JURNAL TEKNOLOGI DAN OPEN SOURCE
Vol. 8, No. 1, June 2025, pp. 27~34
e-ISSN: 2622-1659, accredited Four Grade by Kemenristekdikti, Decree No: 152/E/KPT/2023
DOI: 10.36378/jtos.v8i1.3997

Implementation of Data Mining in Measuring Student Satisfaction at IAIN Kerinci

Rian Rafiska[1]
[1]Ilmu Perpustakaan dan Informasi Islam, Institut Agama Islam Negeri Kerinci, Indonesia

Article Info

Article history:

Received 12 30, 2024
Revised 04 28, 2025
Accepted 05 10, 2025

Keywords:

Data Mining
Satisfaction
Students

ABSTRACT

Measuring student satisfaction is crucial, especially considering the increasing competition in the field of education along with the advancement of knowledge and technology. It is essential to assess whether the services expected by students align with the services they actually receive. Evaluating student satisfaction can significantly help higher education institutions improve service quality, which in turn may lead to an increase in student enrollment.This study employs a quantitative method using one of the data mining techniques—classification—through the C4.5 algorithm to measure student satisfaction levels. The population of this research includes active students at IAIN Kerinci, with a sample size of 100 respondents. The students serve as the subjects providing evaluations or opinions on variables characterized by Tangibles, Reliability, Responsiveness, Assurance, and Empathy.The data is processed using data mining classification techniques, with testing conducted through RapidMiner software. The results of the analysis and testing indicate that data mining effectively classifies the variables in measuring student satisfaction, generating 10 decision tree rules with an accuracy rate of 98.22%. These resulting rules are expected to serve as a foundation for making informed decisions on actions needed to enhance student satisfaction.

Corresponding Author:

Rian Rafiska
Ilmu Perpustakaan dan Informasi Islam
Institut Agama Islam Negeri Kerinci
Sungai Penuh, Indonesia
Email: rianrafiska@gmail.com

1. Introduction

In today's era of technological advancement, data has become a highly valuable commodity. Every sector, including education, must be able to utilize data effectively. Higher education institutions are required to manage their ever-growing volumes of data to continue evolving and providing optimal services. One technique that can be used to process large amounts of data is data mining. Linguistically, "data mining" can be interpreted as "mining data"—similar to how traditional mining extracts valuable minerals from the earth, data mining extracts valuable "information" from large datasets. Data mining refers to the discovery of knowledge in databases [1]. It is the process of extracting useful information from massive datasets using algorithms and extraction methods from statistics, machine learning, and database management systems [2]. Among the many data mining methods available, one of the most popular is the C4.5 algorithm [3]. This algorithm is commonly used to generate decision trees in classification techniques and is characterized by the calculation of entropy and gain values [4].

Student satisfaction plays a crucial role in the sustainability of academic systems in higher education institutions [5]. High student satisfaction with the academic services they receive can serve as an indicator of institutional quality and becomes a competitive advantage in attracting prospective students [6]. Satisfaction itself refers to a feeling that reflects the outcome of comparing a product or service with one's expectations [7]. It can also be described as a sense of contentment or pleasure derived from consuming a product or service [8].

Therefore, satisfaction can be measured based on the responses of service recipients—whether they feel satisfied with the quality of services provided. In this context, student satisfaction refers to how students perceive the academic services offered by the university [9]. Higher education institutions can meet students' expectations only if they understand what students truly want. Thus, evaluating student satisfaction with academic services is essential to help universities improve service quality and identify which types of services influence student satisfaction in the future. Providing quality services that align with student expectations also serves as a form of promotion, enhancing the institution's image and attracting more students [10].

Although research on student satisfaction and its impact on service quality has been widely conducted, most studies tend to examine general satisfaction factors without considering the specific characteristics of academic services in higher education. Moreover, the application of data mining techniques, such as the C4.5 algorithm, in analyzing student satisfaction is still relatively rare. Most existing studies employ descriptive statistics or basic regression methods, whereas data mining has the potential to yield deeper insights by uncovering hidden patterns that conventional analysis might overlook.

