

Low Level Design (LLD)

THYROID DISEASE DETECTION

Revision Number: 1.6

Last date of revision: 31st AUGUST 2024

ADITYA MARNE

SOHAM JAGTAP

ADNAN SHEIKH

SUFIYAN SHAIKH

Document Version Control

Date Issued	Version	Description	Author
20 th Aug 2024	1.1	First Draft	Soham Jagtap
22 th Aug 2024	1.2	Added Workflow chart	Aditya Marne
24 th Aug 2024	1.3	Added Exception Scenarios Overall, Constraints	Adnan Sheikh
25 st Aug 2024	1.4	Added user I/O flowchart	Sufiyan Shaikh
26 st Aug 2024	1.5	Added dataset overview and updated user I/O flowchart.	Aditya Marne
27 st Aug 2024	1.6	Restructure and reformat LLD	Soham Jagtap

Contents

Document Version Control

Abstract

1	Introduction	4
1.1	Why this Low-Level Design Document?	
1.2	Scope	
1.3	Constraints	
1.4	Risks	
1.5	Out of Scope	
2	Technical specifications	6
2.1	Dataset	
2.2	Predicting Disease	
2.3	Logging	
2.4	Database	
2.5	Deployment	
3	Technology stack	8
4	Proposed Solution	8
5	Model training/validation workflow	9
6	User I/O workflow	10
7	Exceptional scenarios	11
8	Test cases	11
9	Performance	11

Abstract

Thyroid disease a very common problem in India, more than one crore people are suffering with the disease every year. Especially it is more common in female. Hyperthyroidism and hypothyroidism are the most two common diseases caused by irregular function of thyroid gland. Thyroid disorder can speed up or slow down the metabolism of the body. In the world of rising new technology and innovation, healthcare industry is advancing with the role of Artificial Intelligence. Machine learning algorithms can help to early detection of the disease and to improve the quality of the life. This study demonstrates the how different classification algorithms can forecasts the presence of the disease. Different classification algorithms such as Logistic regression, Random Forest, Decision Tree, Naïve Bayes, Support Vector Machine have been tested and compared to predict the better outcome of the model.

1 Introduction

1.1 Why this Low-Level Design Document?

The purpose of this document is to present a detailed description of the Thyroid Disease Detection System. It will explain the purpose and features of the system, the interfaces of the system, what the system will do, the constraints under which it must operate and how the system will react to external stimuli. This document is intended for both the stakeholders and the developers of the system and will be proposed to the higher management for its approval.

The main objective of the project is to predict if a person is having compensated hypothyroid, primary hypothyroid, secondary hypothyroid or negative(no thyroid).

Thyroid Disease Detection is a vital part of health IT and can:

- Contain a patient's medical history, diagnoses, medications, treatment plans, allergies and laboratory and test results.
- Allow access to evidence-based tools that providers can use to make decisions about a patient's care.
- Automate and streamline provider workflow

A Thyroid Disease Detection System contains patient health information, such as:

- Patient demographics
- Patient's age
- Vital signs
- Medical histories

- Diagnoses
- Medications
- Lab and test results

This project shall be delivered in two phases:

Phase 1: All the functionalities with PyPi packages.

Phase2: Integration of UI to all the functionalities.

1.2 Scope

This software system will be a Web application This system will be designed to detect the thyroid disease and whether it is a compensated, primary or secondary hypothyroid for better disease management, improved interventions, and more efficient health-care resource allocation. More specifically, Early detection of any preventable diseases is important for better disease management. This system is designed to predict the thyroid disease from patient information such as age, disease history, lab results, procedures and medications.

1.3 Constraints

The thyroid disease detection application must be user friendly, as automated as possible and users should not be required to know any of the workings.

1.4 Risks

Document specific risks that have been identified or that should be considered.

1.5 Out of Scope

Delineate specific activities, capabilities, and items that are out of scope for the project.

