THYROID DISEASE CLASSIFICATION USING

MACHINE LEARNING

INTRODUCTION

1. OVERVIEW
2. PURPOSE

* PROJECT FLOW
* PROJECT STRUCTURE

1. PROBLEM DEFINITION & DESIGN THINKING

* EMPATHY MAP
* BRAINSTORMING MAP

4. RESULT

5. ADVANTAGES

6. DISADVANTAGES

7. APPLICATIONS

8. CONCLUSION

9. FUTURE SCOPE

## 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 occurs in 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 and 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.

## PURPOSE

The Thyroid is defined as the “power house” of our body: if something in this gland fails, the whole body suffers as well. Therefore, the early diagnosis of a possible malfunction plays a fundamental role, as well as the prediction of the course of treatment of a patient with Hypothyroidism.

Which can be of great help for doctors who have patients under treatment. In this study we proposed an approach to predict the thyroid disease treatment .

This approach is intended to be a machine learning –based decision support system for endocrinologists treating patients with thyroid disease. The proposed feature model is evaluated using 10 different machine learning classifiers, belonging to the class of enhancement algorithms.

## PROJECT FLOW

* The user interacts with the UI to enter the input.
* Entered input is analyses by the model which is integrated.
* Once the model analyses the input the prediction is showcased on the UI.

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

* Define problem/problem understanding
* Specify the business problem
* Business requirements
* Literature Survey
* social or Business Impact
* Data Collection & Preparation
* Collect the dataset
* Data Preparation
* Exploratory Data Analysis
* Descriptive statistical
* Visual Analysis
* Model Building
* Training the model in multiple algorithms
* Testing the model
* Performance Testing & Hyper parameter Tuning
* Testing model with multiple evaluation metrics
* Comparing model accuracy before & after applying hyperparameter tuning
* Model Deployment
* Save the best model
* Integrate with web Framework
* Project Demonstration & Documentation
* Record explanation Video for project end to end solution
* Project Documentation step by step project development procedure

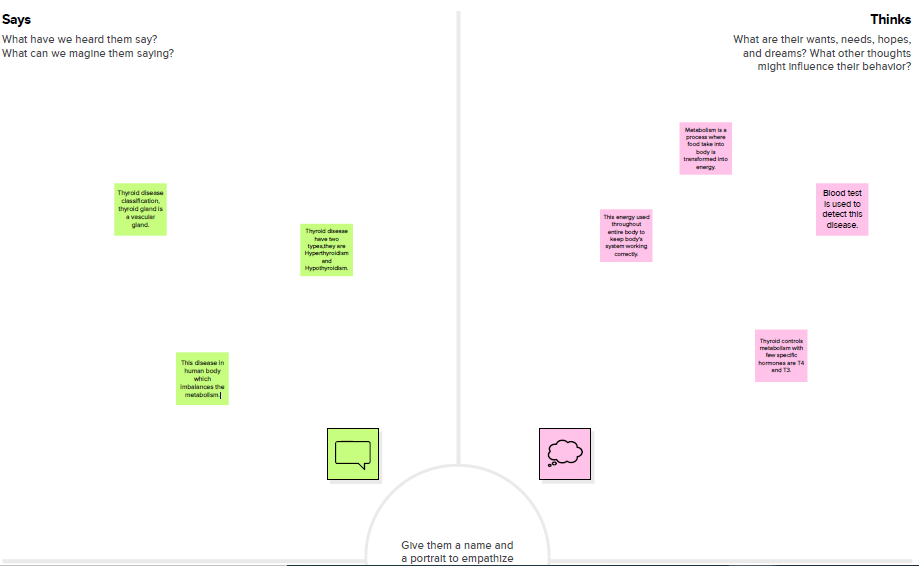
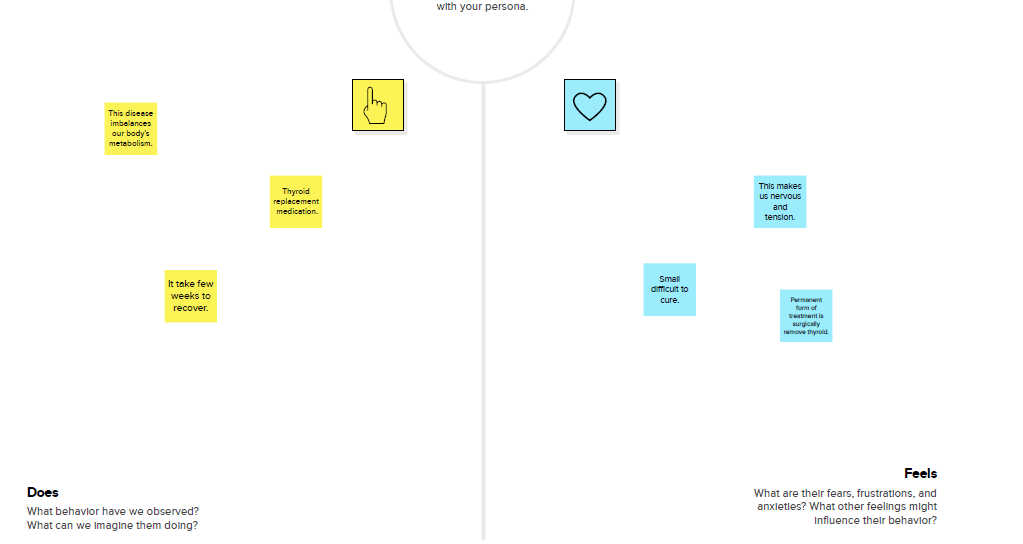
## PROJECT STRUCTURE

Create the project folder which contains files as shown below

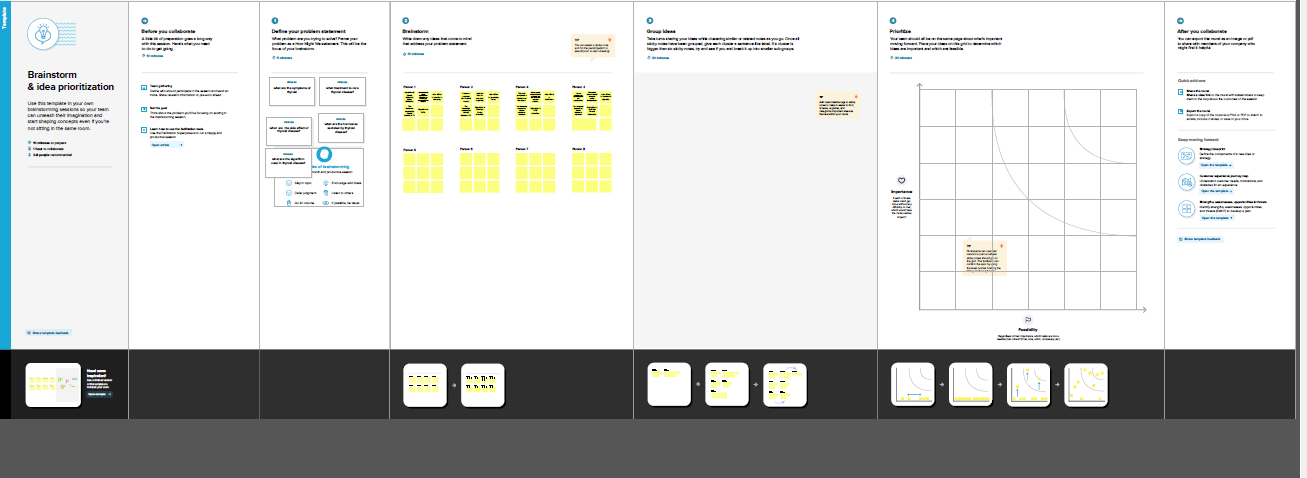
* We are building a flask application which needs HTML pages stored in templates folder and a python script app.py for scripting.
* thyroid\_1\_model.pk1 is our saved model. Further we will use this model for flask integration.
* Training folder contains model training files and the training\_ibm folder contains IBM deployment files

