JOIV : Int. J. Inform. Visualization, 5(4) - December 2021 448-455

INTERNATIONAL JOURNAL ON INFORMATICS VISUALIZATION

journal homepage : www.joiv.org/index.php/joiv

Neural Collaborative with Sentence BERT for News Recommender System

Budi Juarto [a,b,*], Abba Suganda Girsang [b]

[a] Computer Science Department, School of Computer Science, Bina Nusantara University, Jakarta, Indonesia
[b] Computer Science Department, Master of Computer Science, Binus Graduate Program, Bina Nusantara University, Jakarta, Indonesia
[*] Corresponding author: budi.juarto@binus.ac.id

***Abstract*— The number of news produced every day is as much as 3 million per day, making readers have many choices in choosing news according to each reader's topic and category preferences. The recommendation system can make it easier for users to choose the news to read. The method that can be used in providing recommendations from the same user is collaborative filtering. Neural collaborative filtering is usually being used for recommendation systems by combining collaborative filtering with neural networks. However, this method has the disadvantage of recommending the similarity of news content such as news titles and content to users. This research wants to develop neural collaborative filtering using sentences BERT. Sentence BERT is applied to news titles and news contents that are converted into sentence embedding. The results of this sentence embedding are used in neural collaboration with item id, user id, and news category. We use a Microsoft news dataset of 50,000 users and 51,282 news, with 5,475,542 interactions between users and news. The evaluation carried out in this study uses precision, recall, and ROC curves to predict news clicks by the user. Another evaluation uses a hit ratio with the leave one out method. The evaluation results obtained a precision value of 99.14%, recall of 92.48%, f1-score of 95.69%, and ROC score of 98%. Evaluation measurement using the hit ratio@10 produces a hit ratio of 74% at fiftieth epochs for neural collaborative with sentence BERT which is better than neural collaborative filtering (NCF) and NCF with news category.**

***Keywords*— Recommender system; news; neural networks; sentence BERT, neural collaborative filtering.**

Manuscript received 20 Sep. 2021; revised 28 Oct. 2021; accepted 22 Nov. 2021. Date of publication 31 Dec. 2021.

I. INTRODUCTION

Information technology has an incredibly significant role in everyday life. The role of information technology has attracted the attention of experts and researchers because the number of internets and social media users has increased from year to year. In January 2021, the Global digital population has 4.66 billion active internet users, which takes around 59.5 percent of the global population. The number of active internet users who access the internet via mobile phones is around 93.2 percent, or around 4.32 billion active internet users [1].

The online news portal The Wall Street Journal has an average circulation of close to 3 million per day. The data used is the news portal The Wall Street Journal, which focuses on business, economics, and politics. Data from 2018 to 2020 shows that there are 754,000 newspapers printed every day, and online readership reaches 2.26 million people [2]. This large number of online visitors would be assisted in selecting the desired news with a news recommendation system to assist in providing certain news recommendations.

There are two types of recommendation systems commonly used, namely Collaborative Filtering (CF) and Content-Based Filtering (CBF). Both recommendation systems are used depending on the design used to recommend items to the user. Collaborative filtering recommends something to a user based on the interests and preferences of other users who have the same preferences. Collaborative filtering has drawbacks. The data has a high sparsity because the user cannot ratio all items if there are a lot of them. Content-based filtering provides new recommendations based on content in providing recommendations. For example, there are titles, categories, and news content in the news. This recommendation system provides recommendations on the user's suitability with the content that the item has. Content-based filtering has a weakness for providing unexpected recommendations such as news content categories that users have previously read.

Collaborative filtering usually uses a neural collaborative method, but this method has a disadvantage in capturing the user preferences from the category and item content. Therefore, this study implements a sentence BERT in neural collaborative filtering that recommends based on the content of item too likes news title and news body.

A. Related works

Related Works consist of three parts: recommendation system with content-based filtering, recommendation system with collaborative filtering, and hybrid recommender system that combines both content-based filtering and collaborative filtering. Putri *et al.* [3] researched a recommendation system for exclusive pen products using content-based filtering in 2019. This study used 258 product codes, consisting of eight categories and thirty-three keywords. The method used in this research is TF-IDF weighting. This study used nine parameters: price, gender, clip color, usability, name carving, body, age, color, and occupation. The evaluation used in this research is 100 system test data and interviews and a questionnaire. The accuracy obtained from this study is 96.5% of data collection from 61 women and 39 men.

Noorhidayah *et al.* [4] researched news recommendations using TF-IDF weighting and cosine similarity in 2019. The dataset used in this research is the news dataset from the Radar Banjarmasin website. The data set was collected from 50 news data from January to March 2018. The evaluation carried out in this study was that the precision achieved in this study was 76%.

Wahyudi *et al.* [5] studied recommendation systems in hotels using content-based filtering in 2017. This study uses the features of the category as well as the city. The dataset used is the Kaggle dataset which consists of 10,000 lines. The evaluation obtained was a precision of 88.9%, an accuracy of 85%, and a recall of 80%.

Abraham *et al.* [6] researched news recommendation systems using the KNN method in 2017. The dataset used is news with three categories: political news, international news, and business economic news. In this study, an evaluation was carried out using RMSE. The results of the study obtained an RMSE value of 14%. After several iterations, the researcher found articles with recommendations that did not meet the specified KNN method limits requirements.

Nastiti [7] researched implementing a content-based filtering recommendation system in food crops in 2020. The data used in this research is 1000 agricultural land data consisting of the following information: farmer groups, types of food crops, varieties, planting dates, harvest dates, location (longitude, latitude), yields. The evaluation measure compares farmer groups that are relevant to traders, namely farmer groups selected by traders with farmer groups as a result of system recommendations. Evaluation of 10 profiles with 15 recommendations from the top farmer groups obtained an average precision of 78.40%.

