

Decision Support System Design for Informatics Student Final Projects Using C4.5 Algorithm

Rafika Sari ^{1,*}, Hasan Fatoni ¹, Khairunnisa Fadhillah Ramdhanian ¹

* Correspondence Author: e-mail: rafika.sari@dsn.ubharajaya.ac.id

¹ Informatics; Universitas Bhayangkara Jakarta Raya; Jl. Raya Perjuangan No. 81 Margamulya, Bekasi Utara, Bekasi, Indonesia; telp. (021) 88955882; e-mail: rafika.sari@dsn.ubharajaya.ac.id, hasanfatoni96@gmail.com, khairunnisa.fadhillah@dsn.ubharajaya.ac.id

Submitted : **10/02/2023**
Revised : **24/02/2023**
Accepted : **10/03/2023**
Published : **31/03/2023**

Abstract

Academic consultation activities between students and academic supervisors are necessary to help students carry out academic activities. Based on the transcript of grades obtained, many students do not choose the appropriate final project/thesis specialization fields based on their academic abilities, resulting in a lot of inconsistencies between the course grades and the final project specialization fields. The purpose of this research is to minimize the subjectivity aspect of students in choosing their final project academic supervisors and minimize the inconsistencies between the course grades and the final project specialization fields. The method used in this research is classification data mining using the Decision Tree and C4.5 Algorithm methods, with the attributes involved being courses, course grades, and specialization courses. The C4.5 Decision Tree algorithm is used to transform data (tables) into a tree model and then convert the tree model into rules. The implementation of the C4.5 Decision Tree algorithm in the specialization field decision support system has been successfully carried out, with an accuracy rate of 70% from the total calculation data. The data used in this research is a sample data from several senior students in the Informatics program at Ubhara-Jaya. The results of the research decision support system can be used as a good recommendation for the Informatics program and senior students to direct their final project research. It is expected that further research will use more sample data so that the accuracy rate will be better and can be implemented in website or mobile-based applications.

Keywords: C4.5 algorithms, student final project, field courses, decision support system

1. Introduction

The learning process in lectures is supported by a curriculum that outlines the objectives, content, learning materials, and methods for implementing learning activities to achieve national education goals. The curriculum is a set of plans and arrangements for each semester, and the Operational Curriculum, created by the study program, determines the distribution and prerequisites of courses. By consulting the Operational Curriculum, students can determine which courses will be taken in a given semester, along with prerequisite courses that must be completed with certain grade requirements

(Sari et al., 2022). If students wish to take courses with specific prerequisites, they must have fulfilled those requirements in the previous semester and met certain score requirements (Sutrisno & Claudia, 2018).

This research proposes a tool for directing students to carry out research on linear final assignments, based on competencies acquired in specialization courses where they have performed well. Ultimately, this will help the study program achieve a predetermined graduate profile and enable students to apply their competencies in various fields after graduation.

Among several methods that can be used for classification is the decision tree method or Decision Tree. The decision tree method is a method that can turn very large facts into a decision tree that represents the rules. Rules can be easily understood in natural language. A decision tree is a structure that can be used to divide large data sets into smaller record sets by applying a set of decision rules. With each set of divisors, the members of the result set become similar to one another (Meng et al., 2020).

The data in the decision tree is usually expressed in the form of a table with attributes and records. Attribute states a parameter that is created as a criterion in the formation of the tree. The process in the decision tree is to change the shape of the data (table) into a tree model, change the tree model into a rule, and simplify the rule. There are many algorithms that can be used in the formation of a decision tree, including ID3, CART, and C4.5. The C4.5 algorithm is the development of the ID algorithm (Kretschmann et al., 2001; Marlina & Siahaan, 2016).

