizer and Cosine Similarity to carry out the pre-training process for the model. The TF-IDF Vectorizer method is used to convert text to numeric vectors, as a weighting of words by providing a score indicating how important a word is in the dataset. This helps in identifying words that are distinctive and informative to users. In addition, to reduce the dimensions of text data in order to reduce the number of unnecessary features so that it is easy to analyze. Meanwhile, the Cosine Similarity method is used to measure similarities between texts, forming a similarity matrix to find similar texts based on the content of the user's text and in the context of a recommendation system to suggest items that are similar to those that the user liked or visited previously.

After carrying out the pre-training model process, the researcher created a training hyperparameter for the Collaborative Filtering and Content Based Filtering algorithm models, as in table 3. This aims to ensure that the trained model obtains the smallest error predictions.

Table 3. Training Hyperparameter Setting of the Collaborative Filtering and Content Based Filtering

Models	Hyperparameter
Collaborative Filtering (CF)	Epochs = 100; Optimizer = Adam; Embedding Size = 50; Dropout Rate = 0.2; Learning Rate = 10^{-3}; Learning Rate Decay Steps = 10^{5}; Learning Rate Decay Rate = 0.96; Regularization L2 = 10^{-6}.
Content Based Filtering (CBF)	Epochs = 100; Optimizer = Adam; Dropout Rate = 0.5; Learning Rate = 10^{-3}; Batch Size = 32; Validation Split = 0.1.

Hyperparameter settings as in table 3, for the Collaborative Filtering model include the use of the Adam optimizer, the number of epochs is 100, the embedding size is 50, and the dropout rate is 0.2. Meanwhile, the Content-Based Filtering model uses the Adam optimizer, dropout rate is 0.5, number of epochs is 100, batch size is 32, and validation division is 0.1. These settings are based on experimental results and careful tuning to ensure optimal model performance.

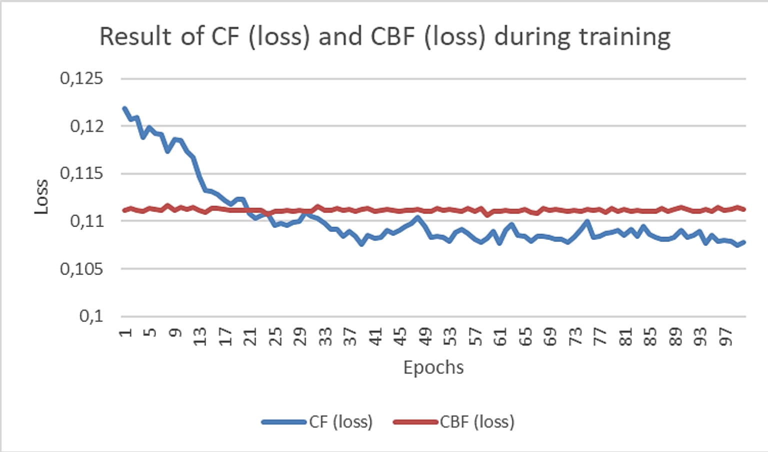

Figure 9. Training Losses of the Two Models

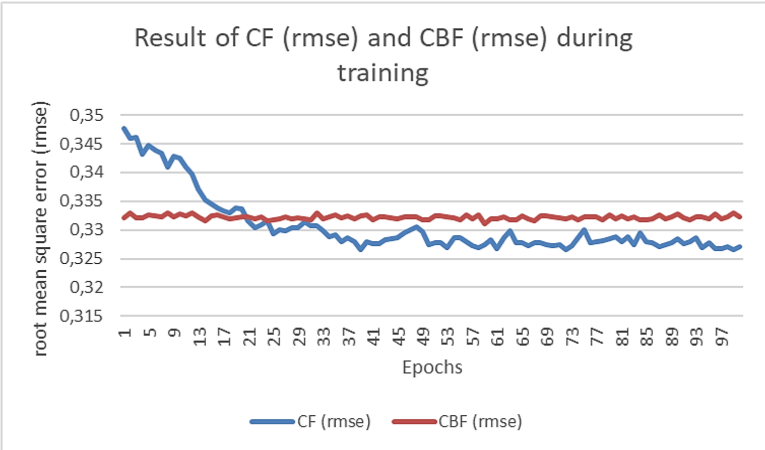

Figure 10. Training RMSE of the Two Models

Analysis of the results of the Collaborative Filtering (CF) model with Content Based Filtering (CBF) in terms of loss and RMSE during a total of 100 epochs of training. Figure 9 shows that both models have a decreasing trend in loss as epochs increase. While figure 10 shows the CF model gradually decreases over training epochs with some fluctuations, which indicates the CF model learns and improves its prediction accuracy which appears to improve to around 0.3265. In addition, the CBF model throughout training the RMSE value is quite stable and remains in the range of 0.33 which indicates that the model learning has quickly reached a stable point given by the features and complexity of the model.