## PROBLEM DEFINITION & DESIGN THINKING

* EMPATHY MAP

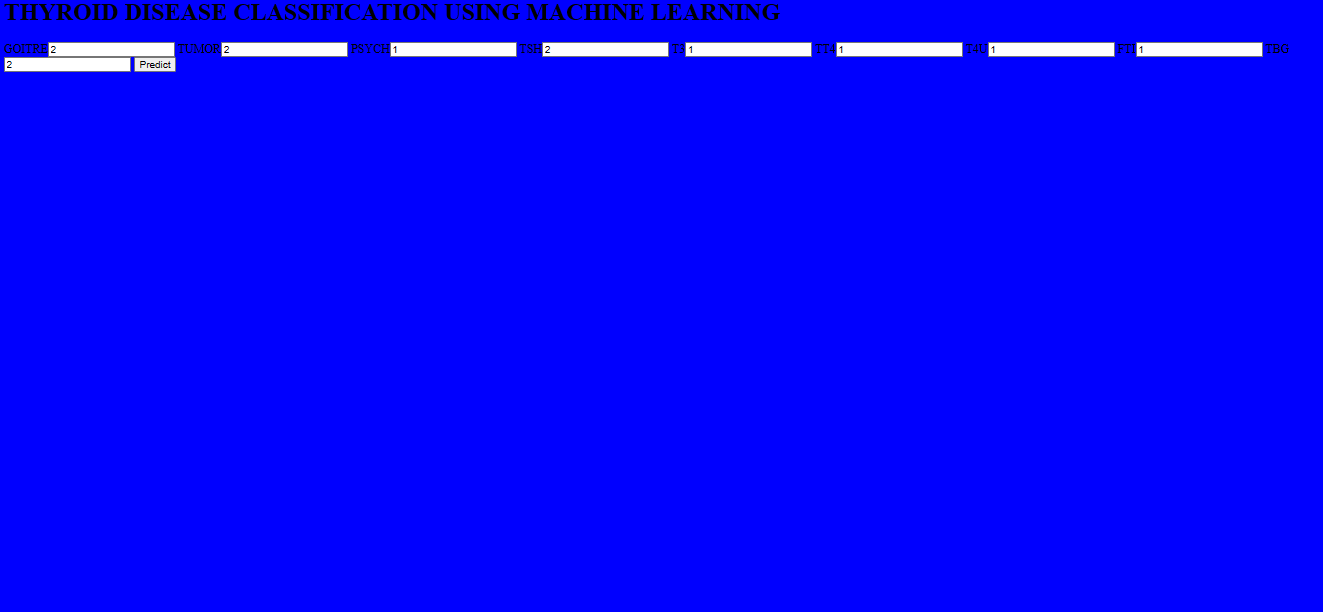
 

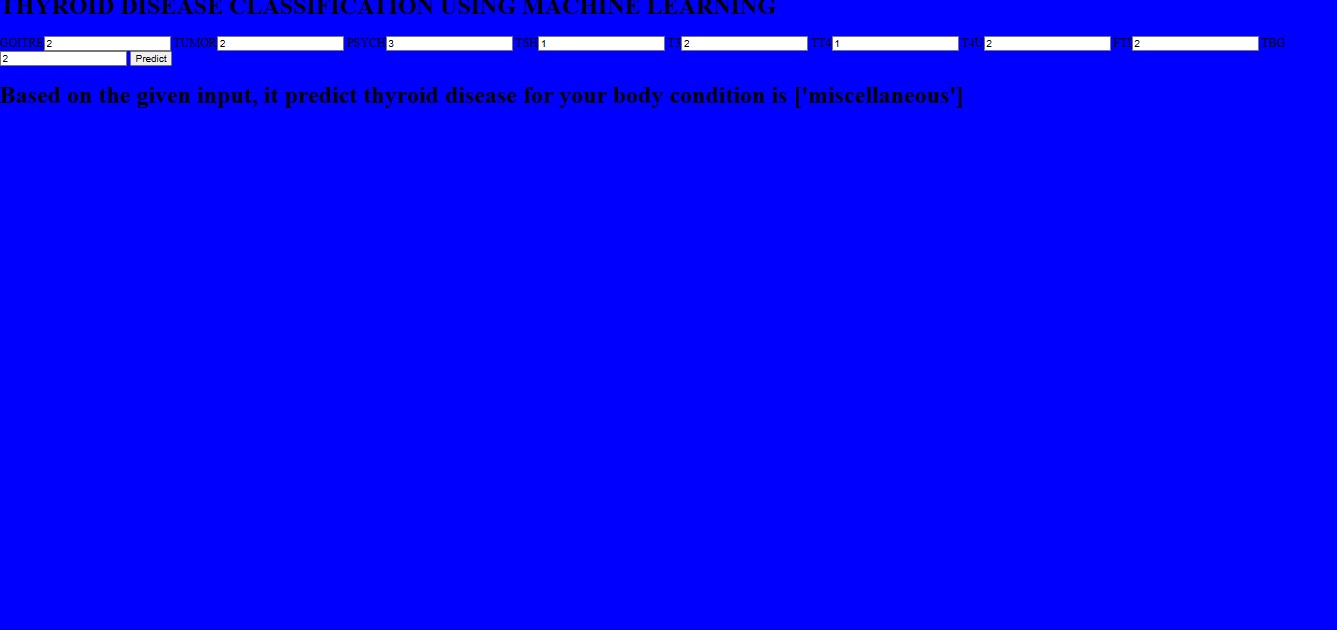
* BRAINSTORMING MAP



## RESULT

app.html





## 

## ADVANTAGES

1. Machine learning can be very effective in predicting and classifying the thyroid diseases in earlier as well later stages.
2. Machine learning methodology to classify thyroid disease is so simple but very clear to mark the differences in ranges.
3. Thyroid disease classification is done with the machine learning methodology in short span of time as well as less cost of spending.
4. Thyroid disease classification can be done easily with the help of machine learning in all age groups.
5. By using Machine learning accurate classification of thyroid disease can be using emerging and innovative methods

# **DISADVANTAGES**

1. In thyroid disease classification machine learning can be used. but it is a newly growing methodology.
2. There is no empirical evidences in classifying thyroid disease in accurate levels.
3. Error levels cannot be predicted in this machine learning methods.
4. Machine learning is newly developing concept. It is effectiveness will only come through after various research methods.
5. Emerging machine learning concepts needs various skilled and qualified human resourses in thyroid disease classification.

## APPLICATION

* Machine Learning has different algorithm that applies to classify

the thyroid disease for the effective treatment

* Machine Learning application of decision tree method helps to

classify the thyroid disease in wast variety of datas like tree to indentify

and apply treatment acconding to the classification

* Machine Learning Random forest algorithm places data sampling

and

## CONCLUSION

Thyroid disease is one of the disease that afflict the world’s Population, and the number of cases of this disease is increasing .Because of medical reports that show series imbalance in thyroid disease ,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-querry\_thyroxin 2-querry\_hypothyroid 3-querry\_hyperthyroid.Here we have included the increased accuracy of some algorithms , as well as the Retension of the accuracy of others .It was observed that the accuracy of Naïve Bayes algorithms increased the accuracy by 90.67.The highest precision of the MLP algorithm was 96.4 accuracy.

## FUTURE SCOPE

There are many thyroid disease affecting people all over the world. Many disease affect the thyroid gland, like hyperthyroidism, hypothyroidism, and thyroid cancer

.

Thyroid disease classification in the future, to better generalize our finding it is necessary to further expand the set of the data and attributes considered. With more data the training process is likely to produce more effective classifiers also allowing a more reliable estimate of the exhibited performance.

Finally, another aspect that could be investigated concerns the presence of any secondary thyroid disease linked to the patient, to understand if there is a particular additional thyroid disease that can affect hypothyroidism. In fact, it often happens that patients are suffering from more than one thyroid disease at the same time

## APPENDEX

Thyroid disease.ipynb

{

"cells": [

{

"cell\_type": "markdown",

"source": [

"# Data Collection & Preparation"

],

"metadata": {

"id": "6FQ\_QPmYkdXp"

}

},

{

"cell\_type": "code",

"execution\_count": null,

"metadata": {

"id": "pm-0QxL3kIkn"

},

"outputs": [],

"source": [

"import pandas as pd\n",

"import numpy as np\n",

"import matplotlib.pyplot as plt\n"

]

},

{

"cell\_type": "code",

"execution\_count": null,

"metadata": {

"colab": {

"base\_uri": "https://localhost:8080/"

},

"id": "89-WfRTyk2r9",

"outputId": "5b071d13-16ce-478c-f7ef-6030a7fea1de"

},

"outputs": [

{

"output\_type": "stream",

"name": "stdout",

"text": [

"Mounted at /content/drive\n"

]

}

],

"source": [

"from google.colab import drive\n",

"drive.mount(\"/content/drive\")"

]

},

{

"cell\_type": "code",

"execution\_count": null,

"metadata": {

"id": "uadDruTVl\_o0"

},

"outputs": [],

"source": [

"data = pd.read\_csv(\"/content/thyroidDF.csv\")"

]

},

{

"cell\_type": "code",

"execution\_count": null,

"metadata": {

"colab": {

"base\_uri": "https://localhost:8080/",

"height": 300

},

"id": "8jQla9Q9mPJg",

"outputId": "9670321b-2f60-4856-9016-38e529cf8e45"

},

"outputs": [

{

"output\_type": "execute\_result",

"data": {

"text/plain": [

" age sex on\_thyroxine query\_on\_thyroxine on\_antithyroid\_meds sick pregnant \\\n",

"0 29 F f f f f f \n",

"1 29 F f f f f f \n",

"2 41 F f f f f f \n",

"3 36 F f f f f f \n",

"4 32 F f f f f f \n",

"\n",

" thyroid\_surgery I131\_treatment query\_hypothyroid ... TT4 T4U\_measured \\\n",

"0 f f t ... NaN f \n",

"1 f f f ... 128.0 f \n",

"2 f f f ... NaN f \n",

"3 f f f ... NaN f \n",

"4 f f f ... NaN f \n",

"\n",

" T4U FTI\_measured FTI TBG\_measured TBG referral\_source target patient\_id \n",

"0 NaN f NaN f NaN other - 840801013 \n",

"1 NaN f NaN f NaN other - 840801014 \n",

"2 NaN f NaN t 11.0 other - 840801042 \n",

"3 NaN f NaN t 26.0 other - 840803046 \n",

"4 NaN f NaN t 36.0 other S 840803047 \n",

"\n",

"[5 rows x 31 columns]"

