UNIVERSITAS BHAYANGKARA JAKARTA RAYA
FAKULTAS ILMU KOMPUTER
Kampus I: Jl. Harsono RM No. 67, Ragunan, Pasar Minggu, Jakarta Selatan, 12550
Telepon: (021) 27808121 – 27808882
Kampus II: Jl. Raya Perjuangan, Marga Mulya, Bekasi Utara, Jawa Barat, 17142
Telepon: (021) 88955882, Fax.: (021) 88955871
Web: fasilkom.ubharajaya.ac.id, E-mail: fasilkom@ubharajaya.ac.id

SURAT TUGAS

Nomor: ST/182/III/2023/FASILKOM-UBJ

1. Dasar: Kalender Akademik Ubhara Jaya Tahun Akademik 2022/2023.

2. Dalam rangka mewujudkan Tri Dharma Perguruan Tinggi untuk Dosen di Universitas Bhayangkara Jakarta Raya maka dihimbau untuk melakukan penelitian.

3. Sehubungan dengan hal tersebut di atas, maka Dekan Fakultas Ilmu Komputer Universitas Bhayangkara Jakarta Raya menugaskan:

NO.	NAMA	NIDN	JABATAN	KETERANGAN
1.	Rafika Sari, S.Si., M.Si.	0329098902	Dosen Tetap Prodi Informatika	Sebagai Penulis Pertama
2.	Khairunnisa Fadhilla Ramdhania, S.Si., M.Si.	0328039201	Dosen Tetap Prodi Informatika	Sebagai Penulis Ketiga

Membuat Artikel Ilmiah dengan judul "***Decision Support System Design for Informatics Student Final Projects Using C4.5 Algorithm***" dengan menerima LoA pada tanggal 10 Maret 2023 untuk dipublikasikan di media Penelitian Ilmu Komputer, Sistem *Embedded and Logic* (PIKSEL), Vol. 11, No. 1, Maret 2023, Hal. 123-134, p-ISSN: 2303-3304, e-ISSN: 2620-3553.

4. Demikian penugasan ini agar dapat dilaksanakan dengan penuh rasa tanggung jawab.

Jakarta, 10 Maret 2023
DEKAN FAKULTAS ILMU KOMPUTER

Dr. Dra. Tyastuti Sri Lestari, M.M.
NIP. 1408206

Articles

Storyboard Design of Android-Based Learning Multimedia Integration Application Using Standard Process Tutorial Model

Santi Purwanti 1-10

PDF

DOI : https://doi.org/10.33558/piksel.v11i1.5893

Abstract View: 1, PDF Download: 0

Pi Hole on SOE Computer Network using Raspberry Pi 3 Model B+ to Optimize Bandwidth Management and Improve Employee Performance

Rahmat Novrianda Dasmen, Darwin Darwin, Irham Irham, Bima Riansyah 11-22

PDF

DOI : https://doi.org/10.33558/piksel.v11i1.5911

Abstract View: 0, PDF Download: 2

The Influence of Youtube Ads on Purchase Intention

Willy Kristian, RA Dyah Wahyu Sukmaningsih, Eric Gunawan, Rafy Pranadya Annaufal, Rafi Giffari 23-34

PDF

DOI : https://doi.org/10.33558/piksel.v11i1.5919

Abstract View: 1, PDF Download: 0

Optimization of Random Forest Prediction for Industrial Energy Consumption Using Genetic Algorithms

Sartini Sartini, Luthfia Rohimah, Yana Iqbal Maulana, Supriatin Supriatin, Dewi Yuliandari 35-44

PDF

DOI : https://doi.org/10.33558/piksel.v11i1.5886

Abstract View: 0, PDF Download: 1

Cluster Analysis Using Principal Component Analysis Method and K-Means to Find Out the Compliance Group of Property Tax

Rully Pramudita, Nining Rahaningsih, Sekar Puspita Arum, Medina Aprilia Putri, Sok Piseth 45-54

PDF

DOI : https://doi.org/10.33558/piksel.v11i1.5924

EfficientNetV2M for Image Classification of Tomato Leaf Deseases

Arazka Firdaus Anavyanto, Maimunah Maimunah, Muhammad Resa Arif Yudianto, Pristi Sukmasetya

PDF

DOI : https://doi.org/10.33558/piksel.v11i1.5925

Identification of Website-Based Product Sales Frequency Patterns using Apriori Algorithms and Eclat Algorithms at Rio Food in Bekasi

Salwa Nabiila Pramuhesti, Herlawati Herlawati, Tyastuti Sri Lestari

PDF

DOI : https://doi.org/10.33558/piksel.v11i1.5941

The Weighted Product Method and the Multi-Objective Optimization on the Basis of Ratio Analysis Method for Determining the Best Customer

Mugiarso Mugiarso, Rasim Rasim

PDF

DOI : https://doi.org/10.33558/piksel.v11i1.6325

Factors Influencing Students' Intention to use Online Tutoring Applications in Jakarta

RA Dyah Wahyu Sukmaningsih, Adam Kurniawan, Ronald Ronald

PDF

DOI : https://doi.org/10.33558/piksel.v11i1.6318

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

Rafika Sari, Hasan Fatoni, Khairunnisa Fadhilla Ramdhania

PDF

DOI : https://doi.org/10.33558/piksel.v11i1.5954

View All Issues ›

..:: JOURNAL IDENTITY ::..

Journal Name	PIKSEL : Penelitian Ilmu Komputer Sistem Embedded and Logic
ISSN	2303-3304 (*print*), 2620-3553 (*online*)
DOI	*prefix:* 10.33558
Publisher	Department of Computer Engineering, Universitas Islam 45
Publication Schedule	March and September
Website	http://jurnal.unismabekasi.ac.id/index.php/piksel

INDEXED BY :

P-ISSN 2303-3304
E-ISSN 2620-3553
Penelitian Ilmu Komputer, Sistem Embedded & Logic | PIKSEL
sinta S3

Editorial Team

EDITOR IN CHIEF

Rahmadya Trias Handayanto, S.T., M.Kom., Ph.D., *(Scopus ID:* 55014574400, *Google Scholar Universitas Islam 45)*

DEPUTY EDITOR IN CHIEF

Inna Ekawati, S.T., MMSI, *(Scopus ID: 57221501629, Google Scholar Universitas Islam 45)*

BOARD OF EDITORS

Maimunah, S.Si., M.Kom. *(Scopus ID:* 57215528459, *Google Scholar, Universitas Muhammadiyah Magelang)*

Deshinta Arrova Dewi. *(Scopus ID: 55012068200, INTI International University,* ***Malaysia****)*

Retno Nugroho Whidhiasih, S.Kom., M.Kom. *(Scopus ID:* 55613478500, *Google Scholar, Universitas Islam 45)*

