Available *online* at : http://bit.ly/InfoTekJar

InfoTekJar : Jurnal Nasional Informatika dan Teknologi Jaringan

ISSN (Print) 2540-7597 | ISSN (Online) 2540-7600

Data Mining

Combination of TOPSIS Method with Attribute Weighting of Information Gain in Decision-Making

Ahmad Rozy, Erna Budhiarti Nababan, Syahril Efendi

Faculty of Computer Science and Information Technology, Universitas Sumatera Utara, Medan, Indonesia

KEYWORDS

TOPSIS, Information Gain

CORRESPONDENCE

Phone: +6282366482029

E-mail: pescob12@gmail.com

A B S T R A C T

In this research, a combination of the Technique for Order Preference by Similarity to Ideal Solution (TOPSIS) lalgorithm was carried out with the attribute weighting of the Information Gain method to obtain better decision support results. The data processed in this study is the Indian Liver Patient Dataset (ILPD) dataset obtained lfrom UCI Machine Learning Repository which has 583 instances, 11 attributes and 1 class label. The class label is a text type that consists of two values, namely a liver patient and a non-liver patient. The experimental results show that TOPSIS' running time and information gain combination algorithm is 1.13 seconds. The result of the accuracy value obtained with a final threshold value greater than 0.5 is 91.25%.

INTRODUCTION

Technique for Order Preference by Similarity to Ideal Solution (TOPSIS) is a method that can help the decision-making process optimally to solve decision problems practically. This is because the concept is simple and easy to understand, the computation is efficient, and can measure the relative performance of decision alternatives in simple mathematical formula. Also, this method can provide a solution of some possible alternatives, by comparing each alternative with the best alternative and the worst alternative that exists. [1].

In some cases in the Decision Support System with the TOPSIS method, the process of weighting attribute values is often obtained based on the assumptions of the decision maker or by asking an expert or expert in determining the weight value of the criteria or attributes, so that it is considered less precise and accurate. [2].

Research by [3] those who examine the application of the TOPSIS method on the recommendation of telecommunication towers with the results obtained, namely the TOPSIS method can perform calculations in the recommendation of selecting telecommunication towers with various attributes. However, the authors observed that in this study the attribute weights were still determined manually.

Then, research by [4] examining the selection of outstanding students at Pematang Siantar High School by applying the TOPSIS method. The results of these studies indicate that the TOPSIS method can determine outstanding students with the specified criteria.

Then, research by [5] examining the feasibility of giving credit using the TOPSIS method in CS Finance leasing companies. The results of this study indicate that applying the TOPSIS method in this case is not an absolute decision because the final decision is still determined by the decision-making management. However, the system is used to assist in considering credit granting decisions.

Then, Information Gain is a feature selection method by ranking attributes. This can help reduce noise from irrelevant features [6]. Then, the process of weighting the attribute values in this study is no longer done manually so that a more objective alternative can be produced.

https://doi.org/10.30743/infotekjar.v6i1.4017

METHOD

The stages in the proposed methodology are the data information used, the attribute weighting process, the calculation using TOPSIS and the calculation of the accuracy of TOPSIS preferences.

Dataset Used

The data used in this research is the Indian Liver Patient Dataset (ILPD) dataset obtained from the UCI's Machine Learning Repository with the Comma Separated Value (csv) type and is a ranking-oriented dataset in determining the best alternative from a set of data, so that it can be tested on the method. decision-making. This dataset consists of 583 instances and 11 attributes with 1 class label which is a text type consisting of two values, namely liver patients and non-liver sufferers. The liver dataset serves as alternative information for ranking and calculating validation with the TOPSIS algorithm and weighting the Information Gain attribute.

Attribute Weighting with Information Gain

The attribute weighting used in this study is Information Gain which can reduce bias in the attribute weighting process [7]. The steps in determining Information Gain [8]:

Step 1: Calculating the entropy value from all tested data and to calculate the entropy value with equation (1):

$$Entropy\ (S) = \sum_{i=1}^{n} -\ pi * \log_2 pi \tag{1}$$

Step 2: Calculating Information Gain to select attribute testing and the highest Information Gain value will be selected as an attribute test of a node with equation (2):

$$Information\ Gain\ (S,A) = Entropy\ (S) - \sum_{i=1}^{n} \frac{|S_i|}{|S|} \times Entropy\ (S_i) \tag{2}$$

Step 3: Calculating the value of the Gain Ratio weight using the min-max normalization equation [9] with the lowest weight is 0.1 and the highest weight is 1 with equation (3):

$$W_i = \frac{(G_i - Min\ (G))}{Max\ (G) - Min\ (G)} \times (0.9) + 0.1 \tag{3}$$

Nomenclature for your equations must be presented after the References section of your article.

