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/1334/XII/2022/FASILKOM-UBJ

Tentang

PENUGASAN DOSEN MENJADI PENULIS DALAM KEGIATAN SEMINAR

DEKAN FAKULTAS ILMU KOMPUTER UNIVERSITAS BHAYANGKARA JAKARTA RAYA

Menimbang : Bahwa dalam rangka Pengembangan Dosen Universitas Bhayangkara Jakarta Raya, maka dipandang perlu mengeluarkan Surat Tugas.

Mengingat :
1. Undang-Undang No. 20 Tahun 2003 tentang Sistem Pendidikan Nasional.
2. Undang-Undang No. 14 Tahun 2005 tentang Guru dan Dosen.
3. Undang-Undang No. 12 Tahun 2012 tentang Pendidikan Tinggi.
4. Permendikbud No. 3 Tahun 2020 tentang Standar Nasional Pendidikan Tinggi.
5. Kalender Akademik Universitas Bhayangkara Jakarta Raya Tahun Akademik 2022/2023.
6. Rencana Kerja dan Anggaran Pembelanjaan Universitas Bhayangkara Jakarta Raya Tahun 2022/2023.

DITUGASKAN

Kepada : **Khairunnisa Fadhilla Ramdhania, S.Si., M.Si.**
NIDN. 0328039201

Untuk :
1. Melaksanakan tugas sebagai **Penulis *Paper* dalam kegiatan Seminar Nasional TIK APTIKOM 2022 yang berjudul "*Decision-Making System for Field of Student's Final Project Using C4.5 Algorithm*"**.
2. Yang dilaksanakan pada:
   Hari : Kamis s.d. Jum'at
   Tanggal : 08 Desember 2022 s.d. 09 Desember 2022
   Tempat : Bali
   Penyelenggara : Asosiasi Pendidikan Tinggi Informatika dan Komputer (APTIKOM)
3. Melaporkan hasil pelaksanaan kepada Dekan Fakultas Ilmu Komputer Universitas Bhayangkara Jakarta Raya.
4. Melaksanakan tugas ini dengan penuh rasa tanggung jawab.

Selesai.

Ditetapkan di : Jakarta
Pada tanggal : 06 Desember 2022
DEKAN FAKULTAS ILMU KOMPUTER

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

SEMNASTIK

APTIKOM 2022

SERTIFIKAT PARTISIPASI

APTIKOM

ITB STIKOM BALI

Menyatakan bahwa:

Rafika Sari, Hasan Fatoni, Khairunnisa Fadhilla Ramdhania, Dwi Budi Srisulistiowati, Andy Achmad Hendarsetiawan, Ajif Yunizar Pratama Yusuf

telah menulis paper berjudul:

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

pada Seminar Nasional TIK APTIKOM 2022
yang dilaksanakan pada tanggal 08-09 Desember 2022 di Bali

Ketua Umum APTIKOM

APTIKOM
ASOSIASI PENDIDIKAN TINGGI INFORMATIKA DAN KOMPUTER

Zainal A. Hasibuan

Ketua Pelaksana SEMNASTIK 2022

SEMNASTIK
APTIKOM 2022

Husni Teja Sukmana

disponsori oleh:

Ruijie ORACLE EDUFECTA Kinobi Solusi Kampus Indonesia SYNNEX INFORMATIKA EVIMA

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

1st Rafika Sari
Informatic, Faculty of Computer Science
Bhayangkara Jakarta Raya University
Bekasi, Indonesia
rafika.sari@dsn.ubharajaya.ac.id

2nd Hasan Fatoni
Informatic, Faculty of Computer Science
Bhayangkara Jakarta Raya University
Bekasi, Indonesia
hasanfatoni96@gmail.com

3rd Khairunnisa Fadhilla Ramdhania
Informatic, Faculty of Computer Science
Bhayangkara Jakarta Raya University
Bekasi, Indonesia
khairunnisa.fadhilla@dsn.ubharajaya.ac.id

4rd Dwi Budi Srisulistiowati
Informatic, Faculty of Computer Science
Bhayangkara Jakarta Raya University
Bekasi, Indonesia
dwi.budi@dsn.ubharajaya.ac.id

5th Andy Achmad Hendarsetiawan
Informatic, Faculty of Computer Science
Bhayangkara Jakarta Raya University
Bekasi, Indonesia
andy.achmad@dsn.ubharajaya.ac.id

