Comparison of naive Bayes classifier and C4.5 algorithms in predicting student study period

Y A Gerhana¹*, I Fallah¹, W B Zulfikar¹, D S Maylawati¹² and M A Ramdhani¹

¹Department of Informatics, UIN Sunan Gunung Djati Bandung, Indonesia
²Department of Informatics, Sekolah Tinggi Tekologi Garut, Garut, Indonesia

*Corresponding author’s email: yanagerhana@uinsgd.ac.id

Abstract. This study aims to compare Naive Bayes Classifier algorithm with C4.5 in predicting the study period of student. The data for this study is collected from Department of Informatics UIN Sunan Gunung Djati Bandung. Software development method that used is Rational Unified Process (RUP). Comparison of both algorithms is the accuracy of predicting student study period and the speed of data processing. The attributes that used are student id, name, gender, GPA, entry point, tahfidz (cited the holy Quran as a special attribute on the case study), previous school, and extra activities during lecture. The result of this study shows that Naive Bayes Classifier and C4.5 algorithm can be applied to predict the study period of students well and accurate enough. The accuracy of Naive Bayes Classifier algorithm is around 88% better slightly than the C4.5 algorithm that has accuracy around 87%. However, the processing time of algorithm C4.5 is better than the algorithm Naive Bayes Classifier, C4.5 has processing time 0.003 nanosecond and the Naive Bayes Classifier has 12.7 nanosecond.

1. Introduction

The application of information technology in higher education can produce abundant data about students and the learning process [1]. One of the information technology tools that has been used to assist in the implementation of education is the IS (IS). IS is a combination of human activities and computerized technology that process the information [2], which is generally used to support the operational and management [3].

IS is a system that processes the data that organized well [4], which was designed based on user needs [5]. The IS has a high flexibility that allows it to be developed into a better system [6]. Moreover, IS is able to overcome the problem of data processing that has high complexity [1]. In many research before, IS has several advantages in terms of: data accessibility [7], time efficiently [8], accurate result [9], support the decision precisely [10], more economical [11], widely used [12], enhance the user understanding [13], enhance productivity [14], provide good data and information [15], and as data storage media [16].

Department of Informatics UIN Sunan Gunung Jati Bandung has an average on-time graduation presentation that is 13.5%. Whereas every year, in the beginning of the new school year the student quota received is increasing. From the information, it is known that not all students can graduate on time according to the study period taken, so that the number of students is increasing which also has an impact on increasing the amount of data. One of technique that can be used to optimize data in order to get insight of data and become information and knowledge is by data mining [15,17–20].
Data mining is a computer learning techniques to analyse and extract knowledge (insight knowledge) automatically [21]. Other definitions of data mining is the process of establishing the definitions of the common concept by observing specific examples of concepts that will be learned [22].

This study aims to compare the Naive Bayes Classifier algorithm with C4.5 based on the data of students from Department of Informatics UIN Sunan Gunung Djati Bandung. Similar studies have been conducted including: a machine learning approach for student assessment in e-learning using Quinlan's C4.5, Naive Bayes and Random Forest algorithms [23]; implementation of data mining to evaluate students' academic performance using the Naive Bayes Classifier algorithm [24]; the implementation of C4.5 algorithm that is based on ada boost for the prediction of heart disease [25]; implementation of C4.5 and Naive Bayes methods based on ada boost to predict creditworthiness [26]; and lung cancer survivability prediction based on performance using classification techniques of Support Vector Machines, C4.5 and Naive Bayes Algorithms for healthcare analytics [26]. This study uses and compares between the Naive Bayes Classifier algorithm and C4.5 algorithm to predict graduation, with the attributes that used are Student ID, name, gender, GPA, entry point, tahfidz (cited the Holy Quran), school origin, and extra activities during the lecture period. Where the criteria for students who are graduated on time are students who have studied for 4 years.

2. Methods
The software development method that used for this research is Rational Unified Process (RUP). RUP is a software engineering method or software development life cycle that collecting various best practices contained in the software development industry [27]. The main characteristic of this method is to use use-case driven and iterative approaches to the software development life cycle. RUP uses an object oriented concept, with activities that focus on developing models using the Unified Model Language (UML) [7]. Reliability of system that verified by an expert among others the logic, concept, and operational of the system [28].

