Development of Student's Dropout Early Warning System Using Analytical Hierarchy Process

N Ariqah, Y Anang

Department of Statistical Computing, Politeknik Statistika STIS, Jakarta, Indonesia

*Corresponding author's e-mail: 221709883@stis.ac.id, anang@stis.ac.id

Abstract. As a higher education institution, Politeknik Statistika STIS also faces the same problems as universities in general, those are student failing to compare that year courses thus have to repeat those courses or student dropping out. To overcome this problem, this research proposes a *Dropout Early Warning System* (DEWS) that can provide early warnings for dropouts and repeat a class. With this system, it is hoped that it can help institutions to identify students who have the potential to drop out or repeat a class. The purpose of making this system is to help academic supervisors and decision makers from Polstat STIS in knowing the potential for student. The potential for students to drop out and repeat a class is measured by a potential score obtained from the results of an assessment of 5 criteria consisting of GPA scores, gender, economic factors, violation points, and record of repeating class. Prediction results are presented in three categories consisting of low potential, medium potential, and high potential which are calculated from the results of weighting calculations using the *Analytical Hierarchy Process* (AHP). The system is tested and verified using *Black Box* test and the evaluation of the calculation method using *confusion matrix*. Based on the test results, the functions that exist in the system can function properly and can supply the needs.

1. Background

Politeknik Statistika STIS (next referred to as Polstat STIS) is an official college under the control of the national statistical office of Indonesia. Students studying at Polstat STIS have the status of an official bond to the government and study assignment. Students with official bond status, after graduating will be appointed as candidate for civil servants within BPS or will be placed in other ministries or institutions according to their placement. There are three departments at Polstat STIS, namely the 3-year diploma and 4-year diploma of statistics and 4-year diploma of statistical computing.

As an educational institution, Polstat STIS needs to solve problems related to student academics. One of them is the problem of dropping out and repeating a class, where it is recorded that every year there are 5 to 10 students who are set to drop out or repeat a class from around 2200 students. Although the number of students who drop out and repeat a class is still small, it is worth paying attention to in order to avoid bigger problems in the future.

The strategy that is usually applied to reduce the number of dropout students is to arrange a good teaching pattern until provide support either in the form of moral or material. However, this strategy cannot be seen from the condition of each student. If we look at the condition of students, there are booster factors that can increase the potential of students to drop out or repeat a class. Based on research conducted by Imran, et al [1], Factors related to dropout students from 13 explanatory

variables studied in IPB university students using survival analysis produce the best model with three explanatory variables, consists of GPA, gender, and faculty. The gender factor explains that male students have a greater chance of dropping out than female students. As for Khoirunnisak and Iriawan [2] who conducted research on Institut Teknologi Sepuluh November (ITS) using Bayesian mixture survival analysis to model the factors that influence student dropouts, found significant factors, consist of age, differences in regional origin, differences in parental income, differences in faculties and entry pathways as well as GPA and new student first semester GPA. Based on the two studies that have been described, it can be taken which factors are in suitable with the conditions and environment of the Polstat STIS.

Evaluation of learning outcomes at Polstat STIS based on Director Regulation Number 2 Polstat STIS in 2021 [5], to continue the next semester has several provisions, consisting of a GPA ≥ 2.5, does not have an E value, has a value of at least C for general compulsory subjects, and has a violation score that does not exceed the specified minimum limit. If one of these conditions is not fulfilled, then class I will be declared a dropout, while class II, III, and IV can repeat at the same level with one opportunity. From all this information can be used as a strategy to reduce the number of students dropping out and repeating a class by predicting students who have this potential.

To identify students who have the potential to drop out and repeat a class, this study proposes a *Dropout Early Warning System* (DEWS). The *dropout early warning system* is a system that can help institutions to identify students who have the potential to drop out so that decision makers can take further action to the students relevant. Students who have the potential to drop out or repeat a class are likely not carefully considering whether they are doing the right thing in their academic activities. Early warnings informed by DEWS can lead students towards graduation and lead them to a better future. Because of this problem, many foreign governments have developed early warning systems for dropouts. For example, the US state of Wisconsin developed DEWS to identify student dropouts [3]. In South Korea, DEWS was also developed to help schools identify students who are at risk of dropping out and not at risk of dropping out [4]. Students who have the potential and do not have the potential to drop out can be mapped through the *dropout early warning system*. The *dropout early warning system* makes it possible to identify students who have the potential to drop out by using a decision-making method.

One of the decision-making methods is *Analytical Hierarchy Process* (AHP) method. AHP is a decision-making method developed by Saaty [7]. AHP will solve a complex and unstructured problem into groups which are then arranged into a hierarchical form so that the problem will be more structured and systematic. In a study conducted by Imanda and Andono [6], the AHP method was used as a weighting for predicting flood-prone areas in the city of Semarang. This study has the same purpose as this study, that is to identify or predict an object. So that makes AHP can be used as a weighting to predict students who have the potential to drop out or stay in class. In determining the existing criteria, the AHP method can manage input values according to criteria that have a certain weight value. By using DEWS, the system can make an assessment in the form of scores obtained from measures that affect dropout. This score is generated from the AHP process which will be categorized into 3 categories, namely low potential, medium potential, and high potential to drop out or repeat a class. The results of these scores will be displayed to the academic supervisor of the student concerned.

