

## Comparison of Naive Bayes Method, K-NN (K-Nearest Neighbor) and Decision Tree for Predicting the Graduation of 'Aisyiyah University Students of Yogyakarta

Tikaridha Hardiani<sup>1,\*</sup>

Department of Information Technology, Science and Technology Faculty University 'Aisyiyah Yogyakarta, Indonesia

<sup>1</sup>tikaridha@unisayogya.ac.id\*

\* corresponding author

Submission date: 10 Juli 2018, Receipt date: 10 Oktober 2019, Publication date: 1 Juli 2020

### Abstract

*The students of Universitas 'Aisyiyah Yogyakarta have been increasing including the number of students in the Faculty of Health Sciences. In 2016 the total number of UNISA students was 1851. The increasing number of students every year leads to great numbers of data stored in the university database. The data provide useful information for the university to predict student graduation or student study period whether they graduate on time with a study period of 4 years or late with a study period of more than 4 years. This can be processed by using a data mining technique that is the classification technique. Data needed in the classification technique are data of students who have graduated as training data and data of students who are still studying in the university as testing data. The training data were 501 records with 10 goals and the testing data were 428 records. Data mining process method used was the Cross-Industry Standard Proses for Data Mining (CRISPDM). The algorithms used in this study were Naive Bayes, K-Nearest Neighbor (KNN) and Decision Tree. The three algorithms were compared to see the accuracy by using Rapidminer software. Based on the accuracy, it was found that the K-NN algorithm was the best in predicting student graduation with an accuracy of 91.82%. The K-NN algorithm showed that 100% of the students of Nursing study program of Universitas Aisyiyah Yogyakarta are predicted to graduate on time.*

**Keywords:** data mining, prediction, student, graduation, decision tree, naive bayes, K-NN

---

### INTRODUCTION

Universitas Aisyiyah Yogyakarta (UNISA) is one of the private universities in Yogyakarta. UNISA was established as Aisyiyah Midwife School in 1963 and later became the Midwifery Academy and in 2016 'Aisyiyah STIKES was changed into Universitas 'Aisyiyah Yogyakarta. There are three faculties in UNISA namely the Faculty of Health Sciences, the Faculty of Science and Technology and the Faculty of Social and Humanities. The number of UNISA students increases from year to year. In 2014 the total number of UNISA students was 1191 people. In 2015 the total number of students increased to 1294 people, and in 2016 the total number of UNISA students was 1851. The number of students in the Faculty of Health Sciences also increases every year.

The University has the student data including academic data and student graduation data. The graduation data are included in the study program and university

---



accreditation points. The on- time students' graduation data are important. Universitas Aisyiyah does not have a predictive pattern for students who graduate on time later used as the reference to predict the number of students who graduate on time.

In accordance with the academic regulation book of Universitas 'Aisyiyah Yogyakarta Health Sciences Academic Year 2017/2018, a regular Bachelor (S-1) program is an academic education program after secondary education which has 144-160 credits scheduled for 8 semesters and at last 14 semesters.

One of the management systems in Universitas 'Aisyiyah Yogyakarta is Integrated Higher Education Management Information System (SIMPTT). The SIMPTT processes data and information related to students including the semester registration card (KRS) process, scheduling process, the lecture process, library information, financial information and students' biodata. So far, the data in SIMPTT have not been used optimally. Data of hundreds to thousands of students can be explored more deeply to be important information. To process these data, the data mining technique is needed. Data mining is a process of finding interesting and hidden patterns from a large collection of data stored in a database, data warehouse, or other data storage area. Data will be processed into information. The ultimate goal of data mining is to use the information to make efficient and right decisions [1]. The data mining algorithms used are Naive Bayes, KNN (K-Nearest Neighbor) and Decision Tree. The algorithm will be compared to the level of accuracy using rapidminer software.

The purpose of this study is to determine the prediction of the study period of the students of the Faculty of Health Sciences, Aisyiyah University Yogyakarta using Naive Bayes, KNN and Decision Tree method and to determine the best algorithm performance from the three algorithms to predict student graduation.

### 1.1. Data Mining

Data mining is a technique to explore hidden information to obtain more benefits from the existing data [2]. According to Han and Kamber, the development of data mining was prompted by large numbers of data that can actually be used to produce useful information [3]. Data mining is known as Knowledge Discovery in Database (KDD). The stages in the use of data mining called KKD process are shown in figure1.

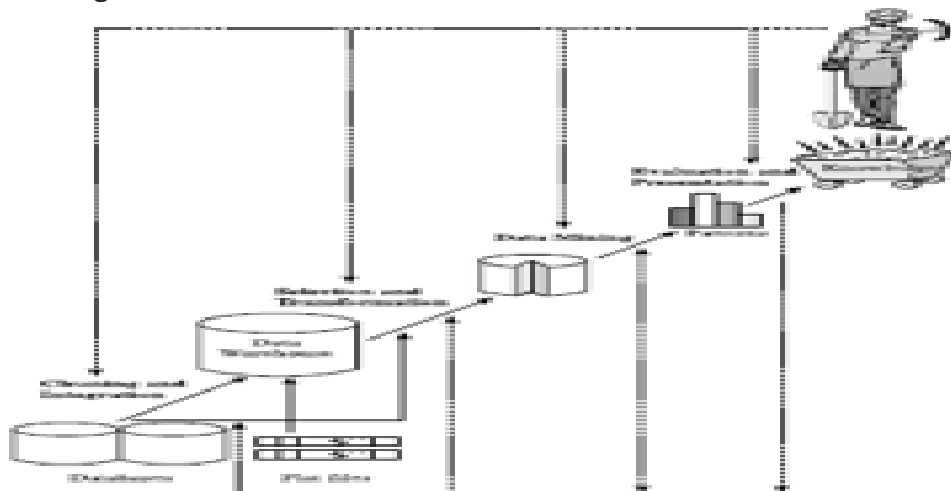


Figure 1. Data Mining Stages

## 1.2 CRISP-DM

CRISP-DM is a standard developed in 1996 by analysts from several industries such as DaimlerChrysler, SPSS, and NCR as a problem-solving strategy of a business or research unit [4]. CRISP DM is a standard data mining process as a problem-solving strategy in general of a business or research unit. The stages of CRISP-DM are illustrated in Figure 2.

