E-ISSN : 2540 - 8984

JIPI (Jurnal Ilmiah Penelitian dan Pembelajaran Informatika)
Volume 07, Nomor 02, Juni 2022 : 408– 418

HYBRID RECOMMENDER SYSTEM USING SINGULAR VALUE DECOMPOSITION AND SUPPORT VECTOR MACHINE IN BALI TOURISM

I K. G. Aryadi Pramarta[1)], Z. K. A. Baizal[2*)]
[1, 2)]Informatics, School of Computing, Telkom University
Jl. Telekomunikasi No 1, Bandung
e-mail: aryadipramarta@student.telkomuniversity.ac.id [1)], baizal@telkomuniversity.ac.id[2)]
*Correspondence Author

ABSTRAK

Saat akan melakukan kunjungan pada suatu daerah wisata, wisatawan harus menentukan tempat yang dia ingin kunjungi. Sementara itu tempat yang diinginkan memiliki beberapa kategori dan jenis. Banyaknya jenis tempat wisata membuat wisatawan kebingungan dalam menentukan pilihannya. Oleh karena itu kami berfokus dalam pembangunan sistem pemberi rekomendasi hybrid dengan menggabungkan beberapa pendekatan rekomendasi yaitu Collaborative Filtering, Content Based Filtering, dan Demographic Filtering. Sistem pemberi rekomendasi ini dibangun untuk menyelesaikan cold start problem yang sering muncul pada sistem pemberi rekomendasi Collaborative Filtering dan Content Based Filtering. Dalam penelitian ini teknik weighted dan switching dipilih menjadi metode hibridisasi. Kedua teknik ini digunakan untuk menangani kelemahan masing masing teknik, sehingga menjadi sistem pemberi rekomendasi yang lebih baik. Algoritma Singular Value Decomposisition (SVD) dipilih untuk digunakan pada Collaborative Filtering sementara itu, Content Based Filtering menggunakan perhitungan nilai cosine similarity dan untuk Demographic Filtering menggunakan algoritma Support Vector Machine (SVM). Data yang digunakan pada penelitian ini adalah data destinasi wisata yang ada di daerah Bali yang didapat dari crawling pada situs TripAdvisor. Pada penelitian ini Root Mean Square Error (RMSE) dan Mean Absolute Error (MAE) digunakan sebagai pengukur tingkat akurasi prediksi rating yang dihasilkan. Hasil percobaan yang dilakukan menunjukkan bahwa hybrid method yang dibangun menghasilkan hasil prediksi akurasi yang lebih baik dibandingkan saat dijalankan secara terpisah dengan nilai rata-rata Mean Absolute Error (MAE) sebesar 0.6660 dan untuk nilai Root Mean Square Error (RMSE) sebesar 0.8644.

***Kata Kunci**: Hybrid Recommender System, Singular Value Decomposition, Support Vector Machine, Switching, Weighted*

ABSTRACT

When going to make a visit to a tourist area, tourists must determine the place they want to visit. Meanwhile, the desired place has several categories and types. The many types of tourist attractions make tourists confused in determining their choice. Therefore, we focus on developing a hybrid recommendation system by combining several recommendations approaches, namely collaborative filtering, content-based filtering, and demographic filtering. This recommendation system was built to solve the cold start problem that often appears in collaborative filtering and content-based filtering. In this study, weighted and switching techniques were chosen as the hybridization method. These two techniques are used to overcome the weaknesses of each technique so that it becomes a better recommendation system. The singular value decomposition (SVD) algorithm was chosen to be used in collaborative filtering, meanwhile, content-based filtering uses the calculation of cosine similarity values , and demographic filtering uses the support vector machine (SVM) algorithm. The data used in this study is data on tourist destinations in the Bali area obtained from crawling on the TripAdvisor site. In this study, the root mean square error (RMSE) and mean absolute error (MAE) was used to measure the accuracy of the resulting rating prediction. The results of the experiments carried out show that the hybrid method that was built produces better accuracy prediction results than when run separately with an average mean absolute error (MAE) of 0.6660 and a root mean square error (RMSE) of 0.8644.

***Keywords**: Hybrid Recommender System, Singular Value Decomposition, Support Vector Machine, Switching, Weighted*

I. INTRODUCTION

The rapid development of technology has affected all sectors. Tourism is one sector that is affected by this. Currently, traveling is an activity that is usually done by people on holidays [1]. The presence of the internet as a source of information is very helpful in the world of tourism. By using the internet, tourists get convenience in finding the desired tourist attractions. Tourism managers also use the internet as a marketing medium. So that tourists get a lot of information about the tourist destinations that are displayed. Most tourists do planning in the selection of tourist attractions by using the internet.

There are so many tourist attractions that do marketing through the internet, of course, making it more difficult

E-ISSN : 2540 - 8984

JIPI (Jurnal Ilmiah Penelitian dan Pembelajaran Informatika)
Volume 07, Nomor 02, Juni 2022 : 408– 418

for tourists to make their choices. The solution for choosing tourist attractions can be by using a recommendation system technology that is able to recommend tourist attractions in an area [2]. Various approaches and types of recommendation systems have been applied by many people in the tourism sector such as the collaborative filtering method [3], content-based filtering [4], demographic filtering [5], and several other approaches. The most frequently used approaches are collaborative filtering and content-based filtering.

L. Gang [3] built a system for recommending tourist attractions using a collaborative filtering approach. Meanwhile, O. Alnogaithan also implemented content-based filtering by utilizing user reviews [4]. Demographic filtering was also applied by Y. Wang [5] in building a tourist spot recommendation system by applying several machine learning methods such as naïve bayes, bayesian network, and support vector machine (SVM). In this study, an evaluation of the prediction results using RMSE was carried out. The support vector machine is the method that has the best performance when tested in this research. But demographic filtering is not able to run alone to obtain good results [5].

