



## Implementation of Word Recommendation System Using Hybrid Method for Speed Typing Website

Melinda Melinda<sup>1</sup>, Maulana Imam Muttaqin<sup>2</sup>, Yudha Nurdin<sup>3</sup>, Al Bahri<sup>4</sup>

<sup>1,2,3,4</sup>Department of Electrical and Computer Engineering, Faculty of Engineering, University of Syiah Kuala

<sup>1</sup>melinda@unsyiah.ac.id, <sup>2</sup>maulanaimammuttaqin2000@gmail.com, <sup>3</sup>yudha.nurdin@unsyiah.ac.id, <sup>4</sup>albahri@unsyiah.ac.id

### Abstract

*Typing is one of the most frequently done activities in society therefore a medium is necessary to help train typing words that are often mistyped. Methods used in this research are the Content-Based Filtering Algorithm to gather the words that have a similar pattern to the words that are often mistyped based on the user's previous typing records and the Collaborative Filtering Algorithm that uses other users typing pattern to recommend the words. The result of this study shows the Collaborative Filtering Algorithm was able to gather words that are hard to type by the user with an accuracy of 49.2%, dan the Collaborative Filtering able to predict the score on how difficult for the user to type a word with the result of Root Mean Square Error (RMSE) value of 0.82 and with the Root Mean Square Percentage Error (RMSPE) value of 30% from the actual value, and a website which is the combination of the two algorithms with the result of 28% of the total word that is recommended was indeed difficult to type by the user with the typing speed of 103 WPM, and 72.3% for the user that has a typing speed of 39 WPM.*

*Keywords: recommendation system, collaborative filtering, content based filtering, speed typing, website*

### 1 Introduction

Typing is a very important skill in today's society because typing skill is used in communication on current technology, and in this century it will make a mass impact on different aspects of human life [1]. Therefore, increasing typing speed is very necessary because it can help them complete various tasks more efficiently [2].

Studies on building a medium to increase typing speed have also been carried out, in previous studies an experiment was carried out to increase typing speed using the drill method which means training continuously and repetitively [3]. Another study to increase typing speed is the creation of a game-based typing website using the Touch-Typing method [4], which is a method of typing without looking at the keyboard. Where this method works by increasing the typing focus but on the reading source at the same time the typing fingers press the letter keys on the keyboard synchronously.

In some previous studies [3], [4], it is still using a collection of random words or text which is used as a typing test material. Alternately, the system can also be developed by collecting and providing some words according to the needs of the user, namely words that

are difficult to type by the user, system that has this ability is a recommendation system.

Many other studies related to the recommendation system have also been carried out previously where the study used the Collaborative Filtering algorithm [5]–[7] used to recommend movies [8] and books [9], the collaborative algorithm itself works by predicting the behavior of a user by comparing it with similar users, the detection results will help in finding suitable items for users. [10].

Another study on the type of Content Based Filtering algorithm [11] also a lot has been done which is used to recommend similar things like movies [12], [13], foods [14] and also career [15], Content-Based filtering algorithm works by collecting an attribute contained in an item and trying to find items that have the same attributes, where similar items will be recommended to users.

Each of the algorithm [5], [11] has both advantages and disadvantages depending on the case, for example in the case of word recommendation in speed typing, the Collaborative Filtering has a problem with data sparsity where there will be a data (Word) that is not rated by the user causing inaccuracy in the recommendation system and for the Content Based Filtering the

algorithm has a problem with overspecialization of the item that can cause to narrow the range of types of the item that will recommended[16]. Therefore, both of the algorithm will be used to cover the disadvantage from each algorithm where the Content Based Filtering will cover the data sparsity of the Collaborative Filtering Algorithm and the Collaborative Filtering Algorithm will cover the overspecialization problem.

In this study, a word recommendation system will be built by combining two types of algorithms, namely Content-Based Filtering and Collaborative filtering or Hybrid which are used to process and recommend words that are difficult to type by users in the hope that they can be used as training material for users to reduce weaknesses in typing.

