PalArch's Journal of Archaeology of Egypt / Egyptology

EVALUATION OF HYBRID RECOMMENDER TECHNIQUES ON MOVIELENS DATASET

Tengku Azfar[1] *Su-Cheng Haw*[2]

[1,2] Faculty of Computing & Informatics, Multimedia University

Jalan Multimedia, Cyberjaya

E-mail: [2]sucheng@mmu.edu.my

Tengku Azfar And Su-Cheng Haw. Evaluation Of Hybrid Recommender Techniques On Movielens Dataset-- Palarch's Journal Of Archaeology Of Egypt/Egyptology 17(10), 890-902 ISSN 1567-214x

Keywords: Recommender System, Recommender Technique, Cosine Similarity, Hybrid Technique, Evaluation.

ABSTRACT

Recommending items can be very helpful and make things easier within minutes. This paper consists of the introductory and explanation on the techniques of hybrid recommendation system. Subsequently, we propose three hybrid combinations, which are (i) Cosine Similarity with k-Nearest Neighbors (KNN), (ii) Term Frequency-Inverse Document Frequency (TF-IDF) with Singular Value Decomposition (SVD) matrix factorization, and (iii) k-Means clustering with Jaccard similarity. Experimental evaluations are then carried out via accuracy measure and data visualization on the MovieLens dataset to determine which combination yields better accuracy.

INTRODUCTION

Recommendation systems are widely used by the big tech companies such as Google, Amazon, Spotify, Netflix, and many others with the purpose of accomplishing the recommendation of their items to the users. Recommender systems consist of algorithms that can be used to suggest the similar and related items to the user by sorting the data from the information base. The recommendation systems provide multiple ways to suggest items such that discovering data patterns in the dataset by understanding the user's searches and choices that will then process the related items based on their interests and needs.

With the help of the Internet, watching and download movies online have been easier to access by users worldwide. With tons and countless number of movies available on the internet, it makes a bit more difficult to explore new movies that suits our taste and preference. That is where movie recommender system comes in. However, there are many recommender systems that frequently produce unsatisfactory recommendations due to the dependent on multitude factors.

There are two main groups of recommender systems, namely collaborative filtering and content-based filtering. The content-based filtering utilizes discrete and specific characteristics of an item so that it could suggest additional items with similar properties (characteristic information). However, the collaborative filtering approach concentrate on creating a model based on a user's history behaviour such as items previously purchased or based on given item ratings (user-item interactions).

Contrast to content-based filtering system, the collaborative filtering is based on the assumption algorithms that connect the users with similar preferences and interest by modeling their similarities between the related profiles to make suggestions. This group of users are called neighborhood. Item-based and user-based nearest neighbor similarities can collaborate to improve the suggestions accuracy and to overcome the cold start problem. The hybrid approach combines both content-based and collaborative filtering aiming to avoid problems that are generated when implementing one of the approaches.

As such, this research aims to evaluate and discover new methods to produce a hybrid recommender system that can overcome the problems of cold-start, sparsity and other issues of the conventional recommender techniques.

Generally, this paper explains on the insights on recommender system, the advantages and disadvantages of implementing collaborative filtering or content-based filtering, past related works on recommendation techniques, and also the implementation of the hybrid algorithms. The hybrid algorithms will then be evaluated using the evaluation metrics of Root Mean Square Error (RMSE), Mean Absolute Eror (MAE), and Mean Square Error (MSE) to determine the accuracy and errors of the three proposed hybrid algorithms.

LITERATURE REVIEW

A recommendation (or recommender) system filters the gathered data using different algorithms and suggests the most relevant items to users. Explicit user ratings after listening to a song, reading an e-book or watching a movie, queries from implicit search engine and purchase histories, or even from other source of knowledge about the specific users or items; support the data required for the recommender systems to operate efficiently.

Giant companies such as Netflix, YouTube, Spotify, Amazon, Facebook, and many others have implemented recommender systems on their websites to help boost sales and improve their customer satisfaction. For istnace, recommender system for online advertisement helps suggest the appropriate contents that match the user preferences. An e-commerce website like Amazon, uses this recommender system to help suggest items or products that could interest the users.