To improve accuracy and complete the shortcomings between methods, you can use an effective hybridization method or make it a hybrid recommender system [6]. The implementation of a hybrid recommendation system for tourist attractions has been carried out by several previous studies. Like the hybrid recommender system built by Kbaier using three methods [7]. The hybrid system that was built combines collaborative filtering, content-based filtering, and demographic filtering. In this study, the use of the k-nearest neighbor algorithm and the decision tree algorithm was chosen. The prediction results were evaluated using MAE, NMAE, and RMSE. The research carried out succeeded in ensuring that there were no cold start problems in the recommendation system that was built and resulted in better accuracy. Other research on the hybrid recommender system was also carried out in Bai's research, namely making a recommendation system with hybrid filtering using content-based filtering, collaborative filtering, and demographic filtering to extract PoIs (Point of Interest) which will be recommended [8].

In this research, we want to build a recommendation system model that has better accuracy and complements the shortcomings between methods by using an effective hybridization method [6]. The chosen technique is the weighted technique and the switching technique. The cold start problem and better accuracy results are the goals of choosing this technique. The choice of the algorithm used is the main differentiator between the existing research and the research we will be doing. For the chosen algorithm in the system built, namely, collaborative filtering with singular value decomposition (SVD) because it can improve prediction accuracy results and has good performance [9]. In content-based with the calculation of cosine similarity between profile items and user profiles. For demographic filtering with support vector machine (SVM) because the algorithm has the highest accuracy according to existing research [6]. Demographic data owned by users can help in solving the cold start problem [10]. In this study, we chose the mean absolute error (MAE) and root mean square error (RMSE) to evaluate the prediction results. By choosing an algorithm that has high accuracy from each approach, of course, in this study, we can get better.

II. RESEARCH METHODOLOGY

In this research, we build a hybrid recommendation system by combining singular value decomposition (collaborative filtering), cosine similarity (content-based filtering), and support vector machine (demographic filtering). It is hoped that the selection of this algorithm will produce better accuracy. The hybridization technique chosen is using the weighted and switching technique as found in previous studies [7]. We chose this technique with the aim of overcoming the cold start problem which is often the main problem when building a recommendation system using collaborative filtering and content-based filtering. The dataset used in this study was obtained through crawling on the TripAdvisor website. In the crawling process, we use the help of the WebHarvy application. This application is easy to use and can help in crawling data quickly. In general, the crawling results obtained are still not good, so it is necessary to do preprocessing first to ensure the dataset is ready to be used by the recommendation system model that was built. The design of the system built in this study is depicted in Figure 1.

E-ISSN : 2540 - 8984

JIPI (Jurnal Ilmiah Penelitian dan Pembelajaran Informatika)
Volume 07, Nomor 02, Juni 2022 : 408– 418

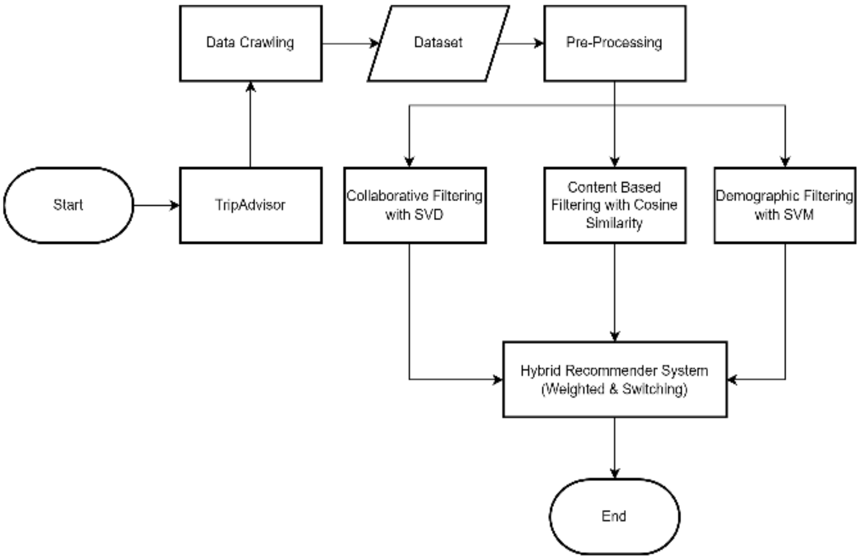

Fig.1. The Design of System

A. *Dataset*

The research dataset is sourced from the TripAdvisor website, obtained by crawling the website using an application called webharvy. This application helps in crawling and finding the data or information you want to get. For the selected destination location, namely the area in Bali, Indonesia. Bali was chosen because it has many famous tourist destinations in the world. With the help of the webharvy application, the crawled data is automatically exported to microsoft excel in CSV format. The extracted data, of course, needs to be cleaned first so that it can be used properly by the recommendation system that was built.

There are three types of datasets obtained from the crawling process. Among them is the rating dataset which contains information about the rating that has been given by the user to a particular destination, the user dataset demographic information from the user, and the destination dataset contains information about tourist destinations. The attributes of each of these datasets can be seen in Table 1 below.

TABLE I
DATASET AND ATTRIBUTES

Dataset Name	Dataset Attributes
dataset rating	Rating,Review,destination_id,user_id
dataset user	User_id,travel_style,Region,County,Member_Since
dataset destination	Destination_id,Destination Name,Rating,Total Review,Category,Location,Duration,Latitude,Longitude,Regency,Description

From table 1, we can see that there are quite a number of attributes in each dataset, so it is necessary to select attributes to match the model you want to build. This dataset will later be used in a recommendation system that is built based on the approach and also the type of technique that will be chosen.