## 2 Research Method

This study uses the Algorithm method of User-Based Collaborative Filtering and Content-Based Filtering to recommend words that are difficult for users to type, while the stages of this study can be seen in the following

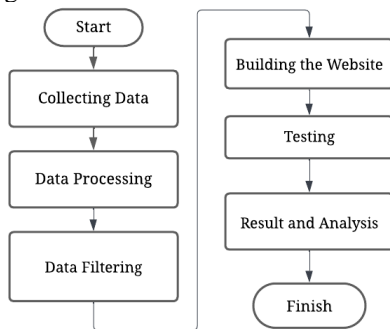


Figure 1. Research process

Based on Figure 1, this proposed study started by collecting the data necessary for the next process which is data processing, the data processing is the step where the algorithm will be used to train model from the collected data, and the trained model will be used to create filtering system to filter the upcoming new data. The filtering system will be used in the backend part of the website in the “building the website” step and also the step where the Web UI created for the user to interact with the system. The finished website will be tested and analyzed to check its performance. The detailed method of this study can be seen in the following.

### 2.1 Data Collection

The study begins with collecting data used in the study, there are two types of data used in this study, namely a collection of Indonesian words that will be recommended by the system and also user typing performance data.

The collection of Indonesian words was collected from sources such as the Electrical Engineering department of Syiah Kuala University final research report collection and the adjective dataset obtained from the Kaggle website [17]. Word retrieval is carried out using a Python program that traces all non-scientific and Indonesian words on each page of the research report file.

Typing performance data is the result of a typing speed test conducted by users which is collected using a fast typing test website application [18]. The data stored is the rhythm data of a user in typing a word, where this rhythm is a collection of time needed to type the right letter to the next right letter which is calculated in milliseconds. The form of the typing rhythm data can be seen in the following illustration.

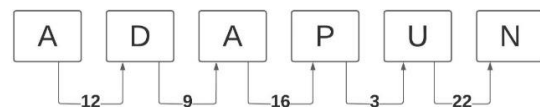


Figure 2. Illustration Rhythm Typing words from a user

### 2.2 Data Processing

This stage is data processing from data that has been collected previously. The data processing of the typing test results aims to change the data from Figure 2 into a single numeric data that has a value range between 1 to 5, where this value is the basis for determining how difficult it is for a user to type a word. The data processing is done by calculating the Standard Deviation from the rhythm data that has been collected [19], [20].

$$s = \sqrt{\frac{\sum_{i=1}^n (x_i - \bar{x})^2}{n-1}} \quad (1)$$

The results obtained from Equation (1) will be calculated again using the equation 2.

$$score = \left( \frac{Standard\ Deviation}{70} \times 4 \right) + 1 \quad (2)$$

From equation (1),  $\bar{x}$  is the average,  $x_i$  is the value  $x$  to  $i$ ,  $n$  sample size and  $s$  is the standard deviation, while in Equation (2) value “70” is the maximum standard deviation value that is allowed on the system created, if a standard deviation is obtained which has a maximum limit, the system will immediately give the highest score of 5, the value “4” is a number that changes the results of the previous division to 0 - 4, and the sum the number 1 makes the data range change to 1 – 5.

Indonesian word collection data processing is done by extracting attributes from each word, the attributes in question are all possible letter arrangement patterns that make the characteristics of a word, such as the word “abadi” will have a collection of attributes, namely “aai aa ai abadi abad aba ab badi bad ba adi ad di”. Each attribute that has been extracted from each word will be

weighted using Term Frequency and Inverse Document Frequency [21]–[23].

$$TF_{i,j} = \frac{n_{i,j}}{\sum_k n_{i,j}} \quad (3)$$

$$idf_j = \log \left[ \frac{n}{df_j} \right] \quad (4)$$

On equation (3)  $n_{i,j}$  is the number of equal attributes that appear in a word, and  $\sum_k n_{i,j}$  is the number of attributes contained in a word. In the Equation (4),  $n$  is the sum of all documents and  $df_j$  is the number of times a word appears in a set of documents. The weighting will produce a matrix containing the weight of each attribute for each word as shown in the following table.

