THIRUVALLUVAR UNIVERSITY
PERIYAR ARTS COLLEGE
CUDDALORE – 607001.

DEPARTMENT OF COMPUTER APPLICATIONS

MACHINE LEARNING WITH PYTHON

Project Title   : Thyroid disease classification using Machine Learning

Team Id   : NM2023TMID35506

Team Leader : ELAVARASAN T(tvu20CA18)

Team member: ABINESH J(tvu20CA01)

Team member: ARUN J(tvu20CA07)

Team member: BARATHAN P(tvu20CA12)

1.Introduction:

Thyroid disease is a subset of endocrinology which is one of the most misunderstood and undiagnosed diseases [1] [2]. Thyroid gland diseases are among the most prevalent endocrine disorders in the world, second only to diabetes, according to the World Health Organization. Hyper function hyperthyroidism and hypothyroidism affect about 2% and 1% of individuals, respectively. Men have about a tenth of the prevalence of women. Hyper-and hypothyroidism may be caused by thyroid gland dysfunction, secondary to pituitary gland failure, or tertiary to hypothalamic malfunction. Due to dietary iodine deficiency, goiter or active thyroid nodules may become prevalent in some regions, with a prevalence of up to 15%. The thyroid gland can also be the location of different kinds of tumors and can be a dangerous place where endogenous antibodies wreak havoc (autoantibodies) [3]. Early disease detection, diagnosis, and care, according to doctors, are vital in preventing disease progression and even death. For several different forms of anomalies, early identification and differential diagnosis raises the odds of good treatment. Despite multiple trials, clinical diagnosis is often thought to be a difficult task [4].

The thyroid gland is a butterfly-shaped gland situated at the base of the throat. It comprises two active thyroid hormones, levothyroxine (T4) and triiodothyronine (T3), which are involved in brain functions such as body temperature control, blood pressure management, and heart rate regulation. Likewise, thyroid disease is one of the most prevalent diseases worldwide, and it is mostly caused by a deficiency of iodine, but it may also be caused by other factors. The thyroid gland is an endocrine gland that secretes hormones and passes them through the bloodstream. It is situated in the middle of the front of the body. Thyroid gland hormones are responsible for aiding in digestion as well as maintaining the body moist, balanced, and so on. Thyroid gland treatments such as T3 (triiodothyronine), T4 (thyroid hormone), and TSH (thyroid stimulating hormone) are used to assess thyroid activity (thyroid stimulating hormone). Thyroid disorder is classified into two types: hypothyroidism and hyperthyroidism. Data mining [5] is a semi-automated method of looking for correlations in massive datasets. Machine learning algorithms are one of the best solutions to many problems that are difficult to solve [6] Classification is a data extraction technique (machine learning) used to predict and identify many diseases, such as thyroid disease, which we researched and classified here because machine learning algorithms play a significant role in classifying thyroid disease and because these algorithms are high performing and efficient and aid in classification [7]. Although the application of computer learning and artificial intelligence in medicine dates back to the early days of the field [8], there has been a new movement to consider the need for machine learning-driven healthcare solutions. As a result, analysts predict that machine learning will become commonplace in healthcare in the near future [9].

Hyperthyroidism is a disorder in which the thyroid gland releases so many thyroid hormones. Hyperthyroidism is caused by an increase in thyroid hormone levels [10]. Dry skin, elevated temperature sensitivity, hair thinning, weight loss, increased heart rate, high blood pressure, heavy sweating, neck enlargement, nervousness, menstrual cycles shortening, irregular stomach movements, and hands shaking are some of the signs [11]. Hypothyroidism is a condition in which the thyroid gland is underactive Hypothyroidism is caused by a decline in

thyroid hormone production. Hypo means deficient or less in medical terms. Inflammation and thyroid gland injury are the two primary causes of hypothyroidism. Obesity, low heart rate, increased temperature sensitivity, neck swelling, dry skin, hand numbness, hair issues, heavy menstrual cycles, and intestinal problems are some of the symptoms. If not treated, these symptoms can escalate over time [12].

2.Literature Review:

Chandel, Khushboo [13] Thyroid disorder is classified using different classification models based on parameters such as TSH, T4U, and goiter in this study. Several grouping methods, such as K-nearest neighbor, are used to justify this argument. The Naive Bayes and support vector machines algorithms are employed. The experiment was carried out using the Rapid miner instrument, and the findings indicate that K-nearest neighbor is more effective than Naive Bayes in detecting thyroid disease. To diagnose thyroid disorder, the researchers used data mining classifiers. Thyroid disorder is a vital factor to consider when diagnosing a disease. KNN and Naive Bayes classifiers were used in this study. The Rapidminer tool is used to compare these two classifiers. The findings revealed that the K-nearest neighbor classifier is the most reliable, with a 93.44 percent accuracy, while the Naive Bayes classifier has a 22.56 percent accuracy. The proposed KNN technique improves classification accuracy, which contributes to improved results. As a result, Naive Bayes can only have a linear, elliptic, or parabolic decision boundary, so the decision boundary consistency of KNN is a huge plus. KNN outperforms most methods since the factors are interdependent.Banu, G. Rasitha [14] Thyroid disease is one of the most common illnesses that humans suffer from. The hypothyroid data used in this study came from the data repository at the University of California, Irvine (UCI). The platform Waikato Environment of Information Analysis will be used for the whole research project (WEKA). The J48 technique was found to be more effective than the decision stump tree technique. In the world of health care, disease diagnosis is a difficult challenge. In the decision-making method, a number of data mining methods are used.

In this analysis, we used dimensionality reduction to pick a subset of attributes from the original results, and we used J48 and decision stump data mining classification techniques to define hypothyroidism. The uncertainty matrix is used to assess classifier output in terms of precision and error rate. The J48 Algorithm has 99.58 percent accuracy, which is higher than decision stump tree accuracy, and it also has a smaller error rate than Decision stump.

3.Methodology:

Data Collection:

   Machine learning algorithms are used in the rapid and early diagnosis of thyroid diseases and other diseases, as they now in a significant position in the health field and help us in diagnosing and classifying diseases for this reason we were able to collect a good amount of data on thyroid diseases and we are working in our study on the classification of diseases using this data The data that I used in our study is a set of data taken from external hospitals and laboratories specialized in analyzing and diagnosing diseases, and the sample taken from the data is the data of the Iraqi people and the type of data taken related to thyroid disease, where data were taken on 1250 people between males and females, and their ages range from 1 year to 1 year. 90 years as these samples contain people with thyroid disease who suffer from hyperthyroidism and hypothyroidism and normal people who do not suffer from thyroid disease. The data were collected over a period of one to four months, and the main goal of collecting the data was to classify thyroid diseases using machine learning algorithms.