Kusuma *et al.* [8] researched the recommendation system in selecting a special thesis topic for the Computer Science department at Gadjah Mada University in 2021. The datasets used in this study are datasets from lecturers' publications in the last three years and the UGM Computer Science course syllabus. From the dataset used, the word embedding was carried out using TFIDF, and similarity measurements were carried out using Euclidian distance. The evaluation carried out in this study is to measure the accuracy of the questionnaire with four aspects, namely relevance 77.33%, novelty 90.67%, serendipity 80%, and increasing recommendation diversity 84% of the data from the questionnaire given to respondents. In addition, time measurements were also carried out in displaying recommendations, with an average time of 7.26 seconds.

Parwita *et al.* [9] conducted research on content-based filtering that focused on testing accuracy in 2019. The recommendations given are supervisors based on published data that previous lecturers have done. This study focuses on the use of stop words to see the effect of using stop words and not using stop words on the evaluation values of precision, recall, and f-measure. Experiments also use a minimum similarity of 5% to 45%. This study concludes that the value of the recommendation system with the stop word removal process is still superior to the recommendation system without the stop word removal process.

Girsang *et al.* [10] researched recommendation system journalists for getting top news based on Twitter with content-based filtering in 2019. The data used in this recommendation system is comment tweets for users and search by similarity. The similarity is measured with cosine similarity. The group is formed when the cosine similarity scores more than 0.7. Data training use 24.583 records with ten categories. The method uses SVM for the classification category. The accuracy classification is 80% in this research.

B. Recommendation System with Collaborative Filtering

Wang *et al.* [11] conducted item-based collaborative filtering research using BERT in 2020. This research was conducted on an e-commerce case study by an employee of eBay. The data used were 8 million pairs of items. The feature used in this research is the name of the product from e-commerce. The results obtained are precision @ 10 is 0.079 and recall @ 10 is 0.791.

Girsang *et al.* [12] conducted research on anime film recommendations with collaborative filtering in 2020. The dataset used is a general dataset from Kaggle, consisting of 73,516 users and 12,294 movie titles. This study uses the alternating least squares method and obtains the root mean square error of 2.53. This research has the advantage of being able to provide movie recommendations to a user on the recommendations of other similar users.

Prayogo *et al.* [13] conducted research on the movie lens rating dataset using collaborative filtering in 2020. The dataset used is movie lens data which consists of 100,000 ratings by 610 users on 9742 films. This collaborative filtering method would predict the rating with Matrix Factorization and K-Nearest Neighbor. The evaluation used in this research is using MAE and RMSE. The Matrix Factorization method obtained MAE = 0.6371 and RMSE = 0.8305, while the K-Nearest Neighbor method obtained MAE = 0.6742 and RMSE = 0.8863. So this study shows the results of Matrix Factorization are better than K-Nearest Neighbor.

Yulian *et al* [14] conducted research on correlation-based similarity without using a rating in 2015. The dataset used in this study is news which does not have a rating. News would be labeled 1 if read and 0 if not read. So that a special correlation-based similarity formula was changed by using binary numbers. The result of the evaluation in this study is the speed of computing. The new formula for specific correlation-based similarity gets faster times than the general correlation-based similarity formula.

Priyono *et al.* [15] conducted research on recommendation systems using collaborative filtering and a priori algorithms in 2017. The dataset used is tourism data in Yogyakarta based on the rating given. The evaluation carried out in this study was by using precision and recall. This study's precision and recall ranged from 60% to 67% of the ten recommended tourism items.

Kurniawan *et al.* [16] conducted research on a recommendation system on shoe products using collaborative filtering in 2016. The dataset used in this study is a dataset of 6 users who give a rating to 6 shoes. The similarity of each item is measured by using the adjusted cosine similarity of items. Evaluation in this study using MAE and obtained a value of 0.9568.

Muliadi *et al.* [17] conducted research on a recommendation system for places to eat using the Typicality Based Collaborative Filtering Algorithm in 2019. The dataset used in this study is the distance from the user to the nearest food stall by using the google maps API and also using the rating from the user to the restaurant the user has visited and rated this distance data is used a weighting in determining the similarity of each user and also performs rating prediction of each user. The evaluation result of this research is using the mean absolute error with a value of 1.366. The error value in this study is still high because the data sparsity is still high where there are still many users who have not researched the item

Salakhutdinov *et al.* [18] conducted research on a film recommendation system using Collaborative Filtering in 2007. This research develops on collaborative filtering, which is not good for large datasets. So this research uses a neural network in the form of two-layer undirected graphical models called Restricted Boltzmann Machines. The dataset used is 100 million user/movie ratings data which can beat the current Netflix system with an error ratio of 6% better than SVD models, which lower RMSE by 0.005 from 0.9514.

Indriawan *et al.* [19] conducted research on the sale of agricultural products using item-based collaborative filtering in 2020. The dataset used is agricultural products displayed on the mobile application. This research focuses on the development of mobile applications where the user can rate each item. Recommendations are given from cosine similarity and recommend items with a similarity level above the similarity value of 0.7.

C. Recommendation System with Hybrid Method

Wairegi and the team conducted research on a dataset of articles obtained in Kaggle using a hybrid approach method that combines content-based and collaborative filtering in 2020 [20]. The results obtained by the hybrid method have a recall performance of 0.54, so it is better than the collaborative filtering method with a value of 0.47 or content-based with a value of 0.5.

