Accredited Ranking SINTA 2
Decree of the Director General of Higher Education, Research, and Technology, No. 158/E/KPT/2021
Validity period from Volume 5 Number 2 of 2021 to Volume 10 Number 1 of 2026

Published online on: **http://jurnal.iaii.or.id**

JURNAL RESTI
(Rekayasa Sistem dan Teknologi Informasi)
Vol. 6 No. 5 (2022) 702 - 709 ISSN Media Electronic: 2580-0760

Memory-based Collaborative Filtering on Twitter Using Support Vector Machine Classification

Anang Furkon Rifai[1], Erwin Budi Setiawan[2]
[1,2]Informatics, School of Computing, Telkom University
[1]anangrifai@student.telkomuniversity.ac.id, [2]erwinbudisetiawan@telkomuniversity.ac.id

Abstract

Nowadays, watching films at home is one of people's entertainment. Netflix is a service provider for watching films and provides many types of film genres. However, of the many films available, it makes users confused to choose which film to watch first. The solution to the problem is a system that provides recommendations for the best films to watch based on user ratings. Twitter is still people's favorite social media to express their feelings, thoughts, and criticisms. In this system, tweets serve as input data that will be processed into data with rating values. This research implemented a recommendation system based on user ratings from tweets using collaborative filtering combined with Support Vector Machine (SVM) classification and implemented it on user-based and item-based. The test results in this study show that Collaborative Filtering gets the best RMSE value results on item-based 0.5911 and 0.8162 on user-based. The Support Vector Machine (SVM) classification algorithm using hyperparameter tuning produces item-based values with a precision of 85.03% and recall of 90.71%, while user-based values with a precision of 87.75% and recall of 88.95%.

Keywords: *Recommender System, User-based, Item-based, Collaborative Filtering, Support Vector Machine.*

1. Introduction

The world of technology and information is growing rapidly. It can be seen in the daily lives of many people today who use technology for information or communication media, not least in the field of music or film. The film is entertainment that most people will know. Films have various genres and titles to watch. People today can watch movies not only through cinemas but also through digital platforms such as Netflix.

Netflix is a company that was founded in 1997. Netflix is a company that offers online subscription rental for movies and television series [1]. Subscribed users can watch movies on Netflix anytime and anywhere. There are many films available for subscription. However, the many films on Netflix will confuse users when choosing what film to watch first. The solution to this problem can be solved with a recommendation system for films on Netflix that can provide the best film recommendations for users.

The increasing number of people who access social media is one of the reasons for the rapid growth of the technology world. A popular social media platform is Twitter, which provides features that users use to express feelings, ideas, and thoughts [2]. Twitter, since 2006 has shown rapid growth, sending 250 million tweets daily [3]. Twitter is a valuable source for research because of the amount of new and relevant data [4]. For example, many people review Netflix movies on Twitter.

In recommender systems, various techniques include collaborative filtering, content-based, knowledge-based, and association rules-based recommendation. Among the available methods, the most successful approach is collaborative filtering, which has two categories: memory-based and model-based [5]. Memory-based is an approach using similarity patterns between users, often called user-based, and between services from historical data, called item-based [6]. However, collaborative filtering has two fundamental problems, data sparsity and scalability. Data sparsity occurs when R misses a lot of data [7]. From these problems, additional algorithms are needed to minimize if there is high sparsity data in the film recommendation system. Support Vector Machine (SVM) is a classifier algorithm that can handle high sparsity data [7]. And the SVM model can solve the conversion of classification problems on two sides of the user, namely items and users [8].

Accepted: 08-07-2022 | Received in revised: 31-07-2022 | Published: 01-10-2022

In previous studies in overcoming data sparsity and scalability, researchers combined collaborative filtering with Transductive Support Vector Machine (TSVM) based on Active Learning (AL) and SVMCF4R [5], [6]. Therefore, in this study, the authors will combine collaborative filtering techniques with the Support Vector Machine (SVM) classification method using hyperparameter tuning with grid search to improve performance. To the author's knowledge, no research uses hyperparameter tuning with grid search as a process to improve the performance of Support Vector Machine (SVM) classification added to collaborative filtering.

This research aims to implement a system combining collaborative filtering techniques with the Support Vector Machine (SVM) classification method. With the hope that the application of classification using SVM after being processed using collaborative filtering can produce a good film recommendation model and provide accurate predictions for recommended and non-recommended films from the process of adding classification methods. Collaborative filtering results using RMSE as a model evaluation to determine the best form of user-based and item-based classification using Support Vector Machine (SVM) with precision and recall as a benchmark for film recommendation results and films that are not recommended.

