

CLASSIFICATION C4.5 METHOD IN SELECTION OF PASKIBRA MEMBERS

Rika Nofitri

Sistem Komputer, Sekolah Tinggi Manajemen Informatika dan Komputer Royal

email: nofitririka15@gmail.com

Abstract: Paskibra selection is an annual agenda held by schools, district disporas, provincial disporas, and the Ministry of Youth and Youth. This activity aim select the best students who will serve as flag raisers at independence commemoration ceremonies and other holidays. Paskibra participants have to choose. This reliable, formidable and attractive students are choosen. So the flag raising can be carried out solemnly and well. This school has a large number of students. Paskibra extracurricular enthusiasts also have many enthusiasts. This school Paskibra is a favourite extracurricular and quite prestigious among students. The selection of Paskibra has been running so well, but the decisions taken are sometimes not by the standards procedures. Therefore, a system that will be a reference for objective decision-making need to created . The decision-making system using the data mining algorithm C4.5. This study aim to create a system that makes it easier for the selection committee to select eligible Paskibra members. The method used is a qualitative method in the form of the results of interviews with data from participants in the last three years of the Paskibra selection. As a result, there are three main criteria and five rules of the selection system.

Keywords: C4.5 algorithm; data mining; paskibra

Abstrak: Seleksi paskibra merupakan agenda tahunan yang diadakan oleh sekolah, dispora kabupaten, dispora provinsi, dan kemenpora untuk memilih siswa-siswi terbaik yang akan bertugas sebagai pengibar bendera pada upacara peringatan kemerdekaan dan hari-hari besar lainnya. Peserta paskibra perlu diseleksi agar terpilih siswa-siswi yang handal, tangguh dan menarik sehingga pengiabaran bendera dapat dilaksanakan dengan khidmad dan baik. Sekolah ini adalah salah satu sekolah menengah atas yang menjadi favorit para lulusan SMP. Oleh karena itu sekolah ini mempunyai siswa yang sangat banyak. Dengan banyaknya siswa yang masuk, maka peminat ekskul paskibra juga memiliki peminat yang banyak, selain itu di sekolah ini paskibra merupakan ekskul favorit dan cukup bergengsi di kalangan siswa. Seleksi paskibra yang berjalan selama ini sudah dilakukan dengan baik, akan tetapi keputusan yang diambil terkadang tidak sesuai dengan standar yang telah ditetapkan. Oleh karena itu perlu dibuatkan sebuah sistem yang akan menjadi acuan untuk pengambilan keputusan yang objektif. Sistem pengambilan keputusan akan dibuat menggunakan data mining algoritma C4.5. Tujuan dari penelitian ini adalah membuat sebuah sistem yang memudahkan panitia penyeleksi untuk memilih anggota paskibra yang layak. Metode yang digunakan adalah metode kualitatif berupa hasil wawancara data peserta seleksi paskibra tiga tahun terakhir. Hasilnya berupa tiga kriteria utama dan lima aturan sistem seleksi.

Kata kunci: Algoritma C4.; data mining; paskibra;

INTRODUCTION

One of the good impacts of the presence of information technology in the community is its ability to take over human work. Technology can make it easier for humans to do anything with the help of technology because almost all aspects of life already use technology, especially information technology. In other words, information and communication technology can facilitate and enhance the quality of human life.

Decision making implements information technology. Decision-making that is previously still done manually and tended to use feelings rather than calculating, can now be taken over by computers, especially information technology. Many methods can be used in decision-making, one of which is using C4.5 classification data mining. The C4.5 algorithm generates a decision tree form. Decision trees are used in the method of classification and prediction[1]. Decision trees utilized to expose data and find hidden relationships between several prospective input variables and a target variable.

Every year the school, Dispora Kabupaten, Dispora Provinsi dan Kemenpora hold a selection paskibra. These agenda aim to select the best students who will serve as flag raisers at in-dependence commemoration ceremonies and other holidays[2]. Paskibra participants need to be chosen so that reliable, tough and attractive students are selected so that the flag raising can be carried out solemnly and well. In the beginning of each new school year, the school will begin to select students who will become members of paskibra, including SMAN 1 Kisaran. The selection of paskibra that

has been running so far has been done well, but the decisions taken are sometimes not in accordance with the standards that have been set. And it is necessary to create a system that will be a reference for objective decision making.

