# High Level Design (HLD)

# Thyroid detection

Revision Number: 2.0

Last date of revision: 03-01-2022

Revision Number 1.0

Last Date of revision : 31/12/2021

# Document Version Control

|  |  |  |  |
| --- | --- | --- | --- |
| Date Issued | Version | Description | Author |
| 10-3-2022 | 1 | Initial HLD -V1.0 | Rohan |
| 10-03-2022 | 2 | Update the data | Rohan |
|  |  |  |  |
|  |  |  |  |
|  |  |  |  |
|  |  |  |  |
|  |  |  |  |

# Content

Document Version Control

Abstract

1. Introduction………………………………………………………………………………….5
   1. Why this High-Level Design Document?.......................................................................5
   2. Scope…………………………………………………………………………………...5
2. General Description…………………………………………………………………………6
   1. Product Perspective…………………………………………………………………….6
   2. Problem Statement……………………………………………………………………...6
   3. Proposed Solution………………………………………………………………………6
   4. Further Improvements…………………………………………………………………..6
   5. Technical Requirements………………………………………………………………...6
   6. Data Requirements………………………………………………………………………7
   7. Tools Used………………………………………………………………………………7
   8. Constraints………………………………………………………………………………8
3. Design Details………………………………………………………………………………..9
   1. Process Flow……………………………………………………………………………9
      1. Model Training and Evaluation…………………………………………………10
      2. Deployment Process…………………………………………………………….11
4. Performance………………………………………………………………………………….12
   1. Reusability………………………………………………………………………………12
   2. Application Compatibility………………………………………………………………12
   3. Resource Utilization…………………………………………………………………….12
   4. Deployment……………………………………………………………………………..12

5 Conclusion……………………………………………………………………………………13

# ABSTRACT

This is the age of the internet where the amount of data being generated is so huge that man alone is not able to process the data. Many machine learning techniques hence have been discovered for this purpose. We are trying to predict the sales of a store using different machine learning techniques and trying to determine the best algorithm suited to our particular problem statement. We have implemented normal regression techniques as well as boosting techniques in our approach and have found that the boosting algorithms have better results than the regular regression algorithms.

# 

# 1 Introduction

# Why this High-Level Design Document?

The purpose of this High-level Design(HLD) document is to add the necessary detail to the current project description to represent a suitable model for coding. This document is also intended to help detect contradictions prior to coding and can be used as a reference manual for how the modules interact at a high level.

The HLD will :

* Present all of the design aspects and define them in detail
* Describe the user interface being implemented
* Describe the performance requirements
* Include design features and the architecture of the project
* List and describe the non-functional attributes like :
  + - Security
    - Relatability
    - Maintainability
    - Portability
    - Reusability
    - Application Compatibility
    - Resource utilization
    - Serviceability

# Scope

The HLD documentation presents the structure of the system, such as the database architecture, application architecture (layers), application flow (navigation), and technology architecture. The HLD uses non-technical to mildly-technical terms which should be understandable to the administrators of the system.

# General Description

# 2.1 Product Perspective

Thyroid detection is the solution for the doctors to identify wether the person has a risk of thyroid or not

# 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

We will perform EDA to find the important relation between different attributes and will use a machine-learning algorithm to predict the future sales demand. The client will be filled the required feature as input and will get results through the web application. The system will get features and it will be passed into the backend where the features will be validated and preprocessed and then it will be passed to a hyperparameter tuned machine learning model to predict the final outcome.

2.4 Data Requirements

The data required for the building of the project is already available on the dashboard. The client will send data in multiple sets of files in batches at a given location.

Data will contain Thyroid level and 31 columns of different values of the report.

# 2.4 Further Improvements

We can save the User's History of Prediction and identify the more related features. It will automatically be assigned for production.

# 2.5 Technical Requirements

Any Device with Internet Access and runs in any browser.

# 2.6 Data requirements

The data required for the building of the project is already available on the dashboard. The client will send data in multiple sets of files in batches at a given location.