Endang Retnoningsih, S.Kom., M.Kom. *(Scopus ID:* 57215526966, Google Scholar Institut Bisnis Muhammadiyah Bekasi*)*

Fata Nidaul Khasanah, S.Kom., M.Eng. *(Scopus ID:* 57189353040, *Google Scholar, Universitas Bhayangkara Jakarta Raya)*

Yopi Handrianto, S.Kom., M.Kom. *(Scopus ID:* 57215294416, *Google Scholar Universitas Bina Sarana Informatika)*

Richard, S.Kom., M.M. *(Scopus ID:* 56638189100, *Google Scholar Universitas Bina Nusantara)*

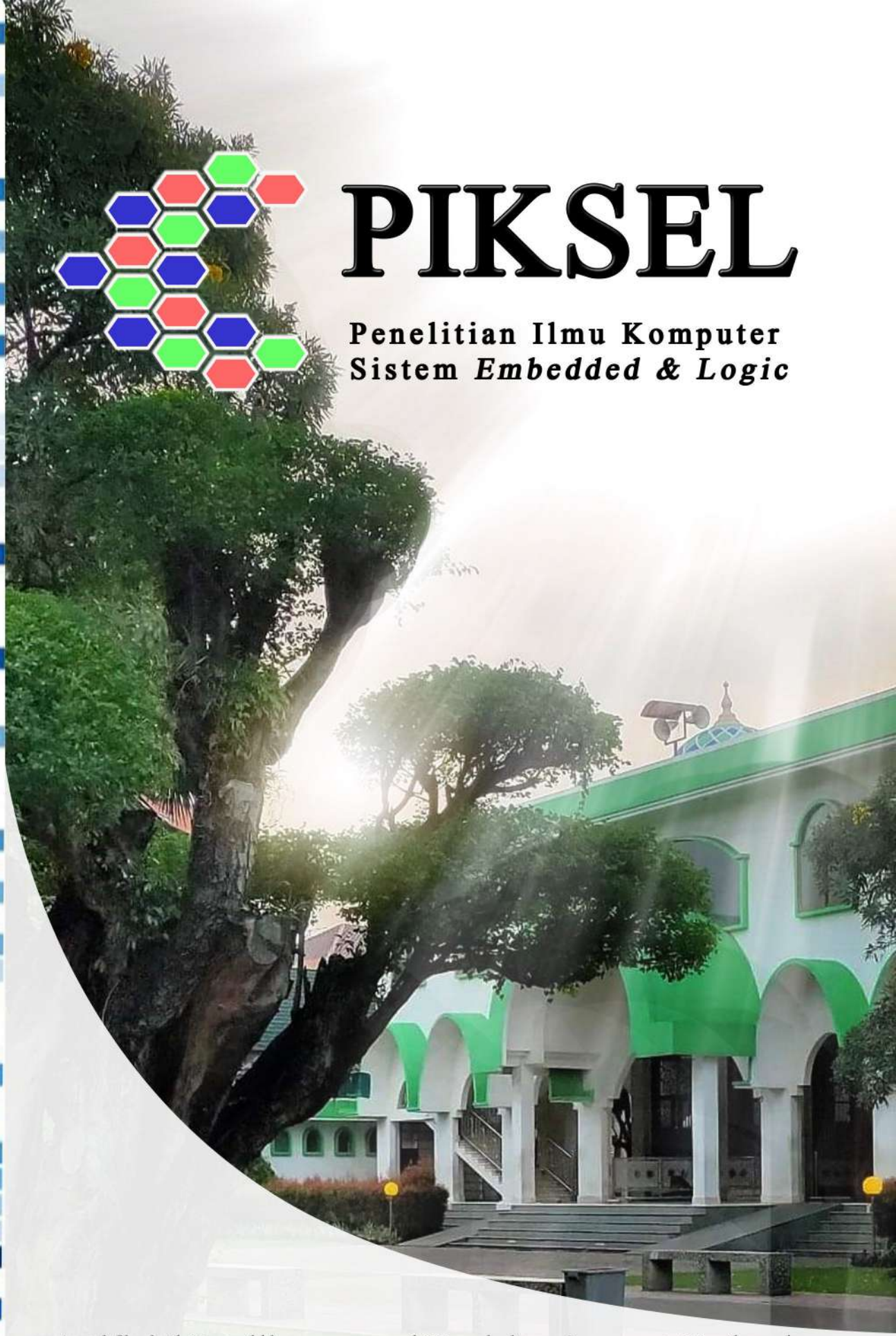
Ben Rahman, B.Sc., S.Kom., M.MSI. *(Scopus ID:* 57215525641, *Google Scholar Universitas Nasional)*

TECHNICAL EDITOR

Irwan Sukandar

Muryanti

Sumarlin



PIKSEL

Penelitian Ilmu Komputer
Sistem *Embedded & Logic*

Artificial Intelligence and Decision Support Technology to Improve Quality of Life

Department of Computer Engineering
Universitas Islam "45" Bekasi

Vol. 11 No. 1 March 2023
PIKSEL
Penelitian Ilmu Komputer
Sistem Embedded and Logic

Penelitian Ilmu Komputer, Sistem Embedded and Logic
p-ISSN: 2303-3304, e-ISSN: 2620-3553
Vol. 11 (1): 123 – 134 (March 2023)
https://doi.org/10.33558/piksel.v11i1.5954

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

Rafika Sari [1,*], Hasan Fatoni [1], Khairunnisa Fadhilla Ramdhania [1]

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

[1] Informatics, Faculty of Computer Science; Universitas Bhayangkara Jakarta Raya; Jl. Raya Perjuangan, Bekasi Utara, Jawa Barat, Indonesia; telp.021-88955882; e-mail: rafika.sari@dsn.ubharajaya.ac.id, hasanfatoni96@gmail.com, khairunnisa.fadhilla@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 cources, 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 ***|Si|*** is the number of cases on the ith 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	Transkip 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 Weight	Courses 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.

