Kinetik: Game Technology, Information System, Computer Network, Computing, Electronics, and Control
Journal homepage: http://kinetik.umm.ac.id
ISSN: 2503-2267
Vol. 5, No. 3, August 2020, Pp. 179-186

Various implementation of collaborative filtering-based approach on recommendation systems using similarity

Zainur Romadhon*[1], Eko Sediyono[2], Catur Edi Widodo[3]
Universitas Diponegoro, Indonesia[1,3]
Universitas Kristen Satya Wacana, Indonesia[2]

Article Info
Keywords:
Recommendation System, Collaborative Filtering, Cosine Similarity

Article history:
Received 06 April 2020
Revised 31 May 2020
Accepted 15 July 2020
Published 31 August 2020

Cite:
Romadhon, Z., Sediyono, E., & Widodo, C. (2020). Various Implementation of Collaborative Filtering-Based Approach on Recommendation Systems using Similarity. Kinetik: Game Technology, Information System, Computer Network, Computing, Electronics, and Control, 5(3). doi:https://doi.org/10.22219/kinetik.v5i3.1062

*Corresponding author.
Zainur Romadhon
E-mail address:
zainur.rmdn@gmail.com

Abstract
The Recommendation System plays an increasingly important role in our daily lives. With the increasing amount of information on the internet, the recommendation system can also solve problems caused by increasing information quickly. Collaborative filtering is one method in the recommendation system that makes recommendations by analyzing correlations between users. Collaborative filtering accumulates customer item ratings, identifies customers with common ratings, and offers recommendations based on inter-customer comparisons. This study aims to build a system that can provide recommendations to users who want to order or choose fast food menus. This recommendation system provides recommendations based on item data calculations with customer review data using a collaborative filtering approach. The results of applying cosine similarity calculation to determine fast food menu recommendations obtained for the item-based recommendation is Pizza Frankfurter BBQ Large with a value of 1.0, item-based with genre recommendation is Calblend Float with value 1.0 and user-based recommendation is Pizza Black Pepper Beef / Chicken Large with mean score 2.5.

1. Introduction

The exponential growth of information that can be accessed through the network, a person can obtain more information and have a variety of choices via the Internet, which provides comfort and convenience for people's daily lives. However, the amount of information available often makes a person confused and difficult to make the right choice [1][2]. A traditional search engine can ease the requirements of user information retrieval to some extent [1]. Even so, search engines can only present the same sorting results to all users and cannot provide personalized services according to different user interests. Therefore, a personalized recommendation system emerges naturally. The recommendation system can automatically suggest information or items that might be liked to users [3]. Recommendation systems are becoming more and more popular with the onset of the World Wide Web and big data [4]. Specifically, according to various data collection and analysis of user preferences, a personalized recommendation system studies the interests and patterns of user behavior to recommend an information and service needed for the user [1].

In general, there are two variants of the recommendation approach: the content-based approach and collaborative filtering-based approach (CF) [3][5]. But we only focus on Collaborative filtering (CF) approach, where recommendations are made based on the user’s ratings of the items. Users with similar ratings are called nearest neighbors and items with high predictive ratings will be recommended to users [6][7]. The collaborative filtering approach can be further grouped into model-based CFs and memory-based CFs [3][5][8]. Among these recommendation approaches, collaborative filtering is generally considered as one of the most used and most successful recommendation technologies in the recommendation system, especially e-commerce websites such as Amazon.com, Netflix and Google News [7]. In addition, collaborative filtering has also been applied to restaurant recommender system, which is useful for providing recommendations to users who will choose or buy a specific culinary menu based on the ratings given by other users [9][10][11][12].

2. Research Method

At present, many applications present information on a variety of foods or food culinary, for example food menus, food prices and restaurant names of food vendors. The popular system used in food culinary applications is a recommendation system [9][12]. Various studies related to the recommendation system have been carried out. Recommendation system is an effective application to help users get information that is useful and in accordance with the interests of users [9][11]. With the increasing amount of information on the internet, the recommendation system can also solve problems caused by increasing information quickly [9][13][14]. For example, applications that use the recommendation system are Youtube, Amazon, Netflix, and Facebook.

The recommendation system is of several types, in this study we use a type of collaborative filtering, this system can advise users about which information is most relevant to them. Collaborative filtering has proven to be the most effective technique to help users find content or information that is in line with what is expected [5][10].