The C4.5 algorithm is the algorithm used to generate a decision tree. The basic idea of this algorithm is making a decision tree based on the selection of the attribute that has the highest priority or can be called the highest gain value based on the entropy value of the attribute as the axis of the classification attribute (Dai & Ji, 2014) this stage the C4.5 algorithm has 2 working principles, namely: Making a decision tree, and making rules (rule model). The rules formed from the decision tree will form a condition in the form of "if then" (Sudrajat, 2022).

There are several advantages of the C4.5 classification algorithm, including the results of the analysis in the form of a decision tree that is easy to understand, requires less data, is able to process nominal and continuous data, uses statistical techniques so that it can be validated, computation time is faster, and the resulting accuracy can match other classification techniques (Kretschmann et al., 2001).

2. Research Method

2.1. Implementation of C4.5 Algorithm

The process of applying the C4.5 algorithm to build a decision tree includes: selecting the root attribute, creating a branch for each value, dividing cases into branches, and repeating the process for each branch until all cases in the branch have the same class (Lee, 2019). Calculation of attribute values in the C4.5 algorithm with the following stages:

- a. Calculate the value of entropy.
- b. Calculate the gain ratio value for each attribute.
- c. The attribute with the highest gain is selected as the root and the attribute with the lower gain ratio is used as a branch.
- d. Calculate the value of the gain ratio of each attribute except the root.
- e. The highest gain ratio attribute is selected as a branch.
- f. Repeat the previous step until the resulting gain = 0.
- g. To calculate the entropy value can be calculated by equation (1) (Cherfi et al., 2018)

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

and (2)

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

where **S** is the case set, **A** is the attribute, **n** is the number of case partitions and **|S_i|** is the number of cases on the *i*th partition and **|S|** the number of cases **S**. The equation to calculate split info uses equation (3).

$$Split Info (S, A) = - \sum_{i=1}^n \frac{S_i}{S} * \log_2 \frac{S_i}{S} \quad (3)$$

where **S** is the case set, **A** is the attribute, **S_i** is the number of samples for the attribute *i*. Equation (4) is used to calculate the gain ratio.

$$Gain Ratio (S, A) = \frac{Gain(S, A)}{Split Info (S, A)} \quad (4)$$

where **S** is case set, **A** is attribute, **Gain(S,A)** is gain info on attribute **A** and **Split-Info (S,A)** is split info on attribute **A**.

2.2. Data

The data used in this study was obtained from final year students in the Informatics study program at Bhayangkara Jakarta Raya University (Ubhara-Jaya). The study program offers three fields of concentration for completing the final project (thesis).

However, many students do not choose fields that align with their academic abilities, which can result in difficulties during the learning and thesis-writing process (Adhatrao et al., 2013). Fields of concentration should facilitate students in completing their studies by allowing them to focus on a specific concentration within their field. To address issues related to academic consulting activities, a system was created to provide recommendations for student interests using data mining classification methods, specifically the Decision Tree C4.5 algorithm. The research data was sourced from final grade data from 2016-2018 students in the Informatics Study Program at Ubhara-Jaya, sampled from three batches with 10 students each. Data from several courses in semesters 4-7 was processed and classified based on the chosen field of concentration. Table 1 shows classification of students' field of interest.

Table 1. Classification of students' field of interest

Field	Field Courses			
	4st Semester	5st Semester	6st Semester	7st Semester
Software Development	Web Programming	Mobile Programming	Framework-based Programming	Application Projects
Network and Infrastructure	Computer Networks II	Computer Network Security and Information Systems	Computer Forensics	Integrated Systems
Data Science	Statistics and Probability	Intelligent Systems	Data Mining	Deep Learning

Source: Research Result (2023)

Table 2 show the data for student transcripts for the 2016 – 2018 academic year of UBJ Informatics Study Program along with the areas of interest taken by students can be seen in the Table 2. The variables in Table 2 are student data with codes STD 1-30, code names of field courses including PW: Web Programming, PM: Mobile Programming, PBF: Framework-based Programming, KP: Application Projects (practical work), JK: Computer Networks, KJK: Computer Network Security and Information Systems, KF: Computer Forensics, ST: Integrated Systems, SDP: Statistics and Probability, SC: Intelligent Systems, DM: Data Mining, and DL: Deep Learning. The code for the field of the student's final project is SD: Software Development, JI: Network and Infrastructure, and DS: Data Science. In the testing process, data transformation is carried out by classifying value attributes into three variables based on the value weights

which can be seen in Table 3. With Value weight category are: (1) 3.50 – 4.00 is High; (2) 2.75 – 3.50 is Medium; and (3) < 2.75 is Low.