These data include gender, age, analysis of T3 (triiodothyronine), T4 (thyroid hormone), TSH (thyroid stimulating hormone), and a host of other characteristics. As the data obtained consist of 17 variables or attributes where all the attributes were taken in our study which consist of (id,age,gender,querythyroxine,on_antithyroid_medication,sick,pregnant,thyroid_surgery,query_hypothyorid,query_hyperthyroid, TSH_M, TSH, T3_M, T3, T3, T4, Category).

```
                    ┌─────────────────────────┐
                    │        Dataset          │
                    └─────────────────────────┘
                                 │
                                 ▼
                    ┌─────────────────────────┐
                    │   Data Pre-Processing    │
                    └─────────────────────────┘
                                 │
                                 ▼
          ┌──────────────────────────────────────────┐
          │     Divided the dataset to two dataset    │
          └──────────────────────────────────────────┘
           │                                        │
           ▼                                        ▼
   ┌──────────────┐        ┌──────────────────────────────────────────┐
   │ All features │        │ Without query_thyroxine&query_hypothyorid& │
   └──────────────┘        │           query_hyperthyroid              │
           │               └──────────────────────────────────────────┘
           ▼                                        ▼
   ┌─────────────────────────────────────────────────────────────────┐
   │              Machine Learning Algorithms                          │
   │ ┌────┐ ┌─────┐ ┌────┐ ┌────┐ ┌────┐ ┌─────┐ ┌─────┐ ┌─────┐      │
   │ │ DT │ │ SVM │ │ RF │ │ NB │ │ LR │ │ LAD │ │ KNN │ │ MLP │      │
   │ └────┘ └─────┘ └────┘ └────┘ └────┘ └─────┘ └─────┘ └─────┘      │
   └─────────────────────────────────────────────────────────────────┘
                                 │
                                 ▼
   ┌─────────────────────────────────────────────────────────────────┐
   │               Classification Accuracy Result                      │
   └─────────────────────────────────────────────────────────────────┘
```
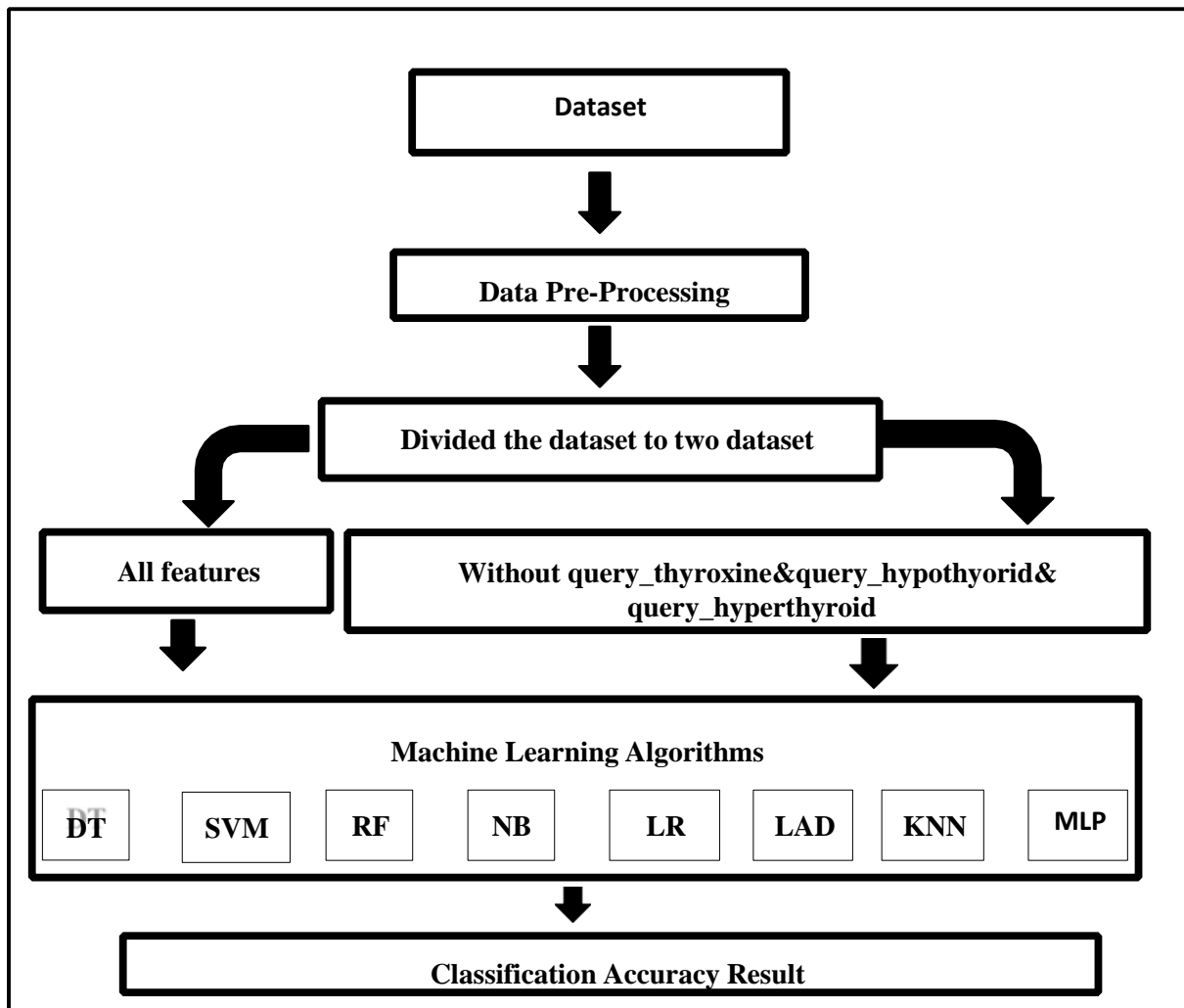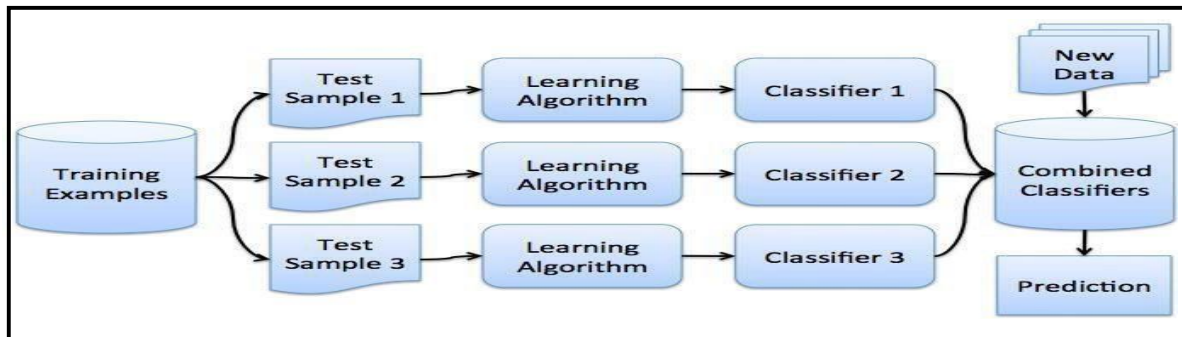
Data Preprocessing:

The process of pre-processing the data is very important and it is a major step in data mining, as it has a good effect on the data, as the pre-processing process is used to reveal the data through analyzing the data and discovering the lost data, as it examines the data with great care. The pre-processing process includes cleaning the data, preparing the data, etc. In this stage or step we did is to clean and arrange the data that we were able to obtain, where we identified a set of missing data in this data where the missing features were identified, and among these properties that were missing T4 by number 151 and T3 by number 112, where we were able to Processing this lost data by replacing it with the value of the mediator, and after working in this way we were able to obtain the data in a good and better way and free from lost data, as the data became arranged and good and free from any defect or problem so that we can work on it smoothly and well. We also used normalization technical with the MLP algorithm.

Data Machine Learning Techniques:

The key aim of using machine learning algorithms is to differentiate between three forms of thyroid disease. The first is hyperthyroidism, the second is hypothyroidism, and the third is stable patients who do not have any thyroid issues. Random forest:

The random forest computes the mean response of every predictor for energy consumption. Then, for each sample, a random forest adds the absolute distance each response was from the mean of each predictor for a total sum of the distance that each answer was from the means of the data. A high distance value will signify individuals who were consistently far away from the mean response in each sample. Detecting rates who repeatedly classify the samples was simple--a function that calculated the mode of each response was used. If the mode of a response was over 90% of the total number of questions, the research marked the response as potentially high in energy consumption. There are many responses marked. It was clear from a visual examination of these responses that the individuals had sampled with the same response.

4. Theortical Analysis:

3.1 Block Diagram:
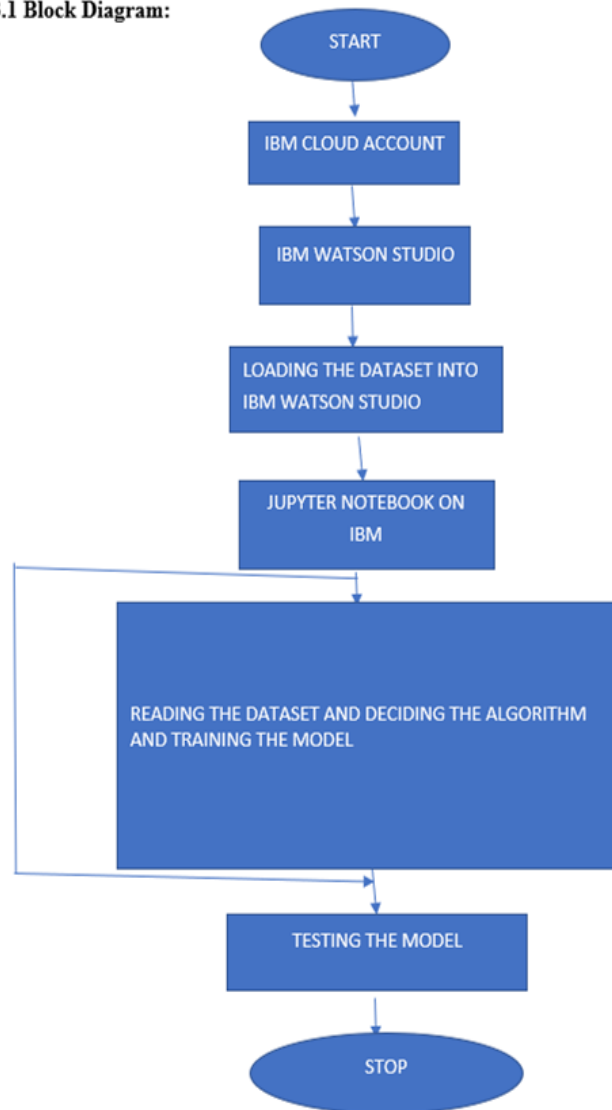
Project Work Flow:

1. Data Collection

2. Data Pre-processing

3. Model Building

4. Application Building

HARDWARE AND SOFTWARE REQUIREMENTS IN THE PROJECT:

For running a machine learning model on the system you need a system with minimum of 16 GB RAM in it and you require a good processor for high performance of the model. In the list of **software requirements** you must have:

- Jupyter Notebook for programming, which can be installed by Anaconda IDE. ● Python packages
- A better software for running the html and css files for application building phase e.g. spyder.