B. *Data Preprocessing*

Problems often arise when we already have data. Several problems that arise such as the number of attributes that are excessive, missing values are found and data formats that do not match are found [11]. This is very influential on the next process that will be run, especially when the training model is being built. Data preprocessing is an important and main step in the data mining process [12]. In this study, different handling steps were carried out for each dataset. This step is taken because it adapts to the needs of the recommendation system that will be built.

In the destination dataset that is used for content based, first, combine several attributes, namely regency, description, and duration attributes. Aims to obtain item information in order to meet the needs of a recommender system using a content-based approach. After the process is carried out, it is continued with handling for the metadata column that has been obtained. The stages of handling carried out shown in Table 2.

E-ISSN : 2540 - 8984

JIPI (Jurnal Ilmiah Penelitian dan Pembelajaran Informatika)
Volume 07, Nomor 02, Juni 2022 : 408– 418

TABLE II
STAGES METADATA PREPROCESSING

Step Name	**Purpose**
Case Folding	To change all characters to lowercase with the aim that the handling of two words such as "Hello" and "hello" is not treated differently because they are the same word
Cleaning Text	To clean up text from links/urls, mentions, hashtags, emoticons, tabs, and numbers.
Remove Punctation	Remove punctuation in the text.
Remove Single Char	Eliminate words containing only one letter.
Tokenize	Split into words to make it easier to check.
Stopword	Word selection to remove words that appear frequently and do not provide information such as the words “yang”, “at”, and “from”.
Join Text	Put it back together into a complete sentence.

The rating dataset and user dataset are preprocessed to ensure that there are no null values. If a null value is found, the data will be dropped. In this study, the merger between the rating dataset and the user dataset was also carried out into a dataset. This dataset will be used in a recommendation system with a demographic filtering approach. The importance of this preprocessing data in order to get good data quality is because the quality of training data has an important role in the success of the algorithm or model being built [10].

C. Collaborative Filtering with Singular Value Decomposition (SVD)

Collaborative filtering is an approach to a recommendation system that makes recommendations based on the calculation of the similarity between users and the predicted value results [13]. Roughly speaking, the algorithms in collaborative filtering can be grouped into two parts, namely memory-based collaborative filtering and model-based collaborative filtering. In this research, we build collaborative filtering using the SVD algorithm. We take advantage of the surprise library. Surprise is a library in python that supports the use of the SVD algorithm. SVD is a factoring matrix by splitting a matrix into two unitary matrices and a diagonal matrix which has a value based on its factorization value [14]. The explanation of the algorithm can be seen in formula 1 [15].

$$R = U \sum V^T \tag{1}$$

In the above formula, the value of R is the user rating matrix, U and V are the results of solving the R matrix which is still one unit, and the sigma value is the diagonal matrix value of a single value. There are four stages that are passed in this collaborative filtering with singular value decomposition model:

1. Enter the user id and destination id for which you want to predict the rating.
2. By using the help of the surprise library, a model for training was developed. For example, the training data sample used is as shown in Table 3 below.

TABLE IIIII
SAMPLE DATA FOR TRAINING

User id	**Destination id**	**Rating**
3031	41	5
3109	21	5
1913	18	5

3. Call the SVD function from the surprise library.
4. Gives a return in the form of a predictive rating result from the input that has been entered previously.

D. Content-Based Filtering with Cosine Similarity

Content-based filtering is a recommendation system approach that tries to create a profile of users to predict ratings or ratings on items that have not been assigned. The recommendation results obtained depend on the previous item selection by the user, the items obtained are adjusted to the search for item similarities between items that have been assessed by the user and existing items [16]. Measuring the usefulness of content-based filtering is usually with a heuristic function such as the cosine similarity metric. In content-based filtering the rating and habits of the user are combined with information about the content available on the item.

By using the destination dataset that has been preprocessed, we use the TF-IDF calculation to calculate the similarity between all existing tourist destinations. After obtaining the results of the similarity, then proceed with building a user profile. This user profile is built by multiplying the matrix between the average rating of the user and the

E-ISSN : 2540 - 8984

JIPI (Jurnal Ilmiah Penelitian dan Pembelajaran Informatika)
Volume 07, Nomor 02, Juni 2022 : 408– 418

TF-IDF vector of each destination. Like the content-based approach in general, we calculate the cosine similarity between the vector user profile that has been built and the vector item profile TF-IDF to find the cosine similarity value. Prediction ratings can then be done after going through all the stages. In calculating the Prediction Rating, the formula as stated in formula 2 [17] is used.

$$P_{u,i} = (x - t)sim(b_u, a_i) + t \tag{2}$$

Where the value of x is the limit of the highest rating value that can be given by the user to a tourist destination, the value of t is the threshold value which is the distinguishing point of the rating considered to be liked or not, in this study we chose to use a threshold with a value of 4 and the value of $sim(b_u, a_i)$ is the calculation of cosine similarity between user u and the item i. Examples of calculations that occur in the Content Based Filtering model are as follows:

1. Receive input in the form of user id and destination for which you want to find the prediction rating
2. The user profile is created by multiplying the TF-IDF vector of each destination matrix with the rating weight matrix of users at rated destinations. To calculate the weight by using formula 3 below.

$$Weight = \frac{r}{rmax} \tag{3}$$

From the above formula, the value of r is the existing rating value, while the rmax value is the maximum rating value in the dataset. In the case of this study, the rmax value is 5. An example of the calculation is if the rating value is 4, the weight value is 0.8.

