ISSN: 2581-1274


# Animal Nutrition Selection Recommendation System with Content Based Filtering Method

Andrian Setya Putra[a,1,*], Pipin Widyaningsih[b,2], Sri Sumarlinda[b,3]

[a] *Faculty of Computer Science, Informatics Engineering, Duta Bangsa University, Indonesia*
[b] *Faculty of Computer Science, Information Systems, Duta Bangsa University, Indonesia*
[1] *202030364@mhs.udb.ac.id*;* [2] *pipin_widya@udb.ac.id;* [3] *sri_sumarlinda@udb.ac.id*
*corresponding author

ARTICLE INFO

*Article history:*
Published

*Keywords:*
Content-Based Filtering
Cosine Similarity
Animals
Nutrition
Recommendation System

ABSTRACT

In this digital age, technology can make everyday life easier, including finding and providing food for livestock and pets. Hilda Pakan Supplier is a business that provides various types of nutrition, such as feed, vitamins, and veterinary medicine. At the time of observation, some customers were often confused in determining the right feed, vitamins, or medicine for their animals due to the lack of information they had. Customers usually have to ask the shopkeeper, which can take a long time especially when the shop is crowded. To overcome these problems, the purpose of this project is to use the content-based filtering method to create a recommendation system for animal nutrition choices. This method was chosen because the recommendation process is simpler and does not require rating data from other users. The system can provide recommendations in accordance with the needs and preferences of animals. In its application, the cosine similarity algorithm is used to measure product similarity based on text descriptions. The results show that this recommendation system is able to produce 10 products with the highest similarity value, which is 0.331060048. In the future, this system can be further developed by adding more products to provide more accurate insurance results.



## I. Introduction

In this era of digital technology, technology has become an integral part of everyday life, for example in the field of animal husbandry and the livestock industry. One important aspect in this industry is the selection of proper nutrition for livestock and pets. Farm animals are animals produced for the purposes of food production, industry, services and agriculture[1]. Pets are animals whose lives are determined by humans for specific reasons. Common pets are cats, fish, chickens, turtles, birds, dogs, etc [2]. To face the need for online commerce in this digital era, many stores choose to expand their services by conducting online sales transactions[3]. The utilization of this technology can be applied to the pet food retail industry, for example by creating an application or system that makes it easier for customers to find pet food. One solution is to use a web-based recommendation system to help customers easily find the right nutrition their pets need.

Based on research conducted by Fajriansyah and his colleagues entitled Implementation of Content-Based Filtering Using TF-IDF and Cosine Similarity for Cuisine Recipe Recommendation System. In this study, to obtain recommended results, the Content-Based algorithm is used by looking at the similarity of word weights in the vocabulary of the pre-processing results of movie synopsis and movie titles. The calculation is done using the standard TF-IDF method. After that, the weighting results will be selected by type after passing through the cosine-similarity level to determine the similarity weight. An accurate value was achieved using the average accuracy @K (MAP@K) of 0.823254 for one type of query and 0.7500556 for several types of seed query based on test results conducted on a total of 4000 movie titles. According to these findings, the single query type outperforms the many seeds query type in terms of recommendations[4].

Hilda Feed Supplier is a business engaged in trading business. This business offers various types of nutritious food such as feed, vitamins and various veterinary medicines to attract customers and compete with other businesses in pet shops or suppliers. In the current business process, customers come directly and the salesperson will ask the customer's needs.

There are various types of animal nutrition such as feed, vitamins, and medicines that can support growth or treat various types of diseases in animals, both pets and livestock. Nutrition comes from the word “nutrition” which means healthy food[5]. The body requires trace amounts of vitamins for several metabolic functions[6]. Medicine is a substance that can affect life processes and its composition used to prevent, treat, diagnose a disease, or cause a certain condition[7].

Based on observations, customers sometimes experience confusion in determining the type of nutrition required by their animals. Usually customers solve their confusion by asking the shopkeeper directly or looking for information on search engines on the internet. Currently, the service at the Hilda Pakan Supplier store still uses manual methods. The manual process is considered less efficient in making sales where when customers are confused and need advice to determine which product is suitable they will ask the seller, but when the shop is busy it will slow down the performance of the shop owner whose impact makes other customers wait. In order to solve these issues, a system of recommendations must be in place that may counsel or suggest the best nutrition for consumers to select when seeking references, taking into account their individual needs and purposes.