Table 2. Classification of Student's Score and Field of Interest

STD	Transkrip Student Grades											Final Project Field	
	PW	PM	PBF	KP	JK	KJK	KF	ST	SDP	SC	DM	DL	
STD 1	3.33	3.00	3.00	3.00	3.67	3.33	3.00	2.67	4.00	3.00	3.00	3.00	SD
STD 2	3.67	3.00	3.00	4.00	3.67	3.33	3.67	3.67	3.67	3.33	3.00	3.33	SD
STD 3	0.00	2.67	2.67	4.00	3.70	2.33	3.33	0.00	3.00	3.67	3.33	3.67	DS
STD 4	3.67	4.00	4.00	4.00	4.00	3.67	3.33	3.33	3.00	4.00	3.33	4.00	JI
STD 5	4.00	4.00	4.00	4.00	4.00	3.33	3.67	3.00	3.00	3.67	3.67	3.67	SD
STD 6	0.00	1.00	1.00	0.00	3.70	3.70	0.00	2.33	0.00	3.67	0.00	3.67	JI
STD 7	1.00	3.00	1.00	4.00	4.00	0.00	3.67	0.00	3.00	1.00	0.00	0.00	JI
STD 8	2.67	3.00	3.00	4.00	3.70	3.33	3.33	2.33	3.00	3.67	2.67	3.67	JI
4STD 9	0.00	2.67	2.67	4.00	3.70	2.00	3.33	3.33	3.00	0.00	2.67	0.00	JI
STD 10	2.00	2.67	2.67	4.00	2.00	3.00	0.00	2.67	3.00	3.67	0.00	3.67	DS
STD 11	3.33	3.33	2.67	4.00	3.33	3.00	3.33	4.00	3.70	2.33	4.00	2.33	SD
STD 12	3.00	3.33	2.33	4.00	3.67	2.67	3.67	2.67	4.00	3.00	1.00	4.00	SD
STD 13	2.67	2.67	4.00	4.00	3.33	3.00	4.00	3.00	3.70	3.67	2.33	3.67	JI
STD 14	2.67	3.33	4.00	4.00	3.33	3.00	4.00	3.00	3.00	3.33	3.33	4.00	SD
STD 15	3.67	3.33	4.00	4.00	2.00	3.00	3.67	3.67	3.70	3.33	2.00	3.33	SD
STD 16	2.67	2.67	4.00	4.00	0.00	3.00	4.00	3.00	3.70	3.00	2.67	2.67	SD
STD 17	3.33	3.33	4.00	4.00	3.00	3.00	4.00	3.00	3.00	3.33	3.00	0.00	SD
STD 18	4.00	4.00	4.00	4.00	3.67	2.67	4.00	4.00	3.70	2.33	4.00	4.00	SD
STD 19	3.33	3.33	4.00	3.33	4.00	3.00	4.00	3.33	3.00	3.33	3.33	4.00	DS
STD 20	3.33	3.33	3.67	4.00	4.00	3.33	4.00	3.33	4.00	3.67	3.33	3.67	JI
STD 21	3.33	3.67	3.33	4.00	4.00	4.00	4.00	4.00	3.00	3.67	4.00	3.67	DS
STD 22	3.67	4.00	3.00	3.67	4.00	4.00	3.33	3.00	1.00	3.67	4.00	3.33	SD
STD 23	3.33	3.67	3.33	4.00	4.00	3.33	4.00	3.33	3.00	3.33	4.00	3.00	JI
STD 24	3.33	4.00	4.00	4.00	4.00	4.00	4.00	4.00	3.67	4.00	4.00	2.67	DS
STD 25	3.33	2.67	3.33	4.00	4.00	3.67	3.33	3.33	3.00	3.33	4.00	4.00	DS
STD 26	3.67	3.67	3.67	4.00	4.00	4.00	3.33	3.33	3.00	3.33	3.67	3.67	SD
STD 27	3.33	3.00	3.67	4.00	4.00	4.00	3.00	4.00	3.00	4.00	4.00	4.00	SD
STD 28	3.00	3.33	3.67	4.00	4.00	3.33	3.00	4.00	3.00	4.00	3.00	4.00	SD
STD 29	3.67	3.67	3.67	4.00	4.00	3.00	3.33	3.67	3.00	3.67	4.00	4.00	SD
STD 30	3.33	3.67	3.67	4.00	4.00	3.33	3.33	2.67	3.00	2.67	3.67	3.33	SD

