Analysis and implementation of the bi-polar slope one algorithm with the content base filtering method in producing culinary place recommendations in kuningan regency

Endra Suseno*, Rio Priantama, and Dede Irawan
Universitas Kuningan, Kuningan, Indonesia

*endra@unik.ac.id

Abstract. The recommendation system is an analysis model to obtain recommendations that use filtering techniques to produce recommendations as desired. One of the prediction algorithms is Bipolar Slope One which is a development of the Slope One algorithm to predict an event. Previous studies have shown that the results of the recommendations on the Bipolar Slope One algorithm show a mismatch between the type of recommended item and the type of item that has been rated by the active user, because the rating prediction process in the Bipolar Slope One algorithm does not pay attention to the type or content of the item, but pay more attention to the similarity of rating patterns. This study uses the Content-Based Filtering method to form a profile based on the attributes of an item which will make an assessment based on the analysis of the similarity of the user profile with the vector space model. Application users get information about which culinary delights have an item profile that is expected by him. Culinary players get information about what item profile I expect by the User.

1. Introduction
Currently there are many optimization techniques developed in obtaining recommendations that can be provided by a system with various domains with hundreds of thousands of products on the market from entertainment, automotive, daily necessities and even tourist attractions, one of which is culinary tourism. Food is one of the primary human needs. [1] This has led to an increase in the need for food needed, both primary and secondary food needs, with this population growth not only will the need for primary food increase but will provide a stimulus for an increase in food business actors in Kuningan regency, with a large number of eating places continues to increase making it difficult for many people to determine food places that can be recommended. So far, recommendations for tourist destinations usually word of mouth, mass media, browsing, and social media. It makes not all tourist objects can be visited, due to lack of information and location of tourism objects which are scattered so that it is difficult to reach by public transportation. [2]

The recommendation system is one of the innovations in the internet world revolution which has caused the amount of information to increase sharply [3]. The recommendation system was built to provide a model as a facility that can fulfill the wishes of system users [4] and is periodically used to provide suggestions / predictions of items based on user preferences in the past [5]. The statistical
model consisting of the Vector Space Model and the probabilistic model becomes the basis for the recommendation system algorithm to retrieve relevant information [6].

Bipolar Slope One which is an algorithm that can be used to predict an event. This algorithm is a development of the Slope One Algorithm where the data will be broken down to get each polar axis, then the calculation of each polar data is carried out, this recommendation algorithm requires a method that can produce accurate recommendations with the number of indicators that determine it. There are several methods that can be used, namely, classic filtering, collaborative filtering, content base filtering.

In an effort to provide recommendations in determining food culinary places in the Kuningan district and the consideration that the culture of the Kuningan community in choosing a food culinary place is based on recommendations from someone who is already known so that it is still very subjective, this research was carried out by applying the Bipolar Slope One Algorithm by using the method. Content-Based Filtering to create a profile based on the attributes that make up an item which will make an assessment based on the analysis of the similarity of the user profile with the vector component that makes up the item. If the item will be liked, then the item will be recommended. So, this application is expected to help the community in determining culinary choices in Kuningan district based on existing culinary categories.

2. Methodology

2.1 Literature review

Research conducted by Thorat et. al. (2015) in a scientific publication entitled "Survey on Collaborative Filtering, Content-based Filtering and Hybrid Recommendation System" provides a holistic overview of the recommendation system which includes collaborative filtering methods, content-based filtering and hybrid recommender system approaches. [7] Revealed that to improve the quality of recommendations, several hybrid approaches are used, especially for collaborative filtering and content based filtering. Hybrid algorithms are used to integrate location information into existing recommendation algorithms. Meanwhile, Jovita et. al. (2015) in a scientific publication entitled "Using Vector Space Model in Question Answering System" uses the vector space model method to represent knowledge and retrieve answers to each question. [8], where each query will be compared with knowledge based on the measurement of their similarity.

Another research related to recommendation systems is research by Vega Aristyanto (2011) in the Telkom University Journal entitled "Analysis and Implementation of Bipolar Slope One Scheme Collaborative Filtering on Recommender System"[9], using the Collaborative Filtering method on the Bipolar Slope.