This research aims to design an animal nutrition selection recommendation system application with the content-based filtering method. The reason for choosing content-based filtering is because the recommendation process is considered simpler because it does not require rating data from other users (reviews) as a basis for making recommendations. The system is expected to provide nutritional recommendations that are tailored to the needs and preferences of animals, while helping animal owners easily choose foods that are more suitable for their animals.

## II. Method

The method used in this research is the Content Based method with Cosine Similarity and TF-IDF. The following is an explanation of the method:

### *A. Cosine Similarity*

Cosine Similarity is a technique that uses the angle's cosine to determine how similar two lines are to one another. The degree of resemblance of papers is determined using the Cosine resemblance approach, which does not require frequency[8]. The following formula is used to find the similarity angle listed in Formula 1:

$$similarity = \cos(\theta) = \frac{(A \,.\, B)}{||A||\,||B||} \quad (1)$$

Where:

- A = The vector that needs to have its similarity determined
- B = The vector that needs to have its similarity determined
- A. B = The vectors A and B's dot product (inner product).
- ||A|| = vector A's length
- ||B|| = vector B's length

### *B.* TF-IDF

One widely used weighting technique that is thought to be inappropriate for text clustering is TF-IDF[9]. TF-IDF is a modelling and natural language processing technique. The term frequency in a text is measured by TF. One popular method for calculating TF is to divide the total number of words in the manuscript by the number of times a word appears. IDF takes into account the word's significance in the document in the interim. The number of papers with that name divided by the total number of documents yields the IDF. The scale is then adjusted by taking the result logarithmically. The TF-IDF approach generates a word weight for each word in the document by multiplying the TF

and IDF values. The significance of each word in the text is indicated by its weight in relation to a broader collection of words[10]. The formula for calculating TF (t, d) and IDF (t, D) is as follows:

$$\mathrm{TF(t,d)} = \frac{\mathrm{f(t,d)}}{Td} \tag{2}$$

$$\mathrm{IDF(t,D)} = \mathrm{Log}\left(\frac{N}{\mathrm{DF(t)}}\right) + 1 \tag{3}$$

Where:

- f(t, d) = The number of times the term t appears in document d
- Td = Document d's word count
- N = is the total count of documents in set D.
- DF(t) = Total number of documents in collection D that contain the word t

TF-IDF is calculated by the following formula:

$$\mathrm{TF-IDF(t,d,D)} = \mathrm{TF(t,d)} \times \mathrm{IDF(t,D)} \tag{4}$$

Where:

- t = The word or phrase being considered
- D = Document set
- d = Document being assessed
- TF (t, d) = Word t frequency in document d
- IDF (t, D) = Word t in document set's inverse document frequency D

## III. Results and Discussion

The process of selecting animal feed recommendations is carried out through several stages such as word processing, TF-IDF measurement, and cosine similarity calculation. This research will calculate the number of products that will be recommended if the customer chooses the keyword "animal feed for adult chickens":

### *A. Preprocessing*

Preprocessing is a step of selecting raw data to be processed on each document, including tokenization, text folding, filtering, and derivation. The main purpose of this step is to be able to represent each document with strokes in a vector by separating the words that make up the document.

Perform text preprocessing on product name, age, feed category, animal type, and description so that the data used can be processed into numbers with TF-IDF and cosine similarity. The stages in this process start from cleaning, case folding, tokenization, lemmatization and stop word removal. Pre-processing is done on items that only contain keywords so that the search process does not take a long time, which in this calculation uses product data as many as 10 documents. The results of preprocessing can be seen in the following table.

Table 1. Text Preprocessing Result Table

| ***Doc*** | ***Product*** | ***Preprocessing Result*** |
|---|---|---|
| Q | animal feed for adult chickens | feed animal adult chicken |
| 1 | Egg Stimulan 5 Gram Medion | egg stimulant gram medion adult vitamin chicken egg stimulant medicine vitamin acu egg productivity rate chicken poultry profit egg stimulant good egg production rate ration efficiency resistant egg production sickness long egg production rate fertile egg quality rate chicken seedlings |
| 2 | Vita Chicks 5 Gram | vita chicks gram vitamin chickens vita chicks ready powder sibling combination multivitamin complete prebiotic sodium butyrate formulation support the work of chickens fast growing sodium butyrate function energy source auxiliary motility level motion villi press grow bad bacteria gastrointestinal tract vitamins a e c antioxidant good protect the body free |