Besides engaging the users to the items and contents with personalized recommendations, the recommender system provide reports to the companies; giving them accurate and up to date reporting allows them to make solid decisions on the websites and the direction of a campaign.

Recommendations able to help companies gain and retain customers by sending out emails and notifications with links to offer new items or suggest movies that might meet the users' interests. By understanding the user wants and needs, the company gains a competitive advantage and decline the threat of losing a customer to a competitor [1].

Bonjori and Moradi [2] implement k-Nearest Neighbor (KNN) and Bayesian method to come up with an experiment suggesting that such a combination can obtain high precision values to recommend movies to the users. Their purpose is to produce a personal recommender system that can learn users' preferences and recommend list of items that are suitable to them automatically. Pearson correlation coefficient is the social recommender used as the item-based collaborative filtering to determine the similarity between two items using preference data of all users. Decision templates are implemented to combine the base recommenders for classification and regression with the bayesian nearest neighbour to predict the preferences of the user.

Amini et al. [3] suggested a hybrid method to recommend movies have combined clustering method with four classification models. This includes Multi-Layer Perceptron Neural Network (MLP), Naïve Bayes, Decision Tree, and Spiking Neural Network (SNN). They have tabulated their result based on the precision, recall, accuracy and classification error and found that among the four classification models, Naïve Bayes has the highest accuracy result and lowest error.

Filipa et al. [4] proposed a recommender system that will integrate unrated movie reviews with movie ratings on the Internet. Also, sentiment analysis was created to make analysis on user preferences where movie reviews did not directly associated with the explicit rating. He proposed a collaborative filtering method, Singular Value Decomposition (SVD) matrix factorization that was tested on the IMDb dataset where it contained 53112 reviews and 50% unrated movies.

Wakil et al. [5] also integrated matrix factorization with nearest neighbors to conduct the experiment on the recommender system. Their target is to build a recommender system that is based on emotions with the use of MovieLens dataset.

Combination of data clustering and computational intelligence were experimented by Katarya and Verma [6], where they leverage k-means clustering with implementing cuckoo search optimization on the MovieLens dataset. They have calculated their results using MAE, RMSE, Standard

Deviation, and t-value for the recommender system.

Geetha et al. [7] proposed a hybrid recommender by combining Pearson Correlation Coefficient and k-Means clustering on the movie dataset. Their proposed approach resulted to overcome the drawbacks of each algorithm and help to improve the overall performance of the system. Their evaluation contains the details of the users such as the age, occupation, directors, actors, and many others.

Kalimuthu and Vellaichamy [9] have taken a slightly different approach of developing a recommender system; that is to combine Fuzzy C Means (FCM) clustering with the help of Bat algorithm together. Fuzzy C Means is used to cluster the users in similar groups, whereas, Bat algorithm is used to acquire the initial position of the clusters for high performance recommendation quality. The experiment made was compared to MAE, recall and precision.

Yucebas [9] had made an analysis and proposed a hybrid recommender system that combines multi-layer Artificial Neural Network (ANN) with k-Means and x-Means clustering. He had grouped the movie genres in clusters and make some result tabulation that involve clusters by training cycle, learning rate, momentum, accuracy, recall, and precision of the recommendation.
Table 1 summarizes the techniques reviewed earlier.

Table 1 Best results

References	Hybrid method	Advantages	Disadvantages
Bonjori and Moradi [2]	Bayesian + K Nearest Neighbours	Obtained high precision values.	Not recommended for large datasets.
Amini et al. [3]	Clustering + Naïve Bayes	Naïve Bayes gives high accuracy.	Implementation can be confusing.
Filipa [4] et al.	Distance Similarity + SVD	Overcome sparsity.	Inaccuracy of recommendation due to unlabelled data.
Wakil et al. [5]	Matrix Factorization + Nearest Neighbours	New method of recommending movies to users.	Detecting emotions could lead to false suggestion.
Katarya and Verma [6]	K-Means Clustering + Cuckoo search optimization	Obtained reasonable RMSE and MAE results.	Low efficiency.
Geetha et al. [7]	Pearson Correlation + K-Means	Produce a high correlation score.	Less accurate in producing recommendations.
Kalimuthu and	Fuzzy C Means Clustering	Overcome the scalability	Bat optimization might not produce

Vellaichamy [8]	(FCM) + Bat optimization	problem.	accurate results at times.
Yucebas [9]	Artificial Neural Network + Clustering	Able to tackle cold-start problem.	Moderate performance values.

