Application of Content-Based Method for Indonesian Food Recommendation System

Ganesh Ekatata Buana
School of Computer Science
Bina Nusantara University
Jakarta, Indonesia
ganesh.buana@binus.ac.id

Indah Bahrina Mursyida
School of Computer Science
Bina Nusantara University
Jakarta, Indonesia
indah.mursyida@binus.ac.id

Kelly Lanjaya
School of Computer Science
Bina Nusantara University
Jakarta, Indonesia
kelly.lanjaya@binus.ac.id

Ajeng Wulandari
School of Computer Science
Bina Nusantara University
Jakarta, Indonesia
ajeng.wulandari@binus.ac.id

Abstract *– Indonesian food covers a really wide variety of unique recipes throughout Indonesia. Meanwhile, Indonesians are often still confused in picking food, despite the numerous different Indonesian food recipes they can try. This issue can be mainly caused by lack of knowledge to Indonesian food recipes and we are trying to resolve this by creating a food recommendation system which provide users with recommendations on which Indonesian recipes they should try. We created this recommendation system by using a content based method based on user preferences and food ingredients. Using CountVectorizer to vectorize the ingredients of each recipe and Cosine Similarity to calculate similarity between the recipes, this recommendation system is aimed to solve user's problems by recommending similar recipes based on recipe ingredients similarity to each other according to the user's personal recipe preferences. The system successfully covered recommendations based on Indonesian food recipes using main ingredients such as chicken, fish, lamb, beef, tofu, eggs, tempeh and shrimp.*

***Keywords*—Recommendation System · Indonesian Food · Content-Based Method · User Preferences · Food Ingredients**

I. Introduction

Indonesia is an archipelagic country that has various races, cultures, and special foods in each region. Food that is spread in various regions has its own taste and uniqueness. Indonesian people have many types of food and they use spices as the main ingredient for their cooking. We know that in ancient times these spices were valuable items for Indonesia. So, now a lot of people who like Indonesian food. In fact, people from abroad also like Indonesian food.

Sometimes, in choosing food someone must be confused by the choice of food they want to eat. In addition, there are also several groups of people who need help in determining food because they have allergies to certain food ingredients and groups of people who have a desire to avoid certain foods in order to maintain their health. This problem arises because the choice of food is very large. So, a person becomes difficult in making decisions about what food to eat. In addition, technology has developed very rapidly and has influenced human life in various aspects. One of the impacts is the influence of technology on human food access. Online food delivery services are now very much appearing in our applications. This can help us in ordering the food and drinks we want whenever and wherever we are. However, on the other hand this also makes it difficult for us to make decisions.

To help Indonesians choose food, we can use a food recommendation system to provide recommendations according to user preferences. Not only that, users can also use this system to get the composition of the food in each food they want to eat. In making a recommendation system, the Content Based method will be used in providing recommendations based on user preferences. This method is a method that adjusts food recommendations according to the tastes of each user. By using the content-based method, we can analyze information about preferences and descriptions of items desired by users and get food recommendations that are suitable for users. From the existing problems, we took the initiative to try to make a simple Indonesian food recommendation system to help Indonesian people in determining their food choices.

II. Literature Review

Changes in the current industrial sector and increasing competition between companies encourage companies to innovate in implementing the right strategy to retain loyal customers and bring in new customers [1]. Nowadays users are also given a lot of choices, so there is a need to filter and convey information efficiently to minimize the problem of information overload [2]. Food recommendation systems are important for several reasons. One of them is because users can get cooking inspiration through digital sources that are currently popular and there are recommendations for types of food such as food in restaurants or products in supermarkets [3][4]. The recommendation system can also be used to recommend products to users based on interests, preferences and various other factors [5]. In addition, researchers have studied such systems for years to tell us that this recommendation system can not only to be a means of helping people who want to find the food they really want to eat, but it can also help them to eat healthier [6]. However, of course there are various kinds of challenges involved in designing a food recommendation system [7].