2. Research Purposes

The general purpose of this research is to build an information system that can assist Polstat STIS decision makers and academic supervisors in knowing the potential for dropout and repeat a class for Polstat STIS students. This system is expected to be used by academic supervisors as a reference to provide direction or guidance to students who have the potential to drop out or repeat a class.

The specific purpose to be achieved in this research are as follows:

1. Creating a prediction model for students who have the potential to drop out and repeat a class using the AHP method to produce a student potential score.

2. Building an early warning information system for dropouts and repeat a class using AHP method to generate scores that are divided into 3 categories, namely low potential, medium potential, and high potential for dropout and repeat a class.
3. Building an information system that is able to provide information on potential dropouts and repeat a class for student that can be used as student evaluation materials.

3. Research methods

3.1. Data Collection Method

In this study, the data collection methods used were interviews and literature studies. Literature study method, conducted a study on AHP and cases related to the *dropout early warning system*. In addition, interviews and literature studies were conducted to obtain the criteria used as a reference in measuring the potential for dropout and repeat a class of STIS Polstat students in this study.

3.2. Analysis Method

The analytical method used is the AHP method based on a literature study conducted by Imanda and Andono [6]. Where the AHP method will arrange a hierarchy according to the purpose of the analysis. The criteria used for calculations by AHP are factors that affect the potential of students to drop out and repeat a class.

3.3. System Development Method

The system development method used in this research is the *System Development Life Cycle* (SDLC) method with the waterfall model. The stages in the development of this system are:

a. System Planning, collecting information about the system to be built, comes from the literature study conducted.
b. System analysis, identification of systems that have been running, so that it can be apply in Polstat STIS.
c. Designing business process diagrams, designing databases, and designing interfaces.
d. Implementation of the design that has been made to the tools that suit the needs.
e. System evaluation, done after the system is complete by testing the overall functionality of the system and each component of the system.

3.4. Evaluation Method

Evaluation in system development aims to check whether the functions in the system function properly or not. The suitability of the input and results of the functions used will be tested using the Black Box test and accuracy calculations using the *confusion matrix*.

4. Results and Discussion

4.1. Proposed System Design

4.1.1. Business process of the proposed system. Figure 1 describes the proposed system's business process to solve the existing problems. For the system itself, the process is carried out by implementing AHP. The initial process begins with giving priority by using a priority questionnaire that will compare the importance of each criterion with one another. Then the priority will be obtained for each criterion. After the priority is determined, students fill in the required data. After the data is collected, the potential score for dropout and repeat a class is updated for the new semester, the value of each criterion is given a score based on the classification of each criterion and a score of potential students for dropout and repeat a class is obtained. For the evaluation process, the department coordinator gets a report that can be used as evaluation material, while students need to meet their respective academic supervisors to get an evaluation of the results of potential dropouts or repeat a class.