The Data mining methods are utilized to decide covered data that is valuable to healthcare specialists with effective analytic decision making [2].

The meaning of Data Mining has the overall objective of extracting information (with intelligent methods) from data sets and converting information into understandable structures for further use[3]. The C4.5 algorithm can support new student admission decision-making through the rules generated. The testing process with RapidMiner resulted in an accuracy of 90.50% [4].

An application built using the K-Nearest Neighbor (KNN) and Simple Additive Weighting (SAW) methods can recommend the names of participants who pass and do not pass the selection based on the results of ranking the scores of each participant [5]. The K-Nearest Neighbor method was used to classify chosen participants. The Simple Additive Weighting method was used to perform ranking. This application built using PHP programming language.

Determining the class of prospective students at web-based English course institutions also implements the C4.5 decision tree algorithm. The results of the classification of students as much as 45 data played by an active class are 8, and ultimate are 6 data. Then produced a basic are 15 data, prov are 5 data. [6]

Improving students' English skills also applies Algorithm C4.5. From this study it was found "hearing from the

environment" as an influential factor in improving students' English skills". [7]

The criteria assessed in the selection of paskibra members at SMAN 1 Range are physical strength, height, agility, discipline and marching ability. The study aims to conduct a decision support system for the selection of paskibra members using C4.5 classification data mining. The C4.5 classification data mining process data on student assessment criteria for prospective paskibra members is described, resulting in a final grade that can be used as a basis for decisions for the implementing committee.

METHOD

C4.5 Method

Many algorithms that can be used formation of decision trees include ID3, CART, and C4.5. Algo-rhythm C4.5 is a development of ID3. Because of this development, the C4.5 algorithm has the same basic work principles as ID3 algo-rhythms. It's just that in the C4.5 algorithm, the selection of attributes using the Gain Ratio. Resulting in the Decision Tree C4.5 algorithm has the good accuracy value [8] [9].

Decision Tree C4.5 C4.5 method is to change the tree generated in several rules. The number of rules is equal to the number of paths that might be built from the root to the leaf node [6]. In general, C4.5 algorithm to build a decision tree with the following general steps: a) Select the attribute as the root b) Create a branch for each value c) Divide cases in branches d) Repeat the process for each branch until all cases in the branch have the same class [10].

To select attributes as roots is based on the highest Gain score of the available

attributes [11]. To count the Gain, the formula as seen in Formula 1.

$$Entropy(Total) = \sum_{i=1}^n - p_i * \log_2 p_i \quad (1)$$

With: S is Set of cases; A is Attributes; N is Number of Partition Attribute A; |Si| is Number of cases in partition; |S| is Number of cases in S. While the calculation of Entropy Scores can be seen in Formula 2.

$$Gain(Total, SA) = Entropy(Total) - \sum_{i=1}^n \frac{|SA_i|}{|Total|} * Entropy(SA_i) \quad (2)$$

With: S is Set of cases; A is Features; N is Number of partition S; Pi is Proposal of Si against S.

Data Analysis

This study involved SMA Negeri 1 Kisaran as a system user to determine the graduation of students in paskibra selection with criteria including height, physical strength, agility, discipline and the ability to line up with each weight or value against the percentage of excellence possessed by these criteria. In this case, result of data mining algorithm C4.5 is implemented in the decision support system .

In this study, the data collected are a. Student data of selection participants; b. Data recapitulation of the assessment of selection participants from 2020. Some of the criteria needed in determining the passing selection of paskibra participants can be seen in table 1.

In this section, the research design is no longer contained but is focused on the result of the research that has been carried out.

The result of the study must be explained clearly and concisely. The result should summarize the findings

(scientific) rather than provide data in great detail.

The discussion must explore the significance of the research result. It is best to quote from previous research that can support the results of your research [12].

Table 1. Selections Criterias

No	Data Name	Category
1	Height	>165cm, <=165cm
2	Strength	Strong, weak
3	Agility	Swift, less swift
4	Discipline	Discipline, undisciplined
5	Marching capability	Good, Ordinary, Poor

The results of these filtering activities can be seen in table 2.