5. **Experimental Investigations:**

Data Preprocessing:-

(i) First check is there any null value in the dataset
(ii) Check whether the dataset contains numerical or categorical features
(iii) Dividing dataset into dependent and independent features
(iv) Plotting Correlation **(v)** Training and testing data

6. **Flow Chart:**
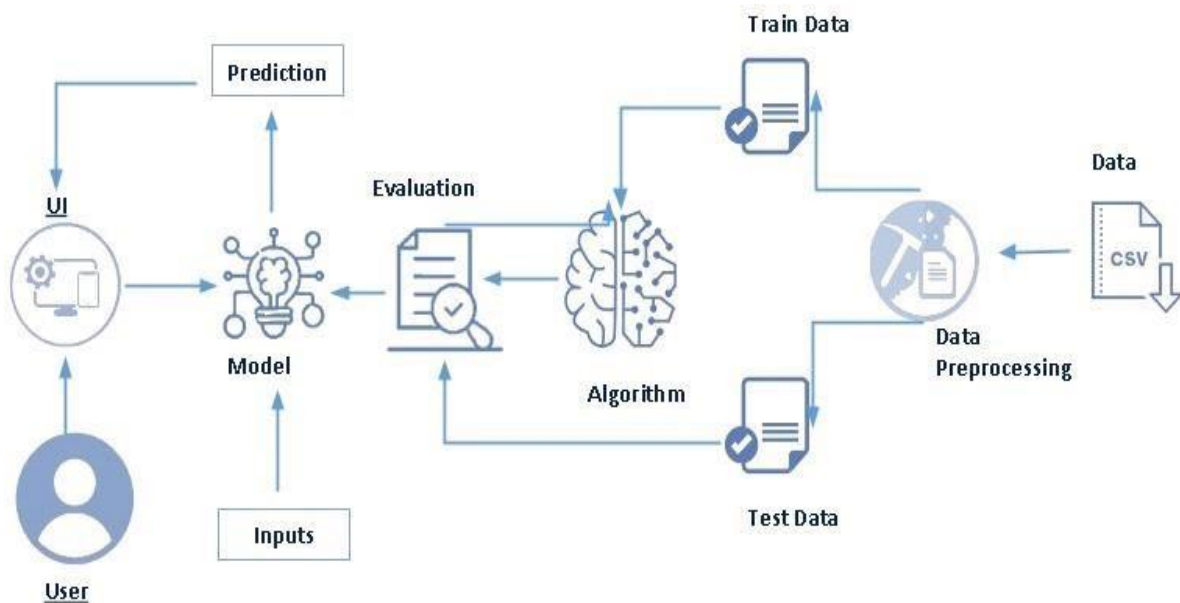
7. **Pre requisites:**

To complete this project, you must required following software's, concepts and packages

- **Anaconda navigator and pycharm:**
- Refer the link below to download anaconda navigator
- Link : https://youtu.be/1ra4zH2G4o0
- **Python packages:**
- Open anaconda prompt as administrator
- Type "pip install numpy" and click enter. • Type "pip install pandas" and click enter.
- Type "pip install scikit-learn" and click enter.
- Type "pip install matplotlib" and click enter.

- Type "pip install scipy" and click enter.
- Type "pip install pickle-mixin" and click enter.
- Type "pip install seaborn" and click enter.
- Type "pip install Flask" and click enter.

8. Prior Knowledge:

You must have prior knowledge of following topics to complete this project.

- **ML Concepts**
- Supervised learning: https://www.javatpoint.com/supervised-machine-learning
- Unsupervised learning: https://www.javatpoint.com/unsupervised-machine-learning
- Regression and classification
- Decision tree: https://www.javatpoint.com/machine-learning-decision-tree-classification-algorithm
- Random forest: https://www.javatpoint.com/machine-learning-random-forest-algorithm
- KNN: https://www.javatpoint.com/k-nearest-neighbor-algorithm-for-machine-learning
- Xgboost: https://www.analyticsvidhya.com/blog/2018/09/an-end-to-end-guide-to-understand-the-math-behind-xgboost/
- Evaluation metrics: https://www.analyticsvidhya.com/blog/2019/08/11-important-model-evaluation-error-metrics/
- **Flask Basics** : https://www.youtube.com/watch?v=lj4I_CvBnt0

9. Project Flow:

- User interacts with the UI to enter the input.
- Entered input is analyzed by the model which is integrated.
- Once model analyses the input the prediction is showcased on the UI

To accomplish this, we have to complete all the activities listed below,

- Data collection

- Collect the dataset or create the dataset
- Visualizing and analysing the data
- Importing the libraries
- Read the Dataset
- Data pre-processing
- Handling missing values
- Descriptive analysis
- Splitting the dataset as x and y

- Handling Categorical Values
- Checking Correlation
- Converting Datatype
- Splitting dataset into training and test set
- Handled Imbalanced Data
- Applying StandardScaler
- Model building
- Import the model building libraries
- Performing Feature Importance
- Selected Output Columns
- Model Building on Selected Columns
- Random Forest Classifier
- XGBoost Classifier
- SVC
- Evaluating performance of model & Save the model
- Application Building
- Create an HTML file
- Build python code