3. Calculation of cosine similarity between the vector user profile formed into a two-dimensional matrix with the existing tfidf destination matrix. The result will return the cosine similarity value for each destination id based on the user inputted. In table 4 we can see the value of the results generated by the calculation of cosine similarity.

TABLE IVV
VALUES SAMPLE COSINE SIMILARITY OF USER PROFILE WITH TF-IDF MATRIX

Destination id	Cosine Similarity
1	0.080207
2	0.245446
3	0.002649

4. The final calculation of the prediction rating using the formula 2 that has been described. The following is an example of a calculation based on table 3 data for destination id 2. The existing cosine similarity value is 0.245446. The highest five rating value is 5 and for the threshold value using a value of 4 so that the predicted rating results obtained are 4.245446. Of course, the value of the prediction rating results will be greatly influenced by the value of the user's cosine similarity which is inputted with the TF-IDF matrix.

E. Demographic Filtering with Support Vector Machine

Demographic filtering is a recommendation system method that produces recommendations by categorizing users with the demographic attributes they have in the data. Demographic filtering uses user attributes as demographic data to get recommendations, for example, giving recommendations based on age, gender, language, area of origin or things related to demographic attributes [18]. Support vector machine is a classification technique. This method is also considered one of the most powerful and accurate among other well-known approaches [19]. The main principle of SVM is to find the best hyperplane from several hyperplanes.

In this research, SVM development uses the help of the scikit-learn library. In the user dataset containing user demographics, the travel style, region, and country attributes are selected as the main reference for user demographic information. One hot encoding is applied to the selected attributes, then we build an SVM model where the rating becomes the class or target to be predicted. There are 4 stages that are passed in the demographic filtering that is built, namely:

1. Identify destinations that have not been rated by users.
2. For each destination that has not been rated, an SVM model will be built based on filtering the destination item data that you want to rate from the combined user and rating data set.
3. Each SVM model will adapt to hyperparameter optimization with the help of grid search CV.
4. After obtaining the best parameters and the model has been trained, the SVM model is applied to get the predicted destination value by the user. The prediction results from the rating are the results of the model that has been built.

E-ISSN : 2540 - 8984

JIPI (Jurnal Ilmiah Penelitian dan Pembelajaran Informatika)
Volume 07, Nomor 02, Juni 2022 : 408– 418

F. Hybrid Recommender System

The hybrid recommendation system is a method that is currently trending by combining two or more recommendation techniques to improve performance [7]. The main purpose of the merger is of course to get more results than when run separately. In this research, a combination of SVD (collaborative filtering), cosine similarity (content-based filtering), and SVM (demographic filtering) will be carried out. There are several techniques in hybridization to build a hybrid recommender system including: weighted, mixed, switching, feature combination, feature augmentation, cascade, and meta-level [20]. In this study we chose to use two hybridization techniques, namely weighted and switching. The selection of these two techniques is in accordance with previous research that has been done [7]. Combining these techniques can solve the cold start problem.

1. *Weighted Method*

   Weighted is a hybridization technique in the recommendation system that calculates the predictive value based on the results of all recommendation approaches used and considers it as a variable in a linear combination [17]. Combining the results of all existing recommendations and then calculating the weight or score of the recommended items to obtain results that have better accuracy and appropriate weights [20]. The formula used in combining the ratings obtained is as contained in the formula 4 [7]

   $$r_w = \alpha \cdot r_{DF} + \beta \cdot r_{CB} + \gamma \cdot r_{CF} \quad (4)$$

   Where the explanation of the above formula is, r_w is the rating prediction result obtained from a combination of formula calculations, r_{DF} is the rating prediction result from the DF (SVM) approach, r_{CB} is the rating prediction result from the CB approach, r_{CF} is the rating prediction result from the CF (SVD) approach, and for the values of $\alpha,\ \beta,$ and γ are values that represent the weights for each approach. In finding the best coefficient value, we do iteration with the selected coefficient value, namely [0.2,0.4,0.6,0.8]. After that, the best RMSE and MAE evaluation metrics are calculated. The selection of this coefficient value is considered based on the two evaluation metrics so that we look for the value that produces the best average between the resulting RMSE and MAE. The best coefficient values obtained are α = 0.4, β = 0.2, and γ = 0.4 which will later be used in the weighted model. The following is an example of a simple calculation on the weighted method. The values obtained from each method are for example 4 for demographic filtering, 4.5 for the predicted value of rating content-based filtering and a value of 4.8 for collaborative filtering. Based on this value, the prediction rating is calculated using formula 4 and the results obtained for the prediction rating in the following example are 4.42. The results of the weighted are very influential with the specified coefficient value.

2. *Double Hybridization Combine Switching and Weighted*

   Switching is a technique that selects the existing recommendation system based on existing conditions or circumstances so that it can choose the best recommendation system based on existing conditions [20]. To overcome the cold start problem and take advantage of the advantages of each recommendation system we use a combination of two hybridization techniques between weighted and switching.

   This selection is of course in accordance with previous research [7]. We try to replace some of the algorithms that have been selected with other algorithms that have better accuracy results. The switching technique that was built combines the results of the rating weighted hybrid method prediction, the prediction results of the content-based rating and the predicted results of the demographic filtering rating. Further explanation of the switching technique used can be seen in Figure 2.