Table 2. Final Results

No	Decission
1	Qualify
2	Not Qualify

The next stage is making a decision tree to determine whether or not the selection participants pass the selection of paskibra members at SMA Negeri 1 Kisaran.

For more details, here are the data processing steps:

- First step is to find the criteria used as roots based on the highest gain value. The trick is to calculate the

number of cases for the decision to pass, the number of cases for the decision not to pass, and the entropy of all cases based on the criteria of height, physical strength, agility, discipline and marching ability. After that, the gain calculation for each attribute is carried out. The calculation results are in table 3.

The entropy value in table 4 is count as follow:

Entropy (Total)

$$= \left(-\frac{17}{35} * \log_2 \left(\frac{17}{35} \right) \right) + \left(-\frac{18}{35} * \log_2 \left(\frac{18}{35} \right) \right) = 0,999$$

For height criteria calculated as follows:

$$Entropy(ya) = \left(-\frac{17}{24} * \log_2 \left(\frac{17}{24} \right) \right) + \left(-\frac{7}{24} * \log_2 \left(\frac{7}{24} \right) \right) = 0,871$$

$$Entropy(tidak) = \left(-\frac{0}{11} * \log_2 \left(\frac{0}{11} \right) \right) + \left(-\frac{11}{11} * \log_2 \left(\frac{11}{11} \right) \right) = 0,9183$$

Table 3 Node 1 Calculation

node 1		jumlah	Qualify	Not Qualify	entropy	gain
Total		35	17	18	0,999	
	Height					
	Yes	24	17	7	0,871	0,4022
	No	11	0	11	0	
Strength	Weak	29	16	13	0,992	0,0658
	Strong	6	1	5	0,650	
Agility	Swift	22	13	9	0,976	0,0552
	Less Swift	13	4	9	0,890	
Discipline	Discipline,	20	11	9	0,993	0,0160
	Undisciplined	15	6	9	0,971	
Marching capability	Good	20	13	7	0,934	0,1508
	Ordinary	12	4	8	0,918	
	Poor	3	0	3	0	

The criteria strength, agility, discipline and marching are count as the height calculation.

After the entropy value is count, the next step is count the gain value using the following equation:

Gain value of Heigh :

$$\begin{aligned} \text{Gain}(\text{Total, tinggi}) &= 0.99 \\ &- \left(\left(\frac{24}{35} * 0.871 \right) + \left(\frac{11}{35} * 0 \right) \right) \\ &= 0.4022 \end{aligned}$$

The gain physical strength, agility, discipline and marching are count as the height calculation.

Based on the calculation of the first iteration described earlier, the obtained decision tree is in Image 1:

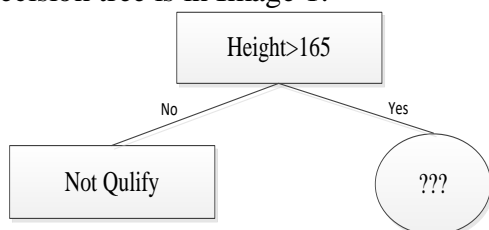


Image 1. Decision Tree of Node 1 Calculation Results (first iteration)

The calculation continues to the next iteration again and again, until there is no more criteria can be used as the next branch, and all attributes in the last branch, namely discipline, have classified cases into one. From the calculations carried out produce a decision tree as in Image 2.

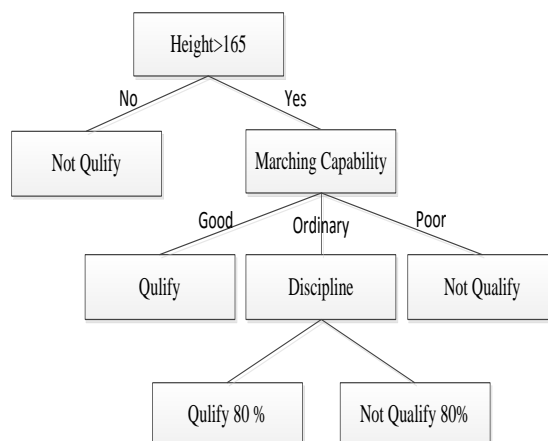


Image 2. Decision Tree of Node 3 Calculation Results (Last Iteration)