Therefore, this study aims to address this gap in the literature by applying the C4.5 algorithm to analyze specific variables related to student satisfaction—namely, Tangibles, Reliability, Responsiveness, Assurance, and Empathy [11]. This approach allows for more accurate identification of which variables significantly influence student satisfaction, providing data-driven recommendations for improving academic services. The findings of this study are expected to offer deeper understanding of the factors affecting student satisfaction in higher education and contribute to the advancement of data mining applications in the context of educational research.

2. Research Method

This research adopts a quantitative approach integrated with data mining techniques, specifically utilizing the C4.5 classification algorithm, to assess student satisfaction levels at IAIN Kerinci. The methodology commences with the collection of primary data through structured questionnaires, which are meticulously designed to evaluate five key dimensions of academic service quality: Tangibles, Reliability, Responsiveness, Assurance, and Empathy. These dimensions are adapted from the SERVQUAL framework, which is widely recognized for assessing service quality in educational and other service-oriented sectors. The deployment of questionnaires enables direct engagement with students as primary respondents, ensuring the acquisition of relevant and measurable data that reflect their personal experiences and perceptions.

The target population for this study encompasses all currently enrolled students at IAIN Kerinci who have actively interacted with various academic service systems, including but not limited to administrative services, access to academic information, and utilization of educational facilities. These students are considered ideal participants as they serve as the end users of campus services and thus can provide authentic insights into service quality. Given the size and diversity of the student body, the total population is estimated to range from several hundred to several thousand individuals, subject to the latest institutional enrollment records.

To ensure the relevance and validity of the data collected in this study, a purposive sampling technique is employed, involving the deliberate selection of 100 respondents who meet predefined criteria. This non-probability sampling method is chosen specifically to focus on students who have had direct and meaningful interactions with the academic services provided by the institution. By targeting individuals with actual experience in areas such as administrative support, academic advising, and the use of campus facilities, the study is better positioned to collect accurate, context-specific data that truly reflects the students' perspectives on service quality and satisfaction.

The selection process was conducted through a multi-faceted recruitment strategy, including in-class announcements, digital outreach via university platforms, and direct engagement at service locations such as the academic administration office and university library. This combination of methods was strategically designed to include a diverse group of students—covering different departments, years of study, and frequencies of service utilization. This diversity is essential to achieving a comprehensive understanding of student satisfaction, as it allows the study to account for various expectations and experiences across the academic community.

Moreover, the research process was carried out in a structured and systematic manner, beginning from instrument design and data collection to classification and analysis using the C4.5 algorithm. This methodical approach not only increases the credibility of the study but also provides a clear methodological framework for future research efforts within similar academic environments.

Ultimately, the insights derived from this research are intended to serve as a valuable resource for institutional decision-makers. By identifying which factors most significantly influence student satisfaction, the university can develop targeted improvements in academic services—thereby enhancing the student experience, promoting retention, and strengthening the institution's reputation in the long term.

3. Result and Discussion

The low rate of student enrollment growth poses a significant challenge for the institution in determining effective strategies to increase the number of students substantially. Despite numerous efforts—such as external promotional campaigns—the impact has remained limited. This condition underscores the necessity for the institution to implement more strategic and evidence-based approaches, one of which is measuring student satisfaction, an initiative that, notably, has not been previously undertaken.

The primary objective of assessing student satisfaction is to determine whether the services provided by the institution align with the expectations of its students. This is particularly important in light of the assumption that academic services and student experiences function as the most effective form of promotion—more impactful than external advertising—because word-of-mouth and internal reputation often influence prospective students' decisions. Additionally, through this evaluation, the institution can identify which service areas most significantly affect student satisfaction, allowing it to focus on improving those specific aspects to enhance overall student experiences.