The validation method is carried out by dividing the dataset into training data and validation data in a ratio of 80:20. This validation process ensures that the model developed can provide accurate and relevant tourist destination recommendations based on user preferences.

Table 4. Evaluation Performance Result of the Two Models

Model Algorithm	Metrics Evaluation					
	MSE	RMSE	Precision	Recall	F1-Score	NDCG
Collaborative Filtering	0.13	0.36	0.62	0.97	0.76	0.89
Content Based Filtering	0.12	0.35	0.40	0.64	0.49	0.91

The results of the model that has been trained can provide recommendations for tourist destinations to users with various functions. The use of the Collaborative Filtering model, as in Figure 11, displays recommendations for tourist destinations based on preferences that have been previously given by the user. Meanwhile, using the Content Based Filtering model, as in Figure 12, displays recommendations for tourist destinations based on specific items, which allows users to find out several destination recommendations based on certain items, such as category, price, location, and others.

```
List of recommendations for: User 221
===========================================

----------------------------------------------------------
Places with the highest travel ratings from users
----------------------------------------------------------
Taman Bunga Cihideung : Nature Reserve
Situ Rawa Gede Kota Bekasi : Nature Reserve
Kawah Papandayan : Nature Reserve
Taman Kota : Amusement Park
Curug Koleangkak (Curug Biru) : Nature Reserve

----------------------------------------------------------
Top 7 place recommendation
----------------------------------------------------------
1 . Alun-alun Sumedang
     Amusement Park , Entry Ticket Price  20404 , Tourist Rating  4,6

2 . Batu Mahpar Wisata Alam Galunggung
     Nature Reserve , Entry Ticket Price  46869 , Tourist Rating  4,3

3 . Taman Kota Tasikmalaya
     Amusement Park , Entry Ticket Price  39090 , Tourist Rating  4,5

4 . Curug Ciparay Tasikmalaya
     Nature Reserve , Entry Ticket Price  41983 , Tourist Rating  4,5

5 . CIMAHI Convention Hall
     Cultural , Entry Ticket Price  0 , Tourist Rating  4.5

6 . Masjid Kubah Emas Dian Al-Mahri Depok
     Nature Reserve , Entry Ticket Price  0 , Tourist Rating  4.7

7 . Taman Rekreasi Wiladatika
     Amusement Park , Entry Ticket Price  0 , Tourist Rating  4.4

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

Figure 11. Recommendation Destination Tourism Result with Collaborative Based Filtering

A recommendation system using Collaborative Filtering suggests tourist destinations based on user preferences. The system will provide a predicted rating for places that have not been visited by the user, then select and display the top seven recommendations based on that rating. Apart from that, it displays information about places that users have visited and liked.

	name	category	description	city
0	SITU JANAWI DESA PAYUNG	Nature Reserve	SITU JANAWI DESA PAYUNG mungkin merujuk pada d...	Majalengka
1	Goa Maria Fatimah Sawer Rahmat	Nature Reserve	Goa Maria Fatimah Sawer Rahmat adalah sebuah t...	Kuningan
2	Lawang Saketeng	Nature Reserve	Lawang Saketeng mungkin merujuk pada suatu are...	Majalengka
3	Cagar Budaya Pulo Majeti	Nature Reserve	Situs cagar budaya dengan nilai sejarah, menaw...	Banjar
4	Gedung Perundingan Linggarjati	Cultural	Gedung Perundingan Linggarjati adalah situs be...	Kuningan
5	Benteng Warna	Cultural	Benteng Warna mungkin merujuk pada benteng sej...	Karawang
6	Patilasan Prabu Siliwangi(Tapakan)	Cultural	Patilasan Prabu Siliwangi (Tapakan) adalah tem...	Majalengka
7	Tugu Ciporang Maleber	Amusement Park	Tugu Ciporang Maleber mungkin adalah landmark ...	Kuningan
8	Woodland	Amusement Park	Woodland mungkin merujuk pada hutan atau area ...	Kuningan
9	Monumen Bumi Patra	Amusement Park	Monumen Bumi Patra mungkin adalah monumen atau...	Indramayu
10	Pasanggrahan Prabu Siliwangi	Amusement Park	Pasanggrahan Prabu Siliwangi mungkin adalah te...	Majalengka

Figure 12. Recommendation Destination Tourism Result with Content-Based Filtering

A recommendation system with Content Based Filtering, provides recommendations that are not only similar to the place the user is looking for, but also ensures that the recommendations are diverse in terms of category and location. This is especially useful where diversity of experience is key to satisfying a wide range of user preferences.