3. Result and Discussion

3.1. Data Analysis
The research data is primary data, the parameters are taken from student graduation data and student main data which are perceived to affect student graduation. The research data was obtained from the graduation data and the main data if Informatics Department at UIN Sunan Gunung Djati Bandung. Graduation data were taken from 2011 and 2012 class students that have total 109 data, which included:

a. Student Main Data. In the student main data there are many attributes, but in this case is selected attributes that have an influence on the student's study period, among others:
   1) The university entrance way is used to determine the relationship between study students and the entrance way that taken by students.
   2) The origin of school is used to determine the relationship between the period of study of students with schools of origin at the high school level.
   3) Gender is used to find out the comparison between men and women who are influential in the study period.
   4) Activities are used as a comparison between students who are active in organizations, either in the department or faculty, and who are not active in the organizations.

b. Student Graduation Data, among others:
   1) GPA is used to determine the relationship between the level of academic achievement of the student study period.
   2) The graduation of the tahfidz exam that is used to find out the relationship between the student's study period in a timely or late manner in taking the tahfidz exam at the Faculty of Science and Technology. Indicators of students on time are the maximum to take the final exam one year after the tahfidz certificate comes out.
3.2. Data Transformation
Data transformation is the process of converting data into a suitable format for processing in data mining. This research has seven attributes, among others class or label, *tahfidz* status, GPA, activity status, the entrance way, origin school, and gender. There are 5 categories of graduation statement (on time and late); 2 categories for *tahfidz* status (on time and late); 3 categories of GPA (with compliments, very satisfactory, and satisfying); 3 categories of activity status (very active, active, and not active); 6 categories of entrance way (SBMPTN, SNMPTN, Written examination, SPMBTAIn, PPA, and SNU); 5 categories of origin school (natural science, social science, language, technique, and the others); and the last is 2 categories of gender (female (F) and male (M)).

3.3. Data Mining
Data mining is the process of automatically finding useful information in large data collections. Data mining techniques are used to examine large databases as a way to find new and useful patterns. Data mining techniques can be used to improve the ability of a system, such as for information retrieval [29], and the other technique, among others:

a. Decision Support System (DSS), DSS is a systematic approach for problem solving, gathering mature determinants of alternatives that produced and taking the most appropriate action or decision [30][31]. DSS is part of a IS including a knowledge-based system or knowledge management that is used to support decision making [32].

b. Classification, classification is a systematic approach to building a classification model from a set of input data. In the classification there is a target category. For example the classification of study periods on time can be separated into two categories, namely graduating on time and graduating late. Then to determine the study period of a student, used the method of classification in data mining [33].

c. Decision Tree (C4.5), the C4.5 algorithm is one of the algorithms that used to build decision trees. The C4.5 algorithm is the development of the ID3 algorithm [34]. Trees are built by dividing the data recursively until each part consists of data from the same class. [29]. The decision tree starts with a root node that is used by the user to take action. From this root node, the user breaks it according to the decision tree algorithm. The end result is a decision tree with each branch showing possible scenarios of the decisions taken. The decision tree method has several advantages then the other methods for large databases which have relatively fast processing time, can be converted into rules of classification easily and simply, and can use SQL queries to access the database [35]. The steps of C4.5 algorithm for building decision trees are among others [36]:

1) Select an attribute as root, the selection of attributes as root is based on the highest gain value of the existing attribute. To calculate the highest gain value, the equation (1) is used:

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

Where:  
- \( S \) = set of cases  
- \( A \) = attribute  
- \( n \) = number of attribute \( A \) partition  
- \( |S_i| \) = number of cases on the \( i \) partition  
- \( |S| \) = number of cases in \( S \)

Entropy value can be calculated using equation (2):

\[ \text{Entropy}(S) = \sum_{i=1}^{n} - p_i * \log_2 p_i \]  (2)

Where:  
- \( S \) = set of cases  
- \( N \) = number of \( S \) partition  
- \( P_i \) = Proposition of \( S \), to \( S \)

2) Create a branch for each value.
3) Devide cases in the branch.
4) Repeat the process for each branch until all cases in the branch have the same class.
d. Naive Bayes Classifier (NBC), NBC uses a branch of mathematics known as probability theory to find the greatest opportunities for possible classification, by looking at the frequency of each classification in training data. NBC is a statistical classification method that can be used to predict the probability of class membership. The NBC has been proved have high accuracy and fast processing time when applied to a database with large data [37]. This method only requires a small amount of training data to determine the parameter estimates that required in the classification process. NBC often works much better in most complex real-world situations than the expected [38].

