

# **Recommender System Based on Tweets with Singular Value Decomposition and Support Vector Machine Classification**

**Rafi Anandita Wicaksono\* , Erwin Budi Setiawan**

Informatics Study Program, Telkom University, Bandung, Indonesia

Email: <sup>1\*</sup>rafianadita@student.telkomuniversity.ac.id, <sup>2</sup>erwinbudisetiawan@telkomuniversity.ac.id

Submitted: 10/08/2022; Accepted: 23/08/2022; Published: 30/08/2022

**Abstract**—In modern times, the movie industry is growing rapidly. Netflix is one of the platforms that can be used to watch movies and provides many types of genres and movie titles. With so many genres and movie titles sometimes making it difficult for people to choose a movie to watch, one solution to the problem is a recommendation system that can recommend movies based on user ratings. One method in the recommendation system is collaborative filtering. One of the algorithms contained in collaborative filtering is singular value decomposition. Twitter is one of the places where people often write their opinions about the movies they have watched, from people's tweets on Twitter will be processed into rating value data. In this system, tweets become input that is processed into data that has a rating. This research implements a user-based recommendation system based on ratings from tweets using collaborative filtering combined with the Singular Value Decomposition (SVD) algorithm and Support Vector Machine (SVM) classification and implemented it on user-based and item-based. This research aims to implement a system that combines collaborative filtering techniques with the Singular Value Decomposition (SVD) algorithm and Support Vector Machine (SVM) classification. With the hope of producing a good movie recommendation model and providing accurate predictions for recommended and non-recommended movies. The test results in this study show that Collaborative Filtering gets the best RMSE value of 0.8162 on user-based and 0.5911 on item-based. The combination of Singular Value Decomposition (SVD) algorithm and Support Vector Machine (SVM) classification using hyperparameter tuning resulted in 81% precision and 81% recall for user-based while 80% precision and 80% recall for item-based.

**Keywords:** Recommender System; Collaborative Filtering; Singular Value Decomposition; Netflix; Twitter; SVM

## **1. INTRODUCTION**

In this modern era, the development of the film industry and technology is very rapid. Movies are entertainment media that have various genres and titles to watch. Nowadays, people can watch movies not only through theaters but can use the Netflix platform. Netflix is an application that provides online streaming services, namely movies with various genres and titles [1]. On Netflix there are very many movies available, sometimes it will cause users confusion to choose the movie to watch. The solution to this problem is to have a movie recommendation system on Netflix that can recommend movies that have high ratings to users.

More and more people are accessing social media. This is one of the reasons why the world of technology is growing so rapidly. One of the most popular social media at the moment is Twitter. This is because users can express their feelings, ideas, and thoughts to their followers in the form of tweet updates[2].

There are several types of recommender systems, including collaborative filtering, content-based filtering, knowledge-based, and association-ruled-based recommendation. The most successful among the available techniques is collaborative filtering. There are two categories namely model-based and memory-based[3]. Collaborative filtering is a method that provides a recommendation based on the similarity of users [4]. Singular Value Decomposition (SVD) is one of the collaborative filtering algorithms that factorizes a non-zero matrix.[5].

Support Vector Machine (SVM) introduced by Vapnik is one of the supervised learning methods with multidimensional estimation classification based on SVM development time. This method goes beyond learning theory. The method can be seen as a general approach to expressing functions in a higher dimensional space, and in many cases can overcome the "curse of dimensionality" [6].

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 [3], [7]. Therefore, in this research, the author will combine collaborative filtering technique with Singular Value Decomposition (SVD) algorithm and Support Vector Machine (SVM) classification technique with classification method using hyperparameter tuning with grid search to improve performance. To the author's knowledge, there has been no research using hyperparameter tuning with grid search as a process to improve the performance of combining the Singular Value Decomposition (SVD) algorithm with Support Vector Machine (SVM) classification added to collaborative filtering.