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

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

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

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

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

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

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							
Atribute	**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							
Atribute	**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)

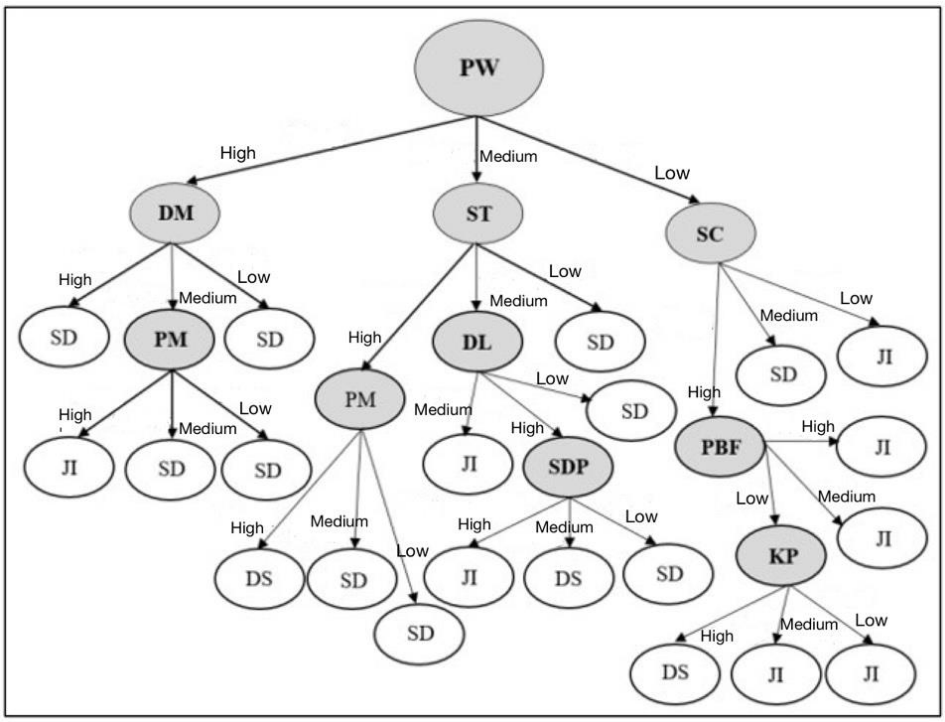
Accuracy = ((TP+TN)/(TP+TN+FP+FN))*100% = ((21+0)/(21+0+9+0))*100% = (21/30)*100% = 70%

Precision = (TP/(TP+FP))*100% = (21/(21+9))*100% = (21/30)*100% = 70%

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., Nouira, 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., Prasetiyo, 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 Decission 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

SERTIFIKAT

Direktorat Jenderal Pendidikan Tinggi, Riset dan Teknologi
Kementerian Pendidikan, Kebudayaan, Riset dan Teknologi Republik Indonesia

Kutipan dari Keputusan Direktorat Jenderal Pendidikan Tinggi, Riset dan Teknologi
Kementerian Pendidikan, Kebudayaan, Riset, dan Teknologi Republik Indonesia

Nomor 225/E/KPT/2022
Peringkat Akreditasi Jurnal Ilmiah periode III Tahun 2022

Nama Jurnal Ilmiah

PIKSEL : Penelitian Ilmu Komputer Sistem Embedded and Logic

E-ISSN: 26203553

Penerbit: Universitas Islam 45 Bekasi

Ditetapkan Sebagai Jurnal Ilmiah

TERAKREDITASI PERINGKAT 3

Akreditasi Berlaku selama 5 (lima) Tahun, yaitu
Volume 10 Nomor 1 Tahun 2022 sampai Volume 14 Nomor 2 Tahun 2026

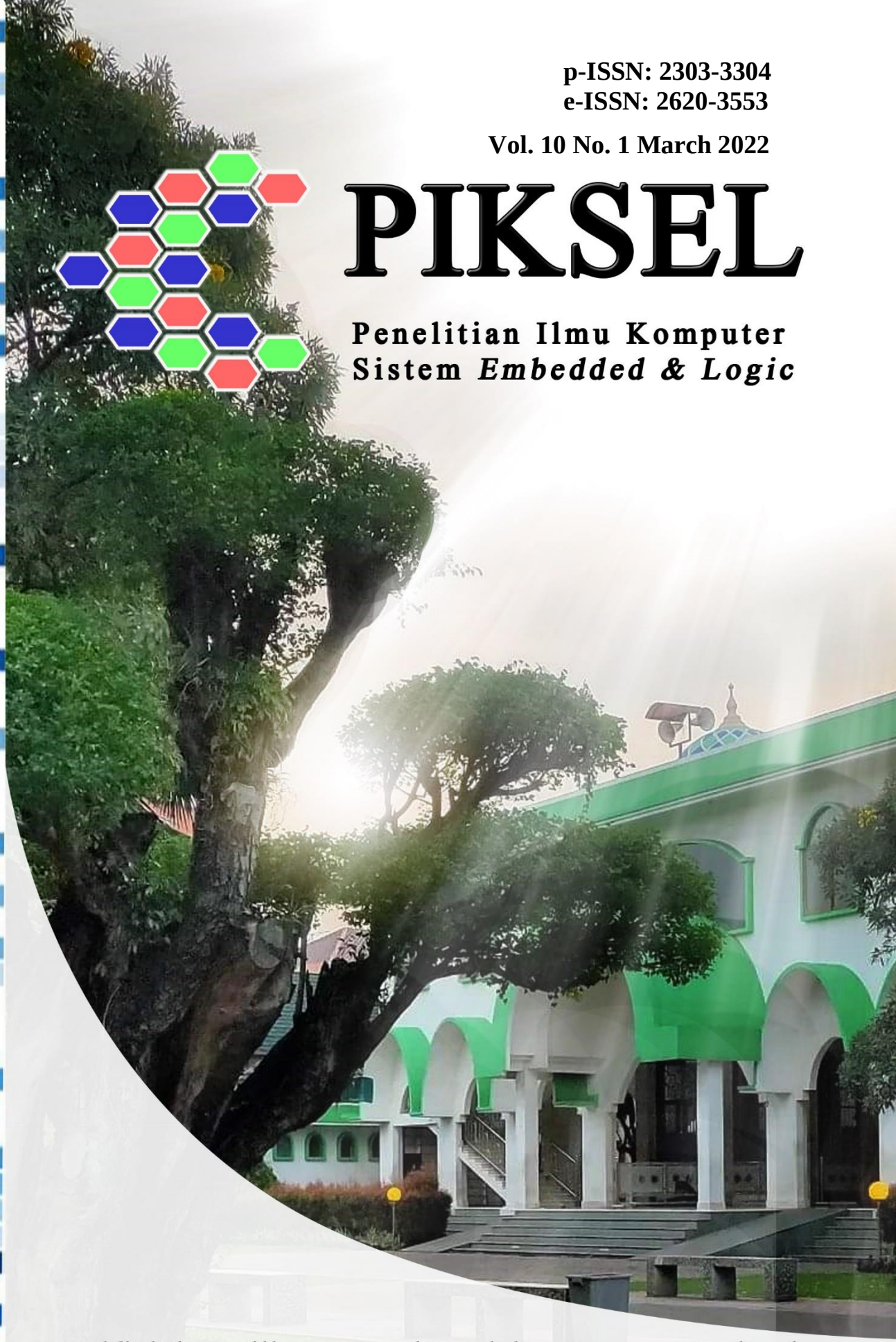
Jakarta, 07 December 2022
Plt. Direktur Jenderal Pendidikan Tinggi,
Riset, dan Teknologi