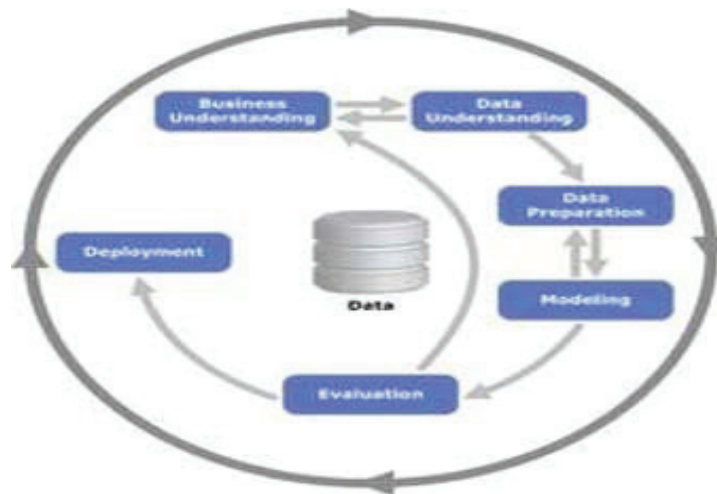


Figure 2. Stages of CRISP-DM

## 1.3 KNN

A K-Nearest Neighbor algorithm is an algorithm whose learning is based on examples where training data are stored so that the new classification can be compared by taking the most similar records in training data [5]. Classification method that classifies new data based on the distance of the new data to some of the nearest data/neighbors. The following is the sequence of work processes of K-NN [6]:

- 1) Determining the k parameter (the closest number of neighbors).
- 2) Calculate the square of the Euclidean distance of each object (training data) on the given sample data.

$$d_i = \sqrt{\sum_{i=1}^p (x_{2i} - x_{1i})^2}$$

**x1 = Sample Data**

**x2 = Test Data**

**i = Variable Data**

**d = Distance**

**p = Data Dimension.**

- 3) Sort these objects into groups that have the smallest Euclidean distance.
- 4) Collect y categories (classes of the nearby objects) as many as k.

## 1.4 Naive Bayes

Naive Bayes is a simple probabilistic classification that calculates a set of probabilities by summing up the frequency and combination of values from a given dataset. The algorithm uses the Bayes theorem and assumes all independent or non-interdependent attributes given by values in class variables [7]. The equation of Naive Bayes is as follows [8].

$$P(C_i|X) = \frac{P(X|C_i)P(C_i)}{P(X)}$$

Description:

$X$  : Criteria of a case based on inputs

$C_i$  : Solution class of pattern i-th, where i is the number of class labels

$P(C_i|X)$  : Probability of occurrence of class label  $C_i$  with input criteria  $X$

$P(X|C_i)$  : Probability of input criteria  $X$  with class label  $C_i$

$P(C_i)$  : Probability of class label  $C_i$

## RESEARCH METHODS

The research method used was Cross Standard Industry Process for Data Mining (CRISP- DM). CRISP-DM was developed in 1996 which aimed to be a standard data mining application. The stages in CRISP-DM are as follow [10] [11]:

### 1. Business Understanding

The purpose of the research was determined in this stage in detail in the business scope or research unit. Subsequently, the translation and analysis of the objective and the limitation of the study of data mining problems were conducted, and then a strategy was prepared to achieve the objective of the study.

### 2. Data Understanding

In this stage, the data were collected to be analyzed and investigated so that the initial data pattern structure is known to obtain an overview of the initial knowledge that was generated from data mining. Furthermore, the quality data was evaluated to eliminate the data of missing value, duplication, and typography. If possible, a small portion of the data group that contains the pattern of the problem was selected. In this phase, the data were selected from the UNISA Nursing Study Program students and the scope of the attributes needed for data mining.

### 3. Data Preparation

Data preparation was performed in this stage by selecting cases and variables to be analyzed in accordance with the analysis type to be carried out. The data were examined whether there is a need for changes in some variables. It is expected that the data would be ready and meet the criteria for modelling. The additional data or other information that is possible to facilitate the data mining process was performed. This stage allows optimization of attribute determination so that the attributes that significantly improve the accuracy of data mining processes were obtained.

### 4. Modeling

Modeling involves situations and specification setting so that the data can be processed by using planned data mining methods. The tools or coding using certain programming languages were required to process the data mining by a computer system by using Rapidminer. Rapidminer is an open source software. The modeling used Naive Bayes, KNN and decision tree algorithms.

### 5. Evaluation

To evaluate one or more models used in modeling to obtain quality and effectiveness before used, to determine whether there is a model that meets the objectives, and to evaluate and validate the performance (accuracy) of the three algorithms using ROC matrix and curve configuration.