Many variations of hybrid systems have been used to improve the base recommender systems that contain many drawbacks, depending on the algorithm implemented. Each of them has their dedicated algorithms and techniques to tackle the problems of the conventional recommender systems. The newly proposed hybrid methods will then be reviewed and elaborated in Section 3.

PROPOSED APPROACH

Overview on the hybrid combination selection

The hybrid system is a mixture of different recommendation techniques to achieve higher accuracy performance and to prevent some drawbacks of the conventional recommendation systems. By combining the techniques, this will provide effective and more accurate recommendations compared to a single algorithm because this can control the shortcoming of an individual method in the hybrid model. These are some of many ways to combine various approaches in the hybrid system:

- Utilize content-based filtering in collaborative approach or vice versa.
- Separate the algorithms and combine the results.
- Develop a unified system which combines both techniques.

In this research, we have selected three hybrid combinations for evaluation purpose. These hybrid systems are as follows:

- H1: Cosine similarity and KNN
- H2: Term Frequency-Inverse Document Frequency (TF-IDF) and SVD Matrix Factorization
- H3: K-Means Clustering and Jaccard similarity

For the first hybrid algorithm, the cosine similarity and KNN are chosen because they are widely used by other researches to produce recommendations. The cosine similarity is used to calculate the rating vectors and the KNN is to find groups of similar users based on common movie ratings and generate suggestions using the average rating. This hybrid algorithm is used as a benchmark as it is popularly used by other researches to make this evaluation.

The second hybrid algorithm, TF-IDF is known to have a high accuracy in information retrieval by calculating the similarity between movies where it converts tags and titles of a movie, and convert the unstructured text into a vector structure. The combination of TF-IDF vector structure with the matrix factorization SVD can be implemented easily as they are both in their vector form. This will generate a more accurate result with SVD having the tool to find some latent features within the TF-IDF, regarding the interactions between the movies and the users. Matrix factorization was used for the Netflix Prize contest algorithm to produce a high accuracy recommendation of movies. Combining these two algorithms may provide a higher accuracy and reduce some of their drawbacks.

The third hybrid combination of k-means clustering with Jaccard similarity is to determine the result when a clustering method is combined with a collaborative similarity measure. For recommending items or movies to users, the movies will be pre-processed into categories of movie genres where the Jaccard distance will help to determine the categories of movies available using its binary similarities. By having the categorized movies specified, selected movie genres will be used to make movie recommendations based on the same genres listed. K-means clustering can help to determine the groups of movies that have similar genres. Therefore, Jaccard similarity can help clustering method to categorize the genres easier with its tool.

Building the evaluation engine

Figure 1 shows the overview process of building the hybrid engine.

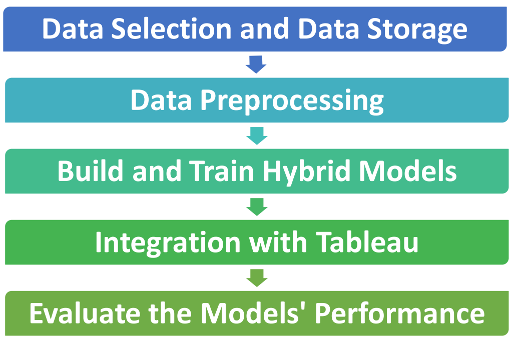

Figure 1 Hybrid Evaluation Engine

Dataset selection

The dataset used for developing this prototype is called the MovieLens dataset. This dataset can be obtained from the GroupLens website itself. As stated earlier, over 100,000 amounts of data will be used for this prototype to determine the accuracy of the hybrid algorithms. There are also other datasets that contain more data such as tag genome data, more number of users, ratings, tags and movies. All of these depend on how we want to use the dataset for. GroupLens has categorized some dataset for suggestions of usage, whether it is for new research or for education and development.

The dataset that is used for this recommender system is categorized under the education and development category of usage. The data that have been provided in the selected dataset include: ratings, user ID, movie ID, movie tags, movie title, movie genre, and the timestamp. These data resources have the sufficient attributes to find the similarities between the users and the movies. In Python and Tableau, the dataset is linked to the Pandas data frame also it is stored in the SQLite for future use.