Prof. Ir. Nizam, M.Sc., DIC, Ph.D., IPU, ASEAN Eng
NIP. 196107061987101001

p-ISSN: 2303-3304
e-ISSN: 2620-3553
Vol. 10 No. 1 March 2022

PIKSEL

Penelitian Ilmu Komputer
Sistem *Embedded & Logic*



Artificial Intelligence and Decision Support Technology to Improve Quality of Life

Department of Computer Engineering
Universitas Islam "45" Bekasi

p-ISSN: 2303-3304
e-ISSN: 2620-3553
Vol. 10 No. 1 March 2022

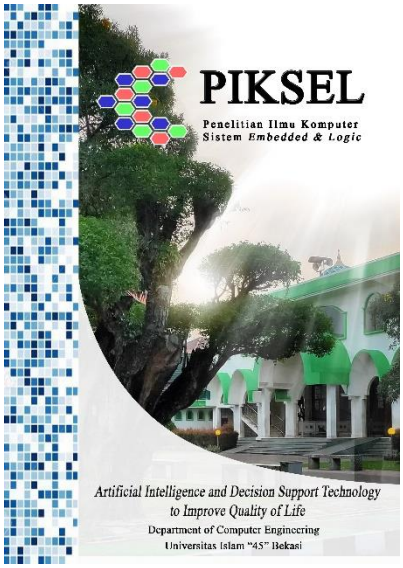
PIKSEL

Penelitian Ilmu Komputer
Sistem *Embedded & Logic*

Artificial Intelligence and Decision Support Technology to Improve Quality of Life

Department of Computer Engineering
Universitas Islam "45" Bekasi

p-ISSN: 2303-3304
e-ISSN: 2620-3553
Vol. 10 No. 1
March 2022



PIKSEL status is accredited by the Directorate General of Research Strengthening and Development No. 225/E/KPT/2022 with Indonesian Scientific Index (SINTA) journal-level of S3, starting from Volume 10 (1) 2022 to Volume 14 (2) 2026.

First publish in 2013.
Available online since 2018.

From Editor-in-Chief

Rahmadya, Ph.D.
Editor-in-Chief

السَّلاَمُ عَلَيْكُمْ وَرَحْمَةُ اللهِ وَبَرَكَاتُهُ

.

Best wishes to all the members of Editorial Board, Reviewers Panel, Authors and Readers of PIKSEL for a very happy, and stay healthy in current global situation.

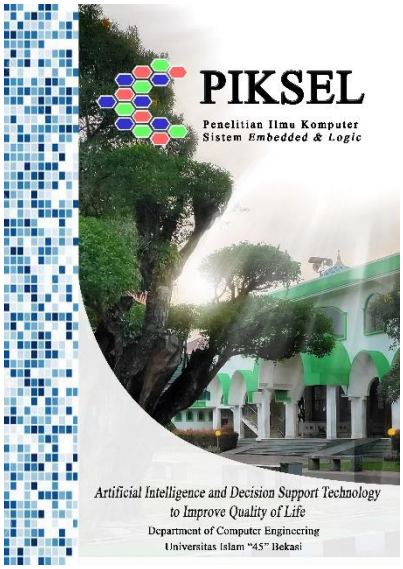
The computer science research is still needed in post-pandemic/endemic situation. Quality of life can be improved through the implementation of current computer science and information systems methods.

Although the face-to-face learning has just been started, the research to support teaching and learning, especially for scheduling is very useful as well as in business areas, e.g., helpdesk ticketing system or purchasing in a company. Also, a decision support technology to predict the major of high school students using forward chaining is presented. Other computer science methods, e.g., generative adversarial method (GAN), C4.5 algorithms, and genetic algorithm (GA) are discussed in this volume.

I hope this issue contribute to support nation after pandemic situation. And once again, thank you to members of Editorial Board, Reviewers Panel, Authors and Readers of PIKSEL (Penelitian Ilmu Komputer, *Sistem Embedded & Logic*).

Publisher: LPPM Universitas Islam 45
Office:
Fakultas Teknik Universitas Islam 45
Jl. Cut Meutia No. 83 Margahayu Kecamatan Bekasi Timur
Kota Bekasi Jawa Barat Indonesia 17113
Telp. (021) 8802015
e-mail: piksel.unisma@gmail.com
e-mail: piksel@unsimabekasi.ac.id
website: http://jurnal.unismabekasi.ac.id/index.php/piksel

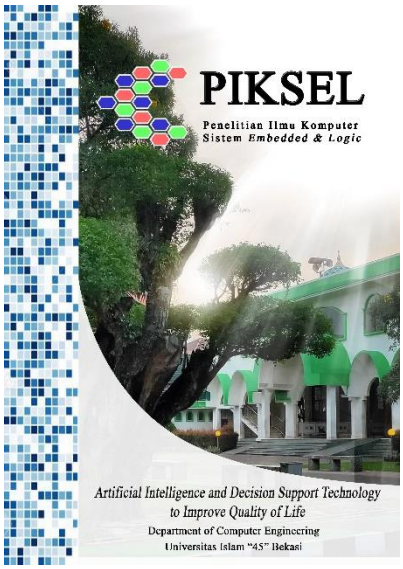
p-ISSN: 2303-3304
e-ISSN: 2620-3553
Vol. 10 No. 1
March 2022



PIKSEL status is accredited by the Directorate General of Research Strengthening and Development No. 225/E/KPT/2022 with Indonesian Scientific Index (SINTA) journal-level of S3, starting from Volume 10 (1) 2022 to Volume 14 (2) 2026.

Editor Board Journal PIKSEL

p-ISSN: 2303-3304
e-ISSN: 2620-3553
Vol. 10 No. 1
March 2022



PIKSEL status is accredited by the Directorate General of Research Strengthening and Development No. 225/E/KPT/2022 with Indonesian Scientific Index (SINTA) journal-level of S3, starting from Volume 10 (1) 2022 to Volume 14 (2) 2026.