This study is designed to provide comprehensive information on the key criteria that influence student satisfaction with both the facilities and the academic services offered by the institution. The input for this research is derived from questionnaire responses collected from students, while the output is a classification of the students' satisfaction levels into two categories: Satisfied and Not Satisfied. By analyzing these outputs, the institution gains insight into the underlying factors that determine whether students are content with their campus experience.

To process the data and extract meaningful patterns, this research utilizes data mining techniques, specifically the classification method, which is suitable for handling categorical variables. The classification process will help uncover hidden patterns within the data and generate actionable insights that can inform institutional strategies. By applying the C4.5 algorithm, the study is expected to produce a set of decision rules that clearly outline the conditions under which students report satisfaction or dissatisfaction, thus serving as a valuable decision-support tool for institutional planning and service enhancement.

a. Data Selection, which involves integrating data from multiple sources.
b. Data Cleaning, which aims to remove inconsistent and noisy data.
c. Data Transformation, which transforms the data into a suitable format for mining.
d. Data Mining, the essential process where intelligent methods are applied to extract data patterns, followed by Pattern Evaluation to identify truly interesting patterns that represent valuable knowledge based on specific interestingness measures.

In this study, five criteria were established to measure student satisfaction levels, namely: Tangible, Assurance, Reliability, Responsiveness, and Empathy. Once all the criteria were classified, the data collected from student questionnaires were then transformed in accordance with the classification results for each criterion.

Based on the transformed data, a decision tree will be constructed to evaluate student satisfaction levels using the predefined criteria. The process of building a decision tree using the C4.5 algorithm involves the following four steps [13]:

a. Preparing the training data. The training data is obtained from previously collected data that has been grouped into specific classes, usually referred to as historical data.
b. Calculating the root of the tree. The root is selected based on the gain value of each attribute, where the attribute with the highest gain is used as the first root. To obtain the gain value, the entropy of each attribute is calculated first using the appropriate formula.:

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

Explanation of the formula:
S = Set of cases
n = Number of partitions in set S
pi = Proportion of subset Si to the total set S

After obtaining the entropy value, the next step is to calculate the gain using the following formula :

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

Explanation of the formula:

S = Set of cases
A = Attribute (Feature)
n = Number of partitions of attribute A
Si| = Number of cases in the i-th partition
|S| = Total number of cases in set S

Repeat the previous process for each branch until all cases within a branch belong to the same class. This recursive process is essential in constructing an effective and accurate decision tree, where each branch represents a pathway of decisions based on specific attribute values, and each leaf node corresponds to a final classification—either “Satisfied” or “Not Satisfied.”

To build the decision tree, several critical steps must be carried out systematically. First, the total number of cases is identified, followed by determining how many of those cases fall under the “Satisfied” decision class (S1) and how many are categorized as “Not Satisfied” (S2). Then, the dataset is partitioned based on the five key attributes: Tangibles, Assurance, Reliability, Responsiveness, and Empathy, which represent the main dimensions of service quality evaluated in the study.

For each attribute, entropy values are calculated to measure the level of impurity or disorder in the dataset. Using these entropy values, the information gain is then computed to identify how much each attribute contributes to improving classification accuracy. The attribute with the highest information gain is selected as the root node of the decision tree. This ensures that the most informative attribute is prioritized, forming the basis for a more efficient and meaningful classification structure.

Table 1. Highest Gain Calculation

Node 1	Total	Satisfied	Not Satisfied	Entrophy	Gain
Total	100	82	18	0,6801	
Tangible					0,2894
High	36	36	0	0,0000	
Medium	48	42	6	0,5436	
Low	16	4	12	0,8113	
Assurance					0,3761
High	52	52	0	0,0000	
Medium	39	30	9	0,7793	
Low	9	0	9	0,0000	
Reliability					**0,4771**
High	48	48	0	0,0000	
Medium	36	33	3	0,4138	
Low	16	1	15	0,3373	
Responsiveness					0,4256
High	53	53	0	0,0000	
Medium	33	28	5	0,6136	
Low	14	1	13	0,3712	
Empathy					0,3863
High	48	48	0	0,0000	
Medium	35	31	4	0,5127	
Low	17	3	14	0,6723	

Based on the calculations in the table above, it can be seen that the attribute with the highest gain is Reliability, with a gain value of 0.4771. Therefore, this attribute is selected as the root node (Node 1). The values Medium and Low under this attribute still require further calculations.