],

"text/html": [

"\n",

" <div id=\"df-d822ef1b-668e-4f67-91e5-3cc9d3bbadc7\">\n",

" <div class=\"colab-df-container\">\n",

" <div>\n",

"<style scoped>\n",

" .dataframe tbody tr th:only-of-type {\n",

" vertical-align: middle;\n",

" }\n",

"\n",

" .dataframe tbody tr th {\n",

" vertical-align: top;\n",

" }\n",

"\n",

" .dataframe thead th {\n",

" text-align: right;\n",

" }\n",

"</style>\n",

"<table border=\"1\" class=\"dataframe\">\n",

" <thead>\n",

" <tr style=\"text-align: right;\">\n",

" <th></th>\n",

" <th>age</th>\n",

" <th>sex</th>\n",

" <th>on\_thyroxine</th>\n",

" <th>query\_on\_thyroxine</th>\n",

" <th>on\_antithyroid\_meds</th>\n",

" <th>sick</th>\n",

" <th>pregnant</th>\n",

" <th>thyroid\_surgery</th>\n",

" <th>I131\_treatment</th>\n",

" <th>query\_hypothyroid</th>\n",

" <th>...</th>\n",

" <th>TT4</th>\n",

" <th>T4U\_measured</th>\n",

" <th>T4U</th>\n",

" <th>FTI\_measured</th>\n",

" <th>FTI</th>\n",

" <th>TBG\_measured</th>\n",

" <th>TBG</th>\n",

" <th>referral\_source</th>\n",

" <th>target</th>\n",

" <th>patient\_id</th>\n",

" </tr>\n",

" </thead>\n",

" <tbody>\n",

" <tr>\n",

" <th>0</th>\n",

" <td>29</td>\n",

" <td>F</td>\n",

" <td>f</td>\n",

" <td>f</td>\n",

" <td>f</td>\n",

" <td>f</td>\n",

" <td>f</td>\n",

" <td>f</td>\n",

" <td>f</td>\n",

" <td>t</td>\n",

" <td>...</td>\n",

" <td>NaN</td>\n",

" <td>f</td>\n",

" <td>NaN</td>\n",

" <td>f</td>\n",

" <td>NaN</td>\n",

" <td>f</td>\n",

" <td>NaN</td>\n",

" <td>other</td>\n",

" <td>-</td>\n",

" <td>840801013</td>\n",

" </tr>\n",

" <tr>\n",

" <th>1</th>\n",

" <td>29</td>\n",

" <td>F</td>\n",

" <td>f</td>\n",

" <td>f</td>\n",

" <td>f</td>\n",

" <td>f</td>\n",

" <td>f</td>\n",

" <td>f</td>\n",

" <td>f</td>\n",

" <td>f</td>\n",

" <td>...</td>\n",

" <td>128.0</td>\n",

" <td>f</td>\n",

" <td>NaN</td>\n",

" <td>f</td>\n",

" <td>NaN</td>\n",

" <td>f</td>\n",

" <td>NaN</td>\n",

" <td>other</td>\n",

" <td>-</td>\n",

" <td>840801014</td>\n",

" </tr>\n",

" <tr>\n",

" <th>2</th>\n",

" <td>41</td>\n",

" <td>F</td>\n",

" <td>f</td>\n",

" <td>f</td>\n",

" <td>f</td>\n",

" <td>f</td>\n",

" <td>f</td>\n",

" <td>f</td>\n",

" <td>f</td>\n",

" <td>f</td>\n",

" <td>...</td>\n",

" <td>NaN</td>\n",

" <td>f</td>\n",

" <td>NaN</td>\n",

" <td>f</td>\n",

" <td>NaN</td>\n",

" <td>t</td>\n",

" <td>11.0</td>\n",

" <td>other</td>\n",

" <td>-</td>\n",

" <td>840801042</td>\n",

" </tr>\n",

" <tr>\n",

" <th>3</th>\n",

" <td>36</td>\n",

" <td>F</td>\n",

" <td>f</td>\n",

" <td>f</td>\n",

" <td>f</td>\n",

" <td>f</td>\n",

" <td>f</td>\n",

" <td>f</td>\n",

" <td>f</td>\n",

" <td>f</td>\n",

" <td>...</td>\n",

" <td>NaN</td>\n",

" <td>f</td>\n",

" <td>NaN</td>\n",

" <td>f</td>\n",

" <td>NaN</td>\n",

" <td>t</td>\n",

" <td>26.0</td>\n",

" <td>other</td>\n",

" <td>-</td>\n",

" <td>840803046</td>\n",

" </tr>\n",

" <tr>\n",

" <th>4</th>\n",

" <td>32</td>\n",

" <td>F</td>\n",

" <td>f</td>\n",

" <td>f</td>\n",

" <td>f</td>\n",

" <td>f</td>\n",

" <td>f</td>\n",

" <td>f</td>\n",

" <td>f</td>\n",

" <td>f</td>\n",

" <td>...</td>\n",

" <td>NaN</td>\n",

" <td>f</td>\n",

" <td>NaN</td>\n",

" <td>f</td>\n",

" <td>NaN</td>\n",

" <td>t</td>\n",

" <td>36.0</td>\n",

" <td>other</td>\n",

" <td>S</td>\n",

" <td>840803047</td>\n",

" </tr>\n",

" </tbody>\n",

"</table>\n",

"<p>5 rows × 31 columns</p>\n",

"</div>\n",

" <button class=\"colab-df-convert\" onclick=\"convertToInteractive('df-d822ef1b-668e-4f67-91e5-3cc9d3bbadc7')\"\n",

" title=\"Convert this dataframe to an interactive table.\"\n",

" style=\"display:none;\">\n",

" \n",

" <svg xmlns=\"http://www.w3.org/2000/svg\" height=\"24px\"viewBox=\"0 0 24 24\"\n",

" width=\"24px\">\n",

" <path d=\"M0 0h24v24H0V0z\" fill=\"none\"/>\n",

" <path d=\"M18.56 5.44l.94 2.06.94-2.06 2.06-.94-2.06-.94-.94-2.06-.94 2.06-2.06.94zm-11 1L8.5 8.5l.94-2.06 2.06-.94-2.06-.94L8.5 2.5l-.94 2.06-2.06.94zm10 10l.94 2.06.94-2.06 2.06-.94-2.06-.94-.94-2.06-.94 2.06-2.06.94z\"/><path d=\"M17.41 7.96l-1.37-1.37c-.4-.4-.92-.59-1.43-.59-.52 0-1.04.2-1.43.59L10.3 9.45l-7.72 7.72c-.78.78-.78 2.05 0 2.83L4 21.41c.39.39.9.59 1.41.59.51 0 1.02-.2 1.41-.59l7.78-7.78 2.81-2.81c.8-.78.8-2.07 0-2.86zM5.41 20L4 18.59l7.72-7.72 1.47 1.35L5.41 20z\"/>\n",

" </svg>\n",

" </button>\n",

" \n",

" <style>\n",

" .colab-df-container {\n",

" display:flex;\n",

" flex-wrap:wrap;\n",

" gap: 12px;\n",

" }\n",

"\n",

" .colab-df-convert {\n",

" background-color: #E8F0FE;\n",

" border: none;\n",

" border-radius: 50%;\n",

" cursor: pointer;\n",

" display: none;\n",

" fill: #1967D2;\n",

" height: 32px;\n",

" padding: 0 0 0 0;\n",

" width: 32px;\n",

" }\n",

"\n",

" .colab-df-convert:hover {\n",

" background-color: #E2EBFA;\n",

" box-shadow: 0px 1px 2px rgba(60, 64, 67, 0.3), 0px 1px 3px 1px rgba(60, 64, 67, 0.15);\n",

" fill: #174EA6;\n",

" }\n",

"\n",

" [theme=dark] .colab-df-convert {\n",

" background-color: #3B4455;\n",

" fill: #D2E3FC;\n",

" }\n",

"\n",

" [theme=dark] .colab-df-convert:hover {\n",

" background-color: #434B5C;\n",

" box-shadow: 0px 1px 3px 1px rgba(0, 0, 0, 0.15);\n",

" filter: drop-shadow(0px 1px 2px rgba(0, 0, 0, 0.3));\n",

" fill: #FFFFFF;\n",

" }\n",

" </style>\n",

"\n",

" <script>\n",

" const buttonEl =\n",

" document.querySelector('#df-d822ef1b-668e-4f67-91e5-3cc9d3bbadc7 button.colab-df-convert');\n",

" buttonEl.style.display =\n",

" google.colab.kernel.accessAllowed ? 'block' : 'none';\n",

"\n",

" async function convertToInteractive(key) {\n",

" const element = document.querySelector('#df-d822ef1b-668e-4f67-91e5-3cc9d3bbadc7');\n",

" const dataTable =\n",

" await google.colab.kernel.invokeFunction('convertToInteractive',\n",

" [key], {});\n",

" if (!dataTable) return;\n",

"\n",

" const docLinkHtml = 'Like what you see? Visit the ' +\n",

" '<a target=\"\_blank\" href=https://colab.research.google.com/notebooks/data\_table.ipynb>data table notebook</a>'\n",

" + ' to learn more about interactive tables.';\n",

" element.innerHTML = '';\n",

" dataTable['output\_type'] = 'display\_data';\n",

" await google.colab.output.renderOutput(dataTable, element);\n",

" const docLink = document.createElement('div');\n",

" docLink.innerHTML = docLinkHtml;\n",

" element.appendChild(docLink);\n",

" }\n",

" </script>\n",

" </div>\n",

" </div>\n",

" "

]

},

"metadata": {},

"execution\_count": 9

}

],

"source": [

"data.head()"

]

},

{

"cell\_type": "code",

"execution\_count": null,

"metadata": {

"colab": {

"base\_uri": "https://localhost:8080/"

},

"id": "ImDazt95mYlr",

"outputId": "d877c136-4752-440c-f689-900df1213df9"

},

"outputs": [

{

"output\_type": "execute\_result",

"data": {

"text/plain": [

"array(['-', 'S', 'F', 'AK', 'R', 'I', 'M', 'N', 'G', 'K', 'A', 'KJ', 'L',\n",

" 'MK', 'Q', 'J', 'C|I', 'O', 'LJ', 'H|K', 'D', 'GK', 'MI', 'P',\n",

" 'FK', 'B', 'GI', 'C', 'GKJ', 'OI', 'D|R', 'E'], dtype=object)"

]

},

"metadata": {},

"execution\_count": 10

}

],

"source": [

"data['target'].unique()"

]

},

{

"cell\_type": "code",

"execution\_count": null,

"metadata": {

"colab": {

"base\_uri": "https://localhost:8080/"

},

"id": "RzoI2VylmsvX",

"outputId": "666f9592-be5e-443e-8bbc-f2eca3c5bd11"

},

"outputs": [

{

"output\_type": "execute\_result",

"data": {

"text/plain": [

"(9172, 31)"

]

},

"metadata": {},

"execution\_count": 11

}

],

"source": [

"data.shape"

]

},

{

"cell\_type": "code",

"execution\_count": null,

"metadata": {

"colab": {

"base\_uri": "https://localhost:8080/"

},

"id": "SPFND-8Tm4BC",

"outputId": "99b85558-002c-441f-d549-82496365871a"

},

"outputs": [

{

"output\_type": "stream",

"name": "stdout",

"text": [

"<class 'pandas.core.frame.DataFrame'>\n",

"RangeIndex: 9172 entries, 0 to 9171\n",

"Data columns (total 31 columns):\n",

" # Column Non-Null Count Dtype \n",

"--- ------ -------------- ----- \n",

" 0 age 9172 non-null int64 \n",

" 1 sex 8865 non-null object \n",

" 2 on\_thyroxine 9172 non-null object \n",

" 3 query\_on\_thyroxine 9172 non-null object \n",

" 4 on\_antithyroid\_meds 9172 non-null object \n",

" 5 sick 9172 non-null object \n",

" 6 pregnant 9172 non-null object \n",

" 7 thyroid\_surgery 9172 non-null object \n",

" 8 I131\_treatment 9172 non-null object \n",

" 9 query\_hypothyroid 9172 non-null object \n",

" 10 query\_hyperthyroid 9172 non-null object \n",

" 11 lithium 9172 non-null object \n",

" 12 goitre 9172 non-null object \n",

" 13 tumor 9172 non-null object \n",

" 14 hypopituitary 9172 non-null object \n",

" 15 psych 9172 non-null object \n",

" 16 TSH\_measured 9172 non-null object \n",

" 17 TSH 8330 non-null float64\n",

" 18 T3\_measured 9172 non-null object \n",

" 19 T3 6568 non-null float64\n",

" 20 TT4\_measured 9172 non-null object \n",

" 21 TT4 8730 non-null float64\n",

" 22 T4U\_measured 9172 non-null object \n",

" 23 T4U 8363 non-null float64\n",

" 24 FTI\_measured 9172 non-null object \n",

" 25 FTI 8370 non-null float64\n",

" 26 TBG\_measured 9172 non-null object \n",

" 27 TBG 349 non-null float64\n",

" 28 referral\_source 9172 non-null object \n",

" 29 target 9172 non-null object \n",

" 30 patient\_id 9172 non-null int64 \n",

"dtypes: float64(6), int64(2), object(23)\n",

"memory usage: 2.2+ MB\n"

]

}

],

"source": [

"data.info()"

]

},

{

"cell\_type": "code",

"execution\_count": null,

"metadata": {

"colab": {

"base\_uri": "https://localhost:8080/"

},

"id": "\_ona2RC6nQmP",

"outputId": "6284b5b0-da43-4f1c-8e1f-6588831234a8"

},

"outputs": [

{

"output\_type": "execute\_result",

"data": {

"text/plain": [

"age 0\n",

"sex 307\n",

"on\_thyroxine 0\n",

"query\_on\_thyroxine 0\n",

"on\_antithyroid\_meds 0\n",

"sick 0\n",

"pregnant 0\n",

"thyroid\_surgery 0\n",

"I131\_treatment 0\n",

"query\_hypothyroid 0\n",

"query\_hyperthyroid 0\n",

"lithium 0\n",

"goitre 0\n",

"tumor 0\n",

"hypopituitary 0\n",

"psych 0\n",

"TSH\_measured 0\n",

"TSH 842\n",

"T3\_measured 0\n",

"T3 2604\n",

"TT4\_measured 0\n",

"TT4 442\n",

"T4U\_measured 0\n",

"T4U 809\n",

"FTI\_measured 0\n",

"FTI 802\n",

"TBG\_measured 0\n",

"TBG 8823\n",

"referral\_source 0\n",

"target 0\n",

"patient\_id 0\n",

"dtype: int64"