10. **Results:**

We have applied our data to a range of machine learning algorithms (Decision Tree, SVM, Random Forest, Naive Bayes, Logistic Regression, Linear Discriminant Analysis, k-Nearest neighbors, Multi-Layer Perceptron) We divided the existing data into two parts, 30% for training and 70% for testing as this training is the first training on this data. In the first step we took all the properties in our data and applied them to a group of algorithms shown in the table below, and after the application process these results appeared to us. This practical part has been implemented on the python platform and is considered a complete and integrated platform. All attributes have been taken which are 16 inputs and one output.
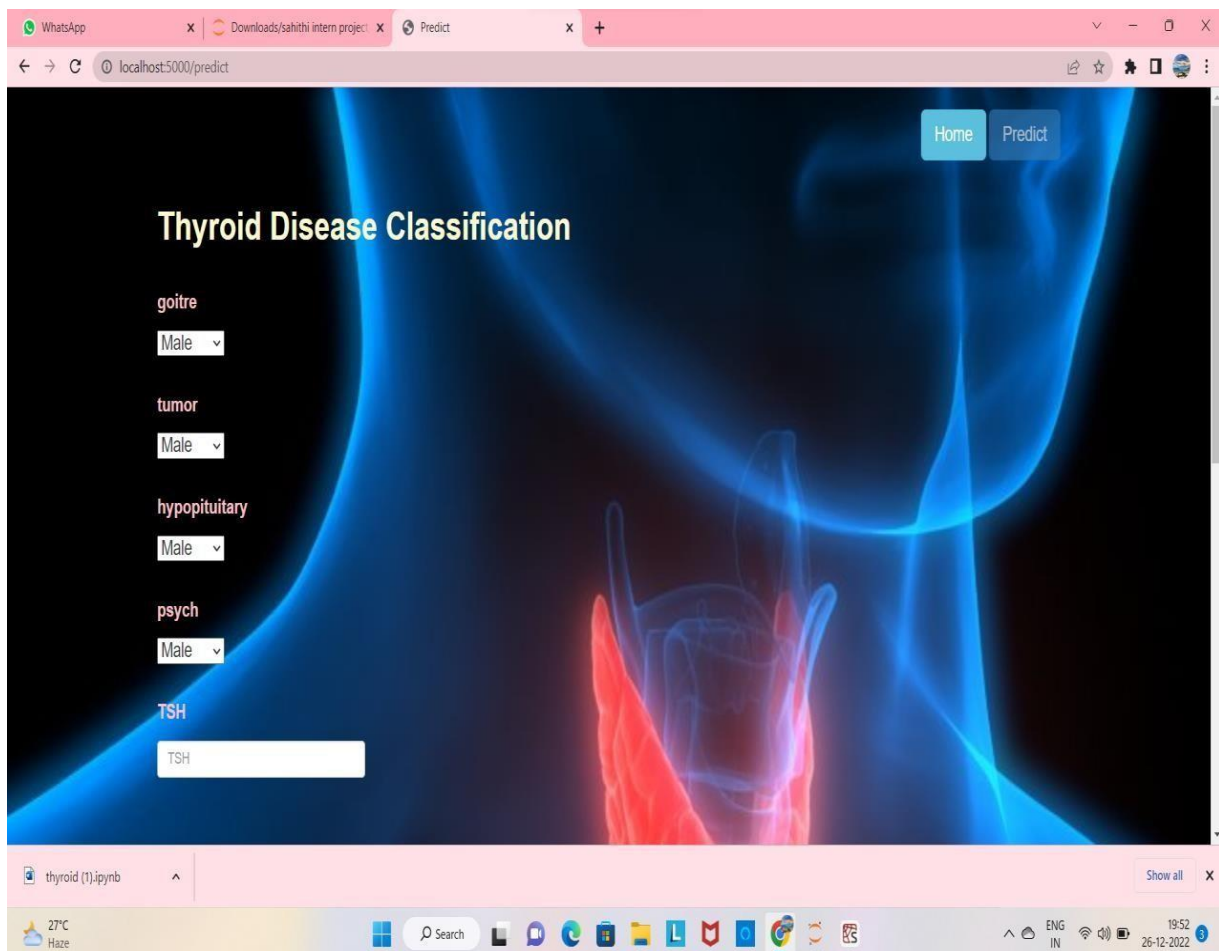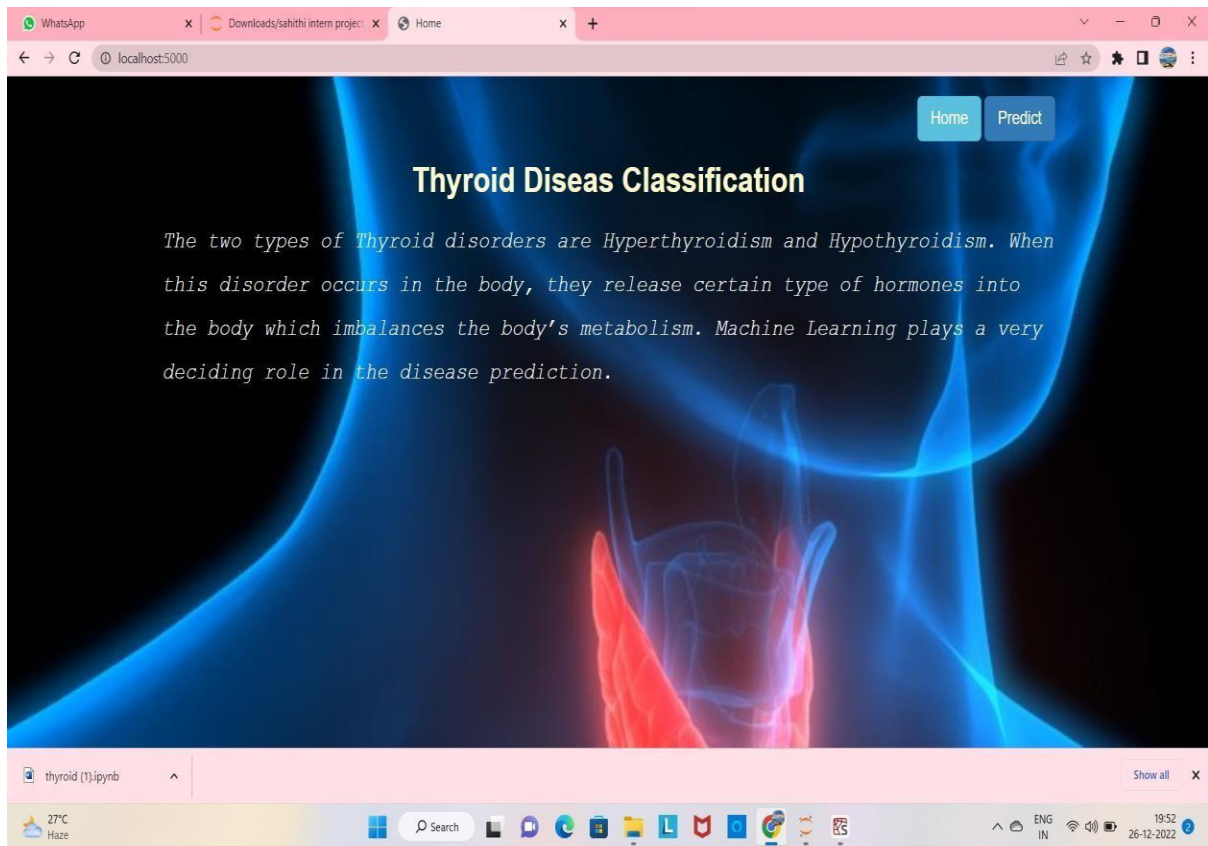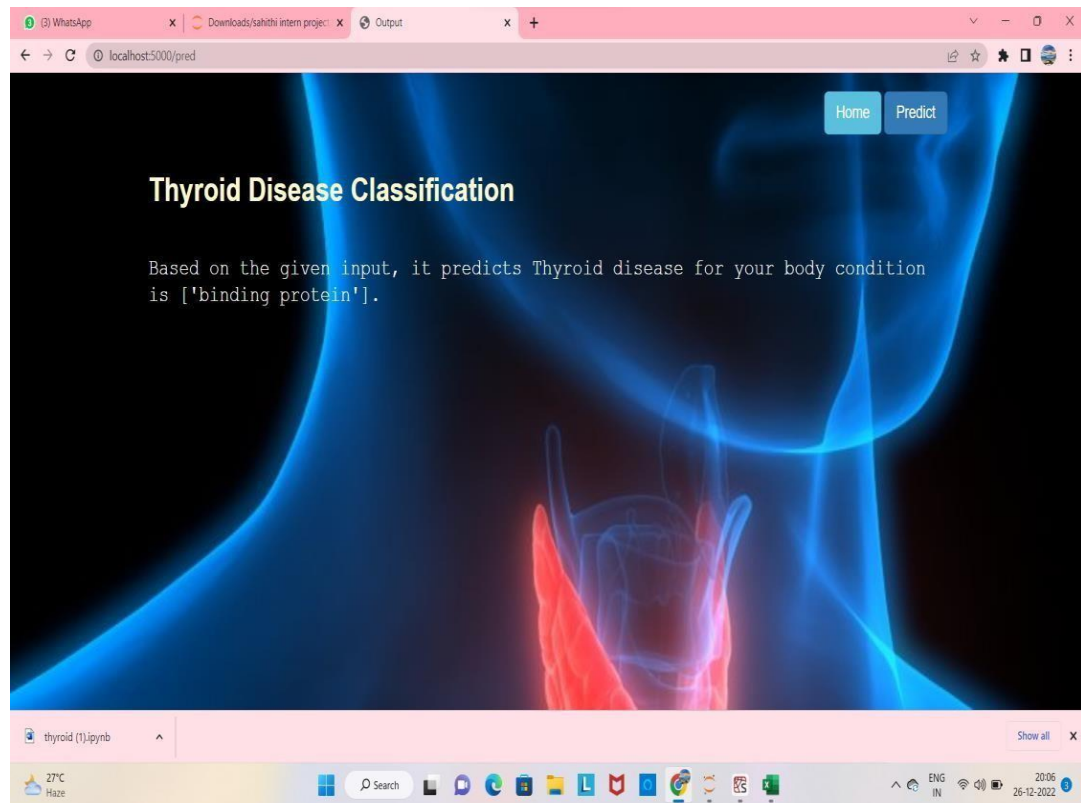
11. Conclusion:

Thyroid disease is one of the diseases that afflict the world's population, and the number of cases of this disease is increasing. Because of medical reports that show serious imbalances in thyroid diseases, our study deals with the classification of thyroid disease between hyperthyroidism and hypothyroidism. This disease was classified using algorithms. Machine learning showed us good results using several algorithms and was built in the form of two models. In the first model, all the characteristics consisting of 16 inputs and one output were taken, and the result of the accuracy of the random forest algorithm was 98.93, which is the highest accuracy among the other algorithms. In the second embodiment, the following characteristics were omitted based on a previous study. The removed attributes were 1-query_thyroxine 2- query_hypothyorid 3-query_hyperthyroid. Here we have included the increased accuracy of some algorithms, as well as the retention of the accuracy of others. It was observed that the accuracy of Naive Bayes algorithm increased the accuracy by 90.67. The highest precision of the MLP algorithm was 96.4 accuracy.

12. References:

[1] Azar, a.T, Hassanien, A.E. and Kim, T. Expert system based on neural fuzzy rules for thyroid diseases diagnosis, Computer Science, Artificial Intelligence, arXiv:1403.0522, Pp. 1-12,2012.

[2] Keles, A. ESTDD: Expert system for thyroid diseases diagnosis, Expert Syst Appl., Vol. 34, No.1, Pp.242–246,2008.

[3] a. c.c.Heuck, "World Health Organization," 2000. [Online]. Available: https://www.who.int/.

[4] Kouroua, K., Exarchosa, T.P. Exarchosa, K.P., Karamouzisc, M.V. andFotiadisa, D.I. (2015) Machine learning applications in cancer prognosis and prediction, Computational and Structural Biotechnology Journal, Vol. 13, Pp.8–17.

[5] Shukla, A. & Kaur, P. (2009). Diagnosis of thyroid disorders using artificial neural networks, IEEE International Advance computing Conference (IACC 2009)– Patiala, India, pp 1016-1020.

[6] Aswad, Salma Abdullah, and Emrullah Sonuç. "Classification of VPN Network Traffic Flow Using Time Related Features on Apache Spark." 2020 4th International Symposium on Multidisciplinary Studies and Innovative Technologies (ISMSIT). IEEE, 2020.

[7] Banu, G. Rasitha. "A Role of decision Tree classification data Mining Technique in Diagnosing Thyroid disease." International Journal of Computer Sciences and Engineering 4.11 (2016): 64-70.

[8] Chandio, Jamil Ahmed, et al. "TDV: Intelligent system for thyroid disease visualization." 2016 International Conference on Computing, Electronic and Electrical Engineering (ICE Cube). IEEE, 2016. [9] Travis B Murdoch and Allan S Detsky. The inevitable application of big data to health care. Jama, 309(13):1351–1352, 2013.

[10] Dr. Srinivasan B, Pavya K "Diagnosis of Thyroid Disease: A Study" International Research Journal of Engineering and Technology Volume: 03 Issue: 11 | Nov – 2016

[11] Aytürk Keleş and Keleş, Ali. "ESTDD: Expert system for thyroid diseases diagnosis." International Research Journal of Engineering and Technology (IRJET) Volume: 03 Issue: 11 | Nov - 2017 34.1 (2017): 242- 246

[12] Khushboo Taneja, Parveen Sehgal, Prerana "Predictive Data Mining for Diagnosis of Thyroid Disease using Neural Network" International Journal of Research in Management, Science & Technology (E-ISSN: 2321- 3264) Vol. 3, No. 2, April 2016

[13] Chandel, Khushboo, et al. "A comparative study on thyroid disease detection using K-nearest neighbor and Naive Bayes classification techniques." CSI transactions on ICT 4.2-4 (2016): 313-319.

