

CHAPTER 3

RESEARCH METHODOLOGY

The stages of the research are a series of sections that describe the flow of research in achieving certain goals, in this research is to create an expert system for diagnosing diabetes mellitus.

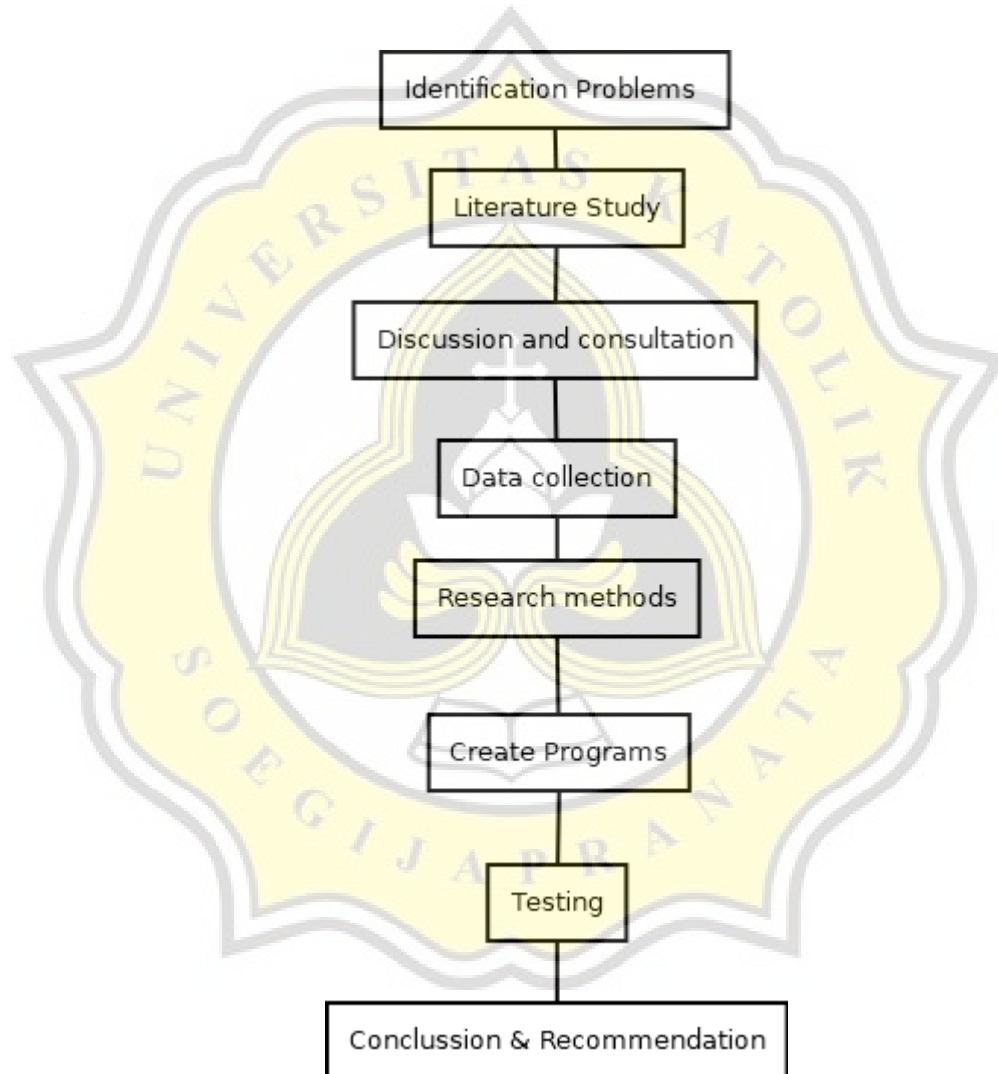


Figure 3.1 Research Methodology

3.1 Identification Problems

The first stage in this research project begins with identifying problems regarding diabetes mellitus regarding the characteristics of people affected by diabetes and types of diabetes mellitus.

