Thyroid Disease Classification Using Machine Learning

Team Members Name : Sathyapriya.E

Sharmiladevi.S

Nivetha.M

Narmatha.J

Class : BSc.Computer Science

1.INTRODUCTION

1.1 Overview:

- The Thyroid gland is a vascular gland and one of the most important organs of the human body. This gland secretes two hormones which help in controlling the metabolism of the body. The two types of Thyroid disorders are Hyperthyroidism and Hypothyroidism. When this disorder occursin the body, they release certain types of hormones into the body which imbalances the body's metabolism. A thyroid-related Blood test is used to detect this disease but it is often blurred an noise will be present. Data cleansing methods were used to make the data primitive enough for the analytics to show the risk of patients getting this disease. Machine Learning plays a very deciding role in disease prediction. Machine Learning algorithms, SVM - support vector machine, Random Forest Classifier, XGB Classifier and ANN - Artificial Neural Networks are used to predict the patient's risk of getting thyroid disease. The web app is created to get data from users to predict the type of disease. Thyroid diseases are increasing in magnitude everyday and spreading all over the world.
- The thyroid gland is a vascular and one of the most important organs of the humans body.
- Thyroid gland secrets two harmones which helps in controlling the metabolism of the body.
- Normal thyroid stimulating hormone levels generally fall between o.4 and 4.0 milliunits per liter. Tsh levels higher than 4. mU/L usually indicate an underactive thyroid (hypothyroidism), and low TSH levels below 0.4 mU/L indicate an overactive thyroid (hyperthyroidism).

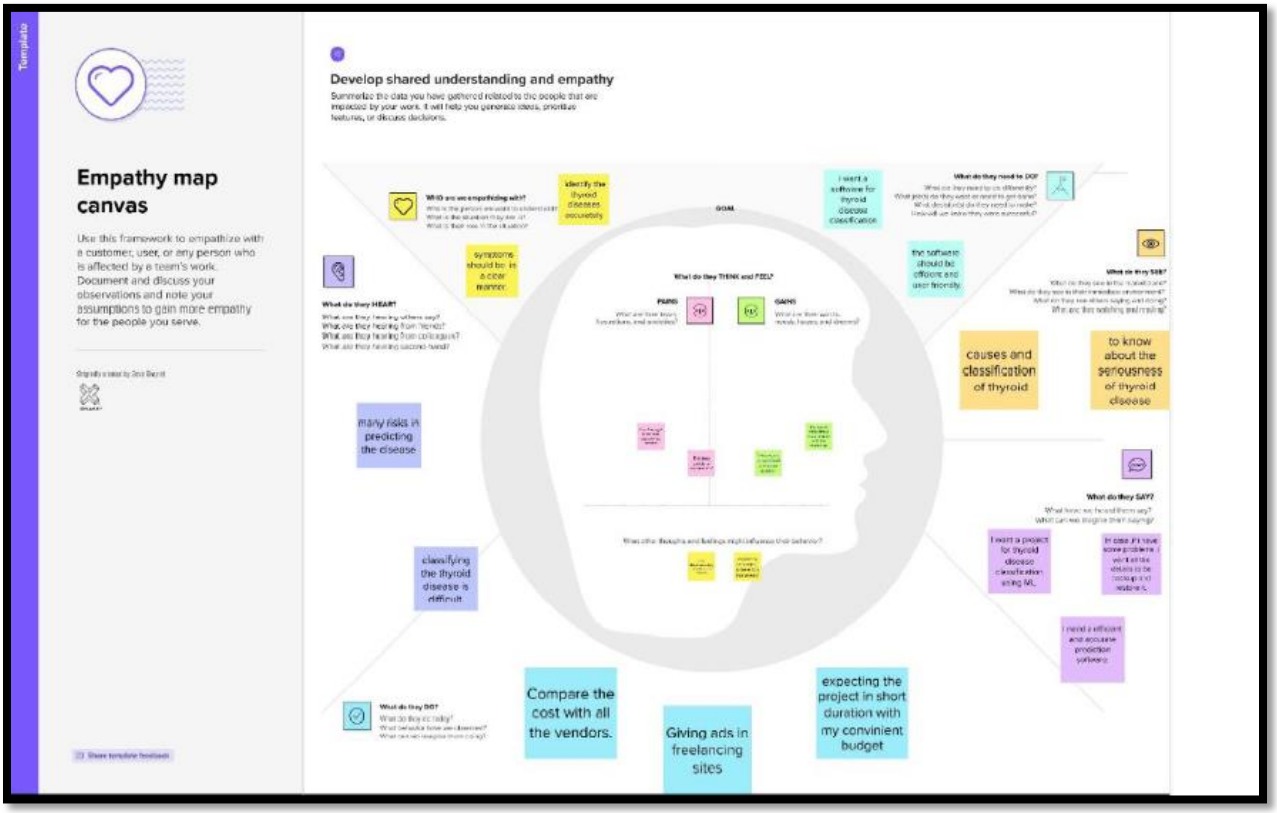
1.2 Purpose:

- The goal of this study is to categorize **thyroid disease** into three categories: **hyperthyroidism**, hypothyroidism, and normal.
- The main purpose is to search the best classification approach for thyroid disease diagnosis by making the comparison of decision tree algorithms.
- In the line of this purpose, the experiments are conducted to compare different kinds of decision tree algorithms given in the previous section.

2.Problem Definition & Design Thinking

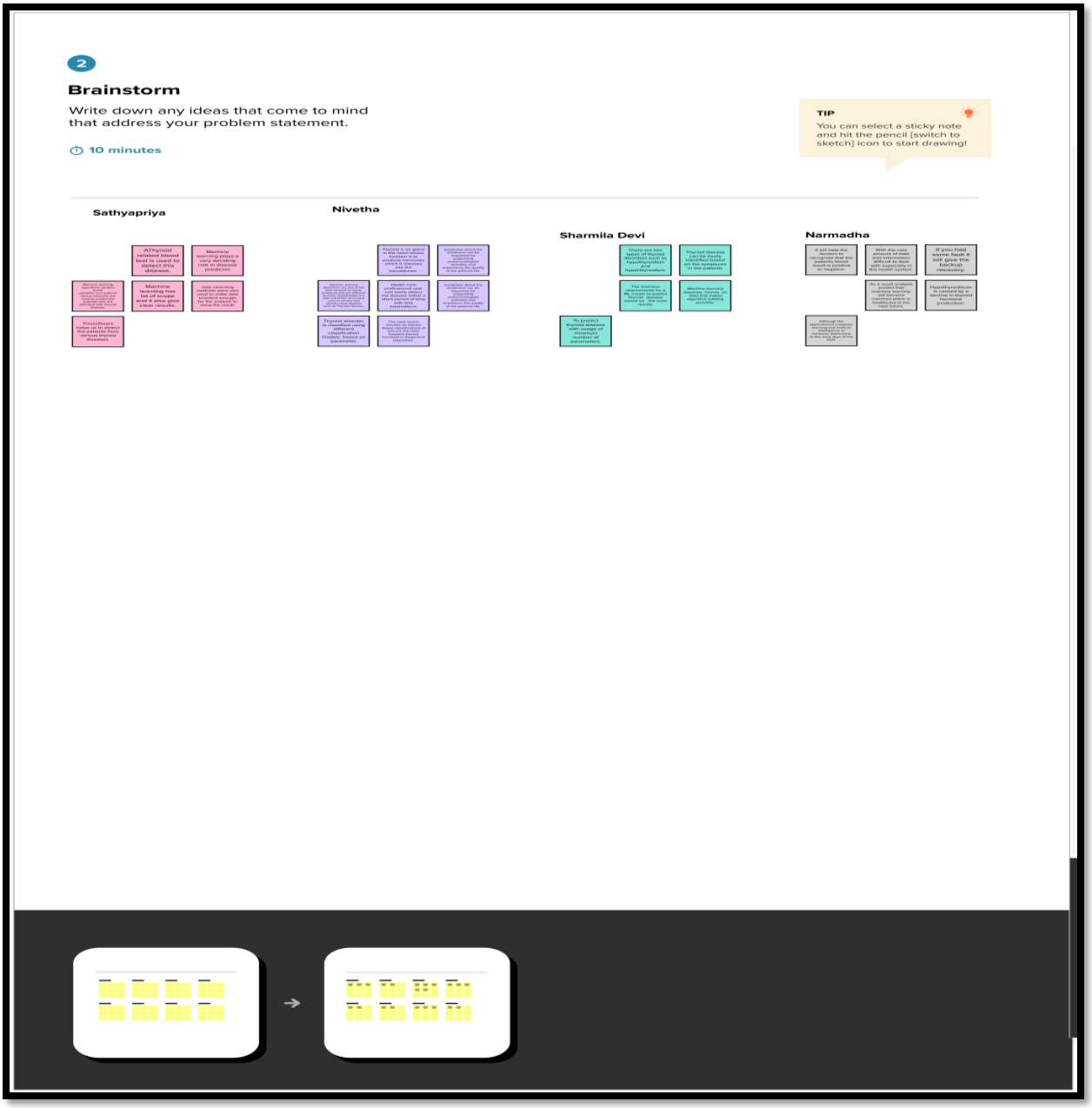
2.1. Empathy map:

In the ideation phase we have empathized have a client placement trends analysis and we have acquired details. Which are represented in the Empathy Map



2.2 Ideation & Brainstroming Map:

- Under this activity our team members or gathered and discussed various ideas to solve our project problem each member contributed 6 to 10 times.
- After gathering all ideas we have asses the impact and feasibility of each point. Finally we have assign the priority for each point based on this values.