[14] Banu, G. Rasitha. "A Role of decision Tree classification data Mining Technique in Diagnosing Thyroid disease." International Journal of Computer Sciences and Engineering 4.11 (2016): 64-70.
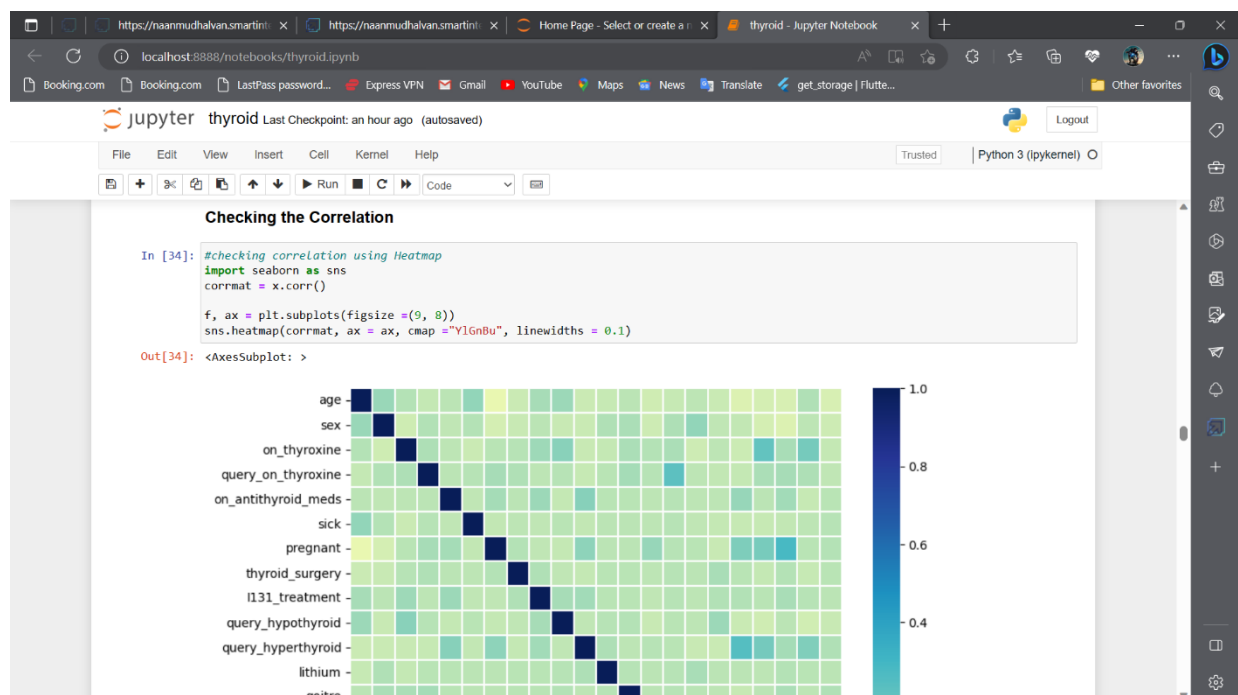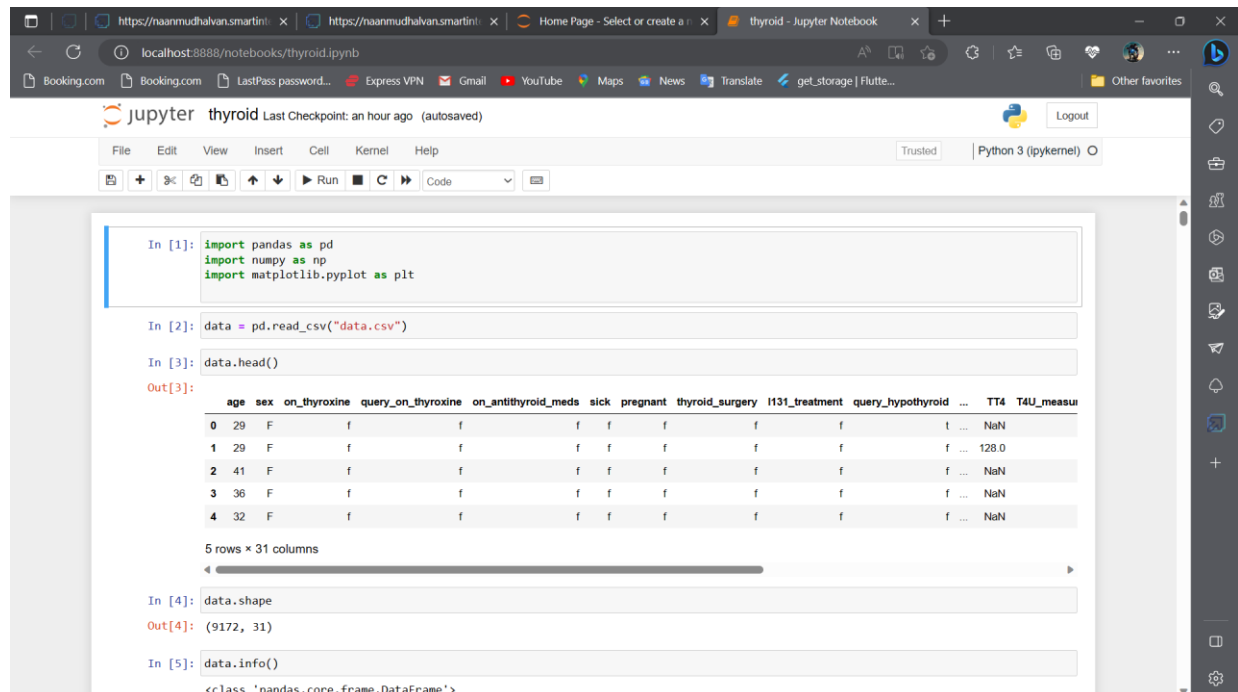
13. Bibiliography:

https://www.researchgate.net/publication/353467967_Thyroid_Disease_Classification_Using_Machine_Learning_Algorithms#:~:text=The%20goal%20of%20this%20study,used%20all%20of%20the%20algorithms

Appendix:

SOURCE CODE:

thyroid.ipynb

RandomForest Model-1

```
In [55]: rfr1 = RandomForestClassifier()
         rfr1.fit(x_bal,y_bal)
         y_pred=rfr1.predict(x_test_bal)
```

```
C:\Users\HP\AppData\Local\Temp\ipykernel_15788\1228087459.py:2: DataConversionWarning: A column-vector y was passed when a 1d a
rray was expected. Please change the shape of y to (n_samples,), for example using ravel().
  rfr1.fit(x_bal,y_bal)
```

```
In [56]: print(classification_report(y_test_bal,y_pred))
```

```
              precision    recall  f1-score   support

           0       0.77      0.08      0.15       122
           1       0.81      0.93      0.86       122
           2       0.88      0.98      0.93       122
           3       0.76      0.83      0.79       122
           4       0.47      0.89      0.62       122
           5       0.87      0.70      0.77       122
           6       0.57      0.49      0.53       122

    accuracy                           0.70       854
   macro avg       0.73      0.70      0.66       854
weighted avg       0.73      0.70      0.66       854
```

```
In [57]: train_score = accuracy_score(y_bal,rfr1.predict(x_bal))
```

```
In [58]: train_score
```

```
In [57]: train_score = accuracy_score(y_bal,rfr1.predict(x_bal))
```

```
In [58]: train_score
```

```
Out[58]: 1.0
```

XGBClassifier Model-2

```
In [59]: from xgboost import XGBClassifier
         xgb = XGBClassifier()
         xgb.fit(x_bal,y_bal)
```

```
Out[59]:                        XGBClassifier
              colsample_bylevel=None, colsample_bynode=None,
              colsample_bytree=None, early_stopping_rounds=None,
              enable_categorical=False, eval_metric=None, feature_types=None,
              gamma=None, gpu_id=None, grow_policy=None, importance_type=None,
              interaction_constraints=None, learning_rate=None, max_bin=None,
              max_cat_threshold=None, max_cat_to_onehot=None,
              max_delta_step=None, max_depth=None, max_leaves=None,
              min_child_weight=None, missing=nan, monotone_constraints=None,
              n_estimators=100, n_jobs=None, num_parallel_tree=None,
              objective='multi:softprob', predictor=None, ...)
```

```
In [60]: y_pred=xgb.predict(x_test_bal)
```

```
In [61]: print(classification_report(y_test_bal,y_pred))
```

https://naanmudhalvan.smartint × | https://naanmudhalvan.smartint × | Home Page - Select or create a n × | thyroid - Jupyter Notebook × | +

localhost:8888/notebooks/thyroid.ipynb

Booking.com   Booking.com   LastPass password...   Express VPN   Gmail   YouTube   Maps   News   Translate   get_storage | Flutte...   Other favorites

jupyter thyroid Last Checkpoint: an hour ago (autosaved)   Logout

File   Edit   View   Insert   Cell   Kernel   Help   Trusted | Python 3 (ipykernel) O

```
In [60]: y_pred=xgb.predict(x_test_bal)
```

```
In [61]: print(classification_report(y_test_bal,y_pred))
```

```
              precision    recall  f1-score   support

           0       0.86      0.30      0.45       122
           1       0.82      0.93      0.87       122
           2       0.96      1.00      0.98       122
           3       0.77      0.84      0.80       122
           4       0.53      0.86      0.66       122
           5       0.83      0.70      0.76       122
           6       0.59      0.53      0.56       122

    accuracy                           0.74       854
   macro avg       0.77      0.74      0.73       854
weighted avg       0.77      0.74      0.73       854
```

```
In [62]: train_score = accuracy_score(y_bal, xgb.predict(x_bal))
         train_score
```

```
Out[62]: 1.0
```

SVC Model-3

```
In [63]: from sklearn.svm import SVC
         from sklearn.metrics import accuracy_score, classification_report

         sv= SVC()
```