4. CONCLUSION

This research succeeded in developing a tourist destination recommendation system using Collaborative Filtering (CF) and Content-Based Filtering (CBF) techniques which were optimized using a Neural Network. Exploratory data analysis revealed important patterns in user preferences and distribution of tourist destinations for model development. The implementation of Integer Encoding in the CF model and the use of the Term Frequency-Inverse Document Frequency Vectorizer and Cosine Similarity in the CBF model show effectiveness in processing text and categorical data into numerical information that can be utilized by the model.

The experimental results show that both CF and CBF models have good abilities in predicting and recommending tourist destinations. This is proven through performance evaluations which include MSE, RMSE, Precision, Recall, F1-Score, and NDCG metrics. The CF model shows better performance in terms of precision and recall, while the CBF model excels in providing diverse recommendations according to user-specific criteria.

The developed recommendation system makes an important contribution to the tourism sector by simplifying the process of searching for tourist destinations and improving the overall user experience. By taking individual preferences into account and overcoming the problem of information overload, this system can help tourists find destinations that best suit their interests. We realize that the adequacy of the parameters used in this research is key in ensuring the validity and reliability of the research results. Therefore, researchers plan to conduct in-depth analysis of additional parameters for the decision-making recommendation system to ensure it covers a wide spectrum to represent diverse situations. The success of the system built in this research marks the first step forward in the use of advanced technology to increase the effectiveness of recommendation systems in the tourism industry.

ACKNOWLEDGEMENTS

We would like to thank all parties who were actively involved in the continuation of this research, so that the results of this research were as expected. Our thanks also go to the publisher organizers who have received this paper in a good way, so that this research can be used as a reference for other researchers.

REFERENCES

[1] P. Pandey, K. Mayank, and S. Sharma, "Recommendation System for Adventure Tourism," in *2023 4th IEEE Global Conference for Advancement in Technology (GCAT)*, 2023, pp. 1–7. doi: 10.1109/GCAT59970.2023.10353339.

[2] C. Huda, A. Ramadhan, A. Trisetyarso, E. Abdurachman, and Y. Heryadi, "Smart Tourism Recommendation Model: A Systematic Literature Review." [Online]. Available: www.ijacsa.thesai.org

[3] K. A. Fararni, F. Nafis, B. Aghoutane, A. Yahyaouy, J. Riffi, and A. Sabri, "Hybrid recommender system for tourism based on big data and AI: A conceptual framework," *Big Data Mining and Analytics*, vol. 4, no. 1, pp. 47–55, 2021, doi: 10.26599/BDMA.2020.9020015.

[4] 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.

[5] L. Esmaeili, S. Mardani, S. A. H. Golpayegani, and Z. Z. Madar, "A novel tourism recommender system in the context of social commerce," *Expert Syst Appl*, vol. 149, p. 113301, 2020, doi: https://doi.org/10.1016/j.eswa.2020.113301.

[6] L. Chen, J. Cao, Y. Wang, W. Liang, and G. Zhu, "Multi-view Graph Attention Network for Travel Recommendation," *Expert Syst Appl*, vol. 191, p. 116234, 2022, doi: https://doi.org/10.1016/j.eswa.2021.116234.

[7] Y. Liu *et al.*, "Interaction-Enhanced and Time-Aware Graph Convolutional Network for Successive Point-of-Interest Recommendation in Traveling Enterprises," *IEEE Trans Industr Inform*, vol. 19, no. 1, pp. 635–643, 2023, doi: 10.1109/TII.2022.3200067.

[8] J. Kustija, A. Ana, and N. Dwi Jayanto, "WEB-BASED AND THINVNC REMOTE LABORATORY IMPLEMENTATION TO SUPPORT STUDENTS SKILLS IN MECHATRONICS COURSE TO FACE THE INDUSTRIAL REVOLUTION 4.0," 2021.

[9] X. Huang, V. Jagota, E. Espinoza-Muñoz, and J. Flores-Albornoz, "Tourist hot spots prediction model based on optimized neural network algorithm," *International Journal of System Assurance Engineering and Management*, vol. 13, no. 1, pp. 63–71, 2022, doi: 10.1007/s13198-021-01226-4.

