# CHAPTER 3 Research Methodology



## RESEARCH DESIGN

Presents the design to which the study was patterned (i.e. Descriptive, Normative, Correlational Study, 2-Pair Group Experimental Design, etc.

## SOURCES OF DATA

Presents the entire population from which the representative group was extracted.

Description of Respondents. It paints a picture of the characteristics of the research participants to whom data was obtained.

Presents the technique utilized in the selection of samples. (Sampling Technique)

## INSTRUMENTATION

## SOFTWARE/HARDWARE TOOLS

### SYSTEM ARCHITECTURE

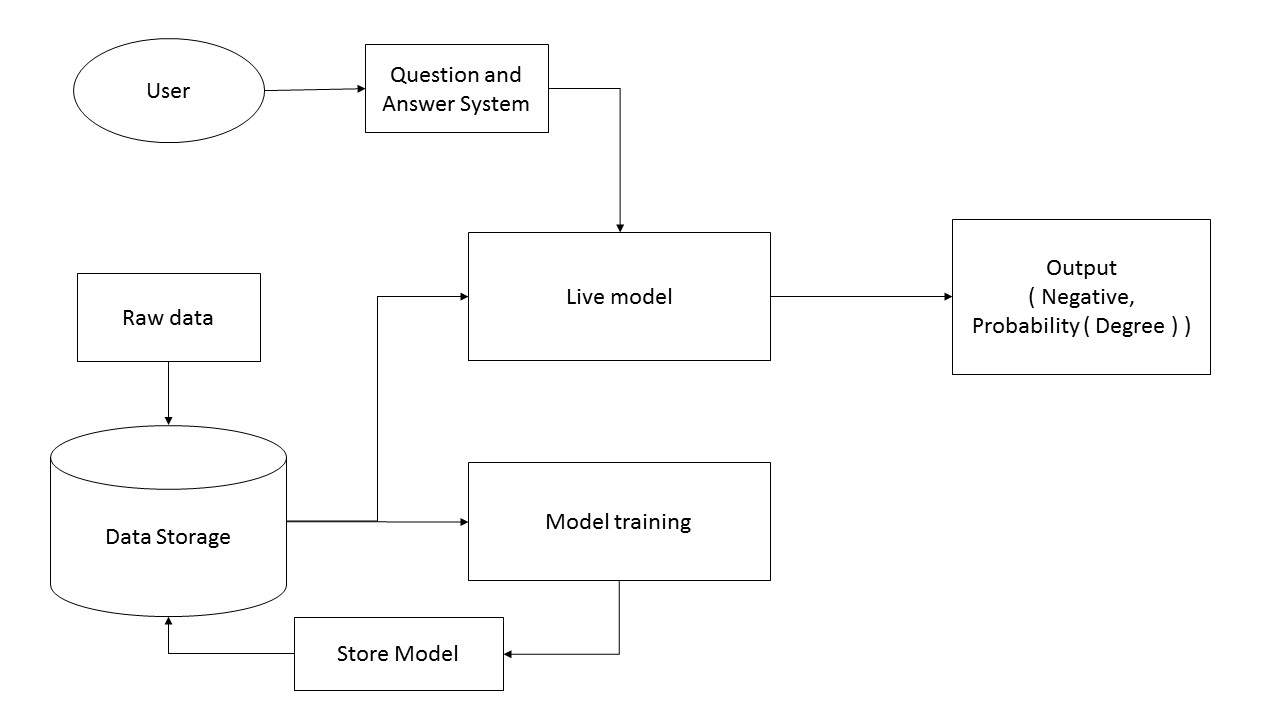


Figure # explains the flow of the system architecture of Celiac Disease Diagnosis Using Decision Tree with Generated Boosting. It begins with user answering the question of the system then the data will pass to the live model then the model will output the result if it is negative or if the result is positive then the system will output the level of the disease. Raw data is consisting of training data and test data and it is stored in the data storage. The raw data is used for training and improving the model which. The trained model will be stored in the data storage which will be used as the live model.

TABLE 1. Clinical Symptoms or Signs Prior to Diagnosis of Celiac Disease

|  |
| --- |
| Abdominal pain, gas, bloating |
| Weight loss |
| Poor growth |
| Diarrhea |
| Extreme weakness |
| Nausea, vomiting |
| Anemia |
| Mood swings/depression |
| Constipation |
| Eczema |
| Bone/joint pain |
| Mouth ulcers |
| Muscle cramps |
| Easy bruising |

These symptoms that were answered by the user will be saved to the working storage and will be processed by the model to generate the result.

Generated Boosting - is a machine learning technique for regression and classification problems, which produces a prediction model in the form of an ensemble of weak prediction models, typically decision trees. It builds the model in a stage-wise fashion like other boosting methods do, and it generalizes them by allowing optimization of an arbitrary differentiable loss function.

Regression analysis - is a statistical process for estimating the relationships among variables. It includes many techniques for modeling and analyzing several variables, when the focus is on the relationship between a dependent variable and one or more independent variables

Decision tree algorithm - is a flow-chart-like structure, where each internal (non-leaf) node denotes a test on an attribute, each branch represents the outcome of a test, and each leaf (or terminal) node holds a class label. The topmost node in a tree is the root node.

### DEVELOPMENT DETAILS

Discuss how the system will be developed. Present all tools that was used in the implementation.

## 3.3.2 RESEARCH INSTRUMENT

Presents the tools used in gathering data, usually in the form of Survey Questionnaires or Experiment Paper. The software/system is not an instrument rather it is the object to which the instrument is used/applied

.

## DATA GENERATION/GATHERING PROCEDURE

Narrative explanation detailing the step-by-step courses of action done in order to obtain the data.

## STATISTICAL TREATMENT OF DATA

The objective of the study is to obtain the accuracy of the system in diagnosing if a person has celiac disease and show the degree if the person is probable of having it. The following formula are used to calculate the system’s accuracy.

The formula for accuracy is used to provide the reliability of the system. It is determined using this formula:

𝐴𝑐𝑐𝑢𝑟𝑎𝑐𝑦 = 100% − 𝐸𝑟𝑟𝑜𝑟 𝑃𝑒𝑟𝑐𝑒𝑛𝑡𝑎𝑔𝑒

Equation 1. Accuracy Formula

Equation 2. Error Percentage Formula

Where:

𝐶𝐷 : Total number of Cases that were correctly diagnosed by the system

(The system’s diagnosis and Expert’s Diagnosis are the same).

𝑇𝑜𝑡𝑎𝑙 : Total number of Cases diagnosed