SVC Model-3

```
In [63]: from sklearn.svm import SVC
         from sklearn.metrics import accuracy_score, classification_report

         sv= SVC()
```

```
In [64]: sv.fit(x_bal,y_bal)
```

C:\Users\HP\AppData\Local\Programs\Python\Python310\lib\site-packages\sklearn\utils\validation.py:1143: DataConversionWarning: A column-vector y was passed when a 1d array was expected. Please change the shape of y to (n_samples,), for example using ravel().
    y = column_or_1d(y, warn=True)

```
Out[64]: ▾ SVC
         SVC()
```

```
In [65]: y_pred = sv.predict(x_test_bal)
```

```
In [66]: print(classification_report(y_test_bal,y_pred))
```

```
              precision    recall  f1-score   support

           0       0.70      0.85      0.77       122
           1       0.78      0.81      0.80       122
           2       0.88      0.92      0.90       122
           3       0.73      0.70      0.72       122
           4       0.73      0.63      0.68       122
           5       0.83      0.53      0.65       122
           6       0.51      0.62      0.56       122
```

Jupyter thyroid Last Checkpoint: an hour ago (autosaved)

Logout

File Edit View Insert Cell Kernel Help

Trusted | Python 3 (ipykernel)

```
In [67]: train_score=accuracy_score(y_bal,sv.predict(x_bal))
         train_score
```

Out[67]: 0.7209584470730968

Grid_Search for RandomForest

```
In [68]: params={
             'n_estimators': [100, 200, 500],
             'criterion': ['gini', 'entropy'],
             'max_depth' : [x for x in range(1,20)]
         }
```

```
In [71]: from sklearn.model_selection import GridSearchCV
         grid_search = GridSearchCV(rfr1, params, scoring='accuracy',cv=5,n_jobs=-1)
```

```
In [72]: grid_search.fit(x_bal,y_bal)
```

C:\Users\HP\AppData\Local\Programs\Python\Python310\lib\site-packages\sklearn\model_selection_search.py:909: DataConversionWar
ning: A column-vector y was passed when a 1d array was expected. Please change the shape of y to (n_samples,), for example usin
g ravel().
  self.best_estimator_.fit(X, y, **fit_params)

Out[72]:
>   **GridSearchCV**
> estimator: RandomForestClassifier
>   RandomForestClassifier

```
In [73]: grid_search.best_params_
```

Out[73]: {'criterion': 'gini', 'max_depth': 18, 'n_estimators': 500}

```
In [74]: rfr_gs=RandomForestClassifier(criterion= 'entropy', max_depth=16, n_estimators= 200)
```

```
In [75]: rfr_gs.fit(x_bal, y_bal)
```

C:\Users\HP\AppData\Local\Temp\ipykernel_15788\1700014317.py:1: DataConversionWarning: A column-vector y was passed when a 1d a
rray was expected. Please change the shape of y to (n_samples,), for example using ravel().
  rfr_gs.fit(x_bal, y_bal)

Out[75]:
>   RandomForestClassifier
> RandomForestClassifier(criterion='entropy', max_depth=16, n_estimators=200)

```
In [76]: y_pred=rfr_gs.predict(x_test_bal)
```

```
In [77]: print(classification_report(y_test_bal,y_pred))
```

```
              precision    recall  f1-score   support

           0       0.69      0.07      0.13       122
           1       0.81      0.94      0.87       122
           2       0.92      0.99      0.96       122
           3       0.78      0.83      0.80       122
           4       0.46      0.89      0.61       122
           5       0.89      0.70      0.78       122
           6       0.61      0.53      0.57       122
```

```
In [78]: train_score= accuracy_score(y_bal,rfr_gs.predict(x_bal))
         train_score

Out[78]: 1.0
```

Grid_Search for XGB

```
In [79]: params={
             'n_estimators': [100, 200, 500],
             'learning_rate': [0.01,0.05,0.1],
             'booster': ['gbtree', 'gblinear'],
             'gamma': [0, 0.5, 1],

         }
```

```
In [80]: grid_xgb = GridSearchCV(xgb, params, scoring='accuracy',cv=3,n_jobs=-1)
```

```
In [81]: grid_xgb.fit(x_bal,y_bal)
```

```
Out[81]:        ▸        GridSearchCV
         ▸ estimator: XGBClassifier
                     ▸ XGBClassifier
```

```
In [82]: grid_xgb.best_params_
```

Grid_Search for XGB

```
In [79]: params={
             'n_estimators': [100, 200, 500],
             'learning_rate': [0.01,0.05,0.1],
             'booster': ['gbtree', 'gblinear'],
             'gamma': [0, 0.5, 1],

         }
```

```
In [80]: grid_xgb = GridSearchCV(xgb, params, scoring='accuracy',cv=3,n_jobs=-1)
```

```
In [81]: grid_xgb.fit(x_bal,y_bal)
```

```
Out[81]:        ▸        GridSearchCV
         ▸ estimator: XGBClassifier
                     ▸ XGBClassifier
```

```
In [82]: grid_xgb.best_params_
```

```
Out[82]: {'booster': 'gbtree', 'gamma': 0, 'learning_rate': 0.1, 'n_estimators': 200}
```

```
In [83]: xgb1=XGBClassifier(booster='gbtree', gamma= 0, learning_rate= 0.1, n_estimators= 500)
```

```
In [84]: xgb1.fit(x_bal,y_bal)
```

jupyter thyroid Last Checkpoint: an hour ago (autosaved)   Logout

File   Edit   View   Insert   Cell   Kernel   Help   Trusted | Python 3 (ipykernel) ○

```
In [84]: xgb1.fit(x_bal,y_bal)
```

Out[84]:
```
                    XGBClassifier
            colsample_bylevel=None, colsample_bynode=None,
            colsample_bytree=None, early_stopping_rounds=None,
            enable_categorical=False, eval_metric=None, feature_types=None,
            gamma=0, gpu_id=None, grow_policy=None, importance_type=None,
            interaction_constraints=None, learning_rate=0.1, max_bin=None,
            max_cat_threshold=None, max_cat_to_onehot=None,
            max_delta_step=None, max_depth=None, max_leaves=None,
            min_child_weight=None, missing=nan, monotone_constraints=None,
            n_estimators=500, n_jobs=None, num_parallel_tree=None,
            objective='multi:softprob', predictor=None, ...)
```

```
In [85]: y_pred= xgb1.predict(x_test_bal)
```

```
In [86]: print(classification_report(y_test_bal,y_pred))
```

```
              precision    recall  f1-score   support

           0       0.83      0.28      0.42       122
           1       0.82      0.93      0.87       122
           2       0.96      1.00      0.98       122
           3       0.76      0.84      0.80       122
           4       0.50      0.80      0.61       122
           5       0.83      0.75      0.78       122
           6       0.59      0.52      0.55       122

    accuracy                           0.73       854
```

```
In [86]: print(classification_report(y_test_bal,y_pred))
```

```
              precision    recall  f1-score   support

           0       0.83      0.28      0.42       122
           1       0.82      0.93      0.87       122
           2       0.96      1.00      0.98       122
           3       0.76      0.84      0.80       122
           4       0.50      0.80      0.61       122
           5       0.83      0.75      0.78       122
           6       0.59      0.52      0.55       122

    accuracy                           0.73       854
   macro avg       0.75      0.73      0.72       854
weighted avg       0.75      0.73      0.72       854
```

```
In [87]: train_score= accuracy_score(y_bal,xgb1.predict(x_bal))
         train_score
```

Out[87]: 1.0

Grid_Serach for SVC

```
In [88]: params = {
             'C' : [0.01, 0.1, 1, 10,100, 1000],
             'kernel': ['rbf', 'linear'],
             'gamma': [1, 0.1, 0.01, 0.001, 0.0001],
         }
```

Jupyter thyroid Last Checkpoint: an hour ago (autosaved)   Logout

File | Edit | View | Insert | Cell | Kernel | Help   Trusted | Python 3 (ipykernel) ○

```
In [89]: grid_svc = GridSearchCV(sv,params, scoring='accuracy',cv=5,n_jobs=-1)
```

```
In [90]: grid_svc.fit(x_bal,y_bal)
```

C:\Users\HP\AppData\Local\Programs\Python\Python310\lib\site-packages\sklearn\utils\validation.py:1143: DataConversionWarning:
A column-vector y was passed when a 1d array was expected. Please change the shape of y to (n_samples,), for example using rav
el().
  y = column_or_1d(y, warn=True)

Out[90]:   ▸ GridSearchCV
           ▸ estimator: SVC
                ▸ SVC

```
In [91]: grid_svc.best_params_
```

