JUMANJI
ISSN (p): 2598-8050
ISSN (e): 2598-8069
| Vol. 7 | No. 1 | Halaman 64 - 75
April 2023

Student Graduation Prediction System Based on Academic and Nonacademic (EQ) Data Using C4.5 Algorithm

Willy Hanafi [1)] Yulison Herry Chrisnanto [2)] Ade Kania Ningsih [3)]

[1,2,3] Department of Informatics Universitas Jenderal Achmad Yani, Cimahi, Indonesia

Email : willyhanafi17@if.unjani.ac.id

ABSTRACT

The graduation profile is an important element for higher education accreditation standards. It reflects the performance of the adopted education system within a certain period. The better the profile graduation, the better the value of the accreditation. Some students are unable to complete their studies on time or even fail to complete their studies because they exceed the specified time limit, which is seven years, and it negatively affects institutions' accreditation. To prevent this from happening, it is necessary to know what obstacles that cause these students could not complete their studies on time. by knowing this information, prevention can be done for students who are potentially unable to complete their studies on time. The purpose of this study was to make a system that can predict the graduation timeline and the factors that influence it. The data used was graduation data from undergraduate students majoring in psychology from 2015 to 2017 at a university in Cimahi. The data had a total record of 461 students, 44 subject value attributes, 13 psychotest attributes, and class attributes. We generated the result by using decision tree method with C4.5 algorithm, which produces 90.32% accuracy. The depth of the tree can also influence the accuracy of the algorithm. This study also found that academic and non-academic (EQ) scores can affect students' graduation time.

Keywords: *C4.5 Algorithm, Graduation On Time, EQ, Decision Tree.*

1. INTRODUCTION

Graduation is a crucial aspect of any educational institution and is highly emphasized by the national standard NSPT PERMENDIKBUD DIKTI no 3 of 2020. Timely graduation not only aids in evaluating an institution's accreditation but also plays a significant role in its overall effectiveness. As a result, colleges strive to make every effort possible to ensure that their students graduate on time. The value of graduation cannot be overstated, and institutions must prioritize this policy issue to maintain their accreditation status[1].

Efforts continue to be made to predict the students' graduation so that they can increase the average graduation at a college institution. The result of predicting graduation on time can provide opportunities for college institutions to minimize student laziness in completing their studies[2] Data mining can also be used to predict the early stages of student dropout, so institutions can identify students who will drop out and can prevent it by providing counseling for them so they can change their minds from the dropout and graduating from college institutions [3].

The decision tree method is one of the popular methods of classification. The decision tree is a classification that generates trees and rules from a dataset that has classes. It has a simple and easy-to-understand structure that shows the relationship between the dataset and the decision[4]. Decision tree can make a model and accurately predict graduate time students [5], and detect a factor that is linked to academic achievement [6]. So, that decision tree is a promising method to use in educational data mining.

According to various research studies, academic and non-academic factors are positively linked[7]. One of the significant non-academic aspects in the education domain is Emotional Quotient (EQ)[8]. Emotional intelligence encompasses the ability to recognize, comprehend, and positively utilize emotions to manage stress, communicate effectively, empathize, handle challenges, and find solutions[9]. Numerous studies have shown that emotional intelligence is a valuable asset in various fields. Some researchers suggest that it is just as important as Intelligence Quotient (IQ), and at times, even more, essential than IQ itself[10].

Gondal and Husain [7] researched that EQ is a major determinant of employee performance and EQ is considered more important than intelligence quotient (IQ) in the workplace. Most people with high IQs and excellent academic records are usually not very good at social interactions. This deficiency is caused by a lack of emotional intelligence, despite having a fairly high IQ. This does not mean that IQ is completely useless but their research shows that EQ has a more important role than IQ in increasing organizational effectiveness.

This study predicts graduation on time by combining data from academics with 44 variables and from psychotest with 13 variables that have EQ variables. This data was obtained from the Department of Psychology at a university in Cimahi. This research has produced a system that can predict graduate on-time students based on academics and non-academics (EQ) using the C4.5 decision tree algorithm

2. METHOD

This research was conducted through several stages. The research method in this study is shown in Figure 1.