3.2 Literature Study

The second stage in this research is a literature study which functions to gather information and deepen issues related to the project to be made. The literature study obtained can be in the form of journals, books, and other references.

3.3 Discussion and consultation

This project requires discussions and consultations with various parties so that this research can run smoothly and successfully.

3.4 Data collection

The data collection stage was carried out using data collection techniques from articles and Kaggle journals regarding diabetes mellitus. This is the link I used to download the dataset available on kaggle

<https://www.kaggle.com/datasets/akshaydattatraykhare/diabetesdataset?resource=download>. This dataset is from Akshay Dattatray Khare which was updated 4 months ago. In this dataset there are 9 columns consisting of BloodPressure, SkinThickness, Insulin, BMI, DiabetesPedigreeFunction, Age, and Outcome. There are 769 data in each column.

3.5 Research Methods

At this stage it discusses what algorithm will be used to solve this project problem, at this stage it uses the Naive Bayes method to test how effective this algorithm is for detecting diabetes.

3.6 Create Programs

After doing Research Methods the next step that must be done is to make a program. Before we make the program, we must first calculate the Naïve Bayes formula

in Microsoft Excel to find probability values from the overall probability value to the probability value for each variable. After knowing the probability value, we test it with data testing and see the results. We will match these results with the source code of the program that we will make, whether the results are the same or not.

Furthermore, here I make a program to implement the Naïve Bayes method in calculating diabetes mellitus. My first step is to make the source code for the initial display when we access the page. The title is "Diabetes Mellitus Prediction System". Then there is a form that contains data that must be filled in by the user if you want to predict whether the user has diabetes or not. The form consists of the number of pregnant, the amount of glucose, blood pressure, etc. When the user has inputted and the user clicks the prediction button, the user will automatically be directed to the next web page which contains data calculated from the Naïve Bayes method.

Next, I created a program that contains calculations for the Naïve Bayes method for detecting diabetes from the data that has been input by the user. Here I connect my php file to a database that has been stored in mysql. Then I describe variable 0 as normal, and variable 1 as sick. Next I created a variable with the name norm and in that variable I entered the `mysqli_query` function to run the query command to the database, I then created a normal variable in that variable I added a function for `mysqli_num_rows` to find out how many rows the normal variable is in the diabetes database table with name diabetes mellitus I did the same steps to find the number of sick variables in the diabetes database table with the name diabetes mellitus.

Next, I also look for the number of columns from x1-x8 that are included in the variable 0 or 1. Next, I create a variable called has1 which contains the sum of all normal variables and has2 which contains the sum of all sick variables. Next, the if else formula is entered with the condition that if $has1 > has2$ then the result that comes out is normal or does not have diabetes.

I repeated the same steps to make the program and test the performance of Naïve Bayes. By testing it against certain conditions. Detailed testing will be explained in the testing sub-chapter and its application will be explained in chapter 4.

3.7 Testing

At this testing stage, the data that has been collected is tested in this project. I tested it into 3 stages of the testing section.

In the first stage, I tested the Naïve Bayes method with too few data cases. The existing dataset is used as training data and then testing data is entered and tested with the least amount of data used as testing data. At this stage I tried to test using 580 testing data, 384 testing data, and 50 testing data then seeing the value of accuracy, precision, and recall. From the first experiment, conclusions were drawn about the influence of the small amount of data in finding the value of accuracy, precision, and recall.

In the second test, I tested the Naïve Bayes method with cases using old test data and new testing data. What is meant by old testing data is data that is already in the training data used as data testing while new data is data that is not in the training data. In this test I used 50 old test data and 50 new test data. The results of accuracy, precision, and recall will appear and conclusions will be drawn regarding the impact or influence of the old testing data and the new testing data.

And in the last test I tested the Naïve Bayes method by looking for the dominant factor. By eliminating the x1-x8 variables and observing the impact of the changes that occur with the reduction of the variables from x1-x8, after the results of accuracy, precision, and recall are recorded, conclusions can be drawn regarding the dominant factor.

3.8 Conclusion & Recommendation

When the results have been obtained, conclusions can be drawn, and suggestions for future research.