Out[91]: {'C': 1000, 'gamma': 1, 'kernel': 'rbf'}

```
In [92]: sv1=SVC(C =1000,gamma=1, kernel= 'rbf')
```
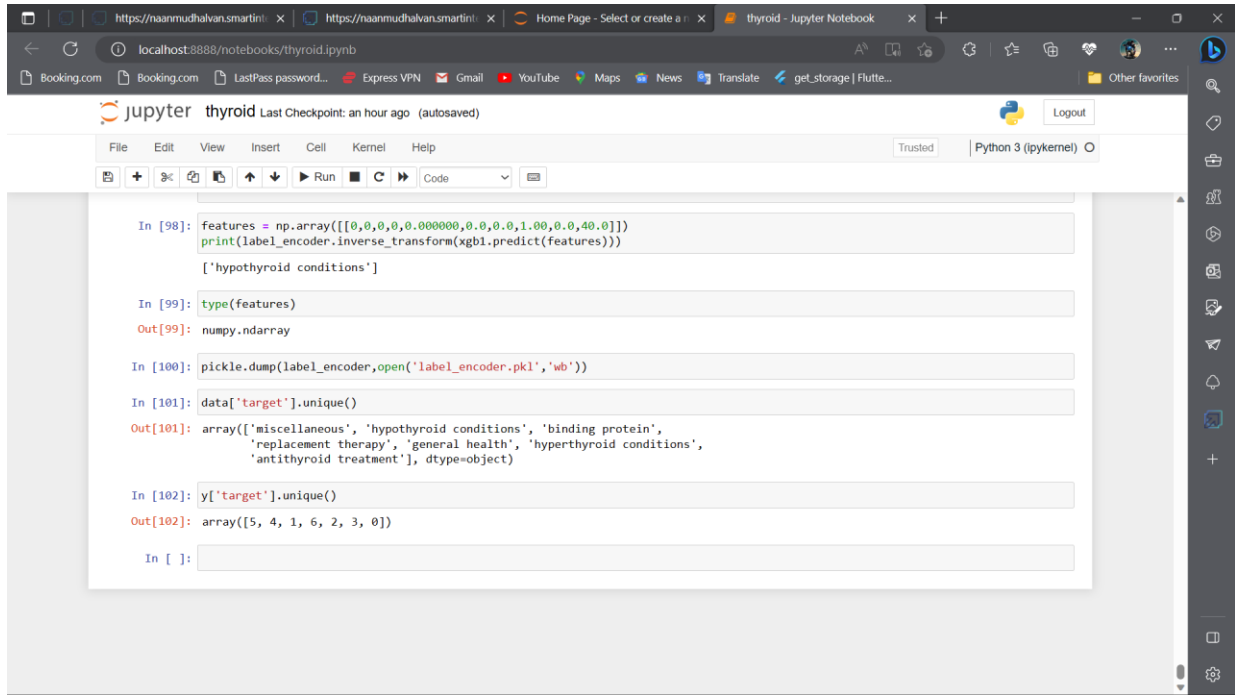
```
In [93]: sv1.fit(x_bal,y_bal)
```

C:\Users\HP\AppData\Local\Programs\Python\Python310\lib\site-packages\sklearn\utils\validation.py:1143: DataConversionWarning:
A column-vector y was passed when a 1d array was expected. Please change the shape of y to (n_samples,), for example using rav
el().
  y = column_or_1d(y, warn=True)

Out[93]:   ▾   SVC
          SVC(C=1000, gamma=1)

```
In [94]: y_pred= sv1.predict(x_test_bal)
```

```
In [95]: print(classification_report(y_test_bal,y_pred))
```

```
              precision    recall  f1-score   support

           0       0.79      0.43      0.56       122
           1       0.65      0.90      0.76       122
           2       0.92      0.89      0.90       122
           3       0.70      0.64      0.67       122
           4       0.59      0.81      0.68       122
           5       0.77      0.69      0.73       122
           6       0.50      0.46      0.48       122

    accuracy                           0.69       854
   macro avg       0.70      0.69      0.68       854
weighted avg       0.70      0.69      0.68       854
```

```
In [96]: train_score= accuracy_score(y_bal,sv1.predict(x_bal))
         train_score
```

Out[96]: 0.9511677282377919

```
In [97]: # saving the model
         import pickle
         pickle.dump(xgb1,open('thyroid_1_model.pkl','wb'))
```

```
In [98]: features = np.array([[0,0,0,0,0.000000,0.0,0.0,1.00,0.0,40.0]])
         print(label_encoder.inverse_transform(xgb1.predict(features)))
```

In [98]: `features = np.array([[0,0,0,0,0.000000,0.0,0.0,1.00,0.0,40.0]])`
`print(label_encoder.inverse_transform(xgb1.predict(features)))`

`['hypothyroid conditions']`

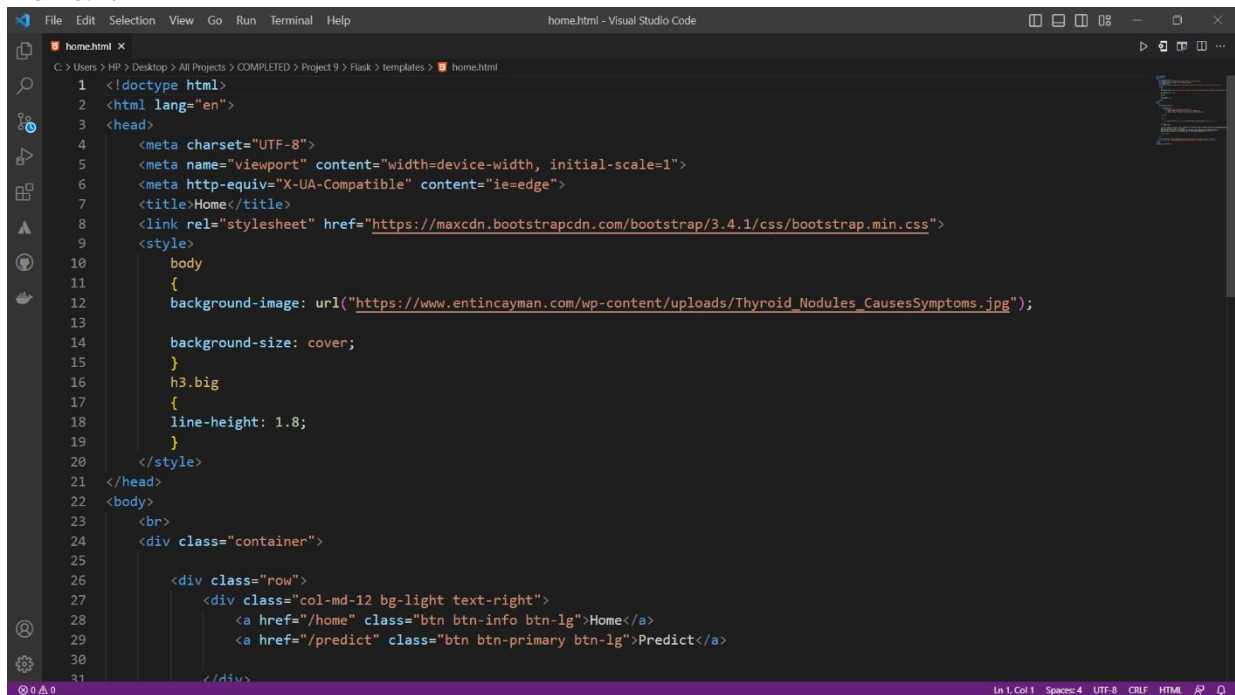In [99]: `type(features)`

Out[99]: `numpy.ndarray`

In [100]: `pickle.dump(label_encoder,open('label_encoder.pkl','wb'))`

In [101]: `data['target'].unique()`

Out[101]: `array(['miscellaneous', 'hypothyroid conditions', 'binding protein',`
`'replacement therapy', 'general health', 'hyperthyroid conditions',`
`'antithyroid treatment'], dtype=object)`

In [102]: `y['target'].unique()`

Out[102]: `array([5, 4, 1, 6, 2, 3, 0])`

In [ ]:

Home.html

```html
<!doctype html>
<html lang="en">
<head>
    <meta charset="UTF-8">
    <meta name="viewport" content="width=device-width, initial-scale=1">
    <meta http-equiv="X-UA-Compatible" content="ie=edge">
    <title>Home</title>
    <link rel="stylesheet" href="https://maxcdn.bootstrapcdn.com/bootstrap/3.4.1/css/bootstrap.min.css">
    <style>
        body
        {
        background-image: url("https://www.entincayman.com/wp-content/uploads/Thyroid_Nodules_CausesSymptoms.jpg");

        background-size: cover;
        }
        h3.big
        {
        line-height: 1.8;
        }
    </style>
</head>
<body>
    <br>
    <div class="container">

        <div class="row">
            <div class="col-md-12 bg-light text-right">
                <a href="/home" class="btn btn-info btn-lg">Home</a>
                <a href="/predict" class="btn btn-primary btn-lg">Predict</a>

            </div>
```

home.html ×

C: > Users > HP > Desktop > All Projects > COMPLETED > Project 9 > Flask > templates > 🥥 home.html > 🔗 p

```html
26          <div class="row">
27              <div class="col-md-12 bg-light text-right">
28                  <a href="/home" class="btn btn-info btn-lg">Home</a>
29                  <a href="/predict" class="btn btn-primary btn-lg">Predict</a>
30
31              </div>
32          </div>
33
34          <center>
35
36              <font color="#FFFDD0"><h1><strong>Thyroid Diseas Classification</strong></h1></font>
37
38          </center>
39          <h3 class="big">
40
41           <p><font face = "courier" color = "#FFFFFF"><em>The two types of Thyroid disorders are Hyperthyroidism and Hypothyroidism
42          When this disorder occurs in the body,
43          they release certain type of hormones into the body which imbalances the body?s metabolism.
44          Machine Learning plays a very deciding role in the disease prediction.</font></p>
45
46          </em></h3><br>
47
48      </div>
49
50      <script src="https://ajax.googleapis.com/ajax/libs/jquery/3.5.1/jquery.min.js"></script>
51      <script src="https://maxcdn.bootstrapcdn.com/bootstrap/3.4.1/js/bootstrap.min.js"></script>
52  </body>
53  </html>
54  <p style="color: ■#FFFFFF">
```