This research aims to implement a system that combines collaborative filtering techniques with the Singular Value Decomposition (SVD) algorithm and Support Vector Machine (SVM) classification. With the hope that the application of classification using SVM after processing using collaborative filtering produces a good movie recommendation model and provides accurate predictions for recommended and non-recommended movies from the process of adding classification methods. The results of collaborative filtering use 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 benchmarks and recall as benchmarks for the results of movie recommendations and movies that are not recommended movies and movies that are not recommended.

## 2. RESEARCH METHODS

### 2.1 Research Stages

The system plan to be built applies two different methods. The first uses collaborative filtering and the second combines collaborative filtering, Singular Value Decomposition (SVD), and Support Vector Machine (SVM) to get movie recommendation results. This process consists of several steps including Data Crawling, Data Preprocessing, Collaborative Filtering User-based and Item-based, Evaluation of Collaborative filtering results, classification using Support Vector Machine (SVM) Algorithm, Evaluation of classification models. The two steps of system design can be seen in Figure 1 and Figure 2.

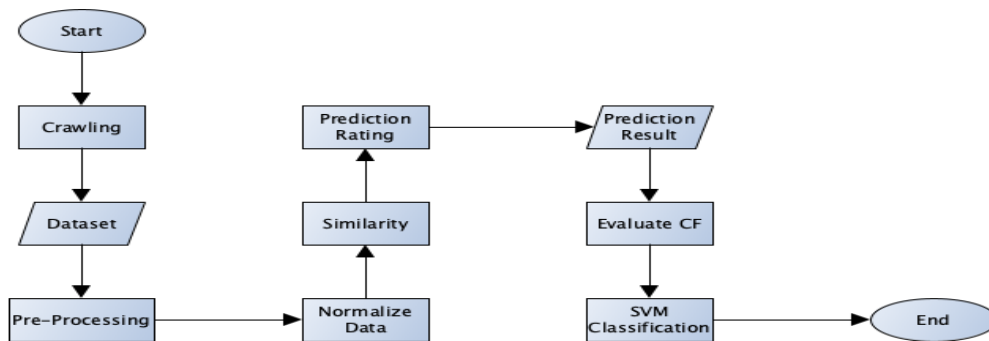


Figure 1. Collaborative Filtering System

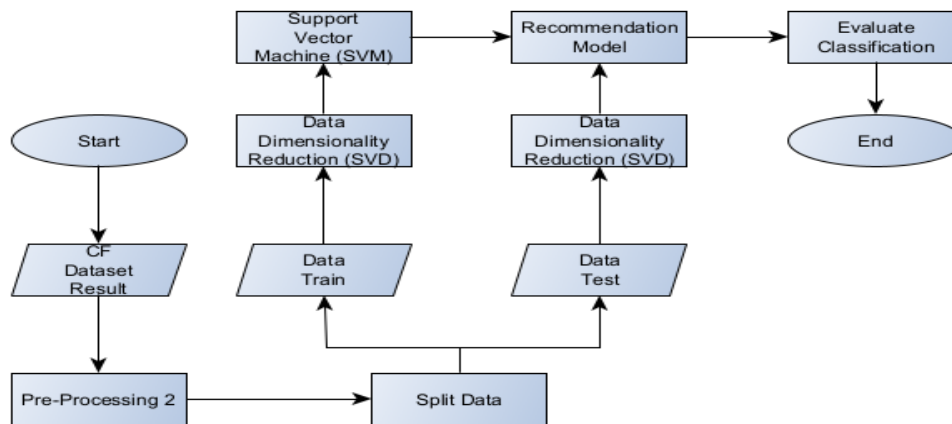


Figure 2. SVM Classification System

### 2.2 Data Crawling

In the data crawling process that we have done on Twitter, namely using the sncrape python library. The data that has been crawled produces several review tweets from each Twitter user who is trusted in reviewing movies. We crawled data sourced from movie titles on the Netflix platform. The movie titles on Netflix that we crawled were movie titles from 2005-2021 and the data we retrieved were id\_tweet, username, date, tweet, movie title.