| *Doc* | *Product* | *Preprocessing Result* |
|---|---|---|
| | | radicals immune system level vitamin b complex carbohydrate metabolism protein fat yield energy level work rate |
| 3 | Hiprovite 781 Fish Pellets | fish pellets hiprovite kids adults fish feed hi pro vite mm kg fish feed floating feed special formulation catfish farming superior quality raw materials select selectively belonging to the nutrition of the protein content according to grow catfish atractant strong catfish appetite stimulant feed size according to open the mouth of the fish easily digested good special formulation level of endurance catfish maximum growth results production watch out strict quality control guarantee good product results net weight grams |
| 4 | BR1-SS Wonokoyo without AGP 1 KG | br ss wonokoyo agp kg kid day chicken feed br ss for chicks broiler meat protein according doc day |
| 5 | Buras Wonokoyo 1 KG | buras wonokoyo kg adult chicken feed buras wonokoyo adult chicken feed package kg kg moisture content max crude protein crude fat min crude fiber max ash max calcium phosphor aflatoxin max ppb ingredients corn meal soybean gluten corn rice bran palm oil |
| 6 | BioJANNA 6 For Bird | biojanna for bird adult vitamin bird medicine product name b ml biojanna vitamin bird chirping level sound agile metabolism digestive system appetite lost odor vitamin category supplement use so chirping bird sounding loud bird getting more agile agile bird body metabolism smooth bird fit healthy bird body healthy odor lost gloss bright bird feathers prevent attack sick less risk of death bird lost pollution odor good digestive system cage healthy appetite level disposable drops drink bird ml water bathing bird type magpie bird magpie bean anis poskay etc ingredients calcium magnesium iron zeng pospho vitamin c |
| 7 | Concentrate 124 Hi-Pro-Vite Pokphand 500 gr | concentrate hi pro vite pokphand gr adult chicken feed concentrate phokpand concentrate used livestock poultry eggs belong to good quality quality use chicken concentrate stable egg corn katul cp mineral |
| 8 | Vitakraft Menu Vital Honey Canaries 500gr | vitakraft vital menu honey canaries gr adult bird feed vitakraft vital menu honey canaries gr content bird feed canaries strong bones endurance level less stress smooth digestion full need fiber nutrition level stamina so bird healthy prime bladder ingredients various seeds herbs select fish oil honey vitamin calcium phosphorus mineral nutrition protein etc use teaspoon gr day bird routine |
| 9 | Sick Bird Digestive Medicine / Bird Respiratory Medicine “SUPER N++” 15ml | sick bird digestion medicine super bird breath medicine n ml adult bird medicine super bird medicine n sick bird medicine netto ml indication medicine for infection of the digestive tract diarrhea poop chalk blood etc. for infection of the respiratory tract shortness of breath flu snoring etc. healed ulcers wounds lethargic birds dull feathers etc. sick medicine caused by microorganisms viruses bacteria etc. shake use store cool avoid sunlight |
| 10 | BR 2 Wonokoyo 1 KG | br wonokoyo kg adult harvest chicken feed br feed protein percent vitamin mineral complete protein feed need fast chicken growth rate chicken day age |

*B. Term Frequency (TF)*

After preprocessing is done, the next stage is word weighting using the TF-IDF method[11]. The term frequency in a text is measured by TF. One popular method for calculating TF is to divide the total number of words in the manuscript by the number of times a word appears. This is the table used to calculate TF.

Table 2. TF Calculation Table

| **Word** | **Term Frequency** | | |
|---|---|---|---|
| | ***Q*** | ***Egg Stimulan 5 Gram Medion*** | ***Vita Chicks 5 Gram*** |
| feed | 0.25 | 0 | 0 |
| animal | 0.25 | 0 | 0 |

| Word | Term Frequency | | |
|---|---|---|---|
| | ***Q*** | ***Egg Stimulan 5 Gram Medion*** | ***Vita Chicks 5 Gram*** |
| chicken | 0.25 | 0.037037 | 0.020833 |
| adult | 0.25 | 0.037037 | 0 |

*C. Inverse Document Frequency (IDF)*

IDF calculates a word's significance within a text. Words with a higher IDF are typically those that are less common in manuscripts. The total number of documents divided by the number of papers that include the phrase yields the IDF.

Table 3. IDF Calculation Table

| Word | DF | IDF |
|---|---|---|
| | | ***Log(n/DF)*** |
| feed | 6 | 0.22184875 |
| animal | 1 | 1 |
| chicken | 6 | 0.22184875 |
| adult | 8 | 0.096910013 |
| ... | ... | ... |

*D. TF-IDF Weighting*

Calculate the weight with the TF x IDF formula, and get the results as in the following table.