Source: Research Result (2023)

3. Results and Analysis

The initial calculation process starts from inputting what attributes will be used in the study. Then calculate the entropy and gain values of each criterion and look for the highest gain value to be used as the root, which will then get the rule or result from the tree (Mazid et al., 2010). The data in table 2 will be transformed into 3 variables based on the weight values presented in Table 3. From table 3 it can be seen the number of

students taking courses based on the existing field fields. The total weighted values of "High", "Medium", and "Low" from each course can be seen in table 4 for calculating the number of cases. Table 5 is the cumulative value of the weighted scores for all cases of field courses. For SD there are 16 cases, for JI there are 18 cases and for DS there are 6 cases.

Table 3. Number of Case

Value	Courses												
	Weight	PW	PM	PBF	KP	JK	KJK	KF	ST	SDP	SC	DM	DL
High		8	10	16	27	23	8	15	9	10	14	12	18
Medium		13	13	7	2	4	17	13	13	18	11	9	6
Low		9	7	7	1	3	5	2	8	2	5	9	6
Total		30	30	30	30	30	30	30	30	30	30	30	30
Final Project Field								Number of Field					
Software Development (SD)								16					
Jaringan dan Insfrastruktur (JI)								8					
Data Science (DS)								6					

Source: Research Result (2023)

3.1. The Calculation Result

The data from the transformation results are then analyzed to produce a decision tree using the C4.5 Algorithm, namely by formulating the Entropy and Gain calculations.

3.1.1. Calculating Entropy (S)

Using equation (1) the entropy value will be calculated as follows:

$$Entropy(S) = \left(-\frac{SD}{Total}\right) * \log_2\left(\frac{SD}{Total}\right) + \left(-\frac{JI}{Total}\right) * \log_2\left(\frac{JI}{Total}\right) + \left(-\frac{DS}{Total}\right) * \log_2\left(\frac{DS}{Total}\right)$$

$$Entropy(Total) = \left(-\frac{16}{30}\right) * \log_2\left(\frac{16}{30}\right) + \left(-\frac{8}{30}\right) * \log_2\left(\frac{8}{30}\right) + \left(-\frac{6}{30}\right) * \log_2\left(\frac{6}{30}\right) = 1.456564763$$

3.1.2. Calculating Entropy

Entropy (S_i) calculations are carried out for all cases, in this case the number of field courses and each course will have 3 Entropy (S_i) values, each with high, medium and small entropy. So there will be as many as 36 values of Entropy (S_i). The following is the calculation of the entropy value for the Web Programming (PW) course, the same thing is also done to calculate the entropy (S_i) of other courses.