2 Technical specifications

2.1 Dataset

```

finalTrain_new2.csv X
notebooks > data > finalTrain_new2.csv > data
1 age,on_thyroxine,goitre,TSH,TSH_measured,T3,TT4_measured,TT4,FTI,referral_source,classes
2 41,f,f,1.3,t,2.5,t,125,109,SVHC,1
3 23,f,f,4.1,t,2.0,t,102,110,other,1
4 46,f,f,0.98,t,2.02,t,109,120,other,1
5 70,t,f,0.16,t,1.9,t,175,110,other,1
6 70,f,f,0.72,t,1.2,t,61,70,SVI,1
7 18,t,f,0.03,t,2.02,t,183,141,other,1
8 59,f,f,4.67,f,2.02,t,72,78,other,1
9 66,f,f,0.6,t,2.2,t,123,132,SVI,1
10 68,f,f,2.4,t,1.6,t,83,93,SVI,1
11 67,t,f,0.03,t,2.02,t,152,153,other,1
12 71,f,f,0.03,t,3.8,t,171,151,other,1
13 59,f,f,2.8,t,1.7,t,97,107,SVI,1
14 28,f,f,3.3,t,1.8,t,109,119,SVHC,1
15 65,f,f,12.0,t,2.02,t,99,87,other,0
16 42,f,f,1.2,t,1.8,t,70,81,other,1
17 63,f,f,1.5,t,1.2,t,117,121,SVI,1
18 28,f,f,2.1,t,2.6,t,121,130,SVHC,1
19 51,t,f,0.1,t,2.02,t,130,151,other,1
20 46,f,f,0.8,t,2.1,t,108,119,other,1
21 54,f,f,3.1,t,2.02,t,104,116,other,1
22 55,f,f,0.2,t,1.8,t,134,131,SVI,1
23 63,f,f,0.03,t,5.5,t,199,190,other,1
24 60,t,f,13.0,t,1.4,t,57,92,other,1
25 25,f,f,0.3,t,3.1,t,129,110,other,1
26 73,f,f,1.9,t,1.5,t,113,106,SVI,1
27 67,f,f,4.67,f,2.02,t,97,102,other,1
28 34,f,f,0.04,t,2.5,t,119,76,other,1
29 70,f,f,2.5,t,2.3,t,84,92,other,1
30 78,f,f,0.5,t,1.9,t,81,98,SVI,1
31 37,f,f,1.7,t,1.9,t,95,90,SVI,1
32 25,f,f,1.1,t,2.02,t,101,94,other,1
33 26,f,f,1.8,t,2.5,t,147,129,SVHC,1
34 58,f,f,0.26,t,2.7,t,120,95,other,1
35 51,f,f,2.8,t,2.02,t,69,91,other,1
36 64,f,f,4.67,f,2.02,t,109,110,other,1
37 44,f,f,45.0,t,1.4,t,39,33,SVI,2

```

2.1.1 Thyroid dataset overview

The thyroid dataset consists of a table with 2800 records and 30 features. Features are distributed as 7 continuous features and 23 categorical features. There are a total 7852 patients in the training set and 1476 patients in the test set.

2.1.2 Input schema

Feature name	Datatype	Size	Null/Required
Age	Int	3	Required

2.2 Predicting Disease

- The system presents the set of inputs required from the user.
- The user gives required information.
- The system then predicts that the user is having thyroid or not.
Also, it tells whether a user is having compensated, primary or secondary hypothyroid based on the lab test results provided by the user.

2.3 Logging

We should be able to log every activity done by the user.

- The System identifies at what step logging required
- The System should be able to log each and every system flow.
- Developers can choose logging methods. You can choose database logging/ File logging as well.
- System should not be hung even after using so many loggings. Logging just because we can easily debug issues so logging is mandatory to do.

2.4 Database

System needs to store every request into the database and we need to store it in such a way that it is easy to retrain the model as well.

1. The User gives required information.
2. The system stores each and every data given by the user or receives on request to the database. Database which we have chosen is MongoDB.

