THYROID DISEASE DETECTION

Detailed Project Report

Umasangavi G
Data Science Intern at Ineuron.ai

INTRODUCTION

At least a person out of ten is suffered from thyroid disease in India. The disorder of thyroid disease primarily happens in the women having the age of 17–54. The extreme stage of thyroid results in cardiovascular complications, increase in blood pressure, maximizes the cholesterol level, depression and decreased fertility. The hormones, **total serum thyroxin (T4)** and **total serum triiodothyronine (T3)** are the two active thyroid hormones produced by the thyroid gland to control the metabolism of body. For the functioning of each cell and each tissue and organ in a right way, in overall energy yield and regulation and to generate proteins in the ordnance of body temperature, these hormones are necessary .

Hyperthyroidism and **Hypothyroidism** are the most two common diseases caused by irregular function of thyroid gland. Thyroid disorder can speed up or slow down the metabolism of the body. In the world of rising new technology and innovation, health care industry is advancing with the role of Artificial Intelligence. Machine learning algorithms can help to early detection of the disease and to improve the quality of the life. This study demonstrates the how different classification algorithms can forecasts the presence of the disease.

OBJECTIVE

The main goal of this project is to predict the risk of hyperthyroid and hypothyroid based on various factors of individuals. Thyroid disease is a common cause of medical diagnosis and prediction, with an on set that is difficult to fore cast in medical research. It will play a decisive role in order to early detection, accurate identification of the disease and helps the doctors to make proper decisions and better treatment.