The general form of Bayes theorem can be seen in the (3) equation.

\[
P(H \mid X) = \frac{P(X \mid H)P(H)}{P(X)}
\]  

Where: 
- \(X\) = Data with unknown classes
- \(H\) = Data hypocryption \(X\) which is a specific class
- \(P(H \mid X)\) = The probability of \(H\) hypocryption based on the \(X\) condition (posteriori probability)
- \(P(H)\) = Probability of \(H\) hypocryption (prior probability)
- \(P(X \mid H)\) = Probability of \(X\) based on the \(H\) condition
- \(P(X)\) = Probability of \(X\)

3.4. Implementation

a. Naive Bayes Classifier

1) Calculate the number of classes from the graduation information column based on the classification that has been formed.
   a) \(C_1\) (graduation information class = "on time") = number of “on time” in the graduation information column = 31/109 = 0.2844
   b) \(C_2\) (graduation information class = "late") = number of “late” in the graduation information column = 78/109 = 0.7155

2) Calculate the same number of cases in each attribute from the graduation information based on testing data. The value of each attribute in the data is the same as the explanation in the 3.2 section, for the examples:
   a) \(P(\text{gender} = "M" \mid \text{graduation information class} = "on time") = 14/31 = 0.4516\)
   b) \(P(\text{gender} = "L" \mid \text{graduation information class} = "late") = 57/78 = 0.7307\)
   c) \(P(\text{GPA} = "Very satisfactory" \mid \text{graduation information class} = "on time") = 14/31 = 0.4516\)
   d) \(P(\text{GPA} = "Very satisfactory" \mid \text{graduation information class} = "late") = 73/78 = 0.9358\)
   e) \(P(\text{entrace way} = "SBMPTN" \mid \text{graduation information class} = "on time") = 12/31 = 0.387\)
   f) \(P(\text{entrace way} = "SBMPTN" \mid \text{graduation information class} = "late") = 7/78 = 0.089\)
   g) \(P(\text{tahfidz} = "on time" \mid \text{graduation information class} = "on time") = 31/31 = 1\)
   h) \(P(\text{tahfidz} = "on time" \mid \text{graduation information class} = "late") = 15/78 = 0.1923\)
   i) \(P(\text{basic in high school} = "natural sciences" \mid \text{graduation information class} = "on time") = 27/31 = 0.8709\)
   j) \(P(\text{basic in high school} = "natural sciences" \mid \text{graduation information class} = "late") = 61/78 = 0.782\)
   k) \(P(\text{activity status} = "active" \mid \text{graduation information class} = "on time") = 5/31 = 0.1612\)
   l) \(P(\text{activity status} = "active" \mid \text{graduation information class} = "late") = 8/78 = 0.1025\)

3) Multiply All Variable Results
   a) All attribute of graduation information class = “on time”
      \(P(\text{X} \mid \text{graduation information class} = "on time") = 0.4516 \times 0.4516 \times 0.387 \times 1 \times 0.8709 \times 0.1612 = 0.0110\)
   b) All attribute of graduation information class = “late”
      \(P(\text{X} \mid \text{graduation information class} = "late") = 0.7307 \times 0.9358 \times 0.089 \times 0.1923 \times 0.782 \times 0.1025 = 0.0009\)
   c) Prior probability multiplication is the same with all attribute of graduation information class = “on time”
P \left( C_i \right) | \text{graduation information class} = \text{"on time"} \times P \left( X \right | \text{graduation information class} = \text{"on time"}) = 0.0110 \times 0.2844 = 0.0031

d) Prior probability multiplication is the same with all attitude of graduation information class = \text{"late"} 
\[ P \left( C_i \right) | \text{graduation information class} = \text{"late"} \times P \left( X \right | \text{graduation information class} = \text{"late"}) = 0.0009 \times 0.7155 = 0.00067

4) Compare the result of each class 
b. Algorithm C4.5
1) Calculation of entropy value (described in Table 1):

| Class          | Frequency | \( Pl * log pi \) |
|----------------|-----------|--------------------|
| On time        | 31        | 0.515904848        |
| Late           | 78        | 0.345477103        |
| Entropy(S)     |           | 0.861381951        |