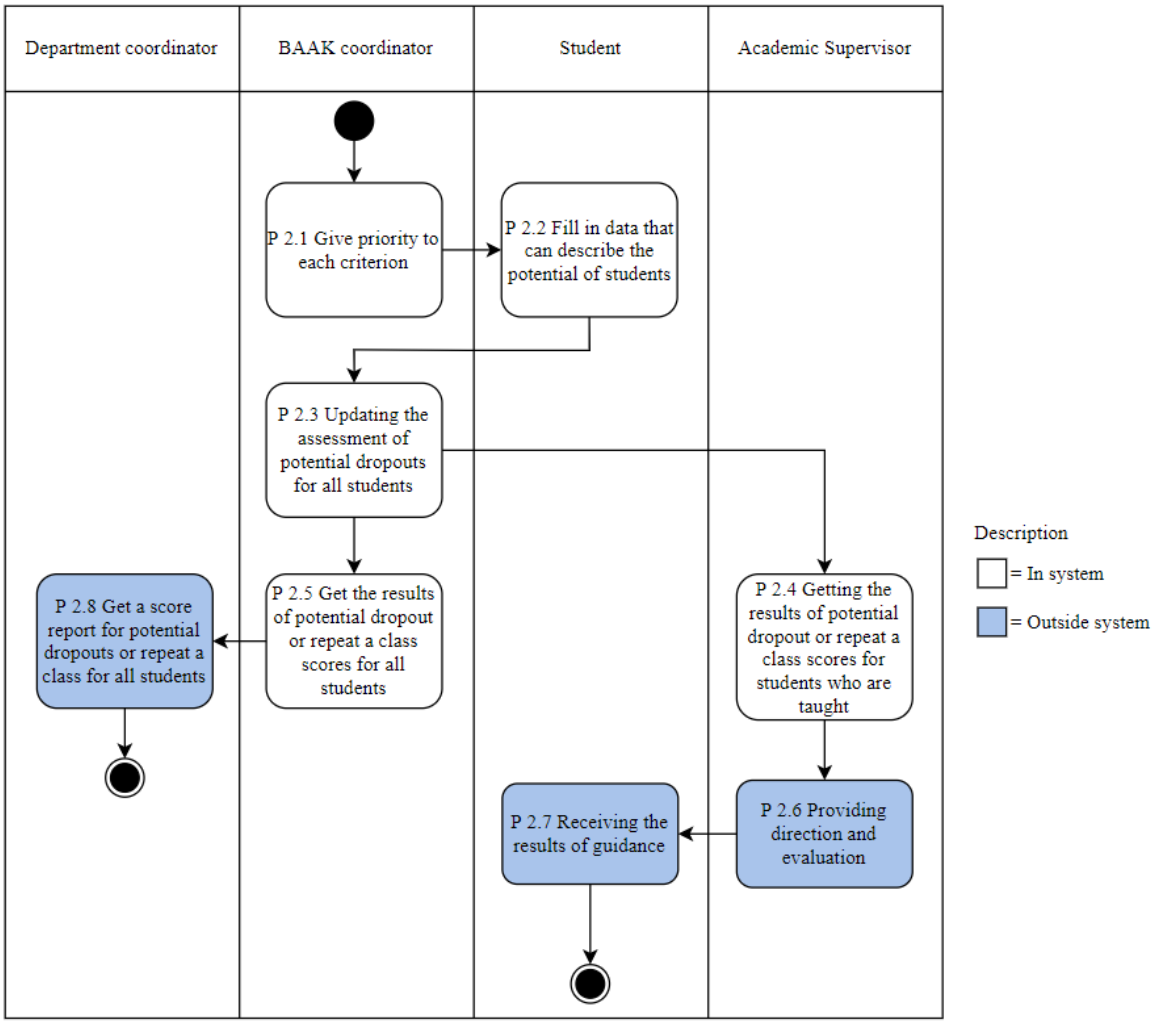

Figure 1. Business process of the proposed system.

4.1.2. Proposed system use case diagram. The diagram in Figure 2 describes the use case of the proposed system.

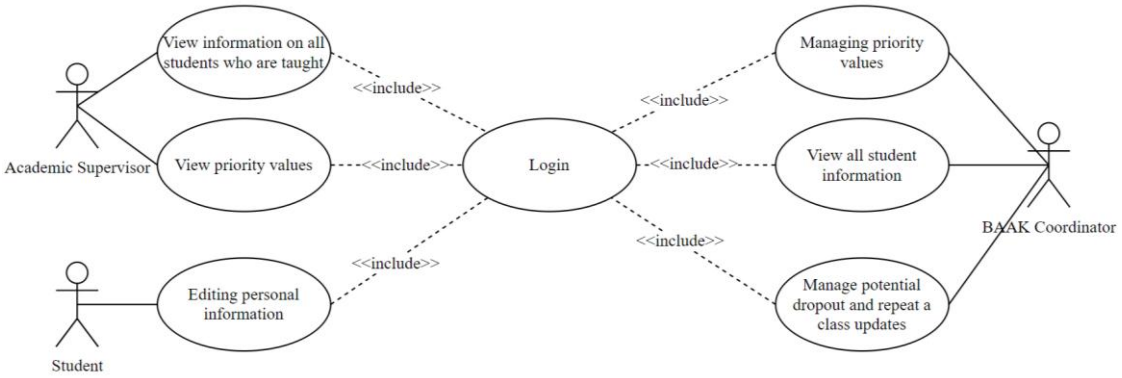

Figure 2. Proposed system use case diagram.

4.1.3. System architecture design. The early warning system is built on a web-based by utilizing a database as a medium for storing information or data. This system has three users consisting of administrators, academic supervisors, and students. For more details, see Figure 3.

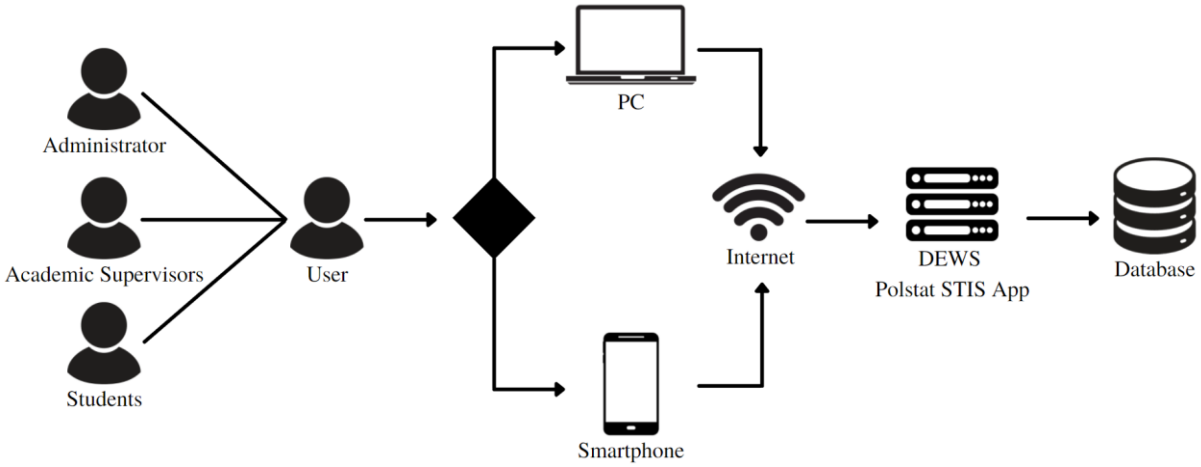

Figure 3. Proposed system architecture design.