After obtaining data containing reviews of movie titles on Netflix, we then select several reviews that match the movie reviews. Then we choose one of the best tweet reviews regarding the discussion of related movie titles. After that, the selected data will be added using rating values from specialized websites for reviewing movies (IMDb, Rotten Tomatoes, and Metacritic) according to the movie titles on Netflix. The results that have been obtained from the data crawling process get results such as Table 1.

Table 1. Crawling Data

User	Movies	Total Data
35	785	6184

### 2.3. Data Preprocessing

Data preprocessing is an important step in data processing, the goal is to get good and efficient data quality. at this stage, the data that was originally in the form of reviews on Twitter will be converted into a 1-5 rating form that can be used as a recommendation system. In the process of converting tweets into ratings, several stages are carried out consisting of Text Processing, Polarity, and Labeling.

Text processing is a stage to get more structured data, at this stage cleaning text that still contains elements of punctuation, numbers, emoticons, URLs, and hashtags.

Polarity is a process that is done to identify a text with a measure of how negative and how positive the text is. Polarity is useful in the process of predicting sentences that have positive or negative words. For example, "the movie is cool" then the word "cool" has a positive meaning. [8]. In this research, polarity is applied by using the library from TextBlob. This library aims to help the processed text data to identify the meaning of words properly. 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 by identifying data from the polarity results to check whether the data is following the existing rating context. Text data becomes a rating with a value of 0 - 5.

Preprocessing 2 is changing the rating data which was originally still 1-5 then made into 0 and 1. In general, at the preprocessing 2 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 into a value of 0 which means the user does not like and ratings with a scale of 3-5 are converted into 1 which means the user likes the movie.

### 2.3. Collaborative Filtering

Collaborative filtering is a recommendation system that provides recommendations based on information from users who search for other users with similar interests. Recommendations are based on an arrangement of users who provide a similarity score.[9]. 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 study are data normalization, calculating similarity values, predicting ratings, and evaluating collaborative filtering models.

Data normalization is a process for 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 user  $u$ .

When calculating the similarity value, there are many ways to find it, one of which is this method. Pearson correlation similarity value between  $u_1$  and  $u_2$  can be calculated as formula 2 [10]:

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

For  $I_{u_1 u_2}$  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 topN which has the smallest RMSE value. On item-based can be applied like formula 3 [10]:

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

For  $\hat{I}$  means the set of  $N$  items that are similar to item  $j$  and have been rated by user  $i$ .  $n$  user-based can be applied as formula 4 [10]:

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

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

### 2.4. Singular Value Decomposition (SVD)

The Singular Value Decomposition (SVD) algorithm is a real factorization on complex matrices, where in the SVD transformation, the original matrix can be decomposed into three matrices of the same size, but if you multiply the three matrices that have been decomposed, it will be the same as the original matrix.[11]. The equation of the SVD algorithm is:

$$A = USV^T \tag{5}$$

With the following visualization:

$$[A]_{m \times n} = [U]_{m \times m} \cdot [S]_{n \times m} \cdot [V^T]_{n \times m} = \sum_{i=1}^{\min(m,n)} \sigma_i \cdot u_i \cdot v_i^T \tag{6}$$

$U$  is an  $m \times n$  matriks,  $S$  is an  $n \times n$  diagonal matriks, and  $V^T$  is an  $n \times n$  Column matrix,  $U$  is called the left singular vector,  $\{U_k\}$  to form the orthonormal basis, for the profile expression test  $U_i \cdot U_j = 1$  for  $i = j$  and otherwise