REVIEWERS

Augustinus Bayu Primawan, D.Tech.Sc.
(Scopus ID: 57204114771, Universitas Sanata Dharma, Yogyakarta, Indonesia)
Petrus Sutyasadi, S.T., M.Eng., D.Eng.
(Scopus ID: 36968351900,, Politeknik Mekatronika Sanata Dharma, Yogyakarta, Indonesia)
Cahyono Sigit Pramudyo, S.T., M.T., D.Eng.
(Scopus ID: 57195353262, Universitas Islam Negeri Sunan Kalijaga, Yogyakarta, Indonesia)
Taqwa Hariguna, S.Kom., M.Kom., Ph.D.
(Scopus ID: 57193771775, Universitas AMIKOM Purwokerto, Purwokerto, Indonesia)
Henriyadi, S.Si., M.Sc.
(Litbang Departemen Pertanian, Jakarta, Indonesia)
Herlawati, S.Si., M.M.,M.Kom
(Scopus ID: 55613443500, Universitas Bhayangkara Jakarta Raya)
Eni Heni Hermaliani, S.Kom., M.M., M.Kom
(Scopus ID: 57200210484, Universitas Bina Sarana Informatika)
Malikus Sumadyo., S.Si., M.T.
(Scopus ID: 57193833463, Universitas Islam 45)
Dadan Irwan, S.T., M.Kom.
(Scopus ID: 55613449700, Universitas Islam 45)
Haryono, S.Kom., M.M.S.I.
(Scopus ID: 55015952700, Universitas Islam 45)
Seta Samsiana, S.T., M.T.
(Scopus ID: 56532498500, Universitas Islam 45)
Hendra Supendar, S.Kom., M.Kom
(Scopus ID: 57210461454, Universitas Bina Sarana Informatika, Indonesia)
Aji Akbar Firdaus, S.T., M.T
(Scopus ID: 56596623100, Universitas Airlangga, Surabaya, Indonesia)

Vol. 11 No. 1 March 2023
PIKSEL
Penelitian Ilmu Komputer
Sistem Embedded and Logic

Penelitian Ilmu Komputer, Sistem Embedded and Logic
p-ISSN: 2303-3304, e-ISSN: 2620-3553
Vol. 11 (1): 123 – 134 (March 2023)
https://doi.org/10.33558/piksel.v11i1.5954

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

Rafika Sari [1,*], Hasan Fatoni [1], Khairunnisa Fadhilla Ramdhania [1]

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

[1] Informatics, Faculty of Computer Science; Universitas Bhayangkara Jakarta Raya; Jl. Raya Perjuangan, Bekasi Utara, Jawa Barat, Indonesia; telp.021-88955882; e-mail: rafika.sari@dsn.ubharajaya.ac.id, hasanfatoni96@gmail.com, khairunnisa.fadhilla@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 cources, 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 ***|Si|*** is the number of cases on the ith 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	Transkip 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 Weight	Courses 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.

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

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

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

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

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

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

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							
Atribute	**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							
Atribute	**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)

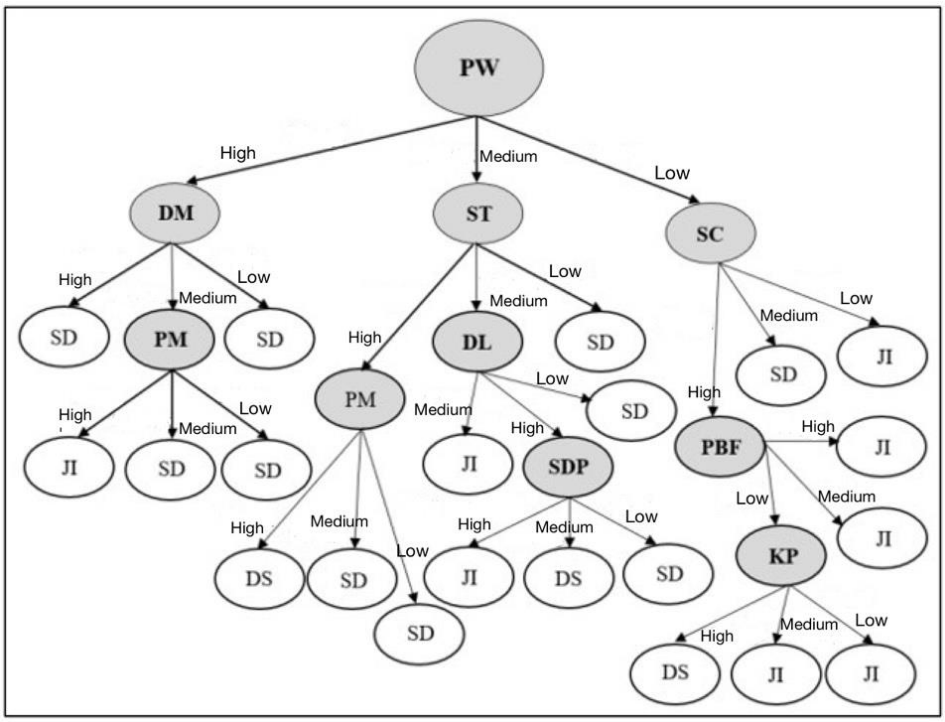
Accuracy = ((TP+TN)/(TP+TN+FP+FN))*100% = ((21+0)/(21+0+9+0))*100% = (21/30)*100% = 70%

Precision = (TP/(TP+FP))*100% = (21/(21+9))*100% = (21/30)*100% = 70%

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., Nouira, 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., Prasetiyo, 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 Decission 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

Camera Ready_Decision-Making System for Field of Student's Final Project Using C4.5 Algorithm

by Rafika Sari

Submission date: 16-Nov-2022 12:34AM (UTC-0600)
Submission ID: 1955594915
File name: a_Decisio_Tree_C4.5_Field_of_Student_s_Final_Project_Rafika.docx (197.91K)
Word count: 4246
Character count: 20409

Decision-Making System for Field of Student's Final Project Using C4.5 Algorithm

***Abstract*—Academic consultation activities between students and academic advisory lecturers are necessary to help students carry out activity's lectures. Based on the transcripts of grades that have been obtained, many students do not choose the field that suits academic ability so there are a lot of nonlinearities between the value of the field course and the field of the final project. The purpose of this research is to minimize aspects of student subjectivity to lecturers regarding the selection of areas of the field for students' final assignments, as well as minimize the nonlinearity between the grades of the courses that have been taken students with the field of the final project to be taken. The method used in this study is Data Mining Classification using the Decision Tree method and the C4.5 Algorithm, with the attributes involved, namely courses, value of field courses and areas of the field. Decision Algorithm Tree C4.5, which is an algorithm to change the shape of the data (table) into a tree model then change the tree model into a rule. The application of the Decision Tree C4.5 algorithm to the decision-making system in the field of specialization has been successfully carried out by producing an accuracy rate of 70% of the total data calculation results. The data used in this study is sampling data from several final-year students in the Ubhara-Jaya Informatics study program. The field decision-making system field as a result of research that has been done can be used as a recommendation for both the Informatics Study Program, Faculty of Computer Science Ubhara-Jaya and for final-year students to direct the field of research in the final project. It is hoped that further research will use more sample data so that the level of accuracy is better and can be implemented in website-based or mobile applications.**

Keywords—decision tree algorithms, student final assignments, field courses, decision-making systems

I. INTRODUCTION

The learning process in lectures requires a curriculum that supports a course in each semester. The curriculum is a set of plans and arrangements regarding the objectives, content, and learning materials as well as the methods used as guidelines for the implementation of learning activities to achieve national education goals. The Operational Curriculum created by the study program aims to determine the distribution and prerequisites of courses in each semester. With the Operational Curriculum, each student can find out what courses will be taken in a certain semester along with the prerequisite courses that must be met with certain grade requirements [1]. If students want to take courses with certain course prerequisites, then the prerequisite courses must be fulfilled in the previous semester and have met certain score requirements [2].