Table 1. Score Matrix TF-IDF

	Attribute 1	Attribute 2	...
Word A	0.04	0.04	...
Word B	0.043	0.043	...
Word C	0.001	0.02	...
...	...	...	...

### 2.3 Data Filtering

At this stage, the words that will be recommended to users are filtered, where the words to be recommended are words that are difficult for users to type. filtering is done using the Content Based Filtering algorithm which will use a data collection of attribute weights from Indonesian words such as Table 1 and the Collaborative Filtering algorithm that uses data on the value of how difficult it is for users to type a word.

In the Content-Based Filtering algorithm, the filtering process is carried out to determine what words have similar attributes determined by using the Cosine Similarity equation (5) as follows [11].

$$\cos \theta(\vec{q}_t, \vec{d}_t) = \frac{\vec{q}_t \cdot \vec{d}_t}{|\vec{q}_t| \times |\vec{d}_t|} \quad (5)$$

Where  $\vec{q}_t$  and  $\vec{d}_t$  is the vector to which the similarity will be compared,  $|\vec{q}_t|$  and  $|\vec{d}_t|$  is the length of the two vectors. The filtering results will produce a collection of words that have the highest similarity value.

In the Collaborative Filtering algorithm, the filtering is done by predicting the score of how difficult it is for a user to type a word, and the word that has the highest prediction score will be recommended to the user. The score prediction process is carried out using the following equation: [24]

$$prediction_{u,i} = \frac{\sum_n \omega_{i,n} * r_{u,n}}{\sum_n |\omega_{i,n}|} \quad (6)$$

From equation (6),  $\omega_{i,n}$  is the similarity value of word  $i$  to other words and  $r_{u,n}$  is the value given by the user to the word whose score will be predicted.

### 2.4 Building the Website

At this stage, a website is developed that can detect words that are difficult for the user to type and recommend relevant words. This stage is the stage of implementation and merging of all systems and algorithms that have been made in the previous stage. The website page created consists of two types, namely the Practice page and the Double (Compete) Player page where both pages require an authentication process from the user.

On the Practice Page, it is useful to practice the user's typing skills by providing words that are predicted to be difficult for the user to type. The workflow of this page starts from the user accessing the page and starting a typing test. The scoring system that has been integrated on the page will rate every word that is typed by the user. The results of the typing test will be stored in the database. When the user performs a retest, the server will read the Database containing the recorded user's typing results and by using the Collaborative Filtering and Content Based Filtering algorithms will filter out words that are thought to be difficult for the user to type, which will be displayed on the Exercise page for use by the user as practice.

On the Dual Player page, users can interact with other users by trying to beat other players in typing all the words as quickly as possible where both users will type the same set of words. The workflow of this page starts with a player looking for a second player who is also looking for a player partner. When the two players are connected, the server will collect the word randomly and display the word set to both users, where the word typed by the user is the same set of words to avoid injustice, and both will type all the words displayed to completion. While the user is typing the displayed word, the word scoring system is also automatically executed. This system will assess how difficult it is for a user to type a word, which will later be sent to the main server for reprocessing on the practice page. The game will be finished when both users finish typing all the words displayed and the player who finishes it first will be the winner. Both players can leave the page and restart the game.

### 2.5 Testing

This stage will test all the systems that have been created. Testing tests the Collaborative Filtering algorithm, carried out by dividing the set of user scores into training data and also test data, where the test data is used as a reference whether the training data is accurate or not. Predicted and tested data is the word score that has been typed by the user.

While the Content Based Filtering algorithm is tested by giving an input in the form of words to be processed by the algorithm to see the output of the input, to see the similarity of the words entered with the output words of

the algorithm. Another test carried out on this algorithm is to calculate how many words that are output from the algorithm are words that are difficult for users to type. Where it is determined by counting how many words have a score greater than a predetermined limit.