2.1 Collaborative Filtering based Recommender System

Collaborative filtering is one method in the recommendation system that makes recommendations by analyzing correlations between users. This method looks for similarities between users to make a prediction [5][10]. For example, if user A is similar to user B, if user A is interested in product A, it can be concluded that user B will also be interested in product A, and product A can be recommended to user B. The collaborative filtering approach requires data about the user's opinion about a products or goods, and identify the environment based on similarities between users [10][15]. Collaborative Filtering is divided into 2 categories [16][17]:

1. Memory-based, it is necessary to (rating) the user's ranking data from the user to the item to calculate the similarity between the user or item.
2. Model-Based generally the data used is not complete and a learning is needed in finding a model of available data which will be used to find similarities of users or items to be predicted.

In this paper, we focus on memory-based of collaborative filtering to get proximity between items and users. Users or items proximity can be calculated using a neighborhood-based algorithm where the similarities between the two items or users will be resulted from the average weight of all ratings.

2.2 K-Nearest Neighbor (KNN)

In collaborative filtering one of the algorithms that can be used is the K-NN algorithm (also known as k nearest neighbors algorithm) which is the most commonly used classification algorithm. K-Nearest Neighbors is a method used to classify new objects based on attributes and training samples that have been classified [18][19]. This algorithm is ideal for solving various category problems. K- Nearest Neighbor is included in the lazy learning algorithm which is easy to implement [18][20]. The following are the steps of the K-NN algorithm:

1. Determine the parameter k (number of nearest neighbors)
2. Calculate distance of training data with all test data
3. Sort the distance based on the smallest value of k.
4. Determine the group of test data based on the majority label on k.

Calculation of metric distances on the KNN algorithm we can use cosine similarity, cosine similarity is a method for measuring the level of similarity between two vectors. Calculations in this method are done by calculating the Cosine value between two vectors [21][22]. The proximity of user characteristics can be found by cosine similarity formula, using the following Equation 1.

$$similarity = \cos(\theta) = \frac{A \cdot B}{\|A\| \|B\|} = \frac{\sum_{i=1}^{n} A_i x\, B_i}{\sqrt{\sum_{i=1}^{n} A_i^2}\; x \sqrt{\sum_{i=1}^{n} B_i^2}} \tag{1}$$

Where:
A = document vector
B = query vector
$A \cdot B$ = multiplication between vector A and vector B
$\|A\|$ = length of vector A
$\|B\|$ = length of vector B
$\|A\|\|B\|$ = multiplication items between $\|A\|$ and $\|B\|$

2.3 Research Overview

The following is a model of the proposed recommendation system. The approach used is based on collaborative filtering, in the process of providing recommendations by using cosine similarity to calculate the level of similarity of two or more vectors, which in this study are applied to item-based, item-based with genre and user-based collaborative filtering. In Figure 1 with the explanation as follows.

1. Preparing data from survey
2. Determining normalization of implicit and explicit data.

3. Calculating the similarity with cosine similarity for each type of dataset based on formulas.
4. Calculating the prediction value for each type of dataset using. User-based, Item-based and item genre-based.
5. Fast food predictions or recommendations