]

},

"metadata": {},

"execution\_count": 13

}

],

"source": [

"data.isnull().sum()"

]

},

{

"cell\_type": "code",

"execution\_count": null,

"metadata": {

"id": "Btl7v52gnapL"

},

"outputs": [],

"source": [

"import pandas as pd\n",

"data = pd.read\_csv(\"/content/thyroidDF.csv\")\n",

"data.drop(['TSH\_measured','T3\_measured','TT4\_measured','T4U\_measured','FTI\_measured','TBG\_measured','referral\_source','patient\_id','T4U\_measured'],axis=1,inplace=True)"

]

},

{

"cell\_type": "code",

"execution\_count": null,

"metadata": {

"colab": {

"base\_uri": "https://localhost:8080/",

"height": 300

},

"id": "49jRB4gSpHfu",

"outputId": "f3bf2ceb-cde4-422e-9c46-c7bd4e54fd02"

},

"outputs": [

{

"output\_type": "execute\_result",

"data": {

"text/plain": [

" age sex on\_thyroxine query\_on\_thyroxine on\_antithyroid\_meds sick pregnant \\\n",

"0 29 F f f f f f \n",

"1 29 F f f f f f \n",

"2 41 F f f f f f \n",

"3 36 F f f f f f \n",

"4 32 F f f f f f \n",

"\n",

" thyroid\_surgery I131\_treatment query\_hypothyroid ... tumor hypopituitary \\\n",

"0 f f t ... f f \n",

"1 f f f ... f f \n",

"2 f f f ... f f \n",

"3 f f f ... f f \n",

"4 f f f ... f f \n",

"\n",

" psych TSH T3 TT4 T4U FTI TBG target \n",

"0 f 0.3 NaN NaN NaN NaN NaN - \n",

"1 f 1.6 1.9 128.0 NaN NaN NaN - \n",

"2 f NaN NaN NaN NaN NaN 11.0 - \n",

"3 f NaN NaN NaN NaN NaN 26.0 - \n",

"4 f NaN NaN NaN NaN NaN 36.0 S \n",

"\n",

"[5 rows x 23 columns]"

],

"text/html": [

"\n",

" <div id=\"df-ead1086d-7e7d-4949-8ae4-94a17210e763\">\n",

" <div class=\"colab-df-container\">\n",

" <div>\n",

"<style scoped>\n",

" .dataframe tbody tr th:only-of-type {\n",

" vertical-align: middle;\n",

" }\n",

"\n",

" .dataframe tbody tr th {\n",

" vertical-align: top;\n",

" }\n",

"\n",

" .dataframe thead th {\n",

" text-align: right;\n",

" }\n",

"</style>\n",

"<table border=\"1\" class=\"dataframe\">\n",

" <thead>\n",

" <tr style=\"text-align: right;\">\n",

" <th></th>\n",

" <th>age</th>\n",

" <th>sex</th>\n",

" <th>on\_thyroxine</th>\n",

" <th>query\_on\_thyroxine</th>\n",

" <th>on\_antithyroid\_meds</th>\n",

" <th>sick</th>\n",

" <th>pregnant</th>\n",

" <th>thyroid\_surgery</th>\n",

" <th>I131\_treatment</th>\n",

" <th>query\_hypothyroid</th>\n",

" <th>...</th>\n",

" <th>tumor</th>\n",

" <th>hypopituitary</th>\n",

" <th>psych</th>\n",

" <th>TSH</th>\n",

" <th>T3</th>\n",

" <th>TT4</th>\n",

" <th>T4U</th>\n",

" <th>FTI</th>\n",

" <th>TBG</th>\n",

" <th>target</th>\n",

" </tr>\n",

" </thead>\n",

" <tbody>\n",

" <tr>\n",

" <th>0</th>\n",

" <td>29</td>\n",

" <td>F</td>\n",

" <td>f</td>\n",

" <td>f</td>\n",

" <td>f</td>\n",

" <td>f</td>\n",

" <td>f</td>\n",

" <td>f</td>\n",

" <td>f</td>\n",

" <td>t</td>\n",

" <td>...</td>\n",

" <td>f</td>\n",

" <td>f</td>\n",

" <td>f</td>\n",

" <td>0.3</td>\n",

" <td>NaN</td>\n",

" <td>NaN</td>\n",

" <td>NaN</td>\n",

" <td>NaN</td>\n",

" <td>NaN</td>\n",

" <td>-</td>\n",

" </tr>\n",

" <tr>\n",

" <th>1</th>\n",

" <td>29</td>\n",

" <td>F</td>\n",

" <td>f</td>\n",

" <td>f</td>\n",

" <td>f</td>\n",

" <td>f</td>\n",

" <td>f</td>\n",

" <td>f</td>\n",

" <td>f</td>\n",

" <td>f</td>\n",

" <td>...</td>\n",

" <td>f</td>\n",

" <td>f</td>\n",

" <td>f</td>\n",

" <td>1.6</td>\n",

" <td>1.9</td>\n",

" <td>128.0</td>\n",

" <td>NaN</td>\n",

" <td>NaN</td>\n",

" <td>NaN</td>\n",

" <td>-</td>\n",

" </tr>\n",

" <tr>\n",

" <th>2</th>\n",

" <td>41</td>\n",

" <td>F</td>\n",

" <td>f</td>\n",

" <td>f</td>\n",

" <td>f</td>\n",

" <td>f</td>\n",

" <td>f</td>\n",

" <td>f</td>\n",

" <td>f</td>\n",

" <td>f</td>\n",

" <td>...</td>\n",

" <td>f</td>\n",

" <td>f</td>\n",

" <td>f</td>\n",

" <td>NaN</td>\n",

" <td>NaN</td>\n",

" <td>NaN</td>\n",

" <td>NaN</td>\n",

" <td>NaN</td>\n",

" <td>11.0</td>\n",

" <td>-</td>\n",

" </tr>\n",

" <tr>\n",

" <th>3</th>\n",

" <td>36</td>\n",

" <td>F</td>\n",

" <td>f</td>\n",

" <td>f</td>\n",

" <td>f</td>\n",

" <td>f</td>\n",

" <td>f</td>\n",

" <td>f</td>\n",

" <td>f</td>\n",

" <td>f</td>\n",

" <td>...</td>\n",

" <td>f</td>\n",

" <td>f</td>\n",

" <td>f</td>\n",

" <td>NaN</td>\n",

" <td>NaN</td>\n",

" <td>NaN</td>\n",

" <td>NaN</td>\n",

" <td>NaN</td>\n",

" <td>26.0</td>\n",

" <td>-</td>\n",

" </tr>\n",

" <tr>\n",

" <th>4</th>\n",

" <td>32</td>\n",

" <td>F</td>\n",

" <td>f</td>\n",

" <td>f</td>\n",

" <td>f</td>\n",

" <td>f</td>\n",

" <td>f</td>\n",

" <td>f</td>\n",

" <td>f</td>\n",

" <td>f</td>\n",

" <td>...</td>\n",

" <td>f</td>\n",

" <td>f</td>\n",

" <td>f</td>\n",

" <td>NaN</td>\n",

" <td>NaN</td>\n",

" <td>NaN</td>\n",

" <td>NaN</td>\n",

" <td>NaN</td>\n",

" <td>36.0</td>\n",

" <td>S</td>\n",

" </tr>\n",

" </tbody>\n",

"</table>\n",

"<p>5 rows × 23 columns</p>\n",

"</div>\n",

" <button class=\"colab-df-convert\" onclick=\"convertToInteractive('df-ead1086d-7e7d-4949-8ae4-94a17210e763')\"\n",

" title=\"Convert this dataframe to an interactive table.\"\n",

" style=\"display:none;\">\n",

" \n",

" <svg xmlns=\"http://www.w3.org/2000/svg\" height=\"24px\"viewBox=\"0 0 24 24\"\n",

" width=\"24px\">\n",

" <path d=\"M0 0h24v24H0V0z\" fill=\"none\"/>\n",

" <path d=\"M18.56 5.44l.94 2.06.94-2.06 2.06-.94-2.06-.94-.94-2.06-.94 2.06-2.06.94zm-11 1L8.5 8.5l.94-2.06 2.06-.94-2.06-.94L8.5 2.5l-.94 2.06-2.06.94zm10 10l.94 2.06.94-2.06 2.06-.94-2.06-.94-.94-2.06-.94 2.06-2.06.94z\"/><path d=\"M17.41 7.96l-1.37-1.37c-.4-.4-.92-.59-1.43-.59-.52 0-1.04.2-1.43.59L10.3 9.45l-7.72 7.72c-.78.78-.78 2.05 0 2.83L4 21.41c.39.39.9.59 1.41.59.51 0 1.02-.2 1.41-.59l7.78-7.78 2.81-2.81c.8-.78.8-2.07 0-2.86zM5.41 20L4 18.59l7.72-7.72 1.47 1.35L5.41 20z\"/>\n",

" </svg>\n",

" </button>\n",

" \n",

" <style>\n",

" .colab-df-container {\n",

" display:flex;\n",

" flex-wrap:wrap;\n",

" gap: 12px;\n",

" }\n",

"\n",

" .colab-df-convert {\n",

" background-color: #E8F0FE;\n",

" border: none;\n",

" border-radius: 50%;\n",

" cursor: pointer;\n",

" display: none;\n",

" fill: #1967D2;\n",

" height: 32px;\n",

" padding: 0 0 0 0;\n",

" width: 32px;\n",

" }\n",

"\n",

" .colab-df-convert:hover {\n",

" background-color: #E2EBFA;\n",

" box-shadow: 0px 1px 2px rgba(60, 64, 67, 0.3), 0px 1px 3px 1px rgba(60, 64, 67, 0.15);\n",

" fill: #174EA6;\n",

" }\n",

"\n",

" [theme=dark] .colab-df-convert {\n",

" background-color: #3B4455;\n",

" fill: #D2E3FC;\n",

" }\n",

"\n",

" [theme=dark] .colab-df-convert:hover {\n",

" background-color: #434B5C;\n",

" box-shadow: 0px 1px 3px 1px rgba(0, 0, 0, 0.15);\n",

" filter: drop-shadow(0px 1px 2px rgba(0, 0, 0, 0.3));\n",

" fill: #FFFFFF;\n",

" }\n",

" </style>\n",

"\n",

" <script>\n",

" const buttonEl =\n",

" document.querySelector('#df-ead1086d-7e7d-4949-8ae4-94a17210e763 button.colab-df-convert');\n",

" buttonEl.style.display =\n",

" google.colab.kernel.accessAllowed ? 'block' : 'none';\n",

"\n",

" async function convertToInteractive(key) {\n",

" const element = document.querySelector('#df-ead1086d-7e7d-4949-8ae4-94a17210e763');\n",