4.1.4. Database design.

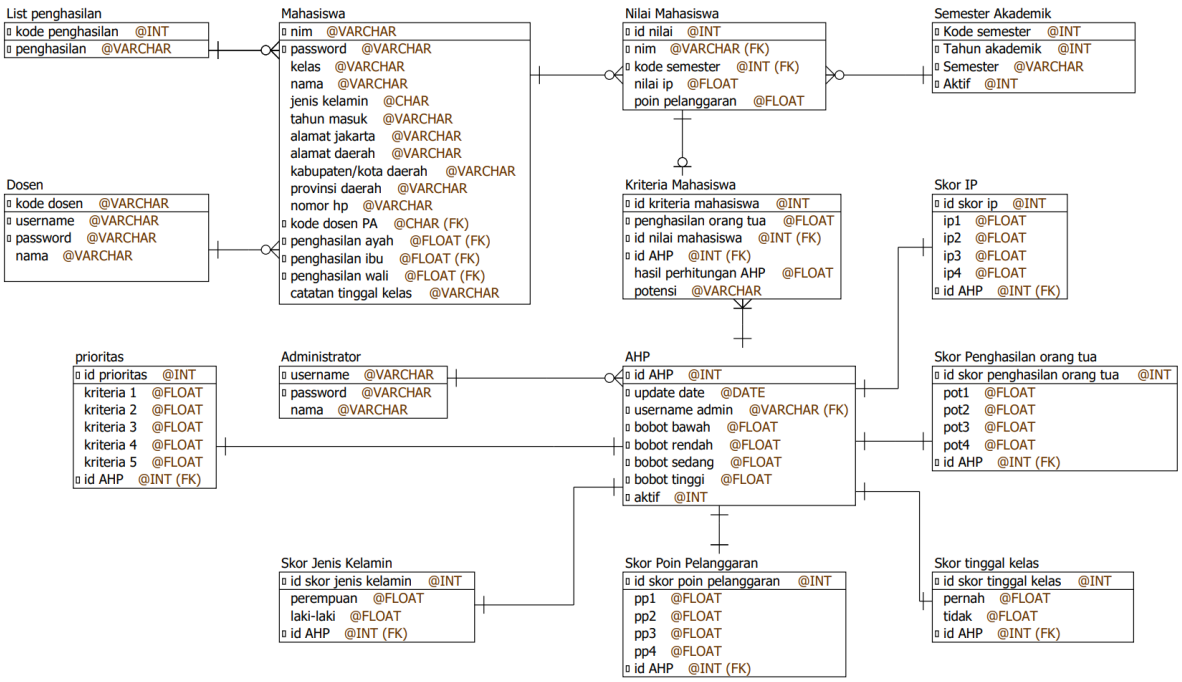

Figure 4. ERD logical system of the proposed design.

The database design that will be used in the proposed system has 14 tables as depicted in Figure 4. The `dosen` table and administrator table contain lecturer and administrator information used to log into the system. The `mahasiswa` table contains general information about students. The `list penghasilan` table contains the coding rules for the source of income of a parent or guardian of a student. The `nilai mahasiswa` table contains information about the student's grades for each semester. The `semester akademik` table contains information about the current semester or not. The `kriteria mahasiswa` table contains student score data used in calculating student potential scores using AHP and this table storing the final student potential scores. The AHP table contains general information about the current AHP method and the previously used AHP method. The `prioritas` table contains priority information for the criteria. The table of `skor jenis kelamin`, `skor IP`, `skor poin pelanggaran`, `skor penghasilan orang tua`, and

`skor tinggal kelas` is an alternative table of criteria that contains information about priorities for alternatives in each criterion.

4.2. Assessment System Implementation

The implementation of AHP in this system is in the form of a hierarchy that aims to measure the potential for students to drop out and repeat a class with the selected criteria. Based on a research literature study from Imran, et al [1] declared that GPA and gender are factors related to students dropping out of school. In addition, according to Khoirunnisak and Iriawan [2], differences in parental income and GPA are factors that affect student dropouts. In Polstat STIS, the GPA, violation points, and class stay records are factors that determine whether students are entitled to continue the semester or not [5]. Thus, the criteria that will be used in the measurement consist of GPA, gender, parent/guardian income, violation points and records of repeating class. These five criteria are then arranged in a hierarchical form with the main objective of measuring the potential of students dropping out and repeat a class. Then, the use of relative measures is included in which the alternatives are given a stratified categorization. For graded categorization for GPA criteria and violation points, it is determined based on the closest value to the provision limit for the Polstat STIS semester further and analysis of the distribution of student score data which is then grouped into 4 categories. Categories for parent/guardian income are based on categorization from research [2]. Gender categorization consists of female and male and repeat a class records consist of students who have repeated a class and students who have never repeated a class. The results of the selection of criteria for calculating student potential can be seen in Figure 5.