2.2 Bipolar Slope One

In a research conducted by Ali Akbar Lubis (2019) in a scientific publication entitled "Bipolar Slope One and BW-Mine Algorithm Accuracy Test in Recommendation Systems" [10], states that the Bipolar Slope One algorithm predicts ratings based on information from other users. In this case the algorithm does not pay attention to the type of film, but does pay attention to the similarity of rating patterns among users. The Bipolar Slope One algorithm takes into account both information from other users who are rated the same item, and from other items rated by the same user. However, the schema also relies on data points that are neither included in the user array nor in the item array, but which are nonetheless important information for ranking predictions.

Much of the power of the approach comes from data that is not taken into account. Especially this paper will use the SLR approach to review research on the Naïve Bayes algorithm with the problem of attribute independence assumptions. Systematic Literature Review (SLR) is a process for identifying, assessing, and interpreting all available research with a view to providing answers to specific RQs[15].
2.3 Weighted Slope One

In many cases, an evenly distributed ranking scale is required for the process to work properly. As in a scientific publication [11], R. Yang et al (2013) entitled "Using Semantic Technology To Improve Recommender System Based On Slope One", states that relevant information can be used to adjust weighting when calculating ranking predictions. This approach is called Weighted Slope One. By using the Weighted Slope One algorithm, we get one prediction of items that the user likes and another prediction uses items that the user doesn't like.

One downside to Slope One is that the number of ratings it observes is not taken into account. Intuitively, to predict the ranking of item user who are given user ratings A for items J and K, if 2000 users rank item pairs J and L while only 20 users rank item pairs K and L, then the user's item J ranks. A tends to be a much better predictor for item L than the user rating A of item K is. So, we defined the Weighted Slope One prediction as following a weighted average.

$$w_{u, j} = \frac{\sum_{i \in S(u)} P_{j,i} P_{j,j} \cdot \sigma_{j,i}}{\sum_{i \in S(u)} \sigma_{j,i}}$$



2.4 Content Based Filtering

Content Based Filtering is a recommendation system utilizing various sources of information to provide users with predictions and recommendations. The recommendation system utilizes the information filtering concept [12]. In information filtering the user already has a profile that represents long-term interests and the system tries to provide each user with the relevant item. Based on the similarity measure between each profile, the system selects and ranks the relevant items, then assigns them to the user. There are two approaches to information filtering, namely collaborative filtering and content-based filtering.

This study will use a content-based filtering approach. The system will select and rank items based on the similarity of the user profile and the item profile. The advantage of this approach is that users gain insight into why an item is considered relevant to them, because the content in each item is known from its representation. However, this approach also has drawbacks, for example the fact that this approach focuses on keyword similarities. This approach is incapable of capturing more complex relationships at a deeper semantic level, based on the various types of attributes associated with the structured object of the text. User interests or preferences are also represented by the same set of features, which are called user profiles. Recommendations are made by comparing user profiles with candidate items expressed in the same feature set. The simplest approach to content-based recommendations is to calculate how the user's profile matches each item. There are several ways to represent so that it can be used as a learning component, one of which is the Vector Space Model.

In this method, document D is represented as an m-dimensional vector. Where each dimension corresponds to a different term and m is the total number of terms used in the document collection. The document vector is written as, $w_i$ is the weight of the terms indicating the level of importance. If document D does not contain the term $t_i$, then the weight of $w_i$ is zero. The term weight can be
determined using the tf-idf scheme [14]. In this approach the weight is calculated based on how often a term appears in a document, and how often it is found in a document collection.

2.5 Vector Space Model

Vector space model is a model used to measure the similarity between a document and a query. In this model, queries and documents are considered as vectors in n-dimensional space, where n is the sum of all terms in the lexicon [15]. The lexicon is a list of all the terms in the index. The vector space model is used because this method allows the document ranking process. This method calculates the cosine of two vectors. The two vectors are the weight of each document and the weight of the query. The document and query weights are used for the document ranking process, as shown in Figure 3 below.