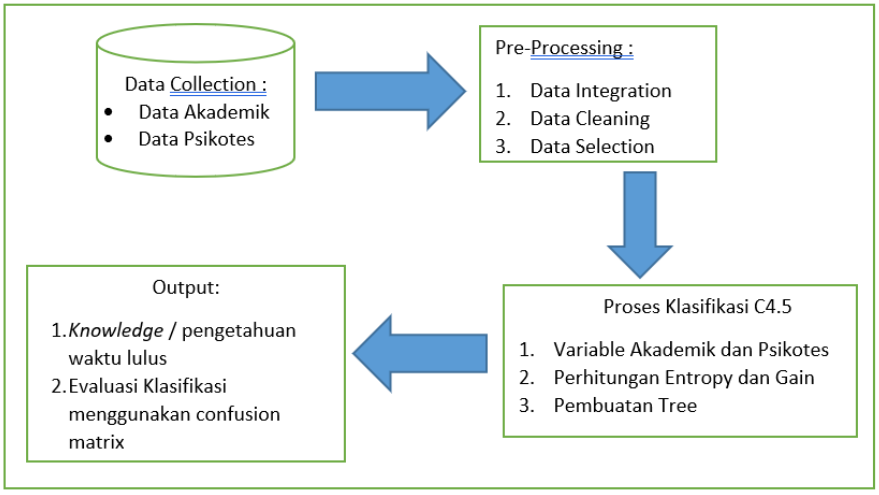

Figure 1. Research Method

2.1. Data Collection

In this research, the graduation record of the student and the psychotest record were obtained from the Department of Psychology at a university in Cimahi. The data used in this research is graduation data that contain attributes of 44 course credits multiplied by the grade score. There's also a psychotest record that consists of 13 attributes in which there is EQ data from the student. The data of the psychotest can be seen in Table 1.

Table 1. Data psychogram

No	PSIKOGRAM												
	INTELIGENT ASPECT						EMOTIONAL ASPECT						
	KI	AS	BP	DO	DB	KB	KS	PD	C	TS	SE	MB	MN
1	4	4	4	4	4	3	3	3	3	4	3	4	2
2	4	3	3	4	4	3	3	3	3	4	2	4	3
3	4	3	3	4	4	3	3	3	3	4	3	3	3
4	4	4	4	4	4	3	3	3	3	3	3	4	3
5	4	3	4	3	4	4	3	3	4	2	3	3	2
6	4	4	4	3	3	3	3	4	4	3	3	4	2
7	4	4	4	3	4	4	3	3	4	3	3	3	2
8	3	3	3	3	3	3	3	3	3	3	2	2	3
9	3	3	3	2	3	3	3	3	3	3	2	3	2
10	3	3	3	3	3	3	3	3	3	3	2	3	3
11	4	3	3	3	3	3	3	3	3	3	2	3	2
12	3	3	4	3	3	3	3	3	2	3	3	3	3
13	3	3	3	3	3	3	3	3	3	2	3	2	3
...	...	...	...	...	...	...	...	...	...	...	...	...	...
1303	1	2	1	2	2	2	3	2	2	2	1	1	3

The description of the data attributes in the table above explains the description of the data attributes that will be used in the psychological test data. 13 data attributes are used as a psychological test with information that can be seen in table 2 below.

Table 2. Description attribute psychotest that are used in this research

1	KI	Intellectual Capacity
2	AS	Sintesa Analytic
3	BP	Practical Thinking
4	DO	Organizational Capabilities
5	DB	Imagination Ability
6	KB	Language Skill
7	KS	Social Sensitivity
8	PD	Adjustment Ability
9	C	Carefulness
10	TS	Stress Tolerance
11	SE	Emotional Stability
12	MB	Achievment Motivation
13	MN	Interest