Wijaya *et al.* [21] conducted research on a recommendation system for laptop sales using a hybrid consisting of collaborative filtering and content-based filtering in 2018. Collaborative filtering uses item ratings, while content-based filtering uses item names and uses the TF-IDF method. The evaluation in this study focuses on the execution time of the recommendation system. The result obtained is that content-based filtering has a faster execution time than collaborative filtering and mixed hybrid

Liu *et al.* [22] conducted research by combining content-based filtering and restricted Boltz man-machine in 2014. This study uses features in the form of demographic information and item categories. However, it is not mentioned in the content-based research using tf-idf or word2vec. Naive Bayes is used to filling in the blank data from the user-item rating matrix. The results obtained using the movie dataset are obtained a Mean Absolute Error (MAE) of 0.674.

Fitrianti *et al.* [23] researched a scientific article recommendation system by combining content-based and collaborative filtering in 2020. The dataset used in this study is a dataset of 100 articles that are entered into the website. Testing and evaluation in this study involved 35 users. After using the recommendation system, everyone would do a rating. The data is then divided into 60 training and 40 testing data. The evaluation was carried out using the mean absolute error, and the result was 0.85.

II. MATERIAL AND METHOD

The research method used in this research is to use neural collaborative filtering for implicit data [24]. This research develops neural collaborative filtering using news content such as news titles, news categories, and news content. In the news titles and news categories, sentence embedding is carried out using the BERT sentence [25] before entering the neural collaborative architecture. As for the user id, item id, and category would enter neural collaborative filtering after embedding it into a latent factor.

Sentence BERT (SBERT) is a modification of BERT using Siamese and triplet networks. Sentence BERT is typically used for large-scale semantic similarity comparison, clustering, and information retrieval via semantic search. The Siamese network of SBERT creates fixed-sized vectors for input sentences can be derived. SBERT can compute sentence embeddings in 5 seconds and compute with cosine similarity in 0.01 seconds to find 10,000 similar sentence pairs, while BERT takes 65 hours.

The SBERT model adds a pooling operation from the BERT output to derive a fixed-sized sentence embedding. The pooling used in the BERT sentence is MEAN pooling. The architecture of sentence BERT is shown in Figure 1. In this study, we did not see the similarity between the two sentences. However, we look at the mean pooling stage of the BERT sentence so that the resulting output is a sentence embedding from the mean pooling results.

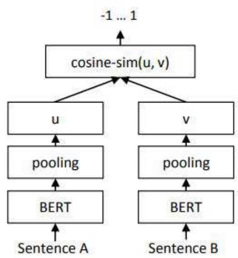


Fig. 1 SBERT architecture for similarity scores.

Sentence BERT has various pre-trained models and has tested the performance and processing time of 5 tasks, including the STS benchmark test set, duplicate question test set from Quora and Sprint, Tweet paraphrase test set, and similar scientific publications. This study used a pre-trained model with the highest average performance, called paraphrase-mpnet-base-v2. The other option to choose a model with faster processing and high quality is the pre-trained model paraphrase-MiniLM-L6-v2. The results of the comparison of models can be seen in Table I.

TABLE I
NEURAL COLLABORATIVE + SBERT ARCHITECTURE

	paraphrase-mpnet-base-v2	paraphrase-MiniLM-L6-v2
Speed	2800	14200
Average Perfomance	76.84	74.81

In the user dataset, there is data on user interactions with items. If there is interaction from the user with the item, then the interaction has a value of one. Meanwhile, if there is no interaction, then the interaction value is 0. User and item interaction is symbolized by y, which is shown in equation 1.

$$y = \begin{cases} 1, \text{if interaction (user, item) is observed;} \\ \quad 0, otherwise \end{cases} \tag{1}$$

The value of one from y means an interaction between user and item, but it does not mean user likes item. Likewise, a value of 0 does not mean the user does not like the item because it can mean that the user is unaware of the item and interacts with it.

It can be seen from Figure 2 that the output of one layer would be the input of the next layer. In this neural collaborative with Sentence BERT method using three feature vectors and one-sentence embedding, such as user latent vector, item latent vector, and category latent vector, the transformation has been carried out into a binary sparse vector one-hot encoding. The sentence embedding used is the BERT sentence in the title and content of the news.

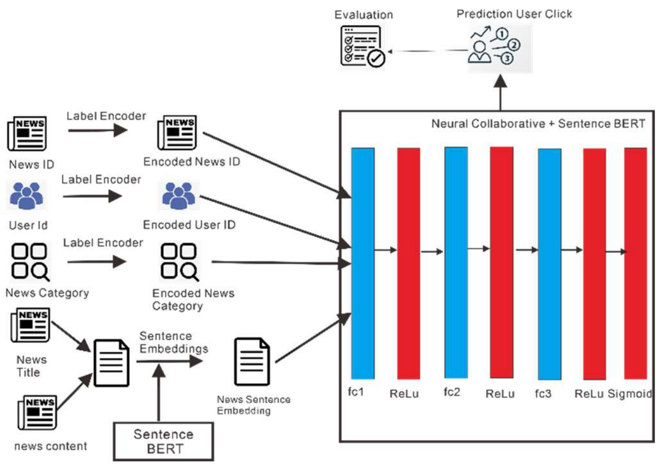


Fig. 2 Neural Collaborative with SBERT Research Flow

Fig. 2 shows the framework of the neural collaborative described. It can be seen that the input layer is the embedding layer. User embedding, item embedding, category embedding, and news sentence embedding are entered into a multi-layer neural architecture called the collaborative filtering layer to map latent vectors to produce score prediction results. Each layer of neural collaborative filtering can be customized to find the latent structure of user-item interactions. We used three connected layers that connect from the first fully connected layer to the third fully connected layer. The architecture combination of a neural collaborative with the embedding of news categories and sentence embedding on the title and news content with SBERT is shown in Figure 3.

