

CHAPTER 4

ANALYSIS AND DESIGN

4.1 Analysis

This project is made using the PHP programming language because PHP applications at Soegijapranata Catholic University mostly use website-based applications with PHP programming. This project uses Mysql database for saving the training data and test data. Data in Mysql is easily processed because it will be included in the two-dimensional array data structure. Server at Soegijapranata Catholic University also uses Mysql as data storage in Sintak application. This project uses the Naïve Bayes algorithm because Thomas Bayes as the inventor scientist of Bayes theorem said prediction about the future opportunities are based on previous experiences.

Bayes Theorem has the general calculations as follows:

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

X = Data with unknown classes

H = Hypotesis of X is a specific class

P(H|X) = Probability of H hypotesis is based on condition X

P(H) = Probability of H hypotesis

P(X|H) = Probability X is based on the condition

P(X) = Probability of X

In this project, Naive Bayes uses three parameters, namely GPA, parents' income, and students who are active in organization. The Naïve Bayes method will divide the data into two clusters which are eligible and ineligible with addition of rank based on the calculation of the value results

will be “*layak*”. If the results from the input value of “*layak*” getting higher, so the data will appear first as the priority data. This project uses the GPA range as follows:

Table 4.1: Table of Classifying the GPA Students

Category	Min	Max
I1	0	2.75
I2	2.76	3.25
I3	3.26	3.50
I4	3.51	4.00

Classification the GPA is intended to classify the training data and test data. Parents’ income is grouped based on the total salary of both parents. Categorizing the total salary of parents as follows:

Table 4.2: Tabel of Categorizing the Total Salary of Parents

Category	Min	Max
P1	0	999.999
P2	999.000	2.899.999
P3	2.900.000	7.899.999
P4	7.999.999	~

Parents’ salary is calculated by adding father’s and mother’s salary as input from scholarship’ applicants with the following formula:

$$\text{Total Salary} = \text{Father's salary} + \text{Mother's salary}$$

Certificate of students’ activity is also used as a determining factor. It has the following calculations:

Table 4.3: Tabel of Categorizing Data based on Number of Certificate

Category	Note	Amount
S1	Tidak Ada	0
S2	Sedikit Ada	1 – 6
S3	Banyak	> 7

From table 4.3 explained that this project will calculate each incoming certificate file to count the number of certificates for each data.

The calculation of each attribute will influence the results of Bayes theorem. The results of each attribute of each class will be multiplied to get the final result which later will determine the data entry into what class, eligible or ineligible. The example of calculation in the following case:

Table 4.4: Tabel of Test

IPK	Gaji Ayah	Gaji Ibu	Jumlah Sertifikat
3.84	Rp. 2.000.000	Rp. 1.500.000	8

Database : 368 Data

Database with Accepted class : 144

Database with Rejected class : 224

Total Salary = Father's salary + Mother's salary

Total Salary = 2.000.000 + 1.500.000 = 3.500.000

Calculation each attribute in Accepted class (*Layak*)

$$P(I4 | Layak) = 39 / 144 = \mathbf{0,2708333333}$$

$$P(P3 | Layak) = 53 / 144 = \mathbf{0,3680555556}$$

$$P(S3 | Layak) = 11 / 144 = \mathbf{0,0763888889}$$

Calculation each attribute in Rejected class (*Tidak Layak*)

$$P(I4 | Tidak Layak) = 50 / 224 = \mathbf{0,2232142857}$$

$$P(P3 | Tidak Layak) = 122 / 224 = \mathbf{0,5}$$

$$P(S3 | Tidak Layak) = 12 / 244 = \mathbf{0,0491803279}$$

Calculation of probability in each class

$$P(\text{Layak}) = P(I4 | Layak) \times P(P3 | Layak) \times P(S3 | Layak)$$

$$P(\text{Layak}) = 0,2708333333 \times 0,3680555556 \times 0,0763888889$$

$$\mathbf{P(\text{Layak}) = 0,0076145753}$$

$$P(\text{Tidak Layak}) = P(I4 | Tidak Layak) \times P(P3 | Tidak Layak) \times P(S3 | Tidak Layak)$$

$$P(\text{Tidak Layak}) = 0,2232142857 \times 0,5 \times 0,0491803279$$

$$P(\text{Tidak Layak}) = 0,0054888759$$

Calculation of probability in each class shows class probability value of “*Layak*” is bigger than probability class of “*Tidak Layak*”; therefore, this decision data goes to the class of eligible (*layak*) to receive the Sandjojo Foundation Scholarship.



