# HIGH LEVEL DOCUMENT

***Name:*** **Anirban Majumder**

***Course:***

**Decode Data Science With Machine Learning**

***Project Title:*** **Thyroid Disease Detection**

***Technologies:*** **Machine Learning Technology**

***Domain:*** **Healthcare**

***Project Difficulty Level:*** **Intermediate**

***Version No.:*** **1.0**

***Last Date of Revision:*** **19th August, 2024**

# DOCUMENT VERSION CONTROL

| Date | Version | Description | Author |
|---|---|---|---|
| 19.08.2024 | 1.0 | An overview that ensures alignment and clarity while directing stakeholders on the project's goals, architecture, and implementation. | Anirban Majumder |

# CONTENTS

# 1. ABSTRACT

Thyroid disease is a common medical problem that has a substantial effect on metabolism and general health. Through the secretion of hormones like thyroxine (T4) and triiodothyronine (T3), the thyroid gland is essential in controlling the body's metabolism. Thyroid hormone imbalances can cause disorders like hyperthyroidism and hypothyroidism, which can lead to major health consequences like cardiovascular illnesses, hypertension, high cholesterol, and decreased fertility. For thyroid illness to be effectively treated and managed, an accurate and timely diagnosis is essential, as delays can make the condition worse and raise the risk of serious health consequences.

Using machine learning approaches, this project tackles the urgent need for reliable thyroid illness identification. Predicting a patient's risk of thyroid disease using past data and a variety of health indicators is the main goal. In order to guarantee high-quality input for model building, the project takes a methodical approach, starting with data exploration and cleaning. Then, feature engineering is used to improve the dataset's predictive capacity, and several machine learning techniques are applied to determine which model is best for precise illness prediction.

The integration of machine learning into healthcare, particularly in the diagnosis of thyroid disease, offers significant potential for improving patient outcomes. By processing large volumes of healthcare data and applying advanced predictive models, this project aims to facilitate early detection and personalized treatment strategies. This approach not only improves diagnostic accuracy but also supports better decision-making in clinical settings, ultimately contributing to the reduction of morbidity and mortality associated with thyroid-related disorders.

# 2. INTRODUCTION

This project investigates the use of machine learning to forecast thyroid disease, a prevalent and significant medical condition. The project intends to improve diagnostic accuracy by using advanced algorithms to analyse patient data. This would enable early detection and personalised treatment methods for better health outcomes in those who may be at risk of thyroid diseases.

## 2.1 Objective of High – Level Document (HLD)

The goal of this High Level Design (HLD) Document is to update the project description with the information required to create a model that is ready for coding. This paper can serve as a reference manual for the high-level interactions between the modules and is also meant to assist in identifying conflicts before coding.

The HLD will serve the following purposes:

- Present all of the design aspects and define them in detail
- Describe the user interface being implemented
- Describe the hardware and software interfaces
- Describe the performance requirements
- Include design feature and the architecture of the project
- List and describe the non-functional attribute like:
  - ✓ Security
  - ✓ Reliability
  - ✓ Maintainability
  - ✓ Portability
  - ✓ Reusability
  - ✓ Application compatibility
  - ✓ Resource utilization
  - ✓ Serviceability

## 2.2 Scope

The HLD document 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.

# 3 GENERAL DESCRIPTION

In order to overcome the difficulties in diagnosing thyroid disorders—which are vital to metabolic balance and general health—the Thyroid Disease Detection project makes use of machine learning. In order to guarantee the quality and dependability of the input data, the project takes a thorough strategy that begins with data cleaning and exploration. By improving the dataset through feature engineering, it becomes more appropriate for predictive modelling. Next, several machine learning techniques are used to create models that can precisely forecast a patient's risk of developing thyroid disease.

The intention is to find trends and risk factors related to thyroid conditions such as hypothyroidism and hyperthyroidism, which, if left untreated, can have serious consequences. The project's goal is to develop a more accurate and dependable early detection approach by analysing massive healthcare datasets. This will lower the risk of a delayed diagnosis and improve patient outcomes.

The project's scope goes beyond simply forecasting thyroid disease; it also showcases machine learning's potential in the medical field and provides insights into how sophisticated data processing and predictive analytics can revolutionise diagnosis. This study advances the field of healthcare technology while highlighting how crucial it is to use machine learning into clinical procedures in order to improve patient care and decision-making.