Table 3. Data Academic

No	NIM	Nama	IPK	Tahun Lulus	Pancasila	PKN	FAAL	...	Psikologi Sosial
1	7111121059	Pramudia Putra	2.86	September 2019	6	8	8	...	8
2	7111121101	Hafsah Hilaliyah	3.03	September 2019	8	8	6	...	8
3	7111121115	Velrini Safriyanti Edwar	2.71	September 2019	8	8	8	...	8
4	7111121181	Retno Kusumaningrum	3.03	September 2019	8	8	6	...	8
5	7111121186	Randi Eka Saputra	2.73	September 2019	8	8	8	...	8
6	7111121201	Inge Rahma Audina	2.74	September 2019	6	8	7	...	8
7	7111131002	Rizky Maulana Yusali	2.52	Maret 2021	8	8	8	...	8
8	7111131017	Sherly Syafria Islami	3.05	September 2020	7	8	5	...	8
9	7111131065	Feisal Hilman Amirullah	3.00	Maret 2021	7	7	8	...	8
10	7111131097	Satria Budi Pratama	2.94	Maret 2021	8	8	8	...	8
11	7111131127	P.A Wibowo Catur Yugo	2.76	September 2019	5	7	5	...	8

12	7111131128	Dea Intan Triana	2.60	September 2019	6	7	5	…	6
13	7111141009	M. Noviyanto Margono	3.04	September 2020	7	8	8	…	7
14	7111141037	Adhitya Eka Damara	2.72	September 2019	8	8	8	…	6
15	7111141047	Melvan Nugraha Laksana	2.79	Maret 2021	8	8	6	…	7
16	7111141049	Risma Fitri Selfira	3.14	September 2019	6	6	4	…	7
17	7111141050	A.D Putra Syarifudin	2.46	Oktober 2021	6	7	5	…	6
…	…	…	…	…	…	…	…	…	…
…	…	…	…	…	…	…	…	…	…
503	7111171225	Moch Andi Nugroho	3.03	Oktober 2021	7	6	4	…	8

In the Academic Data, there are attributes that contain NIM, Name, GPA, Year of graduation and course grades of 44 attributes, so there are a total of 48 attributes.

2.2. Preprocessing Data

1. Data Integration: The preprocessing stage is the stage of changing the data into an alignment. At the data pre-processing stage, the raw data is carried out by the data integration process, which combines data from 2 different sources, which is data psychotest and data academic, and data integration functions to increase the value of accuracy.

Table 4. Data Integration

No	N I M	Nama	IPK	Tahun Lulus	KI	AS	BP	…	Psikologi Sosial
1	7111151001	Desy Wijayanti	3.45	September 2020	4	3	4	…	7
2	7111151003	Hardini Kartikawati	2.78	Mei 2020	3	2	2	…	6
3	7111151004	Annisa Nurasri	2.97	Mei 2020	4	3	3	…	6
4	7111151019	Tiyana Ajeng Indah Sari	3.22	September 2020	2	2	2	…	6
5	7111151024	Trixie Diulva Septarina	3.32	September 2020	3	3	3	…	7
6	7111151026	Gandrik Pangestu	3.35	Mei 2020	3	3	3	…	5
7	7111151032	Alfia Luthfi Handini	3.08	Mei 2020	3	3	2	…	8
8	7111151036	Chevina Janvierta Ramadhani	3.04	Mei 2020	3	3	3	…	6
9	7111151038	Septiarini Rahmadina	2.91	Mei 2020	1	2	2	…	6
10	7111151039	Febrian Widiyanto	2.75	Maret 2022	2	2	3	…	7
11	7111151055	Bagus Fiyan Abdillah Al Ghani	3.27	Mei 2020	3	3	3	…	6

12	7111151057	Aprilia Rizki Maharani	3.43	Mei 2020	4	3	4	...	7
13	7111151058	I Gusti Ayu Amritha Sari	2.65	Maret 2022	4	4	3	...	7
14	7111151060	Yoga Novi Ariyanto	2.72	Oktober 2021	3	4	4	...	7
15	7111151061	Dicky Permana Haryadi	2.82	Oktober 2021	4	3	4	...	5
16	7111151075	Melynda Fatwa	2.91	Maret 2022	3	3	3	...	6
17	7111151081	Fauzan Dimas Salim	2.76	September 2020	2	2	2	...	6
18	7111151082	Achmad Fahri Husaini	3.01	September 2020	3	3	2	...	6
...	...	...	...	...	...	...	...	...	...
503	7111151088	Yusallina Azrani	2.69	Mei 2020	3	2	3	...	5