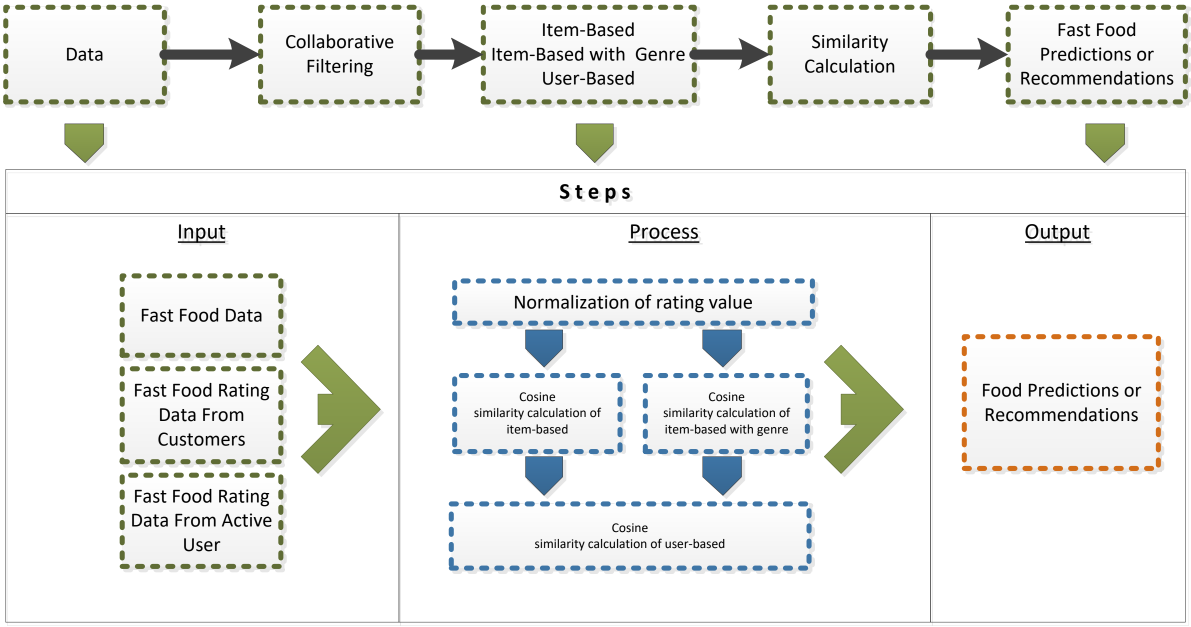


Figure 1. The Structure of Collaborative Filtering Recommender System with Cosine Similarity

3. Results and Discussion

The rapid growth of information on the internet allows one to obtain more information and have a variety of choices via the Internet, which provides comfort and convenience for people's daily lives. However, the amount of information available often makes a person confused and difficult to make the right choice [1][2]. Therefore, the recommendation system can simplify the process of selecting information, the recommendation system can automatically suggest information or items that users might like [3]. In this study we use a collaborative filtering approach to determine a fast food recommendation. The process of providing recommendations by using cosine similarity to calculate the level of similarity of two or more vectors, which is applied to item-based, item-based with genre and user-based collaborative filtering. So we can see the difference in recommendations given by the recommendation system.

3.1 Data Source

The data used is a collection of fast food data and customer review data on fast food restaurant food products in the city of Kudus, Central Java, Indonesia, including KFC, CFC, Pizza Hut Delivery, with 350 data and product ratings. The rating used is integers on a numerical scale from 1 to 5, which if larger integers represent stronger preferences. The following is a table of survey data that has been obtained. The survey data table is shown in Table 1, Table 2 and Table 3.

Table 1. Fast Food Menu Data and Restaurants

Fast Food ID	Menu	Price	Genres	Fast Food Restaurant
1	Chokocha Float	14545	Drink	KFC
2	Mocha Float	8636	Drink	KFC
3	Super Besar 2Pcs HCC	41818	Drink|Food|Packet	KFC
4	CFC Boks Combo	130909	Food|Packet	CFC
5	Super Lima	94545	Food|Packet	CFC
78	Pannacotta Vanilla Kiss	16000	Food|Dessert	PHD
79	Pannacotta Choco Smooch	16000	Food|Dessert	PHD
80	Pizza Splitza Reguler	111000	Food|Snack	PHD

Table 2. Fast Food Rating Data by Customers

No	User ID	Fast Food ID	Rating	Times Rating
1	1	12	5.0	20200320
2	1	19	5.0	20200320
3	1	40	3.0	20200319
4	1	45	5.0	20200319
5	1	62	5.0	20200320
238	40	73	4.0	20200319
239	40	72	5.0	20200319
240	40	71	5.0	20200319

Table 3. Fast Food Rating Data by Active User

No.	Fastfood	Active User (Customer) - 41
1	Chokocha Float	5
2	Mocha Float	4,5
3	Super Besar 1Pc OR	5
4	Super Besar 2Pcs HCC	5
5	Kombo Winger	5
28	Paket Chicken Strips	3
29	Paket California Burger	NaN
30	Chicken Dada/Paha Atas	3

3.2 Recommender System Model with Cosine Similarity

The recommender system (RS) has been widely used in daily life to recommend information of the customer's interest or provide personalized services based on the customer's behavior data, which help the customer quickly obtain the required information from the mass data [7][23][24]. In this paper we've build a recommendation system with cosine similarity. Cosine similarity is a method to measure the difference between two non zero vectors of an inner product space. See the example Table 4 below to understand. Suppose we want to check if customer b and customer c have similar fast food preferences, and we only have two fast food reviews. The reviews are scores from 1 to 5, where 5 is the best score and 1 the worst, and 0 means that a person has not eaten the fast food.