" const dataTable =\n",

" await google.colab.kernel.invokeFunction('convertToInteractive',\n",

" [key], {});\n",

" if (!dataTable) return;\n",

"\n",

" const docLinkHtml = 'Like what you see? Visit the ' +\n",

" '<a target=\"\_blank\" href=https://colab.research.google.com/notebooks/data\_table.ipynb>data table notebook</a>'\n",

" + ' to learn more about interactive tables.';\n",

" element.innerHTML = '';\n",

" dataTable['output\_type'] = 'display\_data';\n",

" await google.colab.output.renderOutput(dataTable, element);\n",

" const docLink = document.createElement('div');\n",

" docLink.innerHTML = docLinkHtml;\n",

" element.appendChild(docLink);\n",

" }\n",

" </script>\n",

" </div>\n",

" </div>\n",

" "

]

},

"metadata": {},

"execution\_count": 15

}

],

"source": [

"data.head()"

]

},

{

"cell\_type": "code",

"execution\_count": null,

"metadata": {

"colab": {

"base\_uri": "https://localhost:8080/"

},

"id": "LwJeXeLfqiFU",

"outputId": "c3b628fa-bb19-4ae3-eba8-69e8c4e8776a"

},

"outputs": [

{

"output\_type": "execute\_result",

"data": {

"text/plain": [

"0 -\n",

"1 -\n",

"2 -\n",

"3 -\n",

"4 S\n",

" ..\n",

"9167 -\n",

"9168 -\n",

"9169 I\n",

"9170 -\n",

"9171 -\n",

"Name: target, Length: 9172, dtype: object"

]

},

"metadata": {},

"execution\_count": 16

}

],

"source": [

"data['target']"

]

},

{

"cell\_type": "code",

"execution\_count": null,

"metadata": {

"id": "TAaCExaGqrIt"

},

"outputs": [],

"source": [

"diagnoses = {'A': 'hyperthyroid conditions',\n",

" 'B': 'hyperthyroid conditions',\n",

" 'C': 'hyperthyroid conditions',\n",

" 'D': 'hyperthyroid conditions',\n",

" 'E': 'hypothyroid conditions',\n",

" 'F': 'hypothyroid conditions',\n",

" 'G': 'hypothyroid conditions',\n",

" 'H': 'hypothyroid conditions',\n",

" 'I': 'binding protein',\n",

" 'J': 'binding protein',\n",

" 'K': 'general health',\n",

" 'L': 'replacement therapy',\n",

" 'M': 'replacement therapy',\n",

" 'N': 'replacement therapy',\n",

" '0': 'antithyroid treatment',\n",

" 'P': 'antithyroid treatment',\n",

" 'Q': 'antithyroid treatment',\n",

" 'R': 'miscellaneous',\n",

" 'S': 'miscellaneous',\n",

" 'T': 'miscellaneous'}\n",

"data['target']=data['target'].map(diagnoses)"

]

},

{

"cell\_type": "code",

"execution\_count": null,

"metadata": {

"colab": {

"base\_uri": "https://localhost:8080/",

"height": 505

},

"id": "KOI5QH1Dr8lt",

"outputId": "5c240b09-6ff2-4b75-df94-9d02f9c244ac"

},

"outputs": [

{

"output\_type": "execute\_result",

"data": {

"text/plain": [

" age sex on\_thyroxine query\_on\_thyroxine on\_antithyroid\_meds sick \\\n",

"0 29 F f f f f \n",

"1 29 F f f f f \n",

"2 41 F f f f f \n",

"3 36 F f f f f \n",

"4 32 F f f f f \n",

"... ... .. ... ... ... ... \n",

"9167 56 M f f f f \n",

"9168 22 M f f f f \n",

"9169 69 M f f f f \n",

"9170 47 F f f f f \n",

"9171 31 M f f f f \n",

"\n",

" pregnant thyroid\_surgery I131\_treatment query\_hypothyroid ... tumor \\\n",

"0 f f f t ... f \n",

"1 f f f f ... f \n",

"2 f f f f ... f \n",

"3 f f f f ... f \n",

"4 f f f f ... f \n",

"... ... ... ... ... ... ... \n",

"9167 f f f f ... f \n",

"9168 f f f f ... f \n",

"9169 f f f f ... f \n",

"9170 f f f f ... f \n",

"9171 f f f t ... f \n",

"\n",

" hypopituitary psych TSH T3 TT4 T4U FTI TBG target \n",

"0 f f 0.3 NaN NaN NaN NaN NaN NaN \n",

"1 f f 1.6 1.9 128.0 NaN NaN NaN NaN \n",

"2 f f NaN NaN NaN NaN NaN 11.0 NaN \n",

"3 f f NaN NaN NaN NaN NaN 26.0 NaN \n",

"4 f f NaN NaN NaN NaN NaN 36.0 miscellaneous \n",

"... ... ... ... ... ... ... ... ... ... \n",

"9167 f f NaN NaN 64.0 0.83 77.0 NaN NaN \n",

"9168 f f NaN NaN 91.0 0.92 99.0 NaN NaN \n",

"9169 f f NaN NaN 113.0 1.27 89.0 NaN binding protein \n",

"9170 f f NaN NaN 75.0 0.85 88.0 NaN NaN \n",

"9171 f f NaN NaN 66.0 1.02 65.0 NaN NaN \n",

"\n",

"[9172 rows x 23 columns]"