### 5.1 Confusion matrix

Confusion matrix is a method used to evaluate and calculate accuracy in data mining [12]. Confusion matrix defines the relationship between values predicted by the model and actual values. The accuracy of the classification results is calculated based on matrix performance [13]. Evaluation with confusion matrix produces accuracy, precision and recall values. The percentage of recorded data classified correctly after the test carried out at the classification results is called accuracy. Precision or confidence is a positive predictive value. Recall or sensitivity is the proportion of true positive cases that are correctly predicted positively [14]. High precision and recall indicate good predictions. The matrix configuration model is presented in Table 1. Data warehouse, or other data storage area. Data will be processed into information. The ultimate goal of data mining is to use the information to make efficient and right decisions [1]. The data mining algorithms used are Naive Bayes, KNN (K-Nearest Neighbor) and Decision Tree. The algorithm will be compared to the level of accuracy using rapidminer software. The purpose of this study is to determine the prediction of the study period of the students of the Faculty of Health Sciences, Aisyiyah University Yogyakarta using Naive Bayes, KNN and Decision Tree method and to determine the best algorithm performance from the three algorithms to predict student graduation.

**Table 1.** Confusion Matrix

| Actual   | Prediction          |                     |
|----------|---------------------|---------------------|
|          | Positive            | Negative            |
| Positive | True Positive (TP)  | True Negative (TN)  |
| Negative | False Positive (FP) | False Negative (FN) |

#### 5.1.1 ROC curve

The method that calculates area under the ROC curve is called AUC (Area Under the ROC Curve). AUC is defined as probability [9]. Confusion matrix is expressed through the ROC curve. The AUC classification is described in Table 2.

**Table 2.** Klasifikasi AUC

| Performance AUC | Classification           |
|-----------------|--------------------------|
| 0,5-0,6         | Failure                  |
| 0,6-0,7         | Poor classification      |
| 0,7-0,8         | Fair classification      |
| 0,8-0,9         | Good classification      |
| 0,9-1,0         | Excellent classification |

### 6. Deployment (Deployment)

A deployment is the use of the resulted model namely a series of methods and representative data that have been processed so as to provide optimal information to interested parties [15]. A deployment in a simple scope is the use of the end result of data mining, for example: report of process result using data mining for student graduation prediction.



## RESULTS AND DISCUSSION

### 1. Business Understanding

The first step in CRISP-DM was understanding the business or research objectives to know the prediction of the study period of Faculty of Health students of Universitas Aisyiyah. The results of this prediction were used as the bases of decision making.

### 2. Data Understanding

The research data were taken from the Data and Information System Center of Universitas Aisyiyah Yogyakarta in the form of excel 1231. The file consisted of the data of students who have graduated and who are still studying. The training data were the students who have graduated from 2012 to 2018, while the testing data were 428 students who are still studying.

### 3. Data Preparation

Data preparation includes all activities to develop student datasets applied to the modeling tools, started by identifying initial raw data to the data mining process. The initial data were 1231 records consisted of 18 attributes, in the form of number, student's number, full name, gender, institution of origin, province of origin, district of origin, number of credits, final GPA, study program, test path, starting date of college, date of graduation, graduation year, month of study, year of study, length of study, activity. The data were selected into 10 attributes, namely number, student's number, full name, gender, school of origin, area of origin, number of credits, GPA, test path, length of study. Either Yes or No was to answer the length of study. A Yes is a student who graduates on time who spends less than or equal to four years to graduate. No is a student who graduates late who spends more than four years to graduate. After data selection was performed, data cleaning was carried out to clear data from noise or missing values. There were 730 data of anullent nursing study program and missing values so that the final data used were 501.