Predict.html

predict.html ×

C: > Users > HP > Desktop > All Projects > COMPLETED > Project 9 > Flask > templates > 🥥 predict.html

```html
1  <!DOCTYPE html>
2  <html lang="en">
3  <head>
4      <meta charset="UTF-8">
5      <title>Predict</title>
6      <link rel="stylesheet" href="https://maxcdn.bootstrapcdn.com/bootstrap/3.4.1/css/bootstrap.min.css">
7      <style>
8          body
9          {
10         background-image: url("https://www.entincayman.com/wp-content/uploads/Thyroid_Nodules_CausesSymptoms.jpg");
11         background-size: cover;
12         }
13         h3.big
14         {
15         line-height: 1.8;
16         }
17     </style>
18  </head>
19  <body>
20      <br>
21      <div class="container">
22
23          <div class="row">
24              <div class="col-md-12 bg-light text-right">
25                  <a href="/home" class="btn btn-info btn-lg">Home</a>
26                  <a href="/predict" class="btn btn-primary disabled btn-lg">Predict</a>
27              </div>
28          </div>
29
30          <br>
31          <font color="#FFFDD0"><h1><strong>Thyroid Disease Classification</strong></h1><br></font>
```

predict.html - Visual Studio Code

predict.html ✕

C: > Users > HP > Desktop > All Projects > COMPLETED > Project 9 > Flask > templates > 5 predict.html > ❤ html > ❤ body > ❤ div.container > ❤ h4 > ❤ form > ❤ div.form-group.mb-3 > ❤ select#hypopituitary.custom-select > ❤ option

```html
31      <font color="#fffbb0"><h1><strong>Thyroid Disease Classification</strong></h1><br></font>
32
33      <h4>
34      <form action="/pred", method="post">
35
36          <div class="form-group mb-3">
37          <div class="input-group-prepend">
38              <p style="color: pink;"><label class="input-group-text" for="goitre">goitre</label></p>
39          </div>
40            <select class="custom-select" name="goitre" id="goitre">
41              <option value="0">Male</option>
42              <option value="1">Female</option>
43            </select>
44          </div><br>
45
46          <div class="form-group mb-3">
47          <div class="input-group-prepend">
48              <p style="color: pink;"><label class="input-group-text" for="tumor">tumor</label></p>
49          </div>
50            <select class="custom-select" name="tumor" id="tumor">
51              <option value="0">Male</option>
52              <option value="1">Female</option>
53            </select>
54          </div><br>
55
56          <div class="form-group mb-3">
57          <div class="input-group-prepend">
58              <p style="color: pink;"><label class="input-group-text" for="hypopituitary">hypopituitary</label></p>
59          </div>
60            <select class="custom-select" name="hypopituitary" id="hypopituitary">
61              <option value="0">Male</option>
62              <option value="1">Female</option>
```

predict.html - Visual Studio Code

predict.html ✕

C: > Users > HP > Desktop > All Projects > COMPLETED > Project 9 > Flask > templates > 5 predict.html > ❤ html > ❤ body > ❤ div.container > ❤ h4 > ❤ form > ❤ div.form-group.row

```html
60              <option value="0">Male</option>
61              <option value="1">Female</option>
62            </select>
63          </div><br>
64
65          <div class="form-group mb-3">
66          <div class="input-group-prepend">
67              <p style="color: pink;"><label class="input-group-text" for="psych">psych</label></p>
68          </div>
69            <select class="custom-select" name="psych" id="psych">
70              <option value="0">Male</option>
71              <option value="1">Female</option>
72            </select>
73          </div><br>
74
75          <div class="form-group row">
76              <div class="col-md-3">
77                  <p style="color: pink;"><label for="TSH">TSH</label></p>
78                  <p style="color: #ff0000;"><input type="text" name="TSH" id="TSH" class="form-control" placeholder="TSH"
79              </div>
80          </div>
81
82           <div class="form-group row">
83              <div class="col-md-3">
84                  <p style="color: pink;"><label for="T3">T3</label></p>
85                  <input type="text" name="T3" id="T3" class="form-control" placeholder="T3" required="required"/><br><br>
86              </div>
87          </div>
88
89            <div class="form-group row">
90              <div class="col-md-3">
```

predict.html - Visual Studio Code

File   Edit   Selection   View   Go   Run   Terminal   Help

predict.html ✕

C: > Users > HP > Desktop > All Projects > COMPLETED > Project 9 > Flask > templates > ⑤ predict.html > ⊘ html > ⊘ body > ⊘ div.container > ⊘ h4 > ⊘ form

```html
            <div class="col-md-3">
                <p style="color: pink;"><label for="TT4">TT4</label></p>
                <input type="text" name="TT4"  id="TT4" class="form-control" placeholder="TT4" required="required"/><br><br
            </div>
        </div>

        <div class="form-group row">
            <div class="col-md-3">
                <p style="color: pink;"><label for="T4U">T4U</label></p>
                <input type="text" name="T4U"  id="T4U" class="form-control" placeholder="T4U" required="required"/><br><b
            </div>
        </div>

        <div class="form-group row">
            <div class="col-md-3">
                <p style="color: pink;"><label for="FTI">FTI</label></p>
                <input type="text" name="FTI"  id="FTI" class="form-control" placeholder="FTI" required="required"/><br><b
            </div>
        </div>

        <div class="form-group row">
            <div class="col-md-3">
                <p style="color: pink;"><label for="TBG">TBG</label></p>
                <input type="text" name="TBG"  id="TBG" class="form-control" placeholder="TBG" required="required"/><br><b
            </div>
        </div>

        <button type="submit" class="btn btn-success btn-lg">Submit</button>
```

```html
        </div>

        <div class="form-group row">
            <div class="col-md-3">
                <p style="color: pink;"><label for="FTI">FTI</label></p>
                <input type="text" name="FTI"  id="FTI" class="form-control" placeholder="FTI" required="required"/><br><b
            </div>
        </div>

        <div class="form-group row">
            <div class="col-md-3">
                <p style="color: pink;"><label for="TBG">TBG</label></p>
                <input type="text" name="TBG"  id="TBG" class="form-control" placeholder="TBG" required="required"/><br><b
            </div>
        </div>

        <button type="submit" class="btn btn-success btn-lg">Submit</button>



    </form>
    <br>
    </h4>
</div>


<script src="https://ajax.googleapis.com/ajax/libs/jquery/3.5.1/jquery.min.js"></script>
<script src="https://maxcdn.bootstrapcdn.com/bootstrap/3.4.1/js/bootstrap.min.js"></script>
</body>
</html>
```

Submit.html

```html
1    <!DOCTYPE html>
2    <html lang="en">
3    <head>
4        <meta charset="UTF-8">
5        <title>Output</title>
6        <link rel="stylesheet" href="https://maxcdn.bootstrapcdn.com/bootstrap/3.4.1/css/bootstrap.min.css">
7        <style>
8            body {
9            background-image: url("https://www.entincayman.com/wp-content/uploads/Thyroid_Nodules_CausesSymptoms.jpg");
10           background-size: cover;
11           }
12           h3.big
13           {
14           line-height: 1.8;
15           }
16       </style>
17   </head>
18   <body>
19       <br>
20           <div class="container">
21
22               <div class="row">
23                   <div class="col-md-12 bg-light text-right">
24                       <a href="/home" class="btn btn-info btn-lg">Home</a>
25                       <a href="/predict" class="btn btn-primary btn-lg">Predict</a>
26
27                   </div>
28               </div>
29               <br>
30               <font color="#FFFDD0"><h1><strong>Thyroid Disease Classification</strong></h1></font><br>
31               <p><font face = "courier" color = "#FFFFFF"><h3>
```

```html
10       background-size: cover;
11       }
12       h3.big
13       {
14       line-height: 1.8;
15       }
16   </style>
17   </head>
18   <body>
19       <br>
20           <div class="container">
21
22               <div class="row">
23                   <div class="col-md-12 bg-light text-right">
24                       <a href="/home" class="btn btn-info btn-lg">Home</a>
25                       <a href="/predict" class="btn btn-primary btn-lg">Predict</a>
26
27                   </div>
28               </div>
29               <br>
30               <font color="#FFFDD0"><h1><strong>Thyroid Disease Classification</strong></h1></font><br>
31               <p><font face = "courier" color = "#FFFFFF"><h3>
32               Based on the given input, it predicts Thyroid disease for your body condition is {{prediction_text}}.
33               </h3>
34               </font></p>
35
36           </div>
37
38   </body>
39   </html>
```

App.py

```python
from flask import Flask, render_template, request
import numpy as np
import pickle
import pandas as pd

model = pickle.load(open(r"thyroid_1_model.pkl", 'rb'))
le = pickle.load(open("label_encoder.pkl", 'rb'))


app = Flask(__name__)


@app.route("/")
def about():
    return render_template('home.html')


@app.route("/predict")
def home1():
    return render_template('predict.html')


@app.route("/pred", methods=['POST', 'GET'])
def predict():
    x = [[float(x) for x in request.form.values()]]

    print(x)
    col = ['goitre', 'tumor', 'hypopituitary', 'psych',
           'TSH', 'T3', 'TT4', 'T4U', 'FTI', 'TBG']
    x = pd.DataFrame(x, columns=col)
```