4.2 Desain

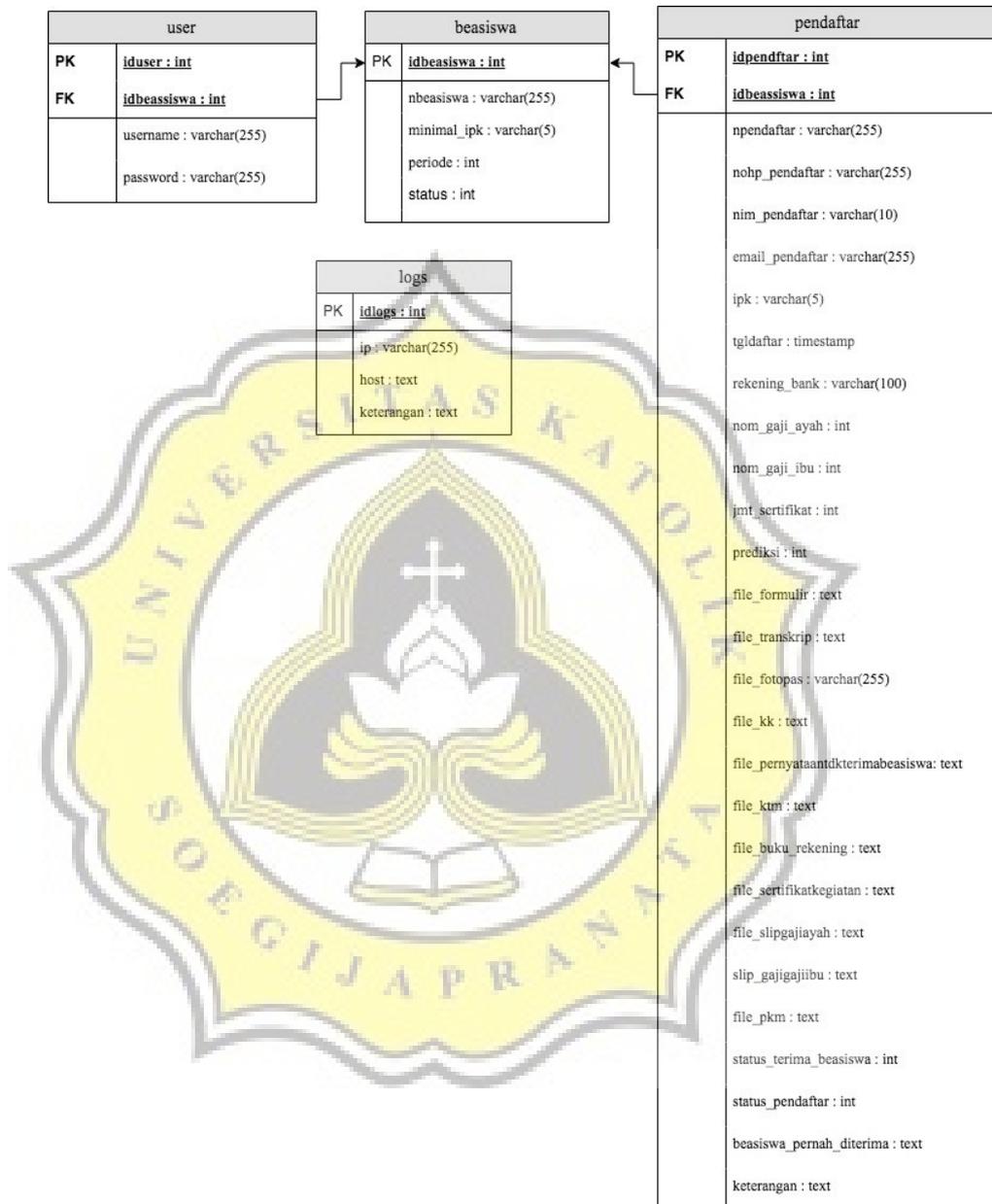


Illustration 4.1: Database Design

This project database uses four tables that have mentioned above in Illustration 4.1. First, user table is used for saving the username data and password for login authentication. Second, scholarship table is used for saving the existing data of scholarship at Soegijapranata Catholic University. Third, registrants table

is used for saving the registrants data. And the fourth, logs table is used for saving the process data which carried out by the user.