The resulting decision tree can be seen in the following figure.

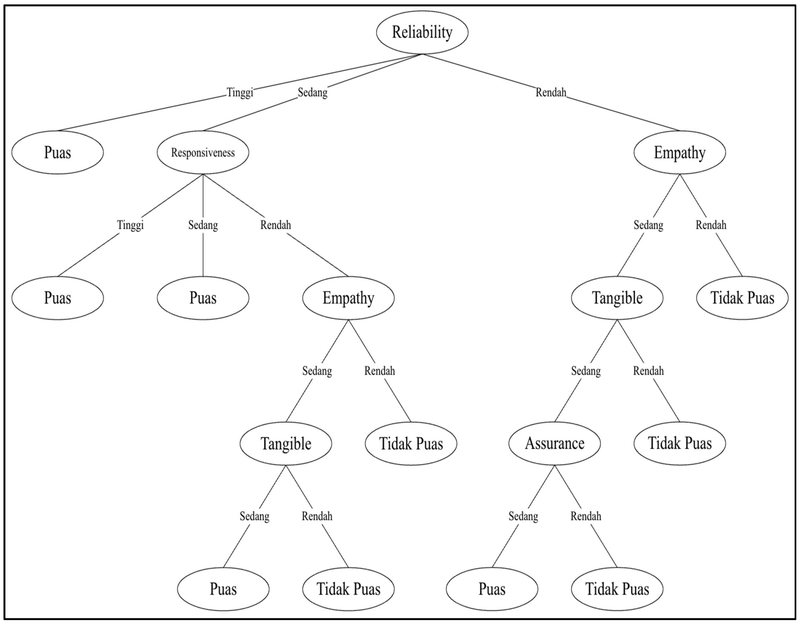

Figure 1. The Generated Decision Tree

The rules derived from the final decision tree are based on the calculations of Entropy and Gain. From this decision tree, a total of ten (10) rules have been identified to assess student satisfaction. These rules serve as logical pathways that map specific attribute conditions to satisfaction outcomes. Each rule is formed by tracing a unique path from the root node to a terminal leaf node in the decision tree, representing a specific classification outcome—either “Satisfied” or “Not Satisfied.” These decision rules can provide valuable insights for institutions to understand which factors most significantly influence student satisfaction, and to what extent certain combinations of service attributes affect student perceptions. The ten derived rules for measuring student satisfaction are as follows:

1. IF Reliability = High THEN Decision = Satisfied
2. IF Reliability = Medium AND Responsiveness = High THEN Decision = Satisfied
3. IF Reliability = Medium AND Responsiveness = Medium THEN Decision = Satisfied
4. IF Reliability = Medium AND Responsiveness = Low AND Empathy = Medium AND Tangibles = Medium THEN Decision = Satisfied
5. IF Reliability = Medium AND Responsiveness = Low AND Empathy = Medium AND Tangibles = Low THEN Decision = Not Satisfied
6. IF Reliability = Medium AND Responsiveness = Low AND Empathy = Low THEN Decision = Not Satisfied
7. IF Reliability = Low AND Empathy = Medium AND Tangibles = Medium AND Assurance = Medium THEN Decision = Satisfied
8. IF Reliability = Low AND Empathy = Medium AND Tangibles = Medium AND Assurance = Low THEN Decision = Not Satisfied
9. IF Reliability = Low AND Empathy = Medium AND Tangibles = Low THEN Decision = Not Satisfied

10. IF Reliability = Low AND Empathy = Low THEN Decision = Not Satisfied

From the ten decision rules generated through the C4.5 algorithm, it can be observed that five rules lead to a Satisfied decision, while the remaining five result in a Not Satisfied decision. This balanced distribution reflects the diversity of student perceptions regarding the quality of academic services, which is influenced by various combinations of the evaluated criteria: Tangibles, Assurance, Reliability, Responsiveness, and Empathy.

