Architecture Design

# Thyroid Prediction

Revision Number – 1.1

## Last Date of Revision : 07-10-2022

Adil Anwar

Document Version Control

|  |  |  |  |
| --- | --- | --- | --- |
| Date | Version | Description | Author |
| 07-10-2022 | 1.0 | Abstract  Introduction  Architecture | Adil Anwar |
| 08-10-2022 | 1.1 | Architecture Design | Adil Anwar |

## Contents

Document Version Control ………………………………………2

Abstract …………………………………………………………... 4

1. Introduction ……………………………………………………………………………. 5
   1. Why this Architecture Design Document ? ………………………………….. 5
2. Architecture

….…………………………………………………………………………………… 5

1. Architecture Design

……………………………………………………………………………………… 5

* 1. Data Collection

……….…………………………………………………………………………. 5

* 1. Data Description

…..…………………………………………………………………………… 6

* 1. Importing Data into Database

…………………………………………………………….. 6

* 1. Exporting Data from Database

…………………………………………………………… 6

* 1. Data Preprocessing

………………………………………………………………………..…. 7

* 1. Modelling Process

………………....………………………………………………...…….….. 7

* 1. UI Integration ……………………………………………………………………………..…….. 7
  2. Data from User

………………………………………………………………………………..… 8

* 1. Data Validation

…….……………………………………………………………………….…... 8

* 1. Rendering the Results

…………….…………………….………….………………………. 8 3.11

Deployment

………………………………………….………………………………………...

8

### Abstract

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.

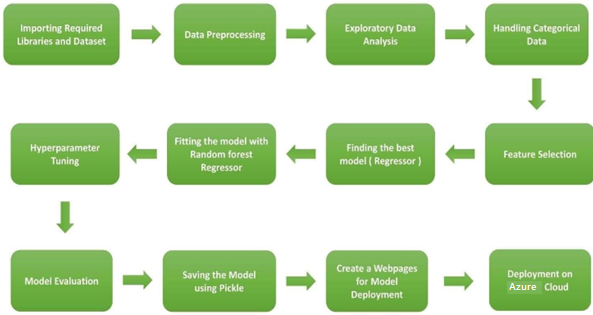
### 1. Introduction

1.1 Why this Architecture Design Document ?

The main objective of the Architecture design documentation is to provide the internal logic understanding of the flight fare prediction code. The Architecture design documentation is designed in such a way that the programmer can directly code after reading each module description in the documentation.

. Architecture Design

3

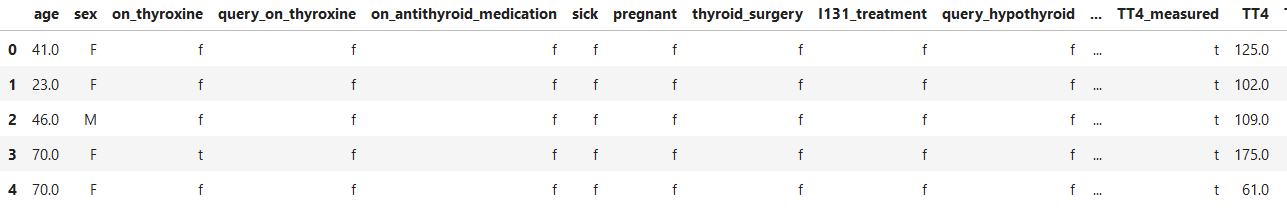


3.1 Data Collection

The data for these project is collected from the Kaggle Dataset, the URL for the dataset is <https://archive.ics.uci.edu/ml/datasets/thyroid+disease> Which is provided by UCI Machine Learning.

3.2 Data Description

Thyroid prediction is 3K+ dataset publicly available on the Kaggle. The information in the dataset is present in two separated excel files named as train.xlsx and test.xlsx. Dataset contains 3772 rows which shows the information such age, sex, on\_thyroxine, query\_on\_thyroxine, on\_antithyroid\_medication, sick, pregnant, thyroid\_surgery, I131\_treatment, query\_hypothyroid and classes. The glance of the Dataset is :



3.3 Importing data into Database

Created associate API for the transfer of the info into the Cassandra info, steps performed are :

* Connection is created with the info.
* Created a info with name ThyroidPrediction.
* cqlsh command is written for making the info table with needed parameters.
* And finally, a cqlsh command is written for uploading the Knowledge Set into data table by bulk insertion.