2) Calculation of entropy value and gain value from each attribute in node 1 (described in Table 2)

| Node          | Classification of each Node | Total | On time | Late | Entropy | Gain    |
|---------------|-----------------------------|-------|---------|------|---------|---------|
| Sex           | Male                        | 71    | 14      | 57   | 0.716258391 | 0.048996106 |
|               | Female                      | 38    | 17      | 21   | 0.991992403  |         |
| GPA           | With compliments            | 24    | 2       | 22   | 0.41381685    | 0.208189438 |
|               | Satisfying                  | 63    | 12      | 51   | 0.702466551   |         |
|               | Very Satisfactory           | 22    | 17      | 5    | 0.77322674    |         |
| Entrance Way  | SBMPTN                      | 19    | 12      | 7    | 0.949452015   | 0.103096145 |
|               | SNMPTN                      | 8     | 2       | 6    | 0.811278124   |         |
|               | Written Exam                | 48    | 8       | 40   | 0.650022422   |         |
|               | PPA                         | 28    | 8       | 20   | 0.863120569   |         |
|               | SPMBPTAIN                   | 3     | 1       | 2    | 0.918295834   |         |
|               | SNU                         | 2     | 0       | 2    | 0           |         |
| Tablids       | On time                     | 46    | 31      | 15   | 0.910878379   | 0.476974562 |
|               | Late                        | 63    | 0       | 63   | 0           |         |
| Senior high school | Natural Science      | 88    | 27      | 61   | 0.88946639    | 0.024193949 |
|               | Social Science              | 2     | 0       | 2    | 0           |         |
|               | Technique                   | 16    | 4       | 12   | 0.811278124   |         |
|               | Language                    | 1     | 0       | 1    | 0           |         |
|               | Others                      | 1     | 0       | 1    | 0           |         |
| Activity status | Not Active                | 82    | 17      | 65   | 0.736323027   | 0.072037295 |
|               | Active                      | 13    | 5       | 8    | 0.961236605   |         |
|               | Very Active                 | 14    | 9       | 5    | 0.940285959   |         |

3) Calculation of entropy value and gain value from each attribute in node 2 (described in Table 3)

| Node          | Classification of each Node | Total | On time | Late | Entropy | Gain    |
|---------------|-----------------------------|-------|---------|------|---------|---------|
| Sex           | Male                        | 24    | 14      | 10   | 0.979868757  | 0.489566933 |
|               | Female                      | 22    | 17      | 5    | 0.77322674    | 0.559086342 |
| GPA           | With compliments            | 6     | 2       | 4    | 0.918295834   | 0.514124609 |
|               | Satisfying                  | 22    | 12      | 10   | 0.994030211   |         |
|               | Very Satisfactory           | 18    | 17      | 1    | 0.309543429   |         |
| Entrance Way  | SBMPTN                      | 14    | 12      | 2    | 0.591672779   |         |
|               | SNMPTN                      | 3     | 2       | 1    | 0.918295834   |         |
|               | Written Exam                | 12    | 8       | 4    | 0.918295834   |         |
|               | PPA                         | 14    | 8       | 4    | 0.985228136   |         |
|               | SPMBPTAIN                   | 2     | 1       | 1    | 0           |         |
|               | SNU                         | 0     | 0       | 0    | 0           |         |
| Senior High School | Natural sciences   | 40    | 27      | 13   | 0.909736123   | 0.527533833 |
|               | Social sciences             | 1     | 0       | 1    | 0           |         |
|               | Technique                   | 4     | 4       | 0    | 0           |         |
|               | Language                    | 0     | 0       | 0    | 0           |         |
|               | Others                      | 0     | 0       | 0    | 0           |         |
| Activity status | Not Active                | 26    | 17      | 9    | 0.930586129   | 0.480042054 |
|               | Active                      | 8     | 5       | 3    | 0.954434003   |         |
4) Calculation of entropy value and gain value for each attribute in node is repeated until all of entropy and gain value are produced.

5) After all of entropy value and gain value are calculated, then create decision tree of C4.5 algorithm (shown in Figure 1).

![Decision Tree](image)

**Figure 1.** Decision Tree.

3.5. Comparison between NBC and C4.5 algorithm

Based on training and testing process using 109 sample data of students at Informatics Department UIN Sunan Gunung Djati Bandung, NBC algorithm has an accuracy value around 88.07%, it means that NBC can classify correctly 96 from 109 data. Whereas, C4.5 algorithm has an accuracy value around 87.15% which is meant C4.5 can classify correctly 95 from 109 data.

4. Conclusion

The Naive Bayes Classifier and C4.5 algorithms that are applied to predict the student study period are going well and quite accurately. Although not significant, NBC algorithm has a better accuracy compared to the C4.5 algorithm which is 88% and 87%. However, C4.5 algorithm has a better processing time than NBC, it is 0.003 nanosecond and NBC has 12.7 nanoseconds. For the further research, it can be more training and testing data so that the result of prediction can be accurate.