Data preprocessing

For the preprocess of data, there are different methods of preprocessing data on different hybrid models as they require specific alteration of data for certain algorithms to work efficiently. For example, the third hybrid algorithm that contains K-Means clustering and Jaccard similarity implement the separation of movie genres and selected only some of the genres for evaluation.

Before the implementation of the hybrid algorithms, the dataset are filled the null values with average overall ratings so that the algorithms can compute smoothly. Also, unrequired data like the links and timestamp were removed as they will not be used in the recommender system (see Figure 2).

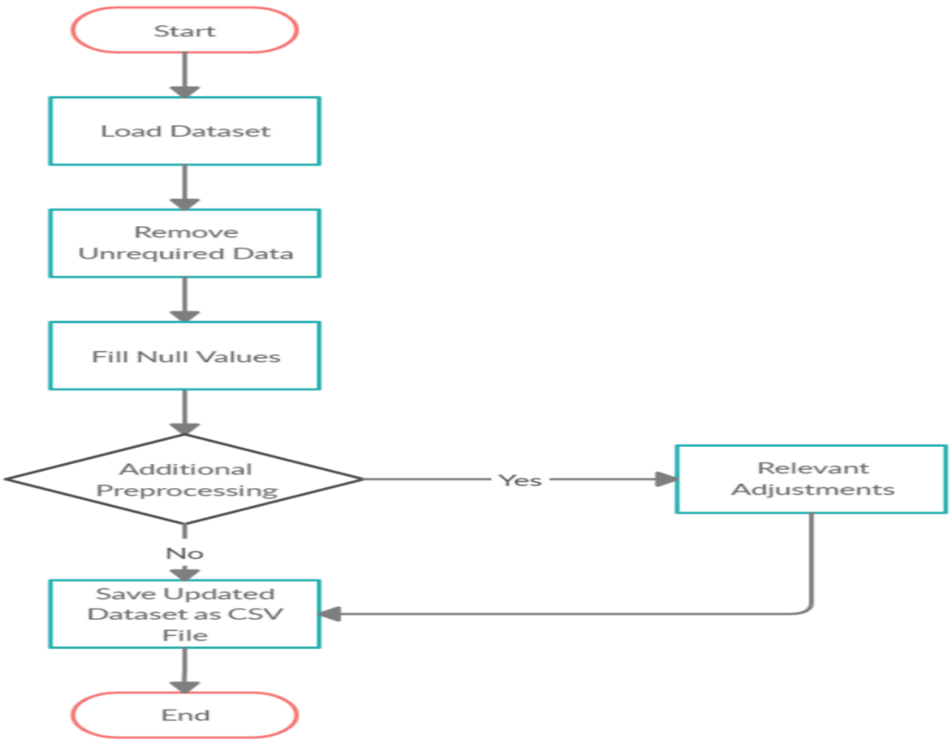

Figure 2 Flow chart of Data Preprocessing

Implementation of the hybrid model

To implement the hybrid recommender system, the programming language Python version 3.7.4 is used to assemble and mix the algorithms to put them into work.

The algorithms proposed are link to cosine similarity, matrix factorization, KNN, Jaccard similarity, K-Means Clustering and TF-IDF. These are the selected algorithms is because cosine similarity can determine the similarities with less complexity. Nearest neighbors can make predictions based on the similarities of the related users that rated the similar movies. Matrix factorization is proposed because it has the advantage to decompose the user and movies interaction matrix. Next, Jaccard similarity comes in handy when the vectors contain binary values. Clustering has the advantage of grouping users or movies based on their similarities more efficiently. Lastly, the TF-IDF is efficient to use when determining the keyword in a title of a movie, genres and the tags used.

In summary, the hybrid combinations are as follows:

1.H1: Cosine similarity and KNN
2.H2: TF-IDF and SVD Matrix Factorization
3.H3: K-Means Clustering and Jaccard similarity

Tableau is used as the user interface dashboard for this project where the user can navigate through it to visualize the MovieLens dataset briefly. Then, the users are able to view the three separate hybrid algorithms in the dashboard where they individually show the RMSE, MAE, and MSE results.