2
Brainstorm
Write down any ideas that come to mind that address your problem statement.
10 minutes
TIP
You can select a sticky note and hit the pencil [switch to sketch] icon to start drawing!
Sathyapriya
Nivetha
Sharmila Devi
Narmadha

3. RESULT:

Read the Dataset:

	age	sex	on_thyroxine	query_on_thyroxine	on_antithyroid_meds	sick	pregnant	thyroid_surgery	I131_treatment	query_hypothyroid	...	TT4	T4U_measured	T4U	FTI_measured	FTI	TBG_measured	TBG	referral_source	target	patient_id
0	29	F	f	f	f	f	f	f	f	t	...	NaN	f	NaN	f	NaN	f	NaN	other	-	840801013
1	29	F	f	f	f	f	f	f	f	f	...	128.0	f	NaN	f	NaN	f	NaN	other	-	840801014
2	41	F	f	f	f	f	f	f	f	f	...	NaN	f	NaN	f	NaN	t	11.0	other	-	840801042
3	36	F	f	f	f	f	f	f	f	f	...	NaN	f	NaN	f	NaN	t	26.0	other	-	840803046
4	32	F	f	f	f	f	f	f	f	f	...	NaN	f	NaN	f	NaN	t	36.0	other	S	840803047
...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
9167	56	M	f	f	f	f	f	f	f	f	...	64.0	t	0.83	t	77.0	f	NaN	SVI	-	870119022
9168	22	M	f	f	f	f	f	f	f	f	...	91.0	t	0.92	t	99.0	f	NaN	SVI	-	870119023
9169	69	M	f	f	f	f	f	f	f	f	...	113.0	t	1.27	t	89.0	f	NaN	SVI	I	870119025
9170	47	F	f	f	f	f	f	f	f	f	...	75.0	t	0.85	t	88.0	f	NaN	other	-	870119027
9171	31	M	f	f	f	f	f	f	f	t	...	66.0	t	1.02	t	65.0	f	NaN	other	-	870119035

9172 rows × 31 columns

Checking for null values:

	age	sex	on_thyroxine	query_on_thyroxine	on_antithyroid_meds	sick	pregnant	thyroid_surgery	I131_treatment	query_hypothyroid	...	TT4	T4U_measured	T4U	FTI_measured	FTI	TBG_measured	TBG	referral_source	target	patient_id
0	29	F	f	f	f	f	f	f	f	t	...	NaN	f	NaN	f	NaN	f	NaN	other	-	840801013
1	29	F	f	f	f	f	f	f	f	f	...	128.0	f	NaN	f	NaN	f	NaN	other	-	840801014
2	41	F	f	f	f	f	f	f	f	f	...	NaN	f	NaN	f	NaN	t	11.0	other	-	840801042
3	36	F	f	f	f	f	f	f	f	f	...	NaN	f	NaN	f	NaN	t	26.0	other	-	840803046
4	32	F	f	f	f	f	f	f	f	f	...	NaN	f	NaN	f	NaN	t	36.0	other	S	840803047
...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
9167	56	M	f	f	f	f	f	f	f	f	...	64.0	t	0.83	t	77.0	f	NaN	SVI	-	870119022
9168	22	M	f	f	f	f	f	f	f	f	...	91.0	t	0.92	t	99.0	f	NaN	SVI	-	870119023
9169	69	M	f	f	f	f	f	f	f	f	...	113.0	t	1.27	t	89.0	f	NaN	SVI	I	870119025
9170	47	F	f	f	f	f	f	f	f	f	...	75.0	t	0.85	t	88.0	f	NaN	other	-	870119027
9171	31	M	f	f	f	f	f	f	f	t	...	66.0	t	1.02	t	65.0	f	NaN	other	-	870119035

9172 rows × 31 columns

	age	sex	on_thyroxine	query_on_thyroxine	on_antithyroid_meds	sick	pregnant	thyroid_surgery	I131_treatment	query_hypothyroid	...	TT4	T4U_measured	T4U	FTI_measured	FTI	TBG_measured	TBG	referral_source	target	patient_id
0	29	F	f	f	f	f	f	f	f	t	...	NaN	f	NaN	f	NaN	f	NaN	other	Z	840801013
1	29	F	f	f	f	f	f	f	f	f	...	128.0	f	NaN	f	NaN	f	NaN	other	Z	840801014
2	41	F	f	f	f	f	f	f	f	f	...	NaN	f	NaN	f	NaN	t	11.0	other	Z	840801042
3	36	F	f	f	f	f	f	f	f	f	...	NaN	f	NaN	f	NaN	t	26.0	other	Z	840803046
4	32	F	f	f	f	f	f	f	f	f	...	NaN	f	NaN	f	NaN	t	36.0	other	S	840803047
...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
9167	56	M	f	f	f	f	f	f	f	f	...	64.0	t	0.83	t	77.0	f	NaN	SVI	Z	870119022
9168	22	M	f	f	f	f	f	f	f	f	...	91.0	t	0.92	t	99.0	f	NaN	SVI	Z	870119023
9169	69	M	f	f	f	f	f	f	f	f	...	113.0	t	1.27	t	89.0	f	NaN	SVI	I	870119025
9170	47	F	f	f	f	f	f	f	f	f	...	75.0	t	0.85	t	88.0	f	NaN	other	Z	870119027
9171	31	M	f	f	f	f	f	f	f	t	...	66.0	t	1.02	t	65.0	f	NaN	other	Z	870119035

9172 rows × 31 columns

```
age                        0
sex                      307
on_thyroxine               0
query_on_thyroxine         0
on_antithyroid_meds        0
sick                       0
pregnant                   0
thyroid_surgery            0
I131_treatment             0
query_hypothyroid          0
query_hyperthyroid         0
lithium                    0
goitre                     0
tumor                      0
hypopituitary              0
psych                      0
TSH_measured               0
TSH                      842
T3_measured                0
T3                      2604
TT4_measured               0
TT4                      442
T4U_measured               0
T4U                      809
FTI_measured               0
FTI                      802
TBG_measured               0
TBG                     8823
referral_source            0
target                     0
patient_id                 0
dtype: int64
```

	age	sex	on_thyroxine	query_on_thyroxine	on_antithyroid_meds	sick	pregnant	thyroid_surgery	I131_treatment	query_hypothyroid	...	hypopituitary	psych	TSH	T3	TT4	T4U	FTI	referral_source	target	patient_id
0	29	F	f	f	f	f	f	f	f	t	...	f	f	0.3	NaN	NaN	NaN	NaN	other	Z	840801013
1	29	F	f	f	f	f	f	f	f	f	...	f	f	1.6	1.9	128.0	NaN	NaN	other	Z	840801014
2	41	F	f	f	f	f	f	f	f	f	...	f	f	NaN	NaN	NaN	NaN	NaN	other	Z	840801042
3	36	F	f	f	f	f	f	f	f	f	...	f	f	NaN	NaN	NaN	NaN	NaN	other	Z	840803046
4	32	F	f	f	f	f	f	f	f	f	...	f	f	NaN	NaN	NaN	NaN	NaN	other	S	840803047
...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
9167	56	M	f	f	f	f	f	f	f	f	...	f	f	NaN	NaN	64.0	0.83	77.0	SVI	Z	870119022
9168	22	M	f	f	f	f	f	f	f	f	...	f	f	NaN	NaN	91.0	0.92	99.0	SVI	Z	870119023
9169	69	M	f	f	f	f	f	f	f	f	...	f	f	NaN	NaN	113.0	1.27	89.0	SVI	I	870119025
9170	47	F	f	f	f	f	f	f	f	f	...	f	f	NaN	NaN	75.0	0.85	88.0	other	Z	870119027
9171	31	M	f	f	f	f	f	f	f	t	...	f	f	NaN	NaN	66.0	1.02	65.0	other	Z	870119035