Calculation Technique for Order Preference by Similarity to Ideal Solution (TOPSIS) Method

TOPSIS is a method that can help the decision-making process optimally, is easy to understand, the computation is efficient, and can provide a solution of some alternatives by comparing each of the best and worst lalternatives. The steps to solve problems with TOPSIS [10] :

Step 1: Input Dataset to process data used in decision making.

Step 2: Input the attribute weight value and the importance category of the attribute (Benefit / Cost).

Step 3: Calculating the normalized decision matrix with equation (4):

$$r_{ii} = \frac{x_{ii}}{\sum mi = 1xii2} \tag{4}$$

Step 4: Calculating a weighted normalized decision matrix with equation (5):

$$y_{ii} = w_i r_{ii} \tag{5}$$

Step 5: Determine a matrix of positive ideal solutions and negative ideal solutions with formula:

The ideal positive solution is denoted A +, with the formula:

$$A^+ = (y1^+, y2^+, \dots, y^+) \tag{6}$$

The ideal negative solution is denoted A +, with the formula:

$$A^- = (y1^-, y2^-, \dots, yn^-) \tag{7}$$

Step 6: Calculating Alternative Distances

The ideal positive solution, with the formula:

$$D_i^+ = \sqrt{\sum_{j=1}^{n} (y_j^+ - y_{ij})^2} \tag{8}$$

The ideal negative solution, with the formula:

$$D_i^- = \sqrt{\sum_{j=1}^{n} (y_{ij} - y_j^-)^2} \tag{9}$$

Step 7: Calculating the Preference Value of Each Alternative

$$V_i = \frac{D_i^-}{D_i^- + D_i^+} \tag{10}$$

Step 8: Alternative Ranking Acquisition

The alternatives are sorted from the largest Vi value to the smallest Vi value. The alternative with the largest Vi value is the best solution.

RESULTS AND DISCUSSION

This section lwill ldiscuss lthe lresults lof the TOPSIS test and the Information Gain and the accuracy of each weighting. In testing this research was assisted by Jupyter Notebook and Rapid Miner Studio.

Attribute Weighting Information Gain

For Information Gain weights are obtained based on calculations in equations (1) to (3) which in the calculation process are assisted by Rapid Miner Studio. The Information Gain weight value of each attribute obtained is shown in Table 1.

 https://doi.org/10.30743/infotekjar.v6i1.4017

Table 1. Information Gain Weight Value

No.	Attribute	Weight Value
1.	Age (C1)	0.201
2.	Gender (C2)	0
3.	TB (C3)	0.981
4.	DB (C4)	1
5.	Alkphos (C5)	0.749
6.	SGPT (C6)	0.680
7.	SGOT (C7)	0.769
8.	TP (C8)	0.003
9.	ALB (C9)	0.193
10.	A/G (C10)	0.292

TOPSIS Result

The results of the TOPSIS preference with the attribute weighting test obtain preference values based on calculations based on equations (4) to (10) and with the results of the multiplication of the Information Gain attribute weights that have been obtained and the results of the preferences are shown in Table 2:

Table 2. Best Alternative Ranking Results

Rank	Alternative	Sex	Preference (V)	Real	Valid (V > 0.5)	Running Time
1.	483	Female	0.500830276	Liver	Liver	
2.	233	Female	0.500745067	Liver	Liver	
3.	539	Female	0.500725776	Liver	Liver	
4.	403	Male	0.5007159	Liver	Liver	
5.	360	Male	0.500695784	Non-Liver	Liver	
6.	402	Female	0.500695488	Liver	Liver	
7.	95	Female	0.500670599	Liver	Liver	1.13 Second
8.	566	Female	0.500663833	Liver	Liver	
9.	313	Female	0.500653069	Non-Liver	Liver	
10.	573	Female	0.500639526	Liver	Liver	
⋮	⋮	⋮	⋮	⋮	⋮	
⋮	⋮	⋮	⋮	⋮	⋮	
583.	270	Female	0.498565362	Liver	Non-Liver	

Value of Method Performance Accuracy

Then in this section displays the results of the accuracy of the TOPSIS test with Information Gain obtained from each of the resulting preference values and the results of the accuracy are as follows.