],

"text/html": [

"\n",

" <div id=\"df-8ca7cc93-51c6-4956-8e3b-2cc610409600\">\n",

" <div class=\"colab-df-container\">\n",

" <div>\n",

"<style scoped>\n",

" .dataframe tbody tr th:only-of-type {\n",

" vertical-align: middle;\n",

" }\n",

"\n",

" .dataframe tbody tr th {\n",

" vertical-align: top;\n",

" }\n",

"\n",

" .dataframe thead th {\n",

" text-align: right;\n",

" }\n",

"</style>\n",

"<table border=\"1\" class=\"dataframe\">\n",

" <thead>\n",

" <tr style=\"text-align: right;\">\n",

" <th></th>\n",

" <th>age</th>\n",

" <th>sex</th>\n",

" <th>on\_thyroxine</th>\n",

" <th>query\_on\_thyroxine</th>\n",

" <th>on\_antithyroid\_meds</th>\n",

" <th>sick</th>\n",

" <th>pregnant</th>\n",

" <th>thyroid\_surgery</th>\n",

" <th>I131\_treatment</th>\n",

" <th>query\_hypothyroid</th>\n",

" <th>...</th>\n",

" <th>tumor</th>\n",

" <th>hypopituitary</th>\n",

" <th>psych</th>\n",

" <th>TSH</th>\n",

" <th>T3</th>\n",

" <th>TT4</th>\n",

" <th>T4U</th>\n",

" <th>FTI</th>\n",

" <th>TBG</th>\n",

" <th>target</th>\n",

" </tr>\n",

" </thead>\n",

" <tbody>\n",

" <tr>\n",

" <th>0</th>\n",

" <td>29</td>\n",

" <td>F</td>\n",

" <td>f</td>\n",

" <td>f</td>\n",

" <td>f</td>\n",

" <td>f</td>\n",

" <td>f</td>\n",

" <td>f</td>\n",

" <td>f</td>\n",

" <td>t</td>\n",

" <td>...</td>\n",

" <td>f</td>\n",

" <td>f</td>\n",

" <td>f</td>\n",

" <td>0.3</td>\n",

" <td>NaN</td>\n",

" <td>NaN</td>\n",

" <td>NaN</td>\n",

" <td>NaN</td>\n",

" <td>NaN</td>\n",

" <td>NaN</td>\n",

" </tr>\n",

" <tr>\n",

" <th>1</th>\n",

" <td>29</td>\n",

" <td>F</td>\n",

" <td>f</td>\n",

" <td>f</td>\n",

" <td>f</td>\n",

" <td>f</td>\n",

" <td>f</td>\n",

" <td>f</td>\n",

" <td>f</td>\n",

" <td>f</td>\n",

" <td>...</td>\n",

" <td>f</td>\n",

" <td>f</td>\n",

" <td>f</td>\n",

" <td>1.6</td>\n",

" <td>1.9</td>\n",

" <td>128.0</td>\n",

" <td>NaN</td>\n",

" <td>NaN</td>\n",

" <td>NaN</td>\n",

" <td>NaN</td>\n",

" </tr>\n",

" <tr>\n",

" <th>2</th>\n",

" <td>41</td>\n",

" <td>F</td>\n",

" <td>f</td>\n",

" <td>f</td>\n",

" <td>f</td>\n",

" <td>f</td>\n",

" <td>f</td>\n",

" <td>f</td>\n",

" <td>f</td>\n",

" <td>f</td>\n",

" <td>...</td>\n",

" <td>f</td>\n",

" <td>f</td>\n",

" <td>f</td>\n",

" <td>NaN</td>\n",

" <td>NaN</td>\n",

" <td>NaN</td>\n",

" <td>NaN</td>\n",

" <td>NaN</td>\n",

" <td>11.0</td>\n",

" <td>NaN</td>\n",

" </tr>\n",

" <tr>\n",

" <th>3</th>\n",

" <td>36</td>\n",

" <td>F</td>\n",

" <td>f</td>\n",

" <td>f</td>\n",

" <td>f</td>\n",

" <td>f</td>\n",

" <td>f</td>\n",

" <td>f</td>\n",

" <td>f</td>\n",

" <td>f</td>\n",

" <td>...</td>\n",

" <td>f</td>\n",

" <td>f</td>\n",

" <td>f</td>\n",

" <td>NaN</td>\n",

" <td>NaN</td>\n",

" <td>NaN</td>\n",

" <td>NaN</td>\n",

" <td>NaN</td>\n",

" <td>26.0</td>\n",

" <td>NaN</td>\n",

" </tr>\n",

" <tr>\n",

" <th>4</th>\n",

" <td>32</td>\n",

" <td>F</td>\n",

" <td>f</td>\n",

" <td>f</td>\n",

" <td>f</td>\n",

" <td>f</td>\n",

" <td>f</td>\n",

" <td>f</td>\n",

" <td>f</td>\n",

" <td>f</td>\n",

" <td>...</td>\n",

" <td>f</td>\n",

" <td>f</td>\n",

" <td>f</td>\n",

" <td>NaN</td>\n",

" <td>NaN</td>\n",

" <td>NaN</td>\n",

" <td>NaN</td>\n",

" <td>NaN</td>\n",

" <td>36.0</td>\n",

" <td>miscellaneous</td>\n",

" </tr>\n",

" <tr>\n",

" <th>...</th>\n",

" <td>...</td>\n",

" <td>...</td>\n",

" <td>...</td>\n",

" <td>...</td>\n",

" <td>...</td>\n",

" <td>...</td>\n",

" <td>...</td>\n",

" <td>...</td>\n",

" <td>...</td>\n",

" <td>...</td>\n",

" <td>...</td>\n",

" <td>...</td>\n",

" <td>...</td>\n",

" <td>...</td>\n",

" <td>...</td>\n",

" <td>...</td>\n",

" <td>...</td>\n",

" <td>...</td>\n",

" <td>...</td>\n",

" <td>...</td>\n",

" <td>...</td>\n",

" </tr>\n",

" <tr>\n",

" <th>9167</th>\n",

" <td>56</td>\n",

" <td>M</td>\n",

" <td>f</td>\n",

" <td>f</td>\n",

" <td>f</td>\n",

" <td>f</td>\n",

" <td>f</td>\n",

" <td>f</td>\n",

" <td>f</td>\n",

" <td>f</td>\n",

" <td>...</td>\n",

" <td>f</td>\n",

" <td>f</td>\n",

" <td>f</td>\n",

" <td>NaN</td>\n",

" <td>NaN</td>\n",

" <td>64.0</td>\n",

" <td>0.83</td>\n",

" <td>77.0</td>\n",

" <td>NaN</td>\n",

" <td>NaN</td>\n",

" </tr>\n",

" <tr>\n",

" <th>9168</th>\n",

" <td>22</td>\n",

" <td>M</td>\n",

" <td>f</td>\n",

" <td>f</td>\n",

" <td>f</td>\n",

" <td>f</td>\n",

" <td>f</td>\n",

" <td>f</td>\n",

" <td>f</td>\n",

" <td>f</td>\n",

" <td>...</td>\n",

" <td>f</td>\n",

" <td>f</td>\n",

" <td>f</td>\n",

" <td>NaN</td>\n",

" <td>NaN</td>\n",

" <td>91.0</td>\n",

" <td>0.92</td>\n",

" <td>99.0</td>\n",

" <td>NaN</td>\n",

" <td>NaN</td>\n",

" </tr>\n",

" <tr>\n",

" <th>9169</th>\n",

" <td>69</td>\n",

" <td>M</td>\n",

" <td>f</td>\n",

" <td>f</td>\n",

" <td>f</td>\n",

" <td>f</td>\n",

" <td>f</td>\n",

" <td>f</td>\n",

" <td>f</td>\n",

" <td>f</td>\n",

" <td>...</td>\n",

" <td>f</td>\n",

" <td>f</td>\n",

" <td>f</td>\n",

" <td>NaN</td>\n",

" <td>NaN</td>\n",

" <td>113.0</td>\n",

" <td>1.27</td>\n",

" <td>89.0</td>\n",

" <td>NaN</td>\n",

" <td>binding protein</td>\n",

" </tr>\n",

" <tr>\n",

" <th>9170</th>\n",

" <td>47</td>\n",

" <td>F</td>\n",

" <td>f</td>\n",

" <td>f</td>\n",

" <td>f</td>\n",

" <td>f</td>\n",

" <td>f</td>\n",

" <td>f</td>\n",

" <td>f</td>\n",

" <td>f</td>\n",

" <td>...</td>\n",

" <td>f</td>\n",

" <td>f</td>\n",

" <td>f</td>\n",

" <td>NaN</td>\n",

" <td>NaN</td>\n",

" <td>75.0</td>\n",

" <td>0.85</td>\n",

" <td>88.0</td>\n",

" <td>NaN</td>\n",

" <td>NaN</td>\n",

" </tr>\n",

" <tr>\n",

" <th>9171</th>\n",

" <td>31</td>\n",

" <td>M</td>\n",

" <td>f</td>\n",

" <td>f</td>\n",

" <td>f</td>\n",

" <td>f</td>\n",

" <td>f</td>\n",

" <td>f</td>\n",

" <td>f</td>\n",

" <td>t</td>\n",

" <td>...</td>\n",

" <td>f</td>\n",

" <td>f</td>\n",

" <td>f</td>\n",

" <td>NaN</td>\n",

" <td>NaN</td>\n",

" <td>66.0</td>\n",

" <td>1.02</td>\n",

" <td>65.0</td>\n",

" <td>NaN</td>\n",

" <td>NaN</td>\n",

" </tr>\n",

" </tbody>\n",

"</table>\n",

"<p>9172 rows × 23 columns</p>\n",

"</div>\n",

" <button class=\"colab-df-convert\" onclick=\"convertToInteractive('df-8ca7cc93-51c6-4956-8e3b-2cc610409600')\"\n",

" title=\"Convert this dataframe to an interactive table.\"\n",

" style=\"display:none;\">\n",

" \n",

" <svg xmlns=\"http://www.w3.org/2000/svg\" height=\"24px\"viewBox=\"0 0 24 24\"\n",

" width=\"24px\">\n",

" <path d=\"M0 0h24v24H0V0z\" fill=\"none\"/>\n",

" <path d=\"M18.56 5.44l.94 2.06.94-2.06 2.06-.94-2.06-.94-.94-2.06-.94 2.06-2.06.94zm-11 1L8.5 8.5l.94-2.06 2.06-.94-2.06-.94L8.5 2.5l-.94 2.06-2.06.94zm10 10l.94 2.06.94-2.06 2.06-.94-2.06-.94-.94-2.06-.94 2.06-2.06.94z\"/><path d=\"M17.41 7.96l-1.37-1.37c-.4-.4-.92-.59-1.43-.59-.52 0-1.04.2-1.43.59L10.3 9.45l-7.72 7.72c-.78.78-.78 2.05 0 2.83L4 21.41c.39.39.9.59 1.41.59.51 0 1.02-.2 1.41-.59l7.78-7.78 2.81-2.81c.8-.78.8-2.07 0-2.86zM5.41 20L4 18.59l7.72-7.72 1.47 1.35L5.41 20z\"/>\n",

" </svg>\n",

" </button>\n",

" \n",

" <style>\n",

" .colab-df-container {\n",

" display:flex;\n",

" flex-wrap:wrap;\n",

" gap: 12px;\n",

" }\n",

"\n",

" .colab-df-convert {\n",

" background-color: #E8F0FE;\n",

" border: none;\n",

" border-radius: 50%;\n",

" cursor: pointer;\n",

" display: none;\n",

" fill: #1967D2;\n",

" height: 32px;\n",

" padding: 0 0 0 0;\n",

" width: 32px;\n",

" }\n",

"\n",

" .colab-df-convert:hover {\n",

" background-color: #E2EBFA;\n",

" box-shadow: 0px 1px 2px rgba(60, 64, 67, 0.3), 0px 1px 3px 1px rgba(60, 64, 67, 0.15);\n",

" fill: #174EA6;\n",

" }\n",

"\n",

" [theme=dark] .colab-df-convert {\n",

" background-color: #3B4455;\n",

" fill: #D2E3FC;\n",

" }\n",

"\n",

" [theme=dark] .colab-df-convert:hover {\n",

" background-color: #434B5C;\n",

" box-shadow: 0px 1px 3px 1px rgba(0, 0, 0, 0.15);\n",

" filter: drop-shadow(0px 1px 2px rgba(0, 0, 0, 0.3));\n",

" fill: #FFFFFF;\n",

" }\n",

" </style>\n",

"\n",

" <script>\n",

" const buttonEl =\n",

" document.querySelector('#df-8ca7cc93-51c6-4956-8e3b-2cc610409600 button.colab-df-convert');\n",

" buttonEl.style.display =\n",

" google.colab.kernel.accessAllowed ? 'block' : 'none';\n",

"\n",

" async function convertToInteractive(key) {\n",

" const element = document.querySelector('#df-8ca7cc93-51c6-4956-8e3b-2cc610409600');\n",

" const dataTable =\n",

" await google.colab.kernel.invokeFunction('convertToInteractive',\n",

" [key], {});\n",

" if (!dataTable) return;\n",

"\n",

" const docLinkHtml = 'Like what you see? Visit the ' +\n",

" '<a target=\"\_blank\" href=https://colab.research.google.com/notebooks/data\_table.ipynb>data table notebook</a>'\n",

" + ' to learn more about interactive tables.';\n",

" element.innerHTML = '';\n",

" dataTable['output\_type'] = 'display\_data';\n",

" await google.colab.output.renderOutput(dataTable, element);\n",

" const docLink = document.createElement('div');\n",

" docLink.innerHTML = docLinkHtml;\n",

" element.appendChild(docLink);\n",

" }\n",

" </script>\n",

" </div>\n",

" </div>\n",

" "

]

},

"metadata": {},

"execution\_count": 18

}

],

"source": [

"data"

]

},

{

"cell\_type": "code",

"execution\_count": null,

"metadata": {

"colab": {

"base\_uri": "https://localhost:8080/"

},

"id": "DQCKOF9CscJq",

"outputId": "134a40f7-4a43-4fa8-c15c-a41af21d1531"

},

"outputs": [

{

"output\_type": "execute\_result",

"data": {

"text/plain": [

"age 0\n",

"sex 307\n",

"on\_thyroxine 0\n",

"query\_on\_thyroxine 0\n",

"on\_antithyroid\_meds 0\n",

"sick 0\n",

"pregnant 0\n",

"thyroid\_surgery 0\n",

"I131\_treatment 0\n",

"query\_hypothyroid 0\n",

"query\_hyperthyroid 0\n",

"lithium 0\n",

"goitre 0\n",

"tumor 0\n",

"hypopituitary 0\n",

"psych 0\n",

"TSH 842\n",

"T3 2604\n",

"TT4 442\n",

"T4U 809\n",

"FTI 802\n",

"TBG 8823\n",

"target 6949\n",

"dtype: int64"

]

},

"metadata": {},

"execution\_count": 19

}

],

"source": [

"data.isnull().sum()"

]

},

{

"cell\_type": "code",

"execution\_count": null,

"metadata": {

"id": "2RSqG0hBsiMm"

},

"outputs": [],

"source": [

"data.dropna(subset=['target'],inplace=True)"

]

},

{

"cell\_type": "code",

"execution\_count": null,

"metadata": {

"colab": {

"base\_uri": "https://localhost:8080/"

},

"id": "zVZQ4a1ytMZg",

"outputId": "22c597ba-6d19-4356-ba94-efdcdc1d6329"

},

"outputs": [

{

"output\_type": "execute\_result",

"data": {

"text/plain": [

"hypothyroid conditions 593\n",

"general health 436\n",

"binding protein 376\n",

"replacement therapy 336\n",

"miscellaneous 281\n",

"hyperthyroid conditions 182\n",

"antithyroid treatment 19\n",

"Name: target, dtype: int64"

]

},

"metadata": {},

"execution\_count": 21

}

],

"source": [

"data['target'].value\_counts()"

]

},

{

"cell\_type": "code",

"execution\_count": null,

"metadata": {

"colab": {

"base\_uri": "https://localhost:8080/"

},

"id": "lFJJ0HmjtWTh",

"outputId": "15d2fa3d-3fb2-45b7-a8a7-79e85cf91839"

},

"outputs": [

{

"output\_type": "execute\_result",

"data": {

"text/plain": [

"0"

]

},

"metadata": {},

"execution\_count": 22

}

],

"source": [

"data['target'].isnull().sum()"

]

},

{

"cell\_type": "code",

"execution\_count": null,

"metadata": {

"colab": {

"base\_uri": "https://localhost:8080/",

"height": 352

},

"id": "k3l1H5U9tjLB",

"outputId": "2fb23aa1-cbe5-4fae-f3f8-0898ad5e6428"