9172 rows × 24 columns

```
age                        0
sex                      307
on_thyroxine               0
query_on_thyroxine         0
on_antithyroid_meds        0
sick                       0
pregnant                   0
thyroid_surgery            0
I131_treatment             0
query_hypothyroid          0
query_hyperthyroid         0
lithium                    0
goitre                     0
tumor                      0
hypopituitary              0
psych                      0
TSH                      842
T3                      2604
TT4                      442
T4U                      809
FTI                      802
referral_source            0
target                     0
patient_id                 0
dtype: int64
```

```
age                         0
sex                         0
on_thyroxine                0
query_on_thyroxine          0
on_antithyroid_meds         0
sick                        0
pregnant                    0
thyroid_surgery             0
I131_treatment              0
query_hypothyroid           0
query_hyperthyroid          0
lithium                     0
goitre                      0
tumor                       0
hypopituitary               0
psych                       0
TSH                       842
T3                       2604
TT4                       442
T4U                       809
FTI                       802
referral_source             0
target                      0
patient_id                  0
dtype: int64
```

```
age                    0
sex                    0
on_thyroxine           0
query_on_thyroxine     0
on_antithyroid_meds    0
sick                   0
pregnant               0
thyroid_surgery        0
I131_treatment         0
query_hypothyroid      0
query_hyperthyroid     0
lithium                0
goitre                 0
tumor                  0
hypopituitary          0
psych                  0
TSH                    0
T3                     0
TT4                    0
T4U                    0
FTI                    0
referral_source        0
target                 0
patient_id             0
dtype: int64
```

```
Data columns (total 24 columns):
 #   Column                Non-Null Count  Dtype
---  ------                --------------  -----
 0   age                   9172 non-null   int64
 1   sex                   9172 non-null   float64
 2   on_thyroxine          9172 non-null   object
 3   query_on_thyroxine    9172 non-null   object
 4   on_antithyroid_meds   9172 non-null   object
 5   sick                  9172 non-null   object
 6   pregnant              9172 non-null   object
 7   thyroid_surgery       9172 non-null   object
 8   I131_treatment        9172 non-null   object
 9   query_hypothyroid     9172 non-null   object
 10  query_hyperthyroid    9172 non-null   object
 11  lithium               9172 non-null   object
 12  goitre                9172 non-null   object
 13  tumor                 9172 non-null   object
 14  hypopituitary         9172 non-null   object
 15  psych                 9172 non-null   object
 16  TSH                   9172 non-null   float64
 17  T3                    9172 non-null   float64
 18  TT4                   9172 non-null   float64
 19  T4U                   9172 non-null   float64
 20  FTI                   9172 non-null   float64
 21  referral_source       9172 non-null   object
```

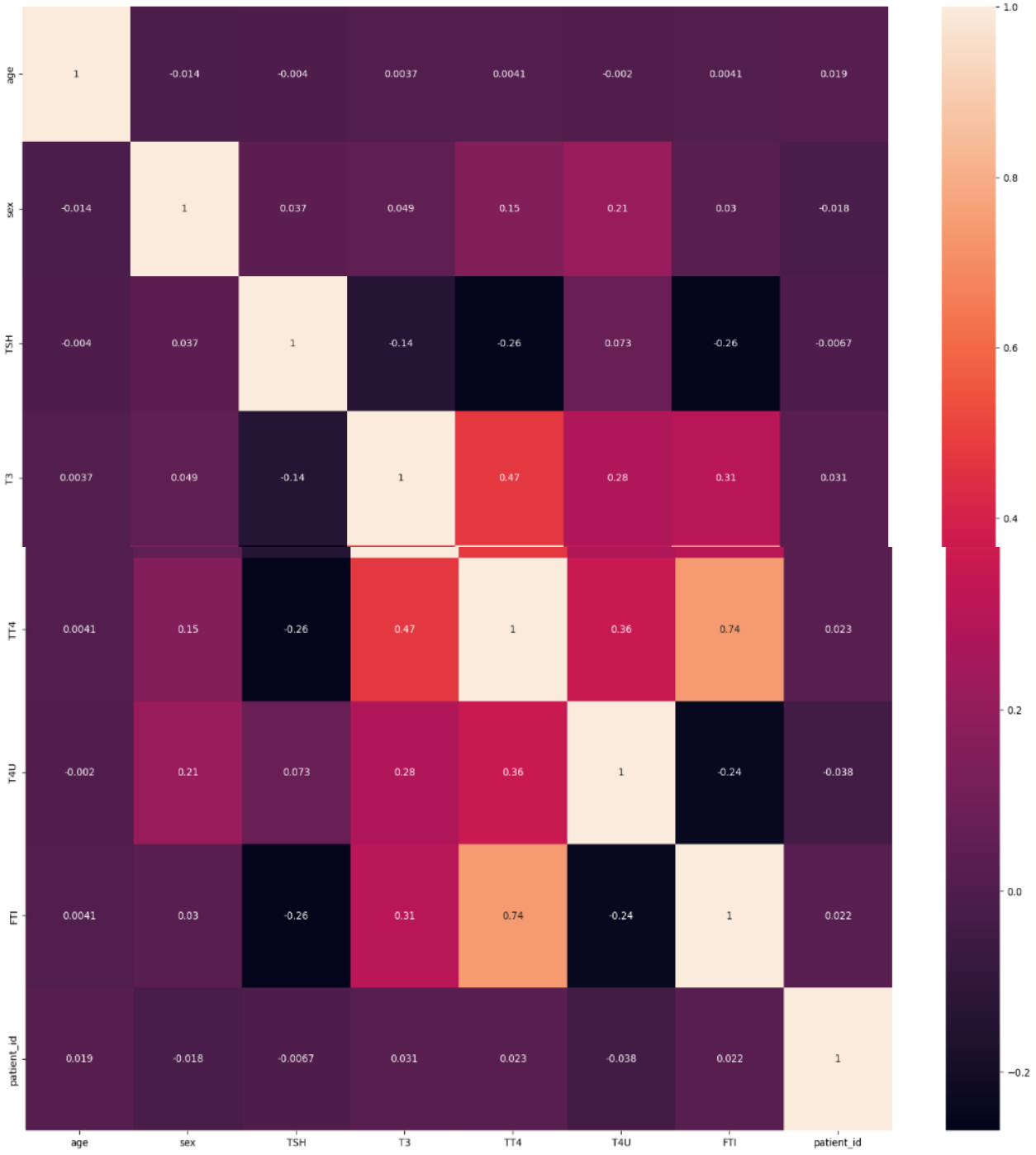
```
 22  target                9172 non-null   object
 23  patient_id            9172 non-null   int64
dtypes: float64(6), int64(2), object(16)
memory usage: 1.7+ MB
```

Exploratory Data Analysis:

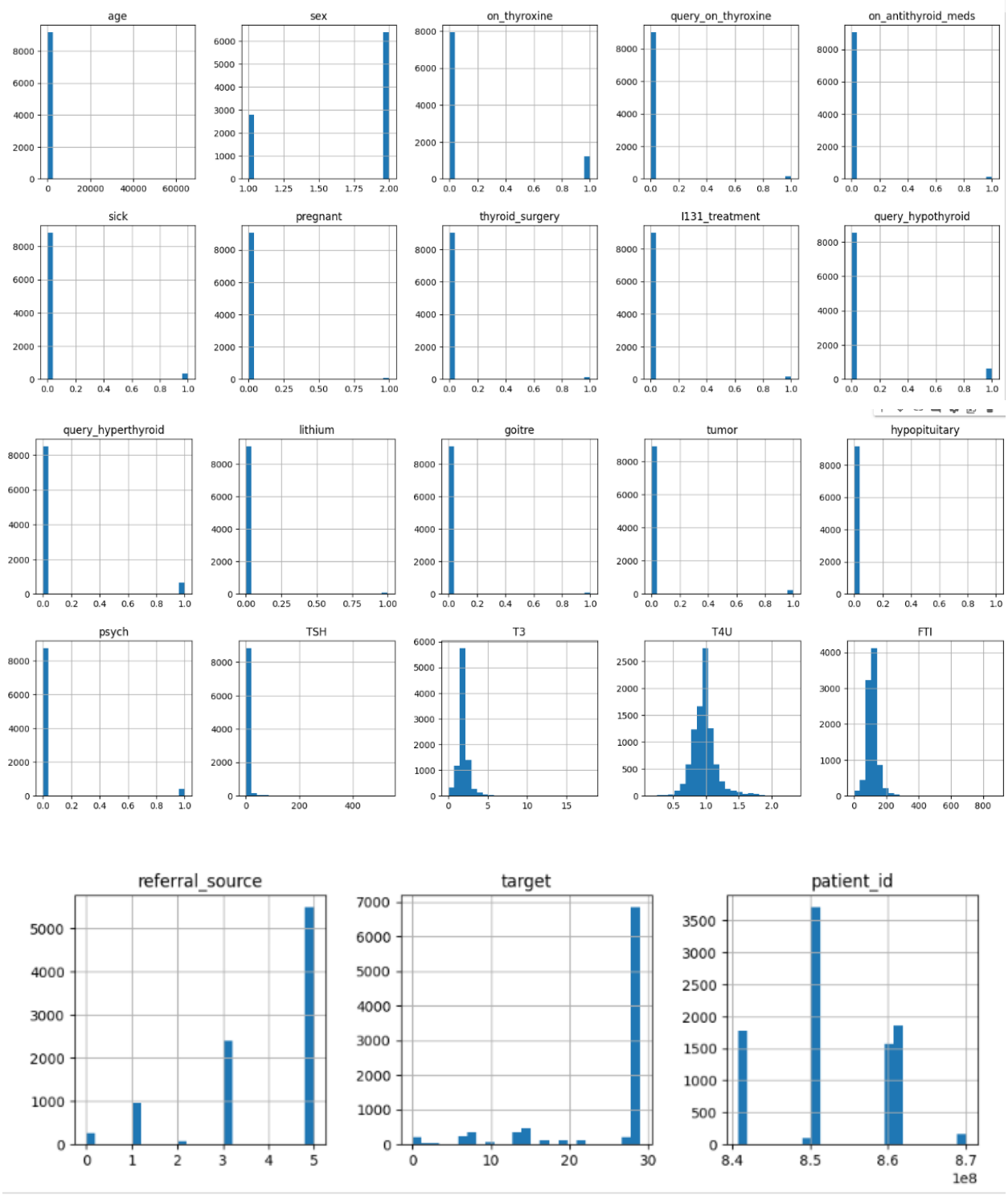
Descriptive analysis:

```
age                        0
sex                        0
on_thyroxine               0
query_on_thyroxine         0
on_antithyroid_meds        0
sick                       0
pregnant                   0
thyroid_surgery            0
I131_treatment             0
query_hypothyroid          0
query_hyperthyroid         0
lithium                    0
goitre                     0
tumor                      0
hypopituitary              0
psych                      0
TSH                        0
T3                         0
TT4                        0
T4U                        0
FTI                        0
referral_source            0
target                     0
patient_id                 0
dtype: int64
```

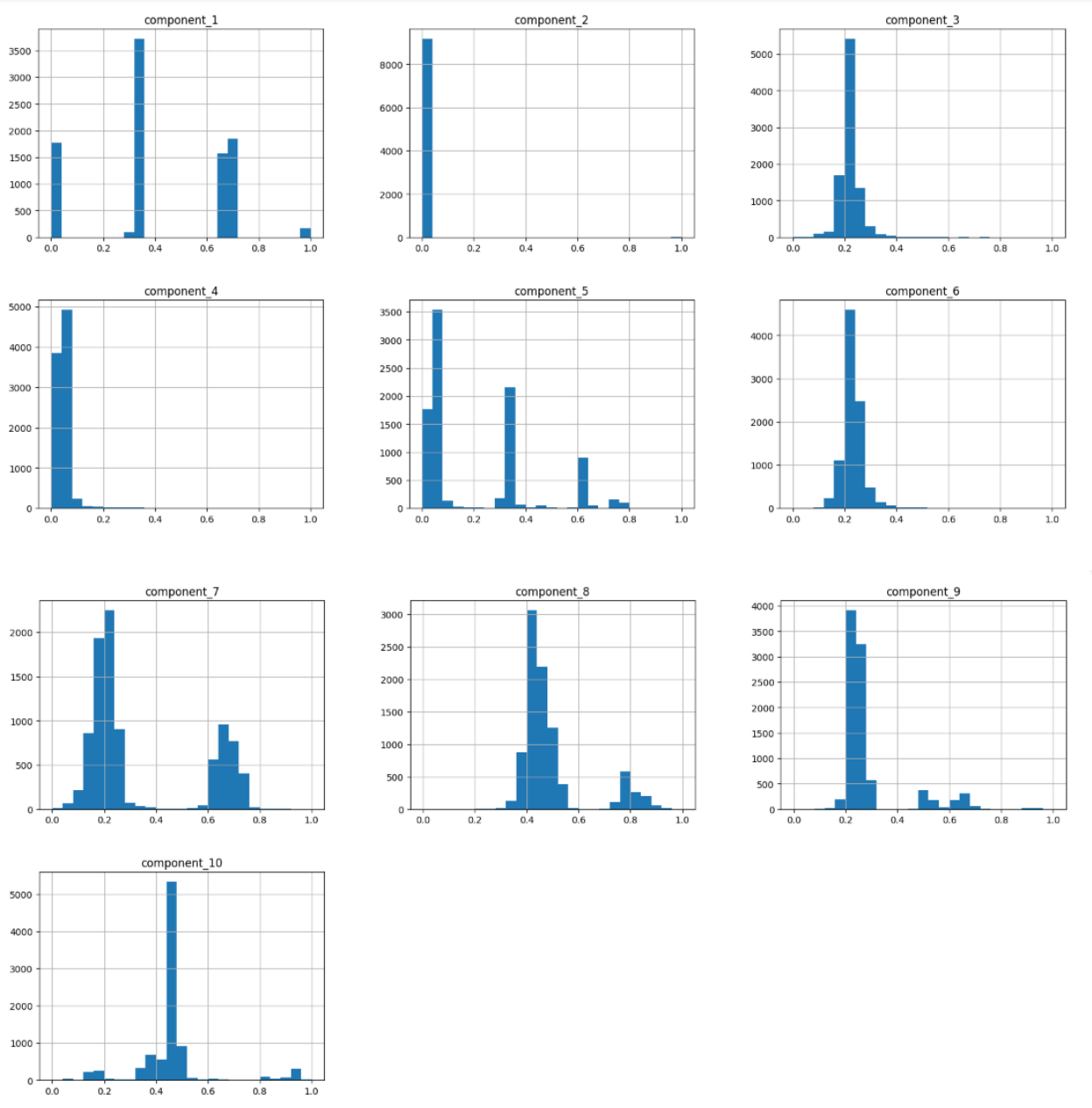
Visual analysis:


age
sex
TSH
T3
TT4
T4U
FTI
patient_id
1 -0.014 -0.004 0.0037 0.0041 -0.002 0.0041 0.019
-0.014 1 0.037 0.049 0.15 0.21 0.03 -0.018
-0.004 0.037 1 -0.14 -0.26 0.073 -0.26 -0.0067
0.0037 0.049 -0.14 1 0.47 0.28 0.31 0.031
0.0041 0.15 -0.26 0.47 1 0.36 0.74 0.023
-0.002 0.21 0.073 0.28 0.36 1 -0.24 -0.038
0.0041 0.03 -0.26 0.31 0.74 -0.24 1 0.022
0.019 -0.018 -0.0067 0.031 0.023 -0.038 0.022 1
1.0
0.8
0.6
0.4
0.2
0.0
−0.2

	age	sex	TSH	T3	TT4	T4U	FTI	patient_id
age	1.000000	-0.014178	-0.003978	0.003693	0.004122	-0.002045	0.004108	0.018506
sex	-0.014178	1.000000	0.036707	0.048889	0.150475	0.214717	0.030148	-0.018141
TSH	-0.003978	0.036707	1.000000	-0.136613	-0.264755	0.072962	-0.257900	-0.006718
T3	0.003693	0.048889	-0.136613	1.000000	0.471792	0.281336	0.312401	0.031492
TT4	0.004122	0.150475	-0.264755	0.471792	1.000000	0.362280	0.738279	0.023354
T4U	-0.002045	0.214717	0.072962	0.281336	0.362280	1.000000	-0.238927	-0.037535
FTI	0.004108	0.030148	-0.257900	0.312401	0.738279	-0.238927	1.000000	0.022164
patient_id	0.018506	-0.018141	-0.006718	0.031492	0.023354	-0.037535	0.022164	1.000000