|    | C                          | D             | E                        | F                         | G                  | H      | I           | J          | K |
|----|----------------------------|---------------|--------------------------|---------------------------|--------------------|--------|-------------|------------|---|
| 1  | nama lengkap               | jenis kelamin | ASAL SEKOLAH             | ASAL DAERAH               | jumlah ipk (akhir) | jumlah | jalur ujian | lama studi |   |
| 2  | AGUSTIN RETNO DEWI         | P             | SMAN 1 KESAMBEN BLITAR   | Prop. Jawa Timur          | 3,47               | 3,25   | TPA         | YES        |   |
| 3  | AKHMAD KHAYKUN THORIQ IL   | L             | MAN 2 KUDUS              | Prop. Jawa Tengah         | 3,47               | 3,28   | TPA         | YES        |   |
| 4  | ALAN YANUAR                | L             | SMAN SIDAREJA OLACAP     | Prop. Jawa Tengah         | 3,47               | 3,45   | TPA         | YES        |   |
| 5  | AMANDA FE RUBY INDAH PL P  | P             | SMAN 1 TEMON KULON PRO   | Prop. D.I. Yogyakarta     | 3,47               | 3,46   | TPA         | YES        |   |
| 6  | ANITA NURFAJRI             | P             | SMAN 1 KOTA BINA NUSA TE | Prop. Nusa Tenggara Barat | 3,47               | 3,34   | PMOK        | YES        |   |
| 7  | ANNISA SEPTIANI            | P             | SMAN 1 SENTOLO KULON PR  | Prop. D.I. Yogyakarta     | 3,47               | 3,34   | TPA         | YES        |   |
| 8  | ANIWAR ARIFIN KUSUMA AREL  | L             | SMAN 1 SEWON BANTUL      | Prop. D.I. Yogyakarta     | 3,47               | 3,63   | TPA         | YES        |   |
| 9  | ARI KURNIAWAN HUTOMO       | L             | SMAN 1 LEDO KALIMANTAN I | Prop. Kalimantan Barat    | 3,47               | 3,24   | TPA         | YES        |   |
| 10 | ARISKA PUJI HAKIKI         | L             | SMK KESEHATAN YARSI MATI | Prop. Nusa Tenggara Barat | 3,47               | 3,13   | TPA         | YES        |   |
| 11 | ARUM TINI SARAS WATI       | P             | SMAN 1 KARANGANOM KLAT   | Prop. Jawa Tengah         | 3,47               | 3,43   | PMOK        | YES        |   |
| 12 | ASEP PURNOMOSIDI           | L             | SMA WIDYA KUTOARJO       | Prop. Jawa Tengah         | 3,47               | 3,45   | TPA         | YES        |   |
| 13 | ASTRID GHEAFANI LAZUARDI P | P             | SMA MUHAMMADYAH 1 KOC    | Prop. D.I. Yogyakarta     | 3,47               | 3,33   | PMOK        | YES        |   |
| 14 | BAGUS SETYAJI              | L             | SMAN 1 PURWAREJA BANDAR  | Prop. Jawa Tengah         | 3,47               | 3,31   | PMBU        | YES        |   |
| 15 | DIAN ADI SAPUTRO           | L             | SMAN 1 KARANGMULJO GURI  | Prop. D.I. Yogyakarta     | 3,47               | 3,31   | TPA         | YES        |   |
| 16 | DWI PEMTIKATI ARYUNA SAR P | P             | SMAN 1 SEYEGAN SLEMAN    | Prop. D.I. Yogyakarta     | 3,47               | 3,33   | TPA         | YES        |   |
| 17 | DWI PURWATI                | P             | SMAN 9 MUARD JAMBI       | Prop. D.K.I. Jakarta      | 3,47               | 3,43   | TPA         | YES        |   |
| 18 | ERLINDA PERMATASARI        | P             | SMAN 1 KARANGANOM KLAT   | Prop. Jawa Tengah         | 3,47               | 3,64   | PMOK        | YES        |   |
| 19 | HANI HANFAH                | P             | SMA PGRI 1 DENPASAR      | Prop. Bali                | 3,47               | 3,44   | PMOK        | YES        |   |
| 20 | HELGA DWI ARDIKANTO        | L             | SMAN 1 SIDAREJA OLACAP   | Prop. Jawa Tengah         | 3,47               | 3,13   | TPA         | YES        |   |
| 21 | IIN FAJAR CAHYANI          | P             | MAN PURWOREJO            | Prop. Jawa Tengah         | 3,47               | 3,25   | TPA         | YES        |   |
| 22 | IRAWAN DANAR NURUNICO L    | L             | SMAN 1 PENGASIH KULON PR | Prop. D.I. Yogyakarta     | 3,47               | 3,1    | PMBU        | YES        |   |
| 23 | LINTANG TITISARI           | P             | SMA MUHAMMADYAH 2 KOC    | Prop. D.I. Yogyakarta     | 3,47               | 3,79   | TPA         | YES        |   |
| 24 | LISA SEPTIANI              | P             | SMAN 1 BOYOLALI          | Prop. Jawa Tengah         | 3,47               | 3,25   | TPA         | YES        |   |
| 25 | LUKMAN ARIF NURCOTO        | L             | SMA BUDI MULIA 2         | Prop. D.I. Yogyakarta     | 3,47               | 3,25   | TPA         | YES        |   |
| 26 | MAHYAWATI                  | P             | SMK KESEHATAN YARSI MATI | Prop. Nusa Tenggara Barat | 3,47               | 3,52   | TPA         | YES        |   |
| 27 | MARETA FITRIA WULANDARI P  | P             | SMAN TUGUNULYO SUMATER   | Prop. Sumatera Selatan    | 3,47               | 3,41   | TPA         | YES        |   |

Figure 3. Data used in the study

#### 4. Modeling

The study used rapidminer software. The algorithms used were naive bayes, K-NN and Decision trees.