The structure of this research is as follows. Section 2 describes the research method. Section 3 shows the results and discussion of the research conducted. Furthermore, in section 4, conclusions and suggestions are based on the experimental results.

2. Research Methods

The system plan to be built on film recommendation uses two different methods. The first uses collaborative filtering, and the second combines collaborative filtering with Support Vector Machine (SVM) to get film recommendation results. The first process consists of several steps: Crawling Data, Data Preprocessing 1, Collaborative Filtering User-based and Item-based, and Evaluation of Collaborative filtering results. The first system design can be seen in Figure 1.

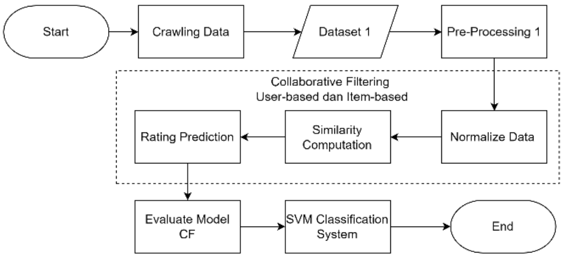

Figure 1. Collaborative Filtering System

The second process, which is continued from Figure 1, consists of classification using the Support Vector Machine (SVM) algorithm and evaluation of the classification model. The second system design can be seen in Figure 2.

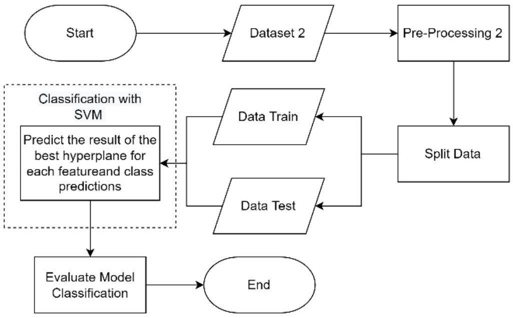
Start
Dataset 2
Pre-Processing 2
Classification with SVM
Predict the result of the best hyperplane for each featureand class predictions
Data Train
Data Test
Split Data
Evaluate Model Classification
End

Figure 2. SVM Classification System

2.1. Crawling Data

We crawled on Twitter using the SNScrape python library in the data crawling process. The data crawled is in the form of tweet reviews from every user who is trusted in reviewing films. Additionally, we crawled data based on film titles on the Netflix platform. The Netflix film titles we crawled were film titles from 2005-2021. The data were taken in the form of id_tweet, username, date, tweet, and movie title.

After obtaining data containing film reviews on film titles on Netflix, we select appropriate reviews containing film reviews. Then choose the best tweet review regarding the discussion of related film titles. After that, the data will be added with rating values from websites specifically for reviewing films, such as IMDb, Rotten Tomatoes, and Metacritic, according to the film titles on Netflix. The results of the data crawling process get results like Table 1.

Table 1. Crawling Data

User	Movie Title	Total Data
35	785	6184

2.2. Data Preprocessing

Data preprocessing is an important technique in getting high-quality and efficient data. In this research, preprocessing is divided into two. First is the Preprocessing 1 stage. Here the data that was originally in the form of reviews on Twitter is converted into a 1-5 rating form that can be used as a recommendation system. Several steps are performed in converting tweet sentences into rating values: Text Processing, Polarity, and Labeling.

Text processing is a stage to get more structured data in selecting text data. At this stage, text cleaning is carried out, which still contains elements of punctuation, numbers, emoticons, URLs, and hashtags.

Polarity is the process of identifying a text with a measure of how negative and how positive the text is. Polarity is useful in the method of predicting sentences

DOI: https://doi.org/10.29207/resti.v6i5.4270

that have positive or negative phrases. For example, "The film is amazing." then the word "amazing" has a positive context [9]. In this research, we apply polarity by using the library from TextBlob. This library helps the processed text data to be good at identifying the meaning of the word. For example, text data that has a polarity value close to -1 means that the rating will be made between 0-2.4, then data whose polarity is close to 1 will be made a rating between 2.6-5 and data that produces a polarity value of 0 becomes a rating of 2.5.