Food recommendation system is also an important domain for both individuals and society, so user preferences must remain a focus in the food domain because experiments described in other studies show that regardless of the source, user feedback is necessary, namely ratings, tags or comments. However, this alone cannot provide satisfactory performance [8]. The food recommendation system has currently become one of the most

popular software developments topics among researchers. In addition, in general the structure of a food recommendation system can be divided into 4, namely an information gathering layer that collects data on user preferences and nutritional needs of the user, a user profile dataset that describes a user's profile that will be used as input in food recommendations, an intelligent system that focuses on receiving a user profile and process it until it becomes an output in the form of recommended food, and a user interface that will deliver the results of the entire process to the user [9].

There are 3 common approach used for the recommendation system, namely collaborative filtering, content-based, and hybrid [10][11][12][13][14][15]. Meanwhile, the most commonly used techniques are Artificial Intelligence techniques such as Knapsack Algorithm, Rule-Based Reasoning, Semantic Web, and Food Ontology [16].

Collaborative filtering is a recommendation based on another comparable customer profile for a very diverse thing settings where it is calculated using one of the different proximity measures to estimate the inspection of customer collections that are not in the data set [11]. In the Collaborative Filtering approach, there are 2 techniques in filtering, namely memory-based filtering and model-based filtering. Where the memory-based filtering technique is a technique that uses the entire nearest neighbor database. This technique is further divided into 2, namely based on user-based and item-based. Meanwhile, the model-based filtering technique is a technique used to provide recommendations by creating a user rating model that can be built using Machine Learning models such as grouping and classification [15].

The content-based approach is a recommendation based on user preferences using the Vector Space Model (VSM) and TF-IDF (Term Frequency-Inverse Document Frequency) [13]. In a content-based approach we can use cosine similarity to receive input from two objects then calculate the similarities between the two objects and return it in the form of real numbers [1]. Other studies also use this method as in research [13]. The author uses this method in creating a recommendation system to provide accurate recommendations with less computational complexity and offer general recommendations for each user based on the popularity and genre of the film. Meanwhile, the hybrid approach is a combination of content-based and collaborative approaches [5].

Research in the topic of this food recommendation system can be divided into several groups based on the purpose of the application made and the references used to recommend food. Some research groups focus on recommendation systems based on user preferences [17][18], several other research groups focus on recommendation systems based on user nutritional needs or nutritional content in food [19][20], while the other research group incorporates users' nutritional preferences and needs as a consideration in their recommendation system [21][22][9].

The food recommendation system based on user preferences uses data from users in the form of food composition that the user likes [17][18], as well as user history and ratings of food recipes that have been cooked before [18]. To process this input, [17] uses a content-based filtering algorithm, while [18] uses its own algorithm, namely MF-T which is a combination of using matrix factorization and tags. They also succeeded in proving that the use of tags can significantly increase the accuracy and accuracy of the recommendation system.

In designing a food recommendation system, the social situation and recommendations for groups also need to be considered more concretely. The history of the user's food choices and how this can be affected also needs to be considered by the technology system [23]. Another reason for the importance of the role of the food recommendation system is social and economic reasons that have become the focus of research in food recommendations aimed at nutritional health [24][25]. Researchers have proposed various methods for including nutrients (nutritional components in algorithms, meal plans, and encouragement) [26]. The food recommendation system based on the user's nutritional needs uses data on the user's specific nutritional needs to suggest foods that are suitable for the specific needs or diseases of each user [19][20]. [19] proposed the K-DLRS (K-clique Deep Learning Classifier-Based Recommendation System) technique in their research and it was proven that this technique had better accuracy and performance than other techniques such as MLP, RNN, Logistic Regression, Naïve Bayes . Meanwhile [20] apply Fuzzy Logic, JENA Inference, and Knapsack Algorithm in their recommendation system.