Architecture
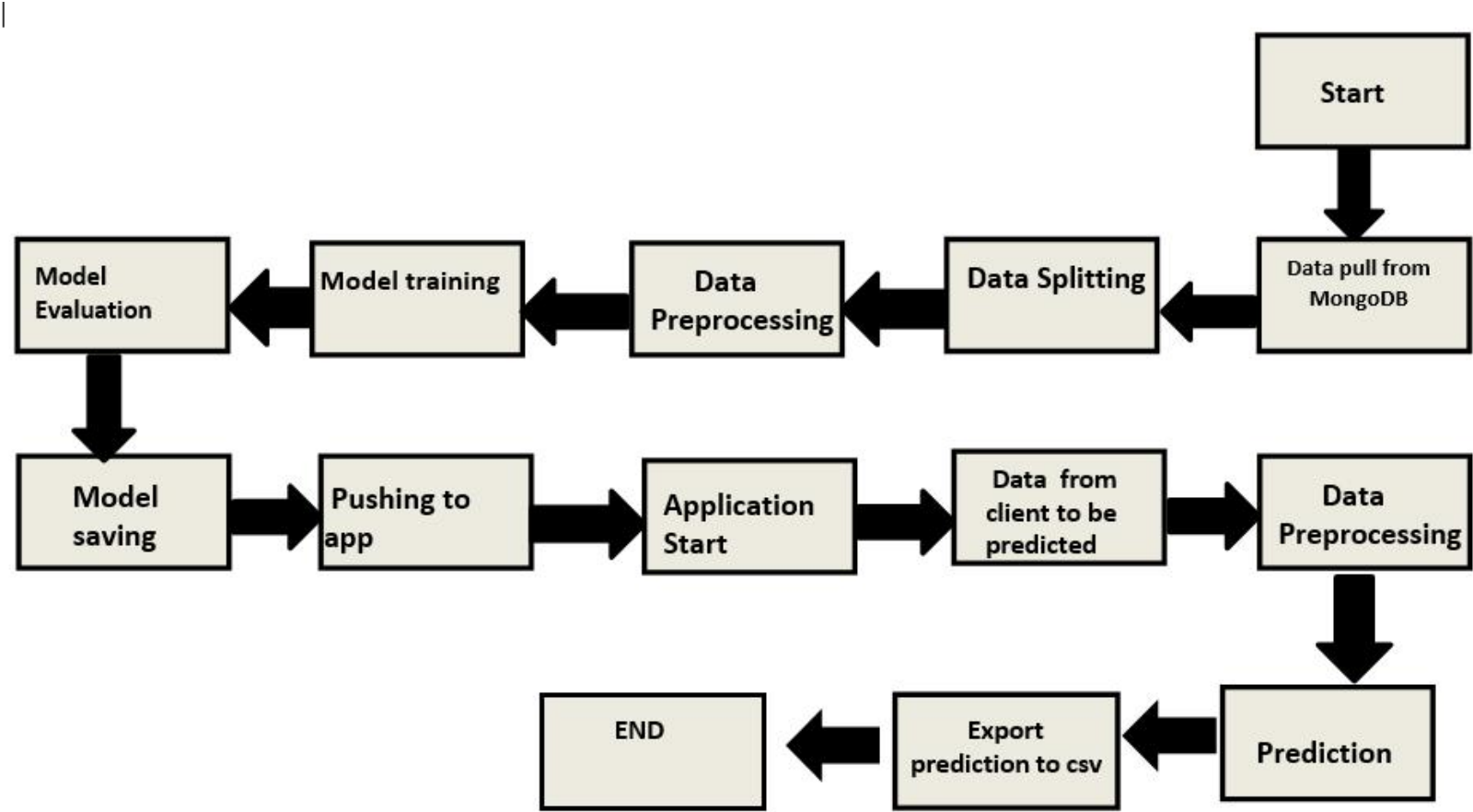
DATASET

age:	continuous,?.
sex:	M,F,?.
on_thyroxine:	f,t.
query_on_thyroxine:	f,t.
on_antithyroid_medication:	f,t.
thyroid_surgery:	f,t.
query_hypothyroid:	f,t.
query_hyperthyroid:	f,t.
pregnant:	f,t.
sick:	f,t.
tumor:	f,t.
lithium:	f,t.
goitre:	f,t.
TSH_measured:	f,t.
TSH:	continuous,?.
T3_measured:	f,t.
T3:	continuous,?.
TT4_measured:	f,t.
TT4:	continuous,?.
T4U_measured:	f,t.
T4U:	continuous,?.
FTI_measured:	f,t.
FTI:	continuous,?.
TBG_measured:	f,t.
TBG:	continuous,?.

