Thyroid Disease Detection

High Level Design (HLD)

Contents

1. Introduction
    1.1. Why High Level Design Document?

2. General Description
    2.1. Product Perspective
    2.2. Problem Statement
    2.3. Proposed Solution
    2.4. Technical Requirements
    2.5. Data Requirements
    2.6. Tools Used
    2.7. Constraints

3. Design Details
    3.1. Process Flow
    3.2. Deployment Process

4. Performance
    4.1. Re-usability
    4.2. Application Compatibility
    4.3. Resource Utilization
    4.4. Deployment
    4.5. User Interface

5. Conclusion

Document Version Control

Version	Date	Author	Comments
1.0	22-Oct-2021	Ahmad Taquee, Priyanka Gupta, Diksha Sharma	Introduction & Architecture defined

Abstract

Thyroid gland plays a major role in maintaining the metabolism of human body. Data mining in health care industry provides a systematic use of the medical data. Thyroid diseases are most common today. Early changes in the thyroid gland will not affect the proper working of the gland. By the early identification of thyroid disorders, better treatment can be provided in the early stage thus can avoid thyroid replacement therapy and thyroid removal up to an extent.

1. Introduction

1.1. Why High-Level Design Document?

The purpose of this High-Level Design (HLD) Document is to add the important details about this project. Through this HLD Document, I'm going to describe every small and big thing about this project.

2. General Description

2.1. Product Perspective

The purpose of this model is to detect the type of Thyroid disease.

2.2 Problem statement

Thyroid disease is a common cause of medical diagnosis and prediction, with an onset that is difficult to forecast in medical research. The thyroid gland is one of our body's most vital organs. Thyroid hormone releases are responsible for metabolic regulation. Hyperthyroidism and hypothyroidism are one of the two common diseases of the thyroid that releases thyroid hormones in regulating the rate of body's metabolism. The main goal is to predict the estimated risk on a patient's chance of obtaining thyroid disease or not.

2.3 Proposed Solution

In the proposed solution, we used Random Forest Classifier machine learning model to classify the different types thyroid disease. Here, first we are performing Data pre-processing step, in which data transformation, handling missing values, feature transformation, feature selection, steps are performed and then we are going to build the machine learning model and will be deployed it on cloud platform.

2.4 Technical Requirements

Following are the requirements of this project:

- Model should be deployed on cloud (Azure, AWS, GCP, Heroku).
- Cassandra database should be integrated in this project for any kind of user input.

2.5 Data Requirements

Data Requirement completely depend on our problem.

- For training and testing the model, we are using Thyroid disease dataset which is available on UCI portal.
- From user we are taking following input

Feature Names:

- Class attribute (2 = normal, 1 = hypothyroid, 0 = hyperthyroid)
- T3-resin uptake test (a percentage)
- Total Serum thyroxin as measured by the isotopic displacement method
- Total serum triiodothyronine as measured by radioimmune assay
- basal thyroid-stimulating hormone (TSH) as measured by