2. Data Cleaning: After integrating the graduation data with the psychological test data, then cleaning the data is carried out and eliminating missing value data because clean data from missing values results in a good classification

Table 5. Data Cleaning

7111171093	Alfiany Tanisha Fadhilah	3.34	Oktober 2021	5	4	3	4	3
7111171020	Alfira Yuanita Indradi	3.38	Oktober 2021	Null	Null	Null	Null	Null
7111171190	Alifvia Septiani Putri Iskandar	3.43	Oktober 2021	3	2	3	3	3
7111171095	Aliya Haniifah Nurdevita	3.36	Oktober 2021	**4**	**3**	**3**	**4**	**4**
7111171163	Alnida Destiana Nishfathul Medynna	3.66	Oktober 2021	5	4	5	4	3
7111171085	Alya Nufus Shafira	3.22	Oktober 2021	3	3	3	3	3
7111171216	Ananta Prayoga	3.25	Oktober 2021	3	2	2	2	1
7111171106	Anisa Nurul Latifa	3.15	Oktober 2021	4	3	2	4	4
7111171094	Anna Kurniati	3.49	Oktober 2021	3	2	3	3	3
7111171014	Annisa Iscmah Nur'Aini	3.28	Oktober 2021	Null	Null	Null	Null	Null
7111171078	Antika Rizki Pratiwi	3.4	Oktober 2021	**3**	**3**	**3**	**2**	**2**
7111171022	Ardita Rizky Agustien	3.5	Oktober 2021	Null	Null	Null	Null	Null
7111171003	Arfhi Listiadewi	3.34	Oktober 2021	Null	Null	Null	Null	Null

3. Data Selection: The data selection stage is the addition and removal of several variables. The added variable is the graduate time. The graduate time variable is obtained based on the time gap between the student batch and the time of graduation, students whose study duration is 4 years can be said to have graduated on time and if it exceeds 4 years it is stated that they have not graduated on time. The omitted variables are NIM, name, and year of student graduation. So, this research has 44 attributes subject scores,13 psychotest attribute scores, and class.

Table 5. Data Cleaning

Pancasila	FAAL	B. Inggris	B indonesia	Agama	Sosiologi	Filsafat Umum	Pengantar Psikologi	Psikologi Fungsi Dasar
8	6	7	6	5	7	7	8	6
8	5	6	6	7	6	7	12	5
8	4	5	7	4	7	8	8	4
5	4	6	7	4	8	5	10	4
7	4	8	7	4	5	5	10	4
6	5	8	7	5	6	5	8	5
8	5	7	7	4	4	5	10	5
6	5	8	7	2	4	5	8	5
7	4	7	7	4	7	5	8	4
6	7	8	7	5	5	5	10	7
7	4	4	7	4	5	5	8	4
7	4	5	7	5	4	7	12	4
8	5	4	7	4	7	7	8	5
8	4	5	7	7	5	7	10	4

2.3. Decision Tree Implementation

The decision tree method is a popular data mining technique that involves creating a tree-like model to classify data. One of the commonly used algorithms in the decision tree method is C4.5, which is an improvement over its predecessor, ID3[11]. To construct a decision tree using the C4.5 algorithm, the first step is to calculate the entropy and gain values for each attribute to determine the one with the highest gain value. Entropy is calculated by summing the results of all the grades on an attribute. Once the attribute with the highest gain value is identified, it is used as the initial root, and the process is repeated to identify sub-roots until all the attributes are partitioned or used as sub-roots in the decision tree[12]. The calculation of entropy and gain is indicated by the formula.