5. References

[1] Darmalaksana W, Ramdhani M A, Cahyana R and Amin A S 2018 Strategic Design of IS Implementation at University *Int. J. Eng. Technol.*, 7 2.29 787–791

[2] Pamoragung A, Suryadi K and Ramdhani M A 2006 Enhancing the implementation of e-Government in Indonesia through the high-quality of virtual community and knowledge portal *Proceedings of the European Conference on e-Government, ECEG* 341–348

[3] Ramdhani M A 2013 *Metodologi Penelitian untuk Riset Teknologi Informasi.* (Bandung: UIN Sunan Gunung Djati)

[4] Maylawati D S, Darmalaksana W and Ramdhani M A 2018 Systematic Design of Expert System Using Unified Modelling Language *IOP Conf. Ser. Mater. Sci. Eng.* 288 1 012047

[5] Ramdhani M A, Maylawati D S, Amin A S and Aulawi H 2018 Requirements Elicitation in Software Engineering *Int. J. Eng. Technol.* 7 2.29 772–775

[6] Aulawi H, Ramdhani M A, Slamet C, Ainissyifa H and Darmalaksana W 2017 Functional Need Analysis of Knowledge Portal Design in Higher Education Institution *Int. Soft Comput.* 12 2 132–141

[7] Maylawati D S, Ramdhani M A and Amin A S 2018 Tracing the Linkage of Several Unified
Modelling Language Diagrams in Software Modelling Based on Best Practices Int. J. Eng. Technol. 7 2.97 776–780

[8] Slamet C, Rahman A, Sutedi A, Darmalaksana W, Ramdhani M A and Maylawati D S 2018 Social Media-Based Identifier for Natural Disaster IOP Conf. Ser. Mater. Sci. Eng. 288 1 012039

[9] Slamet C, Andriani R, Maylawati D S, Darmalaksana W and Ramdhani M A 2018 Web Scraping and Naïve Bayes Classification for Job Search Engine IOP Conf. Ser.: Mater. Sci. Eng 288 1 1–7

[10] Gerhana Y A, Zulfikar W B, Ramdani A H and Ramdhani M A 2018 Implementation of Nearest Neighbor using HSV to Identify Skin Disease IOP Conf. Ser. Mater. Sci. Eng. 288 1 012153

[11] Rahman C, Slamet C, Darmalaksana W, Gerhana Y A and Ramdhani M A 2018 Expert System for Deciding a Solution of Mechanical Failure in a Car using Case-based Reasoning IOP Conf. Ser. Mater. Sci. Eng. 288 1 012011

[12] Slamet C, Rahman A, Ramdhani M A and Darmalaksana W 2016 Clustering the Verses of the Holy Qur’an Using K-Means Algorithm Asian J. Inf. Technol. 15 24 5159–5162

[13] Maylawati D S, Ramdhani M A, Zulfikar W B, Taufik I and Darmalaksana W 2017 Expert system for predicting the early pregnancy with disorders using artificial neural network 2017 5th Int. Conf. Cyber IT Serv. Manag. CITSM 2017

[14] Zulfikar W B, Jumadi, Prasetyo P K and Ramdhani M A 2018 Implementation of Mamdani Fuzzy Method in Employee Promotion System IOP Conf. Ser. Mater. Sci. Eng. 288 1 012147

[15] Maylawati D S, Ramdhani M A, Rahman A and Darmalaksana W 2017 Incremental technique with set of frequent word item sets for mining large Indonesian text data 2017 5th Int. Conf. Cyber IT Serv. Manag. CITSM 2017 1–6

[16] Taufik A, Ismail N, Gerhana Y A, Komarujaman K and Ramdhani M A 2018 Design of Smart System to Detect Ripeness of Tomato and Chili with New Approach in Data Acquisition IOP Conference Series: Materials Science and Engineering 288 1 012018

[17] Slamet C, Atmadja A R, Maylawati D S, Lestari R S, Darmalaksana W and Ramdhani M A 2018 Automated Text Summarization for Indonesian Article Using Vector Space Model in IOP Conference Series: Materials Science and Engineering 288 1 012037

[18] Maylawati D S and Saptawati G A P 2017 Set of Frequent Word Item sets as Feature Representation for Text with Indonesian Slang Journal of Physics: Conference Series