The food recommendation system based on the user's nutritional preferences and needs uses the user's nutritional needs as a reference with additional user preferences as an additional determinant, this recommendation system is more widely used because it has better accuracy in recommending nutritious food which is liked by users [21][22][9]. These studies have succeeded in creating a food recommendation system that can adjust the nutritional needs of users without forgetting the different food preference factors for each person. The techniques and algorithms used also vary, such as [21] which uses the Ant Colony Optimization Algorithm, [22] which uses Harris-Benedict equations, Restriction Rules, and Heuristic Functions, and [9] using AHPSort, Pairwise Comparison, Optimization-Based Approach, and Probabilistic Approach.

Apart from research that focuses on creating food recommendation systems, there are also studies that focus on comparing the most optimal algorithms, techniques or methods for use in food recommendation systems. Such as research [16] which summarizes the techniques and methods commonly used by food recommendation systems in other studies as a whole, and research [27] which found Cascaded Clustered Multi-Step Weighted Bi-Partite Graph Projection as the most effective technique in predicting the restaurant or food that customers like based on similar user reviews of other restaurants or foods.

III. METHODOLOGY

In this study, we used several research methods such as content-based filtering, natural language processing (NLP), and cosine similarity.

1. Content-Based Filtering

   Content-based method is a method that adjusts food recommendations according to the tastes of each user. Content-based methods provide the recommendations by analyzing the description of the items that have been rated by the user and the description of items to be recommended [28]. From the information that has been provided, the system will make appropriate food recommendations for users. So, content-based filtering is to tag products using certain keywords, understand what the user likes, look up those keywords in the database and recommend different products with the same attributes [29].

2. Natural Language Processing

   Natural Language Processing (NLP) is a theoretically motivated range of computational techniques for analyzing and representing naturally occurring texts at one or more levels of linguistic analysis for the purpose of achieving human-like language processing for a range of tasks or applications [30]. Natural language refers to how

human communicate with each other, using words and sentences through either texts or vocal. Thus it can be taken that Natural Language Processing is a process of manipulating natural languages into a form in which computers can understand, interpret, and manipulate.

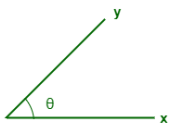
Natural Language Processing is often implemented for text and data analytics, along with several advanced usage such as spell check, predictive texts, voice chats, spam filters, related keywords on search engines, language translation, and smart assistants. Therefore, Natural Language Processing is a wide field to explore and it consists of various distinct techniques available depending on the aim of a project.

Further into text analytics, text feature extraction is often used to tokenize the collection of text documents, learn the vocabulary, and encode new documents [31]. Bag-of-Words (BoW) model is a common method for text feature extraction. It can represent texts in the form of word occurrences within a document by creating a list of known words and calculating the numbers of those words. Bag-of-Words ignores the order of words in the document and depends only with the word occurrence. The Scikit-learn library in Python provides several tools in making Bag of Words, such as CountVectorizer and TfIdfVectorizer. CountVectorizer is a basic vectorizer used to count total occurrences of words in a document. Whilst TfIdfVectorizer is a vectorizer used to calculate the weight of each words in a document according to how relevant the word is in the document, the more frequent the word is, the less impactful the word will be, resulting in a lower weight of the word and vice versa.

3. Cosine Similarity

The algorithm we use to get the similarity of one food to another is cosine similarity. Cosine similarity is used to measure how similar the data are irrespective of their size. Mathematically, cosine similarity is measured by the cosine of the angle between two vectors projected in a multi-dimensional space. In this context, the two vectors are arrays containing the word counts of two documents.

Cosine Similarity between two vectors:



In Cosine Similarity, even though the two data are far apart from the Euclidean distance due to the size of the data, the two data can still get a smaller angle. Smaller the angle, higher the similarity.

Formula of Cosine Similarity [32]:

$$Cos(x, y) = x \cdot y / \|x\| * \|y\| \quad (1)$$

Where,

x . y = product (dot) of the vectors ‘x’ and ‘y’.

||x|| and **||y||** = length of the two vectors ‘x’ and ‘y’.

||x|| * ||y|| = cross product of the two vectors ‘x’ and ‘y’.