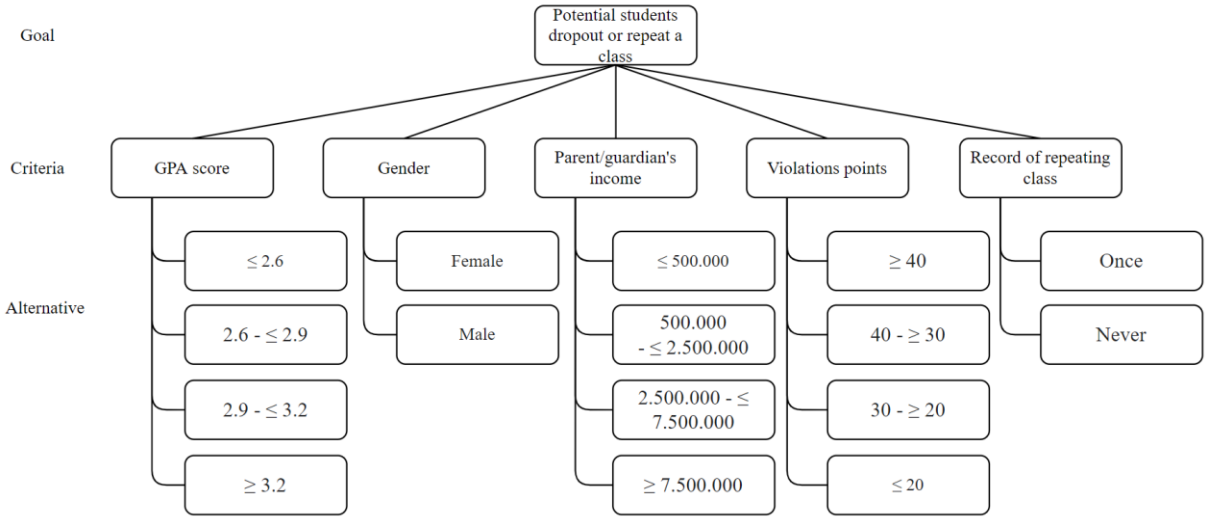

Figure 5. Implementation of AHP in measuring the potential for students to drop out and repeat a class.

The criteria assessment was carried out through pairwise comparisons. The scale used is a scale of 1 to 9 introduced by Saaty [7]. For example, in determining the priority of each criterion, suppose the following comparison is obtained:

Table 1. Example of a comparison matrix.

Criteria	GPA	Gender	Parent/guardian's income
GPA	1	7	5
Gender	0.14	1	0.33
Parent/guardian's income	0.20	3	1
Total	1.34	11	6.33

After obtaining the comparison matrix, each element of the comparison matrix will be divided according to the number of columns. The results of the division are seen in Table 2.

Table 2. Priority result from comparison matrix.

Criteria	GPA	Gender	Parent/guardian's income	Weight
GPA	0.75	0.64	0.79	0.73
Gender	0.10	0.09	0.05	0.08
Parent/guardian's income	0.15	0.27	0.16	0.19
Total	1	1	1	1

The results from Table 2 itself have produced the weight of the criteria by taking the average of each row. After getting the weight values, the next step is to check the consistency of the existing comparisons. The consistency check begins by adding up the product of the values in each column of the comparison matrix table with the weighted values in each row, in order to get λmax value.

$$\lambda max = (1{,}34 * 0{,}73) + (11 * 0{,}08) + (6{,}33 * 0{,}19) = 3{,}06 \quad (1)$$

Then calculate the Consistency Index (CI) and Consistency Ratio (CR) as follows:

$$CI = \frac{(\lambda\_max - n)}{(n-1)} = \frac{(3{,}06 - 3)}{(3-1)} = 0{,}03 \quad (2)$$

$$CR = \frac{CI}{RI} = \frac{0{,}03}{0{,}58} = 0{,}05 \quad (3)$$

From the CR value above, the value of 0.05 is obtained. With a CR value <= 0.1, then the above priorities are consistent.

After obtaining the priority value for each criterion, the next step is setting priorities for the alternatives for each criterion. The steps for determining priority values for alternatives are no different from determining priorities for criteria. After all priorities have been obtained, the next step is to categorize the value of each student into alternative value limits for each criterion. Based on the calculation of the AHP method with the following equation:

$$\sum_{i=1}^{n} X_i = W_i * V_i \quad (4)$$

Description:
$\sum_{i=1}^{n} X_i$ = potential dropout or repeat a class value
W_i = Criteria-i weight
V_i = Criteria-i alternative weights
n = number of criteria

So that the results of the combined sum of these values can be classified based on the potential level of dropout and repeat a class. Classification is done by dividing the assessment into 3 categories, namely low potential, medium potential, and high potential.