Website testing that has been created is done by testing directly using a computer or laptop device by typing all the words displayed on the website page. Where the word already has the three algorithms that have been made previously. The tests carried out on the three algorithms will be tested for the accuracy of the predictions that have been made to determine whether the algorithm is successful in recommending the desired word.

## 2.6 Result and Analysis

At this stage, the results of the tests that have been carried out previously will be analyzed. In the Content-Based Filtering algorithm, the analysis is carried out by comparing the entered words with a collection of words that are the result of processing the algorithm whether the resulting word collection has a wording similar to the input word, and counting the total words that have a higher score than the threshold. a predetermined score using the following equation.

$$\text{percentage} = \frac{s}{n} \times 100\%. \quad (7)$$

in equation (7),  $s$  is the total word that has a score of more than 1.5 while  $n$  is the total word recommended to users. In the Collaborative Filtering algorithm, the analysis is carried out by comparing the predicted value results with the actual value and calculating the error value of the difference between the two values using the Root Mean Square Error and Root Mean Square Percentage Error equations.

$$RMSE = \sqrt{\frac{\sum(f-o)^2}{n}} \quad (8)$$

$$RMSPE = \sqrt{\frac{\sum\left(\frac{f-o}{f}\right)^2}{n}} \quad (9)$$

In equation (8) (9)  $f$  is the predicted value,  $o$  is the actual value and  $n$  is the total value compared.

On the typing website that has been made, the analysis is carried out by calculating how many words are displayed and have been typed have a score exceeding the limit that determines whether a user is difficult or not to type the word. Where the total calculation will be compared with the total words that have been typed to calculate the percentage of how many words are difficult for users to type.

## 2.7 Conclusion

At this stage, conclusions are made from the studies that have been carried out, where the results from the

previous stages serve as a reference for making conclusions.

## 3 Results and Discussion

### 3.1 Word Rating System

The results obtained from the analysis of the word scoring system are a pattern of the user's typing rhythm. When a user manages to type a word smoothly, the rhythm graph obtained is relatively stable, such as Figure 3 and has a relatively low standard deviation. Meanwhile, when the rhythm of a user when typing is disturbed or chaotic, the graph will become unstable like Figure 4 and results in relatively higher standard deviations and scores than when typing fluently. An unstable rhythm can be caused by the user incorrectly typing the next letter or stopping because they don't know where the next letter key is located. Therefore, in this study, it can be concluded that the standard deviation of the user's typing rhythm can be used as the basis for a scoring system where the value is converted back into a narrower range using equations.

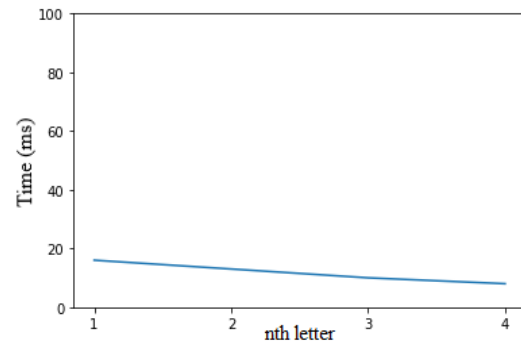


Figure 3. User Rhythm Chart When Fluently Typing on the word "CEKAT"

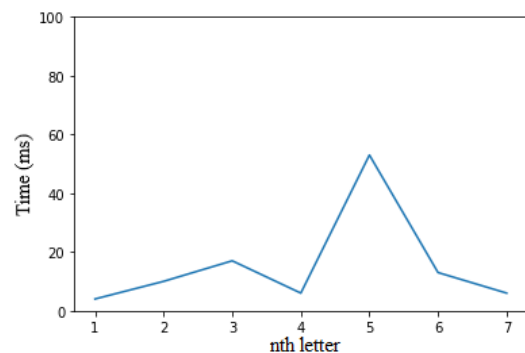


Figure 4. User Rhythm Chart When User Fails Typing the word "TERBEBAS" smoothly

### 3.2 Collaborative Filtering Algorithm

The result from testing this algorithm shows the RMSE result with the value of 0.82 calculated using the Equation (8) and RMSPE with the value of 30% calculated using the Equation (9). The calculation of RMSE and RMSPE is done by which collecting first all