The term weighting in a document vector can be determined in many ways. The common approach used in this final project is the TF-IDF method [12, 13]. In this method, the term weight is determined by two factors: how often term j is contained in document i (term frequency \( t_i, j \)) and how often it appears in the entire document collection (document frequency \( df_j \)). The exact weight of term j in document i is formulated as follows:

\[
wi, j = t_i, j \times idf_j = t_i, j \times \log \frac{N}{df_j}
\]  

\( N \) is the number of documents contained in the document collection. \( idf \) stands for inverse document frequency. This method gives high weight to terms that often appear in a few documents in the document set. After the term weights are calculated, a ranking function is needed to measure the similarity between the query and the vector document. A similarity measure commonly known as a cosine measurement. This measure determines the angle between the document and query vectors when represented in Euclidean V-dimensionality, where \( v \) is the size of the vocabulary. The exact similarity between the \( D_i \) document and the \( Q \) query is defined as follows:

\[
sim(Q, D_i) = \frac{\sum_{i=1}^{v} w_i^Q \times w_i^D}{\sqrt{\sum_{i=1}^{v} w_i^Q \times \sum_{i=1}^{v} w_i^D}}
\]

2.6 Data Collection and Analysis

Literature study and observation are carried out to gather information. Literature study is carried out by studying the theories and literature that support this research, especially those related to recommender systems, bipolar slope one and content-based filtering, as well as the software used to build the system. Observations were made on the object of research, namely culinary places in Kuningan Regency. As the test sample in this study are culinary places which are grouped into several categories. Each culinary place consists of information in the form of: Taste (C), Price (H) and Location (L) which is applied in a Likert scale instrument as in Table xx.
Table 1. Components of Culinary Profile Forming Items

| Item       | 1        | 2        | 3        | 4        | 5        |
|------------|----------|----------|----------|----------|----------|
| Citarasa (C) | Sangat Tidak | Lezat | Sangat Lezat | | |
| Harga (H)      | Sangat Mahal | Sangat Murah | Sangat | | |
| Lokasi (L)     | Sangat Sulit | Sangat Mudah | Sangat | | |

Information from each culinary item becomes a component item obtained from the rating given by the previous user (active user) and will be the data that is tested for its similarity with the component item expected by the user (user).

2.7 Implementation of the Bipolar Slope One Algorithm with the Content Based Filtering Method
The BiPolar Slope One algorithm predicts rating ratings based on information obtained from users. In this case the algorithm does not pay attention to the type of culinary delicacy but analyzes the similarity of user patterns based on profiles. The following is an overview of the implementation of the BiPolar Slope One algorithm by applying the Content Based Filtering method in predicting culinary delights in the form of an Algorithm flowchart.

![Figure 4. Flowchart of bipolar slope one algorithm implementation with a content based filtering approach](image)

2.8 System Development
Management information system development is carried out using the GRAPPLE methodology (Guidelines for Rapid Application Engineering). The GRAPPLE methodology is a simplification of the methodology framework for RAD (Rapid Application Development) system development, so that management information systems can be designed quickly.

3. Result and Discussion
System analysis and design needs to be done before program implementation, as a reference and to produce a good program.

3.1 System Overview
As stated in the background of the problem, the system to be created is used to provide information on recommendations for culinary places in the Kuningan district. The target users of this system are
people who want to know information about culinary places in Kuningan Regency by considering the results of the rating of three component items, namely: Taste, Price and Location. The system provides flexibility for users to get recommendations, give ratings and or both. Each recommendation information from a culinary place that is accessed by certain users will change along with the number of ratings given by other users on the same culinary place object. In the development of this agricultural land recommendation system, the authors collected precise culinary data in Cidahu District, Kuningan Regency.

3.2 Calculation of Recommendations

An example of the calculation is carried out at a culinary place with the category "Restaurant / Rice Stalls" with the code RWN. There are 8 culinary spots in the research object area which are categorized under the RWN code. The User Column in Table 2 is a user query. In Table 2, it can be seen that the user (User) wants a restaurant / rice shop that has a taste scale of 5, a price scale of 4 and a scale of 4.