E-ISSN : 2540 - 8984

JIPI (Jurnal Ilmiah Penelitian dan Pembelajaran Informatika)
Volume 07, Nomor 02, Juni 2022 : 408– 418

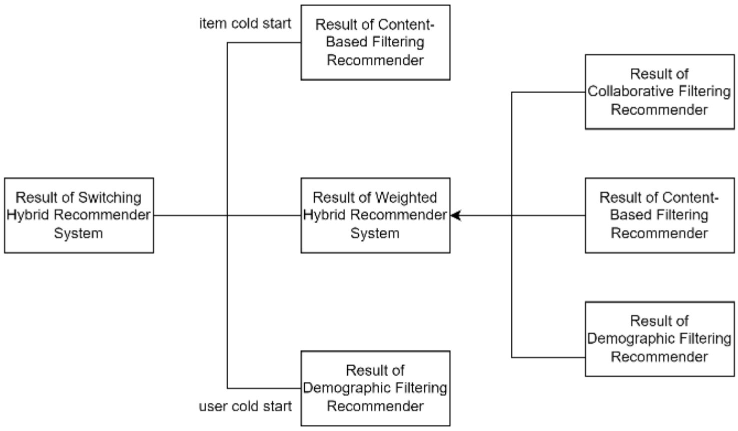

Fig. 2. The Switching Method

In Figure 2 above, it can be seen that the switching results are given based on the conditions faced by the system. When a cold start item occurs where the item is new and does not have a rating at all, a rating prediction result from content-based filtering will be given. If the condition faced is a cold start user where the user has not given a rating at all to the existing item, the rating prediction result from demographic filtering will be given. The last condition is when there is no cold start item and a cold start user, the system will provide rating prediction results based on the hybrid weighted method.

III. Result and Discussion

This study used a dataset that is the result of crawling from the TripAdvisor website. For an explanation of the attributes contained in the dataset, it can be seen in table 1. The dataset must of course go through the data preprocessing stage first so that it becomes data that is ready to be processed by the model to be used as explained in data preprocessing process. After going through the preprocessing stage, the data is divided into two parts, namely training data and testing data with the application of 10-fold cross-validation. Training data is data that will be a reference for the model to learn how to make predictions later, while testing data is data that is used as a test to predict the rating of the data.

At the testing stage, testing is carried out on testing data which is done blind to the existing rating. This blind rating is done to make it easier to compare the predicted rating results with the blind rating. Prediction of rating results obtained based on testing with 10-fold cross-validation on each model built, namely collaborative filtering (SVD), content-based filtering (cosine similarity), demographic filtering (SVM), and double hybridization combine switching and weighted models. To find out the difference in rating between the test data before being tested and after being tested, MAE and RMSE are used in the calculation of the comparison. Scenarios are run to find the difference in the results of each model that has been built. By calculating the MAE and RMSE values of each model, we can find out which method produces the best predictive results by considering the two metric evaluations. For each fold, 10-fold cross-validation was performed alternately 1,696 for test data and 15,255 for training data to be tested on each model. The results of each fold will certainly have a big impact on how the performance of the built model will be. The calculation of MAE and RMSE as a comparison between the testing data that has been tested with those that have not been tested will be carried out on each fold.

A. Test Result for Collaborative Filtering (SVD)

Tests are carried out on the collaborative filtering model using testing data whose ratings have been blinded. The rating from the testing data will later be predicted to test the prediction rating that can be done by the recommender system with collaborative filtering model. To find a comparison between the variables before tested and after tested, testing data that has not been blinded in the rating is used with the prediction results from testing data that already tested. To assist in finding the difference between the results before and after being tested, the MAE and RMSE evaluation metrics were used. The results of the MAE and RMSE evaluation metric values obtained from the model with the collaborative filtering (singular value decomposition) approach are based on evaluations with 10-fold cross-validation as follows.

E-ISSN : 2540 - 8984

JIPI (Jurnal Ilmiah Penelitian dan Pembelajaran Informatika)
Volume 07, Nomor 02, Juni 2022 : 408– 418

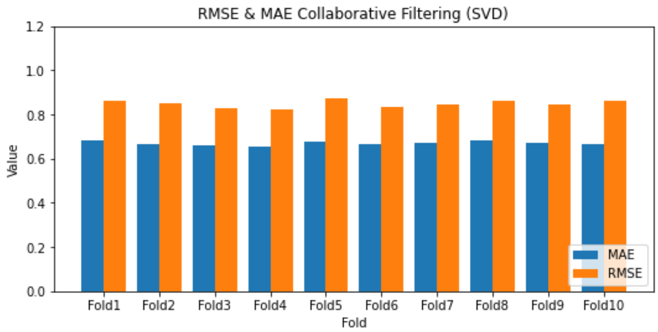

Fig. 3. Collaborative Filtering Test Results

In Figure 3 above, it can be seen for the MAE and RMSE values from model testing using the Collaborative Filtering (Singular Value Decomposition) approach. There are no significant changes between one-fold and another. In this model, the lowest MAE value is achieved in fold 4 with a value of 0.655 while the highest value is achieved in fold 1 with a value of 0.682. In this test, the RMSE value is higher than the MAE value obtained. The lowest RMSE value obtained in this test is the same as the fold obtained by the lowest MAE value, namely in fold 4 with a value of 0.823 while the RMSE value obtained in fold 5 with a value of 0.871. Based on the MAE and RMSE values that we can see, the comparison between the rating testing data before and after being tested for the recommender system with model collaborative filtering approach is best when run on fold 4, producing the best MAE and RMSE values. By paying attention to the results of MAE and RMSE on testing the resulting recommender system collaborative filtering model, information can be obtained about the difference between testing data before tested and testing data after tested which has a value that is not too far away. The difference between one test data and other test data is also not significant so that it can produce a good RMSE value. The results of this test show that Collaborative filtering is able to produce a stable value prediction rating for each fold so that the values in each fold are not too far from each other.