2.6 Tools Used

2.7 Data Requirements

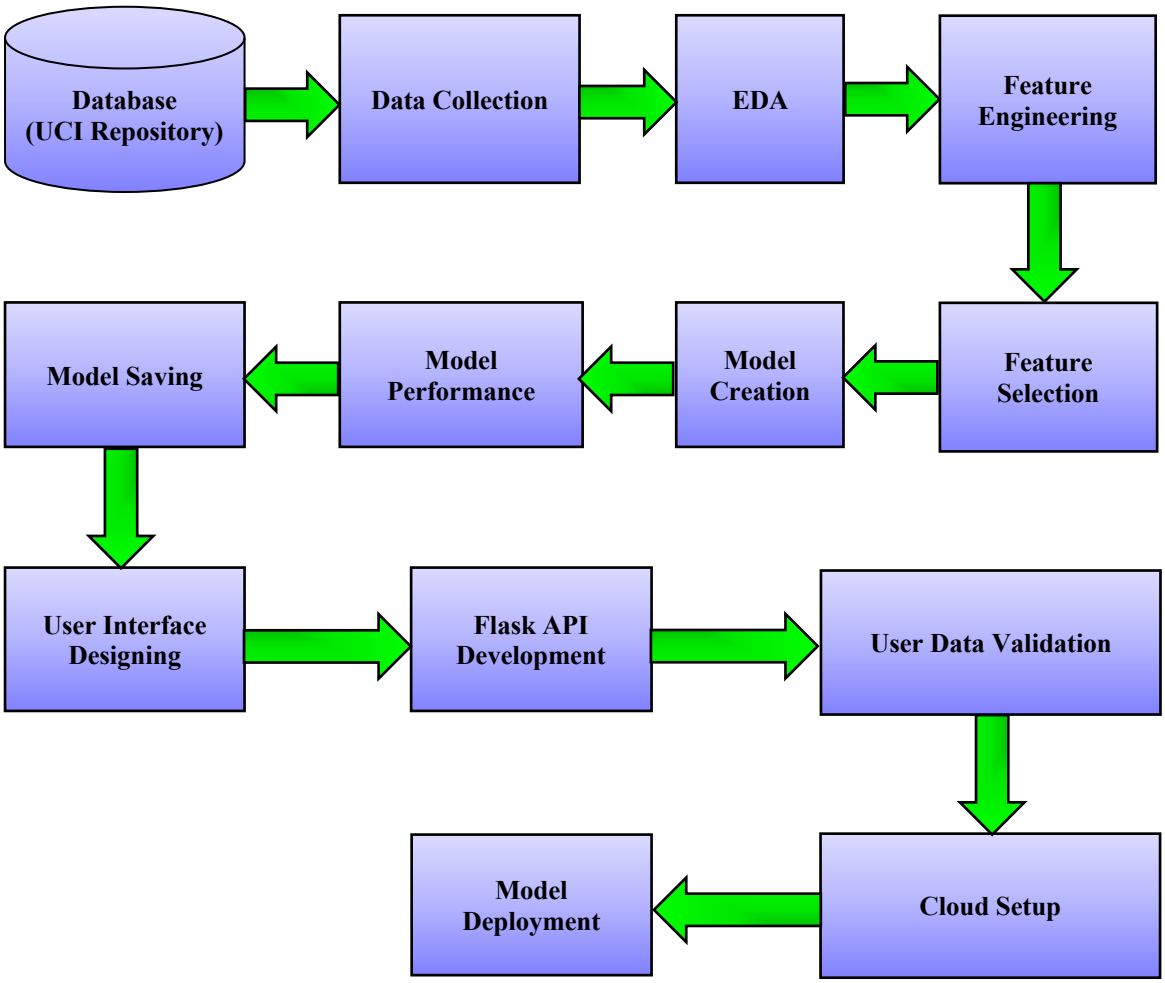
- PyCharm is used as IDE.
- For visualization of the plots, Matplotlib, Seaborn are used.
- Heroku is used for deployment of the model.
- MongoDB is used to retrieve, insert, delete, and update the database.
- Front end development is done using HTML, CSS, Bootstrap, Flask is used for backend development and for API development.
- GitHub is used as version control system