$U_i, U_j = 0$ . The rows of  $V^T$  contain elements of the right singular vector  $\{V_k\}$  and form the orthonormal basis for gene transcription responses. The elements of  $S$  are only zero on the diagonal and are called singular values, hence  $S = \text{diag}(s_1, \dots, s_n)$  henceforth  $S_k > 0$  for  $1 \leq k \leq r$  and  $S_1 = 0$  for  $(r + 1) \leq k \leq n$  with the sorting convention of the order of singular values of the singular vectors determined by high to low, the sorting of singular values starts with the highest singular value at the top left of the index  $S$  matrix [5]. In SVD calculation, we first need to calculate the eigenvalues and eigenvectors of  $AA^T$  and  $A^T A$ . The eigenvectors of  $AA^T$  form a column of  $U$ , while the eigenvectors of  $A^T A$  form a column of  $V$ . In addition, the singular values ( $SV_s$ ) in  $S$  are the square roots of the eigenvalues of  $A^T A$  atau  $AA^T$ .

## 2.5. 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 Minimization (SRM) which guarantees a low classification error rate.[12]. SVM develops a hyperplane or n-hyperplane that is useful for keeping some of the data points in the class. [13]. The best hyperplane is the one that is at the maximum distance from the supported points of each class. In particular to reduce the global error the margin should be higher.

SVM is an appropriate classification used in separating two classes in the input space because the purpose of the SVM algorithm can find the best hyperplane. [14]. The two data separated by the hyperplane are the first class worth 1, and the next class worth -1 as in formulas 5 and 6. [14].

$$X_i W + b \geq 1 \text{ for } Y_i = 1 \tag{7}$$

$$X_i W + b \leq -1 \text{ for } Y_i = -1 \tag{8}$$

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, the first is in the user-item matrix, all items are used as features. [12]. The main principle of SVM is linear classification, but it has been developed to overcome non-linear problems by using a trick kernel. Trick kernels such as formula 7[15].

$$K(x_i, x_k) = \phi_j \cdot \phi_k \tag{9}$$

The use of kernels can optimize the process of Support Vector Machine (SVM) classification by knowing the kernel function to be used. The following kernels can be used as formulas 8, 9, 10, and 11 [15].

$$\text{Kernel Linear: } K(x_i, x_k) = x_k^T x \tag{10}$$

$$\text{Kernel Polynomial: } K(x_i, x_k) = (x_k^t + 1)^d \tag{11}$$

$$\text{Kernel RBF: } K(x_i, x_k) = \exp(-\|x - x_k\|^2 + 2\sigma^2) \tag{12}$$

$$\text{Kernel Sigmoid: } K(x_i, x_k) \tanh(kx_k^T x + \theta) \tag{13}$$

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

Hyperparameter tuning is a method for algorithm optimization. Techniques in hyperparameter tuning that can be used are grid search, random search, evolutionary algorithm, and sequential model-based optimization. [16]. In this study, the authors used hyperparameter tuning with grid search. The grid search algorithm works by trying all combinations of parameter values and returning combinations with high values.[17].

## 2.6. Performance Evaluation

Evaluate CF model is the process to calculate the value of RMSE in CF method. Root Mean Square Error (RMSE) calculates the larger difference for large errors in prediction ratings. [18]. The closer the RMSE value is to 0, the better. To find RMSE as formula 12 [18].

$$\text{RMSE} = \sqrt{\frac{\sum_{i=1}^n (\hat{r}_i - r_i)^2}{n}} \tag{14}$$

Furthermore, to Evaluate the classification model, performance measurements on classification accuracy metrics can be calculated using Confusion Matrix. 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 the amount of data that is positive and correctly predicted as positive, False Positive (FP) is the amount of data that is negative but predicted as positive, False Negative (FN) is the amount of data that is positive but predicted as negative, True Negative (TN) is the

amount of data that is negative and predicted as negative. From the indicators on the confusion matrix to detect classes can be obtained by calculating accuracy, precision, recall, and F1-Score.