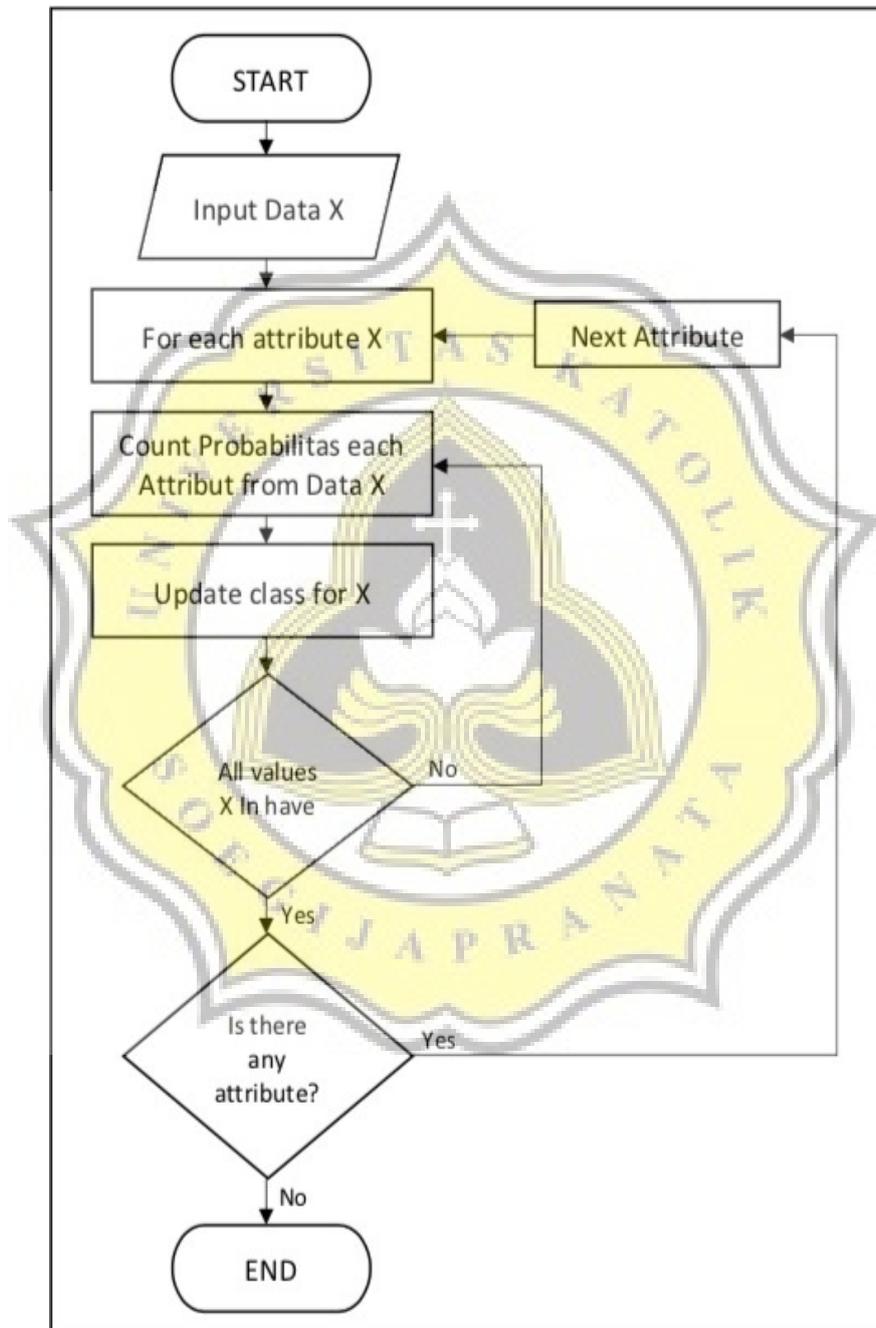
