DETAILED PROJECT REPORT

THYROID DISEASE DETECTION

Project By: BHARATH.N.S

E-Mail: bharathksg100@gmail.com

Phone Number: 8431547679

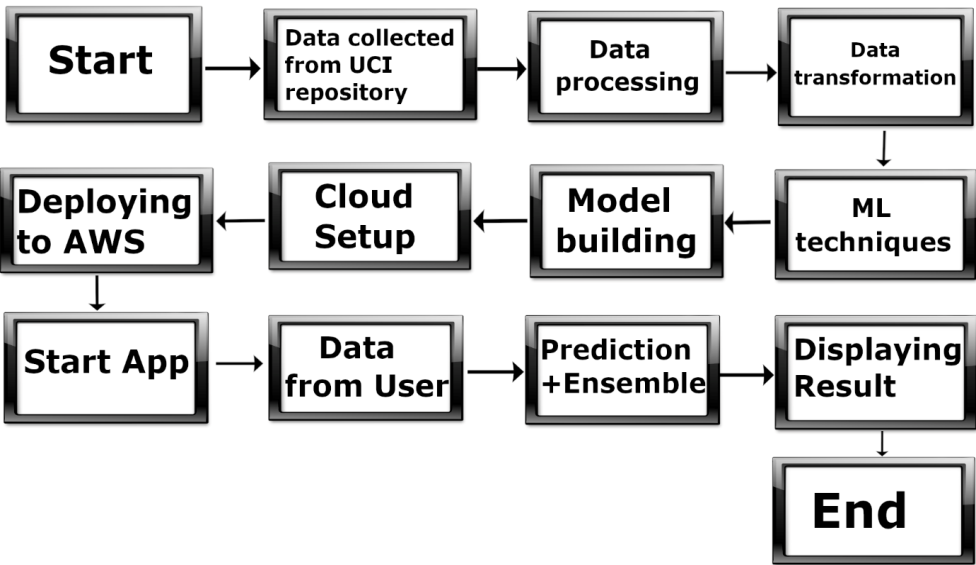
**Objective**

Development of a predictive model for detecting thyroidal disease for medical teams. The model will determine whether a patient has the risk of thyroidal disease.

**Benefits**

* Detection of thyroid disease
* Eliminates human error
* Higher efficiency
* Time efficient

**Architecture**



**Data Transformation**

* Removing of outliers
* Deletion of unnecessary features
* Imputation of null values
* Feature Selection
* Re-scaling the data into a range of 0 and 1
* Fixing the imbalanced data

**Model Training**

After data processing, data is sampled into training, testing and validation. Then training data is fitted into 6 different models and they are tested on testing data. Finally, Random Forest model are selected based on performance metrics.

**Model Selection**

Training data is used to train multiple models. Single models selected based on performance metrics. That is Random Forest .

**Prediction**

The Test data is used as a final test for our models. The output of the model is ensembled and the final values are tested.

**QUESTION AND ANSWER**

Q1) What’s the source of the data?

The entire data is collected from the UCI machine learning repository.

Q2) What was the type of data?

The data was a combination of numerical, categorical and null values.

Q3) How logs are managed?

Once the code is executed a log file will be created in the same repository.

Q4) What techniques were used for data processing?

* Removing unwanted features/columns
* Visualizing relation of independent variable with each other through heatmaps.
* Removing outliers
* Encoding the categorical features
* Fixing the imbalanced data
* Scaling the data

Q5) How training was done and what models were used?

After data processing, the training data was fitted into 6 different models. These included Logistic Regression, K-nearest Neighbours, Support vector machine, XGBoost, Decision Tree and Random Forest. Finally based on performance metrics Random Forest selected.

Q6) How prediction was done?

The final evaluation was made with the untouched test dataset. Same data processing steps were applied and prediction of model was ensembled and tested.

Q7) What are the different stages of deployment?

Streamlit app

* Once the models were tuned, they were saved as a pickle file
* Created model integrate with Streamlit app
* Entire file push into github
* Streamlit account was Created
* Requirement file was created which contained all the necessary libraries needed for the code to execute.
* Streamlit cloud integrate with Github
* Then entire model deploy in streamlit cloud

Flask app

* Flask app Created and also created html file .
* Then Random Forest model integrate with Flask application.
* Then deploy local environment Sucessfully.
* I try to deploy in AWS EC2 instance and GCP but it failed due to 502 Bad Gateway interface.