4.3. Information System Implementation

The implementation of the system is carried out using PHP programming language with CodeIgniter framework. As the data storage, the system developed uses MySQL. The following is a partial implementation of the Polstat STIS *Dropout Early Warning System* interface.

4.3.1. Login page. The login page interface is divided into three, consisting of logins for academic supervisor, BAAK coordinator (administrators), and students. The implementation of the interface for the academic supervisor login page can be seen in Figure 6.

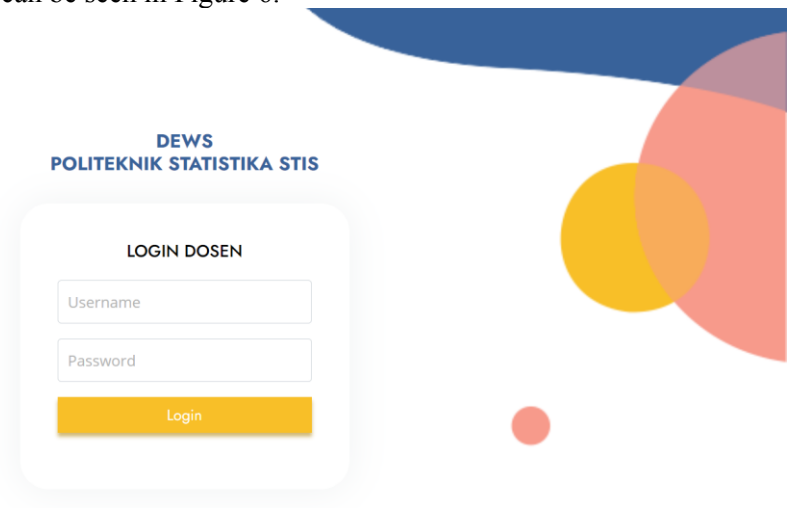

Figure 6. Implementation of the academic supervisor login page interface.

4.3.2. Home page. The implementation of the interface for the academic supervisor's home page shows the accumulation of students who are taught based on the category of potential dropouts and repeats a class. Just like the home page of the academic supervisor, the home page for the BAAK coordinator is an accumulation of all Polstat STIS students. The implementation of the interface for the home page of academic supervisor can be seen in Figure 7.

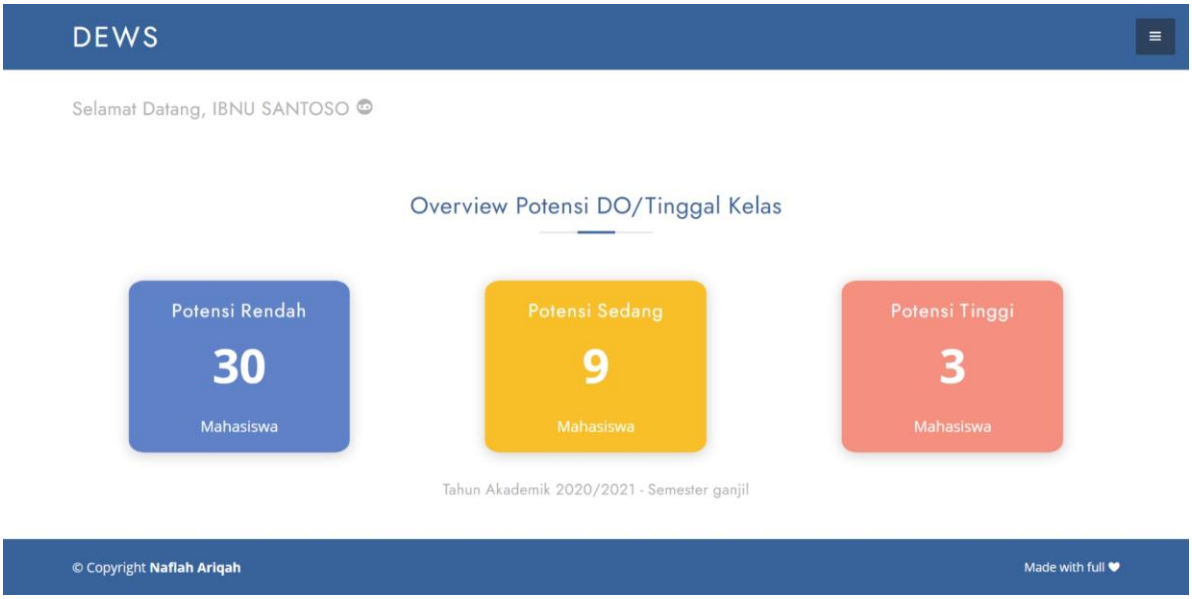

Figure 7. Implementation of the academic supervisor's home page interface.

The student home interface shows the main profile of the student and information about the grades associated with that student. Figure 8 shows the implementation of the interface for the student home page.

Figure 8. Implementation of the student home page interface.

4.3.3. AHP assessment edit page. The priority edit page in Figure 9 serves to determine the priority value of each criterion. Next, there is one more form to determine the alternative priority of each criterion. The results of this AHP assessment form will have a major influence on the final grades of potential dropouts and repeats a class student. This page can only be accessed by the BAAK coordinator or the system administrator.