The recommender system require the user to select a movie that is on the list provided and it will recommend the top 5 movies that is similar to the movie picked based on the hybrid algorithm chosen. The conclusion section in the dashboard will show the overall evaluation metrics for the three hybrid algorithms.

Figure 3 shows the dashboard for data visualization. The Dataset section is the overview of the MovieLens dataset that is used for the recommender system. Mostly are the descriptions and data visualization of the dataset generally. The word cloud represents the genres available and the rate of use in movies listed in the dataset. The chart however, depicts the number of movies in genres categories by year.

Figure 4 shows the analysis on the density histogram of movie ratings grouped by genres. Also, the right chart above shows the average movie ratings by year given by the users in the dataset.

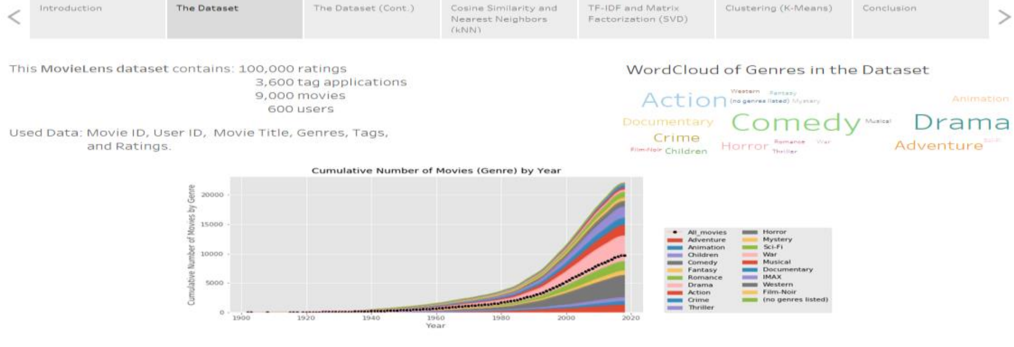

Figure 3 Tableau Interface of the Dataset Overview

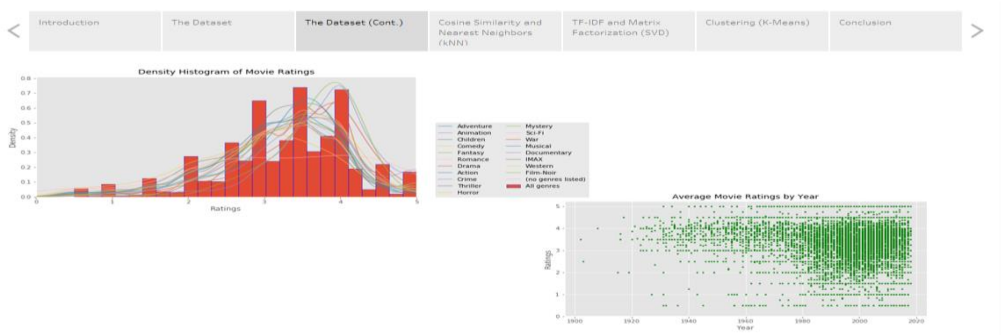

Figure 4 Dataset Overview Continuations

Figure 5 illustrates the dashboard for the first hybrid recommendation. The dashboard contains the results for RMSE, MAE, and MSE to determine the accuracy of the hybrid algorithms. Below it represents the linear regression of the predictions comparing the actual ratings and the predicted ratings.

The example of Similar Users' Ratings table shows the data that is in Python integration to process the similarities between users. Also, in this dashboard a list of movies based on the previously user's pick of the movie will be recommended.