To validate the accuracy and reliability of the manually derived decision rules, it is essential to conduct testing and verification using a data mining tool. This ensures that the analytical results are consistent and align with expected outcomes. In this study, the validation process was carried out using RapidMiner Studio, a robust data mining software capable of performing classification analysis using the C4.5 algorithm.

The primary objective of this testing phase is to confirm whether the manual calculation and rule derivation match the outcomes generated by the software. Such testing is crucial for identifying any discrepancies, reinforcing the credibility of the analysis, and ensuring that the model performs as intended when applied to real data.

The results of the testing conducted through RapidMiner Studio, including the decision tree visualization and classification outcomes, are presented in the following figure :

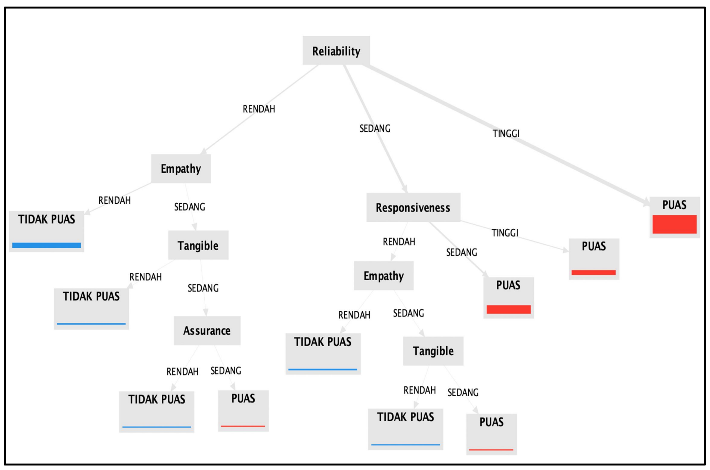

Figure 2. Decision Tree in RapidMiner

Based on the results of the data mining tests using the C4.5 algorithm, both through manual calculation and testing with RapidMiner Studio, it can be concluded that the resulting decision trees and rules from both methods are highly consistent and reliable, as the rules produced are identical. This consistency between manual and automated results validates the robustness of the methodology and strengthens the credibility of the findings. The C4.5 algorithm is considered to be highly effective for data classification because it clearly reveals the characteristics of the classified data, whether in the form of a decision tree structure or in if-then rule format. These rule sets help in understanding which attributes most influence outcomes—in this case, student satisfaction—making them an essential tool in institutional decision-making.

Furthermore, the decision tree output provides an intuitive visual representation that can easily be interpreted by both technical and non-technical users. This enhances its utility for a broader audience, such as academic administrators and policy makers, who may rely on such insights to improve service delivery. Overall, the C4.5 algorithm not only delivers high accuracy but also ensures transparency and interpretability in the classification process, making it a powerful instrument for knowledge discovery and strategic planning in educational environments..

4. Conclusion

Based on the findings of this study, it can be concluded that data mining using the C4.5 algorithm has successfully classified the variables involved in measuring student satisfaction, resulting in 10 decision rules with an accuracy rate of 98.22%. These generated rules are expected to serve as a foundation for making informed decisions regarding actions that should be taken to improve student satisfaction. The implementation of the C4.5 algorithm using RapidMiner Studio software proves to be more effective and efficient in processing data. Moreover, the decision tree generated by RapidMiner Studio matches the manually calculated results, confirming the consistency and reliability of the method used.

Referensi

[1] A. Fauzi, A. B. H. Yanto, and N. Indriyani, "Data Mining Implementation of Organic Vegetable Sales Using Apriori Algorithm," *J. Teknol. Dan Open Source*, vol. 6, no. 1, pp. 98–109, 2023, doi: 10.36378/jtos.v6i1.3049.