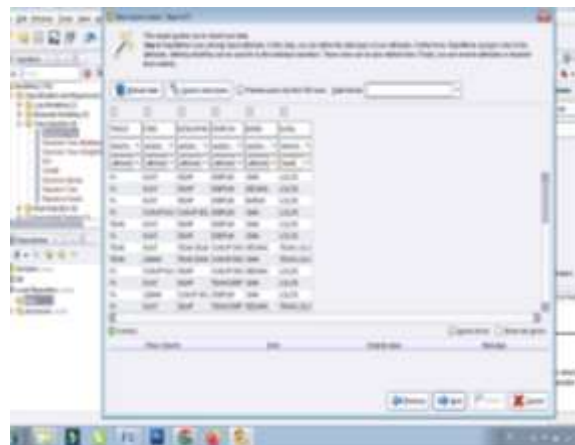


Image 3. Data Import Wizard

RESULT AND DISCUSSION

Then this system is imported into a tool that has been determined by researchers, namely using rapidminer software. The import data in question are as seen in Image 3.

In Image 3, the condition attributes of the selection of paskibra members, namely height, physical strength, agility, discipline and marching displayed. The decision attribute is a result marked by a label mark on the wizard form. After that, in the application of rapidminer software, the connection process will be carried out between the imported data and the decision tree method operator in the operator menu. Then drag the data and operators or move the process into the process window. So you can see the connection in Image 4.

From the stages are in Image 4, the application process is running by clicking the run button in the application window. So from the results of the process run, the results are decision trees that describe the relationship between criteria that cause the passing or not of the paskibra selection participants.

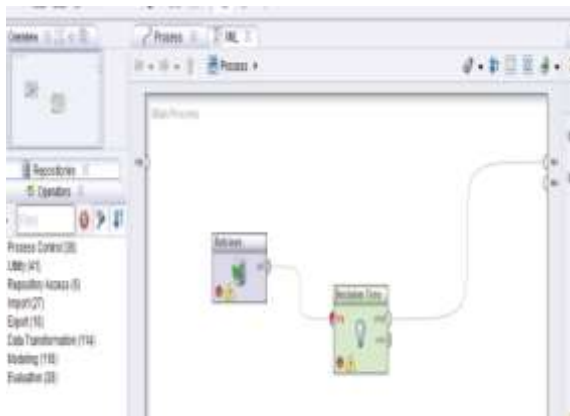


Image 4. Import Data Connection with Method C4.5

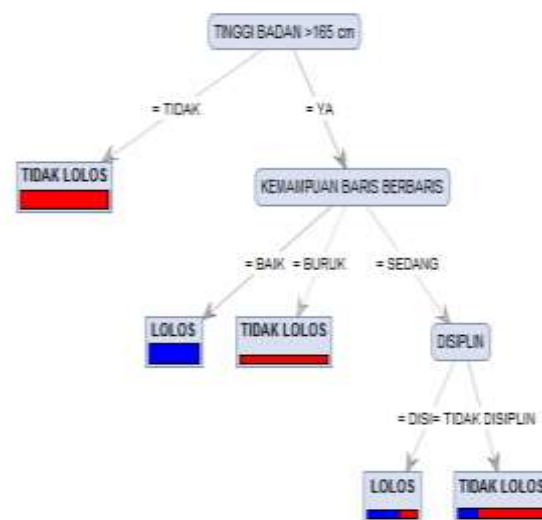


Image 5. Decision Tree Results with Method C4.5

From the picture above, we get some rules as IF height does not more than 165 cm THEN decision does not pass; IF height more than 165 cm AND good marching ability THEN decision passes; IF height more than 165 cm AND poor marching ability THEN decision does not pass; IF height more than 165 cm AND medium marching ability AND discipline THEN decision passes 80%; IF height more than 165 cm AND medium marching ability AND undisciplined THEN decision does not pass 80%.

CONCLUSION

From the analysis and discussion, the following conclusions can be drawn. The application of this method is very simple and very easy to understand because each stage helps in making decisions. The application of the decision support system built can assist the selection committee for paskibra members at SMA Negeri 1 Kisaran in determining the right formula for decision making in accepting paskibra members. The final result obtained from this decision support system will provide a decision tree to determine the success rate of the participants. The first criterion that becomes the root is height, followed by the ability to line up and discipline.