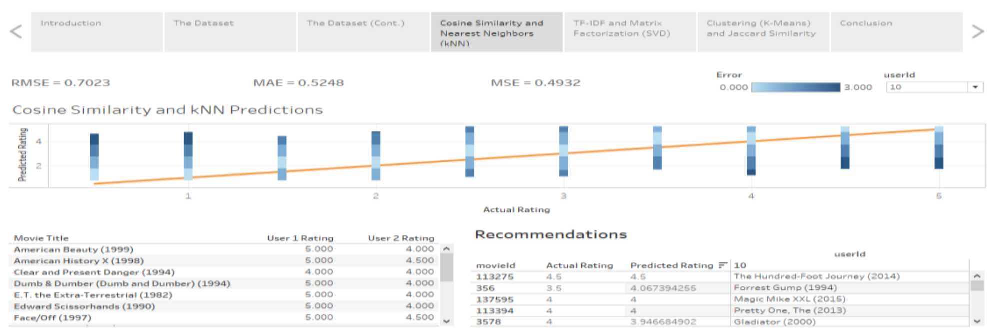

Figure 5 Cosines and KNN Hybrid Dashboard

Figure 6 represent the second hybrid part of the dashboard containing TF-IDF and SVD matrix factorization. Similar to the first hybrid dashboard, it has the RMSE value, MAE value, and the MSE value. The linear regression chart shows the accuracy of the predictions based on the actual ratings and predicted ratings. The error hue colour bar represents the error presentation in the chart. The darker the colour, the more error it has.

The table at the bottom left of the dashboard shows the data and the metadata containing the movie ID, movie title, and the genres and tags combined together for determining the frequency later on in Python.

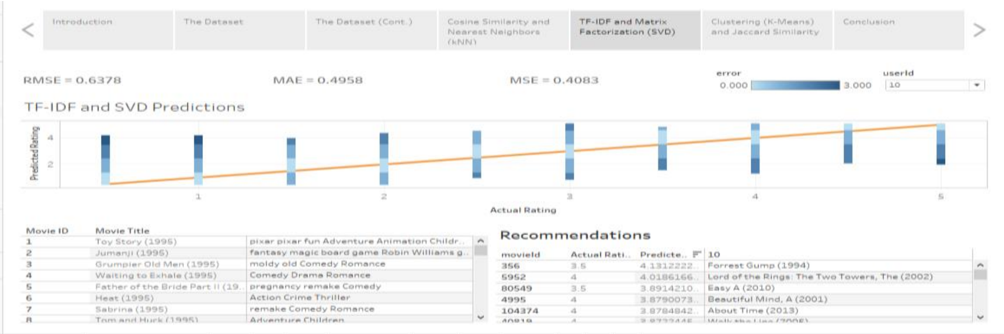

Figure 6 TF-IDF and SVD Hybrid Dashboard

Figure 7 shows the dashboard for the third hybrid algorithm containing as similar as the other two hybrid algorithms. The RMSE, MAE, and MSE values are depicted in the dashboard as well. The linear regression chart may look a bit different from the other two algorithms because the ratings are clustered into groups based on the movie genres. However, the error contain in this algorithm has only the range up to 2.0 instead of 3.0 like the other algorithms.
At the bottom left of the dashboard is the example presentation of the clustering of movie genres. It has been separated into three genres: drama, comedy and action. The genres are indicated as different colors in the chart while at the same time, we can see the similarities between movies that have the same genres in groups.

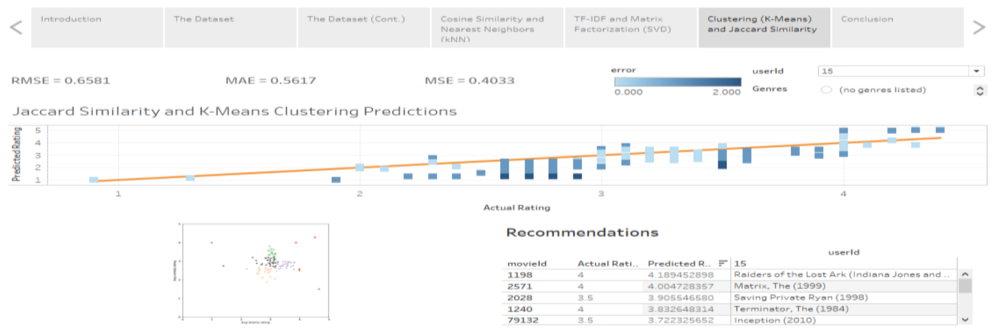

Figure 7 K-Means Clustering and Jaccard Similarity Hybrid Dashboard.