This research uses precision and recall. Precision is the accuracy between the requested data and the prediction results in the model. Recall describes the success of the model in information. In calculating precision and recall as formulas 13 and 14 [19].

$$\text{Precision} = \frac{TP}{TP+FP} \tag{15}$$

$$\text{Recall} = \frac{TP}{TP+FN} \tag{16}$$

### 3. RESULT AND DISCUSSION

In this research, the first step that must be done is to calculate the predicted rating value using the collaborative filtering memory-based method, namely user-based and item-based, then the data is evaluated using the RMSE value to select the best topN. The second step is the classification process using the Support Vector Machine (SVM) algorithm and then optimized using grid search hyperparameters which aims to produce films that will be recommended or not recommended, after which it can be evaluated using precision and recall values.

#### 3.1 Data Result

Based on crawling data on Twitter, we used 785 movie titles on Netflix and 30 Twitter users using the snsrape library. After that, we selected one tweet review from thousands of review data about each related movie title. The results we have obtained after selecting the best review is that we get a total data of 3134. After the process of selecting one of the best reviews, it will go to the next step, namely the text processing stage. Text processing is cleaning punctuation, numbers, emoticons, URLs, and hashtags. The following Table 2 displays the crawling result data that has gone through the selection of 1 review that matches the movie title and has gone through text processing.

**Table 2.** Data After Text Processing

id tweet	username	date	tweet
1.28E+18	Cenayang Film	2020- 07-14 05:00: 00+00 :00	Film zombie tapi lebih berasa seperti film drama tentang bapak- anak. Kebanyakan film zombie tuh si lakonnya punya senjata kayak bedil untuk survive, nah di sini lakonnya ga punya senjata. Plus lokasi di kereta doang Ada di Viu
.... 1.40E+18	.... HabisNo ntonFilm	.... 2021- 06-03 13:17: 17+00 :00	.... Masih takjub sama bagusnya adegan ini di The Conjuring

Then proceed to the polarity stage, which makes the value into -1 to 1 and changes to a rating of 0 - 5. The pattern can be seen in table 3.

**Table 3.** Data Polarity Result

id tweet	...	tweet	polarity	Rating
1.28E+18	....	Film zombie tapi lebih berasa seperti film drama tentang bapak-anak. Kebanyakan film zombie tuh si lakonnya punya senjata kayak bedil untuk survive, nah di sini	0.5	3.75

		lakonnya ga punya senjata. Plus lokasi di kereta doang Ada di Viu		
....	....	Masih takjub sama bagusnya adegan ini di The Conjuring	....	....
1.40E+18	....		0.7	4.25

Then a new dataset is created in the form of a 2-dimensional pivot table matrix containing User,Movie, and rating containing 5988 filled rating data and 21487 empty ratings as in table 4.

**Table 4.** Matrix Dataset

User	Anak	Cenayang	....	Sir_
Film	Nonton	Film		amirsyarif
#Alive	0.0	2,78	....	0.0
....	....	....	....	....
Young	2.5	2.50	....	3.92
Adult				

**3.2 Collaborative Filtering Result**

Using the dataset that is already in the matrix, the normalization process is carried out to make it non-redundant and flexible. The normalization process takes place item-based and user-based. Data that has been normalized for user-based can be seen in table 5 and item-based can be seen in table 6.

**Table 5.** Normalization Data User-based

Film	#Alive	17 Again	....	Zombielan
User				d
Elbert_Reyner	0.455487	-0.448864	....	0.362758
....	....	....	....	....
IMDB	0.044859	0.414773	....	0.266140

**Table 6.** Normalization Data Item-based

User	Anak	Cenayang	....	Sir_
Film	Nonton	Film		amirsyarif
#Alive	0.0	-0.020326	....	0.0
....	....	....	....	....
Young Adult	2.5	2.50	....	3.92

After normalizing user-based and item-based. then calculate the similarity value using Pearson correlation. As can be seen for the calculation of user-based similarity in table 7 and item-based in table 8.