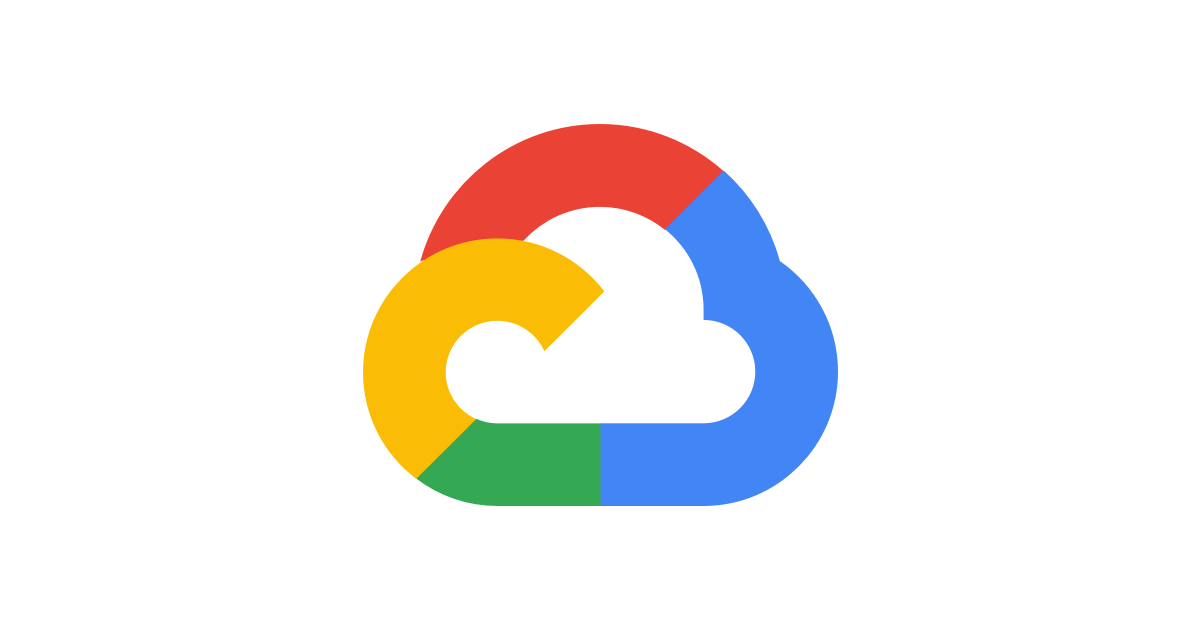
Data will contain Thyroid level and 31 columns of different values of the report.

|  |  |  |
| --- | --- | --- |
| Age | Float | Age of the person |
| Sex | Boolean | Sex of the person |
| on\_thyroxine | Boolean | Person is on thyroxine or not |
| on\_antithyroid\_medication | Boolean | Person is on anti thyroid medication or not |
| L131 treatment | Boolean | The person had done l131 treatment or not |
| Query hypothyroid | Boolean | Has query on hypothroid or not |
| Goitre | Boolean | Goitre present or not |
| Psych | Boolean | Has psychological disfunction or not |
| Tsh measured | Boolean | Is tsh measured or not |
| Tsh | float | Value or tsh |
| T3 measured | Boolean | Is t3 measured or not |
| T3 | Float | Value or t3 |

# 2.7 Tools Used

Python Programming language and frameworks such as Numpy, Pandas,

Scikit-learn, Google cloud are used to build the whole model. 



* pycharm and jupyter notebook is used as IDE.
* For visualization of the plots Matplotlib, Seaborn are used.
* Flask is used for deployment of the model.
* Front end Development is done using HTML/CSS, Bootstrap 4.
* Python is used for backend development.
* Github is used as a version control system.

# 2.8 Constraints

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

# Design Details

# Process Flow

For Identifying the different types of anomalies, we will use a machine learning model. Below is the process flow diagram as shown below.

Proposed Methodology



# Model Training and Evaluation

# 

# Deployment Process

Start the application

Enter the details

Submit the data

Execute by model

Predicted result

# 4 Performance

The Performance of the model depends on the dataset. We have done in depth preprocessing of the dataset for greater accuracy and much closer prediction with less error.

# Reusability

The code and the module are created during the time of building the project should maintain all coding guidelines and full project code is written in a Modular fashion. Our system should have the flexibility to work properly from any location. And it should handle any improper input value from the user and should give a meaningful error message so the user can correct his/her mistake and enter valid input to get the result. And the system should be reusable in every manner with different types of inputs values that are all are it has been trained.

# Application Compatibility

This Project will be using Python as an interface between them. Each Component will have its own task to perform, and it is the job of the python to ensure proper transfer of information.

# Resource Utilization

When any task is performed, it will likely use all the processing power available until that function is finished

# Deployment

# 

# 5 Conclusion

The main goal is to predict the estimated risk on a patient's chance of obtaining thyroid disease or not.