$$Entropy(S_1) = \left(-Pi\frac{SD}{Total}\right) * \log_2\left(Pi\frac{SD}{Total}\right) + \left(-Pi\frac{JI}{Total}\right) * \log_2\left(Pi\frac{JI}{Total}\right) + \left(-Pi\frac{DS}{Total}\right) * \log_2\left(Pi\frac{DS}{Total}\right)$$

$$Entropy(High) = \left(-\frac{7}{8}\right) * \log_2\left(\frac{7}{8}\right) + \left(-\frac{1}{8}\right) * \log_2\left(\frac{1}{8}\right) + \left(-\frac{0}{8}\right) * \log_2\left(\frac{0}{8}\right) = 0.543564$$

$$Entropy(S_2) = \left(-Pi\frac{SD}{Total}\right) * \log_2\left(Pi\frac{SD}{Total}\right) + \left(-Pi\frac{JI}{Total}\right) * \log_2\left(Pi\frac{JI}{Total}\right) + \left(-Pi\frac{DS}{Total}\right) * \log_2\left(Pi\frac{DS}{Total}\right)$$

$$Entropy(Medium) = \left(-\frac{7}{13}\right) * \log_2\left(\frac{7}{13}\right) + \left(-\frac{2}{13}\right) * \log_2\left(\frac{2}{13}\right) + \left(-\frac{4}{13}\right) * \log_2\left(\frac{4}{13}\right) = 1.419556$$

$$Entropy(S_3) = \left(-P_i \frac{SD}{Total}\right) * \log_2\left(P_i \frac{SD}{Total}\right) + \left(-P_i \frac{JI}{Total}\right) * \log_2\left(P_i \frac{JI}{Total}\right) + \left(-P_i \frac{DS}{Total}\right) * \log_2\left(P_i \frac{DS}{Total}\right)$$

$$Entropy(Low) = \left(-\frac{2}{9}\right) * \log_2\left(\frac{2}{9}\right) + \left(-\frac{5}{9}\right) * \log_2\left(\frac{5}{9}\right) + \left(-\frac{2}{9}\right) * \log_2\left(\frac{2}{9}\right) = 1.435520503$$

3.1.3. Calculating Gain

Gain (S,A) calculation is carried out for all cases, in this case the number of field courses. So there will be 12 Gain (S,A) values. The following is the calculation of the Gain value for the Web Programming (PW) course, the same thing is also done to calculate the Gain (S,A) for other courses.

$$Gain(PW) = Entropy(Total) - \frac{|Jumlah\ Kasus\ (Tinggi)|}{|Total|} * Entropy(Tinggi) + \frac{|Jumlah\ Kasus\ (Sedang)|}{|Total|} * Entropy(Sedang) + \frac{|Jumlah\ Kasus\ (Kecil)|}{|Total|} * Entropy(Kecil)$$

$$= 1.456564763 - \left[\left(\frac{8}{30} * 0.543564443\right) + \left(\frac{13}{30} * 1.419556299\right) + \left(\frac{9}{30} * 1.435520503\right)\right] = 0.265817032$$

Table 4 is the result of calculating entropy and gain at node 1 using the C4.5 decision tree algorithm. The highest gain value will be obtained from the tree root (Mardi, 2017). The stages of the calculation process are presented in table 4 are: (1) Calculating Entropy and Gain; (2) Selection of the highest Gain as the root (Node), in the calculation the highest data obtained is in the case of the Web Programming (PW) course; (3) Repeat the process of calculating Entropy and Gain to find branches until all cases in the branch have the same class, ie when all variables have become part of the decision tree or each variable has a leaf or decision; and (4) Create Rules based on decision trees (Muslim et al., 2018)