## 3.1 Product Perspective

The Thyroid Disease Detection solution system is a data science-based machine learning model which help us to detect the thyroid disease in people and take necessary action.

## 3.2 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.

## 3.3 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 this case, Data Preprocessing Step is performed firstly, where procedures such as Data Transformation, Data Imputation, Data Encoding, Feature Selection are performed, followed by Building, Training, Evaluation and Selection of Machine Learning Model.

### 3.4 Further Improvements

There are many use cases in the healthcare sector for the thyroid illness detection solution. In order to provide an additional layer of assurance for those who exhibit mild indications of thyroid disease, TDD solutions can also be synchronised with other health care domain solutions.

### 3.5 Data Requirements

Data requirements solely depend on our problem. The following attributes are required:

- Age
- Gender
- Thyroxin Treatment
- Antithyroid Medication
- Pregnancy
- Sick at the time of diagnosis
- Iodine Test
- Lithium Test
- Goitre Test
- Tumor Test
- TSH Level Measure
- T3 Level Measure
- T4 Level Measure
- Free Thyroxin
- Thyroxin – Binding Globulin (TBG)

### 3.6 Tools




The tools used in this project include the following:

- Python being the language for coding, Database Operations, Data Analytics, Machine Learning and creating the Flask App.
- Flask used for creating the Web App.
- Numpy and Pandas are Libraries used for Data Analysis.
- Matplotlib and Seaborn are Libraries that are essential for Data Visualization.
- HTML, CSS and JavaScript are used for the development of the WebPage.
- AWS is a cloud based platform to deploy the model.
- Scikit-learn is a Library that is essentially used for performing Machine Learning operations.
- GitHub is being used as a Repository.
- MongoDB being the Database from which the data has been retrieved.
- JSON has been used for transmission of data in web application.
- Jupyter Notebook is an IDE, used to execute the .ipynb files.
- Visual Studio Code (popularly known as VSCode) is an IDE that has been used extensively in this project, to write codes for each language, irrespective of the file type (.py, .ipynb, .html, .css, .js, .json, .txt, .md files have been written).
- This is the link of the Database: <Thyroid Disease - UCI Machine Learning Repository>

### 3.7 Constraints

The Thyroid Disease Detection solution system must be correct enough that it not mislead any report and as automated as possible and users should not be required to know any of the workings.

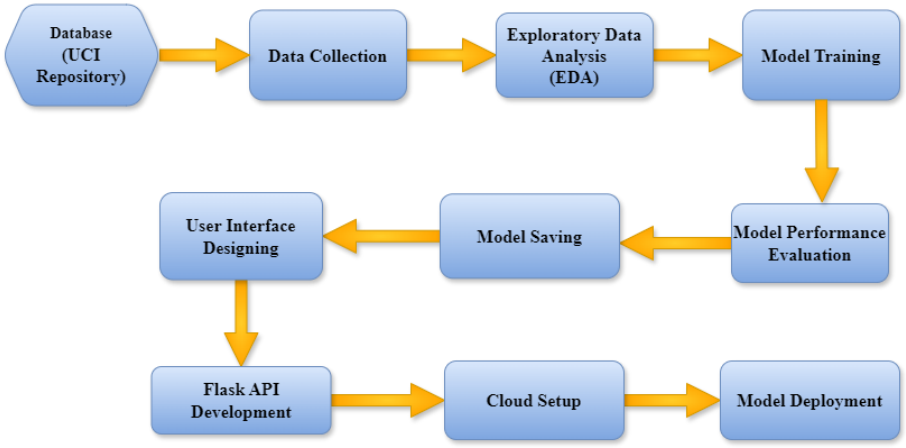
### 3.8 Assumption

The main objective of the project is to implement the use cases as previously mentioned for new dataset that comes through Hospitals which has this solution install in their campus to capture people reports.