* 1. Exporting Data from Database

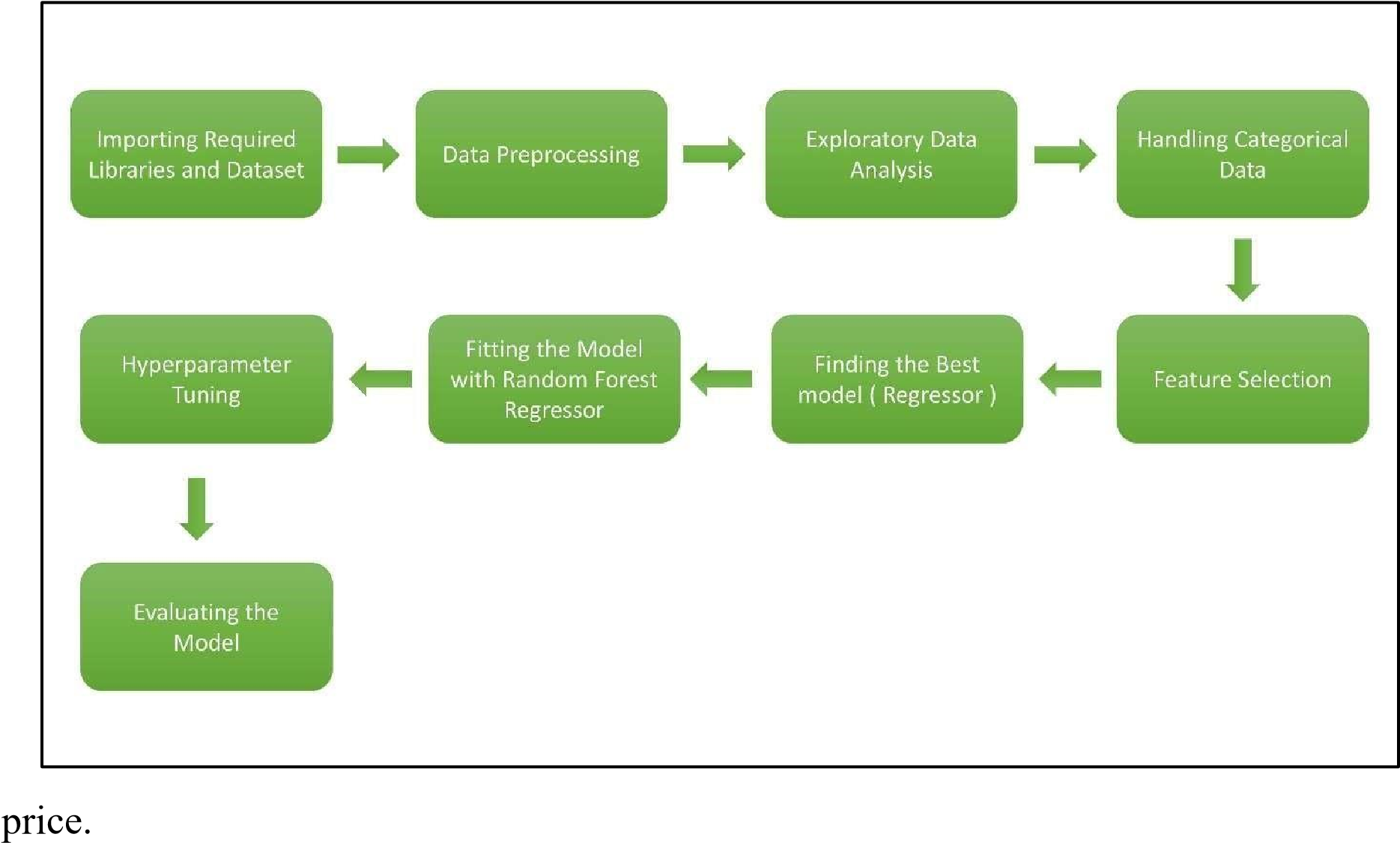
In the above created API, the download URL is also being created, which downloads the data into a csv file format.

* 1. Data Preprocessing
* Checked for info of the Dataset, to verify the correct datatype of the Columns.
* Checked for Null values, because the null values can affect the accuracy of the model.
* Converted all the desired columns into Datetime format.
* Performed One – Hot encoding on the desired columns.
* Checking the distribution of the columns to interpret its importance.

Now, the info is prepared to train a Machine Learning Model.

* 1. Modelling Process

After preprocessing the data, We visualize our data to gain insights and then these insights are randomly spread and split into two parts, train and test data. After splitting the data, we use Random Forest Regressor to model our data to predict the Flight Fare



* 1. UI Integration

Both CSS and HTML files are being created and are being integrated with the created machine learning model. All the required files are then integrated to the app.py file and tested locally.

* 1. Data from User

The data from the user is retrieved from the created HTML web page.

* 1. Data Validation

The data provided by the user is then being processed by app.py file and validated. The validated data is then sent to the prepared model for the prediction.

* 1. Rendering the Results

The data sent for the prediction is then rendered to the web page.

* 1. Deployment

The tested model is then deployed to Azure. So, users can access the project from any internet devices.