Table 4. Represent each person's reviews in a separate vector

Name	Chokocha Float	Mocha Float
Customer B	4	3
Customer C	5	5

In accordance with the product valuation data by consumers in Table 4, it is then converted into a vector form, as shown in Figure 2.

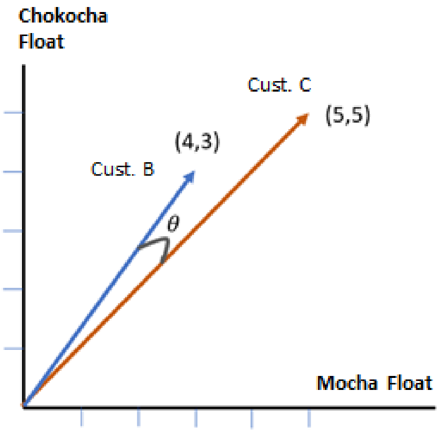
$$\vec{b} = \begin{bmatrix} 4 \\ 3 \end{bmatrix} \qquad \vec{c} = \begin{bmatrix} 5 \\ 5 \end{bmatrix}$$

Figure 2. Vector b Represents Customer B and Vector c Represents Customer C

The cosine similarity will measure the similarity between these two vectors which is a measurement of how similar are the preferences between these two people, as shown in Figure 3.

In the example above the similarity 0.989 is close to the maximum value of 1, this means that given only two fast food reviews the two users have similar preferences.

Theoretically, the cosine similarity can be any number between -1 and +1 because of the image of the cosine function, but in this case, there will not be any negative fast food rating so the angle θ will be between 0º and 90º bounding the cosine similarity between 0 and 1. If the angle θ = 0º =>cosine similarity = 1, if θ = 90º => cosine similarity =0.



Calculating :

$$b.c = \sum_{i=1}^{n} b_i c_i = (4\,x\,5) + (3\,x\,5) = 35$$

$$\|b\| = \sqrt{4^2 + 3^2} = 5$$

$$\|c\| = \sqrt{5^2 + 5^2} = 5\sqrt{2}$$

$$similarity = \frac{35}{5\,x\,5\sqrt{2}} \sim 0.989$$

Figure 3. Example Calculating Similarity

3.2.1 Item-Based Recommender

Item-to-item or item-based recommender is calculated based on the similarity between items using people's ratings on those items. In this paper, each item is represented by a fast food. For example, in order to calculate the similarity between Chokocha Float and other menu, we created two vectors with all the reviews for each fast food and then calculated the vector's cosine similarity. The vector contains several 0 values to fill in null values when a move was not rated by a user. We calculated the cosine similarity between all fast food data. In the Figure 4 below we've compared 5 popular fast foods with Chokocha Float.

	menu	cosine_sim
1	Pizza Frankfurter BBQ Large	1.000000
2	Paket Chicken 2Pcs	0.666667
3	Pizza Black Pepper Beef/Chicken Large	0.666667
4	SUPER Family 5Pcs OR	0.666667
5	Pizza Deluxe Cheese Personal	0.333333

Figure 4. Cosine Similarity Between Chokocha Float and 5 Popular Fast Foods

If active user or customer liked eating Chokocha Float, the information above shows that Pizza Frankfurter BBQ Large, Paket Chicken 2Pcs, Pizza Black Pepper Beef/Chicken Large and SUPER Family 5Pcs OR are the most similar fast foods with 1.0, 0.67, 0.67, 0.67 and 0.67 similarity scores. And ranked last, Pizza Deluxe Cheese Personal with similarity score 0.33. The similarity scores relies on only 240 user ratings, which is a small sample for creating a robust recommender system.

3.2.2 Item-Based with Genre Recommender

In the previous example, we noticed that Pizza Frankfurter BBQ Large was a better recommendation than Paket Chicken 2Pcs. But if we look into each fastfood's genre we find that they are slightly different. Chokocha Float's genre is Drink. Pizza Frankfurter BBQ Large's genre is Food and Snack, as shown in Figure 5.

	menu	cosine_sim	genre_cosine_sim
35	Calblend Float	0.000000	1.0
7	Lite Orange Drink	0.333333	1.0
24	Lite Blackcurrant Drink	0.000000	1.0
42	Cozy Cappucino Jelly	0.000000	1.0
41	Cozy Blackcurrant Shake	0.000000	1.0

Figure 5. Cosine Similarity Using Ratings and Genre

We added a new layer to the recommender, first we will find fast foods with similar genres and then select the best rating similarities. In order to do that we've added a new column containing the genre cosine similarity. So the now the Calblend Float will come first in the recommendation order because it has a 1.0 genre similarity and has similar name.