Labeling in this research process is to identify the polarity result data to be rechecked to whether it is by the existing rating context. Text data becomes a rating with a value of 0 to 5.

Preprocessing 2 is changing the rating data, which was originally still 1-5, then made into 0 and 1. Generally, at the preprocessing two stages, the data used is data that has gone through the process stages of collaborative filtering. Ratings with a scale of 0-2 are converted to a value of 0 which means the user does not like the movie, and ratings with a scale of 3-5 are converted into 1, which means the user likes the film.

2.3. Collaborative Filtering

Collaborative filtering is a recommendation system that provides results based on information from users looking for other users with similar interests. Recommendations are given based on user preferences that provide similarity values [10]. In collaborative filtering, to make recommendations can use similarity based on an item (item-based) and similarity based on the user (user-based).

The steps taken in the collaborative filtering system in this research are data normalization, calculating the similarity value, rating prediction, and evaluating the collaborative filtering model.

Data normalization is a process of grouping data attributes that form simple, non-redundant, and flexible entities. So that data that has been normalized has good quality. The formula used for data normalization is like formula 1.

$$nr_{i,u} = r_{i,u} - \bar{r}_u \tag{1}$$

For $nr_{i,u}$ is the normalized rating of item i by user u. Then for $r_{i,u}$ the actual rating of item i from user u and $\bar{r}_u$ is the average rating of the items rated by users.

When calculating the similarity value, there are many ways to find it, one of which is this method. For example, Pearson correlation similarity value between u_1 and u_2 can be calculated as formula 2 [11]:

$$sim(u_1, u_2) = \frac{\sum_{i \in I_{u_1 u_2}} (x_{u_1,i} - \bar{x}_{u_1})(x_{u_2,i} - \bar{x}_{u_2})}{\sqrt{\sum_{i \in I_{u_1 u_2}} (x_{u_1,i} - \bar{x}_{u_1})^2 (x_{u_2,i} - \bar{x}_{u_2})^2}} \tag{2}$$

$I_{u_1 u_2}$ is used to designate a set of items covered by u_1 and u_2. $\bar{x}_{u_1}$ denotes the average rating of user u_1.

Next, the rating prediction stage of this process predicts the rating on the empty rating value. Using the value of *n* in *top-n*, which has the smallest RMSE value. An item-based can be applied like formula 3 [11]:

$$x_{i,j} = \bar{x}_j + \frac{\sum_{i' \in \hat{I}} sim(j,i') \times (x_{i,i'} - \bar{x}_{i'})}{\sum_{i' \in \hat{I}} |sim(j, i')|} \tag{3}$$

For $\hat{I}$ means the set of N items that are similar to item j and have been rated by user i. In user-based, it can be applied as formula 4 [11]:

$$x_{i,j} = \bar{x}_i + \frac{\sum_{u' \in \widehat{U}} sim(i,u') \times (x_{u',j} - \bar{x}_{u'})}{\sum_{u' \in \widehat{U}} |sim(i,u')|} \tag{4}$$

For $\widehat{U}$ means the set of N nearest neighbors of user i who has rated item j.

2.4. Support Vector Machine (SVM)

Support Vector Machine (SVM) is a classification algorithm that provides a more efficient and accurate classification process compared to other classification methods because it applies the principle of Structural Risk Minimisation (SRM), which ensures low error in classification [12]. SVM develops a hyperplane or n-hyperplane that is useful for keeping some data points in the class [13]. The optimal hyperplane is the one with the maximum distance from the supported points of each class. In particular, higher margins are needed to reduce global errors.

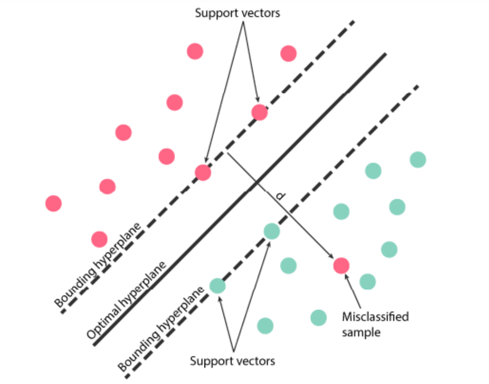

Figure 3. SVM Model Illustration [14]

SVM is an appropriate classification used in separating two classes in the input space because the SVM algorithm's goal is to find the best hyperplane [15]. The two data divided by the hyperplane are the first-class worth 1 and the next class worth -1 as formulas 5 and 6 [15].

$$X_{i.}W + b \geq 1 \; for \; Y_i = 1 \tag{5}$$

$$X_{i.}W + b \leq -1 \; for \; Y_i = -1 \tag{6}$$

For X_i is the *-i* data, *W* is the weight value of the support vector perpendicular to the hyperplane, *b* is the bias value, and Y_i is the *-i* data class.