Probabilistic treatment for neural collaboration with sentences BERT makes recommendations with implicit feedback using a binary classification problem. For negative instances, samples are usually taken from unobserved interactions. The number of ratio comparisons between the observed ratios varies depending on the researcher. The number of comparisons shows results that start well from one observed interaction compared to four unobserved interactions. This study used four unobserved interactions.

The model has hyperparameters that can be set, such as batch size, epochs, learning ratio, and optimizer. In this study, the model uses linear regression, and the loss function is mean square error. In this study, we use batch size = 512, epochs = 5, learning ratio = 0,01, and ADAM optimizer. The neural collaborative architecture used in this study is shown in table II.

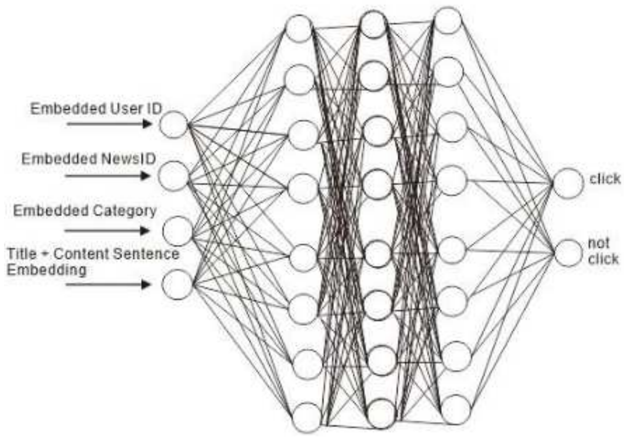


Fig. 3 Neural Collaborative with SBERT Architecture

TABLE II
NEURAL COLLABORATIVE + SBERT ARCHITECTURE

	Name	Type	Params
1	user embedding	Embedding	799.0k
2	Item embedding	Embedding	757.0k
3	Item cat embedding	Embedding	272
4	Fc1	Linear	104.0k
5	Fc2	Linear	8.3k
6	Fc3	Linear	2.1k
7	Output	Linear	33

The data used in this study consist of 49.981 users and 47.313 news, with 5.475.542 interactions between users and news. Before the data can be trained evaluated, there are some pre-processing data for each evaluation metric. Evaluation metric precision, recall, and receiver operating characteristic (ROC) would use label 1 for observed interaction and 0 for unobserved interaction. The data would be split into eighty percent for training and twenty percent for testing. The amount of training data used is 4,380,434 data, and testing data uses 1,095,108 data. Neural collaboration with sentence

BERT would predict the interaction user for each item and evaluate it with the testing label. Figure 4 shows the training and evaluation process using precision, recall, F1-score, and ROC. The data would be split into ratios for each observed's hit ratio evaluation metric. There would be four unobserved items in the training step. The train data and test data use leave one out cross-validation. We choose the latest news user to interact with the user in this case. To evaluate hit ratio @10 in testing data, each user randomly selects 99 items that the user has not interacted with or called unobserved items. Then these 99 items combine with the test items, and then there would be 100 items. After that, the model runs on these 100 items and ranks them to their predicted probabilities. After that, only the top 10 items are selected, and if the test item is in the top 10 items, which means a hit.

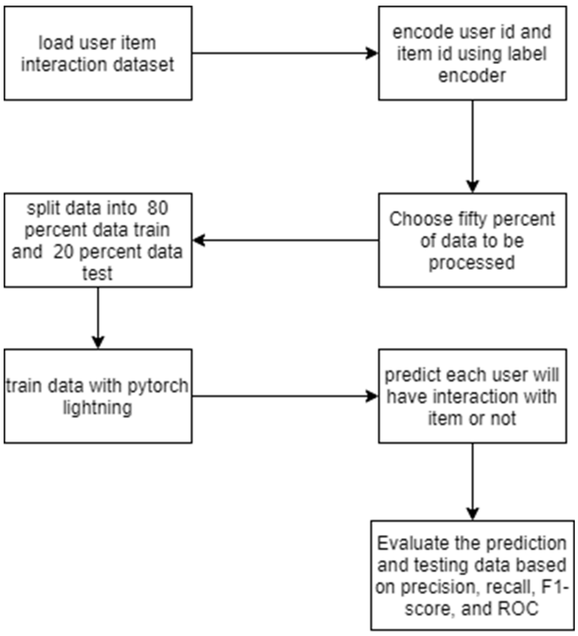


Fig. 4 Training and Evaluating Step for accuracy, precision, recall, F1-score, and ROC