3.2.3 User-Based Recommender

Using two vectors with each person's scores for the 80 fast foods, we can calculate the cosine similarity between these two users. In this example, the active user rated 30 fast foods, we've imported his ratings and found the 4 users or customers that had the most similar cosine similarity. And then calculated the mean score for each fast food considering their rating, as shown in Figure 6.

userId menu	Zainur Romadhon	11	18	10	19	Mean_score
Pizza Black Pepper Beef/Chicken Large	0.0	0.0	0.0	5.0	5.0	2.50
Pizza Black Pepper Beef/Chicken Reguler	0.0	4.0	4.0	0.0	0.0	2.00
Surimi Wrap	0.0	0.0	5.0	0.0	0.0	1.25
Salad Bar	0.0	0.0	0.0	0.0	5.0	1.25
Spaghetti Kari	0.0	0.0	0.0	0.0	5.0	1.25

Figure 6. Comparing Active User's Scores with The Highest Mean Score

The Figure 6 above is filtered in descending order of Mean_score. The numbers 11, 18, 10 and 19 represent users that have similar preference than active user. The last column is the mean of their fast food ratings. The fast food data above were all suggested by active user and had good scores.

4. Conclusion

The recommendation system using a collaborative filtering approach provides a fairly good recommendation. The results of applying cosine similarity calculation to determine fast food menu recommendations obtained for the item-based recommendation is Pizza Frankfurter BBQ Large with a value of 1.0, item-based with genre recommendation is Calblend Float with value 1.0 and user-based recommendation is Pizza Black Pepper Beef / Chicken Large with mean score 2.5. But the problem with the Collaborative Filtering by using cosine similarity such as popularity bias, the system is biased towards items that have the most users interaction (i.e. ratings and reviews). When a new item or fast food data is added to the list, this item will have far less user interaction and therefore is rarely used as a recommendation. Therefore, using more user review data will increase cosine similarity recommendations. To improve user-based recommendations, it's important to have more reviews from many users. Using more qualitative information about each fast food menu can improve the accuracy of recommendations, such as the type of fast food, and price.

Acknowledgement

Thank you to the fast food restaurant in the city of Kudus, Central Java, Indonesia, which has become a place for researchers to develop this journal research. Hopefully, this research can make a major contribution to the advancement of technology in Indonesia.

References

[1] G. Xu, Z. Tang, C. Ma, Y. Liu, and M. Daneshmand, "A collaborative filtering recommendation algorithm based on user confidence and time context," *J. Electr. Comput. Eng.*, Vol. 2019, 2019. https://doi.org/10.1155/2019/7070487

[2] K. Choi and Y. Suh, "A new similarity function for selecting neighbors for each target item in collaborative filtering," *Knowledge-Based Syst.*, 2013. https://doi.org/10.1016/j.knosys.2012.07.019

[3] X. Yang, Y. Guo, Y. Liu, and H. Steck, "A survey of collaborative filtering based social recommender systems," *Comput. Commun.*, 2014. https://doi.org/10.1016/j.comcom.2013.06.009

[4] P. Moradi and S. Ahmadian, "A reliability-based recommendation method to improve trust-aware recommender systems," *Expert Syst. Appl.*, 2015. https://doi.org/10.1016/j.eswa.2015.05.027

[5] G. Adomavicius and A. Tuzhilin, "Toward the next generation of recommender systems: A survey of the state-of-the-art and possible extensions," *IEEE Transactions on Knowledge and Data Engineering*. 2005. https://doi.org/10.1109/TKDE.2005.99

[6] H. R. Zhang, F. Min, Z. H. Zhang, and S. Wang, "Efficient collaborative filtering recommendations with multi-channel feature vectors," *Int. J. Mach. Learn. Cybern.*, 2019. https://doi.org/10.1007/s13042-018-0795-8

[7] J. Feng, X. Fengs, N. Zhang, and J. Peng, “An improved collaborative filtering method based on similarity,” *PLoS One*, Vol. 13, No. 9, Pp. 1–18, 2018. https://doi.org/10.1371/journal.pone.0204003