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

Description:
S: a set of cases
n: the number of partitions S
pi: the proportion of Si to S

$$Gain\ (S, A) = Entropy(S) - \sum_{i=0}^{n} \quad \frac{Si}{S}\ Entropy\ (Si) \tag{2}$$

Description:
S = set of cases
A = attributes
n = number of attribute partitions S
|Si| = proportion Si to S
|S| = case number in S

The formula for calculating entropy, as shown above, involves computing the sum of grades for a given attribute. Once the entropy value is obtained, the gain calculation formula is then applied to determine the attribute with the highest gain value. This process is repeated for each attribute until the one with the highest gain is identified as the initial root for the decision tree. The algorithm then proceeds to compute the sub-roots and partitions until all attributes are used or have become sub-roots in the decision tree. The resulting decision tree can then be used to classify new data based on the learned patterns.

3. RESULT AND DISCUSSION

During the data mining process, classification is performed on the training data to generate classification rules that describe the patterns present in the dataset. These rules can be used to make predictions about new data based on the learned patterns. Figure 2 displays the rule model that represents students who graduate on time, providing a visual representation of the classification rules that have been extracted from the data.

```
|--- teori psikologi sosial <= 6.50
|   |--- Psikologi Fungsi Dasar <= 5.50
|   |   |--- Biopsikologi <= 6.50
|   |   |   |--- Psikologi personel <= 5.50
|   |   |   |   |--- MN <= 2.50
|   |   |   |   |   |--- Biopsikologi <= 5.50
|   |   |   |   |   |   |--- class: 0
|   |   |   |   |   |--- Biopsikologi >  5.50
|   |   |   |   |   |   |--- class: 1
|   |   |   |   |--- MN >  2.50
|   |   |   |   |   |--- TS <= 3.50
|   |   |   |   |   |   |--- class: 0
|   |   |   |--- Psikologi personel >  5.50
|   |   |   |   |--- psikologi pendidikan <= 2.00
|   |   |   |   |   |--- psikologi kognitif <= 7.00
|   |   |   |   |   |   |--- class: 0
|   |   |   |   |   |--- psikologi kognitif >  7.00
|   |   |   |   |   |   |--- class: 1
|   |   |   |   |--- psikologi pendidikan >  2.00
|   |   |   |   |   |--- class: 1
|   |   |--- Biopsikologi >  6.50
|   |   |   |--- Statitiska <= 6.50
|   |   |   |   |--- Asas Manajemen <= 5.50
|   |   |   |   |   |--- class: 1
|   |   |   |   |--- Asas Manajemen >  5.50
|   |   |   |   |   |--- KKN <= 4.50
|   |   |   |   |   |   |--- class: 1
|   |   |   |   |   |--- KKN >  4.50
|   |   |   |   |   |   |--- class: 0
|   |   |   |--- Statitiska >  6.50
|   |   |   |   |--- class: 1
|   |--- Psikologi Fungsi Dasar >  5.50
|   |   |--- class: 1
|--- teori psikologi sosial >  6.50
|   |--- Biopsikologi <= 6.50
|   |   |--- observasi dalam psikologi <= 4.50
|   |   |   |--- KS <= 2.50
|   |   |   |   |--- class: 0
|   |   |   |--- KS >  2.50
|   |   |   |   |--- TS <= 3.50
|   |   |   |   |   |--- class: 1
|   |   |   |   |--- TS >  3.50
|   |   |   |   |   |--- class: 0
|   |   |--- observasi dalam psikologi >  4.50
|   |   |   |--- masalah umum dan gangguan anak <= 7.00
|   |   |   |   |--- class: 1
|   |   |   |--- masalah umum dan gangguan anak >  7.00
|   |   |   |   |--- TS <= 2.50
|   |   |   |   |   |--- class: 0
|   |   |   |   |--- TS >  2.50
|   |   |   |   |   |--- class: 1
|   |--- Biopsikologi >  6.50
|   |   |--- class: 1
```

Figure 2. Generated rule model using python

In data mining, the classification process involves analyzing the training data to generate a tree model and classification rules. This process helps us to understand the relationship between different attributes in the dataset. Using the information gathered during the model construction, we can form a tree model that represents the attribute relationships. This model provides a visual representation of how the attributes are related to each other and can be useful in identifying patterns and making predictions about new data. Figure 3 displays a tree model that has been generated through the classification process