IV. Result and Discussion

The dataset that we use is Indonesian Food Recipes [33] taken from Cookpad which consists of 8 different datasets and grouped based on main proteins such as chicken, fish, goat, beef, tofu, eggs, tempeh, and shrimp. From all existing datasets, we combine all datasets into one dataset. Once combined, the total of all data is 15641 different food recipe lines. Then, we process this dataset using Python in Google Collab. We also use libraries like pandas, numpy, and scikit learn to perform CountVectorizer and Cosine Similarity.

First, import all dataset into the dataframe via Google Collab session storage. Then, all data frames will be combined into one same dataframe using the concat feature. Then, we'll reset the index so that the new dataframe doesn't have multiple and repeated indexes. This needs to be done so that each recipe can be recognized by the index as its unique identifier. So, we need to reset the default index of each data frame so that each recipe has a different index.

Next, we use the CountVectorizer feature from the scikit learn library to count the number of each different word in the ingredients column into a 2-dimensional matrix. After the matrix containing the number of each word is formed, we use the cosine similarity feature of the scikit learn library to calculate the similarity of one recipe to another based on the word composition in the ingredients column of each recipe, where a recipe will have a value of 1 when compared to the recipe itself as the similarity value, the highest, with 0 being the lowest similarity value.

Furthermore, we will input the recipe that is preferred by the user as a reference for this recommendation system, where the system will take the cosine similarity matrix column that has been created according to the recipe index. In this column, you can see the level of similarity of all recipes from index 0 to 15640. From here, we sort the existing similarity values and sort them based on the highest similarity value. Then, we create a loop to display the recipe names in the top position that have the highest similarity to the recipe liked by the user as a recipe recommendation that is suitable for the user. It should be noted that the recipe that the user likes will always rank first with the highest similarity value of 1.

Table 1. Food recipe recommendations for user preference 'Ayam Woku Manado'

RecipeIndex	RecipeTitle	SimilarityValue
0	Ayam Woku Manado	1.0
420	Ayam Garang Asem	0.7309919961943783
2711	Ikan tongkol woku	0.7252363135986415
15312	Sambal goreng kentang dan udang	0.7223460919122201
86	Ayam bumbu Betutu	0.7167142005590647
2882	Woku Ikan Nila	0.6979377209210283
11463	Gulai telur ayam	0.6922321755348866
2118	Ikan Gurame Acar Kuning	0.6902614651513899
9505	Pepes Tahu Putih	0.6873997801514456
11248	Telor balado	0.6873113742890582

The table above shows ten recipes sorted by their similarity value in descending order to the recipe ‘Ayam Woku Manado’.

Table 2. Food recipe recommendations for user preference 'Siomay Udang Ayam'

RecipeIndex	RecipeTitle	SimilarityValue
15640	Siomay Udang Ayam	1.0
13104	Tempe mendoan	0.7450760631780817
14212	Bakso Ayung (Ayam+Udang)	0.7313574508612275
15360	Udang butterfly	0.719870728700614
12995	Bakso Tempe Saus Tiram	0.7168795835326197
14463	Siomay ayam udang	0.7088695043765847
1807	Nugget ayam simple	0.7082882469748284
1392	Nugget Ayam	0.6982379478974613
12781	Tempe goreng sederhana	0.6969627833961297
13249	Tempe Mendoan	0.6919952095899509

The table above shows ten recipes sorted by their similarity value in descending order to the recipe ‘Siomay Udang Ayam’.

V. Conclusion

In this paper, we research food recommendations using the content-based method. We make food recommendations using Indonesian recipes obtained from Kaggle. Here, we use the CountVectorizer and Cosine Similarity algorithms. We have conducted experiments according to user preferences, where users like "Ayam Woku Manado" and the results of the recommendations given by our system are "Ayam Woku Manado" with a similarity level of 1, "Ayam Garang Asem" with a similarity level of 0.7309, and so on. A suggestion that can be further developed from this research is to use hybrid filtering for the food recommendation system.