B. Test Result for Content-Based Filtering (Cosine Similarity)

The tests carried out on the content-based filtering model are the same as those carried out on the collaborative filtering model by using a rating of data testing that has been blind rated. In the comparison stage between rating testing data that has been tested and before being tested, testing data whose ratings have not been blinded are also used to make comparisons between rating testing data before being tested and those that have been tested using MAE and RMSE calculations. The results of the evaluation of the MAE and RMSE models obtained from the model with the Content-Based (Cosine Similarity) approach based on the evaluation with 10-fold cross-validation

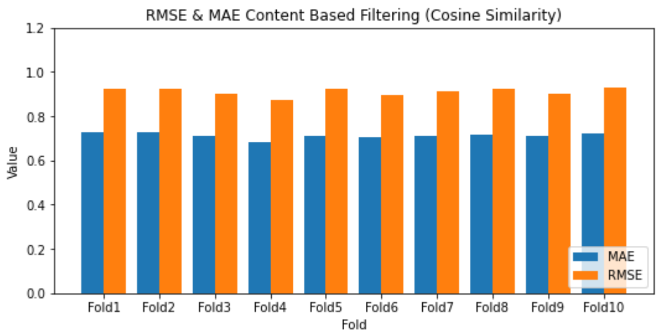

Fig. 4. Content-based filtering test results

In Figure 4 above, you can see the MAE and RMSE results from testing the recommendation system with content-based filtering (Cosine Similarity). The results obtained in each fold did not change significantly. The lowest MAE value was obtained in fold 4 with a value of 0.681 and the highest MAE value is found in fold 1 with a value of 0.7267. The RMSE value obtained in content-based filtering is much higher than the value obtained in the Collaborative Filtering test. The lowest result for the RMSE value obtained is in fold 4 which is also the fold the lowest MAE value with a value of 0.8736 while the highest RMSE value is obtained in fold 10 with a value of 0.9272. Based on the MAE and RMSE results, fold 4 is the best fold that can provide the lowest value in comparison of the results rating between the tested data and the testing data before being tested for the recommendation system using the content-based filtering (cosine similarity) approach. By looking at the results of MAE and RMSE, the comparison between the ratings of the test data results after being tested and before being tested on the content-based

E-ISSN : 2540 - 8984

JIPI (Jurnal Ilmiah Penelitian dan Pembelajaran Informatika)
Volume 07, Nomor 02, Juni 2022 : 408– 418

filtering model is further apart between the predicted rating and the existing rating. The results on this content-based filtering produce results that are not better when compared to the results on the collaborative filtering model.

C. Test Result for Demographic Filtering (SVM)

The tests carried out on the Demographic Filtering model are the same as in the previous model by using testing data that has been blind rated. Of course, this is to help find the difference between the rating of the testing data that has not been tested and the rating of the testing data that has been tested using MAE and RMSE calculations. The results of the evaluation of the MAE and RMSE models obtained from the recommendation system using the Demographic Filtering (Support Vector Machine) approach are based on evaluations with 10-fold cross validation as follows.

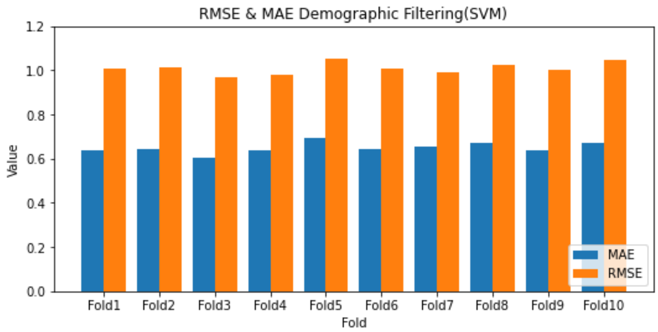

Fig. 5. Demographic filtering test results

In Figure 5 above, it can be seen for the MAE and RMSE values from the testing of the Demographic Filtering (Support Vector Machine) recommendation system model. In testing this model, the MAE and RMSE values in each fold are more diverse when compared to the two previous models that have been tested. The lowest MAE value was obtained in fold 3 with a value of 0.6035 while the highest MAE value was obtained in fold 5 with a value of 0.6932. This test shows that the results are quite far between the MAE and RMSE values in the Demographic Filtering (Support Vector Machine) model when compared to the other two models. The lowest RMSE value was obtained in fold 3 with a value of 0.97 while the highest RMSE value was obtained in fold 5 with a value of 1.052. By paying attention to the MAE and RMSE values generated in this test, it was found that fold 5 was the best fold for the demographic filtering model, while fold 3 was the worst fold when executed. If seen, the MAE generated in the demographic filtering model is quite good, while the RMSE is still relatively high because it is above 1. There are many values that have a high difference between the data before being tested and before being tested so that it gives a high RMSE value in the demographic filtering model.

D. Test Result for Double Hybridization Combine Switching and Weighted

The last test was carried out on the Double Hybridization model. This model is the result of a combination of the previously built models by adding weight and using switching. The test data used as test data in this model is the same as other models that have been tested by doing a blind rating first. In making a comparison between the ratings from the test data before being tested and being tested before being used, the MAE and RMSE calculations are also used. Calculation The results of the evaluation of the MAE and RMSE models obtained from the recommendation of the system with Multiple Hybridization based on the evaluation with 10 times cross-validation are as follows.