**Table 7.** Similarity Data User-based

User	Elbert_Reyner	GilaFilm	...	IMDB
User		ID	.	
Elbert_Reyner	1	-	...	-0.029027
....	....	0.050642	.	....
IMDB	-0.029027	0.000806	...	1

**Table 8.** Similarity Data Item-Based

Film	#Alive	1922	....	Young Adult
Film				
#Alive	1	0.489942	....	0.082278
....	....	....	....	....
Young Adult	0.082278	0.384033	....	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 topN. The prediction rating will be evaluated using RMSE based on the best n value. Correlation between RMSE and n for user-based like Figure 3 and item-based like Figure 4.

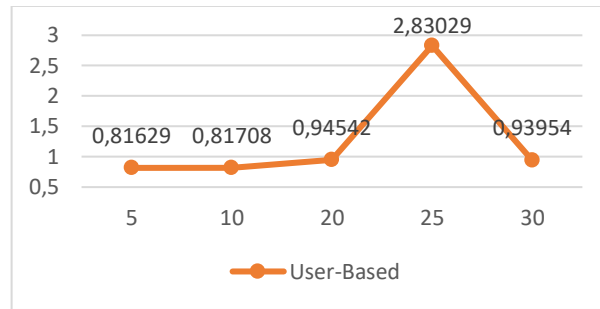


Figure 1. RMSE for User-Based

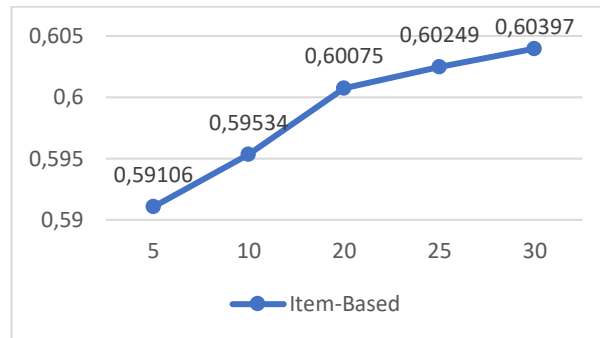


Figure 4. RMSE for Item-based

The better RMSE value is a value close to 0, so the best RMSE value on item-based when topN = 5 with an RMSE value of 0.5911. For user-based, the best RMSE value is when topN = 5 with an RMSE value of 0.81629. Furthermore, each topN that produces the best RMSE value will be selected to calculate the prediction rating and used for the classification process using 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 prediction rating of the best top-n value. As in table 9 the user-based dataset and item-based dataset in table 10.

Table 9. Matrix Dataset User-based

User Film	Anak Nonton	Cenayang Film	....	Sir_ amirsyarif
#Alive	3,173	2,780	....	2,802
....	....	....	....	....
Young Adult	2,500	2,500	....	3,920

Table 10. Matrix Dataset Item-based

Film User	#Alive	17 Again	...	Zombieland
Elbert_Reyner	4,000	2,250	...	4,000
....	....	....	...	....
IMDB	3,150	3,200	...	3,800

### 3.3 SVM Classification Result

The dataset used in this classification process by changing the 0-5 rating value 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. A value of 1 can be assumed that the user likes the movie and a value of 0 can be assumed that the user does not like the movie. As in table 11, the values become 0 and 1 for user-based and table 12 for item-based.

Table 11. Dataset Klasifikasi SVM User-based

User Film	Anak Nonton	Cenayang Film	....	Sir_ amirsyarif
#Alive	1	0	....	1
....	....	....	....	....
Young	0	0	....	1

Adult				
Table 12. Dataset Klasifikasi SVM Item-based				
Film User	#Alive	17 Again	....	Zombieland
Anak Nonton	1	0	....	1
....	....	....	....	....
Cenayang Film	0	0	....	1

After the dataset has been converted to 0 and 1, it will then be processed into the SVD algorithm and SVM classification. In SVM classification, hyperparameter tuning is done using grid search. For the SVM data process can be seen in table 13.