References

[1] Fiarni, C., Maharani, H., & Calista, N. (2019). Product Recommendation System Design Using Cosine Similarity and Content-based Filtering Methods. *International Journal of Information Technology and Electrical Engineering*, *3*(2), 42-48. https://doi.org/10.22146/ijitee.45538

[2] Bhatia, G., Shedge, S., Sahetia, S., Mhatre, D., & Gangwani, S. (2020). Food Dicted: A Restaurant & Food Recommendation System, *International Journal of Future Generation Communication and Networking*, 13(1), 284-290.

[3] Runo, M. (2011). FooDroid: A Food recommendation App for University Canteens. Semester Thesis, ETH Zurich.

[4] Herse, S., Vitale, J., Ebrahimian, D., Tonkin, M., Ojha, S., Sidra, S., Johnston, B., Philips, S., Gudi, S. L. K. C., Clark, J., Judge, W., Williams, M. (2018). Bon Appetit! Robot Persuasion for Food Recommendation. *Conference: ACM/IEEE International Conference on Human-Robot Interaction Companion 2018.* doi:10.1145/3173386.3177028

[5] Jeyabharathi, J., Loheswaran, K., Ramaiah, V., & Kumaravel, T. (2020). Restaurant Recommendation System Using Support Vector Machine and Naive Bayes Classifier Machine Learning Algorithms. *International Journal of Future Generation Communication and Networking*, *13*(4), 3710-3714. http://www.sersc.org/journals/index.php/IJFGCN/article/view/34414

[6] Sahoo, A. K., Pradhan, C., Barik, R. K., Dubey, H. (2019). DeepReco: Deep Learning Based Health Recommender System Using Collaborative Filtering. *Computation 2019, 7(2):25*. doi:10.3390/computation7020025

[7] Jiang, H., Wang, W., Liu, M., Nie, L., Duan, L., Xu, C. (2019). Market2Dish: A Health-Aware Food Recommendation System. *MM '19: Proceedings of the 27th ACM International Conference on Multimedia.* 2188-2190. doi:10.1145/3343031.3350594

[8] Naik, A. P. (2020). Intelligent Food Recommendation System Using Machine Learning. *International Journal of Innovative Science and Research Technology, 5(8).* Dharwad, India.

[9] Toledo, R. Y., Alzahrani, A. A., & Martínez, L. (2019). A food recommender system considering nutritional information and user preferences. *Access IEEE* , vol.*7*, pp. 96695-96711.

[10] Bharati, Y.D. (2019). Recommendation System for Video Streaming Websites Based on User Feedback. *International Journal of Engineering and Advanced Technology (IJEAT)*, *8*(6), 1317-1320.

[11] Nilesh, Kumari, M., Hazarika, P., & Raman, V. (2019). Recommendation of Indian Cuisine Recipes based on Ingredients.*IEEE 35th International Conference on Data Engineering Workshops (ICDEW) 2019,* 96-99. https://doi.org/10.1109/ICDEW.2019.00-28

[12] Rajabpour, N., Mohammadighavam, A., Naserasadi, A., & Estilayee, M. (2018). TFR: A Tourist Food Recommender System based on Collaborative Filtering. *International Journal of Computer Applications*, *181*(11), 31-39.

[13] Singh, R.H., Maurya, S., Tripathi, T., Narula, T., & Srivastav, G. (2020). Movie Recommendation System using Cosine Similarity and KNN. *International Journal of Engineering and Advanced Technology (IJEAT)*, *9*(5), 556-559. http://dx.doi.org/10.35940/ijeat.E9666.069520

[14] Hariyale, I., & Raghuwanshi, M.M. (2020). Design of Recommender System using Content Based Filtering and Collaborative Filtering Technique: A Comparative Study. *International Journal of Advanced Science and Technology*, *29*(5), 4852-4865.

[15] Shah, K. (2019). Book Recommendation System using Item based Collaborative Filtering. *International Research Journal of Engineering and Technology (IRJET)*, *6*(5), 5960-5965.