Sampling data used in this study is the data of final year students in the Informatics study program, Bhayangkara Jakarta Raya University (UBJ). There are three types of field that serve as a reference for completing the final project (thesis). In fact, in the selection of fields offered by the Study Program, many students do not choose fields that are in accordance with their academic abilities so that students experience difficulties in learning until the thesis writing process is in accordance with the chosen field. Field should be able to facilitate students in completing their studies, because field aims to enable students to focus on learning the specific concentration of field [3]. In order to help solve problems encountered related to academic consulting activities, a system was created that can provide analysis of recommendations for student interest using the data mining classification method using the Decision Tree C4.5 algorithm. The source of the research data used is the final grade 2016 - 2018 student grade data obtained from the UBJ Informatics Study Program which was taken by sampling with a weight of three batches with 10 student data in each batch. The data to be processed in this research is taken from several grades of semester 4 - 7 courses and classified based on the field of field.

II. RELATED STUDY

A. State of the Art

The decision-making system is one approach that is often used to analyze a problem in various conditions and has been used in various sectors of life [4]. Many researchers and practitioners have published research results related to decision-making systems. Several studies regarding the application of the C4.5 Algorithm for determining student majors resulted in the Decision Tree C4.5 Algorithm test results being more accurate than Naïve Bayes with an accuracy rate of 93.31% for student majors and 82.64% accuracy for major recommendations [5]. In line with the research predicting student performance using the C4.5 and ID3 classification algorithms. Based on student performance from the beginning of learning, this study can estimate student performance at the time of the exam [3]. Further research on determining customer satisfaction with Java internet providers uses the C4.5 algorithm which is used to calculate the level of customer satisfaction with rental fees, internet access and facilities and services which results in the calculation of the highest gain value of 0.970951[6]. Similar research was also conducted to analyze field recommendations using the Decision Tree method with the C4.5 algorithm which compares the accuracy value of the proportion of the same number of classes with the proportion of different classes showing that the proportion of the number of classes can affect the accuracy of the resulting model tree [1]. Then other research is regarding the classification of Indonesian Youtube Channels Using the C4.5 Algorithm with test results using Cross-Validation obtained an accuracy value of 92.73% with a class precision of Very Good 80.77% and class precision Good 96.43%, and class recall Very Good 87.50% and class Good recall 94.19% [7]. Another related research is news classification using the C4.5 algorithm which produces the highest accuracy value of 84% [8]. The C4.5 algorithm has also been used to classify English emotional in a research that has been done.[9]

B. Decision Tree

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[10]. 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 [3][11].

C. C4.5 Algorithm

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 [12]. At 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"[6].

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 [13].

III. RESEARCH METHOD

A. 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 [14]. Calculation of attribute values in the C4.5 algorithm with the following stages:

- Calculate the value of entropy,
- Calculate the gain ratio value for each attribute,
- The attribute with the highest gain is selected as the root and the attribute with the lower gain ratio is used as a branch.
- Calculate the value of the gain ratio of each attribute except the root.
- The highest gain ratio attribute is selected as a branch.
- Repeat the previous step until the resulting gain = 0.

To calculate the entropy value can be calculated by equation (1) [15] [16]

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

where S is the case set, n is the number of case partitions and P_i is the proportion of S_i to S. The information gain value can be calculated using equation (2).

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

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).

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

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.

$$GainRatio(S, A) = \frac{Gain\ (S, A)}{Split\ Info\ (S, A)} \qquad (4)$$

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.

B. Data Penelitian

Data selection variables used in this study are the grades of compulsory courses for semesters 4 – 7 of the class of 2016 – 2018 which are related to the field of field, with the classification of the field of field which can be seen in table 1.

TABLE I. 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 (practical work)
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

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 II. The variables in table II 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 III.

TABLE II. CLASSIFICATION OF STUDENTS' SCORE AND FIELD OF INTEREST

STD	Transkip 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
STD 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

TABLE III. VALUE WEIGHT CATEGORY

Value weight	Category
3.50 – 4.00	High
2.75 – 3.50	Medium
< 2.75	Low

IV. RESULT AND DISCUSSION

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 [17]. The data in table II will be transformed into 3 variables based on the weight values

presented in table 3. From table 4 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 IV for calculating the number of cases. Table V 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 IV. TRANSFORMATION OF STUDENT'S TRANSCRIPTION

No	Student's Transcription												Final Project Field
	PW	PM	PBF	KP	JK	KJK	KF	ST	SDP	SC	DM	DL	
1	Medium	Medium	Medium	Medium	High	Medium	Medium	Low	High	Medium	Medium	Medium	SD
2	High	Medium	Medium	High	High	Medium	High	High	High	Medium	Medium	Medium	SD
3	Low	Low	Low	High	High	Low	Medium	Low	Medium	High	Medium	High	DS
4	High	High	High	High	High	High	Medium	Medium	Medium	High	Medium	High	JI
5	High	High	High	High	High	Medium	High	Medium	Medium	High	High	High	SD
6	Low	Low	Low	Low	High	High	Low	Low	Low	High	Low	High	JI
7	Low	Medium	Low	High	High	Low	High	Low	Medium	Low	Low	Low	JI
8	Low	Medium	Medium	High	High	Medium	Medium	Low	Medium	High	Low	High	JI
9	Low	Low	Low	High	High	Low	Medium	Medium	Medium	Low	Low	Kecil	JI
10	Low	Low	Low	High	Low	Medium	Low	Low	Medium	High	Low	High	DS
11	Medium	Medium	Low	High	Medium	Medium	Medium	High	High	Low	High	Low	SD
12	Medium	Medium	Low	High	High	Low	High	Low	High	Medium	Low	High	SD
13	Low	Low	High	High	Medium	Medium	High	Medium	High	High	Low	High	JI
14	Low	Medium	High	High	Medium	Medium	High	Medium	Medium	Medium	Medium	High	SD
15	High	Medium	High	High	Low	Medium	High	High	High	Medium	Low	Medium	SD
16	Low	Low	High	High	Low	Medium	High	Medium	High	Medium	Low	Low	SD
17	Medium	Medium	High	High	Medium	Medium	High	Medium	Medium	Medium	Medium	Low	SD
18	High	High	High	High	High	Low	High	High	High	Low	High	High	SD
19	Medium	Medium	High	Medium	High	Medium	High	Medium	Medium	Medium	Medium	High	DS
20	Medium	Medium	High	High	High	Medium	High	Medium	High	High	Medium	High	JI
21	Medium	High	Medium	High	High	High	High	High	Medium	High	High	High	DS
22	High	High	Medium	High	High	High	Medium	Medium	Low	High	High	Medium	SD
23	Medium	High	Medium	High	High	Medium	High	Medium	Medium	Medium	High	Medium	JI
24	Medium	High	High	High	High	High	High	High	High	High	High	Low	DS
25	Medium	Low	Medium	High	High	High	Medium	Medium	Medium	Medium	High	High	DS
26	High	High	High	High	High	High	Medium	Medium	Medium	Medium	High	High	SD
27	Medium	Medium	High	High	High	High	Medium	High	Medium	High	High	High	SD
28	Medium	Medium	High	High	High	Medium	Medium	High	Medium	High	Medium	High	SD
29	High	High	High	High	High	Medium	Medium	High	Medium	High	High	High	SD
30	Medium	High	High	High	High	Medium	Medium	Low	Medium	Low	High	Medium	SD

TABLE V. NUMBER OF CASES

Value Weight	Courses											
	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 Infrastruktur (JI)	8
Dara Science (DS)	6

A. The Calculation Results

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.