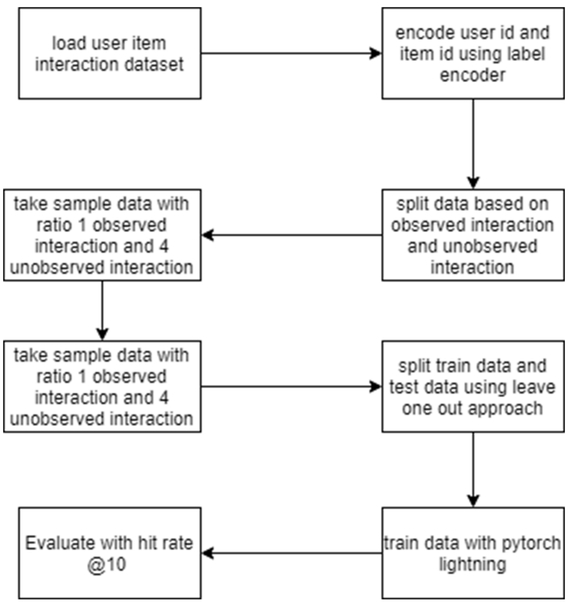


Fig. 5 Training and Evaluating Step for hit ratio@10

III. Result and Discussion

The recommendation system has several evaluation techniques depending on the data generated by user interactions with items. The resulting data can be in the form of implicit data and explicit data. Implicit data is binary data with a value of 1 or 0, where 1 symbolizes the interaction between users and items such as reading, viewing, etc. 0 represents no interaction from the user with the item. While explicit data is data taken directly from users, such as ratings or questionnaires. Explicit data usually has a specific range of one to five in the value of user interactions with items. The evaluation uses precision, recall, F1-score, and hit ratio using implicit data. The explicit data use RMSE and MAE evaluation techniques. The hit ratio evaluation uses a leave-one-out technique where every user who interacts with the last item becomes the testing data. Hit ratio (HR) is a recall-based metric, where HR does not pay attention to the hit position. In this research, data using implicit data, the evaluation carried out in this research uses precision, recall, F1-score, and hit ratio. The ROC curve is produced by calculating and plotting the true positive ratio (TPR) against a single classifier's false-positive ratio (FPR). The Equations are shown in formulas 8 to 13 and the confusion matrix in figure 6. We can evaluate the result of neural collaborative filtering with sentence BERT.

		Predicted Condition	
		Postive (PP)	Negative (PN)
Actual Condition	Positive (P)	True Positive (TP)	False Negative (FN)
	Negative (N)	False Positive (FP)	True Negative (TN)

Fig. 6 Confusion Matrix

$$\text{Precision} = \frac{TP}{TP+FP} \quad (8)$$

$$\text{Recall} = \frac{TP}{TP+FN} \quad (9)$$

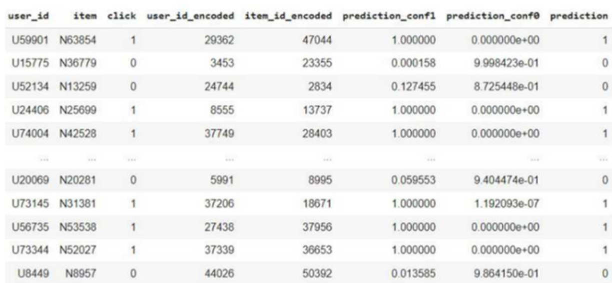
$$\text{F1} = 2x\frac{precision\ x\ recall}{precision+recall} \quad (10)$$

$$\text{TPR} = Sensitivity = \frac{TP}{TP+FN} \quad (11)$$

$$\text{FPR} = 1 - Specifity = \frac{FP}{FP+TN} \quad (12)$$

$$\text{Hit Ratio} = \frac{number\ of\ hits\ in\ test}{number\ of\ user} \quad (13)$$

In the evaluation using precision, recall, and ROC prediction, results of the model have a value of zero to one so that if the predictive value is more than 0.5, it would be labeled as one, while if the predicted value is less than 0.5 it would be labeled zero. An example of the results of prediction with the model is shown in Figure 7.



user_id	item	click	user_id_encoded	item_id_encoded	prediction_conf1	prediction_conf0	prediction
U59901	N63854	1	29362	47044	1.000000	0.000000e+00	1
U15775	N36779	0	3453	23355	0.000158	9.998423e-01	0
U52134	N13259	0	24744	2834	0.127455	8.725448e-01	0
U24406	N25699	1	8555	13737	1.000000	0.000000e+00	1
U74004	N42528	1	37749	28403	1.000000	0.000000e+00	1
...	...	...	...	...	...	...	...
U20069	N20281	0	5991	8995	0.059553	9.404474e-01	0
U73145	N31381	1	37206	18671	1.000000	1.192093e-07	1
U56735	N53538	1	27438	37956	1.000000	0.000000e+00	1
U73344	N52027	1	37339	36653	1.000000	0.000000e+00	1
U8449	N8957	0	44026	50392	0.013585	9.864150e-01	0

Fig. 7 Example of Prediction Results from Model

We can calculate precision, recall, and create a ROC graph based on this model. The results of precision, recall, and ROC is in Figure 8. The results show a good value where all precision and recall are above 90%. The lowest recall is on label one at 92%, and precision on label 0 is 93%. Figure 9. shows the results of the obtained confusion matrix, Figure 10. shows the results of the obtained precision-recall curve, and Figure 11. shows the ROC curve results obtained.

```
              precision    recall  f1-score   support

           1       0.99      0.92      0.96    534399
           0       0.93      0.99      0.96    560709

    accuracy                           0.96   1095108
   macro avg       0.96      0.96      0.96   1095108
weighted avg       0.96      0.96      0.96   1095108
```