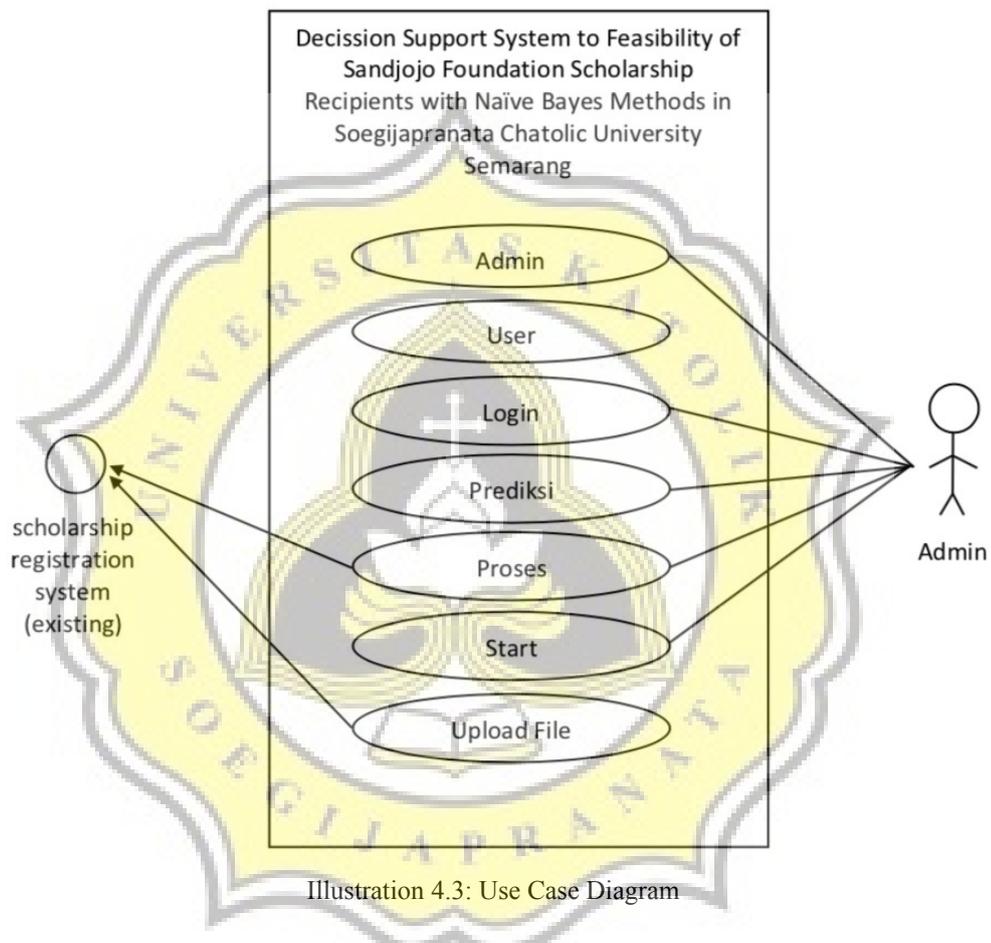


