Thyroid Disease Detection

High Level Design (HLD)

Contents

Abstract

The thyroid gland is crucial for preserving the body's metabolism. The use of medical data is made systematically possible by data mining in the healthcare sector. Today, thyroid disorders are the most prevalent. The thyroid gland's ability to function normally will not be hampered by early modifications. Thyroid replacement therapy and thyroid removal can be avoided to some extent by identifying thyroid diseases early and starting a better course of treatment.

1. Introduction

1.1. Why High-Level Design Document?

This High-Level Design (HLD) Document is meant to add the crucial information about this project. I'm going to go into detail about every detail of this project in this HLD Document.

2. General Description

2.1. Product Perspective

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

2.2 Problem statement

The beginning of thyroid disease is difficult to foresee in medical research, making it a common source of medical diagnosis and prediction. One of the most important organs in our body is the thyroid gland. Releases of thyroid hormone are in charge of controlling metabolism. One of the two thyroid conditions that frequently affect people, hyperthyroidism and hypothyroidism release thyroid hormones that control the body's metabolism. Predicting a patient's anticipated risk of developing thyroid illness is the major objective.

2.3 Proposed Solution

In the suggested remedy, we classified the various types of thyroid ailment using a machine learning model. Before building the machine learning model and

deploying it on a cloud platform, we first complete the data pre-processing stage, during which data transformation, handling missing values, feature transformation, and feature selection procedures are performed.

2.4 Technical Requirements

Following are the requirements of this project:

• Model should be deployed on cloud (Azure, AWS, GCP, Heroku).

2.5 Data Requirements

Data Requirement completely depend on our problem.

• We are utilising the Thyroid Disease dataset from the UCI site to train and test the model.

• From user we are taking following input

Feature Names:

• Class attribute (compensated hypothyroid = 0,

negative = 1, primary hypothyroid = 2,

secondary hypothyroid = 3)

• 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)

2.6 Tools Used

2.7 Data Requirements

• PyCharm is used as IDE.

• For visualization of the plots, Matplotlib, Seaborn and Plotly are used.

• Azure is used for deployment of the model.

•Front end development is done using HTML, Bootstrap, Flask is used for backend development and for API development.

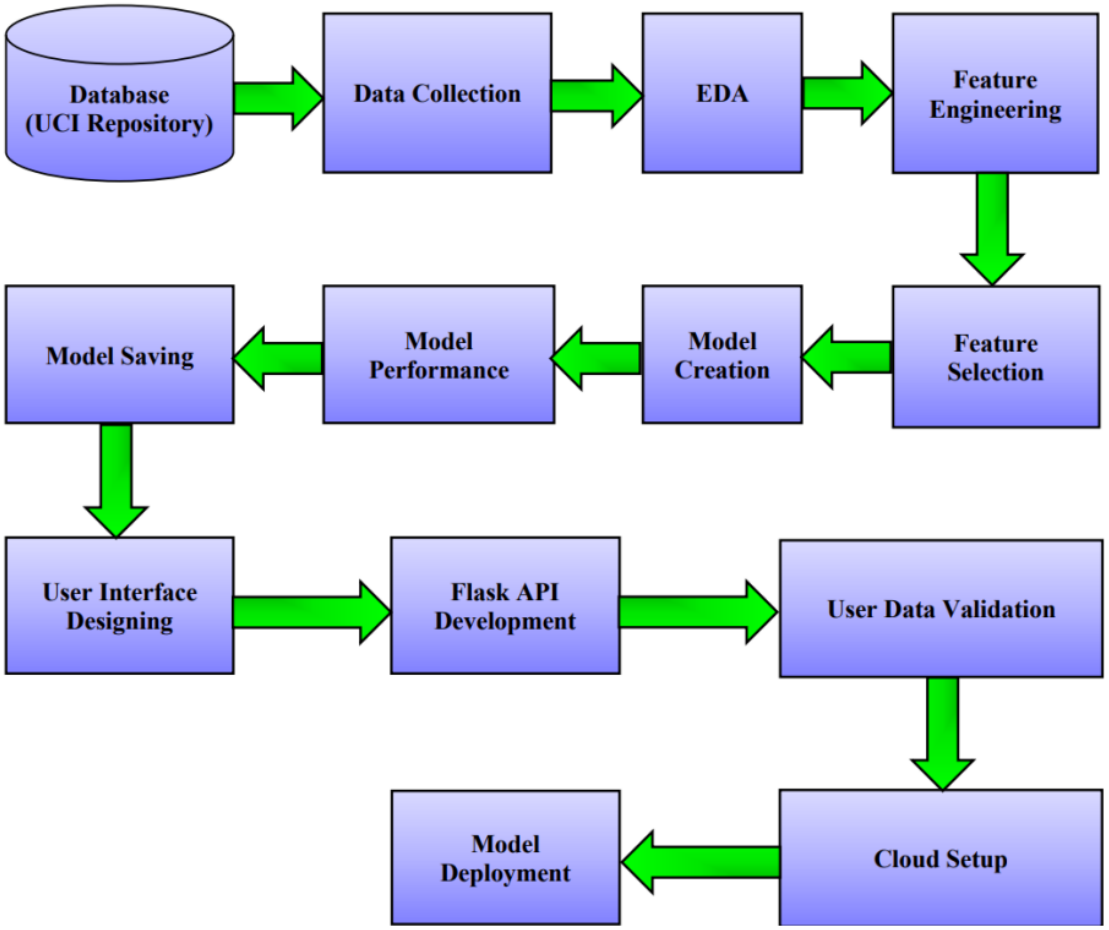
• GitHub is used as version control system