the real scores gathered from user typing a given word and all the predicted scores from the algorithm of the same word, by using the Equation (8) and (9) each score from each word will be compared between its real value and the predicted value which returns both RMSE value and RMSPE value. Based on the results of the test the RMSE and RMSPE value have a relatively high value, this is due to the lack of dataset or users scores to use as a base for predicting the users score when typing a word.

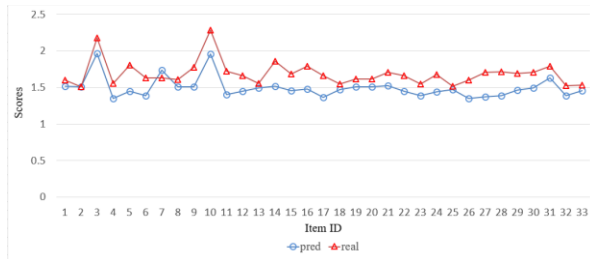


Figure 5. Graph Comparison between Predicted Value and Actual

Figure 5 shows a graphic comparison between the predicted value and the actual value shows that even though they have different values, the pattern from the predicted value tries to follow the graphic pattern of the actual value, for example from Figure 5 the word with Item ID 3, 6 and 10 has a higher value on both prediction and real value other than other ID, although for some ID the value does not match where the predicted value resulting higher value when it supposed to be lower for example Item ID 12, 19, 48, etc. due to lack of data . this pattern is useful for filtering which words are prioritized to be recommended and which are not by recommending the word that has the highest score first

### 3.3 Content-Based Filtering Algorithm

The test results of this algorithm show that the word output from the algorithm has a similar word structure, which can be seen in 2.

Table 2. Testing Result of Content Based Filtering Algorithm

Word Input	Word Output	Total > 1,5
humanistis	humanistis	1,19
	manise	1,17
	mistis	1,20
	semanis	1,52
	komunistis	2,27
	tuman	3,49
	sauvinistis	2,02
	statistis	2,18
	sadistis	1,35
	humoristis	1,26
...	...	...
<b>Total Average</b>		<b>4.92</b>

Table 3. Content-Based Filtering Output

Word Input	Word Output
anggak	panggak
	anggara
	manggah
	ganggang
	menunggak
mengganggu	pengganggu
	lenggang
	benggang
	senggang
	menunjukkan
unjukkan	ditunjukkan
	unjuk
	menunjuk

The results of other tests carried out on this algorithm are measuring the score given by the user to the word that has been recommended which can be seen in

Table 3. The test results show that 49.2% of the total recommended words have a score exceeding the threshold which indicates that the user has difficulty typing the word, where the threshold has the value of 1.5 from the maximum value of 5, this value is determined by observing users hand while typing the word for example at what moment does the users hand start to struggle to type the correct letter, although this method is not the best way to decide a threshold value, the value itself has no effect on performance of the algorithm, its purpose only as based to know the ratio between the word that are hard to type by the user and not, since the algorithm only checking some of the highest score as a base to gather list of word that are hard to type by the user. From the result of the second test, less than half of the word that are recommended are the word that are hard to type by the user which is low, this happens due to the inaccuracy of deciding the threshold and also the lack of the words that has the same structure or patterns

### 3.4 Fast Typing Game Website Interface Design

One of the page from Website is the practice page where the user is asked to type the word that is displayed as well as possible as shown in Figure 6. The graph next to the word collection is a graph that shows the rhythm of a user typing a word, and on the Figure 7 is a typed result that shows the user's typing speed and displays the words that the system considers difficult to type for the user.