},

"outputs": [

{

"output\_type": "execute\_result",

"data": {

"text/plain": [

" age sex on\_thyroxine query\_on\_thyroxine on\_antithyroid\_meds sick pregnant \\\n",

"4 32 F f f f f f \n",

"18 63 F t f f t f \n",

"32 41 M f f f f f \n",

"33 71 F t f f f f \n",

"39 55 F t f f f f \n",

"\n",

" thyroid\_surgery I131\_treatment query\_hypothyroid ... tumor hypopituitary \\\n",

"4 f f f ... f f \n",

"18 f f f ... f f \n",

"32 f f f ... f f \n",

"33 f f f ... f f \n",

"39 f f t ... f f \n",

"\n",

" psych TSH T3 TT4 T4U FTI TBG target \n",

"4 f NaN NaN NaN NaN NaN 36.0 miscellaneous \n",

"18 f 68.000000 NaN 48.0 1.02 47.0 NaN hypothyroid conditions \n",

"32 f 0.050000 1.6 39.0 1.00 39.0 NaN miscellaneous \n",

"33 f 0.050000 NaN 126.0 1.38 91.0 NaN binding protein \n",

"39 f 9.599999 2.4 136.0 1.48 92.0 NaN replacement therapy \n",

"\n",

"[5 rows x 23 columns]"

],

"text/html": [

"\n",

" <div id=\"df-3ba7d39c-c64f-442e-81d3-ca0eece9d968\">\n",

" <div class=\"colab-df-container\">\n",

" <div>\n",

"<style scoped>\n",

" .dataframe tbody tr th:only-of-type {\n",

" vertical-align: middle;\n",

" }\n",

"\n",

" .dataframe tbody tr th {\n",

" vertical-align: top;\n",

" }\n",

"\n",

" .dataframe thead th {\n",

" text-align: right;\n",

" }\n",

"</style>\n",

"<table border=\"1\" class=\"dataframe\">\n",

" <thead>\n",

" <tr style=\"text-align: right;\">\n",

" <th></th>\n",

" <th>age</th>\n",

" <th>sex</th>\n",

" <th>on\_thyroxine</th>\n",

" <th>query\_on\_thyroxine</th>\n",

" <th>on\_antithyroid\_meds</th>\n",

" <th>sick</th>\n",

" <th>pregnant</th>\n",

" <th>thyroid\_surgery</th>\n",

" <th>I131\_treatment</th>\n",

" <th>query\_hypothyroid</th>\n",

" <th>...</th>\n",

" <th>tumor</th>\n",

" <th>hypopituitary</th>\n",

" <th>psych</th>\n",

" <th>TSH</th>\n",

" <th>T3</th>\n",

" <th>TT4</th>\n",

" <th>T4U</th>\n",

" <th>FTI</th>\n",

" <th>TBG</th>\n",

" <th>target</th>\n",

" </tr>\n",

" </thead>\n",

" <tbody>\n",

" <tr>\n",

" <th>4</th>\n",

" <td>32</td>\n",

" <td>F</td>\n",

" <td>f</td>\n",

" <td>f</td>\n",

" <td>f</td>\n",

" <td>f</td>\n",

" <td>f</td>\n",

" <td>f</td>\n",

" <td>f</td>\n",

" <td>f</td>\n",

" <td>...</td>\n",

" <td>f</td>\n",

" <td>f</td>\n",

" <td>f</td>\n",

" <td>NaN</td>\n",

" <td>NaN</td>\n",

" <td>NaN</td>\n",

" <td>NaN</td>\n",

" <td>NaN</td>\n",

" <td>36.0</td>\n",

" <td>miscellaneous</td>\n",

" </tr>\n",

" <tr>\n",

" <th>18</th>\n",

" <td>63</td>\n",

" <td>F</td>\n",

" <td>t</td>\n",

" <td>f</td>\n",

" <td>f</td>\n",

" <td>t</td>\n",

" <td>f</td>\n",

" <td>f</td>\n",

" <td>f</td>\n",

" <td>f</td>\n",

" <td>...</td>\n",

" <td>f</td>\n",

" <td>f</td>\n",

" <td>f</td>\n",

" <td>68.000000</td>\n",

" <td>NaN</td>\n",

" <td>48.0</td>\n",

" <td>1.02</td>\n",

" <td>47.0</td>\n",

" <td>NaN</td>\n",

" <td>hypothyroid conditions</td>\n",

" </tr>\n",

" <tr>\n",

" <th>32</th>\n",

" <td>41</td>\n",

" <td>M</td>\n",

" <td>f</td>\n",

" <td>f</td>\n",

" <td>f</td>\n",

" <td>f</td>\n",

" <td>f</td>\n",

" <td>f</td>\n",

" <td>f</td>\n",

" <td>f</td>\n",

" <td>...</td>\n",

" <td>f</td>\n",

" <td>f</td>\n",

" <td>f</td>\n",

" <td>0.050000</td>\n",

" <td>1.6</td>\n",

" <td>39.0</td>\n",

" <td>1.00</td>\n",

" <td>39.0</td>\n",

" <td>NaN</td>\n",

" <td>miscellaneous</td>\n",

" </tr>\n",

" <tr>\n",

" <th>33</th>\n",

" <td>71</td>\n",

" <td>F</td>\n",

" <td>t</td>\n",

" <td>f</td>\n",

" <td>f</td>\n",

" <td>f</td>\n",

" <td>f</td>\n",

" <td>f</td>\n",

" <td>f</td>\n",

" <td>f</td>\n",

" <td>...</td>\n",

" <td>f</td>\n",

" <td>f</td>\n",

" <td>f</td>\n",

" <td>0.050000</td>\n",

" <td>NaN</td>\n",

" <td>126.0</td>\n",

" <td>1.38</td>\n",

" <td>91.0</td>\n",

" <td>NaN</td>\n",

" <td>binding protein</td>\n",

" </tr>\n",

" <tr>\n",

" <th>39</th>\n",

" <td>55</td>\n",

" <td>F</td>\n",

" <td>t</td>\n",

" <td>f</td>\n",

" <td>f</td>\n",

" <td>f</td>\n",

" <td>f</td>\n",

" <td>f</td>\n",

" <td>f</td>\n",

" <td>t</td>\n",

" <td>...</td>\n",

" <td>f</td>\n",

" <td>f</td>\n",

" <td>f</td>\n",

" <td>9.599999</td>\n",

" <td>2.4</td>\n",

" <td>136.0</td>\n",

" <td>1.48</td>\n",

" <td>92.0</td>\n",

" <td>NaN</td>\n",

" <td>replacement therapy</td>\n",

" </tr>\n",

" </tbody>\n",

"</table>\n",

"<p>5 rows × 23 columns</p>\n",

"</div>\n",

" <button class=\"colab-df-convert\" onclick=\"convertToInteractive('df-3ba7d39c-c64f-442e-81d3-ca0eece9d968')\"\n",

" title=\"Convert this dataframe to an interactive table.\"\n",

" style=\"display:none;\">\n",

" \n",

" <svg xmlns=\"http://www.w3.org/2000/svg\" height=\"24px\"viewBox=\"0 0 24 24\"\n",

" width=\"24px\">\n",

" <path d=\"M0 0h24v24H0V0z\" fill=\"none\"/>\n",

" <path d=\"M18.56 5.44l.94 2.06.94-2.06 2.06-.94-2.06-.94-.94-2.06-.94 2.06-2.06.94zm-11 1L8.5 8.5l.94-2.06 2.06-.94-2.06-.94L8.5 2.5l-.94 2.06-2.06.94zm10 10l.94 2.06.94-2.06 2.06-.94-2.06-.94-.94-2.06-.94 2.06-2.06.94z\"/><path d=\"M17.41 7.96l-1.37-1.37c-.4-.4-.92-.59-1.43-.59-.52 0-1.04.2-1.43.59L10.3 9.45l-7.72 7.72c-.78.78-.78 2.05 0 2.83L4 21.41c.39.39.9.59 1.41.59.51 0 1.02-.2 1.41-.59l7.78-7.78 2.81-2.81c.8-.78.8-2.07 0-2.86zM5.41 20L4 18.59l7.72-7.72 1.47 1.35L5.41 20z\"/>\n",

" </svg>\n",

" </button>\n",

" \n",

" <style>\n",

" .colab-df-container {\n",

" display:flex;\n",

" flex-wrap:wrap;\n",

" gap: 12px;\n",

" }\n",

"\n",

" .colab-df-convert {\n",

" background-color: #E8F0FE;\n",

" border: none;\n",

" border-radius: 50%;\n",

" cursor: pointer;\n",

" display: none;\n",

" fill: #1967D2;\n",

" height: 32px;\n",

" padding: 0 0 0 0;\n",

" width: 32px;\n",

" }\n",

"\n",

" .colab-df-convert:hover {\n",

" background-color: #E2EBFA;\n",

" box-shadow: 0px 1px 2px rgba(60, 64, 67, 0.3), 0px 1px 3px 1px rgba(60, 64, 67, 0.15);\n",

" fill: #174EA6;\n",

" }\n",

"\n",

" [theme=dark] .colab-df-convert {\n",

" background-color: #3B4455;\n",

" fill: #D2E3FC;\n",

" }\n",

"\n",

" [theme=dark] .colab-df-convert:hover {\n",

" background-color: #434B5C;\n",

" box-shadow: 0px 1px 3px 1px rgba(0, 0, 0, 0.15);\n",

" filter: drop-shadow(0px 1px 2px rgba(0, 0, 0, 0.3));\n",

" fill: #FFFFFF;\n",

" }\n",

" </style>\n",

"\n",

" <script>\n",

" const buttonEl =\n",

" document.querySelector('#df-3ba7d39c-c64f-442e-81d3-ca0eece9d968 button.colab-df-convert');\n",

" buttonEl.style.display =\n",

" google.colab.kernel.accessAllowed ? 'block' : 'none';\n",

"\n",

" async function convertToInteractive(key) {\n",

" const element = document.querySelector('#df-3ba7d39c-c64f-442e-81d3-ca0eece9d968');\n",

" const dataTable =\n",

" await google.colab.kernel.invokeFunction('convertToInteractive',\n",

" [key], {});\n",

" if (!dataTable) return;\n",

"\n",

" const docLinkHtml = 'Like what you see? Visit the ' +\n",

" '<a target=\"\_blank\" href=https://colab.research.google.com/notebooks/data\_table.ipynb>data table notebook</a>'\n",

" + ' to learn more about interactive tables.';\n",

" element.innerHTML = '';\n",

" dataTable['output\_type'] = 'display\_data';\n",

" await google.colab.output.renderOutput(dataTable, element);\n",

" const docLink = document.createElement('div');\n",

" docLink.innerHTML = docLinkHtml;\n",

" element.appendChild(docLink);\n",

" }\n",

" </script>\n",

" </div>\n",

" </div>\n",

" "

]

},

"metadata": {},

"execution\_count": 23

}

],

"source": [

"data.head()"