MODEL TRAININGAND VALIDATION WORKFLOW
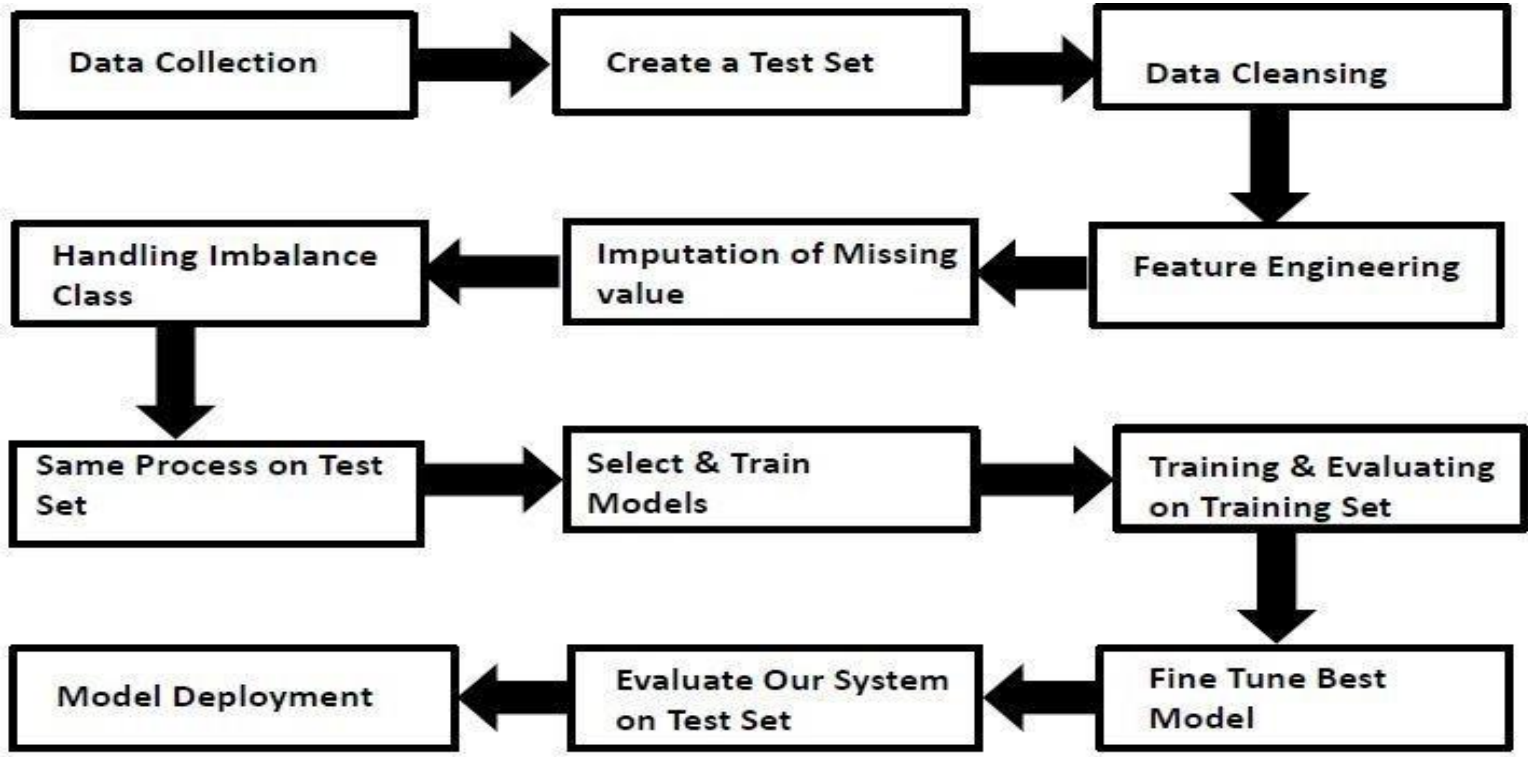
DEPLOYMENT PROCESS
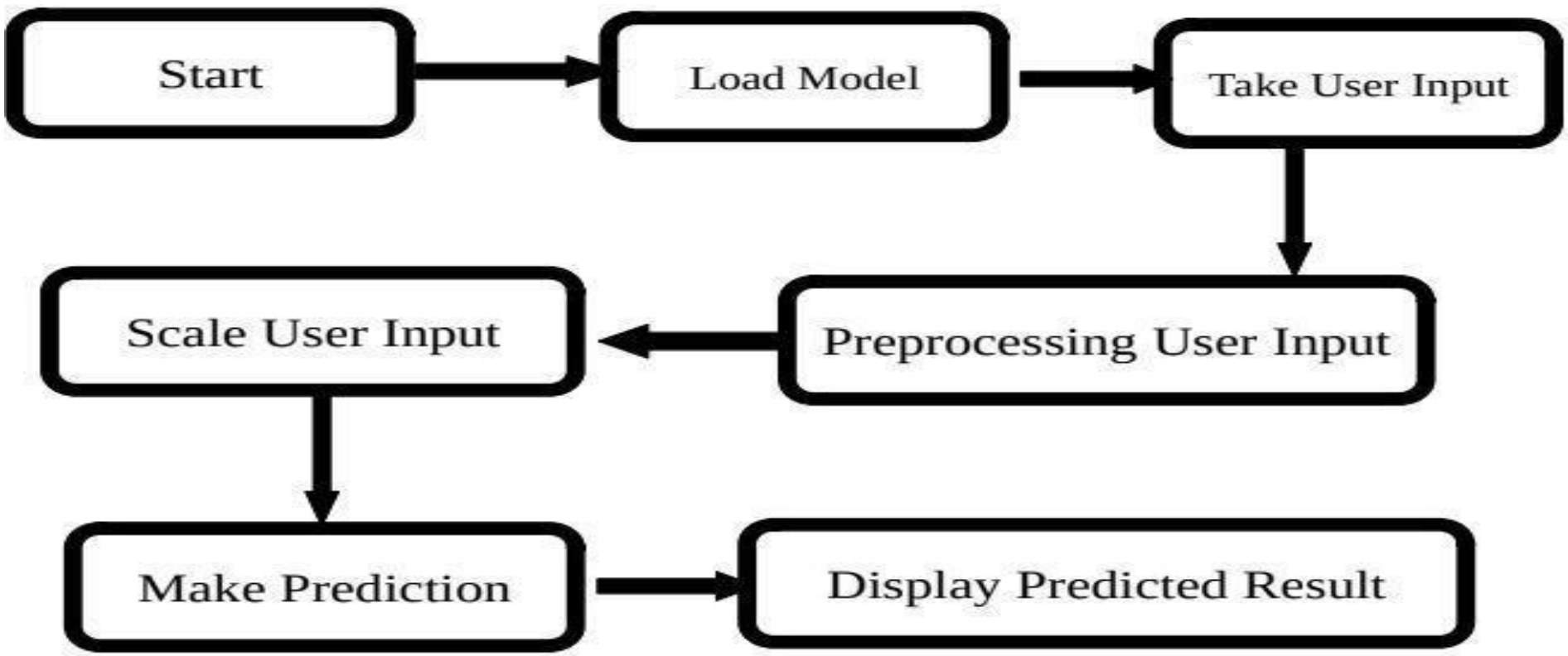
WORKFLOW