Table 4. Entropy and Gain Calculation

Node 1							
Attribute	Weight	Number of Cases (S)	SD (S ₁)	JI (S ₂)	DS (S ₃)	Entropy	Gain
Total		30	16	8	6	1.456564763	
Pemrograman	High	8	7	1	0	0.543564443	0.265817032
Web (PW)	Medium	13	7	2	4	1.419556299	
	Low	9	2	5	2	1.435520503	
Pemrograman	High	10	6	2	2	1.370950594	0.167474123
Mobile (PM)	Medium	13	9	3	1	1.140115679	
	Low	7	1	3	3	1.448815638	
Framework-based	High	16	11	3	2	1.199460293	0.090412811
Programming	Medium	7	3	2	2	1.556656709	
(PBF)	Low	7	2	3	2	1.556656709	
Application Project	High	27	15	7	5	1.426573285	0.10598214
(KP)	Medium	2	1	0	1	1	
	Low	1	0	1	0	0	
Computer	High	23	11	7	5	1.509871837	0.098996355
Networks (JK)	Medium	4	3	1	0	0.811278124	

Node 1							
Attribute	Weight	Number of Cases (S)	SD (S ₁)	JI (S ₂)	DS (S ₃)	Entropy	Gain
Total		30	16	8	6	1.456564763	
	Low	3	2	0	1	0.918295834	
Computer Network	High	8	3	2	3	1.561278124	0.072131895
Security (KJK)	Medium	17	11	4	2	1.260771796	
	Low	5	2	2	1	1.521928095	
Computer	High	15	8	4	3	1.456564763	0.083254753
Forensics (KF)	Medium	13	8	3	2	1.334679142	
	Low	2	0	1	1	1	
Integrated	High	9	7	0	2	0.764204505	0.178085882
Systems (ST)	Medium	13	6	5	2	1.460484683	
	Low	8	3	3	2	1.561278124	
Statistics and	High	10	7	2	1	1.15677965	0.07632591
Probability (SDP)	Medium	18	8	5	5	1.546631617	
	Low	2	1	1	0	1	
Intelligent Systems	High	14	5	5	4	1.577406283	0.156825137
(SC)	Medium	11	8	1	2	1.095795256	
	Low	5	3	2	0	0.970950594	
Data Mining (DM)	High	12	8	1	3	1.188721875	0.144926628
	Medium	9	5	2	2	1.435520503	
	Low	9	3	5	1	1.351644115	
Deep Learning	High	18	8	5	5	1.546631617	0.106751728
(DL)	Medium	6	5	1	0	0.650022422	
	Low	6	3	2	1	1.459147917	

Source: Research Result (2023)

3.2. Validation Test

Validation test is carried out by analyzing the results of calculations using the Confusion Matrix model (Zheng et al., 2021). Table 5 is the result of the calculation of the confusion matrix in the C4.5 algorithm, shown in Table 5.

Table 5. Confusion Matrix

Confusion Matrix	Real TRUE	Real FALSE
Prediction TRUE	(TP) 21	(FP) 9
Prediction FALSE	(TN) 0	(FN) 0

Source: Research Result (2023)

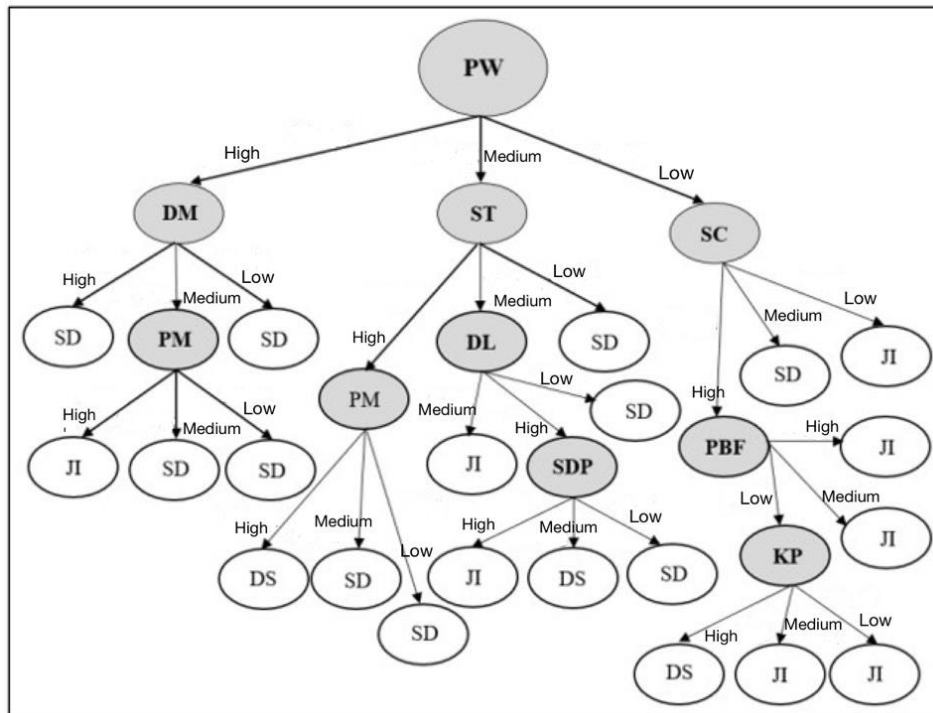
$$\text{Accuracy} = ((TP+TN)/(TP+TN+FP+FN))*100\% = ((21+0)/(21+0+9+0))*100\% = (21/30)*100\% = 70\%$$