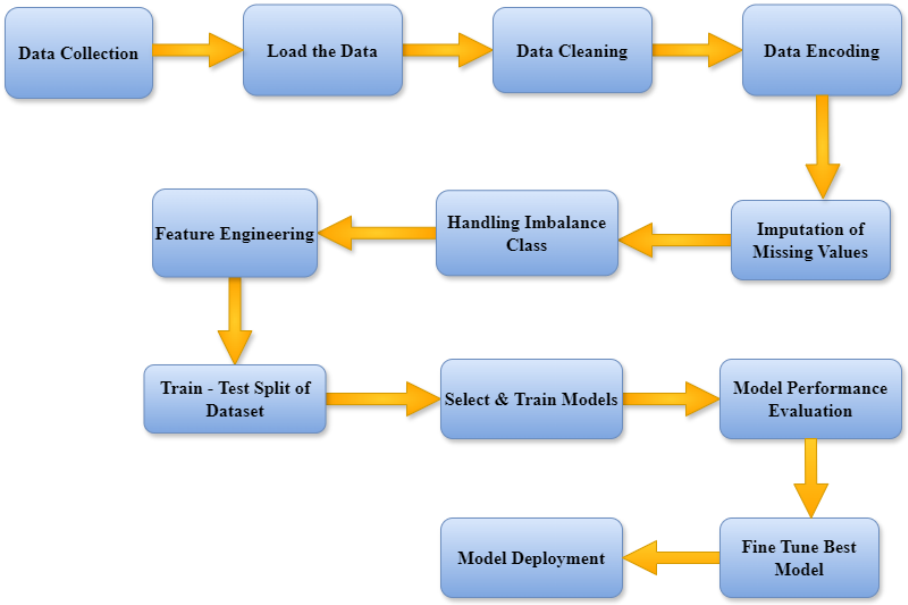
# 4. DESIGN DETAILS

The design details of the Project has been explained as follows:

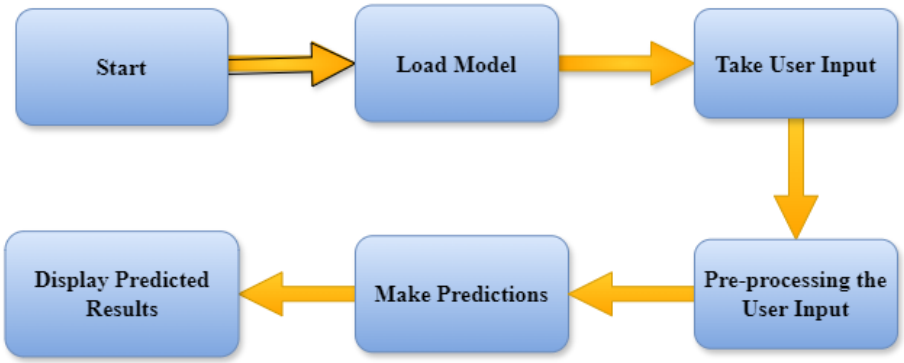
- The process flow has been described by the flowchart given below:




- The workflow of Model Training and Evaluation has been shown below:




- The workflow of the Deployment process is shown below:



# 5. PERFORMANCE

The performance can be described as follows:

- ***Purpose:***- This Machine Learning based Thyroid Disease Detection solution is going to be used for the detection of Thyroid Disorders in patients having symptoms of Thyroid, so that necessary actions can be taken for treatment.

- ***Summary:***- The complete process of this Project can be summarized as follows:
  - ✓ Firstly, the data of the database has been uploaded to MongoDB and has been successfully retrieved from MongoDB, in the form of a CSV File.
  - ✓ The dataset was checked for value consistency and the duplicated values were dropped and missing value imputation was performed.
  - ✓ Data encoding was also done.
  - ✓ Exploratory Data Analysis (EDA) was performed.
  - ✓ The Class Imbalance was handled.
  - ✓ The dataset was split into Training Dataset and Test Dataset.
  - ✓ Using different algorithms, the dataset was trained, accompanied by Cross-Validation and Hyperparameter Tuning.
  - ✓ The Model Performance were evaluated.
  - ✓ The Model with the best performance was saved in the form of a Pickle object.
  - ✓ The Model was deployed on a cloud platform.

- ***Reusability:***- The coding has been done in such a way that it can be reused without any issues.