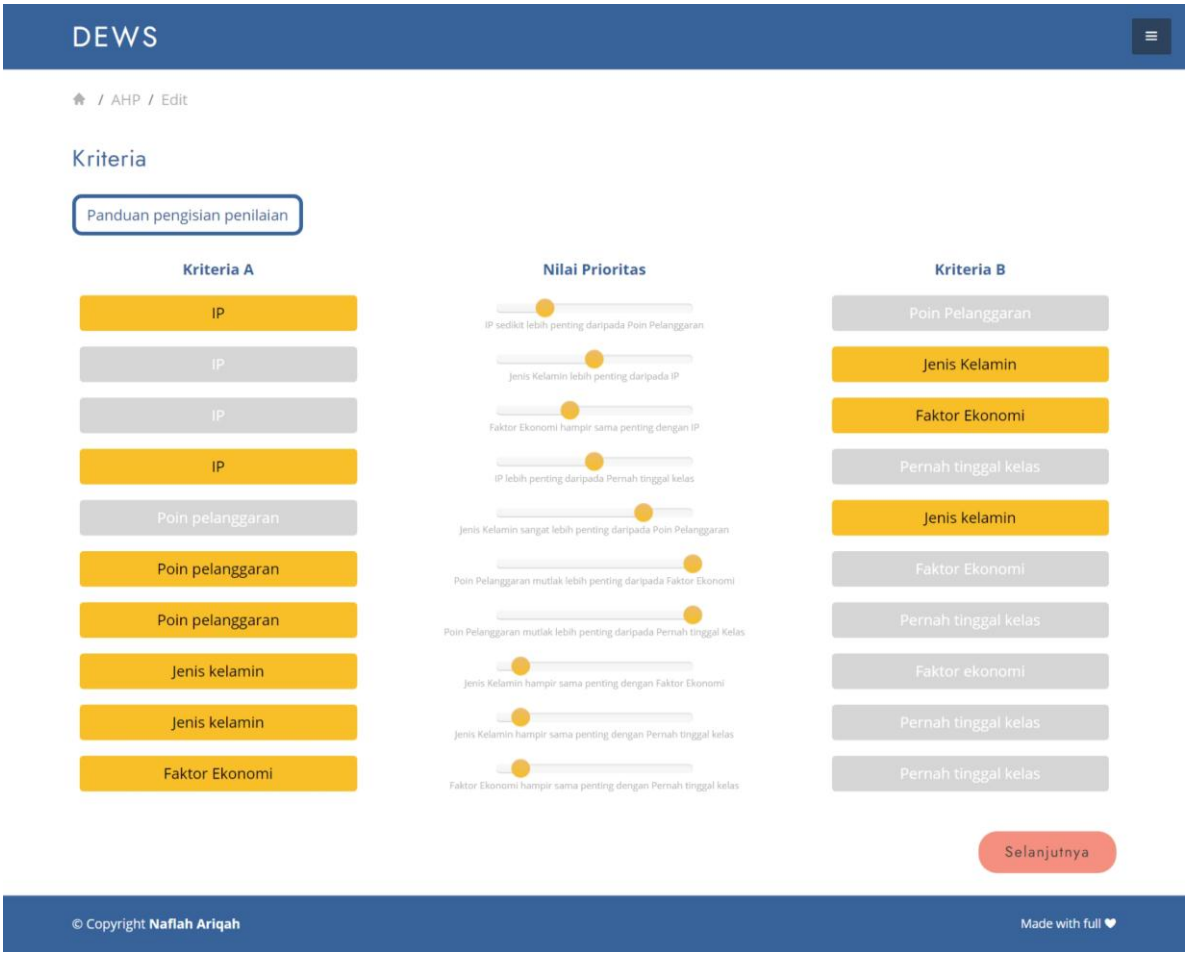

Figure 9. Priority edit page interface implementation.

4.4. System Testing and Evaluation

Testing of this system is done by means of black box testing. This test is intended to test whether the functions that exist in the system can function properly or not. Table 3 shows the results of black box testing.

Table 3. Testing using black box testing.

Test Case	Succeed	Not successful
Users (BAAK Coordinator, students, and academic supervisors) log in according to the username and password contained in the database and are directed to the homepage of each user	✓	
Admin changes priorities for criteria and alternatives from criteria and the information is stored and immediately updated for priorities in academic supervisor and administrator	✓	
Admin updates the results of the potential dropout and repeat a class assessment for all students every semester	✓	
Users (BAAK Coordinator, students, and academic supervisors) log out and are redirected back to the login page of each user	✓	

Test Case	Succeed	Not successful
Students make changes to their profile information	✓	
Academic supervisors see detailed information from their students	✓	

Furthermore, to see which method of prioritizing criteria has better performance in predicting potential dropouts and repeat a class for students, an evaluation is carried out using a *confusion matrix* in terms of accuracy, precision, recall, and F-measure. In this study, 2 models of prioritization of criteria will be used. Model 1 with GPA criteria has the highest priority, then the second priority is the violation points and gender, the third priority is the parent/guardian income, and the fourth priority is the record of repeating class. Then for model 2 GPA criteria with the highest priority, then the second priority is the gender criteria, record of repeating class, parent/guardian income, and violation points. The sample data used are 426 which are the values of Polstat STIS students in the 2016 entry year

Table 4. *Confusion matrix* prioritization of model criteria 1.