SVM is supervised learning that focuses on extracting features from user-profiles and training classifiers for the classification process [12]. The SVM method is usually used in two ways: firstly, in the user-item matrix, all items are used as features [12]. The main principle of SVM is linear classification. Still, it was developed to overcome nonlinear problems using a trick kernel. A trick kernel is like formula 7 [16].

$$K(x_i \cdot x_k) = \phi_j \cdot \phi_k \qquad (7)$$

The use of kernels can optimize the process of Support Vector Machine (SVM) classification by knowing the kernel function to be used. For example, the following kernels such as formulas 8, 9, 10, and 11 [16].

$$Kernel\ Linear: K(x_i, x_k) = x_k^T x \qquad (8)$$

$$Kernel\ Polynomial: K(x_i, x_k) = (x_k^T + 1)^d \qquad (9)$$

$$Kernel\ RBF:$$
$$K(x_i, x_k) = exp\{-\| x - x_k \|^2 + 2\sigma^2\} \qquad (10)$$

$$Kernel\ Sigmoid: K(x_i, x_k) = \tanh\,[kx_k^T x + \theta] \qquad (11)$$

In this study, we also conducted a process to improve the quality of Support Vector Machine (SVM) classification using hyperparameter tuning.

Hyperparameter tuning is a method for optimizing algorithms. Hyperparameter tuning techniques are grid search, random search, evolutionary, and sequential model-based optimization [17]. In this study, the authors used hyperparameter tuning with grid search. The grid search algorithm tries all parameter value combinations and returns the high-value combinations [18].

2.5. Performance Evaluation

Evaluating Collaborative Filtering model is the process of calculating the value of RMSE in the Collaborative Filtering method. Root mean square error (RMSE) calculates the largest difference for large errors in rating prediction [19]. Therefore, if the RMSE value is closer to 0 the better. To find RMSE like formula 12 [19].

$$RMSE = \sqrt{\frac{\sum_{i=1}^{n}(\tilde{r}_i - r_i)^2}{n}} \qquad (12)$$

Furthermore, to Evaluate the classification model, performance measurements on classification accuracy metrics can be calculated using the Confusion Matrix. In order to calculate the ratio of relevant recommendation results, it is necessary to calculate precision and recall.

In the confusion matrix, there are four generated in the table, including True Positive (TP), False Positive (FP), False Negative (FN), and True Negative (TN). True Positive (TP) is a dataset correctly predicted as positive. False Positive (FP) is a negative but expected positive dataset, and False Negative (FN) is a positive but expected negative dataset. True Negative (TN) is the amount of negative data that is expected to be negative. The indicators on the confusion matrix to detection classes can be determined by calculating accuracy, precision, recall, and F1-Score.

This research uses precision and recall. Precision is the accuracy of the information requested by the user with the predicted results in the model. A recall describes the model's success in finding the information again. Precision and recall can be calculated as formulas 13 and 14 [20].

$$Precision = \frac{TP}{TP+FP} \qquad (13)$$

$$Recall = \frac{TP}{TP+FN} \qquad (14)$$

3. Results and Discussions

In this research, the first step is to calculate the predicted rating value using the collaborative filtering memory-based method, namely user-based and item-based, then evaluate using the RMSE value to select the best top-n. The second step is the classification process using the Support Vector Machine (SVM) algorithm, then optimized using grid search hyperparameters to get recommended or not recommended film results, then evaluated using precision and recall values.

3.1. Data

We crawled Twitter data based on 785 film titles on Netflix and 30 Twitter users using the SNScrape library. After that, we selected one tweet from thousands of review data about each related film title. The results of choosing the best review get a total of 3134 data. The following Table 2 displays the crawling result data that has gone through the selection of 1 review that matches the film title.