BIBLIOGRAPHY

- [1] M. I. A. Nasution, "SISTEM PAKAR UNTUK MENGANALISIS KERUSAKAN AIR CONDITIONER DENGAN MENGGUNAKAN METODE FORWARD CHAINING EXPERT SYSTEM TO ANALYZE AIR CONDITIONER DAMAGE USING THE," pp. 177–189.
- [2] H. Oliff and Y. Liu, "Towards Industry 4.0 Utilizing Data-Mining Techniques: A Case Study on Quality Improvement," *Procedia CIRP*, vol. 63, pp. 167–172, 2017, doi: 10.1016/j.procir.2017.03.311.
- [3] T. H. Sinaga, A. Wanto, I. Gunawan, S. Sumarno, and Z. M. Nasution, "Implementation of Data Mining Using C4.5 Algorithm on Customer Satisfaction in Tirta Lihou PDAM," *J. Comput. Networks, Archit. High Perform. Comput.*, vol. 3, no. 1, pp. 9–20, Jan. 2021, doi: 10.47709/CNAHPC.V3I1.923.
- [4] L. M. Fitriani and A. Setyono, "Penerapan Algoritma C4.5 untuk Seleksi Penerimaan Siswa Baru pada SD Islam Terpadu Permata Bunda Demak Implementation of Decision Tree Algorithm for Selection of New Student Admission on Permata Bunda Integrated Islamic Elementary School," *CCIT J.*, vol. 11, no. 2, pp. 158–170, 2018, doi: 10.33050/ccit.v11i2.584.
- [5] A. J. T, D. Yanosma, and K. Anggriani, "Implementasi Metode K-Nearest Neighbor (Knn) Dan Simple Additive Weighting (Saw) Dalam Pengambilan Keputusan Seleksi Penerimaan Anggota Paskibra," *Pseudocode*, vol. 3, no. 2, pp. 98–112, 2017, doi: 10.33369/pseudocode.3.2.98-112.
- [6] Uminingsih, I. Suraya, and I. Nugroho, "Implementasi Algoritma Decision Tree C4.5 Untuk Menentukan Kelas Calon Siswa Di Lembaga Kursus Bahasa Inggris Berbasis Web," *J. Teknol.*, vol. 12, no. 1, pp. 35–44, 2019.
- [7] S. W. Siahaan, K. D. R. Sianipar, P. P. P. A. N. . F. I. R.H Zer, and D. Hartama, "Penerapan Algoritma C4.5 Dalam Meningkatkan Kemampuan Bahasa Inggris Pada Mahasiswa," *Petir*, vol. 13, no. 2, pp. 229–239, 2020, doi: 10.33322/petir.v13i2.1029.
- [8] L. D. Yulianto, A. Triayudi, and I. D. Sholihati, "Implementation Educational Data Mining For Analysis of Student Performance Prediction with Comparison of K-

- Nearest Neighbor Data Mining Method and Decision Tree C4.5,” vol. 4, no. 1, pp. 441–451, 2020.
- [9] F. F. Harryanto and S. Hansun, “Penerapan Algoritma C4.5 untuk Memprediksi Penerimaan Calon Pegawai Baru di PT WISE,” *Maret*, vol. 3, no. 2, p. 95, 2017.
- [10] A. Waluyo, H. Jatnika, M. R. S. Permatasari, T. Tuslaela, I. Purnamasari, and A. P. Windarto, “Data Mining Optimization uses C4 . 5 Classification and Particle Swarm Optimization (PSO) in the location selection of Student Boardinghouses Data Mining Optimization uses C4 . 5 Classification and Particle Swarm Optimization (PSO) in the location sel”, doi: 10.1088/1757-899X/874/1/012024.
- [11] D. Hartama, A. P. Windarto, and A. Wanto, “The Application of Data Mining in Determining Patterns of Interest of High School Graduates The Application of Data Mining in Determining Patterns of Interest of High School Graduates,” 2019, doi: 10.1088/1742-6596/1339/1/012042.
- [12] D. Sutrisno, S. N. Gill, and S. Suseno, “The development of spatial decision support system tool for marine spatial planning,” *Int. J. Digit. Earth*, vol. 11, no. 9, pp. 863–879, 2018.