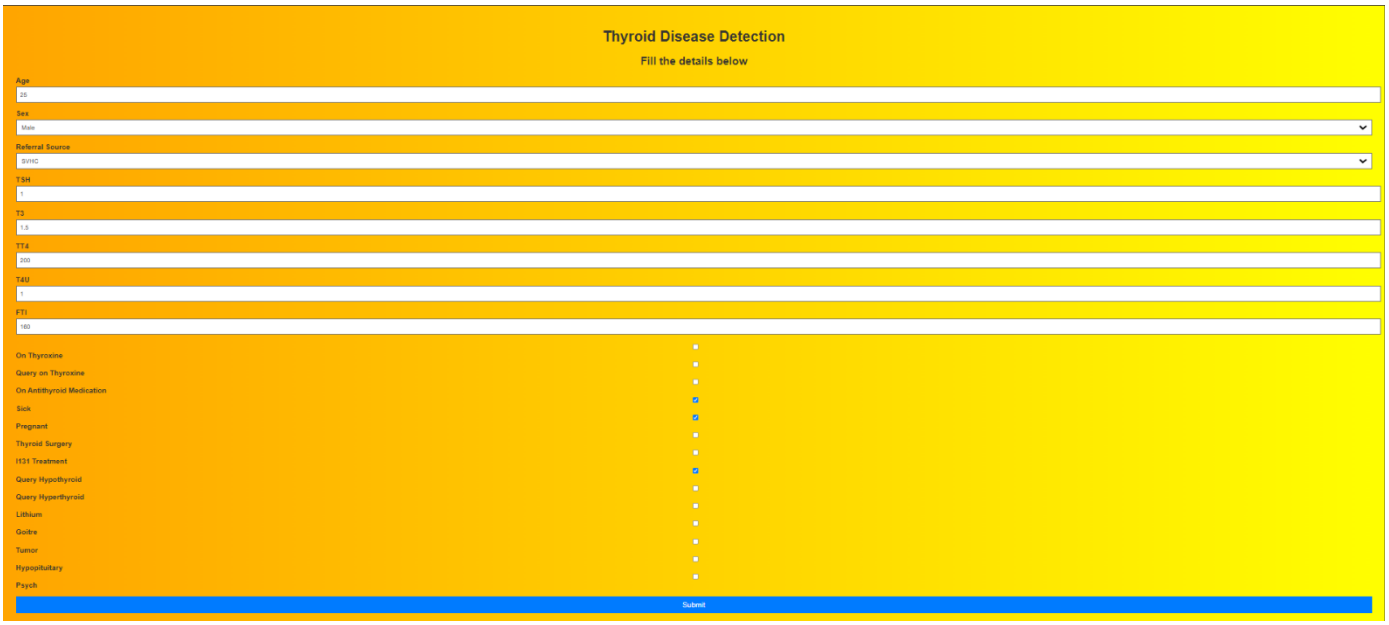
- ***Application Compatibility:***- Python will be used as an interface between the many components of this project. Every component will have a certain task to complete, and the Python will make sure that the information is transferred properly.

- ***Resource Utilization:***- Any task that is carried out will probably consume all of the available processing power until it is completed.

- ***Deployment:-*** The Model can be deployed on any of the given Cloud Platforms:



- ***User Interface:-*** This is the user interface:



## 6. CONCLUSION

The Thyroid Disease Detection project serves as an example of how machine learning may be effectively applied in the medical field to diagnose thyroid problems. The project successfully forecasts the likelihood of thyroid illnesses, including hyperthyroidism and hypothyroidism, by methodically performing data preprocessing, feature engineering, and model selection. Incorporating sophisticated algorithms into the diagnostic procedure not only improves prediction accuracy but also facilitates early identification, which is essential to averting serious consequences and enhancing patient outcomes.

The project's outcomes demonstrate how machine learning has the power to revolutionise conventional medical diagnostics by offering more accurate, data-driven insights. By using this method, medical professionals can make well-informed choices and create individualised treatment programs that are suited to the requirements of each patient. In addition, the architecture and technique of the project can be modified for related healthcare applications, demonstrating the adaptability and scalability of machine learning in this domain.

To sum up, the Thyroid Disease Detection project makes a substantial contribution to the healthcare sector by illuminating the use of machine learning to enhance patient care and diagnostic precision. Such cutting-edge methods will be crucial to solving difficult medical problems and improving patients' quality of life everywhere as healthcare continues to change.