6nd Ajif Yunizar Pratama Yusuf
Informatic, Faculty of Computer Science
Bhayangkara Jakarta Raya University
Bekasi, Indonesia
ajif.yunizar@dsn.ubharajaya.ac.id

***Abstract*—Academic consultation activities between students and academic advisory lecturers very necessary to help students in carrying out activities lectures. Based on the transcripts of grades that have been obtained, many students do not choose the field that suits them their academic ability so that there is a lot of nonlinearity 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 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. As for 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 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. Algorithm application Decision Tree C4.5 on the decision making system in the field of field is successful run by producing a fairly good level of accuracy from the calculation results total data. The data used in this study is sampling data from several final year students in the Ubhara Jaya Informatics study program. Field decision-making system field as a result of research that has been done can be used as a recommendations for both the Informatics Study Program, Faculty of Computer Science UBJ and for final year students to direct the field of research in the final project.**

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 there are students who want to take courses that have 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 attribute that has the highest gain ratio 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]

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

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).

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

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).

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

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.

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

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.

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	Pemrograman Web	Pemrograman Mobile	Pemrograman Berbasis Framework	Proyek Aplikasi
Network and Infrastructure	Jaringan Komputer II	Keamanan Jaringan Komputer dan Sistem Informasi	Komputer Forensik	Sistem Terintegrasi
Data Science	Statistika & Probabilitas	Sistem Cerdas	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, PA: Application Projects, 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	PA	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	PA	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

	Courses

Value Weight	PW	PM	PBF	PA	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)

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

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

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 (*S1*)	JI (*S2*)	DS (*S3*)	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	
		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					
	Pemrograman berbasis Framework (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					
	Proyek Aplikasi (PA)	High	27	15	7	5	1.426573285	0.10598214
		Medium	2	1	0	1	1	
		Low	1	0	1	0	0	
		Total	30					
	Jaringan Komputer (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					
	Keamanan Jaringan Komputer (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					
	Komputer Forensik (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					
	Sistem Terintegrasi (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					
	Statistika dan Probabilitas (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					
	Sistem Cerdas (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.

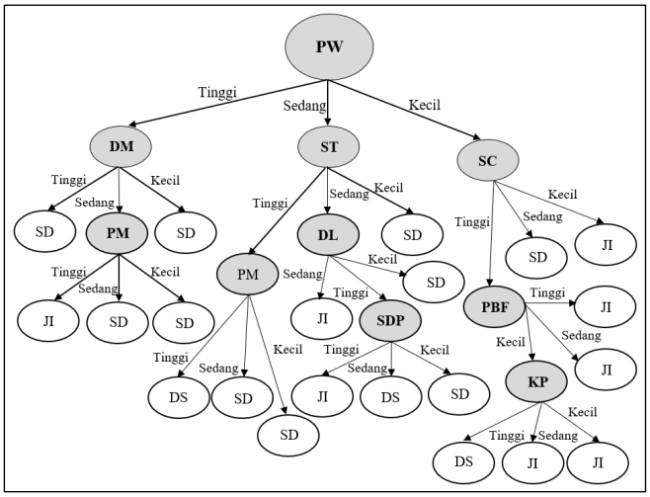


Fig. 1. Decision tree results

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 field 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. test data.

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

[1] Rafika Sari, Khairunnisa Fadhilla Ramdhania, and Rakhmat Purnomo, "Team-Teaching-Based Course Scheduling Using Genetic Algorithm," *Penelitian Ilmu Komputer, Sistem Embedded and Logic* , vol. 10, no. 1, pp. 55–66, Mar. 2022.

[2] S. Claudia and T. Sutrisno, "Analisis Rekomendasi Peminatan Menggunakan Metode Decision Tree Dengan Algotirma C4.5."

[3] K. Adhatrao, A. Gaykar, A. Dhawan, R. Jha, and V. Honrao, "Predicting Students' Performance Using ID3 and C4.5 Classification Algorithms," *International Journal of Data Mining & Knowledge Management Process*, vol. 3, no. 5, pp. 39–52, Sep. 2013, doi: 10.5121/ijdkp.2013.3504.