Fig. 8 Prediction Results from Model

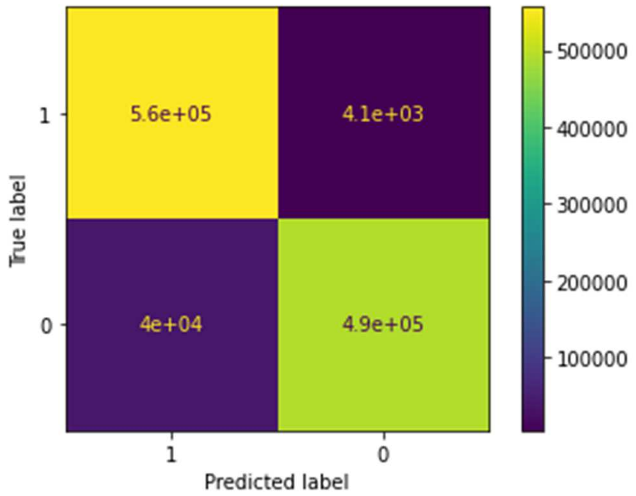


Fig. 9 Confusion Matrix Result

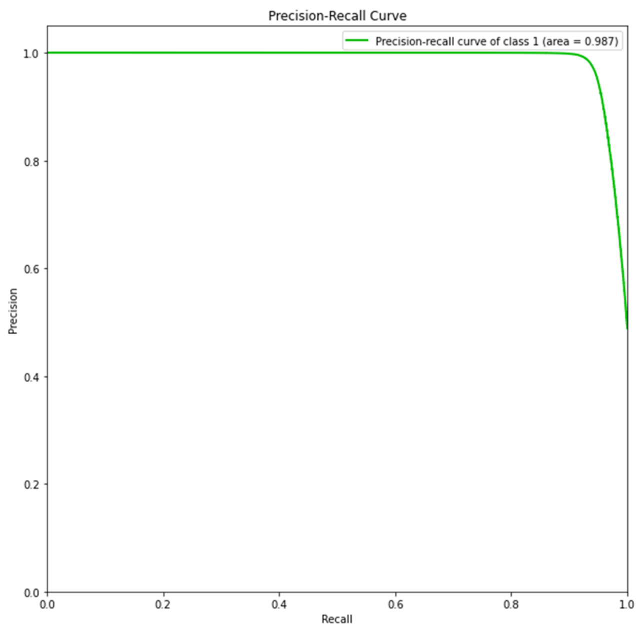


Fig. 10 NCF+SBERT Precision-Recall Curve Result

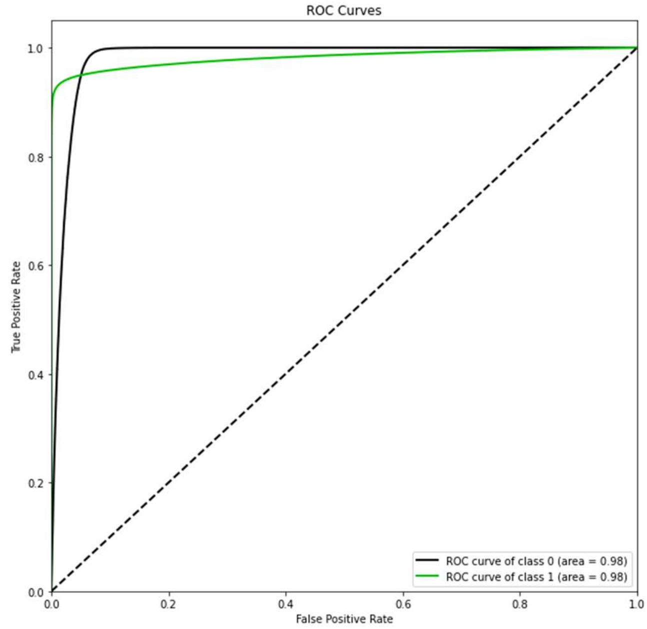


Fig. 11 NCF+SBERT ROC Curve Result

The comparison and analysis of the results of the Neural Collaborative Filtering with Sentence BERT (SBERT) recommendation model, the results obtained would be compared with the Neural Collaborative Filtering method. The Neural Collaborative model used only uses a user dataset that only sees the relationship between the user id, item id, and their interaction whether they read or not news. The first comparison was carried out to evaluate precision, recall, f1-score, and ROC score. Based on Table III, it can be seen that the comparison of a neural collaborative with neural collaborative using sentences BERT. The results show not much difference between the two methods using the precision, recall, and f1-score evaluation techniques. The addition of epochs in training only makes precision and recall a trade-off where if recall increases, precision would decrease and vice versa. So that in this study, only ten epochs were used in the study. The ROC curve also does not show much difference from the hybrid model compared to the Neural Collaborative. The Neural Collaborative Filtering (NCF) ROC results are shown in Figure 12.

TABLE III
NCF+SBERT VS NCF PRECISION, RECALL, F_1, AND ROC COMPARISON

Evaluation Metric	Neural Collaborative	Neural Collaborative + S BERT
Accuracy	95.90%	95.83%
Precision	92.31%	92.69%
Recall	99.21%	98.61%
F1	95.64%	95.56%
ROC	98.00%	98.00%

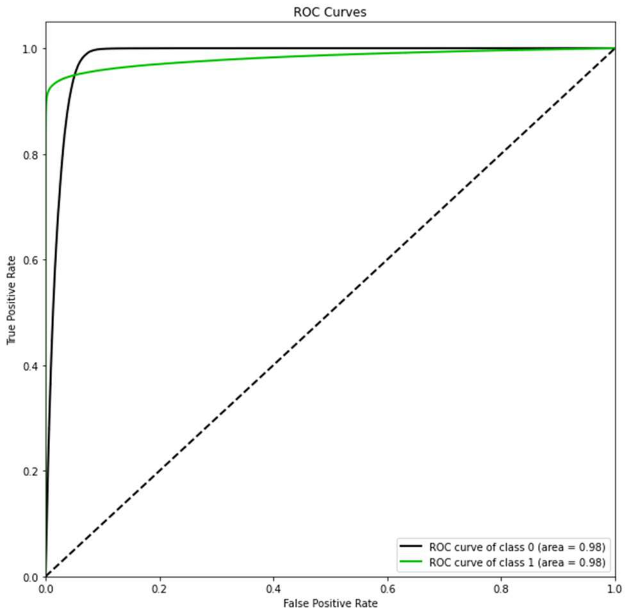


Fig. 12 NCF ROC Curve Result

In the evaluation using the hit ratio on ten items. Each Neural Collaborative method and the hybrid neural collaborative method with BERT sentences were carried out with ten, thirty, and fifty epochs. The results of each method and epoch are shown in Table IV.