Table 2. Crawling Result Data on Twitter

id tweet	username	date	tweet	title
1100	djaycoholyc	2011-09-03 16:25:58+00:00	Nonton IP Man lagi. Berasa beda ya, tetep suka sih. Cuma sekuelnya jelek.	Ip Man
....	....	....	....	....
1400	HabisNontonFilm	2021-06-03 13:17:17+00:00	Masih takjub sama bagusnya adegan ini di The Conjuring. 👏👏 https://t.co/IoIEDvUlLn	The Conjuring

DOI: https://doi.org/10.29207/resti.v6i5.4270

After selecting the best review, it will go to the next step: text processing to clean punctuation marks, numbers, emoticons, URLs, and hashtags. Therefore, can be seen in Table 3 of the text processing process.

Table 3. Text Processing Result Data

id tweet	username	date	tweet	title
1100	djaycoholyc	2011-09-03 16:25:58+00:00	nonton ip man lagi berasa beda ya tetep suka sih cuma sekuelnya jelek	Ip Man
....	....	....	....	....
1400	HabisNontonFilm	2021-06-03 13:17:17+00:00	masih takjub sama bagusnya adegan ini di the conjuring	The Conjuring

Then proceed to the polarity stage, which is used as a step to identify sentences that have positive or negative phrases, and convert them into a value of 0 to 5 as a tweet rating. The pattern can be seen in Table 4.

Table 4. Polarity Result Data

id tweet	....	tweet	polarity	rating
1100	....	nonton ip man lagi berasa beda ya tetep suka sih cuma sekuelnya jelek	-0,35	1.63
....	....	....	....	....
1400	....	masih takjub sama bagusnya adegan ini di the conjuring	0.7	4.25

The dataset from the crawling process on Twitter was then combined with ratings from the IMDb, Rotten Tomatoes, and Metacritic websites. The results of the data merge consisted of 6184 reviews, 35 usernames, and 791 Netflix film titles. The combination of Twitter crawling data with ratings from various websites can be seen in Table 5.

Table 5. Dataset 1

username	film	idUser	idFilm	rating
HabisNontonFilm	The Conjuring	1	605	4.25
rayculz	Black Panther	2	306	3.72
....	....	....	....	....
Metacritic metascore	Cult of Chucky	34	353	3.45
Metacritic User Score	Righteous Kill	35	788	2.45

Then a new dataset is created as a 2-dimensional pivot table matrix containing idUser, idFilm, and rating containing 5988 filled rating data and 21487 empty ratings so that it has 78.20% sparsity data as in Table 6.

Table 6. Matrix Dataset 1

idFilm	1	2	....	....	790	791
idUser						
1	0	2.65	....	....	2.44	0
2	0	0	....	....	2.92	0
....	....	....	....	....	....	....
....	....	....	....	....	....	....
34	0	4.25	....	....	1.70	2.05
35	0	3.75	....	....	3.15	2.25

3.2. Collaborative Filtering Result

The normalization process is carried out using the dataset already in the matrix to make it non-redundant and flexible. The normalization process takes place item-based and user-based. The normalized data for item-based can be seen in Table 7, and for user-based can be seen in Table 8.

Table 7. Item-based Normalized Data

idFilm	1	2	...	...	790	791
idUser			...	...		
1	0	-0.114	...	...	-0.198	0
2	0	0	...	...	-0.040	0
...	...	...	...	...	...	...
...	...	...	...	...	...	...
34	0	0.328	...	...	-0.251	-0.17
35	0	0.159	...	...	0.005	-0.22

Table 8. User-based Normalized Data

idUser	1	2	...	...	34	35
idFilm			...	...		
1	0	0	...	...	0	0
2	-0.225	0	...	...	0.616	0.353
...	...	...	...	...	...	...
...	...	...	...	...	...	...
790	-0.109	0.040	...	...	-0.340	0.112
791	0	0	...	...	-0.076	0.006

After normalizing user-based and item-based, the next step is calculating the similarity value using Pearson correlation, as can be seen for the calculation of user-based similarity in Table 9 and for item-based in Yable 10.

Table 9. Data Similarity Item-based

idFilm	1	2	...	...	790	791
idFilm			...	...		
1	1	-0.166	...	...	-0.316	-0.67
2	0	1	...	...	0.046	-0.10
...	...	...	...	...	...	...
...	...	...	...	...	...	...
790	-0,316	0.046	...	...	1	0.513
791	-0,675	-0,101	...	...	0.513	1

Table 10. Data Similarity User-based

idUser	1	2	...	...	34	35
idUser			...	...		
1	1	0.033	...	...	-0.123	-0.06
2	0.033	1	...	...	-0.025	-0.52
...	...	...	...	...	...	...
...	...	...	...	...	...	...
34	-0.123	-0.025	...	...	1	-0.11
35	-0.062	-0.052	...	...	-0.118	1

If the similarity value is close to 1, then the two items are very similar. Close to -1, then the two items are very different and close to 0 does not have much correlation. Next, use the similarity value to calculate the prediction rating with top-n. Finally, the prediction rating will be evaluated using RMSE based on the best n value. The correlation between RMSE and n for item-based as shown in Figure 4 and for user-based as shown in Figure 5.