[10] I. Ivanova and M. Wald, "Recommender Systems for Outdoor Adventure Tourism Sports: Hiking, Running and Climbing," *Human-Centric Intelligent Systems*, vol. 3, no. 3, pp. 344–365, Jul. 2023, doi: 10.1007/s44230-023-00033-3.

[11] V. Garipelly, P. T. Adusumalli, and P. Singh, "Travel Recommendation System Using Content and Collaborative Filtering - A Hybrid Approach," in *2021 12th International Conference on Computing Communication and Networking Technologies (ICCCNT)*, 2021, pp. 1–4. doi: 10.1109/ICCCNT51525.2021.9579907.

[12] J. Kustija, "SOLUTIONS TO OVERCOME INEQUALITY IN LABORATORY FACILITIES AND LABORATORY SHARING IN SIMILAR INSTITUTIONS REMOTE LABORATORY BASED," 2022.

[13] J. Kumar, B. K. Patra, B. Sahoo, and K. S. Babu, "Group recommendation exploiting characteristics of user-item and collaborative rating of users," *Multimed Tools Appl*, 2023, doi: 10.1007/s11042-023-16799-4.

[14] H. 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: 10.1007/s12652-019-01521-w.

[15] Y. Liu *et al.*, "Interaction-Enhanced and Time-Aware Graph Convolutional Network for Successive Point-of-Interest Recommendation in Traveling Enterprises," *IEEE Trans Industr Inform*, vol. 19, no. 1, pp. 635–643, 2023, doi: 10.1109/TII.2022.3200067.

[16] 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-Resour. Lang. Inf. Process.*, vol. 20, no. 5, Jun. 2021, doi: 10.1145/3453651.

[17] H. Alrasheed, A. Alzeer, A. Alhowimel, N. Shameri, and A. Althyabi, “A Multi-Level Tourism Destination Recommender System,” in *Procedia Computer Science*, Elsevier B.V., 2020, pp. 333–340. doi: 10.1016/j.procs.2020.03.047.

[18] M. Hong and J. J. Jung, “Multi-criteria tensor model for tourism recommender systems,” *Expert Syst Appl*, vol. 170, p. 114537, 2021, doi: https://doi.org/10.1016/j.eswa.2020.114537.

[19] V. Boppana and P. Sandhya, “Web crawling based context aware recommender system using optimized deep recurrent neural network,” *J Big Data*, vol. 8, no. 1, p. 144, 2021, doi: 10.1186/s40537-021-00534-7.

[20] M. Aldayel, A. Al-Nafjan, W. M. Al-Nuwaiser, G. Alrehaili, and G. Alyahya, “Collaborative Filtering-Based Recommendation Systems for Touristic Businesses, Attractions, and Destinations,” *Electronics (Switzerland)*, vol. 12, no. 19, Oct. 2023, doi: 10.3390/electronics12194047.

[21] N. Wayan Priscila Yuni Praditya, A. Erna Permanasari, I. Hidayah, M. Indana Zulfa, and S. Fauziati, “Collaborative and Content-Based Filtering Hybrid Method on Tourism Recommender System to Promote Less Explored Areas How to Cite: Praditya, Ni et.al., 2022.Collaborative and Content-Based Filtering Hybrid Method on Tourism Recommender System to Promote Less Explored Areas,” 2022.

[22] K. V. Dudekula *et al.*, “Convolutional Neural Network-Based Personalized Program Recommendation System for Smart Television Users,” *Sustainability (Switzerland)*, vol. 15, no. 3, Feb. 2023, doi: 10.3390/su15032206.

[23] D. Sharma, E. Banwala, I. Elouaghzani, and R. Katarya, “EasyChair Preprint Analysis of Personalized Tourism Recommender Systems Analysis of Personalized tourism recommender systems,” 2023.

[24] L. Wu, Y. Xia, S. Min, and Z. Xia, “Deep Attentive Interest Collaborative Filtering for Recommender Systems,” *IEEE Trans Emerg Top Comput*, pp. 1–15, 2023, doi: 10.1109/TETC.2023.3286404.

[25] S. Venkatesan, “A Recommender System Based on Matrix Factorization Techniques Using Collaborative Filtering Algorithm,” *neuroquantology*, vol. 21, no. 5, p. 864, 2023.

[26] R. Widayanti, M. Heru, R. Chakim, C. Lukita, U. Rahardja, and N. Lutfiani, “Improving Recommender Systems using Hybrid Techniques of Collaborative Filtering and Content-Based Filtering,” *Journal of Applied Data Sciences*, vol. 4, no. 3, pp. 289–302, 2023.

[27] S. Bhaskaran and R. Marappan, “Enhanced personalized recommendation system for machine learning public datasets: generalized modeling, simulation, significant results and analysis,” *International Journal of Information Technology*, vol. 15, no. 3, pp. 1583–1595, 2023.

[28] G. Chalkiadakis, I. Ziogas, M. Koutsmanis, E. Streviniotis, C. Panagiotakis, and H. Papadakis, “A Novel Hybrid Recommender System for the Tourism Domain,” *Algorithms*, vol. 16, no. 4, Apr. 2023, doi: 10.3390/a16040215.