TABLE IV
NCF+SBERT VS NCF VS NCF+NEWS CATEGORY HIT RATIO@10 COMPARISON

Epoch	Neural Collaborative	Neural Collaborative + S BERT+News Category	Neural Collaborative + News Category	Neural Collaborative + S BERT
10	64%	56%	62%	57%
30	63%	71%	67%	72%
50	61%	72%	69%	74%

Table IV. shows the hit ratio results of both neural collaborative and neural collaborative methods with sentences BERT. The results show that the neural collaborative method with the sentence BERT results better than the neural collaborative at epochs 30th and 50th. In the 30th epoch, the hit ratio value obtained in the neural method with the BERT sentence obtained a hit ratio value of ten items of 71%. The hit ratio value has not increased significantly from epoch 30th to epoch 50th because the hit ratio value produced is not much different. Meanwhile, increasing the number of epochs in the neural collaborative method did not increase the hit ratio obtained. NCF method that uses News Category shows an improvement for hit ratio, but the hit ratio result is still lower than Neural Collaborative Filtering with sentence BERT.

The results of neural collaborative filtering combined with sentences BERT and categories give a smaller hit ratio than neural collaborative filtering combined with sentences BERT alone. Categories can provide better recommendations to users, but the title and content of the news can provide much better recommendations because it is more detailed in sentence embedding. Each reader in the Microsoft News dataset reads many news categories with different amounts of each news category to affect the hit ratio.

So, it can be concluded that the hybrid neural collaborative method using sentences BERT can increase the hit ratio of ten items, and the use of news content in the form of news titles and news abstracts can improve the performance of the neural collaborative model. To prove it, there should be some topic relevance from users reading news on train data and test data. In this case, try to see the topic of words read by user-id U59901. The word cloud visualization is selected from an interaction with a value of one or an interaction between the user and the item, or the user has read the news id. In the train data, the user reads eleven news categories and nine news categories in the test data. The results of the visualization of the word cloud of the training data news title are shown in Figure 13, and the word cloud of the testing data news title is displayed on the testing data shown in Figure 14.

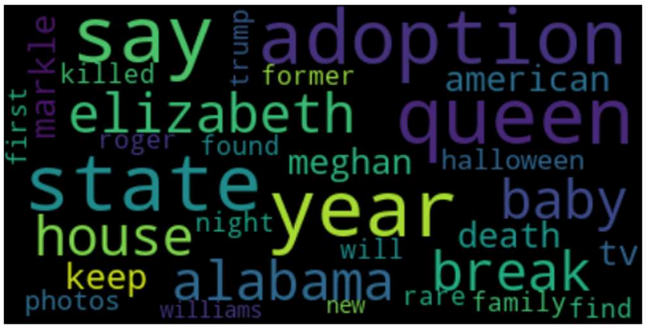


Fig. 13 Sample News Title Word Cloud Training Data

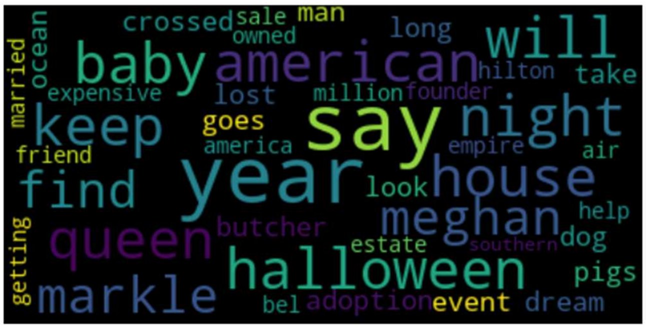


Fig. 14 Sample News Title Word cloud Testing Data

Figure 13 and Figure 14 have some similar words such as year, American, queen, and adoption, where some of the topics that appear in the train appear in the test data so that it shows a relationship between reading topics from training data or previous reading news with testing data or reading news that would be selected next.

IV. CONCLUSION

This study developed a hybrid recommender system model by combining neural collaborative filtering using BERT sentences. The pre-trained model in the BERT sentence used is paraphrase-mpnet-base-v2 because it has the highest average performance among other pre-trained models in the BERT sentence. Pre-trained is used to change the title and content of the news into a sentence embedding in the form of a sentence vector. So that the neural collaborative model would be developed into a hybrid model by combining neural collaborative with content from news, namely news categories that are changed with encoder labels and sentence vectors from news titles and contents. This model would then be evaluated using precision, recall, ROC, and hit ratio. The evaluation of the neural collaborative with the sentence BERT was then compared to the neural collaborative method. The results show that the neural collaborative method using sentences BERT produces precision, recall, and ROC curve values that are not much different from neural collaborative. While the evaluation using the hit ratio on ten items resulted in a higher value than neural collaborative. In the thirtieth epoch, the hybrid method has a hit ratio value of 72% compared to neural collaborative with 63%. The fiftieth epoch also shows that the value of the hybrid method has a hit ratio value of 74% compared to neural collaborative, which gives a value of 61%.

Future research can be developed from several things, such as using different sentence embedding or using different pre-trained models from sentences BERT by considering performance and time. In this study, hyper-parameter tuning has not been carried out. So that in the following research it can perform hyperparameter tuning of PyTorch lightning such as the Ray library using the available dragonfly algorithm or other metaheuristic methods. The evaluation using precision and recall has given good results. Future research can also fine-tune the precision-recall trade-off. Future research can also be applied to use datasets that read similar categories and do not read many categories because it will make sentence embedding more challenging to provide good recommendations with various reading topics.

REFERENCES

[1] Watson, A., (2020). Average circulation of the Wall Street Journal from 2018 to 2020. Retrieved from Statista. [Online]. Available: https://www.statista.com/statistics/193788/average-paid-circulation-of-the-wall-street-journal/.