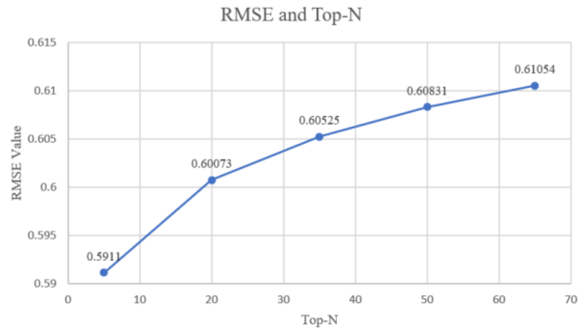

Figure 4. RMSE and Top-N for Item-based

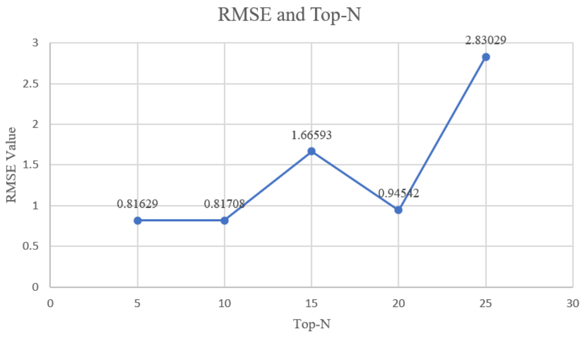

Figure 5. RMSE and Top-N for User-based

The better RMSE value is close to 0, so the best RMSE value on item-based is when top-n = 5 with an RMSE value of 0.5911. For user-based, the best RMSE value is when top-n = 5 with an RMSE value of 0.81629. Furthermore, each top-n that produces the best RMSE value will be selected to calculate the prediction rating and used for the classification process using the Support Vector Machine (SVM) in the next step. Dataset 2 will be used as a user-based and item-based classification process based on the best top-n value prediction rating. As in Table 11, item-based dataset 2, and user-based dataset 2 in Table 12.

Table 11. Matrix Dataset 2 Item-based

idFilm	1	2	...	...	790	791
idUser			...	...		
1	3.89	2.65	...	...	2.44	2.23
2	3.89	3.08	...	...	2.92	2.23
...	...	...	...	...	...	...
...	...	...	...	...	...	...
34	3.89	4.25	...	...	1.70	2.05
35	3.89	3.75	...	...	3.15	2.25

Table 12. Matrix Dataset 2 User-based

idUser	1	2	...	...	34	35
idFilm			...	...		
1	2.94	3.07	...	...	3.28	2.91
2	2.65	3.09	...	...	4.25	3.75
...	...	...	...	...	...	...
...	...	...	...	...	...	...
790	2.44	2.92	...	...	1.70	3.15
791	3.08	3.07	...	...	2.05	2.25

3.3. SVM Classification Results

Dataset 2 will be used in the classification process by changing the 0-5 rating values to 0 and 1. The 0-2 rating value will be changed to a value of 0, and the 3-5 rating value will be changed to a value of 1. The value 1 is assumed to be a value that means the user likes the film. A value of 0 is assumed to be a value that means the user does not like the film. As shown in Table 13, the values will be 0 and 1 for item-based and Table 14 for user-based.

Table 13. Dataset 2 Item-based SVM Classification

idFilm	1	2	...	...	790	791
idUser			...	...		
1	1	0	...	...	0	0
2	1	1	...	...	0	0
...	...	...	...	...	...	...
...	...	...	...	...	...	...
34	1	1	...	...	0	0
35	1	1	...	...	1	0

Table 14. Dataset 2 User-based SVM Classification

idUser	1	2	...	...	34	35
idFilm			...	...		
1	0	1	...	...	1	0
2	0	1	...	...	1	1
...	...	...	...	...	...	...
...	...	...	...	...	...	...
790	0	0	...	...	0	1
791	1	1	...	...	0	0

After the dataset has been converted to 0 and 1, it will then be processed to SVM classification. In SVM classification, hyperparameter tuning is performed using grid search. For the SVM process, item-based data can be seen in Table 15 and user-based data in Table 16.