Table 4. TF-IDF Calculation Table

| Word | TF-IDF | | |
|---|---|---|---|
| | ***Q*** | ***Egg Stimulan 5 Gram Medion*** | ***Vita Chicks 5 Gram*** |
| feed | 0.350411 | 0 | 0 |
| animal | 0.528486 | 0 | 0 |
| chicken | 0.352865 | 0.052276 | 0.029405 |
| adult | 0.349895 | 0.051836 | 0 |
| ... | 0.350411 | 0 | 0 |

*E. Cosine Similarity Calculation*

Similarity vectors are calculated for each document using the cosine similarity formula. Calculate the product of Q and the other 10 documents. The following table displays the results of summing the product of each column with Q.

Table 5. Cosine Similarity Calculation Table

| Word | Q Weight * Document Weight | | |
|---|---|---|---|
| | ***D1*** | ***D2*** | ***D3*** |
| | 0 | 0 | 0.000286145 |
| | 0 | 0 | 0 |
| | 0.000455712 | 0.000256338 | 0 |
| | 0.000086959 | 0 | 0.000054602 |
| **Total** | 0.000542671 | 0.000256338 | 0.000340747 |

*F. Vector Length Calculation*

Calculate the vector length of each document, including Q, by summing the vector results of each word in the total column, then squaring the total. The following table displays the vector length calculation results.

Table 6. Vector Length Calculation Table

| Word | Vector Length | | | |
|---|---|---|---|---|
| | ***\|Q\|*** | ***\|D1\|*** | ***\|D2\|\|*** | ***\|D3\|*** |
| | 0.003076054 | 0 | 0 | 0.000026618 |
| | 0.062500000 | 0 | 0 | 0 |
| | 0.003076054 | 0.000067513 | 0.000021361 | 0 |
| | 0.000586972 | 0.000012883 | 0 | 0.000005079 |
| | ... | ... | ... | ... |
| **Total** | 0.06923908 | 0.03437395 | 0.02042067 | 0.02220585 |
| | | | | |
| **Root** | 0.2631332 | 0.18540213 | 0.1429009 | 0.14901626 |

To determine how similar columns D1 through D10 are to one another, use the cosine similarity formula. The following table displays the computation.

Table 7. Similarity Calculation Table

| ***No.*** | ***Product Name*** | ***Cosine Value*** |
|---|---|---|
| 1 | Egg Stimulan 5 Gram Medion | 0.011123613 |
| 2 | Vita Chicks 5 Gram | 0.006817139 |
| 3 | Hiprovite 781 Fish Pellets | 0.00869005 |
| 4 | BR1-SS Wonokoyo without AGP 1 KG | 0.03100452 |
| 5 | Buras Wonokoyo 1 KG | 0.020864403 |
| 6 | BioJANNA 6 For Bird | 0.001315462 |
| 7 | Concentrate 124 Hi-Pro-Vite Pokphand 500 gr | 0.331060048 |
| 8 | Vitakraft Menu Vital Honey Canaries 500gr | 0.012134155 |
| 9 | Sick Bird Digestive Medicine / Bird Respiratory Medicine “SUPER N++” 15ml | 0.00168597 |
| 10 | BR 2 Wonokoyo 1 KG | 0.045328043 |

The suggested product search order with the keyword adult chicken feed can be seen in the following table.

Table 8. Similarity Calculation Table

| ***Doc*** | ***Product Name*** | ***Cosine Value*** | ***Rank*** |
|---|---|---|---|
| D7 | Concentrate 124 Hi-Pro-Vite Pokphand 500 gr | 0.331060048 | 1 |
| D10 | BR 2 Wonokoyo 1 KG | 0.045328043 | 2 |
| D4 | BR1-SS Wonokoyo without AGP 1 KG | 0.03100452 | 3 |
| D5 | Buras Wonokoyo 1 KG | 0.020864403 | 4 |
| D8 | Vitakraft Menu Vital Honey Canaries 500gr | 0.012134155 | 5 |
| D1 | Egg Stimulan 5 Gram Medion | 0.011123613 | 6 |
| D3 | Hiprovite 781 Fish Pellets | 0.00869005 | 7 |
| D2 | Vita Chicks 5 Gram | 0.006817139 | 8 |
| D9 | Sick Bird Digestive Medicine / Bird Respiratory Medicine “SUPER N++” 15ml | 0.00168597 | 9 |
| D6 | BioJANNA 6 For Bird | 0.001315462 | 10 |

Based on the table above, the order of the most recommended is “Concentrate 124 Hi-Pro-Vite Pokphand 500 gr”, “BR 2 Wonokoyo 1 KG”, “BR1-SS Wonokoyo without AGP 1 KG”, “Buras

Wonokoyo 1 KG", "Vitakraft Menu Vital Honey Canaries 500gr", "Egg Stimulan 5 Gram Medion", "Hiprovite 781 Fish Pellets", "Vita Chicks 5 Gram", "Sick Bird Digestive Medicine / Bird Respiratory Medicine "SUPER N++" 15ml" and "BioJANNA 6 For Bird".