[4] A. Wildan and R. Sari, "Metode Certainty Factor pada Sistem Pakar Identifikasi Penyakit Mental," 2022. [Online]. Available: http://ejurnal.ubharajaya.ac.id/index.php/jiforty

[5] L. Swastina, "Penerapan Algoritma C4.5 Untuk Penentuan Jurusan Mahasiswa," 2013.

[6] B. Sudrajat, "Penggunaan Algoritma C4.5 Untuk Menentukan Kepuasan Pelanggan Pada Warnet Game Victory," *Jurnal Ilmu Teknik dan Komputer*, vol. 6, no. 1, 2022.

[7] A. R. Sukma, R. Halfis, and A. Hermawan, "Klasifikasi Channel Youtube Indonesia Menggunakan Algoritma C4.5," *Jurnal Teknik Komputer AMIK BSI*, vol. 5, no. 1, pp. 21–28, Feb. 2019, doi: 10.31294/jtk.v4i2.

[8] Y. Wulandari, E. Haerani, S. Kurnia Gusti, and D. S. Ramadhani, "Klasifikasi Berita Menggunakan Algoritma C4.5," *Jurnal Nasional Komputasi dan Teknologi Informasi*, vol. 5, no. 2, 2022.

[9] P. V. Ngoc, C. V. T. Ngoc, T. V. T. Ngoc, and D. N. Duy, "A C4.5 algorithm for english emotional classification," *Evolving Systems*, vol. 10, no. 3, pp. 425–451, Sep. 2019, doi: 10.1007/s12530-017-9180-1.

[10] X. Meng, P. Zhang, Y. Xu, and H. Xie, "Construction of decision tree based on C4.5 algorithm for online voltage stability assessment," *International Journal of Electrical Power & Energy Systems*, vol. 118, p. 105793, Jun. 2020, doi: 10.1016/J.IJEPES.2019.105793.

[11] L. Marlina and A. Putera Utama Siahaan, "Data Mining Classification Comparison (Naïve Bayes and C4.5 Algorithms)," *International Journal of Engineering Trends and Technology*, vol. 38, no. 7, 2016, [Online]. Available: http://www.ijettjournal.org

[12] W. Dai and W. Ji, "A mapreduce implementation of C4.5 decision tree algorithm," *International Journal of Database Theory and Application*, vol. 7, no. 1, pp. 49–60, Jan. 2014, doi: 10.14257/ijdta.2014.7.1.05.

[13] E. Kretschmann, W. Fleischmann, and R. Apweiler, "Automatic rule generation for protein annotation with the C4.5 data mining algorithm applied on SWISS-PROT," 2001. [Online]. Available: http://www.cs.waikato.ac.nz/

[14] J. S. Lee, "AUC4.5: AUC-Based C4.5 Decision Tree Algorithm for Imbalanced Data Classification," *IEEE Access*, vol. 7, pp. 106034–106042, 2019, doi: 10.1109/ACCESS.2019.2931865.

[15] A. Cherfi, K. Nouira, and A. Ferchichi, "Very Fast C4.5 Decision Tree Algorithm," *Applied Artificial Intelligence*, vol. 32, no. 2, pp. 119–137, Apr. 2018, doi: 10.1080/08839514.2018.1447479.

[16] S. J. Lee, Z. Xu, T. Li, and Y. Yang, "A novel bagging C4.5 algorithm based on wrapper feature selection for supporting wise clinical decision making," *J Biomed Inform*, vol. 78, pp. 144–155, Feb. 2018, doi: 10.1016/J.JBI.2017.11.005.