Table 13. SVM on User-based

	Test size	Random State	C	Gamma	Kernel	Average Precision	Average recall
User-based	0,2	11	1	scale	rbf	0,8136	0,8771
Item-based	0,1	11	1	scale	rbf	0,7755	0,8637

Hyperparameter tuning using grid search will find the best parameter value to optimize the value of precision and recall. As shown in Table 14, the hyperparameter tuning steps using grid search.

Table 14. Hyperparameter tuning on User-based and Item-based

	Test size	Random State	C	Gamma	Kernel	Average Precision	Average recall
User-based	0,2	11	0,1	1	rbf	0,8158	0,8158
Item-based	0,1	11	0,1	1	poly	0,8062	0,8062

After performing hyperparameter tuning, it can be seen that the best parameters for user-based data are at C = 0.1, Gamma = 1, and Kernel = rbf with the same precision and recall values with an optimal increase after hyperparameter tuning on the precision value. For item-based data, the best parameters are C = 0.1, Gamma = 1, and Kernel = poly with an optimal increase after hyperparameter tuning on the precision value.

#### 4. CONCLUSION

In the research we have done to develop a collaborative filtering method that has been combined with the Singular Value Decomposition (SVD) algorithm and Support Vector Machine (SVM) classification, this research uses crawling datasets from Twitter combined with IMDb, Rotten Tomatoes, and Meta-critic web rating reviews. Then the data is processed to produce movie ratings. Collaborative filtering in rating prediction using user-based and item-based gets the best topN value at user-based topN = 5 with an RMSE value of 0.8162 and item-based topN = 5 with an RMSE value of 0.5910. 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 with the best parameters, namely C = 0.1, Gamma = 1, Kernel = rbf getting for user-based precision value 81% and recall 81%. For item-based with the best parameters, namely C = 0.1, Gamma = 1, Kernel = poly, the precision value is 80% and the recall value is 80%. It can be concluded that the combination of collaborative filtering and the Support Vector Machine (SVM) classification algorithm can be used to make movie recommendations on whether the movie can be recommended or not and by 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 dataset. In addition, in the future, the recommendation system can be created in combination with other methods such as content-based and other classification algorithms.

#### REFERENCES

- [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. Mohamadrezai, "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] 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.
  - [4] H. D. Februariyanti, "IMPLEMENTASI METODE COLLABORATIVE FILTERING UNTUK SISTEM Diterima : Diterbitkan :," vol. IX, no. I, pp. 43–50, 2021.
  - [5] S. Zheng, C. Ding, and F. Nie, "Regularized Singular Value Decomposition and Application to Recommender System," no. 3, 2018, [Online]. Available: <http://arxiv.org/abs/1804.05090>
  - [6] 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.knsys.2017.11.010.
  - [7] 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.
  - [8] 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.
  - [9] V. Subramaniaswamy, R. Logesh, M. Chandrashekar, 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.
  - [10] P. Thakkar, K. Varma, V. Ukani, S. Mankad, and S. Tanwar, *Collaborative Filtering Using Machine Learning*. Springer Singapore, 2019. doi: 10.1007/978-981-13-1747-7.
  - [11] G. Karypis, J. Konstan, and J. Riedl, "Incremental Singular Value Decomposition Algorithms for Highly Scalable Recommender Systems Dimensionality Reduction for," *Science (80- )*, vol. 81, no. 4, pp. 717–733, 2015.
  - [12] N. Samaiya, S. K. Raghuvanshi, and R. K. Pateriya, *Shilling attack detection in recommender system using PCA and SVM*, vol. 813. Springer Singapore, 2019. doi: 10.1007/978-981-13-1498-8\_55.
  - [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] I. Monika Parapat and M. Tanzil Furqon, "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: <http://j-ptiik.ub.ac.id>
  - [15] A. R. D. Pratiwi and E. B. Setiawan, "Implementation of Rumor Detection on Twitter Using the SVM Classification Method," no. 10, pp. 782–789, 2021.
  - [16] 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.
  - [17] 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.
  - [18] 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.
  - [19] 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.