## IV. Conclusion

Content-Based Filtering uses product similarity to recommend specific content to users based on the user's preference profile. This method does not require parameters such as user reviews to make recommendations. The cosine similarity algorithm is suitable for data that has no physical value such as similarity between documents based on the similarity of words in the text. Based on the calculation using content-based filtering, 10 recommendation products are generated with the highest final result with a similarity value of 0.331060048. For further research, perhaps more products will be displayed in this recommendation system to support accurate product recommendation results.

## Acknowledgment

Thank you to Mrs. Pipin Widyaningsih, M.Kom and Mrs. Sri Sumarlinda, M.Kom, who have guided me during the research and writing of this journal. I would also like to thank Hilda Pakan Supplier Store for giving permission for this research.

## References

[1] S. I. Mahmud, "Policy on Management of Livestock Curbing in Gadung Subdistrict," *Journal of Digital Economics and Business (MINISTAL),* vol. 1, no. 2, pp. 71–82, 2022.

[2] L. Aldino Ismail and B. Tjahjono, "Internet of Things-Based Animal Feeding," *ikraith-informatics*, vol. 7, no. 2, pp. 49–57, 2022, doi: 10.37817/ikraith-informatika.v7i2.2250.

[3] G. Syahrani, S. Sevira, A. Yunizar, and P. Yusuf, "Design of Jabodetabek Coffee Shop Recommendation Chatbot by Using Dialogflow Natural Language Processing," *SKANIKA: Computer Systems and Informatics Engineering*, vol. 7, no. 1, pp. 74–84, 2024.

[4] M. Fajriansyah, P. P. Adikara, and A. W. Widodo, "Movie Recommendation System Using Content Based Filtering," *Journal of Information Technology and Computer Science Development*, vol. 5, no. 6, pp. 2188–2199, 2021, [Online]. Available: http://j-ptiik.ub.ac.id

[5] W. Hidayah, "Nutritional Capacity for Early Childhood Growth and Development Levels," *Al Jayyid: Journal of Early Childhood Education*, vol. 1, no. 1, pp. 68–78, 2023.

[6] C. L. Polak, N. S. H. Malonda, and M. D. Amisi, "An Overview Of The Adequacy Of Water-Soluble Vitamins In Semester VI Students At The Faculty Of Public Health, University Of Sam Ratulangi Manado During The Covid-19 Pandemic," *KESMAS Journal*, vol. 10, no. 2, pp. 26–34, 2021.

[7] W. Lisyanto Prabowo, "Theory on Drug Prescribing Knowledge," *Journal of Medika Hutama*, vol. 2, no. 4, pp. 1036–1039, 2021, [Online]. Available: http://jurnalmedikahutama.com

[8] A. Apriani, H. Zakiyudin, and K. Marzuki, "Application of Cosine Similarity Algorithm and TF-IDF Weighting System for New Student Admission at Private Campuses," *Journal of Bumigora Information Technology (BITe),* vol. 3, no. 1, pp. 19–27, Jul. 2021, doi: 10.30812/bite.v3i1.1110.

[9] R. Ramadhan, Y. A. Sari, and P. P. Adikara, "Comparison of Term Frequency-Inverse Document Frequency and Term Frequency-Relevance Frequency Weighting of N-Gram Features in Sentiment Analysis," *Journal of Information Technology and Computer Science Development*, vol. 5, no. 11, pp. 5075–5079, 2021, [Online]. Available: http://j-ptiik.ub.ac.id

[10] D. Septiani and I. Isabela, "Term Frequency Inverse Document Frequency (TF-IDF) Analysis in Information Retrieval on Text Documents," *SINTESIA: Journal of Indonesian Information Systems and Technology*, vol. 1, no. 2, pp. 81–88, 2022.

[11] L. Kurnia Ramadhani and B. Nurul Widyaningrum, "Analysis Sentiment to Religious Based Issues Netizen Comments on President Jokowi 's Instagram using the Naive Bayes Classifier Method," *Jurnal Inotera*, vol. 9, no. 1, pp. 59–66, Feb. 2024, doi: 10.31572/inotera.vol9.iss1.2024.id296.