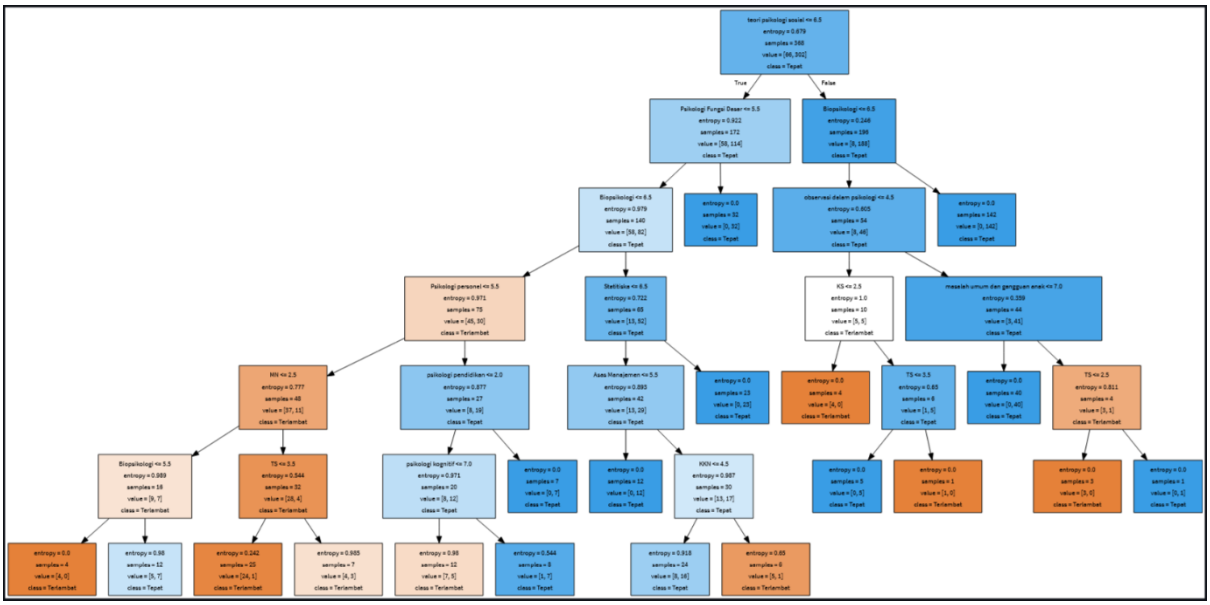

Figure 3. Generated tree model using python

By analyzing Figure 2 and Figure 3, we can see that the attribute "Teori Psikologi Sosial" is the root node that determines whether a student graduates on time. This is based on the calculation of the highest gain value for each attribute, which is used to construct the decision tree. As the algorithm progresses, additional nodes are added to the tree, with the "EQ" attribute being one of them like KS, MN and TS. The results of this study indicate that if someone has good stress tolerance (TS), social sensitivity (KS), and high interest (MN) in that field, it can affect the student's graduation on time. This suggests that while EQ may not be the main factor in determining graduation on time, it still plays a role in the overall outcome.

Furthermore, experiments conducted on this model have shown that the accuracy of the predictions can vary depending on the depth of the decision tree. The results are presented in Table 1, which provides a detailed breakdown of precision, recall, and F1-score. This information can be used to fine-tune the decision tree and improve its accuracy in predicting graduation outcomes.

Table 6. Precision, recall, f1-score, and accuracy at each depth tree.

Depth Tree	Precision	Recall	F1-Score	Accuracy
2	39.24%	50.00%	43.97%	78.49%

3	83.54%	85.89%	84.62%	89.24%
4	81.92%	85.20%	83.36%	88.17%
5	87.22%	82.94%	84.82%	90.32%
6	85.31%	86.57%	86.92%	90.32%
7	86.09%	84.76%	85.40%	90.32%

Based on the table above, it can be seen that the depth of the tree affects the accuracy of the method. and in this research, the best accuracy obtained is at depths 5, 6, and 7 and has an accuracy of 90.32%. Besides that, the graduation time of students is also affected by the EQ factor. can be seen in the rules and trees that have been done, that EQ can affect graduate on-time students.