Illustration 4.2: Application Flowchart

This project flowchart can be seen from Illustration 4.2 above. First input data x is who student submit the scholarship ti UPT with 3 attribute GPA, Parent

Salary, and Certificate, then all of attribute will calculate in each attribute with a loop in every attribute; therefore, if there are three attributes in one data, then the program or probability calculation will repeat three times until all the attributes are fulfilled.



In this project is described on the Illustration 4.3 above, shows the Registrants System is accessed to the Process class and Upload File class is used for entering the data. Admin can access the Admin class, Login class, Prediction class, and Start class. Prediction class is only for Help class when testing the calculation accuracy of Naïve Bayes.

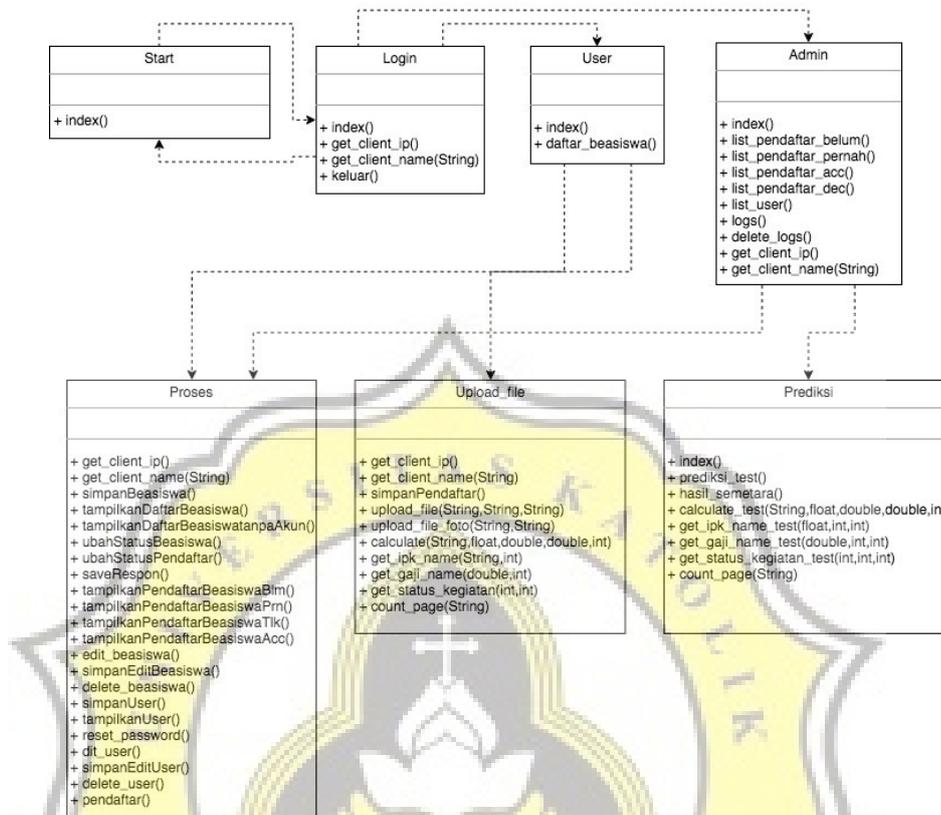


Illustration 4.4: Class Diagram

In Illustration 4.4 explains that this project has seven classes which consists of Start class, Login class, User class, Admin class, Process class, Upload File class, and Prediction class. Prediction class is as a class which used for testing the accuracy of this project and can only be accessed by admin. The use of case diagram is translated in the form of a Class Diagram. In Class Diagram, Admin become Admin class, User become User class, Login become Login class, *Prediksi* become Prediction class, *Proses* become Process class, Start become Start class, and the last, Upload File become Upload File class.

The wireframe shows a login page layout. On the left side, there is a vertical column with the text "Step to Register Yayasan Sandjojo Sholarship". On the right side, there is a box containing the "Unika Soegijapranata Symbol". Below the symbol are two input fields: "Input Username Field" and "Input Password Field". Below these fields are two buttons: "Login Button" and "Show SOP Button". A large, faint watermark of the Unika Soegijapranata logo is visible in the background of the wireframe.

Illustration 4.5: Design Login Page

From Illustration 4.5 above shows the symbol of Soegijapranata Catholic University is located above the page because it is used for showing the ownership of the application. Input Username Field is used to enter the username for login and Input Password Login is used to enter the login password. Login Button is used to enter the application and Show SOP Button is used to show the *SOP (Standar Operasional Prosedur)* registration for Sandjojo Foundation Scholarship. On the left side is the Step to Register the Sandjojo Foundation Scholarship. It shows the steps that must be passed by the registrants in order to register as the registrants of Sandjojo Foundation Scholarship.



Illustration 4.6: Design Content Page

In page of content consists of Title Page displays the real time, Bar Menu on the left side contains menu for moving pages, Navigate shows the page location, Content displays the content of the page, Username Login shows the login session for users, Profile Picture shows the Login picture of user, Logout Button is used to delete the login session, exit the content page, and also return to the main page that is Login Page.

4.3 Scenario Testing

This project will go through the functional test stage, performance test, and calculation accuracy test. The functional test of this project produced a compatibility of design program with output results from this project in each function of buttons and display. Functional test will be done by accessing each page to ensure each display is in accordance with the initial design.

Performance test is used to test how long it will take to get the value result of calculation using Naïve Bayes method. By adding a few lines of code to calculate the time needed for getting the accuracy results.

Accuracy test on this project used different number of test data and database which are tested in 4 times. It is possible to produce different accuracy. By taking 2016 data as the training data and 2017 data as the test data. In each test, there will be a reduction in training data and test data to test the accuracy of each incoming data.