[2] M. Hutasuhut, D. Octavina, and J. Halim, "Penerapan Data Mining dalam Menganalisa Pola Kelayakan Siswa Pada Kelas Unggulan Menggunakan Algoritma Iterative Dichotomiser 3 (ID3) pada SMP N. 2 Rantau Selatan," *J. SAINTIKOM (Jurnal Sains Manaj. Inform. dan Komputer)*, vol. 18, no. 2, p. 154, 2019, doi: 10.53513/jis.v18i2.154.

[3] I. Sulihati, "Penerapan Komparasi Algoritma C4.5 dan Naïve Bayes untuk Menentukan Hasil Seleksi Masuk Perguruan Tinggi," *J. Tecnoscienza*, vol. 6, no. 2, pp. 311–320, 2022, doi: 10.51158/tecnoscienza.v6i2.694.

[4] F. Riandari and A. Simgangunsong, *Penerapan Algoritma C4.5 Untuk Mengukur Tingkat Kepuasan Mahasiswa*. 2019.

[5] A. Kusmin, M. . Monoarfa, and A. Juanna, "Pengaruh Kualitas Layanan Sistem Informasi Akademik (SIAT) Terhadap Tingkat Kepuasan Mahasiswa Program Studi S1 Manajemen Fakultas Ekonomi Universitas …," *YUME J. …*, vol. 6, no. 2, pp. 72–82, 2023.

[6] A. Kurbani, "Pengaruh Kualitas Layanan Akademik dan Fasilitas Pendidikan terhadap Kepuasan Mahasiswa Kuliah pada Universitas PGRI Palembang," *J. Media Wahana Ekon.*, vol. 13, no. 4, pp. 22–35, 2017.

[7] J. Al Fian and T. Yuniati, "Pengaruh Kepuasan Dan Kepercayaan Pelanggan Terhadap Loyalitas Pelanggan Auto 2000 Sungkono Surabaya," *J. Ilmu dan Ris. Manaj.*, vol. 5, no. 6, pp. 1–18, 2016.

[8] A. S. Maulana, "Pengaruh Kualitas Pelayanan Dan Harga Terhadap Kepuasan Pelanggan PT. TOI," *J. Ekon.*, vol. 7, no. 2, pp. 113–125, 2016.

[9] P. Purwaningsih *et al.*, "Tingkat Kepuasan Mahasiswa Terhadap Layanan Akademik Dosen Aplikasi Komputer Prodi S1 Administrasi Rumah Sakit STIKES Mutiara Mahakam Samarinda," *J. Pengabdi. Kpd. Masy. Nusant.*, vol. 5, no. 1, pp. 340–345, 2024, doi: 10.55338/jpkmn.v5i1.2483.

[10] E. A. Sari and L. S. Wijaya, "Strategi Promosi Melalui Direct Marketing Untuk Meningkatkan Jumlah Mahasiswa Baru," *J. Ilmu Komun.*, vol. 17, no. 1, p. 16, 2020, doi: 10.31315/jik.v17i1.2608.

[11] N. L. A. P. Ningsih, D. A. P. N. Widari, and I. M. Artawan, "Analisa Kepuasan Mahasiswa terhadap Kualitas Pelayanan Pendidikan," *WACANA Ekon. (Jurnal Ekon. Bisnis dan Akuntansi)*, vol. 19, no. 1, pp. 24–29, 2020, doi: 10.22225/we.19.1.1403.24-29.

[12] Sugiyono, *Metode Penelitian Kuantitatif, Kualitatif dan R&D*. Bandung: Alfabeta, 2017.

[13] Y. S. Luvia, A. P. Windarto, S. Solikhun, and D. Hartama, "Penerapan Algoritma C4.5 Untuk Klasifikasi Predikat Keberhasilan Mahasiswa Di Amik Tunas Bangsa," *Jurasik (Jurnal Ris. Sist. Inf. dan Tek. Inform.*, vol. 1, no. 1, p. 75, 2017, doi: 10.30645/jurasik.v1i1.12.