The multiplayer mode page as shown in Figure 8 is where users can test their typing skills with other users. From Figure 8 the user is able to see the progress on how far has the other player has typed using the percentage above the word collection.





Figure 6. Exercise Page View

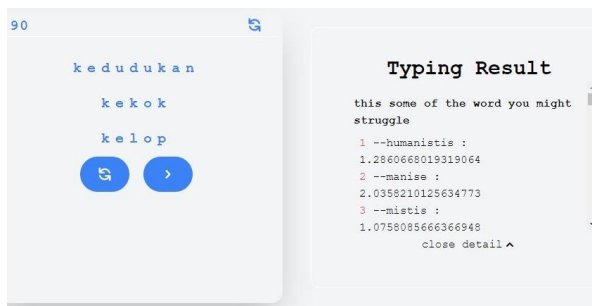


Figure 7. Typing Result Display

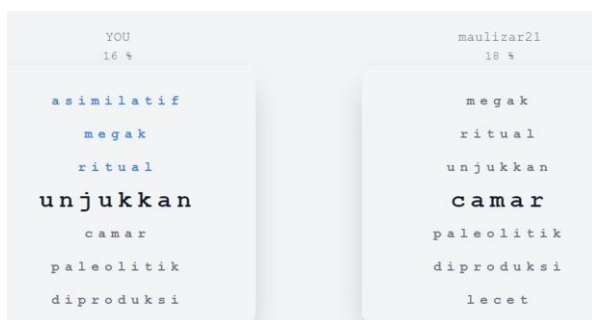


Figure 8. Dual Player Mode Page

The detail page of user typing speed as shown in Figure 9 is where the users can see the progress of typing speed in the form of graphs as well as other data such as highest, lowest and average typing speed that has been done.

Authentication page as shown in Figure 10 is a page where users authenticate themselves using a username and password so that users can use features such as doubles players and practice pages because these features require user data to work.

Register page as shown in Figure 11, aims to register if the user does not have a registered account, where on this page the user will be asked to fill in data such as username, password and password confirmation



Figure 9. Users Typing Speed Details Page

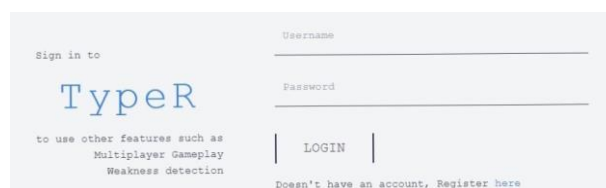


Figure 10. Authentication Page



Figure 11. Register Page

### 3.5 Evaluation of Recommendation System on Fast Typing Website

The results of the merging test of the entire system on the fast-typing website show that for users who have a typing speed of 103 KPM, an average of 29.4% of the total words recommended to users are words that are difficult for users to type, which can be seen in Table 4.

Table 4. Test Results of Website Word Recommendation System with users who have a speed of 103 KPM

Words Total	Total Typo Word	Percentage
90	26	28%
90	21	23%
90	35	38%
90	30	33%
90	23	25%
<b>Average</b>		<b>29.4%</b>

As for users who have a typing speed of 39 KPM, an average of 73.2% of the total recommended words, which can be seen in Table 5.

Table 5. Test Results of Website Word Recommendation System with users who have a speed of 39 KPM

Words Total	Total Typo Word	Percentage
90	69	76%
90	62	68%
90	66	73%
90	68	75%
90	74	80%
<b>Average</b>		73%

The results obtained from the two different users have different averages. This could be due to the different experiences of the two users who have different speeds, where users who have a higher average typing speed will have fewer words that are difficult to type. Meanwhile, users who have a lower average typing speed will have more words that are difficult to type.

#### 4 Conclusion

The results of the studies that have been carried out show that the application of Collaborative Filtering and Content-Based Filtering algorithms has the potential to be applied not only to determining preferences but also to the opposite, namely determining user weaknesses. The test results show that combining the Collaborative Filtering and Content-Based Filtering algorithms, 28% of the total recommended words for users are words that are difficult to type for users who have a typing speed of 103 KPM and 72.3% for users who have a typing speed of 39 KPM.