2.8 Constraints

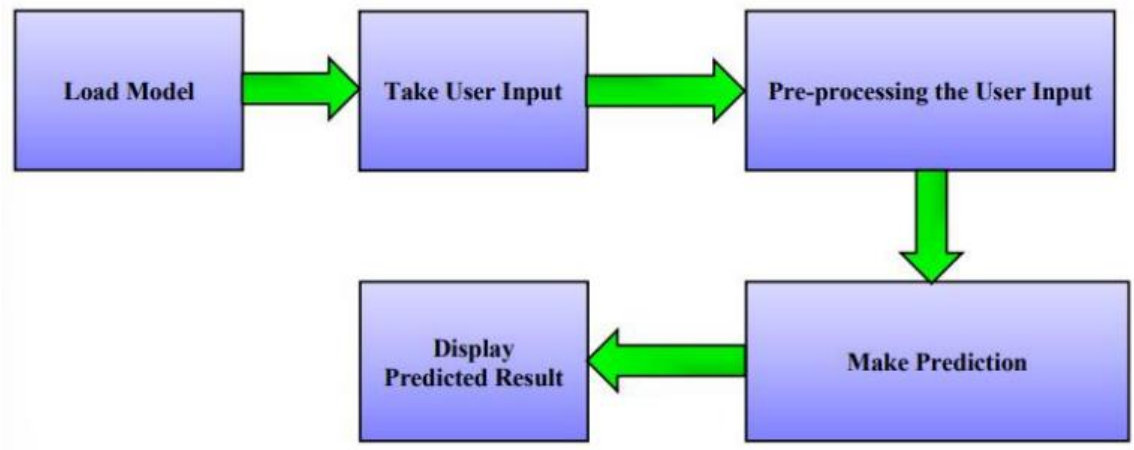
The Thyroid Disease Detection Model system needs to be user-friendly, error-free, and not require users to understand any of the workings of the back-end.

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, I cleaned the dataset properly by removing all null values and duplicate values present in dataset.

ii) After that EDA and feature transformation were performed.

iii) And then I performed feature selection process.

iv) Then I performed the encoding – numerical features and categorical features.

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

vi) After performing above, dataset was trained on different classification algorithm (KNN, Random Forest Classifier, Decision Tree, XG Boost etc.). After training the dataset on different algorithms, highest accuracy of 95.08 % on KNN Classifier was observed.

vii) After that I saved model in pickle file format.

viii) TO increase the accuracy of model further more we selected important features by using Extra trees classifier. And final model was built on selected features only. After feature selection, XGBoost classifier gave highest ROC of 0.988.

ix) After that I saved XGBoost classifier model in pickle file format.

