Thyroid Disease Detection

High Level Design (HLD)

Contents

Document Version Control

Version	Date	Author	Comments
1.0	10-Jan-2022	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

To create an AI solution for detecting thyroid disease and to implement the following use cases.

- To detect thyroid disease and its type in healthy person.
- To detect thyroid disease and its type in unhealthy person.

Here unhealthy person means person already affected by thyroid disease.

2.3 Proposed Solution

The solution proposed here is a data science model based on machine learning can be implemented to perform above mention use cases. In first use case , we will take input from a healthy person who is not suffering from thyroid disease and see whether proposed solution is going to detect it or not. And in second use case, we will take input from an unhealthy person, already suffering from thyroid disease and check our solution whether it is performing or not in right way.

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.

High Level Design

- 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

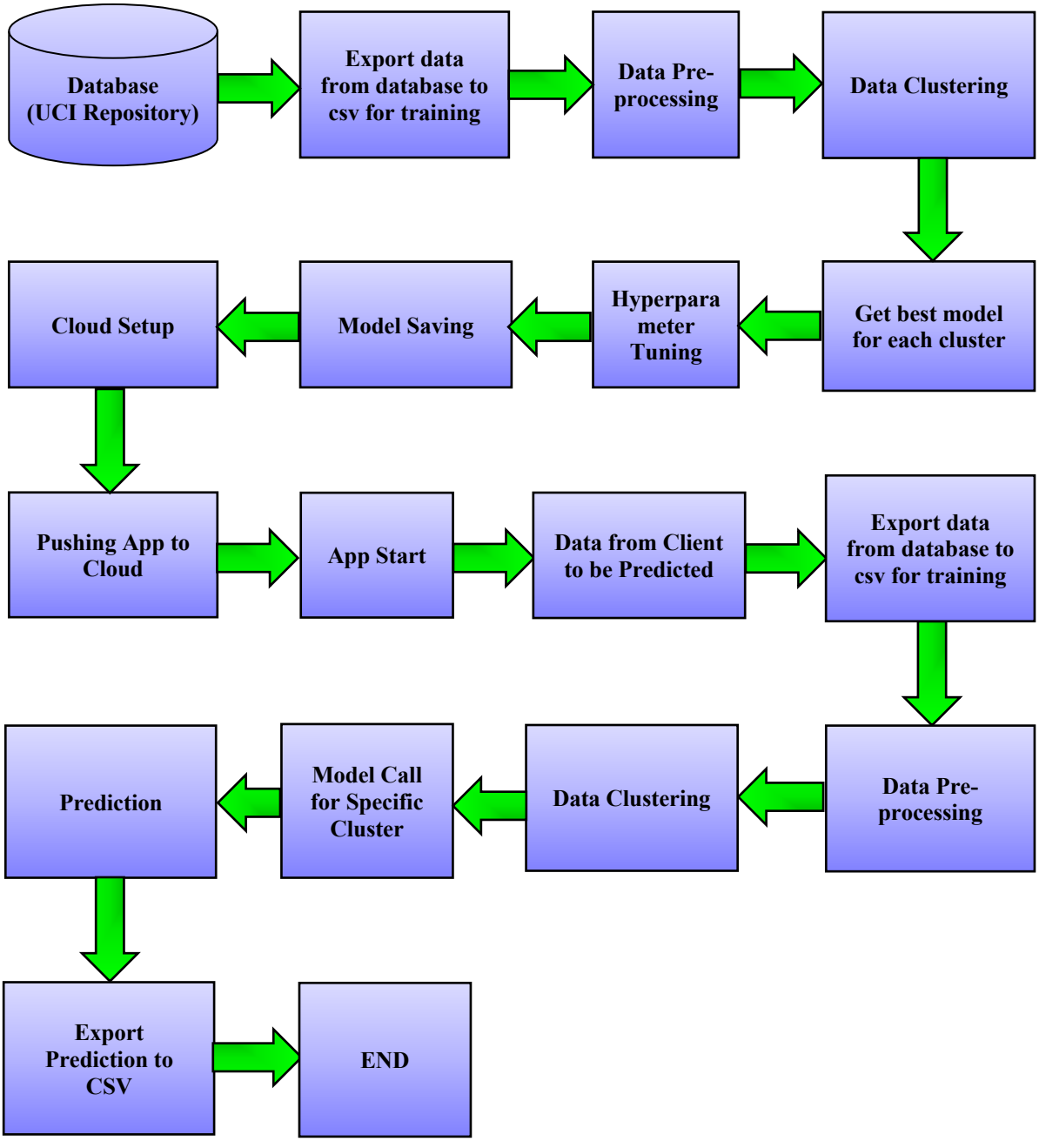
- PyCharm is used as IDE.
- For visualization of the plots, Matplotlib, Seaborn are used.
- Heroku is used for deployment of the model.
- MongoDB is used to retrieve, insert, delete, and update the database.
- Front end development is done using HTML, CSS, Bootstrap, Flask is used for backend development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

2.8 Constraints

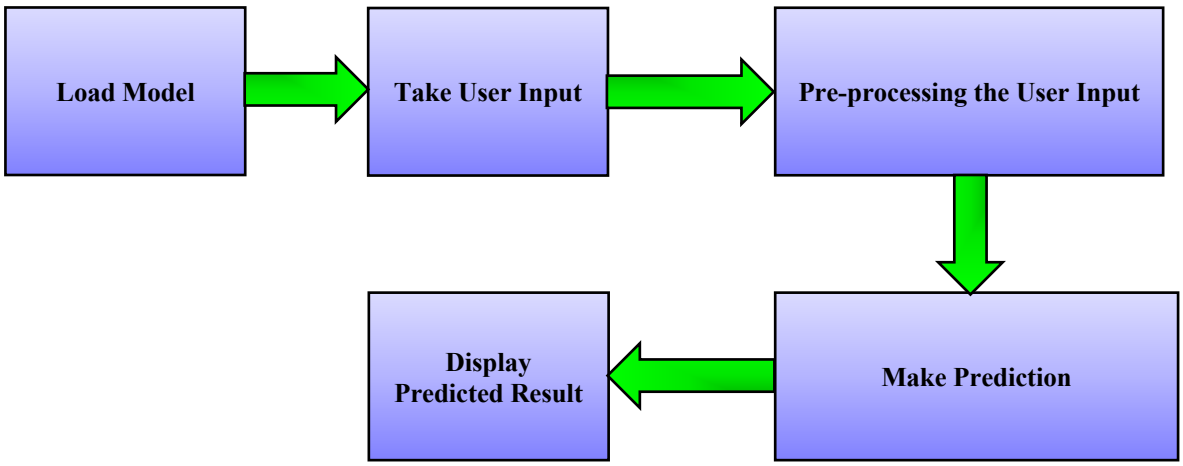
The Thyroid Disease Detection Model system must be user friendly, errors free and users should not be required to know any of the back-end working.

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 has its 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.

5. Conclusion

Thyroid Disease Detection solution will take health-care domain data of those patients who have undergone diagnosis for thyroid to train our machine learning model and will evaluate its performance over use-cases mentioned above. And then leverage its prediction to detect thyroid disease in people having symptoms of thyroid and able to alert people who is on positive side so that medical attention along with treatment will be given to that particular people as soon as possible. This solution should be as accurate as possible, so that chances of misleading reports will be taken good care of.