The Content Based Filtering method used with the vector space model approach is to calculate the angular cosine of two vectors (Cosine Similarity). In this case it is W from the Restaurant / Warung Nasi (RWN) as the item profile with W from the User as the user profile. The value of w has been obtained, as listed in Table 4, where at that stage the weight of the query is multiplied by the weight of the items contained in each RWN record (1-8), then added up for each column.

The next step is to square the query weight. The item weights for each RWN (1-8) are also squared. Then add it up for each column as in Table 5. The final step is to enter the cosine similarity formula as in Table 6.

When viewed from the calculation results, it can be concluded that the closest to the similarity with the merchant profile is the Restaurant / Warung Nasi number 3 (RWN03) with the similarity value reaching 0.99. In the cosine similarity calculation method, the most similar results are those that are close to the value of one.

### Table 2. Weight Before Normalization

| Pengguna | RWN01 | RWN02 | RWN03 | RWN04 | RWN05 | RWN06 | RWN07 | RWN08 |
|----------|-------|-------|-------|-------|-------|-------|-------|-------|
| Citarasa | 5     | 3     | 4     | 4     | 3     | 4     | 3     | 2     |
| Lokerasi | 4     | 2     | 5     | 3     | 4     | 1     | 5     | 4     |

### Table 3. Weight After Normalization

| Pengguna | RWN01 | RWN02 | RWN03 | RWN04 | RWN05 | RWN06 | RWN07 | RWN08 |
|----------|-------|-------|-------|-------|-------|-------|-------|-------|
| Citarasa | 0.6   | 0.8   | 0.8   | 0.6   | 0.6   | 0.8   | 0.6   | 0.4   |
| Lokerasi | 0.8   | 0.4   | 1     | 0.6   | 0.8   | 0.2   | 1     | 0.8   |

### Table 4. Multiplication of Queries and Items.

| Pengguna | P*RWN01 | P*RWN02 | P*RWN03 | P*RWN04 | P*RWN05 | P*RWN06 | P*RWN07 | P*RWN08 |
|----------|---------|---------|---------|---------|---------|---------|---------|---------|
| Citarasa | 1       | 0.6     | 0.8     | 0.6     | 0.6     | 0.8     | 0.6     | 0.4     |
| Lokerasi | 0.32    | 0.8     | 0.48    | 0.64    | 0.16    | 0.8     | 0.64    | 0.8     |

### Table 5. Square of weight

| Pengguna | P*RWN01 | P*RWN02 | P*RWN03 | P*RWN04 | P*RWN05 | P*RWN06 | P*RWN07 | P*RWN08 |
|----------|---------|---------|---------|---------|---------|---------|---------|---------|
| Citarasa | 1       | 0.36    | 0.64    | 0.36    | 0.36    | 0.64    | 0.36    | 0.36    |
| Lokerasi | 0.64    | 0.16    | 1       | 0.36    | 0.64    | 0.04    | 1       | 0.64    |

### Table 6. The calculation results

| RWN01 | RWN02 | RWN03 | RWN04 | RWN05 | RWN06 | RWN07 | RWN08 |
|-------|-------|-------|-------|-------|-------|-------|-------|
| 0.959242 | 0.899271031 | 0.999483871 | 0.955317528 | 0.873158643 | 0.974049244 | 0.976768328 | 0.901230117 |
After ranking, the information that can be used as material to be presented in the user interface is as shown in Table 7.

Table 7. Ranking Recommendations for Users

| Kode  | Cosine     | Rekomendasi Ke- |
|-------|------------|-----------------|
| RWN03 | 0.999483871| 1               |
| RWN07 | 0.976768328| 2               |
| RWN06 | 0.974049244| 3               |
| RWN01 | 0.959241987| 4               |
| RWN04 | 0.955317528| 5               |
| RWN08 | 0.901230117| 6               |
| RWN02 | 0.899271031| 7               |
| RWN05 | 0.873158643| 8               |

3.3 System Architecture
Users interact with the system using a mobile device. Data storage and recommendation process, takes place on a web server.

3.4 Use Case Diagram
Use Case Diagram describes the sequence of activities carried out by actors and the system to achieve certain goals, describes the expected functionality of a system. Use Case Diagrams also present interactions between actors / actors and the system. Figure 6 below explains the Use Case Diagram used in this recommendation system.