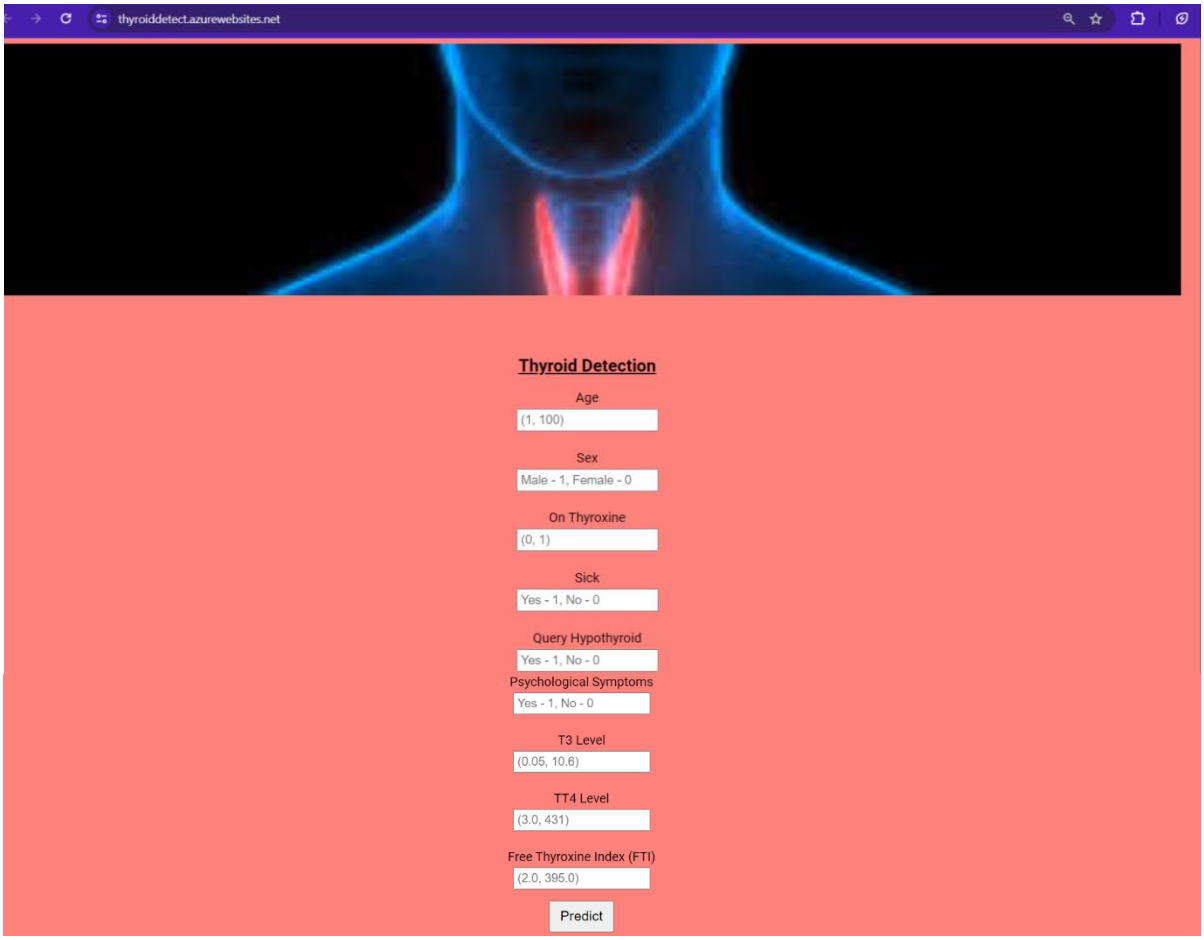
ix) After that model was ready to deploy, I deployed this model on Azure.

c) Deployment:

We have deployed this model on Azure cloud.

URL: https://thyroiddetect.azurewebsites.net/

4.3 User Interface



5. Conclusion

This project offers a system for recognizing the thyroid condition a user is experiencing, along with information about the condition and healthy lifestyle suggestions. For classification, XGBoost Classifier model is deployed.