age
sex
on_thyroxine
query_on_thyroxine
on_antithyroid_meds
sick
pregnant
thyroid_surgery
I131_treatment
query_hypothyroid
query_hyperthyroid
lithium
goitre
tumor
hypopituitary
psych
TSH
T3
T4U
FTI
referral_source
target
patient_id
1e8

	component_1	component_2	component_3	component_4	component_5	component_6	component_7	component_8	component_9	component_10
0	-1.214633e+07	-9.454399	2.468719	-4.748913	-1.097560	0.012287	-0.266622	-0.061928	0.414206	0.835251
1	-1.214633e+07	-9.454499	2.194759	-3.478534	-1.079480	-0.039700	-0.213144	-0.220836	-0.129252	0.012569
2	-1.214630e+07	2.545152	1.431917	0.059278	-1.082378	0.094690	-0.232747	-0.322359	0.698666	-0.519714
3	-1.214430e+07	-2.460643	1.431935	0.058945	-1.065406	0.037189	-0.205197	-0.220305	-0.132886	0.014848
4	-1.214430e+07	-6.460645	1.432483	0.058756	-1.065476	0.037193	-0.205171	-0.220315	-0.132875	0.014839
...	...	...	...	...	...	...	...	...	...	...
9167	1.717168e+07	-67.183880	-37.832040	-7.800883	0.760207	0.169700	0.620459	-0.066421	0.794199	-0.542946
9168	1.717168e+07	-101.181155	-16.321784	-3.169405	0.835236	0.002639	0.640887	0.012621	-0.036148	-0.003599
9169	1.717168e+07	-54.182403	-26.104141	-5.273229	0.811264	0.094784	0.605034	0.048484	-0.033933	-0.015449
9170	1.717168e+07	-76.182503	-27.074623	-5.475111	-1.230923	0.095114	-0.232755	-0.179577	-0.136062	0.007863
9171	1.717169e+07	-92.185377	-49.556822	-10.320205	-1.258967	0.169890	0.673645	0.146192	0.470252	0.817250



	component_1	component_2	component_3	component_4	component_5	component_6	component_7	component_8	component_9	component_10
0	0.000000e+00	0.001582	0.223373	0.038830	0.047337	0.234646	0.214690	0.467087	0.497575	0.943941
1	3.410871e-08	0.001582	0.223089	0.041294	0.049944	0.231583	0.240763	0.404027	0.221711	0.469903
2	9.891539e-07	0.001765	0.222298	0.048155	0.049526	0.239501	0.231206	0.363739	0.641970	0.163195

	component_1	component_2	component_3	component_4	component_5	component_6	component_7	component_8	component_9	component_10
3	6.934301e-05	0.001689	0.222298	0.048154	0.051973	0.236113	0.244638	0.404237	0.219866	0.471215
4	6.937712e-05	0.001628	0.222298	0.048154	0.051963	0.236114	0.244650	0.404234	0.219872	0.471211
...	...	...	...	...	...	...	...	...	...	...
9167	9.999996e-01	0.000702	0.181598	0.032912	0.315193	0.243920	0.647173	0.465304	0.690464	0.149809
9168	9.999996e-01	0.000183	0.203895	0.041894	0.326011	0.234078	0.657132	0.496670	0.268971	0.460586
9169	9.999997e-01	0.000900	0.193755	0.037814	0.322555	0.239506	0.639653	0.510902	0.270096	0.453758
9170	9.999997e-01	0.000564	0.192749	0.037422	0.028108	0.239526	0.231202	0.420400	0.218254	0.467191
9171	1.000000e+00	0.000320	0.169444	0.028026	0.024065	0.243931	0.673103	0.549675	0.526025	0.933568

Model Building:

Random Forest Classifier Model:

```
0.7347208457218368
```

Decision Tree Classifier:

```
0.7978196233894945
```

KNeighbors Classifier:

```
0.804426825239511
```

SVC Model:

```
0.7158903204492897
```

Logistic Regression:

```
0.7403369672943508
```

Performance Testing:

Testing model with multiple &Hyperparameter Tunning:

```
age
[   29    41    36    32    60    77    28    54    42    51    37
16
    43    63    40    75    56    85    71    67    55    61    46
44
    82    64    70    33    59    53    52    49    35    48    27
69
    76    73    68    66    30    88    38    58    21    45    83
62
    25    86    72    14    15    39    26    20    80    90    23
18
    13    78    24    81    92    57    74     9    47    17    11
50
    34     8    79    31    65    84    12    10    19    22     1
2
    97     6    89    87   455    93     7    91     5    94     4
65511
    95 65512     3 65526]



sex
[2. 1.]



on_thyroxine
[0 1]



query_on_thyroxine
[0 1]



on_antithyroid_meds
[0 1]



sick
[0 1]
```

```
pregnant
[0 1]


thyroid_surgery
[0 1]


I131_treatment
[0 1]


query_hypothyroid
[1 0]


query_hyperthyroid
[0 1]


lithium
[0 1]


goitre
[0 1]


tumor
[0 1]


hypopituitary
[0 1]


psych
[0 1]


TSH
[3.00000000e-01 1.60000000e+00 5.21840275e+00 7.00000000e-01
```