Table 15. SVM on Item-based

Test Size	Random State	C	Gamma	Kernel	Average Precision	Average Recall
0,1	8	1	scale	RBF	0.85036	0.90714

Table 16. SVM on User-based

Test Size	Random State	C	Gamma	Kernel	Average Precision	Average Recall
0,3	0	1	scale	RBF	0.85966	0.8927

Hyperparameter tuning using grid search will find the best parameter value in optimizing the value of precision and recall. As in Table 17, the hyperparameter tuning steps use grid search.

DOI: https://doi.org/10.29207/resti.v6i5.4270

Table 17. Hyperparameter Tuning in User-based

Test Size	Random State	C	Gamma	Kernel	Average Precision	Average Recall
0,3	0	10	0.1	RBF	0.87753	0.88958

After performing hyperparameter tuning, it can be seen that the best parameters for item-based data are at C = 0.1, Gamma = 1, Kernel = RBF with the same precision and recall values as SVM before hyperparameter tuning. On the other hand, for user-based data, the best parameters are C = 10, Gamma = 1, and Kernel = RBF, with an optimal increase after hyperparameter tuning on the precision value. By using hyperparameter tuning using grid search can find the best parameters for support vector machine classification.

4. Conclusion

In the research, we have done applying collaborative filtering methods combined with Support Vector Machine (SVM) classification. Using a crawling dataset from Twitter combined with IMDb, Rotten Tomatoes, and Metacritic web rating reviews, which is then processed to produce a film rating. Collaborative filtering in rating prediction using user-based and item-based get the best top-n value on item-based top-n = 5 with RMSE value 0.5911 and user-based top-n = 5 with RMSE value 0.8162. The results of the best top-n are then processed in the Support Vector Machine (SVM) classification algorithm plus optimization of the classification algorithm with hyperparameter tuning using grid-search. Finding the best parameters are C = 0.1, Gamma = 1, Kernel = RBF getting for item-based the value of precision 85,03% and recall 90,71%. For user-based with the best parameters of C = 10, Gamma = 1, Kernel = RBF get a precision value of 87,75% and a recall value of 88,95%.

It can be concluded that the combination of collaborative filtering and a Support Vector Machine (SVM) classification algorithm can be used to make whether a movie can be recommended or not. And using hyperparameter tuning with grid search can find the best parameters to produce optimal values.

Therefore, future research can improve the performance of the recommendation system with a larger data set. In addition, in the future, it can be combined with other methods such as content-based or using other classification algorithms to create a recommendation system.

Reference

[1] M. Chiny, M. Chihab, O. Bencharef, and Y. Chihab, "Netflix Recommendation System based on TF-IDF and Cosine Similarity Algorithms," no. Bml 2021, pp. 15–20, 2022, doi: 10.5220/0010727500003101.

[2] H. Tahmasebi, R. Ravanmehr, and R. Mohamadrezaei, "Social movie recommender system based on deep autoencoder network using Twitter data," *Neural Comput. Appl.*, vol. 33, no. 5, pp. 1607–1623, 2021, doi: 10.1007/s00521-020-05085-1.

[3] N. Monarizqa, L. E. Nugroho, and B. S. Hantono, "Penerapan Analisis Sentimen Pada Twitter Berbahasa Indonesia Sebagai Pemberi Rating," *Jurnal Penelitian Teknik Elektro dan Teknologi Informasi*, vol. 1. pp. 151–155, 2014.

[4] D. Das, H. T. Chidananda, and L. Sahoo, *Personalized movie recommendation system using twitter data*, vol. 710. Springer Singapore, 2018.

[5] X. Wang, Z. Dai, H. Li, and J. Yang, "A New Collaborative Filtering Recommendation Method Based on Transductive SVM and Active Learning," *Discret. Dyn. Nat. Soc.*, vol. 2020, no. 1, 2020, doi: 10.1155/2020/6480273.

[6] L. Ren and W. Wang, "An SVM-based collaborative filtering approach for Top-N web services recommendation," *Futur. Gener. Comput. Syst.*, vol. 78, pp. 531–543, 2018, doi: 10.1016/j.future.2017.07.027.