[16] Norouzi, S., Nematy, M., Zabolinezhad, H., Sistani, S., Etminani, K. (2017). Food recommender systems for diabetic patients: a narrative review. *Reviews in Clinical Medicine*, 4(3), 128-130.

[17] Rana, K. S. (2016). *FOOD RECOMMENDATION SYSTEM BASED ON CONTENT FILTERING ALGORITHM* (Doctoral dissertation). Tribhuvan University, Nepal.

[18] Ge, M., Elahi, M., Fernández-Tobías, I., Ricci, F., & Massimo, D. (2015). Using tags and latent factors in a food recommender system. In *Proceedings of the 5th International Conference on Digital Health 2015* (pp. 105-112).

[19] Manoharan, S. (2020). Patient Diet Recommendation System Using K Clique and Deep learning Classifiers. *Journal of Artificial Intelligence*, *2*(02), 121-130.

[20] Chen, R. C., Lin, Y. D., Tsai, C. M., & Jiang, H. (2013). *Constructing a diet recommendation system based on fuzzy rules and knapsack method. In International Conference on Industrial, Engineering and Other Applications of Applied Intelligent Systems (pp. 490-500). Springer, Berlin, Heidelberg.*

[21] Rehman, F., Khalid, O., Bilal, K., & Madani, S. A. (2017). Diet-right: A smart food recommendation system. *KSII Transactions on Internet and Information Systems (TIIS)*, *11*(6), 2910-2925.

[22] Ribeiro, D., Machado, J., Ribeiro, J., Vasconcelos, M. J. M., Vieira, E. F., & de Barros, A. C. (2017). SousChef: Mobile Meal

Recommender System for Older Adults. In *ICT4AgeingWell* (pp. 36-45).
[23] Maia, R., Ferreira, J. C. (2018). Context-Aware Food Recommendation System. *Proceedings of the World Congress on Engineering and Computer Science, 1*. San Francisco, USA.
[24] Schafer, H., Groh, G., Schlichter, J., Kolossa, S., Daniel, H., Hecktor, R., Greupner, T. (2015). Personalized Food Recommendation. *Proceedings of the 2nd International Workshop on Decision Making and Recommender Systems, 1533*. Bolzano, Italy.
[25] Elahi, M., Ge, M., Ricci, F., Fernandez-Tobias, I., Berkovsky, S., David, M. (2015). Interaction Design in a Mobile Food Recommender system. *Proceedings of the Joint Workshop on Interfaces and Human Decision Making for Recommender Systems co-located with ACM Conference on Recommender Systems, 1438*. Vienna, Austria.
[26] Elahi, M., Ge, M., Ricci, F., Massimo, D., Berkovsky, S. (2014). Interactive Food Recommendation for Groups. Foster City, Silicon Valley.
[27] Sawant, S., & Pai, G. (2013). Yelp food recommendation system. Stanford University, California.
[28] Manjula, R., & Chilambuchelvan, A. (2016). Content Based Filtering Techniques in Recommendation System using user preferences. International Journal of Innovations in Engineering and Technology (IJIET), 7(4), 149-154.
[29] Jagtap, S., Mane, Y., Kadam, T., & Dange, T. (2020). Approach Towards Hybrid Recommendation System using Content Based and Collaborative Filtering Techniques. *International Research Journal of Engineering and Technology (IRJET)*, 7(8), 706-709.
[30] Liddy, E.D. (2001). Natural Language Processing. *In Encyclopedia of Library and Information Science*, 2nd Ed. NY. Marcel Decker, Inc.
[31] Liang, Y. (2017). *Yelp Rating Prediction with Sentiment and Topic Models.* https://doi.org/10.17615/6tz2-yg28
[32] Rao, S. (2020). Cosine Similarity. *GeeksforGeek*s. https://www.geeksforgeeks.org/cosine-similarity/.
[33] Wibowo, P. (2019). Indonesian Food Recipes. *Kaggle*. https://www.kaggle.com/canggih/indonesian-food-recipes.