1.20000000e+00 1.90000000e+00 1.00000000e+00 5.00000000e-01
2.60000000e+00 6.80000000e+01 1.50000000e+00 5.90000000e+00
5.00000000e-02 4.00000000e+00 4.00000000e-01 8.00000000e-01
2.00000000e-01 3.00000000e+00 9.59999900e+00 1.40000000e+02
6.00000000e-01 1.70000000e+00 2.50000000e+00 6.80000000e+00
2.10000000e+00 1.10000000e+00 9.79999900e+00 3.70000000e+00
1.00000000e-01 3.50000000e-01 9.00000000e+01 1.30000000e+00
2.70000000e+00 2.90000000e+00 2.50000000e-01 5.80000000e+00
2.00000000e+00 9.00000000e-01 3.20000000e+00 7.40000000e+00
1.80000000e+00 5.00000000e+00 7.00000000e+01 6.50000000e+00
1.15999990e+01 4.20000000e+00 8.80000000e+01 2.20000000e+00
5.30000000e+01 1.65000000e+01 8.40000000e+00 4.10000000e+00
3.30000000e+00 8.00000000e+01 1.76000000e+02 3.00000000e+01
1.10000000e+01 6.00000000e+00 7.10000000e+01 9.09999900e+00
1.07000000e+01 4.90000000e+00 3.40000000e+00 1.83000000e+00
4.60000000e+00 1.40000000e+00 1.40000000e+01 3.50000000e+00
4.60000000e+01 1.80000000e+01 1.70000000e+02 1.91999990e+01
7.50000000e+00 1.45000000e+02 7.90000000e+00 1.90000000e+01
4.80000000e+01 2.50000000e+01 7.30000000e+00 6.70000000e+00
1.53000000e+02 3.50000000e+01 3.80000000e+00 2.30000000e+00
9.20000000e+00 1.20000000e+01 5.50000000e+00 1.52999990e+01
1.50000000e-01 1.02999990e+01 4.30000000e+02 2.13000000e+02
3.70000000e+01 2.16000000e+02 2.60000000e+01 4.30000000e+00
2.40000000e+00 4.70000000e+01 8.20000000e+00 1.44000000e+01
1.74000000e+01 1.00000000e+02 1.32000000e+01 4.10000000e+01
3.10000000e+00 2.20000000e+01 4.50000000e+01 4.50000000e+00
2.40000000e+01 6.50000000e+01 4.40000000e+00 5.60000000e+01
1.14000000e+01 5.20000000e+00 5.60000000e+00 8.29999900e+00
4.00000000e+02 2.60000000e+02 5.40000000e+00 8.00000000e+00
6.60000000e+00 8.50000000e+00 5.10000000e+00 1.38000000e+02
4.20000000e+01 4.40000000e+01 2.80000000e+00 1.00000000e+01
5.40000000e+01 3.90000000e+00 1.50000000e+02 3.10000000e+01
6.50000000e-01 2.00000000e+02 8.50000000e+01 8.60000000e+01
6.30000000e+00 5.00000000e+01 3.40000000e+01 1.60000000e+01
9.40000000e+00 1.42999990e+01 4.90000000e+01 6.20000000e+00
8.20000000e+01 7.70000000e+01 9.20000000e+01 1.30000000e+01
4.70000000e+00 5.20000000e+01 2.70000000e+01 8.79999900e+00
1.05000000e+00 8.09999900e+00 1.32999990e+01 2.80000000e+01
1.25000000e+02 1.50000000e+01 1.01000000e+00 3.90000000e+01
2.88000000e+02 2.10000000e+01 3.60000000e+00 4.50000000e-01
1.43000000e+02 2.35000000e+02 5.50000000e+01 7.60000000e+00
3.80000000e+01 3.20000000e+01 6.10000000e+01 4.80000000e+00
1.47999990e+01 7.00000000e+00 2.00000000e+01 9.00000000e+00
9.80000000e+01 2.90000000e+01 6.40000000e+00 1.70000000e+01
2.30000000e+01 1.84000000e+01 3.60000000e+01 1.09000000e+02
8.59999900e+00 6.10000000e+00 1.26000000e+02 4.30000000e+01
9.70000000e+00 4.00000000e-02 1.20000000e-01 6.90000000e+00
1.60000000e+02 1.60000000e-01 5.30000000e-01 8.80000000e-01
2.60000000e-01 2.50000000e-02 3.90000000e-01 7.70000000e-01
7.00000000e-02 4.00000000e+01 2.30000000e-01 8.40000000e-01
9.50000000e-01 3.80000000e-01 4.10000000e-01 6.50000000e-02
5.00000000e-03 7.80000000e-01 8.10000000e-01 1.83000000e+02
6.10000000e-01 3.20000000e-01 8.20000000e-01 8.90000000e+00
8.70000000e-01 9.80000000e-01 4.70000000e-01 5.80000000e-01

```
 4.30000000e-01 8.00000000e-02 8.60000000e-01 8.90000000e-01
 9.10000000e-01 5.30000000e+02 1.00000000e-02 5.20000000e-01
 8.30000000e-01 1.50000000e-02 3.00000000e-02 5.90000000e-01
 6.20000000e-01 7.40000000e-01 4.60000000e-01 1.78000000e+02
 2.00000000e-02 9.00000000e-02 1.65000000e+02 6.80000000e-01
 8.50000000e-01 3.50000000e-02 1.70000000e-01 7.10000000e+00
 7.80000000e+00 9.90000000e+00 4.20000000e-01 2.80000000e-01
 3.70000000e-01 4.40000000e+02 2.70000000e-01 6.30000000e-01
 7.10000000e-01 5.70000000e-01 9.20000000e-01 7.20000000e-01
 4.90000000e-01 9.40000000e-01 9.70000000e-01 5.60000000e-01
 5.40000000e-01 7.90000000e-01 1.03000000e+02 7.50000000e-01
 5.73000000e+00 6.70000000e-01 7.20000000e+00 1.10999990e+01
 9.30000000e-01 5.10000000e-01 9.50000000e+00 7.60000000e-01
 7.70000000e+00 6.00000000e-02 6.40000000e-01 5.70000000e+00
 7.30000000e-01 1.99000000e+02 9.90000000e+01 5.80000000e+01
 2.20000000e-01 6.60000000e+01 9.60000000e-01 1.90000000e-01
 2.40000000e-01 5.50000000e-01 1.80000000e-01 2.10000000e-01
 4.80000000e-01 1.02000000e+00 4.40000000e-01 9.90000000e-01
 3.40000000e-01 3.30000000e-01 1.17000000e+02 6.90000000e-01
 8.90000000e+01 7.60000000e+01 1.51000000e+02 5.10000000e+01
 1.39000000e+02 1.40000000e-01 2.90000000e-01 7.80000000e+01
 4.72000000e+02 5.30000000e+00 2.30000000e+02 3.10000000e-01
 1.20999990e+01 4.50000000e-02 5.50000000e-02 6.00000000e+01
 1.30000000e-01 1.08000000e+02 1.88000000e+02 3.05000000e+01
 4.68000000e+02 3.60000000e-01 2.64000000e+01 6.60000000e-01
 2.36000000e+02 3.30000000e+01 4.78000000e+02 9.29999900e+00
 1.16000000e+02 2.55000000e-01 1.10000000e-01 1.14000000e+02
 1.03000000e+00 1.97000000e+02 9.60000000e+01 1.15000000e+01
 1.31000000e+02 6.70000000e+01 1.06000000e+02 1.72000000e+02
 1.64000000e+02 7.50000000e-02 1.41000000e+02 1.04000000e+00
 1.35000000e-01 8.50000000e-02 1.25000000e-01 1.15000000e-01
 7.30000000e+01 4.36000000e+02 9.50000000e-02 1.19000000e+01
 4.94000000e+02 6.20000000e+01 3.93000000e+02 8.70000000e+00
 1.05000000e-01 2.52000000e+02 5.90000000e+01 1.18000000e+02
 1.14000000e+00 1.12000000e+00 1.91000000e+02 1.66000000e+02
 4.60000000e+02 1.92000000e+02 1.36000000e+02 1.77000000e+02
 1.98000000e+02 1.37000000e+01 1.67999990e+01 7.40000000e+01
 5.00000000e+02 4.01000000e+00]
```

```
T3
[ 1.97062881  1.9         2.6         1.8         1.7         2.3
  2.4         2.9         2.          2.1         1.6         0.1
  1.4         1.2         1.5         1.3         2.5         2.7
  2.2         2.8         3.2         0.4         0.8         1.
  1.1         3.7         4.4         3.          3.1         3.6
  7.6         0.9         4.2         0.5         0.6         0.3
  0.7         3.8         0.2         4.1         6.6         4.7
  8.599999    3.3         4.3         0.05        3.4         4.6
  4.9         6.2         3.5         3.9         8.9         4.5
  8.099999    5.          4.8         5.1         5.3         6.7
  7.3         6.1         4.          5.5         5.4         5.7
  7.          6.          7.1         8.5        10.599999    1.44
```

```
  5.2         6.9          6.8         5.6         6.4         18.
  9.5         6.5         13.299999    0.83        0.69         0.93
  5.9         8.        ]
```

```
T4U
[0.97605572 1.02       1.06       0.94       1.08       0.84
 1.13       1.07       0.87       0.89       0.62       0.91
 0.68       1.         1.38       0.79       0.95       1.57
 0.92       1.48       1.1        0.7        1.01       1.05
 0.96       0.78       1.4        0.66       0.86       0.76
 0.9        1.16       1.12       0.98       1.04       1.26
 0.83       0.97       0.93       0.88       0.73       1.29
 1.3        0.75       0.8        1.83       1.03       0.61
 1.44       1.18       0.59       0.81       0.64       1.2
 0.82       1.19       0.99       1.56       1.22       0.71
 1.32       0.67       0.32       1.11       0.85       0.52
 1.15       1.21       0.77       0.69       1.51       1.33
 0.55       1.45       1.24       1.79       0.72       1.73
 1.27       1.68       1.09       1.43       0.35       0.3
 1.28       0.2        1.41       1.14       0.53       1.52
 1.23       0.74       1.53       1.62       1.66       0.4
 1.86       1.59       0.29       0.34       1.17       1.76
 0.57       0.63       1.71       0.31       0.49       1.31
 1.34       0.5        1.75       1.36       0.36       1.42
 0.6        1.74       1.46       1.63       0.28       1.47
 1.25       1.96       1.39       0.48       0.56       1.55
 0.65       1.69       1.64       1.65       1.67       1.77
 1.82       0.19       1.49       1.35       1.5        1.97
 1.94       1.58       0.58       1.61       0.54       1.7
 0.38       0.41       0.25       2.03       1.93       0.944
 1.8        2.12       2.01       1.54       1.88       0.46
 1.84       1.37       0.47       2.32       1.78       1.89
 0.44       0.17       1.6        1.87       0.37       0.45
 1.95       0.51       2.33       0.42       1.81       2.15
 2.02       2.16       2.        ]
```

```
FTI
[113.64074552  47.          85.          84.          96.
 105.          95.         106.         176.         129.
 100.          69.          39.          91.          90.
  93.          66.         121.          92.         173.
 117.          31.         113.          67.         101.
 126.         123.         149.          68.          86.
 132.         131.         116.          97.         124.
 136.         142.         104.           7.5        107.
  73.         110.         130.          88.         128.
 122.         102.         134.         163.          63.
 354.          81.         109.         114.         133.
 170.          99.         111.         108.         161.
  78.         148.          98.         135.          80.
```

```
127.        213.        119.         65.         89.
143.        316.        155.        172.        150.
103.        120.        258.          5.        272.
263.        166.        138.         52.        164.
337.         94.        118.        182.         41.
 70.        144.         10.          4.         13.
 87.        140.         74.        152.         77.
  3.         82.        145.         64.         79.
147.         54.         83.        634.        650.
 12.         61.         11.        115.         35.
 17.        165.        167.        153.         44.
  3.4        55.         71.        253.         75.
  2.5       197.         24.        156.        237.
203.        112.        141.          3.5       190.
 37.         45.        193.         57.         76.
160.          6.        200.        485.         49.
158.        137.        428.        450.        174.
189.        202.        159.        196.        154.
139.         34.        222.        184.        178.
146.        125.         21.        157.         51.
839.        332.        151.        305.        299.
266.         32.         53.        370.         22.
168.         60.        187.        171.        220.
169.        232.        254.        345.        194.
211.        217.        550.         23.        257.
188.        192.        179.        218.        208.
  6.6       240.         43.        259.        265.
  9.        212.        347.        216.         29.
186.        162.        256.        177.         20.
 33.        308.        482.         16.        180.
 59.        264.        881.        204.         15.
  1.4       271.         58.        334.        214.
228.        205.         72.        183.        612.
445.        244.        175.          2.8       283.
 62.          7.6         5.4       199.        209.
227.         26.         48.        221.         28.
195.        395.        206.         46.        198.
215.          8.5        56.         14.        223.
  8.9         8.4        18.          9.099999   50.
207.        185.        291.        235.         36.
224.         19.          7.        312.        247.
 40.        274.        210.        242.        251.
 27.        181.          2.        249.         42.
349.        191.        280.        281.        201.
245.        362.        219.        239.        273.
356.          2.4         8.299999  236.          8.7
262.          3.84      231.        369.          3.1
378.        546.        268.         25.        298.
241.        642.        238.        233.        325.
255.          4.5       519.          4.15      297.
  4.85      321.        226.        290.        243.
  5.5       288.         38.        329.       ]
```

```
referral_source
[5 3 1 0 2 4]

patient_id
[840801013 840801014 840801042 ... 870119025 870119027 870119035]
```