Related research that uses the same method but uses a different object shows better results with a precision value of 96%. This is because the study uses a much larger amount of data of 1 million rows of data due to the easy access to get this type of data. While in this study, it only had about 8 thousand more rows of data due to unusual types of data and requiring data retrieval from the beginning. Therefore, it is necessary to retrieve data with more users and input data from more users to improve the accuracy of predicting words that are difficult for users to type.

#### References

- [1] C.-H. Lin and E. Z.-F. Liu, "A Comparison between Drill-Based and Game-Based Typing Software," in *Transactions on Edutainment III*, vol. 5940, Z. Pan, A. D. Cheok, W. Müller, and M. Chang, Eds. Berlin, Heidelberg: Springer Berlin Heidelberg, 2009, pp. 48–58. doi: 10.1007/978-3-642-11245-4\_5.
- [2] A. Ikhsananto and S. Sutirman, "The Effectiveness of Rapid Typing Software and Module for Improving 10-Finger Typing Skills," *DP*, vol. 13, no. 2, pp. 228–237, Dec. 2018, doi: 10.15294/dp.v13i2.15540.
- [3] H. S. Lestari, "INCREASE TYPING SPEED THROUGH DRILL METHOD," p. 4, 2021.
- [4] M. W. Pamengas, T. Afrianto, and W. S. Wardhono, "Pengembangan Gim Edukasi Mengetik menggunakan Metode Touch-Typing untuk Meningkatkan Kecepatan Mengetik," p. 8.
- [5] Z. Cui *et al.*, "Personalized Recommendation System Based on Collaborative Filtering for IoT Scenarios," *IEEE Trans. Serv. Comput.*, vol. 13, no. 4, pp. 685–695, Jul. 2020, doi: 10.1109/TSC.2020.2964552.
- [6] C.-D. Wang, Z.-H. Deng, J.-H. Lai, and P. S. Yu, "Serendipitous Recommendation in E-Commerce Using Innovator-Based Collaborative Filtering," *IEEE Trans. Cybern.*, vol. 49, no. 7, pp. 2678–2692, Jul. 2019, doi: 10.1109/TCYB.2018.2841924.
- [7] Q. Zhang, L. Cao, C. Zhu, Z. Li, and J. Sun, "CoupledCF: Learning Explicit and Implicit User-item Couplings in Recommendation for Deep Collaborative Filtering," in *Proceedings of the Twenty-Seventh International Joint Conference on Artificial Intelligence*, Stockholm, Sweden, Jul. 2018, pp. 3662–3668. doi: 10.24963/ijcai.2018/509.
- [8] S. M. Ali, G. K. Nayak, R. K. Lenka, and R. K. Barik, "Movie Recommendation System Using Genome Tags and Content-Based Filtering," in *Advances in Data and Information Sciences*, vol. 38, M. L. Kolhe, M. C. Trivedi, S. Tiwari, and V. K. Singh, Eds. Singapore: Springer Singapore, 2018, pp. 85–94. doi: 10.1007/978-981-10-8360-0\_8.
- [9] E. Uko, B. O., and P. O., "An Improved Online Book Recommender System using Collaborative Filtering Algorithm," *IJCA*, vol. 179, no. 46, pp. 41–48, Jun. 2018, doi: 10.5120/ijca2018917193.
- [10] J. Lee, M. Sun, and G. Lebanon, "A Comparative Study of Collaborative Filtering Algorithms," *arXiv:1205.3193 [cs, stat]*, May 2012, Accessed: Oct. 26, 2021. [Online]. Available: <http://arxiv.org/abs/1205.3193>
- [11] M. Benard Magara, S. O. Ojo, and T. Zuva, "A comparative analysis of text similarity measures and algorithms in research paper recommender systems," in *2018 Conference on Information Communications Technology and Society (ICTAS)*, Durban, Mar. 2018, pp. 1–5. doi: 10.1109/ICTAS.2018.8368766.
- [12] Y. Deldjoo, M. Elahi, P. Cremonesi, F. Garzotto, P. Piazzolla, and M. Quadrana, "Content-Based Video Recommendation System Based on Stylistic Visual Features," *J Data Semant*, vol. 5, no. 2, pp. 99–113, Jun. 2016, doi: 10.1007/s13740-016-0060-9.
- [13] S. Reddy, S. Nalluri, S. Kuniseti, S. Ashok, and B. Venkatesh, "Content-Based Movie Recommendation System Using Genre Correlation," in *Smart Intelligent Computing and Applications*, vol. 105, S. C. Satapathy, V. Bhateja, and S. Das, Eds. Singapore: Springer Singapore, 2019, pp. 391–397. doi: 10.1007/978-981-13-1927-3\_42.
- [14] M. Phanich, P. Pholkul, and S. Phimoltare, "Food Recommendation System Using Clustering Analysis for Diabetic Patients," in *2010 International Conference on Information Science and Applications*, Seoul, Korea (South), 2010, pp. 1–8. doi: 10.1109/ICISA.2010.5480416.
- [15] T. V. Yadalam, V. M. Gowda, V. S. Kumar, D. Girish, and N. M., "Career Recommendation Systems using Content based Filtering," in *2020 5th International Conference on Communication and Electronics Systems (ICCES)*, COIMBATORE, India, Jun. 2020, pp. 660–665. doi: 10.1109/ICCES48766.2020.9137992.
- [16] P. B. Thorat, R. M. Goudar, and S. Barve, "Survey on Collaborative Filtering, Content-based Filtering and Hybrid Recommendation System," *IJCA*, vol. 110, no. 4, pp. 31–36, Jan. 2015, doi: 10.5120/19308-0760.
- [17] "Indonesian Adjective Sentiment (Kata Sifat)." <https://www.kaggle.com/datasets/linkgish/indonesian-adjective-sentiment-kata-sifat> (accessed Aug. 14, 2022).
- [18] "Speed Check." <https://speed-check.netlify.app/> (accessed Aug. 14, 2022).
- [19] D. K. Lee, J. In, and S. Lee, "Standard deviation and standard error of the mean," *Korean J Anesthesiol*, vol. 68, no. 3, p. 220, 2015, doi: 10.4097/kjae.2015.68.3.220.
- [20] E. H. Livingston, "The mean and standard deviation: what does it all mean?," *Journal of Surgical Research*, vol. 119, no. 2, pp. 117–123, Jun. 2004, doi: 10.1016/j.jss.2004.02.008.

