i

Architecture

Thyroid Disease Detection System

|  |  |
| --- | --- |
| Written By | Ashwini Kakde |
| Document Version | 1.0 |
| Last Revised Date |  |

**Document Version Control**

**Change Record:**

|  |  |  |  |
| --- | --- | --- | --- |
| **Version** | **Date** | **Author** | **Comments** |
| **1.0** | **12-06-2023** | **Ashwini Kakde** | **Architecture** |
|  |  |  |  |
|  |  |  |  |
|  |  |  |  |

**Reviews:**

|  |  |  |  |
| --- | --- | --- | --- |
| **Version** | **Date** | **Reviewer** | **Comments** |
|  |  |  |  |

**Approval Status:**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Version** | **Review Date** | **Reviewed By** | **Approved By** | **Comments** |
|  |  |  |  |  |

# Contents

Document Version control 2 1 Architecture 4

1. Architecture Description 5
   1. Data Description 5 2.2 Export Data from DB to CSV for training 5
   2. Data Preprocessing 5
   3. Get best model of from different model 5

2.5Hyper parameter Tuning 5

* 1. Model saving 6
  2. Cloud setup 6
  3. Push App to cloud 6 2.9 Data from client side for prediction 6

2.10 Export prediction to CSV 6

Start

Export predicted data to csv

End

Best model saving

Cloud setup

Pushing app to cloud

Application

Start

Data from client to predict

Applying hyper parameter tuning

Applying different algorithm to gest best model

Data Transformation

Data Preprocessing

Export data from database to csv for training

# Architecture Description

3

## Data Description

We will be using Thyroid Disease Data Set present in UCI Machine Learning Repository. This Data set is satisfying our data requirement. Total 7200 instances present in different batches of data.

## Export Data from database to CSV for Training

Here we will be exporting all batches of data from database into one csv file for training.

## Data Preprocessing

We will be exploring our data set here and do EDA if required and perform data preprocessing depending on the data set. We first explore our data set in Jupyter Notebook and decide what pre-processing and Validation we have to do such as imputation of null values, dropping some column, etc and then we have to write separate modules according to our analysis, so that we can implement that for training as well as prediction data.

## Data Transformation

## The data transformation phases involves converting all categorical values into numerical format .Missing value imputation for numerical and categorical data .Numerical values are filled with median values and categorical values are filled with most frequent values and imbalanced data is handled by random over sampler.

## Get best model of different applied algorithm models

Here we will train various model on each cluster which we will obtain in Data Clustering, and then will try to get best model of each cluster.

## Hyper parameter Tuning

After selecting best model for each cluster, we will do hyper parameter tuning for each selected model, and try to increase performance of the models.

## Model Saving

After performing hyper parameter tuning for models, we will save our models so that we can use them for prediction purpose.

## Cloud Setup

Here we will do cloud setup for model deployment. Here we also create our

Flask app and user interface and integrate our model with flask app and UI

## Push app to cloud

After doing cloud setup and checking app locally, we will push our app to cloud to start the application.

## Data from client side for prediction purpose

Now our application on cloud is ready for doing prediction. The prediction data which we receive from client side will be exported from DB and further will do same data cleansing process as we have done for training data using modules we will write for training data. Client data will also go along the same process of **Exporting data from DB**, **Data pre-processing**, **and Data clustering** and according to each cluster number we will use our **saved model** for prediction on that cluster.

## Export Prediction to CSV

Finally when we get all the prediction for client data, then our final task is to export prediction to csv file and hand over it to client.