##### a. Naive Bayes

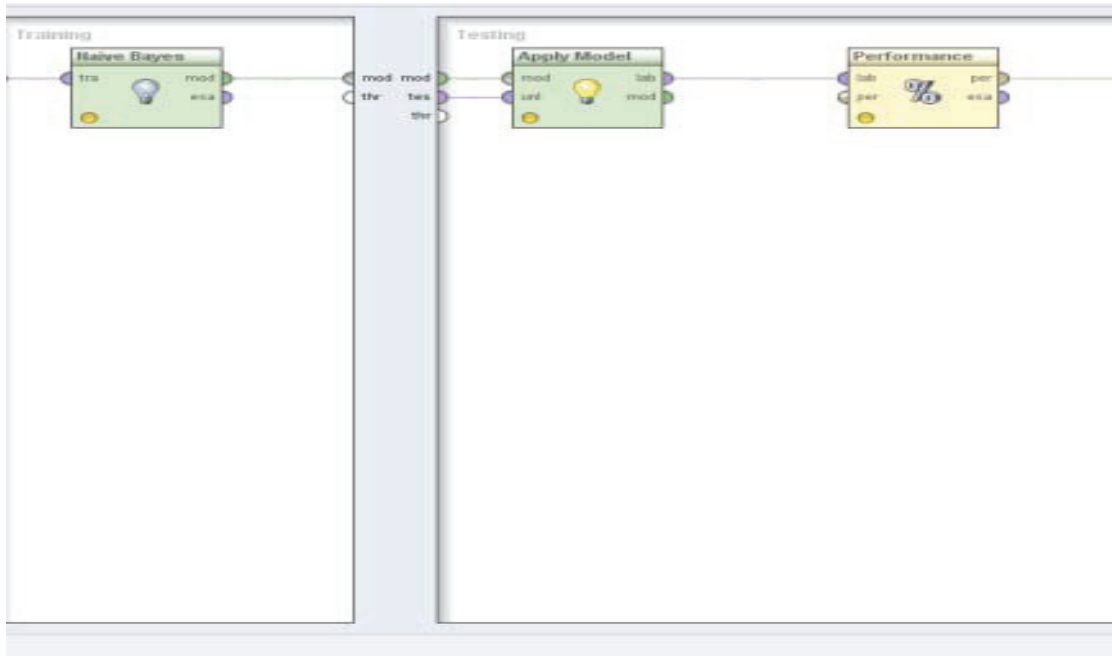


Figure 4. Naive Bayes algorithm on rapidminer

##### b. K-NN

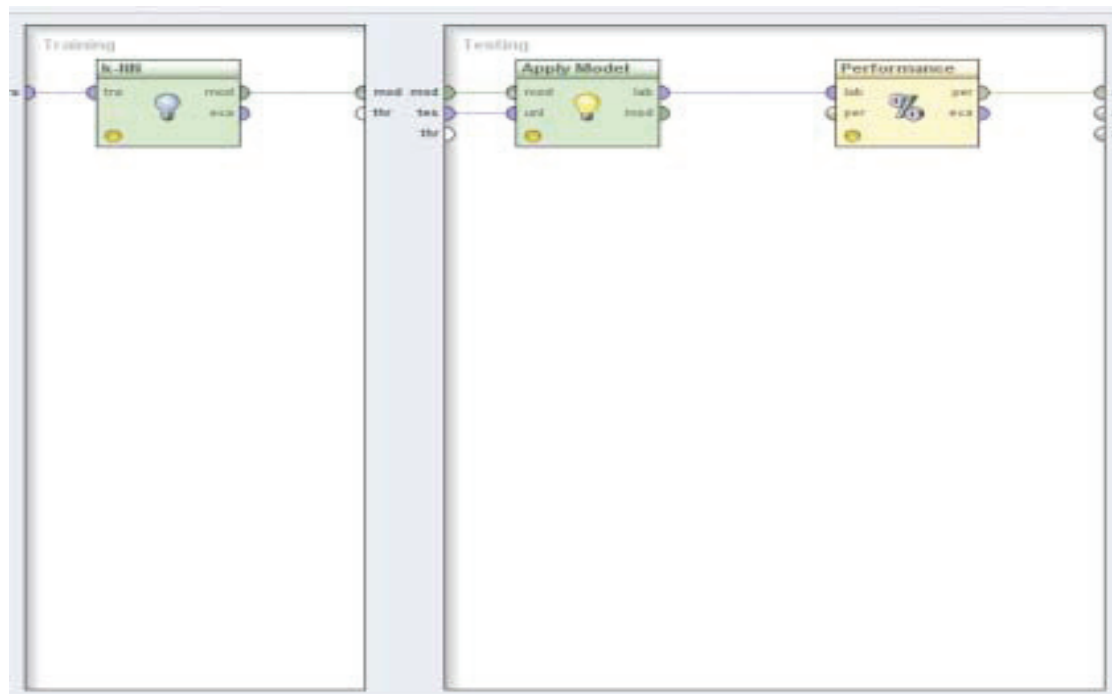


Figure 5. KNN algorithm on rapidminer

c. Decision Tree

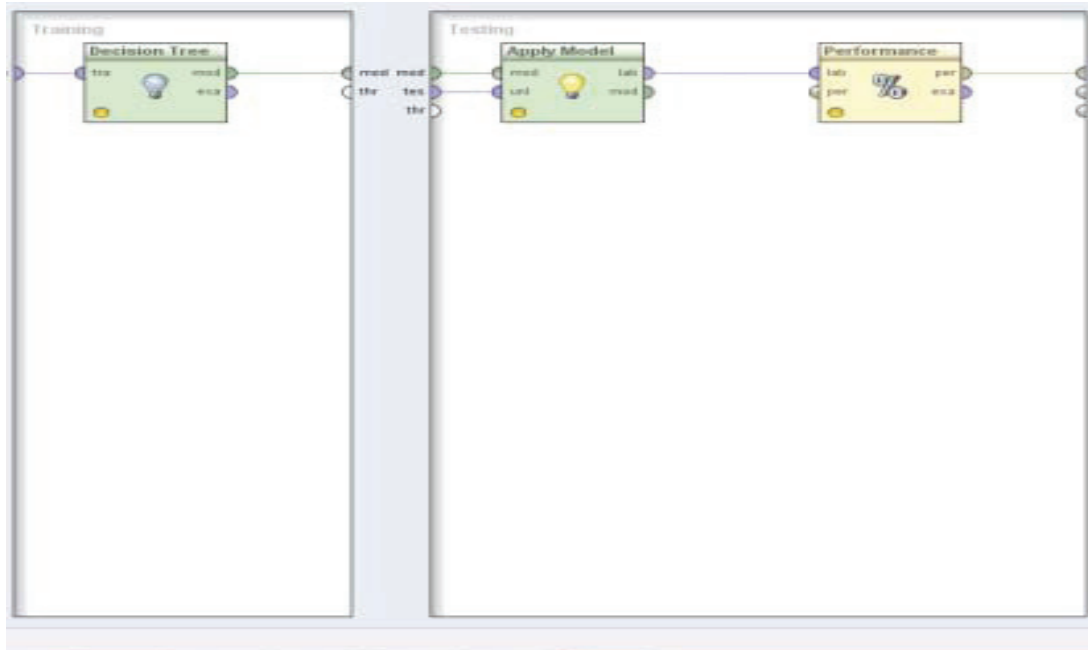


Figure 6. Decision Tree Algorithm on Rapidminer

From the modeling results obtained the following accuracy:

Table 3. Comparison of accuracy

| Naive Bayes | K-NN   | Decision Tree |
|-------------|--------|---------------|
| 82,62%      | 91,82% | 88,82%        |

| Row No. | confidence | confidence | prediction | nim        | nama        | jeniskelamin | ASAL_SEKO   | ASAL_DAER.    | jmlsks | ip/ajahir/tul | jalurujian |
|---------|------------|------------|------------|------------|-------------|--------------|-------------|---------------|--------|---------------|------------|
| 1       | 0.636      | 0.364      | YES        | 1610201001 | ATRIJUN PAI | P            | MAN TEMPE   | Prop. D.I. Yo | 60     | 3.280         | BIDIK MISI |
| 2       | 0.636      | 0.364      | YES        | 1610201002 | LULUK NUR   | P            | SMA N 1 GO  | Prop. D.I. Yo | 60     | 3.580         | BIDIK MISI |
| 3       | 0.636      | 0.364      | YES        | 1610201003 | NUR SAADA   | P            | SMA NEGER   | Prop. D.I. Yo | 60     | 3.870         | BIDIK MISI |
| 4       | 0.636      | 0.364      | YES        | 1610201004 | SUPRIYATI   | P            | SMA N 1 JET | Prop. D.I. Yo | 60     | 3.530         | PMDK       |
| 5       | 0.636      | 0.364      | YES        | 1610201005 | ASTRININGF  | P            | SMK NUHA    | Prop. Jawa 1  | 60     | 3.210         | PMDK       |
| 6       | 0.636      | 0.364      | YES        | 1610201006 | OGI OPRILIA | P            | SMA NEGER   | Prop. Jawa 1  | 60     | 3.340         | PMDK       |
| 7       | 0.636      | 0.364      | YES        | 1610201007 | ANNISA FAT  | P            | SMA N 1 SLU | Prop. Sumat   | 60     | 3.120         | TPA        |
| 8       | 0.636      | 0.364      | YES        | 1610201008 | RIZKI EDIAS | P            | SMA N 1 JEF | Prop. Jawa 1  | 60     | 2.990         | TPA        |
| 9       | 0.636      | 0.364      | YES        | 1610201009 | WIDRIA NOV  | P            | MAS PPPIMI  | Prop. Jawa 1  | 60     | 3.190         | PMDK       |
| 10      | 0.636      | 0.364      | YES        | 1610201010 | NUR HIDAY   | P            | SMA NEGER   | Prop. Jawa 1  | 60     | 3.120         | PMDK       |
| 11      | 0.636      | 0.364      | YES        | 1610201011 | CATUR RET   | P            | SMA N 5 PUJ | Prop. Jawa 1  | 60     | 3.520         | PMDK       |
| 12      | 0.636      | 0.364      | YES        | 1610201012 | NABILLAH A  | P            | SMA NEGER   | Prop. Jawa 1  | 60     | 3.710         | PMDK       |
| 13      | 0.636      | 0.364      | YES        | 1610201013 | DEVI RAHM   | P            | SMA N 13 SE | Prop. Jawa 1  | 60     | 3.240         | TPA        |
| 14      | 0.636      | 0.364      | YES        | 1610201014 | NUR ANI RA  | P            | SMK KESEH   | Prop. Banten  | 60     | 3.440         | TPA        |
| 15      | 0.636      | 0.364      | YES        | 1610201015 | ENDAH PUS   | P            | SMA YABAKI  | Prop. Jawa 1  | 60     | 3.380         | PMDK       |
| 16      | 0.636      | 0.364      | YES        | 1610201016 | SITI ROHIMI | P            | MA YASPIB E | Prop. Subwa   | 60     | 3.500         | PMDK       |
| 17      | 0.636      | 0.364      | YES        | 1610201017 | ELENA NABI  | P            | SMA NEGER   | Prop. Kalim   | 60     | 3.730         | PMDK       |
| 18      | 0.636      | 0.364      | YES        | 1610201018 | NURUL CA    | P            | SMK KESEH   | Prop. Nusa    | 60     | 3.130         | PMDK       |
| 19      | 0.636      | 0.364      | YES        | 1610201019 | BETY RINDA  | P            | SMK KESEH   | Prop. Jambi   | 60     | 3.440         | PMDK       |



**Figure 7.** Prediction results using the KNN algorithm

The graduation prediction result of the semester 2 until semester 10 students of faculty of health, nursing study program, Universitas Aisyiyah selected randomly were 100% graduated on time