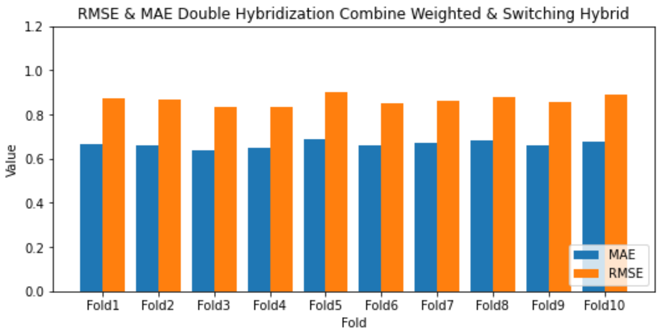

Fig. 6. Double hybridization combine switching and weighted test results

In Figure 6 above, it can be seen for the MAE and RMSE values from testing the Hybrid recommender system model with the Combine Switching Method. In testing this model, the resulting MAE and RMSE values are almost

E-ISSN : 2540 - 8984

JIPI (Jurnal Ilmiah Penelitian dan Pembelajaran Informatika)
Volume 07, Nomor 02, Juni 2022 : 408– 418

like the Collaborative Filtering and Demographic Filtering tests. This happened because a large proportion was chosen when the results of this recommendation system were combined. The lowest MAE value was obtained in fold 3 with a value of 0.6408 while the highest MAE value was obtained in fold 5 with a value of 0.6901. The lowest RMSE value was obtained in fold 3 with a value of 0.8352 while the highest RMSE value was obtained in fold 5 with a value of 0.8992. Based on the MAE and RMSE values generated in this model, information is obtained that the comparison between the existing ratings in the pre-test data and the pre-test data produces a good value because the results obtained in the two existing metric evaluations namely MAE and RMSE are considered good when compared to the model used. other. The selection and calculation of the weights between the coefficient values used in calculating the weights also has a big impact so that they can have more values.

E. Comparative Analysis of Recommended System Model Performance

Tests on each model have been successfully carried out, to find the best model performance, of course, an analysis of each model is needed. At this stage of analysis, the average calculation of the MAE and RMSE values produced by each model is carried out first. This step is taken to obtain an average comparison between the ratings on the testing data that has been tested and those that have not been tested. In a recommendation system, the rating prediction calculation is one way to determine the performance of the recommendation system model. Utilization of MAE and RMSE as evaluation metrics for analysis is used to be able to map clearly the comparison of the results of each model built in calculating the rating of testing data that has been tested with those that have not been tested.

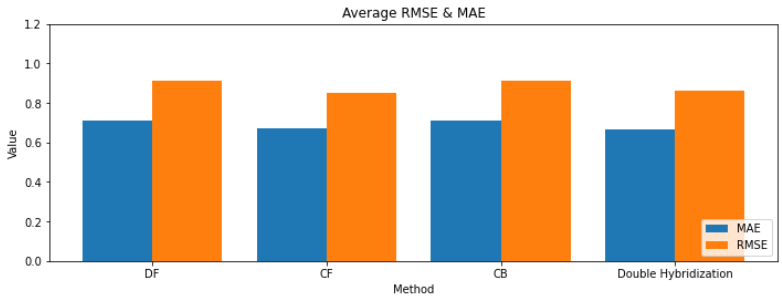

Fig. 7. Average MAE & RMSE

It can be seen in Figure 7 the average results of each model that was built. After the test, the average MAE and RMSE values were calculated for each model. The average results obtained to determine the performance of the recommendation system model that was built. The collaborative filtering model produces a good average value when compared to the other two models between demographic filtering and content-based filtering. The MAE and RMSE values generated in content-based filtering are the worst when compared to the other two models but produce a better RMSE value when compared to demographic filtering.

Focusing on MAE result, it is found that Demographic Filtering produces the lowest average MAE value but produces the largest RMSE value when compared to other models so that the demographic filter model considers it still not good. The value of the RMSE calculation plays an important role in determining the performance value because RMSE provides a larger penalty value when an error occurs in rating predictions between the predicted rating results and the actual rating. When the resulting rating is far from the actual value, RMSE will give a double penalty when compared to MAE. By considering these two aspects of metric evaluation, it was found that the average value of the double hybridization method resulted is a good MAE and RMSE value. The results obtained certainly get a considerable influence from the combination of demographic filtering (SVM) and collaborative filtering (SVD). The combination of methods between the built models plays an important role in providing better test results than when run separately.

IV. CONCLUSION

Based on the research that has been done, the experiment of the double hybridization method with switching and weighted techniques combining singular value decomposition and support vector machine in the recommendation system has been successfully carried out. Evaluation using MAE (mean absolute error) and RMSE (root mean square error) metrics found that the double hybridization method between singular value decomposition (collaborative filtering), cosine similarity (content-based filtering), and support vector machine (demographic filtering) provides accuracy higher prediction than when undergoing the recommendation method separately. Suggestions for future research need to be improvised on the content-based filtering approach by utilizing several algorithms that have better accuracy values. In addition, further research can carry out exploration related to the selected alpha, beta, and gamma values when running weighted on the hybrid method.

REFERENCES

[1] Z. K. A. Baizal, K. M. Lhaksmana, A. A. Rahmawati, M. Kirom, and Z. Mubarok, "Travel route scheduling based on user's preferences using simulated annealing," *Int. J. Electr. Comput. Eng.*, vol. 9, no. 2, pp. 1275–1287, 2019, doi: 10.11591/ijpeds.v9i2.pp1275-1287.

[2] Z. K. A. Baizal, D. Tarwidi, and B. Wijaya, "Tourism Destination Recommendation Using Ontology-based Conversational Recommender System," *Int. Jounal Comput. Digit. Syst.*, vol. 1, no. 1, 2021.

[3] L. Gang, "Personalized Recommendation of Tourist Attractions Based on Collaborative Filtering," in *2020 13th International Conference on Intelligent Computation Technology and Automation (ICICTA)*, 2020, pp. 144–147, doi: 10.1109/ICICTA51737.2020.00038.