The conclusion dashboard in Figure 8 shows the evaluation metrics used to evaluate the hybrid algorithms. As stated earlier, the RMSE, MAE, and MSE were used and here is the comparisons among the three hybrid algorithms. On the left side of the charts, is the hybrid H1 (Cosine similarity with KNN), the second hybrid H2 (TF-IDF with SVD), and the third hybrid H3 (clustering with Jaccard similarity). Also, there are some short descriptions on the algorithms' advantages and disadvantages.

Figure 8 Conclusion Dashboard

Experimental evaluations

Figure 9 shows the three hybrid algorithms ranging from the left is the first hybrid (Cosine similarity with KNN nearest neighbours), the second is the TF-IDF with SVD matrix factorization, and the third is k-Means clustering with Jaccard similarity. The RMSE calculation can be achieved by implementing the script accuracy rmse (function) using the surprise library.

RMSE or Root Mean Square Error is used to predict errors by calculating the standard deviation of the data. It is the measurement of the regression line distance between the data points. It will measure the spread between the data points around the line of best fit. To determine the best value we will have to look at the lower value, the lower the value, higher the accuracy. Based on the figure above, hybrid two which is the TF-IDF with SVD has the lowest RMSE value compared to the first hybrid, the Cosine similarity with KNN nearest neighbors.

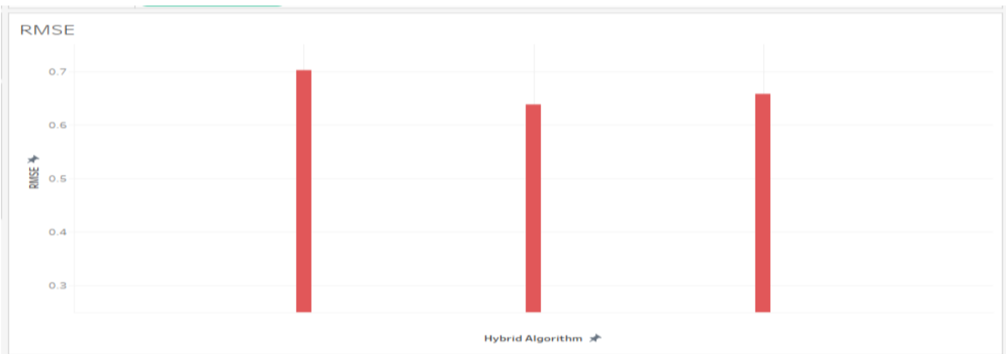

Figure 9 RMSE Evaluation Results.

The arrangement of the hybrid algorithms is similar to the RMSE evaluation chart where the H1, H2 and H3 are situated accordingly. The MAE calculation can be obtained by implementing the script accuracy mae (function) using the surprise library on Python easily.

MAE is the calculation of the average of all absolute errors contain in the hybrid algorithm. From Figure 10, we can say that H2 still remains the most accurate among the other two hybrid algorithms in terms of having the lowest MAE value. However, this time H3 acquire the highest MAE value than the H1 and H2.

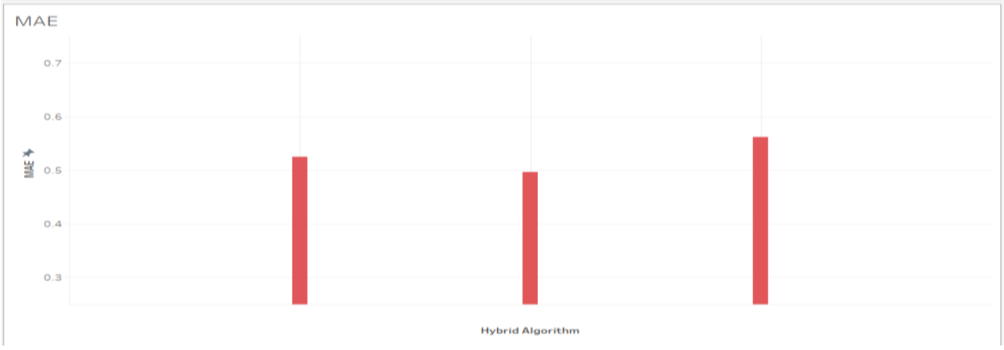

Figure 10 MAE Evaluation Results.

MSE is quite similar to RMSE where it is to determine the closeness of a regression line between the data points. It operates by calculating the distances to the regression line from the data points and squaring the value. It is needed to square the value to remove the unrequired negative values. This method can produce larger differences from gaining more weight on the values.