[7] M. Grčar, D. Mladenič, B. Fortuna, and M. Grobelnik, "Data sparsity issues in the collaborative filtering framework," *Lect. Notes Comput. Sci. (including Subser. Lect. Notes Artif. Intell. Lect. Notes Bioinformatics)*, vol. 4198 LNAI, pp. 58–76, 2006, doi: 10.1007/11891321_4.

[8] X. Yu, Y. Chu, F. Jiang, Y. Guo, and D. Gong, "SVMs Classification Based Two-side Cross Domain Collaborative Filtering by inferring intrinsic user and item features," *Knowledge-Based Syst.*, vol. 141, pp. 80–91, 2018, doi: 10.1016/j.knosys.2017.11.010.

[9] A. P. Gopi, R. N. S. Jyothi, V. L. Narayana, and K. S. Sandeep, "Classification of tweets data based on polarity using improved RBF kernel of SVM," *Int. J. Inf. Technol.*, 2020, doi: 10.1007/s41870-019-00409-4.

[10] V. Subramaniyaswamy, R. Logesh, M. Chandrashekhar, A. Challa, and V. Vijayakumar, "A personalised movie recommendation system based on collaborative filtering," *Int. J. High Perform. Comput. Netw.*, vol. 10, no. 1–2, pp. 54–63, 2017, doi: 10.1504/IJHPCN.2017.083199.

[11] P. Thakkar, K. Varma, V. Ukani, S. Mankad, and S. Tanwar, *Combining User-Based and Item-Based Collaborative Filtering Using Machine Learning*. Springer Singapore, 2019.

[12] N. Samaiya, S. K. Raghuwanshi, and R. K. Pateriya, *Shilling attack detection in recommender system using PCA and SVM*, vol. 813. Springer Singapore, 2019.

[13] A. Tripathi and A. K. Sharma, "Recommending Restaurants: A Collaborative Filtering Approach," *ICRITO 2020 - IEEE 8th Int. Conf. Reliab. Infocom Technol. Optim. (Trends Futur. Dir.*, pp. 1165–1169, 2020, doi: 10.1109/ICRITO48877.2020.9197946.

[14] J. Cardoso-Fernandes, A. C. Teodoro, A. Lima, and E. Roda-Robles, "Semi-automatization of support vector machines to map lithium (Li) bearing pegmatites," *Remote Sens.*, vol. 12, no. 14, 2020, doi: 10.3390/rs12142319.

[15] I. M. Parapat, M. T. Furqon, and Sutrisno, "Penerapan Metode Support Vector Machine (SVM) Pada Klasifikasi Penyimpangan Tumbuh Kembang Anak," *J. Pengemb. Teknol. Inf. dan Ilmu Komput.*, vol. 2, no. 10, pp. 3163–3169, 2018, [Online]. Available: https://j-ptiik.ub.ac.id/index.php/j-ptiik/article/view/2577.

[16] A. R. D. Pratiwi and E. B. Setiawan, "Implementation of Rumor Detection on Twitter Using the SVM Classification Method," *J. RESTI (Rekayasa Sist. Dan Teknol. Informasi)*, no. 10, pp. 782–789, 2020.

[17] G. Y. N. N. Adi, M. H. Tandio, V. Ong, and D. Suhartono, "Optimization for Automatic Personality Recognition on Twitter in Bahasa Indonesia," *Procedia Comput. Sci.*, vol. 135, pp. 473–480, 2018, doi: 10.1016/j.procs.2018.08.199.

[18] Gita Safitri and Erwin Budi Setiawan, "Optimization Prediction of Big Five Personality in Twitter Users," *J. RESTI (Rekayasa Sist. dan Teknol. Informasi)*, vol. 6, no. 1, pp. 85–91, 2022, doi: 10.29207/resti.v6i1.3529.

[19] F. Fkih, "Similarity measures for Collaborative Filtering-based Recommender Systems: Review and experimental comparison," *J. King Saud Univ. - Comput. Inf. Sci.*, no. xxxx, 2021, doi: 10.1016/j.jksuci.2021.09.014.

DOI: https://doi.org/10.29207/resti.v6i5.4270

[20] G. Xu, G. Jia, L. Shi, and Z. Zhang, “Personalized Course Recommendation System Fusing with Knowledge Graph and Collaborative Filtering,” *Comput. Intell. Neurosci.*, vol. 2021, 2021, doi: 10.1155/2021/9590502.