$$\text{Precision} = (TP/(TP+FP))*100\% = (21/(21+9))*100\% = (21/30)*100\% = 70\%$$

$$\text{Recall} = (TP/(TP+FN))*100\% = (21/(21+0))*100\% = (21/21)*100\% = 100\%$$

From the above calculation, the accuracy, precision, and recall values are 70%, 70% precision, and 100% recall respectively. This shows an accurate value. So it can be concluded that the results of this study were successful in implementing the Decision

Tree method using the C4.5 Algorithm properly and it is hoped that it will facilitate study program managers and academic supervisors in helping the development of student studies and determining the field that will be taken based on the value data that has been obtained by students.



Source: Research Result (2023)

Figure 1. Decision Tree Results

From the decision tree that is formed in Figure 1, we get the rules (rule model) in determining recommendations for students' final project field. There are 8 rules formed.

- a. IF (PW='High') AND (DM='High') THEN Label = Software Development
- b. IF (PW='High') AND (DM='Medium') AND (PM='High') THEN Label = Jaringan dan Infrastruktur
- c. IF (PW='High') AND (DM='Medium') AND (PM='Medium') THEN Label = Software Development
- d. IF (PW='Medium') AND (ST='High') AND (PM='High') THEN Label = Data Science
- e. IF (PW='Medium') AND (ST='High') AND (PM='Medium') THEN Label = Software Development
- f. IF (PW='Medium') AND (ST='Medium') AND (DL='High') AND (SDP='High') THEN Label = Jaringan dan Infrastruktur
- g. IF (PW='Medium') AND (ST='Medium') AND (DL='High') AND (SDP='Medium') THEN Label = Data Science
- h. IF (PW='Medium') AND (ST='Medium') AND (DL='Medium') THEN Label = Jaringan dan Infrastruktur.

4. Conclusion

The Decision Tree method using the C4.5 Algorithm was successfully utilized to create rules for a decision support system in the field of student final projects. Training data based on variables such as courses, areas of field, and value of field courses produced an accuracy rate of 70% from the calculation results of data tests. The decision support system resulting from this research can serve as a recommendation for both the Informatics Study Program and the Faculty of Computer Science at Ubhara-Jaya, as well as for final-year students to guide their field of research in the final project. Further research is recommended to include a larger sample size, which could improve the accuracy level and enable implementation in website-based or mobile applications.

Acknowledgements

Our gratitude goes to the informatics study program, Faculty of Computer Science, Bhayangkara University, Jakarta Raya, which has facilitated this research to obtain the real data needed.

Author Contributions

Rafika proposed the topic; Hasan collecting the data, Rafika, Hasan, and Khairunnisa conceived models and designed the experiments; Rafika conceived the optimisation algorithms; Rafika, Hasan and Khairunnisa analysed the result.