[4] O. Alnogaithan, S. Algazlan, A. Aljuraiban, and A. A. Shargabi, "Tourism Recommendation System Based on User Reviews," in *2019 International Conference on Innovation and Intelligence for Informatics, Computing, and Technologies (3ICT)*, 2019, pp. 1–5, doi: 10.1109/3ICT.2019.8910312.

[5] Y. Wang, S. C. F. Chan, and G. Ngai, "Applicability of demographic recommender system to tourist attractions: A case study on TripAdvisor," *Proc. 2012 IEEE/WIC/ACM Int. Conf. Web Intell. Intell. Agent Technol. Work. WI-IAT 2012*, pp. 97–101, 2012, doi: 10.1109/WI-IAT.2012.133.

[6] A. N. Ngaffo, W. El Ayeb, and Z. Choukair, "A Bayesian Inference Based Hybrid Recommender System," *IEEE Access*, vol. 8, pp. 101682–101701, 2020, doi: 10.1109/ACCESS.2020.2998824.

[7] M. E. B. H. Kbaier, H. Masri, and S. Krichen, "A personalized hybrid tourism recommender system," *Proc. IEEE/ACS Int. Conf. Comput. Syst. Appl. AICCSA*, vol. 2017-Octob, pp. 244–250, 2018, doi: 10.1109/AICCSA.2017.12.

[8] M. L. Bai, R. Pamula, and P. K. Jain, "Tourist Recommender System using Hybrid Filtering," *2019 4th Int. Conf. Inf. Syst. Comput. Networks, ISCON 2019*, vol. 1, no. 1, pp. 746–749, 2019, doi: 10.1109/ISCON47742.2019.9036308.

[9] T. Anwar, V. Uma, and G. Srivastava, "Rec-CFSVD++: Implementing Recommendation System Using Collaborative Filtering and Singular Value Decomposition (SVD)++," *Int. J. Inf. Technol. Decis. Mak.*, vol. 20, no. 04, pp. 1075–1093, Apr. 2021, doi: 10.1142/S0219622021500310.

[10] N. Ifada, M. Sophan, I. Syachrudin, and S. Sugiharto, *An Efficient Scheme to Combine the User Demographics and Item Attribute for Solving Data Sparsity and Cold-start Problems*. 2019.

[11] Riska Chairunisa, Adiwijaya, and Widi Astuti, "Perbandingan CART dan Random Forest untuk Deteksi Kanker berbasis Klasifikasi Data Microarray," *J. RESTI (Rekayasa Sist. dan Teknol. Informasi)*, vol. 4, no. 5, pp. 805–812, 2020, doi: 10.29207/resti.v4i5.2083.

[12] M. D. Purbolaksono, M. Irvan Tantowi, A. Imam Hidayat, and A. Adiwijaya, "Perbandingan Support Vector Machine dan Modified Balanced Random Forest dalam Deteksi Pasien Penyakit Diabetes," *J. RESTI (Rekayasa Sist. dan Teknol. Informasi)*, vol. 5, no. 2, pp. 393–399, 2021, doi: 10.29207/resti.v5i2.3008.

[13] G. Geetha, M. Safa, C. Fancy, and D. Saranya, "A Hybrid Approach using Collaborative filtering and Content based Filtering for Recommender System," *J. Phys. Conf. Ser.*, vol. 1000, no. 1, 2018, doi: 10.1088/1742-6596/1000/1/012101.

[14] E. Gutflaish, A. Kontorovich, S. Sabato, O. Biller, and O. Sofer, "Temporal anomaly detection: Calibrating the surprise," *33rd AAAI Conf. Artif. Intell. AAAI 2019, 31st Innov. Appl. Artif. Intell. Conf. IAAI 2019 9th AAAI Symp. Educ. Adv. Artif. Intell. EAAI 2019*, pp. 3755–3762, 2019, doi: 10.1609/aaai.v33i01.33013755.

[15] T. Ayu, W. Tyas, Z. K. A. Baizal, and R. Dharayani, "Tourist Places Recommender System Using Cosine Similarity and Singular Value Decomposition Methods," vol. 5, pp. 1201–1207, 2021, doi: 10.30865/mib.v5i4.3151.

[16] Q. R. Arvianti, Z. K. A. Baizal, and D. Tarwidi, "Tourism Recommender System Using Item-Based Hybrid Clustering Method (Case Study : Bandung Raya Region)," *J. Data Sci. Its Appl.*, vol. 2, no. 2, pp. 95–101, 2019, doi: 10.34818/JDSA.2019.2.35.

[17] S. Suriati, M. Dwiastuti, and T. Tulus, "Weighted hybrid technique for recommender system," *J. Phys. Conf. Ser.*, vol. 930, no. 1, 2017, doi: 10.1088/1742-6596/930/1/012050.

[18] M. Al-Shamri, "User Profiling Approaches for Demographic Recommender Systems," *Knowledge-Based Syst.*, vol. 100, Mar. 2016, doi: 10.1016/j.knosys.2016.03.006.

[19] A. Lawi and F. Aziz, "Classification of credit card default clients using LS-SVM ensemble," *Proc. 3rd Int. Conf. Informatics Comput. ICIC 2018*, pp. 1–4, 2018, doi: 10.1109/IAC.2018.8780427.

[20] Z. Fayyaz, M. Ebrahimian, D. Nawara, A. Ibrahim, and R. Kashef, "Recommendation systems: Algorithms, challenges, metrics, and business opportunities," *Appl. Sci.*, vol. 10, no. 21, pp. 1–20, 2020, doi: 10.3390/app10217748.