MSE values are different compared to the other two evaluation metrics as this method ranging from 0.4 to 0.5 amounts of errors. H1 has the highest value while H3 has the lowest value of MSE compared to the others (see Figure 11).

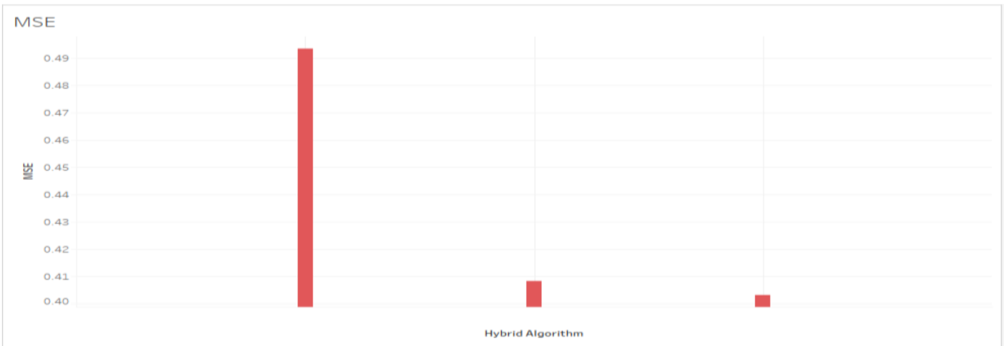

Figure 11 MSE Evaluation Results

CONCLUSION AND FUTURE WORKS

Three hybrid combinations of Cosine Similarity with KNN nearest neighbors, TF-IDF with SVD matrix factorization and k-Means clustering with Jaccard similarity have been proposed. Experimental evaluations have been carried out to find the method which produce high accuracy of the recommendations. The TF-IDF and SVD matrix factorization results have achieved satisfied recommendations.

As for our future work, we would like to experiment hybrid algorithms combining TF-IDF with clustering and SVD matrix factorization with cosine similarity as they have high accuracy in the evaluation results. Moreover, working on larger datasets of other domains could be an improvement in this evaluation as they could achieve a different view of the recommendation. This could include the other movie dataset, flight recommendations, music recommendations, and many others.

REFERENCES

R. Nagamanjula And A. Pethalakshmi, "A Novel Scheme For Movie Recommendation System Using User Similarity And Opinion Mining", International Journal Of Innovative Technology And Exploring Engineering, 8(4s2), Pp. 316-322, 2019.

E. Bojnordi And P. Moradi, "A Novel Collaborative Filtering Model Based On Combination Of Correlation Method With Matrix Completion Technique", In Proc. Of International Symposium On Artificial Intelligence And Signal Processing, 2012, Pp. 191-194.

M. Amini, M. Nasiri, M. Afzali, "Proposing A New Hybrid Approach In Movie Recommender System", International Journal Of Computer Science And Information Security, 12 (8), Pp. 40-45, 2014.

P. Filipa, P. Dias, F.J. Magalhaes, "A Recommender System For The Tv On The Web: Integrating Unrated Reviews And Movie Ratings", Multimedia Systems, 19(6), Pp. 543-558, 2013.

K. Wakil, R. Bakhtyar, And K. Ali, "Improving Web Movie Recommender System Based On Emotions", International Journal Of Advanced Computer Science And Applications, 6(2), 2015.

R. Katarya And O.P. Verma, "An Effective Collaborative Movie Recommender System With Cuckoo Search", Egyption Informatics Journal, 18(2), Pp. 105-112, 2016.

G. Geetha, M. Safa, C. Fancy, D. Saranya, "A Hybrid Approach Using Collaborative Filtering And Content Based Filtering For Recommender System", Journal Of Physics: Conf. Series 1000, Pp. 1-7, 2018.

V. Vellaichamy And V. Kalimuthu, "Hybrid Collaborative Movie Recommender System Using Clustering And Bat Optimization", International Journal Of Intelligent Engineering & Systems, 10(5), Pp. 38-47, 2017.

S.C Yucebas, "Movieann: A Hybrid Approach To Movie Recommender Systems Using Multi-Layer Artificial Neural Networks", Pp. 214-232, 2019.