Data Collection

- Thyroid Disease Data Set from UCI Machine Learning Repository

- For Data Set: https://archive.ics.uci.edu/ml/datasets/thyroid+disease

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 3772 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 Splitting

We filter the columns for splitting the data for train and test for further uses

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, 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 Training

We trained a RandomForestClassifier model in our notebook and was good on it. We trained with our processed data.

Model Evaluation

Model evaluation done by classification and report was saved to .pkl file

Model Saving

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

Model Deployment

The final model is deployed on Heroku using Flask framework

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.

Data processing and Prediction

Client data will also go along the same process **Data pre-processing** and according to that we will predict those data.

Export Prediction to CSV

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

FREQUENTLY ASKED QUESTIONS

Q1) What is the source of data?

The data for training is obtained from famous machine learning repository.

UCI Machine Learning Repository: https://archive.ics.uci.edu/ml/datasets/thyroid+disease

Q2) What was the type of data?

The data was the combination of numerical and Categorical values.

Q3) What's the complete flow you followed in this Project?

Refer slide 7th, 8th and 9th for better understanding.

Q4) After the File validation what you do with incompatible file or files which didn't passthe validation?

Files like these are moved to the Achieve Folder and a list of these files has been shared with the client and we removedthe bad data folder.

Q5) How logs are managed?

We are using different logs as per the steps that we follow in training and prediction like model training log andprediction log etc. And then sub log are inside those folder.

Q 6) What techniques were you using for data pre-processing?

- Removing unwanted attributes

- Visualizing relation of independent variables with each other and output variables

- Checking and changing Distribution of continuous values

- Cleaning data and imputing if null values are present.

- Converting categorical data into numeric values.

Q 7) How training was done or what models were used?

- First Data validation done on raw data and then good data insertion happen in DB.

- Then Data preprocessing done on final CSV file received from DB.
  We trained a RandomForestClassifier model in our notebook and was good on it. We
  trained with our processed data.

Q 8) How Prediction was done?

Our application on cloud is ready for doing prediction. The prediction data which we receive
from client side.

Client data will also go along the same process **Data pre-processing** and
according to that we will predict those data.

Q 9) What are the different stages of deployment?

- After model training and finalizing all models. We created required files for deployment.

- Finally deployed our model over various cloud platforms such as Heroku .

Q 10) How is the User Interface present for this project?

- For this project I have made one user input prediction.

- UI are very user friendly and easy to use.

THANK YOU