[8] X. Su and T. M. Khoshgoftaar, “A Survey of Collaborative Filtering Techniques,” *Adv. Artif. Intell.*, 2009. https://doi.org/10.1155/2009/421425

[9] A. A. Fakhri, Z. K. A. Baizal, and E. B. Setiawan, “Restaurant Recommender System Using User-Based Collaborative Filtering Approach: A Case Study at Bandung Raya Region,” in *Journal of Physics: Conference Series*, 2019. https://doi.org/10.1088/1742-6596/1192/1/012023

[10] X. Ramirez-Garcia and M. García-Valdez, “Post-filtering for a restaurant context-aware recommender system,” *Stud. Comput. Intell.*, 2014. https://doi.org/10.1007/978-3-319-05170-3_49

[11] J. Zeng, F. Li, H. Liu, J. Wen, and S. Hirokawa, “A restaurant recommender system based on user preference and location in mobile environment,” in *Proceedings - 2016 5th IIAI International Congress on Advanced Applied Informatics, IIAI-AAI 2016*, 2016. https://doi.org/10.1109/IIAI-AAI.2016.126

[12] A. Gupta and K. Singh, “Location based personalized restaurant recommendation system for mobile environments,” in *Proceedings of the 2013 International Conference on Advances in Computing, Communications and Informatics, ICACCI 2013*, 2013. https://doi.org/10.1109/ICACCI.2013.6637223

[13] N. Jonnalagedda, S. Gauch, K. Labille, and S. Alfarhood, “Incorporating popularity in a personalized news recommender system,” *PeerJ Comput. Sci.*, 2016. https://doi.org/10.7717/peerj-cs.63

[14] Z. K. A. Baizal, D. H. Widyantoro, and N. U. Maulidevi, “Factors influencing user’s adoption of conversational recommender system based on product functional requirements,” *Telkomnika (Telecommunication Comput. Electron. Control.*, 2016. http://dx.doi.org/10.12928/telkomnika.v14i4.4234

[15] Y. Zheng, L. Li, and F. Zheng, “Context-awareness support for content recommendation in e-learning environments,” in *2009 International Conference on Information Management, Innovation Management and Industrial Engineering, ICIII 2009*, 2009. https://doi.org/10.1109/ICIII.2009.434

[16] A. S. Dharma and T. Samosir, “The User Personalization with KNN for Recommender System,” Vol. 3, No. 2, Pp. 45–48, 2019.

[17] J. S. Breese, D. Heckerman, and C. Kadie, “Empirical Analysis of Predictive Algorithms for Collaborative Filtering,” Pp. 43–52, 2013.

[18] G. A. Pradnyana, I. K. A. Suryantara, and I. G. M. Darmawiguna, “Impression Classification of Endek (Balinese Fabric) Image Using K-Nearest Neighbors Method,” *Kinet. Game Technol. Inf. Syst. Comput. Network, Comput. Electron. Control*, 2018. https://doi.org/10.22219/kinetik.v3i3.611

[19] J. W. Yodha and A. W. Kurniawan, “Pengenalan Motif Batik Menggunakan Deteksi Tepi Canny Dan K-Nearest Neighbor,” *Techno.COM*, 2014.

[20] K. Alkhatib, H. Najadat, I. Hmeidi, and M. K. A. Shatnawi, “Stock Price Prediction Using K-Nearest Neighbor Algorithm,” *Int. J. Business, Humanit. Technol.*, 2013.

[21] R. Saptono, H. Prasetyo, and A. Irawan, “Combination of cosine similarity method and conditional probability for plagiarism detection in the thesis documents vector space model,” *J. Telecommun. Electron. Comput. Eng.*, Vol. 10, No. 2–4, Pp. 139–143, 2018.

[22] F. Mohammadi, “A New Approach To Focused Crawling : Combination of Text summarizing With Neural Networks and Vector Space Model,” Vol. 2, No. 3, Pp. 31–36, 2013.

[23] J. Bobadilla, F. Ortega, A. Hernando, and A. Gutiérrez, “Recommender systems survey,” *Knowledge-Based Syst.*, 2013. https://doi.org/10.1016/j.knosys.2013.03.012

[24] J. Lu, D. Wu, M. Mao, W. Wang, and G. Zhang, “Recommender system application developments: A survey,” *Decis. Support Syst.*, 2015. https://doi.org/10.1016/j.dss.2015.03.008