	True Value	
Predicted Value	True	False
True	20	33
False	4	369

Table 5. *Confusion matrix* prioritization of model criteria 2.

	True Value	
Predicted Value	True	False
True	8	8
False	16	394

Based on the evaluation using the *confusion matrix*, the value of accuracy, precision, recall, and F-measure can be obtained, the results can be seen in Table 6.

Table 6. Comparison of compilation method performance accuracy.

Model	Accuracy	Precision	Recall	F-measure
1	0.9131	0.3774	0.8333	0.5195
2	0.9438	0.5000	0.3333	0.4000

From Table 6, it can be seen that the two models of prioritizing criteria have quite different performance. Compared to model 2, the values of accuracy, precision, recall, and F-measure in model 1 are still better. Although the accuracy value of model 2 is higher, it is not much different. Because the reference in this study is to choose a model that has a high recall value and accuracy. High recall means that the model is better at predicting positive students who drop out or repeat a class who actually do not than the model that incorrectly predicts students who do not drop out or repeat a class when in fact the student drops out or repeats a class. In other words, it is more undesirable for students who actually have the potential but are wrong in their predictions. This is in accordance with the purpose of this study, which is to build an early warning system for dropouts and repeat a class, therefore the speed of information about students who have the potential is prioritized in providing early warning to students and accuracy is the second priority in model selection.

In a study conducted by Lee S [4] who also developed DEWS with precision and recall evaluation methods, produces an accuracy value of 0.898 using a boosted decision tree. Although the accuracy value generated from this study shows better performance, it does not mean that the calculation model of this study is more optimal. The number of data samples used by Lee S is larger as many as 165,715 students, while in this study only 426 data samples were used. The comparison of the number of data samples that far can also affect the performance between models. The results of these predictions can still generate false warnings so that it requires good policies or interpretations from decision makers so that the prediction results can be used properly.

5. Closing

Based on the research that has been done, several conclusions are obtained, that is the author has succeeded in making a predictive model using the AHP method which produces predictions of students who have the potential to drop out and repeat a class by producing a student's potential score. From the model created, it has been successfully applied to the developed information system. The dropout early warning system produces a score that is divided into 3 potential categories, namely low potential, medium potential, and high potential. Potential information generated by the system can be used as evaluation material by the Polstat STIS decision makers and academic supervisor.

There are some limitations in this study, which can be used as material for further research. First, system integration with Sipadu which is the Polstat STIS academic management system. Integration is done to make it easier for users to use this system and make it easy to collect the required data in the calculation of this system. Second, improve system performance so that it can dynamically provide criteria. The system can be improved by adding these additional features in order to be flexible in following changes in criteria selection. Third, strengthen the accuracy of the assessment both in terms of prioritization or selection of criteria to be more precise in predicting students who have the potential to drop out and stay in class. Suggestions for the selection of criteria can see the effect on PMB test scores, differences in regional origin and percentage of attendance. PMB test scores can be used as one of the criteria in assessing potential dropouts for new students.

References

[1] Imran F, Susetyo B and Wigena A H 2013 Identifikasi Faktor-Faktor yang berhubungan dengan Mahasiswa Putus Kuliah di IPB Angkatan 2008 menggunakan Analisis Survival *Journal of Statistics* **2** 1–6

[2] Khoirunnisak M and Iriawan N 2010 Pemodelan Faktor-Faktor yang Mempengaruhi Mahasiswa Berhenti Studi (Dropout) di Institut Teknologi Sepuluh November Menggunakan Analisis Bayesian Mixture Survival *Jurnal FMIPA Universitas Sepuluh Nopember* 1–9

[3] Knowles J E 2015 Of Needles and Haystacks: Building an Accurate Statewide Dropout Early Warning System in Wisconsin *Journal of Educational Data Mining* **7** 18–67

[4] Lee S and Chung J Y 2015 The Machine Learning-Based Dropout Early Warning Sistem for Improving the Performance of Dropout Prediction *Applied Sciences* **9** 3093

[5] Politeknik Statistika STIS 2020 *Peraturan Direktur Politeknik Statistika STIS Nomor 002 Tahun 2020* (Jakarta: Politeknik Statistika STIS)

[6] Imanda R I and Andono P N 2015 *Pemanfaatan Sistem Informasi Geografis dengan Metode Analytical Hierarchy Process (AHP) untuk Prediksi Daerah Rawan Banjir di Kota Semarang* (Semarang: Universitas Dian Nuswantoro)

[7] Saaty T L 2004 Decision making-the analytic hierarchy and network processes (AHP/ANP) *Journal of Systems Science and Systems Engineering* **13** 1-35