2.8 Constraints

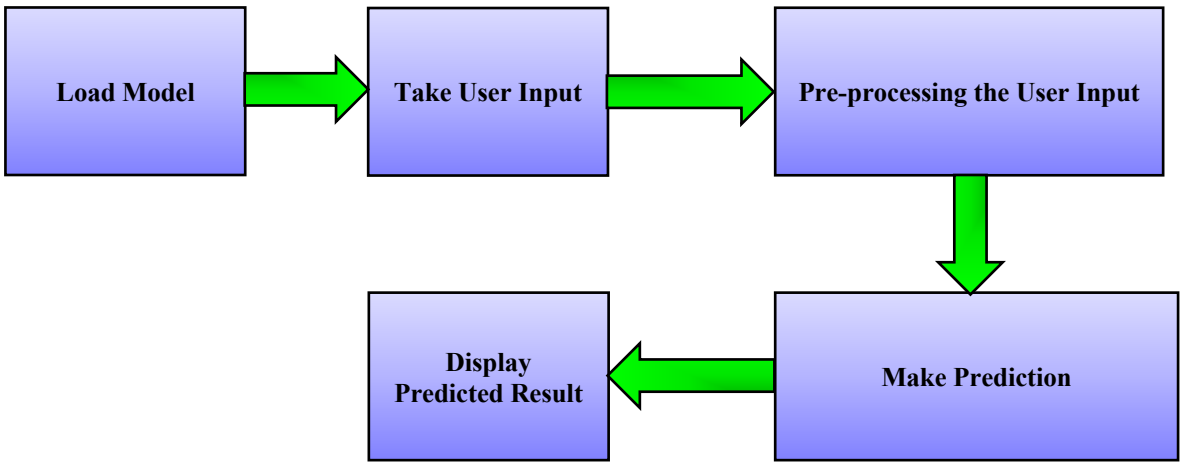
The Thyroid Disease Detection Model system must be user friendly, errors free and users should not be required to know any of the back-end working.

3. Design Details

3.1 Process Flow



3.2 Deployment Process



4. Performance

a) Solution of Thyroid Disease Detection is used to predict the thyroid disease, and it should be as accurate as possible.

b) That's why before building this model we followed complete process of Machine Learning. Here is summary of complete process:

i) First, we cleaned our dataset properly by removing all null value and duplicate value present in dataset.

ii) After that we performed EDA and feature transformation.

iii) And then we performed feature selection process.

iv) Then we performed the encoding – numerical features and categorical features

v) And now, we split the dataset in train-test split.

vi) After performing above, we trained our dataset on different classification algorithm (Logistic, SVM, KNN, Decision Tree Classifier, Random Forest Classifier etc.). After training the dataset on different algorithms, we got highest accuracy of 99.7% on Random Forest Classifier. Here also I got highest accuracy of 96.6% on test dataset by same Random Forest Classifier.

vii) After that we saved our model in pickle file format.

viii) After that our model was ready to deploy, we deployed this model on Heroku.

c) Re-usability

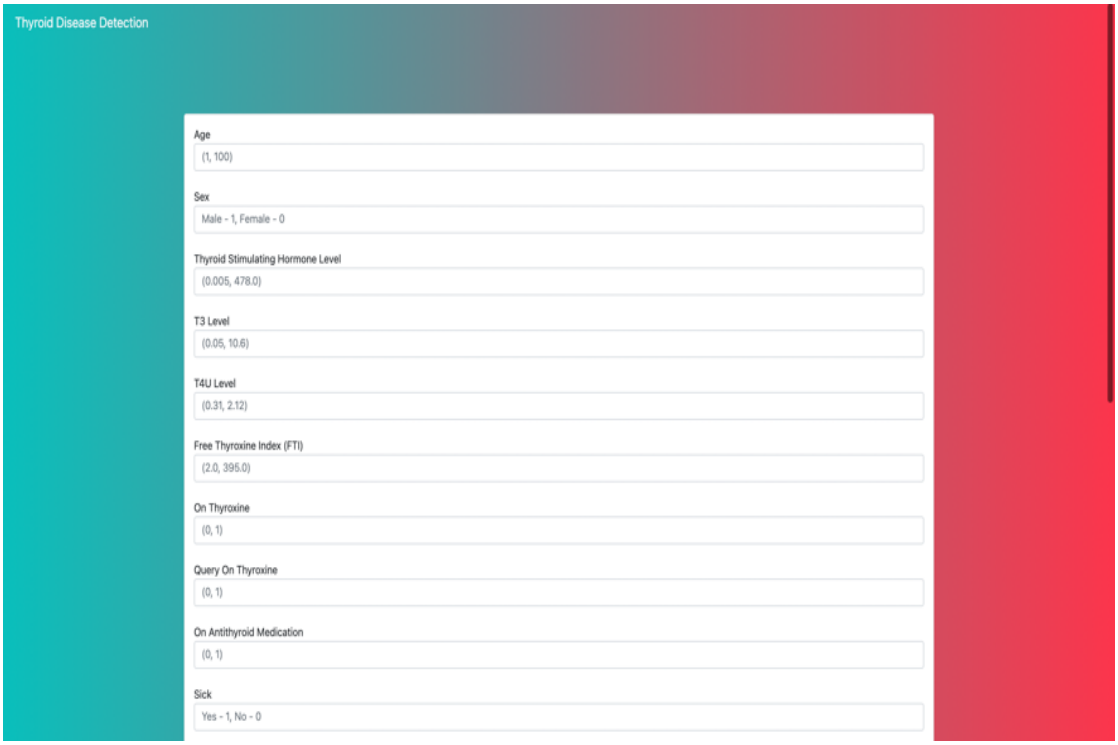
We have done programming of this project in such a way that it should be reusable. So that anyone can add and contribute without facing any problems