Conflicts of Interest

The author declare no conflict of interest.

References

- Adhatrao, K., Gaykar, A., Dhawan, A., Jha, R., & Honrao, V. (2013). Predicting Students' Performance Using ID3 and C4.5 Classification Algorithms. *International Journal of Data Mining & Knowledge Management Process*, 3(5), 39–52. <https://doi.org/10.5121/ijdkp.2013.3504>
- Cherfi, A., Nouria, K., & Ferchichi, A. (2018). Very Fast C4.5 Decision Tree Algorithm. *Applied Artificial Intelligence*, 32(2), 119–137. <https://doi.org/10.1080/08839514.2018.1447479>
- Dai, W., & Ji, W. (2014). A Mapreduce Implementation of C4.5 Decision Tree Algorithm. *International Journal of Database Theory and Application*, 7(1),

- 49–60. <https://doi.org/10.14257/ijdta.2014.7.1.05>
- Kretschmann, E., Fleischmann, W., & Apweiler, R. (2001). Automatic Rule Generation for Protein Annotation with the C4.5 Data Mining Algorithm Applied on SWISS-PROT. *Bioinformatics*, 17(10), 920–926. <https://doi.org/10.1093/bioinformatics/17.10.920>
- Lee, J.-S. (2019). AUC4.5: AUC-Based C4.5 Decision Tree Algorithm for Imbalanced Data Classification. *IEEE Access*, 7, 106034–106042. <https://doi.org/10.1109/ACCESS.2019.2931865>
- Mardi, Y. (2017). Data Mining : Klasifikasi Menggunakan Algoritma C4.5. *Edik Informatika*, 2(2), 213–219. <https://doi.org/10.22202/ei.2016.v2i2.1465>
- Marlina, L., & Siahaan, P. U. (2016). Data Mining Classification Comparison (Naïve Bayes and C4.5 Algorithms). *International Journal of Engineering Trends and Technology*, 38(7).
- Mazid, M. M., Ali, A. B. M. S., & Tickle, K. S. (2010). Improved C4.5 Algorithm for Rule Based Classification. *Proceedings of the 9th WSEAS*.
- Meng, X., Zhang, P., Xu, Y., & Xie, H. (2020). Construction of Decision Tree Based on C4.5 Algorithm for Online Voltage Stability Assessment. *International Journal of Electrical Power & Energy Systems*, 118, 105793. <https://doi.org/10.1016/j.ijepes.2019.105793>
- Muslim, M. A., Rukmana, S. H., Sugiharti, E., Prasetyo, B., & Alimah, S. (2018). Optimization of C4.5 Algorithm-Based Particle Swarm Optimization for Breast Cancer Diagnosis. *Journal of Physics: Conference Series*, 983, 012063. <https://doi.org/10.1088/1742-6596/983/1/012063>
- Sari, R., Ramdhania, K. F., & Purnomo, R. (2022). Team-Teaching-Based Course Scheduling Using Genetic Algorithm. *PIKSEL : Penelitian Ilmu Komputer Sistem Embedded and Logic*, 10(1), 55–66. <https://doi.org/10.33558/piksel.v10i1.4416>
- Sudrajat, B. (2022). Penggunaan Algoritma C4.5 Untuk Menentukan Kepuasan Pelanggan Pada Warnet Game Victory. *Jurnal Ilmu Teknik Dan Komputer*, 6(1).
- Sutrisno, T., & Claudia, S. (2018). Analisis Rekomendasi Peminatan Menggunakan Metode Decision Tree dengan Algoritma C4.5. *Computatio*:

Journal of Computer Science and Information System, 2(2), 178–185.

Zheng, X., Feng, W., Huang, M., & Feng, S. (2021). Optimization of PBFT Algorithm Based on Improved C4.5. *Mathematical Problems in Engineering*, 2021, 1–7. <https://doi.org/10.1155/2021/5542078>