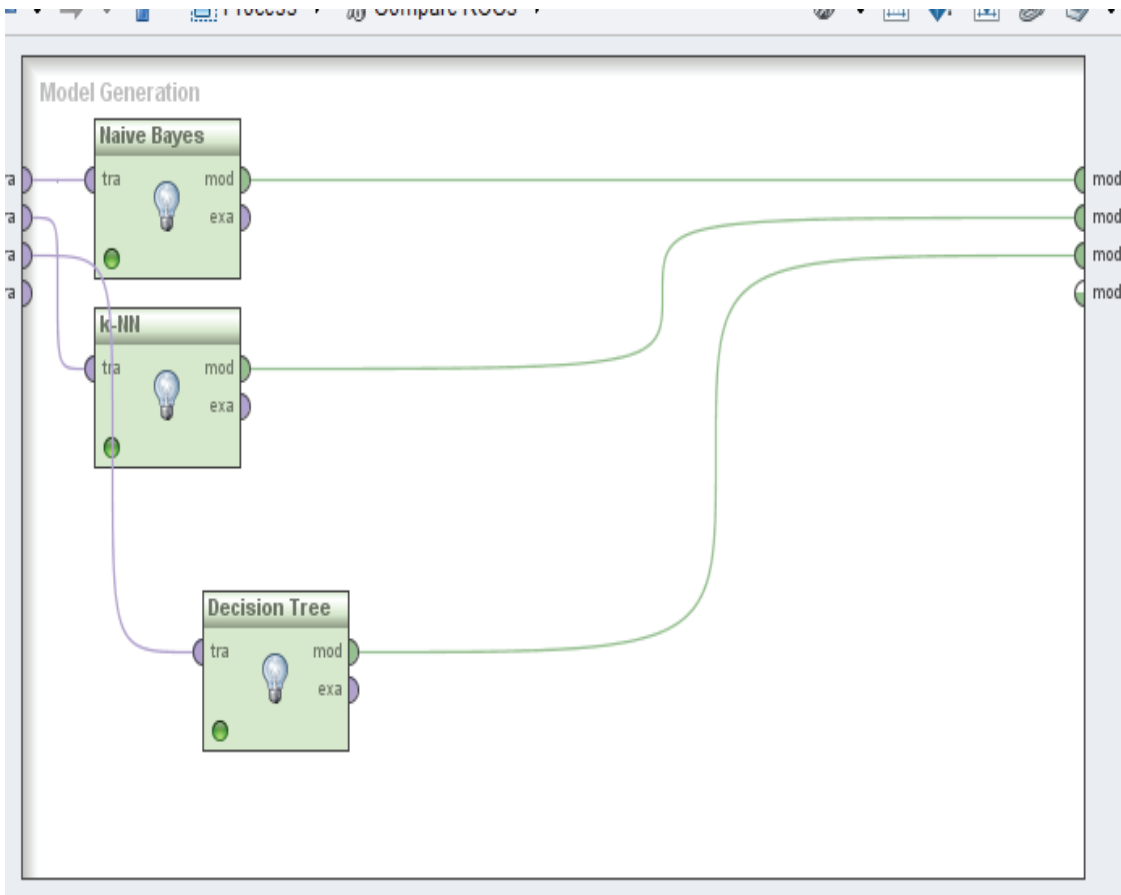
5. Evaluation

Evaluation used matrix and ROC curve configuration. The comparison of evaluations for each algorithm is shown in Table 4.

**Table 4.** Comparative Evaluation

|          | Naive Bayes | KNN    | Decision Tree |
|----------|-------------|--------|---------------|
| Presiasi | 46,01%      | 90,98% | 87,07%        |
| Recall   | 60,89%      | 53,25% | 33,39%        |
| AUC      | 0,845       | 0,899  | 0,77          |

Based on tables 3 and 4, it is obtained that the best accuracy is generated by the KNN algorithm with 91.82%, followed by the Decision tree algorithm with 88.82% and the last Naive Bayes with 82.62%. At the evaluation stage, the highest value of precision, recall and AUC was resulted by the KNN algorithm. The higher the value of precision, recall and AUC, the better the classification. The ROC curve comparison test using the rapidminer framework is shown in Figure 8.



**Figure 8.** ROC curve comparison test model on rapidminer

The comparison result of the ROC curve is shown in figure 9. It is seen in the figure that KNN curve illustrated in blue line has the highest AUC value.