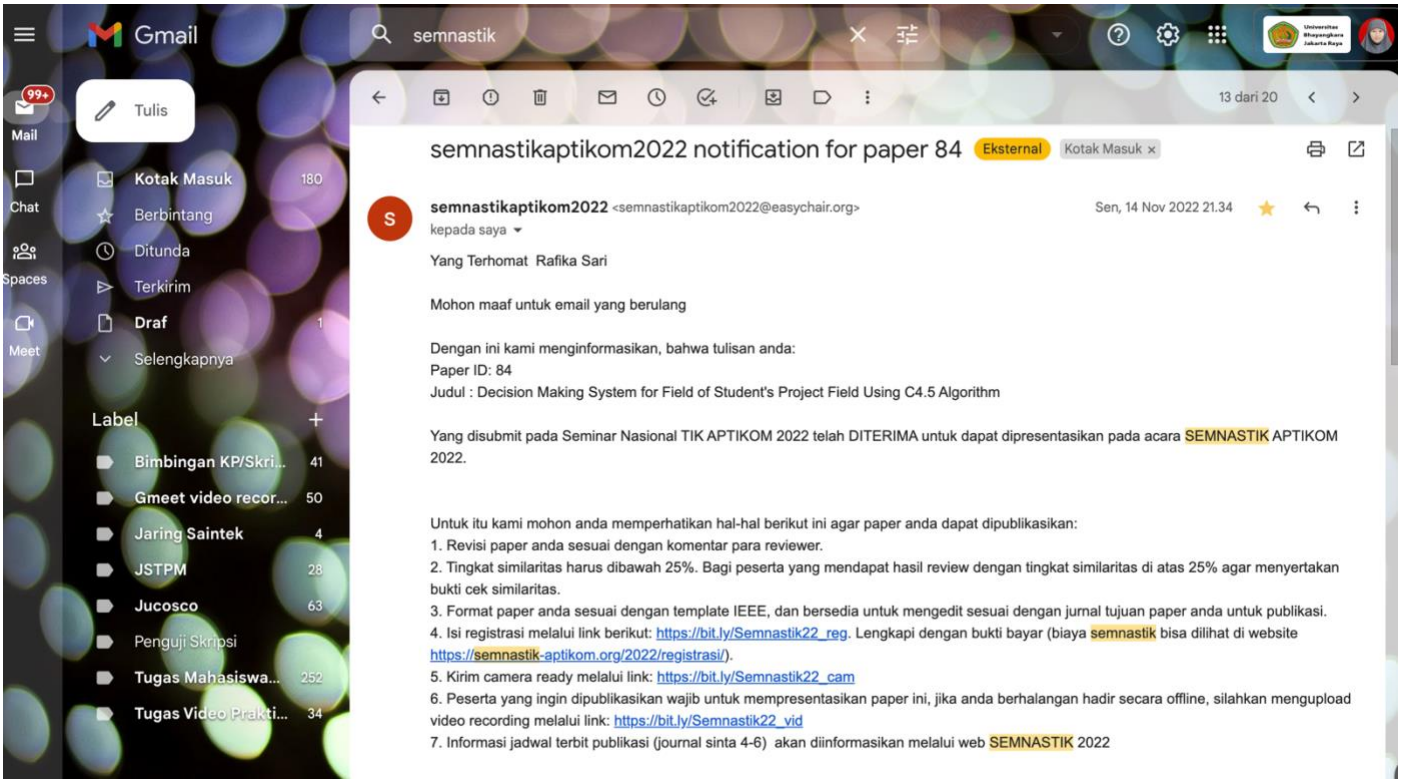
[17] M. M. Mazid, A. B. M. S. Ali, and K. S. Tickle, "Improved C4.5 Algorithm for Rule Based Classification."

[18] Yuli Mardi, "Data Mining: Klasifikasi Menggunakan Algoritma C4.5," *Jurnal Edik Informatika*, vol. 2, no. 2, pp. 213–219.