d) Application Compatibility

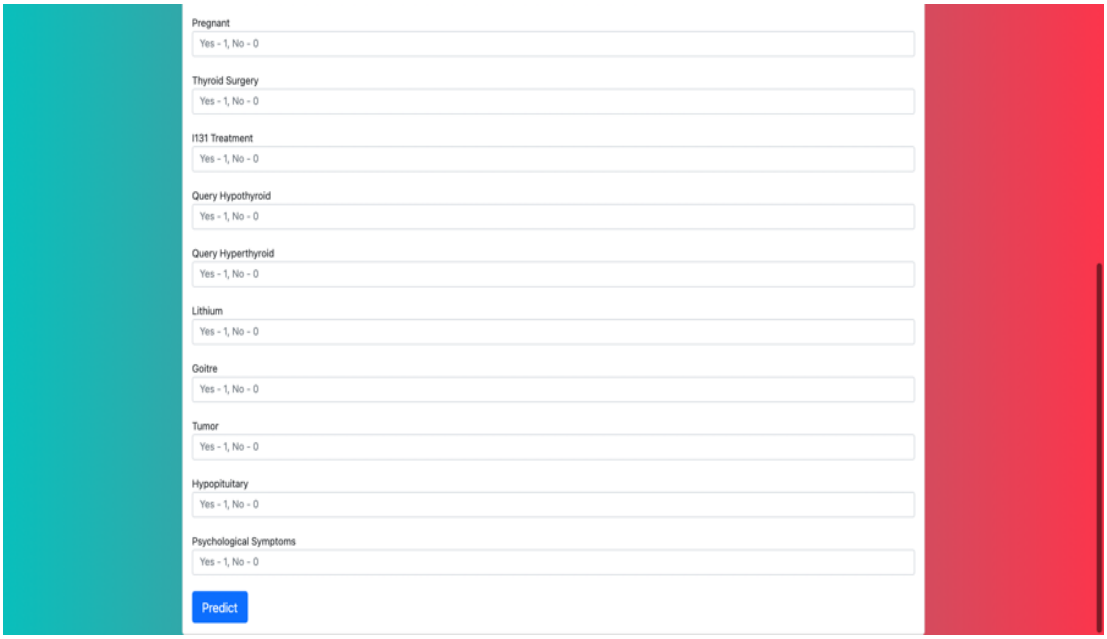
The different module of this project is using Python as an interface between them. Each module have it's own job to perform and it is the job of the Python to ensure the proper transfer of information.

e) Deployment: We have deployed this on Heroku cloud.

4. User Interface



High Level Design



5. Conclusion

This Project proposes a method for the classification of thyroid disease that a user is suffering from along with disease description and healthy advices. Random Forest Classifier is used for classification.