3.5 Use Case Implementation
a. Registration and Login Pages
Registration and Login pages are used by users in order to take advantage of this recommended application.
b. Main Page
The main page gives the user the option to choose the option "to give a rating" or "view recommendations" for each culinary category that has been selected through the options in the scroll menu.

c. Rating Page
The Rating page is used for users who want to give ratings to pre-selected culinary categories in terms of: Taste, Price and Location of Culinary Objects that have been visited / want to be rated by the user.
d. Recommendations page
The recommendation page is used by users to get recommendations for culinary objects according to the selected category based on the aspects of Taste, Price and Location.

![Figure 10. Rating Page](image)

4. Conclusion
As an effort to provide recommendations in determining culinary places for food in the Kuningan district, research by applying the Bipolar Slope One Algorithm using the Content-Based Filtering method can provide recommendations for culinary objects according to user expectations. Users are given the flexibility to give ratings to culinary objects and produce item profiles, also get recommendations for which culinary objects are in accordance with user expectations based on the selected category.

The construction of this system uses a bipolar slope one algorithm combined with a content based filtering method and a vector space model approach, providing higher accuracy between item profiles and query items. In order to maintain system performance after use of this system, it would be better if the system performance testing was carried out using either the precision method or the recall method.
References

[1] Dharma P., Seng H., (2017). Aplikasi Rekomendasi Tempat Makan Menggunakan algoritma Slope One Pada Platform Android. IJCCS, Vol 11
[2] Kurniadi, E., Budianto, H. (2018). Rancang Bangun Aplikasi Wisata Kabupaten Kuningan Berbasis Android Menggunakan Metode Location Based Service (LBS). Cloud Information, Vol 3
[3] Deshpande, M.G., Mudeebihalkar, A.V., Jadhav, A.B. & Kokate, S. (2016). Hybrid Content-Based Filtering Recommendation Algorithm on Hadoop. International Journal of Advanced Research in Computer Engineering & Technology, Vol. 5.
[4] Kalita, J., Balas, V.E., Borah, S., & Pradhan, R. (Eds.). (2019). Recent Developments in Machine Learning and Data Analytics, Vol. 740. Singapore: Springer Singapore. Diakses dari: https://doi.org/10.1007/978-981-13-1280-9
[5] Achakulvisut, T., Acuna, D.E., Ruangrong, T., & Kording, K. (2016). Science Concierge: A fast content-based recommendation system for scientific publications. PLoS ONE, 11(7). Diakses dari: https://doi.org/10.1371/journal.pone.0158423
[6] Singh, J.N. & Dwivedi, S.K. (2012). Analysis of Vector Space Model in Information Retrieval. International Journal of Computer Application, 110(4), 31–36. Diakses dari: https://doi.org/10.5120/19308-0760
[7] Jovita, Linda, Hartawan, A., & Suhartono, D. (2015). Using Vector Space Model in Question Answering System. In Procedia Computer Science, Vol. 59, pp. 305–311. Elsevier. Diakses dari: https://doi.org/10.1016/j.procs.2015.07.570
[8] Aristyanto, Vega. (2011). Analisis dan Implementasi Bipolar Slope One Scheme Collaborative Filtering pada Recommender System. Universitas Telkom. Bandung
[9] Lu, Z., Dou, Z., Lian, J., Xie, X. & Yang, Q. (2015). Content-Based Collaborative Filtering For News Topic Recommendation. In AAAI 2015: Proceedings of the Twenty-ninth AAAI Conference on Artificial Intelligence, pp. 217–223. Diakses dari: https://doi.org/10.1080/714965461
[10] Jing, L., Ng, M.K. & Huang, J.Z. (2010). Knowledge-based vector space model for text clustering. Knowledge and Information Systems, 25(1), 35–55. Diakses dari: https://doi.org/10.1007/s10115-009-0256-5
[11] Gadje, J., Sane, S. & Kekre, H.B. (2015). Performance Analysis of Layered Vector Space Model in Web Information Retrieval. International Journal of Applied Information Systems, 8(5), 7–15. Diakses dari: https://doi.org/10.5120/ijais15-451320