- [21] S. Qaiser and R. Ali, "Text Mining: Use of TF-IDF to Examine the Relevance of Words to Documents," *IJCA*, vol. 181, no. 1, pp. 25–29, Jul. 2018, doi: 10.5120/ijca2018917395.
- [22] A. A. Hakim, A. Erwin, K. I. Eng, M. Galinium, and W. Muliady, "Automated document classification for news article in Bahasa Indonesia based on term frequency inverse document frequency (TF-IDF) approach," in *2014 6th International Conference on Information Technology and Electrical Engineering (ICITEE)*, Yogyakarta, Indonesia, Oct. 2014, pp. 1–4. doi: 10.1109/ICITEED.2014.7007894.
- [23] P. Bafna, D. Pramod, and A. Vaidya, "Document clustering: TF-IDF approach," in *2016 International Conference on Electrical, Electronics, and Optimization Techniques (ICEEOT)*, Chennai, India, Mar. 2016, pp. 61–66. doi: 10.1109/ICEEOT.2016.7754750.
- [24] J. Chiang, "Overview of Item-Item Collaborative Filtering Recommendation System," *Geek Culture*, May 30, 2021. <https://medium.com/geekculture/overview-of-item-item-collaborative-filtering-recommendation-system-64ee15b24bb8> (accessed Aug. 24, 2022).