1) 1) Calculating Entropy (S) using equation (1)

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

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

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.

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

$$Entropy(High) = -\left(\frac{7}{8}\right) * log_2\left(\frac{7}{8}\right) + -\left(\frac{1}{8}\right) * log_2\left(\frac{1}{8}\right) + -\left(\frac{0}{8}\right) * log_2\left(\frac{0}{8}\right) = 0.543564443$$

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

$$Entropy(Medium) = -\left(\frac{7}{13}\right) * log_2\left(\frac{7}{13}\right) + -\left(\frac{2}{13}\right) * log_2\left(\frac{2}{13}\right) + -\left(\frac{4}{13}\right) * log_2\left(\frac{4}{13}\right) = 1.419556299$$

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

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

3) 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 VI 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. [18].

TABLE VI. ENTROPY AND GAIN CALCULATION

Node	ATRIBUTE	WEIGHT	NUMER OF CASES (*S*)	SD (S_1)	JI (S_2)	DS (S_3)	ENTROPY	GAIN
1	Total		30	16	8	6	1.456564763	
	Pemrograman Web (PW)	High	8	7	1	0	0.543564443	**0.265817032**
		Medium	13	7	2	4	1.419556299	

Node	ATRIBUTE	WEIGHT	NUMER OF CASES (*S*)	SD (S_1)	JI (S_2)	DS (S_3)	ENTROPY	GAIN
		Low	9	2	5	2	**1.435520503**	
		Total	30					
	Pemrograman Mobile (PM)	High	10	6	2	2	1.370950594	0.167474123
		Medium	13	9	3	1	1.140115679	
		Low	7	1	3	3	1.448815638	
		Total	30					
	Framework-based Programming (PBF)	High	16	11	3	2	1.199460293	0.090412811
		Medium	7	3	2	2	1.556656709	
		Low	7	2	3	2	1.556656709	
		Total	30					
	Application Projects (practical work) (KP)	High	27	15	7	5	1.426573285	0.10598214
		Medium	2	1	0	1	1	
		Low	1	0	1	0	0	
		Total	30					
	Computer Networks (JK)	High	23	11	7	5	1.509871837	0.098996355
		Medium	4	3	1	0	0.811278124	
		Low	3	2	0	1	0.918295834	
		Total	30					
	Computer Network Security (KJK)	High	8	3	2	3	1.561278124	0.072131895
		Medium	17	11	4	2	1.260771796	
		Low	5	2	2	1	1.521928095	
		Total	30					
	Computer Forensics (KF)	High	15	8	4	3	1.456564763	0.083254753
		Medium	13	8	3	2	1.334679142	
		Low	2	0	1	1	1	
		Total	30					
	Integrated Systems (ST)	High	9	7	0	2	0.764204505	0.178085882
		Medium	13	6	5	2	1.460484683	
		Low	8	3	3	2	1.561278124	
		Total	30					
	Statistics and Probability (SDP)	High	10	7	2	1	1.15677965	0.07632591
		Medium	18	8	5	5	1.546631617	
		Low	2	1	1	0	1	
		Total	30					
	Intelligent Systems (SC)	High	14	5	5	4	1.577406283	0.156825137
		Medium	11	8	1	2	1.095795256	
		Low	5	3	2	0	0.970950594	
		Total	30					
	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	
		Total	30					
	Deep Learning (DL)	High	18	8	5	5	1.546631617	0.106751728
		Medium	6	5	1	0	0.650022422	
		Low	6	3	2	1	1.459147917	
		Total	30					

The stages of the calculation process are presented in table VI are:

- Calculating Entropy and Gain
- 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.
- 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.
- Create Rules based on decision trees [19]

B. *Validasi Test*

Validation test is carried out by analyzing the results of calculations using the Confusion Matrix model [20]. Table 7 is the result of the calculation of the confusion matrix in the C4.5 algorithm.

TABLE VII. CONFUSION MATRIX

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

- $Accuracy = \left(\frac{TP+TN}{TP+TN+FP+FN}\right) * 100\%$
$= \left(\frac{21+0}{21+0+9+0}\right) * 100\% = \left(\frac{21}{30}\right) * 100\% = 70\%$
- $Precision = \left(\frac{TP}{TP+FP}\right) * 100\% = \left(\frac{21}{21+9}\right) * 100\%$
$= \left(\frac{21}{30}\right) * 100\% = 70\%$
- $Recall = \left(\frac{TP}{TP+FN}\right) * 100\% = \left(\frac{21}{21+0}\right) * 100\%$
$= \left(\frac{21}{21}\right) * 100\% = 100\%$

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.

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.

- IF (PW='High') AND (DM='High') THEN Label = Software Development
- IF (PW='High') AND (DM='Medium') AND (PM='High') THEN Label = Jaringan dan Infrastruktur
- IF (PW='High') AND (DM='Medium') AND (PM='Medium') THEN Label = Software Development
- IF (PW='Medium') AND (ST='High') AND (PM='High') THEN Label = Data Science
- IF (PW='Medium') AND (ST='High') AND (PM='Medium') THEN Label = Software Development
- IF (PW='Medium') AND (ST='Medium') AND (DL='High') AND (SDP='High') THEN Label = Jaringan dan Infrastruktur
- IF (PW='Medium') AND (ST='Medium') AND (DL='High') AND (SDP='Medium') THEN Label = Data Science
- IF (PW='Medium') AND (ST='Medium') AND (DL='Medium') THEN Label = Jaringan dan Infrastruktur.