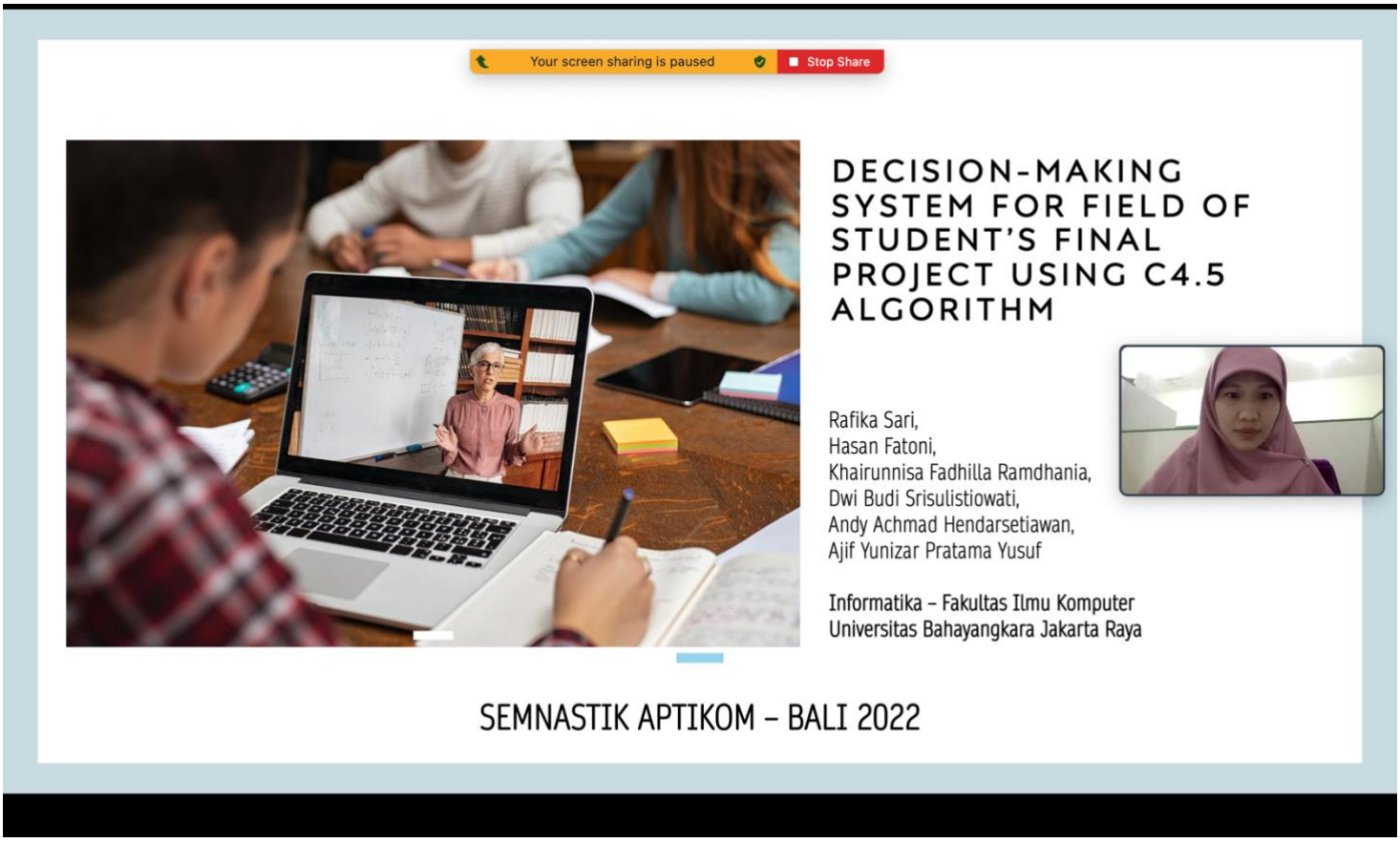
[19] M. A. Muslim, S. H. Rukmana, E. Sugiharti, B. Prasetiyo, and S. Alimah, "Optimization of C4.5 algorithm-based particle swarm optimization for breast cancer diagnosis," in *Journal of Physics: Conference Series*, Apr. 2018, vol. 983, no. 1. doi: 10.1088/1742-6596/983/1/012063.

[20] X. Zheng, W. Feng, M. Huang, and S. Feng, "Optimization of PBFT Algorithm Based on Improved C4.5," *Math Probl Eng*, vol. 2021, 2021, doi: 10.1155/2021/5542078.

1. **Bukti *Acceptence* naskah publikasi SEMNASTIK**



2. **Bukti Presentasi Pada SEMNASTIK secara *hybrid***



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.

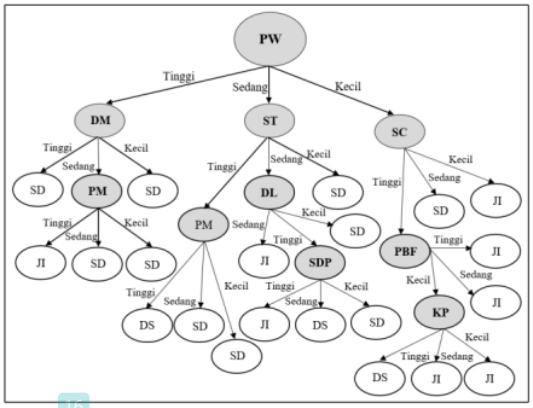


Fig. 1. Decision tree results

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