2.5 Deployment



3 Technology stack

Front End	HTML/CSS
Backend	Python Flask
Database	MongoDB
Deployment	Render

4 Proposed Solution

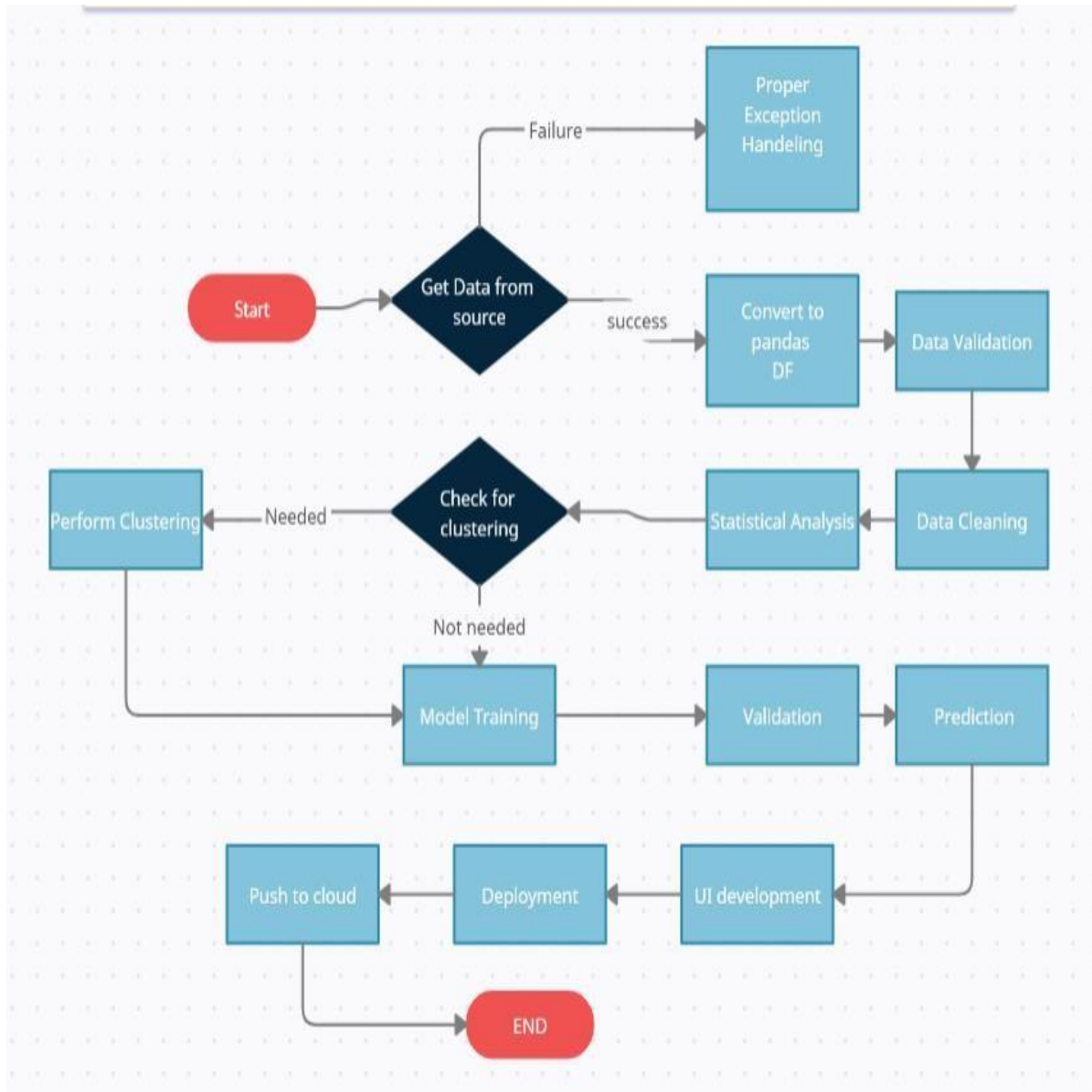
The proposed solution for this project is Machine learning algorithms can be implemented to predict the risk of thyroid disease. Considering various features like age, sex, level of thyroxine, anti-thyroid medication as inputs from the web app, the implemented classification model will predict the output as presence or absence of the thyroid.

Here, we have used Decision Tree algorithm to predict whether the patient is having Thyroid or not.