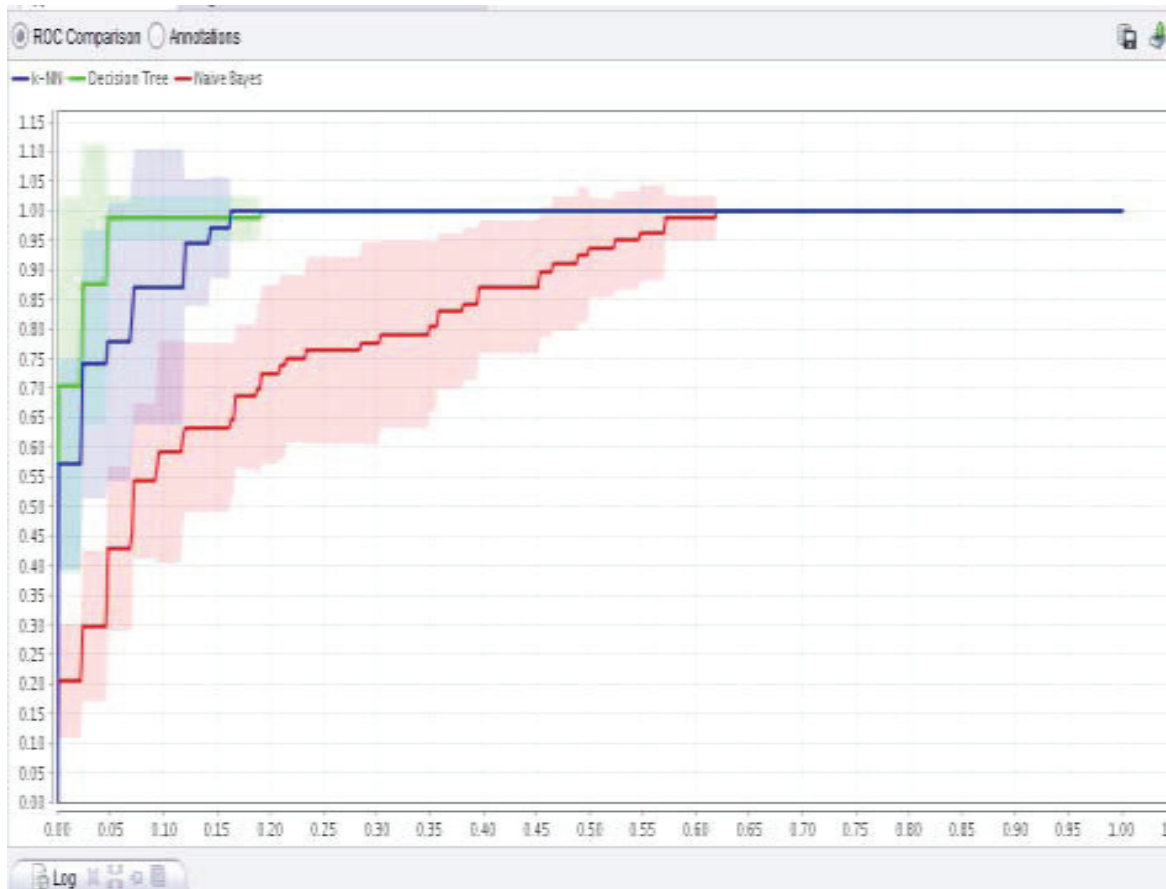


Figure 9. Comparative results of the ROC curve test

## 6. Deployment

The deployment was the final stage in making reports on data mining activities. The final report contained the knowledge gained or pattern recognition on the data in the data mining process. Based on the research conducted, a new pattern, information, and knowledge were produced in the data mining process to classify student graduation based on the data taken from the students of the Faculty of Health Sciences, Aisyiyah University, Yogyakarta. A new pattern, information, and knowledge were developed in accordance with the purpose of data mining, namely data mining calculation patterns that contain training data and testing data and search for the probability of each attribute based on the training data and testing data to produce new information.

The result of the study showed that, based on the data, more students graduated on time than those who late. To test the accuracy level, Rapidminer was used as a tool in the process of testing the accuracy of the classification. It shows that the K-NN algorithm was more accurate than the Naive Bayes and decision trees. The prediction using data testing with KNN algorithm found that the students majoring in nursing were 100% graduated on time.

## CONCLUSION

The conclusions of the study are:

1. More students of Faculty of Health Sciences graduated on time than those who

graduated later.

2. The algorithm comparison shows that the K-NN algorithm is more accurate to predict student graduation compared to Naive Bayes and decision tree that is 91.82%.
3. The result of K-NN algorithm prediction shows that the students majoring in nursing were 100% graduated on time. The data taken were random data of students who are still studying in the university.

Suggestions for further research are:

1. Adding attributes and number of datasets.
2. Using different algorithms to test student datasets to compare the best algorithms.

## REFERENCES

- A. Nadali and H. E. Nosratabadi, "Evaluating the Success Level of Data Mining Projects Based on CRISP-DM Methodology by a Fuzzy Expert System," *IEEE*, pp. 161–165, 2011.
- A. Saleh, "Implementasi Metode Klasifikasi Naïve Bayes dalam Memprediksi Besarnya Penggunaan Listrik Rumah Tangga," *Citec J.*, vol. 2, no. 3, pp. 207–217, 2015.
- A. Rakhman, "Menggunakan Metode Decision Tree Berbasis Particle Swarm Optimization ( PSO )," *Smart Camp*, vol. 6, no. 1, pp. 193–197, 2017.
- C. Catley, K. Smith, C. Mcgregor, and M. Tracy, "Extending CRISP-DM to Incorporate Temporal Data Mining of Multi- dimensional Medical Data Streams : A Neonatal Intensive Care Unit Case Study," pp. 0–4, 2009.
- D. Iskandar and Y. K. Suprpto, "Perbandingan akurasi klasifikasi tingkat kemiskinan antara algoritma C4 . 5 dan Naïve Bayes Clasifier," vol. 11, no. 1, pp. 14–17, 2013.
- D. Sartika, D. I. Sensuse, U. Indo, G. Mandiri, and F. I. Komputer, "Perbandingan Algoritma Klasifikasi Naive Bayes , Nearest Neighbour , dan Decision Tree pada Studi Kasus Pengambilan Keputusan Pemilihan Pola Pakaian," *J. Tek. Inform. dan Sist. Inf.*, vol. 1, no. 2, pp. 151–161, 2017.
- D. T. Larose and C. D. Larose, *Discovering Knowledge in Data*. 2014.
- E. Parilla-ferrer, P. L. F. Jr, and J. T. B. Iv, "Automatic Classification of Disaster-Related Tweets," in *International conference on Innovative Engineering Technologies (ICIET'2014)*, 2015, no. September.
- E. R. Paramita Mayadewi, "Prediksi Nilai Proyek Akhir Mahasiswa Menggunakan Algoritma Klasifikasi Data Mining," *Sist. Inf.*, vol. 11, no. November, pp. 1–7, 2015.
- J. Han, M. Kamber, and J. Pei, *Data Mining: Concepts and Techniques*. 2012.
- L. R. Fielitz and D. K. Scott, "Prediction of Physical Performance Using Association Rule Mining," 2002.
- M. Azmi and F. Sarmadi, "Improving the accuracy of K-nearest neighbour method in long-lead hydrological forecasting," *Sci. Iran.*, vol. 23, no. 3, pp. 856–863, 2016.
- S. Pulakkazhy, "Data Mining In Banking And Its Applications-A Review," *J. Comput. Sci.*, vol. 9, no. 10, pp. 1252–1259, 2013.
- T. Hardiani, "Segmentasi Nasabah Simpanan Menggunakan Fuzzy C Means Dan Fuzzy Rfm ( Recency , Frequency , Monetary ) Pada Bmt Xyz," *Nero*, vol. 3, no. 3, pp. 185–192, 2018.

Y. Kumar, G. Sahoo, and G. Yadav, "Predication of Parkinson's disease using data mining methods: A comparative analysis of tree, statistical, and support vector machine classifiers," *Indian J. Med. Sci.*, vol. 65, no. 6, p. 231, 2011.