[29] M. Das and P. J. A. Alphonse, “A comparative study on tf-idf feature weighting method and its analysis using unstructured dataset,” *arXiv preprint arXiv:2308.04037*, 2023.

[30] B. Kabra and C. Nagar, “Convolutional neural network based sentiment analysis with tf-idf based vectorization,” *Journal of Integrated Science and Technology*, vol. 11, no. 3, p. 503, 2023.

[31] S. Sumathi and M. Suriya, “Certain Investigations on Cognitive based Movie Recommendation system using Pairwise Cosine Similarity,” in *2023 9th International Conference on Advanced Computing and Communication Systems (ICACCS)*, IEEE, 2023, pp. 2139–2143.

[32] C. V. M. Krishna, G. A. Rao, and S. Anuradha, “Analysing the impact of contextual segments on the overall rating in multi-criteria recommender systems,” *J Big Data*, vol. 10, no. 1, p. 16, 2023.

[33] B. A. Hamed, O. A. S. Ibrahim, and T. Abd El-Hafeez, “Optimizing classification efficiency with machine learning techniques for pattern matching,” *J Big Data*, vol. 10, no. 1, p. 124, 2023.

[34] D. Fahrizal and J. Kustija, “Creating an Innovative Mechatronics Learning Experience with Mecha-Learn: A Website-Based Platform,” 2023. [Online]. Available: www.multiresearchjournal.com

[35] D. Park and S. Kim, “Probabilistic Precision and Recall Towards Reliable Evaluation of Generative Models,” in *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 2023, pp. 20099–20109.

[36] H. Shang, J.-M. Langlois, K. Tsioutsiouliklis, and C. Kang, “Precision/Recall on Imbalanced Test Data,” in *International Conference on Artificial Intelligence and Statistics*, PMLR, 2023, pp. 9879–9891.

[37] D. Fourure, M. U. Javaid, N. Posocco, and S. Tihon, “Anomaly detection: How to artificially increase your f1-score with a biased evaluation protocol,” in *Joint European Conference on Machine Learning and Knowledge Discovery in Databases*, Springer, 2021, pp. 3–18.

[38] H. Brama, L. Dery, and T. Grinshpoun, “Evaluation of neural networks defenses and attacks using NDCG and reciprocal rank metrics,” *Int J Inf Secur*, vol. 22, no. 2, pp. 525–540, 2023.

BIBLIOGRAPHY OF AUTHORS

Diki Fahrizal, A student who is pursuing a bachelor's degree in Electrical Engineering Education at the Department of Electrical Engineering Education, Faculty of Technology and Vocational Studies, Universitas Pendidikan Indonesia. Actively involved with other lecturers in conducting research. He research interests are Internet of Things, Artificial Intelligence, Computer Vision, Machine Learning, Data Mining, and VR/AR Technology. He can be contacted at email: diki15@upi.edu

Prof. Dr. Jaja Kustija, M.Sc., Received a bachelor's degree in Electrical Engineering Education at the Indonesian University of Education in 1984, received a master's degree M.Sc. in the field of Instrumentation Physics at the Bandung Institute of Technology in 1992, and received his doctorate in Technology and Vocational Education at the Indonesian University of Education in 2012. The research and publications carried out are very experienced both on a national scale indexed by SINTA and on an international scale indexed by Scopus, and have research umbrellas with other institutions or with students with major topics Design, Manufacture, Testing Monitoring and Remote Control and Learning Media based on Electronics and the Internet of Things. Books that have been published by one of the national scale publishers include; Electromagnetic Field theory and Applications; Mechatronics books for colleges; Electromagnetic Field With Required Engineering Mathematics Enrichment; Electrical Circuit 1 Equipped with Matlab For Calculation Operations. Now he is a professor at the Indonesian University of Education in the field of Electrical and Mechatronic Engineering Education. He can be contacted at email: jaja.kustija@upi.edu

Muhammad Aqil Haibatul Akbar, S.T., Is a postgraduate student majoring in tourism at Universitas Pendidikan Indonesia. The last education taken is a bachelor of engineering majoring in industrial engineering at the Institut Teknologi Nasional, Bandung. He is actively involved with other lecturers in conducting research. His research interests are Industrial Engineering, Ecotourism, and Community-Based Tourism. He can be contacted via email : muhammad007aqil@upi.edu