Comparing model accuracy before & after applying hyperparameter tuning:

```
0.8460521968946151


0.7330690452593327


0.6362735381565907


0.7330690452593327
```

Integrate with web FrameWork:

Building HML pages:

```
<html>

<head>

<center>

<font size="10">

<h1>

<u>Thyroid Disease Classification</u></h1>

<style>

body

{
```

```
background-image:url("123.JPG");

background-repeat:no repeat;

}

h1

{

color:white;

}

form

{

color:white;

}

</style>


</font>

</center>

</head>


<form action="">

Age           &nb
sp;           &nbs
p;         :<input
type="text"><br><br>

Sex           &nb
sp;           &nbs
p;            
;:<input type="radio"value="f">Female


<input type="radio" value="m">Male<br><br>
```

On_thyroxine   &
nbsp;   :<input type="text">

query_on_thyroxine   :<input
type="text">

On_antithyroid_meds   :<input type="text">

Sick   &nb
sp;   &nbs
p;   :<input
type="text">

Pregnant

 :<input type="text">

Thyroid_surgery
; :<input type="text">

I131_treatment
   :<input type="text">

Query_hypothyroid   :<input
type="text">

Query_hyperthyroid   :<input
type="text">

Lithium

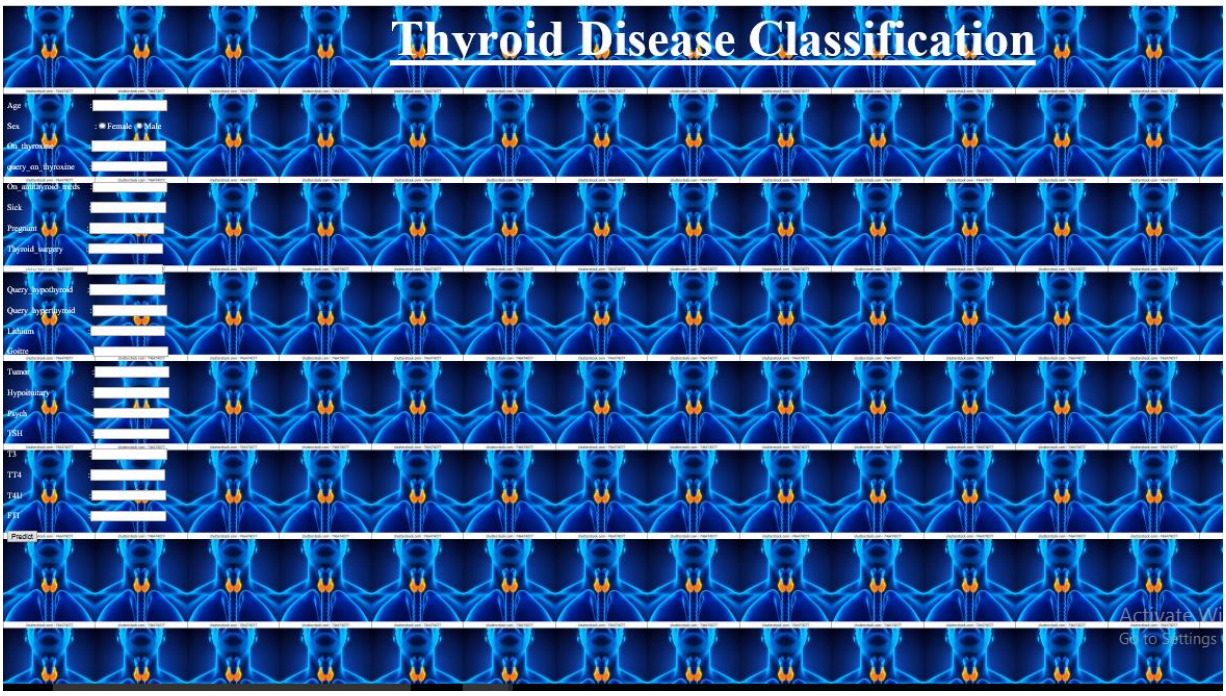
   :<input
type   :"text">

Goitre   &
nbsp;   &n
bsp;   :<input type:"text">

Tumor   &
nbsp;   &n
bsp;   :<input type:"text">

Hypoituitary   &n
bsp;   :<input
type="text">


```
Psych           &
nbsp;          &n
bsp;       :<input type="text"><br><br>

TSH           &n
bsp;          &nb
sp;         :<input
type="text"><br><br>

T3          &nbs
p;            
;           :<input
type="text"><br><br>

TT4          &nb
sp;          &nbs
p;        :<input
type="text"><br><br>

T4U          &n
bsp;          &nb
sp;        :<input
type="text"><br><br>

FTI          &nb
sp;          &nbs
p;         :<input
type="text"><br><br>

<button type="submit" class="btn btn-primary btn-block btn-large">Predict</button>

</font>

</form>



</html>
```


Thyroid Disease Classification
Age
Sex
Female
Male
on_thyroxine
query_on_thyroxine
on_antithyroid_meds
Sick
Pregnant
Thyroid_surgery
Query_hypothyroid
Query_hyperthyroid
Lithium
Goitre
Tumor
Hypopituitary
Psych
TSH
T3
TT4
T4U
FTI
Predict

4. ADVANTAGES & DISADVANTAGES:

Advantages:

- It helps to regulate many body functions by constantly releasing a steady amount of thyroid hormones into the bloodstream.
- Controls how much energy your body uses.
- It is used throughout your entire body to keep many of your body's systems working correctly.

Disadvantages:

- Thyroid disorders can cause puberty and menstruation to occur abnormally early or late.
- In addition, abnormally high or low levels of thyroid hormone can cause very light or very heavy menstrual periods, very irregular menstrual periods, or absent menstrual periods (a condition called amenorrhea).
- This can be dangerous because it can cause your heart to beat faster and weight loss without trying.