[2] Joseph, Johnson., (2021). Worldwide digital population as of January 2021. Retrieved from Statista. [Online]. Available: https://www.statista.com/statistics/617136/digital-population-worldwide/ .

[3] Putri, M. W., Muchayan, A., & Kamisutara, M. (2020). Sistem Rekomendasi Produk Pena Eksklusif Menggunakan Metode Content-Based Filtering dan TF-IDF. *JOINTECS (Journal of Information Technology and Computer Science)*, *5*(3), 229-236..

[4] Noorhidayah, N., Indriani, F., & Faisal, M. R. (2019). Sistem Rekomendasi Berita Online dengan menggunakan Pembobotan TF-IDF dan Cosine Similarity. *Soliter*, *2*, 85-93.

[5] Wahyudi, Kristian & Latupapua, Johanes & Chandra, Ritchie & Girsang, Ganda. (2020). Hotel Content-Based Recommendation System. Journal of Physics: Conference Series. 1485. 012017. 10.1088/1742-6596/1485/1/012017.

[6] Abraham, Sirajuddin, & Yeni Dwi Rahayu. (2017). Sistem Rekomendasi Artikel Berita Menggunakan Metode Knearest Neighbor Berbasis Website. Prosiding Sensei 2017 1, no. 1

[7] Nastiti, P. (2019). Penerapan Metode Content Based Filtering dalam Implementasi Sistem Rekomendasi Tanaman Pangan. Teknika, 8(1), 1-10.

[8] Kusuma, Hans Satria, and Aina Musdholifah. (2021).Recommendation System for Thesis Topics Using Content-based Filtering. *IJCCS (Indonesian Journal of Computing and Cybernetics Systems)* 15, no. 1: 65-74.

[9] Parwita, W. G. S. (2019). Pengujian Akurasi Sistem Rekomendasi Berbasis Content-Based Filtering. Inform. Mulawarman J. Ilm. Ilmu Komputer, 14(1), 27..

[10] Girsang, A. S., Isa, S. M., & Harvy, I. (2021, April). Recommendation System Journalist For Getting Top News Based On Twitter Data. In *Journal of Physics: Conference Series* (Vol. 1807, No. 1, p. 012006). IOP Publishing.

[11] Wang, Tian, and Yuyangzi Fu. "Item-based Collaborative Filtering with BERT." In *Proceedings of The 3rd Workshop on e-Commerce and NLP*, pp. 54-58. 2020.

[12] Girsang, A. S., Al Faruq, B., Herlianto, H. R., & Simbolon, S. (2020, June). Collaborative recommendation system in users of anime films. In Journal of Physics: Conference Series (Vol. 1566, No. 1, p. 012057). IOP Publishing.

[13] Prayogo, J. E., Suharso, A., & Rizal, A. (2020). Analisis Perbandingan Model Matrix Factorization dan K-Nearest Neighbor dalam Mesin Rekomendasi Collaborative Berbasis Prediksi Rating. *Jurnal Informatika Universitas Pamulang*, *5*(4), 506-514.

[14] Yulian, H., & Martina, I. (2015). Penggunaan Correlation-Based Similarity untuk Sistem Rekomendasi Tanpa Rating. Jurnal Telematika, 10(1), 3.

[15] Priyono, A. B. A. (2017). Performa Apriori dan Collaborative Filtering untuk Sistem Rekomendasi. Jurnal Ilmiah Informatika Komputer, 21(1).

[16] Kurniawan, A. (2016). Sistem Rekomendasi Produk Sepatu Dengan Menggunakan Metode Collaborative Filtering. In Seminar Nasional Teknologi Informasi dan Komunikasi (Vol. 2016, pp. 18-19).

[17] Muliadi, K. H., & Lestari, C. C. (2019). Rancang Bangun Sistem Rekomendasi Tempat Makan Menggunakan Algoritma Typicality Based Collaborative Filtering. Techno. Com, 18(4), 275-287.

[18] Salakhutdinov, R., Mnih, A., & Hinton, G. (2007, June). Restricted Boltzmann machines for collaborative filtering. In Proceedings of the 24th international conference on Machine learning (pp. 791-798).

[19] Indriawan, W., Gufroni, A. I., & Rianto, R. (2020). Sistem Rekomendasi Penjualan Produk Pertanian Menggunakan Metode Item Based Collaborative Filtering. Jurnal Siliwangi Seri Sains dan Teknologi, 6(2).

[20] Wairegi, S., Mwangi, W., & Rimiru, R. (2020). A Framework for Items Recommendation System Using Hybrid Approach. In 2020 IST-Africa Conference (IST-Africa) (pp. 1-15). IEEE.

[21] Wijaya, A., & Alfian, D. (2018). Sistem Rekomendasi Laptop Menggunakan Collaborative Filtering Dan Content-Based Filtering. Jurnal Computech & Bisnis, 12(1), 11-27.

[22] Liu, Y., Tong, Q., Du, Z., & Hu, L. (2014, September). Content-boosted restricted Boltzmann machine for recommendation. In International Conference on Artificial Neural Networks (pp. 773-780). Springer, Cham.

[23] Fitrianti, B. S., Fachurrozi, M., & Yusliani, N. (2018). Sistem Rekomendasi Artikel Ilmiah Berbasis Web Menggunakan Content-based Learning dan Collaborative Filtering. Generic, 10(1), 17-23.

[24] He, X., Liao, L., Zhang, H., Nie, L., Hu, X., & Chua, T. S. (2017, April). Neural collaborative filtering. In Proceedings of the 26th international conference on world wide web (pp. 173-182).

[25] Reimers, N., & Gurevych, I. (2019). Sentence-bert: Sentence embeddings using siamese bert-networks. arXiv preprint arXiv:1908.10084.