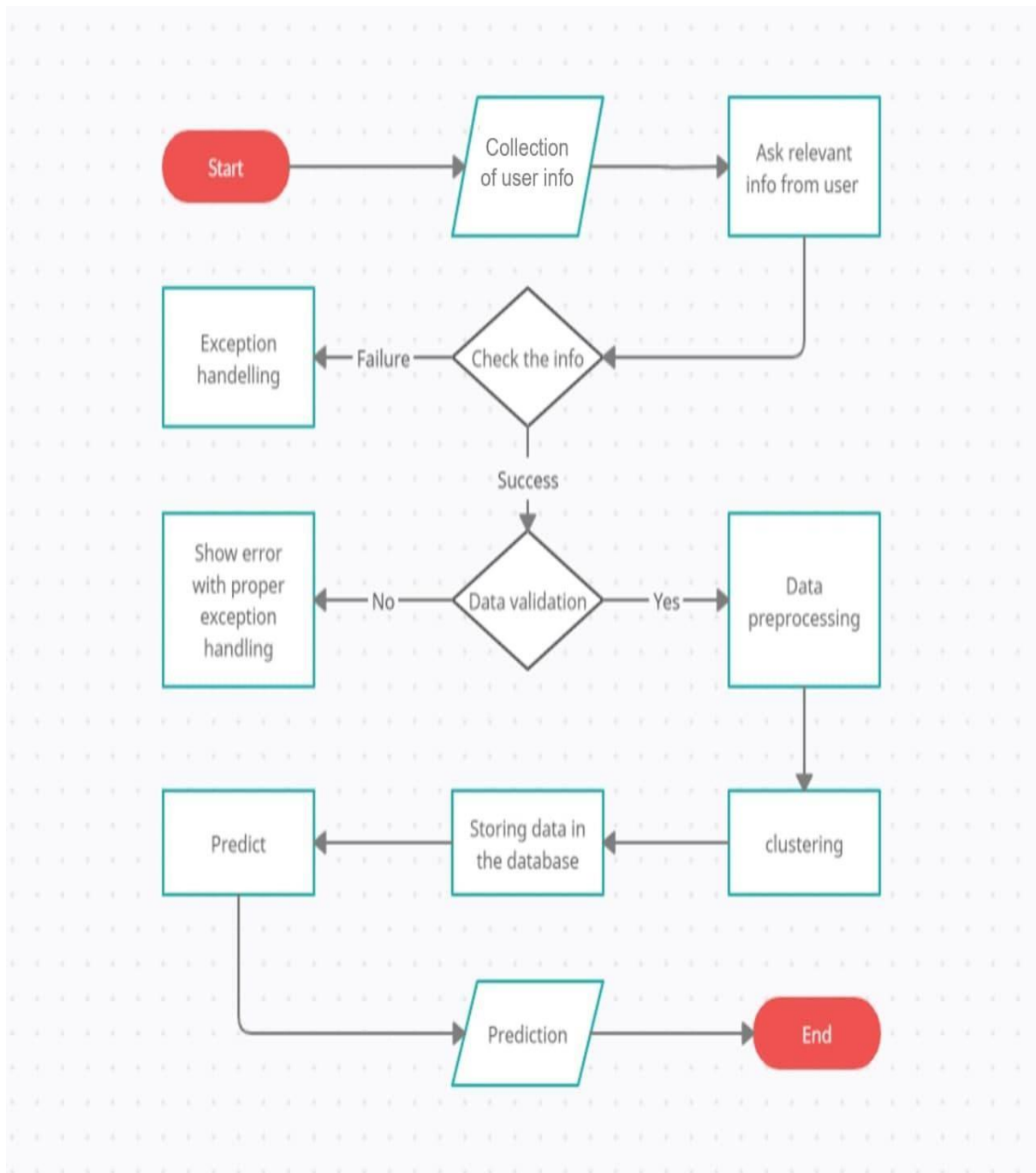
However, drawing a baseline model is important since it tells us how well other models have performed compared to base model. Here, the base model for Thyroid Disease Detection dataset is Logistic Regression.

1. Baseline Model: Logistic Regression
2. Actual Model: Decision Tree

5 Model training/validation workflow



6 User I/O workflow



7 Exceptional scenarios

Step	Exception	Mitigation	Module
20 st Aug 2024	1.1	First Draft	Soham Jagtap
22 st Aug 2024	1.2	Added Workflow chart	Aditya Marne

8 Test cases

Test case	Steps to perform test case	Module	Pass/Fail

9 Performance

We can observe that the accuracy of the predicted output was seen at 99.25% using Decision Tree algorithm. Other classification models such as logistic regression and Random Forest classifier have given good accuracy above 90%.