5. APPLICATIONS:

- Assume a machine learning model can detect thyroid disease in a patient. The thyroid disease can then be easily identified based on the symptoms in the patient's history.
- Currently, models are evaluated using accuracy metrics on a validation dataset that is accessible.
- The thyroid gland is a vital hormone gland: It plays a major role in the metabolism, growth and development of the human body. It helps to regulate many body functions by constantly releasing a steady amount of thyroid hormones into the bloodstream.

6. CONCLUSION:

- The combination of the Recursive Feature Elimination an the Support Vector Machine Technique has proven to be effective.
- The feature set finally used is ‘Age’, ‘Sex’, ‘TSH’, ‘TT4’, ‘T4U’, ‘T3’, ‘FTI’. Age and Sex have been considered as important features because Thyroid disorders are said to occur during a particular age and most particularly in females.
- 2 classes of Thyroid disorders namely Hyperthyroid,Hypothyroidcan be classified.

7. FUTURE SCOPE:

- It helps the doctors to easily diagnosis the patients.
- Easy to predict whether the patient is having hypothyroid or hyperthyroid.
- It helps the Health care professional to easily predict with less expensive.

8. APPENDIX:

SOURCE CODE:

Importing the libraries:

```
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
import seaborn as sns

import pickle
```

Read the Dataset:

```
file = open("/content/thyroidDF (1) (1).csv")
df = pd.read_csv(file)

df
```

Checking for null values:

```
feature_cols = ["age",
                "sex",
                "on_thyroxine",
                "query_on_thyroxine",
                "on_antithyroid_medication",
                "sick",
                "pregnant",
                "thyroid_surgery",
                "I131_treatment",
                "query_hypothyroid",
                "query_hyperthyroid",
                "lithium",
                "goitre",
                "tumor",
                "hypopituitary",
                "psych",
                "TSH measured",
                "TSH",
                "T3_measured",
                "T3",
                "TT4_measured",
                "TT4",
                "T4U_measured",
                "T4U",
                "FTI_measured",
                "FTI",
                "TBG_measured",
                "TBG",
               "target"]

df

   target = df.target
   create = target.str.split('([A-Za-z]+)', expand=True)
   create = create[1]
   target = create.replace({None:'Z'}) #here z is none type
   df.target = target

   df.target.unique()
```

```
array(['Z', 'S', 'F', 'AK', 'R', 'I', 'M', 'N', 'G', 'K', 'A',
'KJ', 'L', 'MK', 'Q', 'J', 'C', 'O', 'LJ', 'H', 'D', 'GK',
'MI', 'P', 'FK', 'B', 'GI', 'GKJ', 'OI', 'E'], dtype=object)
```

```
df
```

```
df = df.replace(['?'],np.nan)
```

```
df.isnull().sum()
```

```
df.drop(['TBG_measured','TBG','T3_measured','TSH_measured','TT
4_measured','T4U_measured','FTI_measured'],axis=1,inplace=True
)
```

```
df
```

```
df.isnull().sum()
```

```
df.sex.replace({'F':2,'M':1},inplace=True)
```

```
round_Values = round(df.sex.mean())
df.sex.fillna(round_Values,inplace=True)
```

```
df.sex.unique()
```

```
array([2., 1.])
```

```
df.isnull().sum()
```

```
from sklearn.impute import KNNImputer
knnimp = KNNImputer(n_neighbors=3)

cols = ['TSH','T3','TT4','T4U','FTI']
for i in cols:
    df[i] = knnimp.fit_transform(df[[i]])
```

```
df.isnull().sum()

df.info()
```

Exploratory Data Analysis:

Descriptive Analysis:

```
df.info()
```

Visual Analysis:

```
plt.figure(figsize=(20,20))
sns.heatmap(df.corr(),annot=True)
```

```
<Axes: >
```

```
df.corr()
```

```
from sklearn.preprocessing import LabelEncoder
le = LabelEncoder()

cols = df.select_dtypes(include=['object'])

for i in cols.columns:
    try:
        df[i] = le.fit_transform(df[i])
    except:
        continue
```

```
for a in range(len(df.corr())):
    for b in range(a):
        if((df.corr().iloc[a,b]) >= 0.7):
            print(df.corr().columns[b])
```

```
TT4
```

```
df.drop('TT4',axis=1,inplace=True)
```

```
df.hist(bins=25,figsize=(20,20));
```

```
X = df.drop('target',axis=1)
y = df.target
df2 = X
```

```
y.unique()
```

```
array([29, 28, 6, 1, 27, 13, 19, 22, 8, 15, 0, 16, 17, 21, 26, 14,
3, 23, 18, 12, 4, 10, 20, 25, 7, 2, 9, 11, 24, 5])
```

```
from sklearn.decomposition import PCA
pca = PCA(n_components=10)
```

```
v = pca.fit_transform(X)
```

```
X_pca = pd.DataFrame(data = v, columns = ['component_1', 'component_
2', 'component_3', 'component_4', 'component_5', 'component_6', 'com
ponent_7','component_8','component_9','component_10'])
```

```
X_pca
```

```
from sklearn.preprocessing import MinMaxScaler
scaler = MinMaxScaler()
```

```
for i in X_pca.columns:
    X_pca[i] = scaler.fit_transform(X_pca[[i]])
```

```
X_pca.hist(bins=25,figsize=(20,20));
```

Model Building:

Training the model in multiple algorithms:

```
from sklearn.model_selection import train_test_split
X_train,X_test,y_train,y_test = train_test_split(X_pca,y,test_size=0
.33,random_state=42)
```

Random Forest Classifier Model:

```
from sklearn.ensemble import RandomForestClassifier
rf = RandomForestClassifier(max_depth=2,n_estimators=200)
rclf = rf.fit(X_train,y_train)
rfpred = rclf.predict(X_test)
accuracy_score(rfpred,y_test)
```

Decision Tree Classifier:

```
from sklearn.metrics import accuracy_score

from sklearn.tree import DecisionTreeClassifier
tree = DecisionTreeClassifier(max_depth=3)
clf = tree.fit(X_train,y_train)
treepredict = clf.predict(X_test)

accuracy_score(treepredict,y_test)
```

KNeighbors Classifier:

```
from sklearn.neighbors import KNeighborsClassifier
neigh = KNeighborsClassifier(n_neighbors=3)
knnclf = neigh.fit(X_train,y_train)
y_pred = knnclf.predict(X_test)

accuracy_score(y_pred,y_test)
```

SVC Model:

```
from sklearn.svm import SVC
svm = SVC(kernel="sigmoid")
sclf = svm.fit(X_train,y_train)
y_pred = sclf.predict(X_test)
accuracy_score(y_pred,y_test)
```

Logistic Regression:

```
from sklearn.linear_model import LogisticRegression
lr = LogisticRegression(max_iter=1000)
lrclf = lr.fit(X_train,y_train)
y_pred = lrclf.predict(X_test)
accuracy_score(y_pred,y_test)
```

Performance Testing:

Testing model with multiple &Hyperparameter Tunning:

```
for i in df2.columns:
    print("\n\n")
    print(i)
    print(df2[i].unique())
```

```
referral_source
[5 3 1 0 2 4]

patient_id
[840801013 840801014 840801042 ... 870119025 870119027 870119035]
```

Comparing model accuracy before & after applying hyperparameter tuning:

```
cols = ['age','sex','TSH','T3','T4U','FTI']
for i in cols:
    df2[i] = scaler.fit_transform(df2[[i]])
```

```
X = df2
y = df['target']
```

```
from sklearn.model_selection import train_test_split
X_train,X_test,y_train,y_test = train_test_split(X,y,test_size=0.33,
random_state=42)


from sklearn.metrics import accuracy_score


from sklearn.tree import DecisionTreeClassifier
tree = DecisionTreeClassifier(max_depth=3)
clf = tree.fit(X_train,y_train)
y_pred = clf.predict(X_test)
accuracy_score(y_pred,y_test)




from sklearn.ensemble import RandomForestClassifier
rf = RandomForestClassifier(max_depth=2,n_estimators=200)
rclf = rf.fit(X_train,y_train)
y_pred = rclf.predict(X_test)
accuracy_score(y_pred,y_test)


from sklearn.neighbors import KNeighborsClassifier
neigh = KNeighborsClassifier(n_neighbors=3)
knnclf = neigh.fit(X_train,y_train)
y_pred = knnclf.predict(X_test)
accuracy_score(y_pred,y_test)


from sklearn.svm import SVC
svm = SVC(kernel="sigmoid")
sclf = svm.fit(X_train,y_train)
y_pred = sclf.predict(X_test)
accuracy_score(y_pred,y_test)
```