4. CONCLUSION

In summary, this study has demonstrated that the decision tree method, specifically the C4.5 algorithm, can be used to predict whether a student will graduate on time based on both academic and non-academic data, such as EQ. The results of the testing phase showed that the decision tree method achieves an accuracy rate of 90.32%, indicating that it is a viable method for predicting graduation outcomes. Moreover, the research findings suggest that EQ is a significant factor that affects a student's ability to graduate on time. To improve the accuracy of the prediction results, it may be beneficial to optimize the method used or consider incorporating additional non-academic variables that are associated with timely graduation. Overall, the results of this study provide valuable insights into the factors that influence graduation outcomes and offer a framework for future research in this area.

REFERENCES

[1] A. Maulana. (2021). "Prediction of student graduation accuracy using decision tree with application of genetic algorithms." *IOP Conf Ser Mater Sci Eng*, 1073(1), 012055. https://doi.org/10.1088/1757-899x/1073/1/012055

[2] Gunawan, Hanes, & Catherine. (2021). "C4.5, K-Nearest Neighbor, Naïve Bayes and Random Forest Algorithms Comparison to Predict Students' On Time Graduation". *Indonesian Journal of Artificial Intelligence and Data Mining (IJAIDM)*, 4(2), 62–71. https://doi.org/10.24014/ijaidm .v4i2.10833

[3] v. Hegde, & P. P. Prageeth. (2018). "Higher Education Student Dropout Prediction and Analysis through Educational Data Mining." *Proceedings of the 2nd International Conference on Inventive Systems and Control*, 694–699. https://doi.org/10.1109/ICISC.2018.8398887

[4] Ks. Priya, & Avs. Kumar Director. (2013). Improving the Student's Performance Using Educational Data Mining". *International Journal of Advanced Networking and Applications*, 4(4), 1680–1685.

[5] D. A. Wulandari, B. N. Sari, & T. N. Padilah. (2022). "Prediction of Student Graduation Accuracy Using C45 Algorithm (Case Study: Fasilkom Unsika),." *SYSTEMATICS*, 4(1), 372.

[6] F. Martínez Abad, & A. A. Chaparro Caso López. (2017). "Data-mining techniques in detecting factors linked to academic achievement." *School Effectiveness and School Improvement*, 28(1), 39–55. https://doi.org/10.1080/09243453.2016.1235591

[7] U. H. Gondal, & T. Husain. (2013). "A Comparative Study of Intelligence Quotient and Emotional Intelligence: Effect on Employees? Performance". *Asian Journal of Business Management*, 5(1), 153–162. https://doi.org/10.19026/ajbm.5.5824

[8] W. Wardani. (2019). "Pengaruh kecerdasan adversitas dan kecerdasan emosional melalui model inkuiri sosial terhadap keterampilan sosial siswa 1." *Jurnal Teori Dan Praksis Pembelajaran IPS*, 4(2), 66–72. https://doi.org/10.17977/um022v4i22019p66

[9] A. S. Drigas, & C. Papoutsi. (2018). "A new layered model on emotional intelligence." *Behavioral Sciences*, 8(5). https://doi.org/10.3390/bs8050045

[10] M. A. Cohn, B. L. Fredrickson, S. L. Brown, J. A. Mikels, & A. M. Conway. (2009). "Happiness Unpacked: Positive Emotions Increase Life Satisfaction by Building Resilience." *Emotion*, 9(3), 361–368. https://doi.org/10.1037/a0015952

[11] M. Wahyudi, & A. Andriani. (2021). "Application of C4.5 and Naïve Bayes Algorithm for Detection of Potential Increased Case Fatality Rate Diarrhea,." *Journal of Physics: Conference Series*, 1830(1). https://doi.org/10.1088/1742-6596/1830/1/012016

[12] E. Purwanto. (2019). "PREDIKSI KELULUSAN TEPAT WAKTU MENGGUNAKAN METODE C4.5 DAN K-NN (Studi Kasus: Mahasiswa Program Studi S1 Ilmu Farmasi, Fakultas Farmasi, Universitas Muhammadiyah Purwokerto)." 20(2), 131–142.