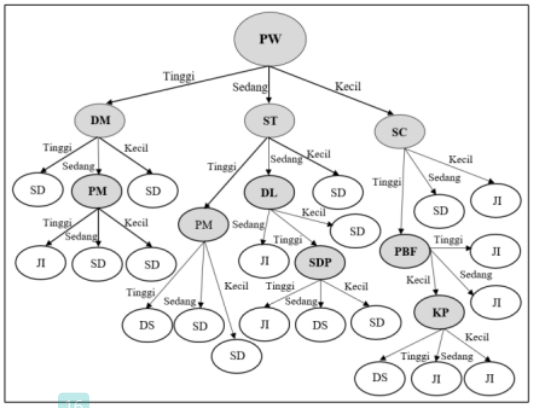


Fig. 1. Decision tree results

The Decision Tree method using the C4.5 Algorithm has been successfully applied in making rules for the decision-making system in the field of student's final project by using training data based on variables: courses, areas of field and the value of field courses, by producing an accuracy rate of 70% from the calculation results of data test. The field decision-making system field as a result of research that has been done can be used as a recommendation for both the Informatics Study Program, Faculty of Computer Science Ubhara-Jaya and for final-year students to direct the field of research in the final project. It is hoped that further research will use more sample data so that the level of accuracy is better and can be implemented in website-based or mobile applications.

ACKNOWLEDGMENT

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.

REFERENCES

Camera Ready_Decision-Making System for Field of Student's Final Project Using C4.5 Algorithm

ORIGINALITY REPORT

20%	17%	18%	13%
SIMILARITY INDEX	INTERNET SOURCES	PUBLICATIONS	STUDENT PAPERS

PRIMARY SOURCES

1	ijcis.net Internet Source	3%
2	login.seaninstitute.org Internet Source	2%
3	image.guardian.co.uk Internet Source	2%
4	www.publichealth.gov.au Internet Source	1%
5	Wahyu Supriyatin. "Palm oil extraction rate prediction based on the fruit ripeness levels using C4.5 algorithm", ILKOM Jurnal Ilmiah, 2021 Publication	1%
6	www.lrl.mn.gov Internet Source	1%
7	D Arifin, A Hadiana. "Computer-based Techniques for Predicting the Failure of Student Studies Using the Decision Tree	1%

method", IOP Conference Series: Materials Science and Engineering, 2019
Publication

8	journal2.uad.ac.id Internet Source	1 %
9	Submitted to Nexford University Student Paper	1 %
10	join.if.uinsgd.ac.id Internet Source	1 %
11	www.ejurnal.ubharajaya.ac.id Internet Source	1 %
12	www.napierplanning.govt.nz Internet Source	1 %
13	Meilin Widyastuti, Agnes Gracella Fepdiani Simanjuntak, Dedy Hartama, Agus Perdana Windarto, Anjar Wanto. "Classification Model C.45 on Determining the Quality of Custumer Service in Bank BTN Pematangsiantar Branch", Journal of Physics: Conference Series, 2019 Publication	<1 %
14	Rina Novita, Supratman Zakir, Agus Nur Khomarudin, Efmi Maiyana, Hamimah Hasyim. "Use of the C4.5 Algorithm in Determining Scholarship Recipients", Journal of Physics: Conference Series, 2021 Publication	<1 %

15	Sucipto, Kusrini, Emha Luthfi Taufiq. "Classification method of multi-class on C4.5 algorithm for fish diseases", 2016 2nd International Conference on Science in Information Technology (ICSITech), 2016 Publication	<1 %
16	Glen Nur Awaludin, Yana Aditia Gerhana, Dian Sa'adillah Maylawati, Wahyudin Darmalaksana et al. "Comparison of Decision Tree C4.5 Algorithm with K-Nearest Neighbor (KNN) Algorithm in Hadith Classification", 2020 6th International Conference on Computing Engineering and Design (ICCED), 2020 Publication	<1 %
17	Eka Irawan, Sumarno, Indra Gunawan, Heru Satria Tambunan, Hendri Qurniawan. "Application of Classification Algorithm C4.5 in Recommendations for Natural Tourism Development in District Simalungun", Journal of Physics: Conference Series, 2019 Publication	<1 %
18	I Made Wirawan, Triyanna Widiyaningtyas, Nurwakiah B. Siti. "Nutritional Status of Infants Classification by Calculating Anthropometry Through C4.5 Algorithm", 2019 International Conference on Electrical,	<1 %

Electronics and Information Engineering (ICEEIE), 2019
Publication

19	Noor Abdul Haris, Muhammad Nidhom, Arif Setia Sandi Ariyanto, Hari Asgar, Kusrini. "KIP Recipient Decision Making For Students Affected by Covid_19 Pendemi Using Fuzzy MADM Method", 2020 3rd International Conference on Information and Communications Technology (ICOIACT), 2020 Publication	<1 %
20	Rini Sovia, Abulwafa Muhammad, Syafri Arlis, Guslendra Guslendra, Sarjon Defit. "Analysis of sales levels of pharmaceutical products by using data mining algorithm C45", Indonesian Journal of Electrical Engineering and Computer Science, 2021 Publication	<1 %
21	journal.utem.edu.my Internet Source	<1 %
22	jurnal.ceredindonesia.or.id Internet Source	<1 %
23	repository.bsi.ac.id Internet Source	<1 %
24	www.info2.dec.state.ak.us Internet Source	<1 %

25	Charles A. Ellis, Robyn L. Miller, Vince D. Calhoun. "An Approach for Estimating Explanation Uncertainty in fMRI dFNC Classification", Cold Spring Harbor Laboratory, 2022 Publication	<1 %
26	Submitted to Universitas Mercu Buana Student Paper	<1 %
27	Submitted to University of Witwatersrand Student Paper	<1 %
28	scitepress.org Internet Source	<1 %
29	Munif Ma'arij Kholil, Farrikh Alzami, M. Arif Soeleman. "AdaBoost Based C4.5 Accuracy Improvement on Credit Customer Classification", 2022 International Seminar on Application for Technology of Information and Communication (iSemantic), 2022 Publication	<1 %
30	Saruni Dwiasnati, Yudo Devianto. "Utilization of Prediction Data for Prospective Decision Customers Insurance Using the Classification Method of C.45 and Naive Bayes Algorithms", Journal of Physics: Conference Series, 2019 Publication	<1 %
31	enerbioscrub.ciemat.es Internet Source	<1 %

32	jurnal.iaii.or.id Internet Source	<1%
33	Ketjie, Viny Christanti Mawardi, Novario Jaya Perdana. "Prediction of Credit Card Using the Naïve Bayes Method and C4.5 Algorithm", IOP Conference Series: Materials Science and Engineering, 2020 Publication	<1%
34	Agung Wibowo, Yuri Rahayu, Andi Riyanto, Taufik Hidayatulloh. "Classification algorithm for edible mushroom identification", 2018 International Conference on Information and Communications Technology (ICOIACT), 2018 Publication	<1%

Exclude quotes	On	Exclude matches	Off
Exclude bibliography	On		

Camera Ready_Decision-Making System for Field of Student's Final Project Using C4.5 Algorithm