[19] Jumadi, Maylawati D S A, Subaeki B and Ridwan T 2016 Opinion mining on Twitter microblogging using Support Vector Machine: Public opinion about State Islamic University of Bandung Proceedings of 2016 4th International Conference on Cyber and IT Service Management, CITSM 2016

[20] Maylawati D S and Saptawati G A P 2017 Set of Frequent Word Item sets as Feature Representation for Text with Indonesian Slang Journal of Physics: Conference Series

[21] Maylawati D S, Rahman A, Saputra M I N, Darmalaksana W and Ramdhani M A 2018 Model of Citation Network Analysis using Sequence of Words as Structured Text Representation IOP Conference Series: Materials Science and Engineering 288 1 012048

[22] Hermawati F A 2013 Data Mining (Yogyakarta: Andi)

[23] Mahboob T, Irfan S and Karamat A 2016 A machine learning approach for student assessment in E-learning using Quinlan’s C4. 5, Naïve Bayes and Random Forest algorithms Multi-Topic Conference (INMIC), 2016 19th International 1–8

[24] Ridwan M, Suyono H and Sarosa M 2013 Penerapan Data Mining untuk Evaluasi Kinerja Akademik Mahasiswa Menggunakan Algoritma Naive Bayes Classifier J. EECCIS 7 1 59–64

[25] Rohman A, Suhartono V and Supriyanto C 2017 Penerapan Algoritma C4. 5 Berbasis Adaboost untuk Prediksi Penyakit Jantung Cyberku J. 13 1 2

[26] Pradeep K R and Naveen N C 2018 Lung Cancer Survivability Prediction based on Performance Using Classification Techniques of Support Vector Machines, C4. 5 and Naïve Bayes Algorithms for Healthcare Analytics Procedia Comput. Sci. 132 412–420

[27] Pressman R S 2011 Software Engineering: A Practitioner’s Approach, 7th ed. (New York: McGraw-Hill)
[28] Suhendi H Y, Ramdhani M A and Irwansyah F S 2018 Verification Concept of Assessment for Physics Education Student Learning Outcome Int. J. Eng. Technol. 7 3.21 321–325

[29] Setiawati D, Taufik I and Zulfikar W B 2016 Klasifikasi Terjemahan Ayat al-Quran tentang Ilmu Sains Menggunakan Algoritma Decision Tree Berbasis Mobile JOIN (Jurnal Online Inform. J) 1 24–27

[30] Eniyati S 2011 Perancangan Sistem Pendukung Pengambilan Keputusan untuk Penerimaan Beasiswa dengan Metode SAW (Simple Additive Weighting) Din. - J. Teknol. Inf. 16 2 171–176

[31] Suryadi K and Ramdhani M A 2002 Sistem Pendukung Keputusan (Bandung: Remaja Rosdakarya)

[32] Fauzan R, Indrasary Y and Muthia N 2017 Sistem Pendukung Keputusan Penerimaan Beasiswa Bidik Misi di POLIBAN dengan Metode SAW Berbasis Web JOIN (Jurnal Online Inform. J) 2 2 79–83

[33] J. S. Informasi, “Klasifikasi Data Karyawan Untuk Menentukan Jadwal Kerja Menggunakan Metode Decision Tree.”

[34] Larose D T 2005 Discovering Knowledge in Data (New Jersey: John Wiley & Sons)

[35] Mashlahah S 2013 Prediksi Kelulusan Mahasiswa Menggunakan Metode Decision Tree dengan Penerapan Algoritma C4.5 Jur. Tek. Inform. Fak. Sains dan Teknol. UIN Maulana Malik Ibrahim Malang

[36] Informasi S, Batam U P, Kunci K, Kerja K and Mining D 2017 Analisa dan Penerapan Algoritma C4.5 Dalam Data Mining Untuk Mengidentifikasi Faktor-Faktor Penyebab Kecelakaan Kerja Kontruksi PT. Arupadhatu Adisesanti JOIN (Jurnal Online Inform.) 2 1 36–41

[37] Budiawan W and Lukman N 2016 Perbandingan Naive Bayes Classifier dengan Nearest Neighbor untuk Identifikasi JOIN (Jurnal Online Inform.) 1 2 82–86

[38] Saleh A 2015 Implementasi Metode Klasifikasi Naive Bayes Dalam Memprediksi Besarnya Penggunaan Listrik Rumah Tangga Creative Information Technology Journal (CITEC Journal) 2 3 207–217

Acknowledgments
The authors expressed the appreciation and thank you to the Research and Publishing Center of UIN Sunan Gunung Djati Bandung, which has provided funding support for the publication of this article.