Dataset : 583
Valid : 532
Error : 51
Accuracy of TOPSIS + Information Gain : (532/583) * 100 = 91.25 %

CONCLUSIONS

[1] Maisari, K. D.,Andreswari, D., Efendi, R. 2017. Implementasi Metode TOPSIS dengan Pembobotan Entropy Untuk Penentuan Calon Penerima Bantuan Siswa Miskin (BSM) APBD Kota Bengkulu (Studi Kasus: SMAN 8 Kota Bengkulu). *Jurnal Rekursif,* ***5***(2), pp. 179-194.

[2] Saputra, M. 2018. Analisis Penentuan Nilai Bobot pada Metode Simple Additive Weighting (SAW) dengan Menggunakan Metode Analytical Hierarchy Process (AHP). *Tesis Magister*. Universitas Sumatera Utara.

[3] Saputra, R. A., Ramdhani, L. S., & Firmansyah, M. 2019. Aplikasi Sistem Pendukung Keputusan Analisis Kelayakan Pemberian Kredit Menggunakan Metode TOPSIS Pada Perusahaan Leasing CS Finance. *CESS (Journal of Computer Engineering System and Science)*, ***4***(2), pp. 137-143.

[4] Situmorang, H., Damanik, B., Sibagariang, S., & Manurung, I. H. G. 2019. Aplikasi Sistem Pendukung Keputusan Analisis Kelayakan Pemberian Kredit Menggunakan Metode TOPSIS Pada Perusahaan Leasing CS Finance. *CESS (Journal of Computer Engineering System and Science)*, ***4***(2), pp. 291-296.

[5] Siregar, V. M. M., Sonang, S., Purba, A. T., Sugara, H., & Siagian, F. 2021. Implementation of TOPSIS Algorithm for Selection of Prominent Student Class. *Annual Conference on Science and Technology Research (ACOSTER) 2020. Journal of Physics: Conference Series*. vol. 1783, pp. 1-6.

[6] Waruwu , Y. F., Zarlis , M., Nababan, E. B., & Ziliwu, M. S. 2018. Seleksi Atribut Pada Algoritma Radial Basis Function Neural Network Menggunakan Information Gain," *Seminar Nasional Royal (SENAR) 2018*, vol. 9986, pp. 21–24.

[7] Priyadarsini, R. P., Valarmathia, M. L., & Sivakumari, S. 2011. Gain Ratio Based Feature Selection Method for Privacy Preservation. *ICTACT Journal on Soft Computing*, ***1***(4), pp. 201–205.

[8] Utama, H. S., Rosiyadi, D., Aridarma, D., Prakoso, B. S. 2019. Sentimen Analisis Kebijakan Ganjil Genap di Tol Bekasi Menggunakan Algoritma Naïve Bayes dengan Optimalisasi Information Gain. *Jurnal PILAR Nusa Mandiri*, *15*(2), pp. 247-254.

[9] Saranya, C., & Manikandan, G. 2013. A Study on Normalization Techniques for Privacy Preserving Data Mining. *International Journal of Engineering and Technology (IJET)*, **5**(3), pp. 2701-2704.

[10] Ramdania, D. R., Manaf, K., Junaedi, F. R., Fathonih, A., & Hadiana, A. 2020. TOPSIS Method on Selection of New Employees Acceptance. *2020 6th International Conference on Wireless and Telematics* (*ICWT*).

AUTHOR(S) BIOGRAPHY

Ahmad Rozy

He has worked as an ICT staff at Dinas Komunikasi dan Informasi Kota Medan. He graduated his master postgraduate in Computer Engineering from Universitas Sumatera Utara in 2021.

 https://doi.org/10.30743/infotekjar.v6i1.4017