[14] K. Adhatrao, A. Gaykar, A. Dhawan, R. Jha, and V. Honrao, "Predicting Students' Performance Using ID3 and C4.5 Classification Algorithms," Int. J. Data Min. Knowl. Manag. Process, vol. 3, no. 5, pp. 39–52, 2013, doi: 10.5121/ijdkp.2013.3504.

[15] J. Al Fian and T. Yuniati, "Pengaruh Kepuasan Dan Kepercayaan Pelanggan Terhadap Loyalitas Pelanggan Auto 2000 Sungkono Surabaya," J. Ilmu dan Ris. Manaj., vol. 5, no. 6, pp. 1–18, 2016.

[16] L. S. B. Ginting, W. Zarman, and I. Hamidah, "Analisis dan Penerapan Algoritma C4.5 Dalam Data Mining untuk Mempredikisi Masa Studi Mahasiswa Berdasarkan Data Nilai Akademik," in Prosiding Seminar Nasional Aplikasi Sains & Teknologi (SNAST), 2014, pp. 263–272.

[17] S. Haryati, A. Sudarsono, and E. Suryana, "Implementasi Data Mining Untuk Memprediksi Masa Studi Mahasiswa Menggunakan Algoritma C4.5 (Studi Kasus: Universitas Dehasen Bengkulu)," J. Media Infotama, vol. 11, no. 2, pp. 130–138, 2015.

[18] N. Hijriana and M. Rasyidan, "Penerapan Metode Decision Tree Algoritma C4.5 Untuk Seleksi Calon Penerima Beasiswa Tingkat Universitas," Sains Dan Teknol., vol. 3, no. 1, pp. 9–13, 2017.

[19] R. Hu, "Data Mining in the Application of Criminal Cases Based on Decision Tree," Int. J. Eng. Sci., vol. 2, no. 2, pp. 24–27, 2013, [Online]. Available: http://waprogramming.com/papers/5117e214b38b50.87913974.pdf

[20] E. R. Maris, "Analisis Kepuasan Pelanggan Menggunakan Algoritma C4 . 5," pp. 1–14, 2019.

[21] M. Mulyana, M. Effendy, and L. Hidayat, "Membangun Kepuasan Mahasiswa Pengguna

Laboratorium Komputer," J. Anal. Sist. Pendidik. Tinggi "JAS-PT," vol. 1, no. 2, pp. 93–101, 2017.
[22] B. R. Patel and K. K. Rana, "A Survey on Decision Tree Algorithm For Classification," Ijedr, vol. 2, no. 1, pp. 1–5, 2014.
[23] A. S. Maulana, "Pengaruh Kualitas Pelayanan Dan Harga Terhadap Kepuasan Pelanggan PT. TOI," J. Ekon., vol. 7, no. 2, pp. 113–125, 2016.
[24] H. Al Rasyid, "Pengaruh Kualitas Layanan Dan Pemanfaatan Teknologi Terhadap Kepuasan Dan Loyalitas Pelanggan Go-Jek," J. Ecodemica, vol. 1, no. 2, pp. 210–223, 2017.
[25] T. Sulastri, "Analisis Kepuasan Mahasiswa Terhadap Kinerja Dosen," J. Ilm. Ekon. Manaj. Dan Kewirausahaan "OPTIMAL," vol. 10, no. 2, pp. 167–184, 2016.
[26] A. Syakur, "Hubungan Kualitas Pelayanan Terhadap Kepuasan Mahasiswa Dan Loyalitas Mahasiswa Ditinjau Dari Model Pembelajaran Di Akademi Farmasi Surabaya," REFORMASI, vol. 8, no. 2, pp. 100–108, 2018, [Online]. Available: www.inherent-dikti.net
[27] S. P. Utari, "Implementasi Metode C4 . 5 Untuk Menentukan Guru Terbaik Pada Smk 1 Percut Sei Tuan Medan," J. Buana Inform., vol. 2, no. April, pp. 82–86, 2015.