]

},

{

"cell\_type": "code",

"execution\_count": null,

"metadata": {

"colab": {

"base\_uri": "https://localhost:8080/",

"height": 300

},

"id": "oRoJMGWwtm9C",

"outputId": "9bb5f895-3396-486d-f0be-5873044688cb"

},

"outputs": [

{

"output\_type": "execute\_result",

"data": {

"text/plain": [

" age TSH T3 TT4 T4U \\\n",

"count 2223.000000 2073.000000 1629.000000 2126.000000 2045.000000 \n",

"mean 52.859199 14.965538 1.963634 116.910470 1.013501 \n",

"std 19.669100 46.343358 1.457743 60.203955 0.281001 \n",

"min 1.000000 0.005000 0.050000 2.000000 0.170000 \n",

"25% 36.000000 0.250000 1.000000 77.000000 0.850000 \n",

"50% 56.000000 2.000000 1.700000 109.500000 0.960000 \n",

"75% 69.500000 8.799999 2.500000 156.000000 1.120000 \n",

"max 95.000000 530.000000 18.000000 600.000000 2.330000 \n",

"\n",

" FTI TBG \n",

"count 2046.000000 98.000000 \n",

"mean 120.928416 47.717347 \n",

"std 70.903791 32.398750 \n",

"min 1.400000 9.299999 \n",

"25% 83.000000 32.000000 \n",

"50% 109.000000 36.000000 \n",

"75% 157.000000 46.750000 \n",

"max 881.000000 200.000000 "

],

"text/html": [

"\n",

" <div id=\"df-8b6d274c-9c69-4929-9925-941dbd0534ca\">\n",

" <div class=\"colab-df-container\">\n",

" <div>\n",

"<style scoped>\n",

" .dataframe tbody tr th:only-of-type {\n",

" vertical-align: middle;\n",

" }\n",

"\n",

" .dataframe tbody tr th {\n",

" vertical-align: top;\n",

" }\n",

"\n",

" .dataframe thead th {\n",

" text-align: right;\n",

" }\n",

"</style>\n",

"<table border=\"1\" class=\"dataframe\">\n",

" <thead>\n",

" <tr style=\"text-align: right;\">\n",

" <th></th>\n",

" <th>age</th>\n",

" <th>TSH</th>\n",

" <th>T3</th>\n",

" <th>TT4</th>\n",

" <th>T4U</th>\n",

" <th>FTI</th>\n",

" <th>TBG</th>\n",

" </tr>\n",

" </thead>\n",

" <tbody>\n",

" <tr>\n",

" <th>count</th>\n",

" <td>2223.000000</td>\n",

" <td>2073.000000</td>\n",

" <td>1629.000000</td>\n",

" <td>2126.000000</td>\n",

" <td>2045.000000</td>\n",

" <td>2046.000000</td>\n",

" <td>98.000000</td>\n",

" </tr>\n",

" <tr>\n",

" <th>mean</th>\n",

" <td>52.859199</td>\n",

" <td>14.965538</td>\n",

" <td>1.963634</td>\n",

" <td>116.910470</td>\n",

" <td>1.013501</td>\n",

" <td>120.928416</td>\n",

" <td>47.717347</td>\n",

" </tr>\n",

" <tr>\n",

" <th>std</th>\n",

" <td>19.669100</td>\n",

" <td>46.343358</td>\n",

" <td>1.457743</td>\n",

" <td>60.203955</td>\n",

" <td>0.281001</td>\n",

" <td>70.903791</td>\n",

" <td>32.398750</td>\n",

" </tr>\n",

" <tr>\n",

" <th>min</th>\n",

" <td>1.000000</td>\n",

" <td>0.005000</td>\n",

" <td>0.050000</td>\n",

" <td>2.000000</td>\n",

" <td>0.170000</td>\n",

" <td>1.400000</td>\n",

" <td>9.299999</td>\n",

" </tr>\n",

" <tr>\n",

" <th>25%</th>\n",

" <td>36.000000</td>\n",

" <td>0.250000</td>\n",

" <td>1.000000</td>\n",

" <td>77.000000</td>\n",

" <td>0.850000</td>\n",

" <td>83.000000</td>\n",

" <td>32.000000</td>\n",

" </tr>\n",

" <tr>\n",

" <th>50%</th>\n",

" <td>56.000000</td>\n",

" <td>2.000000</td>\n",

" <td>1.700000</td>\n",

" <td>109.500000</td>\n",

" <td>0.960000</td>\n",

" <td>109.000000</td>\n",

" <td>36.000000</td>\n",

" </tr>\n",

" <tr>\n",

" <th>75%</th>\n",

" <td>69.500000</td>\n",

" <td>8.799999</td>\n",

" <td>2.500000</td>\n",

" <td>156.000000</td>\n",

" <td>1.120000</td>\n",

" <td>157.000000</td>\n",

" <td>46.750000</td>\n",

" </tr>\n",

" <tr>\n",

" <th>max</th>\n",

" <td>95.000000</td>\n",

" <td>530.000000</td>\n",

" <td>18.000000</td>\n",

" <td>600.000000</td>\n",

" <td>2.330000</td>\n",

" <td>881.000000</td>\n",

" <td>200.000000</td>\n",

" </tr>\n",

" </tbody>\n",

"</table>\n",

"</div>\n",

" <button class=\"colab-df-convert\" onclick=\"convertToInteractive('df-8b6d274c-9c69-4929-9925-941dbd0534ca')\"\n",

" title=\"Convert this dataframe to an interactive table.\"\n",

" style=\"display:none;\">\n",

" \n",

" <svg xmlns=\"http://www.w3.org/2000/svg\" height=\"24px\"viewBox=\"0 0 24 24\"\n",

" width=\"24px\">\n",

" <path d=\"M0 0h24v24H0V0z\" fill=\"none\"/>\n",

" <path d=\"M18.56 5.44l.94 2.06.94-2.06 2.06-.94-2.06-.94-.94-2.06-.94 2.06-2.06.94zm-11 1L8.5 8.5l.94-2.06 2.06-.94-2.06-.94L8.5 2.5l-.94 2.06-2.06.94zm10 10l.94 2.06.94-2.06 2.06-.94-2.06-.94-.94-2.06-.94 2.06-2.06.94z\"/><path d=\"M17.41 7.96l-1.37-1.37c-.4-.4-.92-.59-1.43-.59-.52 0-1.04.2-1.43.59L10.3 9.45l-7.72 7.72c-.78.78-.78 2.05 0 2.83L4 21.41c.39.39.9.59 1.41.59.51 0 1.02-.2 1.41-.59l7.78-7.78 2.81-2.81c.8-.78.8-2.07 0-2.86zM5.41 20L4 18.59l7.72-7.72 1.47 1.35L5.41 20z\"/>\n",

" </svg>\n",

" </button>\n",

" \n",

" <style>\n",

" .colab-df-container {\n",

" display:flex;\n",

" flex-wrap:wrap;\n",

" gap: 12px;\n",

" }\n",

"\n",

" .colab-df-convert {\n",

" background-color: #E8F0FE;\n",

" border: none;\n",

" border-radius: 50%;\n",

" cursor: pointer;\n",

" display: none;\n",

" fill: #1967D2;\n",

" height: 32px;\n",

" padding: 0 0 0 0;\n",

" width: 32px;\n",

" }\n",

"\n",

" .colab-df-convert:hover {\n",

" background-color: #E2EBFA;\n",

" box-shadow: 0px 1px 2px rgba(60, 64, 67, 0.3), 0px 1px 3px 1px rgba(60, 64, 67, 0.15);\n",

" fill: #174EA6;\n",

" }\n",

"\n",

" [theme=dark] .colab-df-convert {\n",

" background-color: #3B4455;\n",

" fill: #D2E3FC;\n",

" }\n",

"\n",

" [theme=dark] .colab-df-convert:hover {\n",

" background-color: #434B5C;\n",

" box-shadow: 0px 1px 3px 1px rgba(0, 0, 0, 0.15);\n",

" filter: drop-shadow(0px 1px 2px rgba(0, 0, 0, 0.3));\n",

" fill: #FFFFFF;\n",

" }\n",

" </style>\n",

"\n",

" <script>\n",

" const buttonEl =\n",

" document.querySelector('#df-8b6d274c-9c69-4929-9925-941dbd0534ca button.colab-df-convert');\n",

" buttonEl.style.display =\n",

" google.colab.kernel.accessAllowed ? 'block' : 'none';\n",

"\n",

" async function convertToInteractive(key) {\n",

" const element = document.querySelector('#df-8b6d274c-9c69-4929-9925-941dbd0534ca');\n",

" const dataTable =\n",

" await google.colab.kernel.invokeFunction('convertToInteractive',\n",

" [key], {});\n",

" if (!dataTable) return;\n",

"\n",

" const docLinkHtml = 'Like what you see? Visit the ' +\n",

" '<a target=\"\_blank\" href=https://colab.research.google.com/notebooks/data\_table.ipynb>data table notebook</a>'\n",

" + ' to learn more about interactive tables.';\n",

" element.innerHTML = '';\n",

" dataTable['output\_type'] = 'display\_data';\n",

" await google.colab.output.renderOutput(dataTable, element);\n",

" const docLink = document.createElement('div');\n",

" docLink.innerHTML = docLinkHtml;\n",

" element.appendChild(docLink);\n",

" }\n",

" </script>\n",

" </div>\n",

" </div>\n",

" "

]

},

"metadata": {},

"execution\_count": 24

}

],

"source": [

"data.describe()"

]

},

{

"cell\_type": "code",

"execution\_count": null,

"metadata": {

"colab": {

"base\_uri": "https://localhost:8080/",

"height": 143

},

"id": "Vt7ewh9\_tuBx",

"outputId": "799c37be-3958-46f2-f347-767e05de194d"

},

"outputs": [

{

"output\_type": "execute\_result",

"data": {

"text/plain": [

"Empty DataFrame\n",

"Columns: [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, TT4, T4U, FTI, TBG, target]\n",

"Index: []\n",

"\n",

"[0 rows x 23 columns]"

],

"text/html": [

"\n",

" <div id=\"df-f93d3ba2-ae96-478c-86f7-685a60172ed7\">\n",

" <div class=\"colab-df-container\">\n",

" <div>\n",

"<style scoped>\n",

" .dataframe tbody tr th:only-of-type {\n",

" vertical-align: middle;\n",

" }\n",

"\n",

" .dataframe tbody tr th {\n",

" vertical-align: top;\n",

" }\n",

"\n",

" .dataframe thead th {\n",

" text-align: right;\n",

" }\n",

"</style>\n",

"<table border=\"1\" class=\"dataframe\">\n",

" <thead>\n",

" <tr style=\"text-align: right;\">\n",

" <th></th>\n",

" <th>age</th>\n",

" <th>sex</th>\n",

" <th>on\_thyroxine</th>\n",

" <th>query\_on\_thyroxine</th>\n",

" <th>on\_antithyroid\_meds</th>\n",

" <th>sick</th>\n",

" <th>pregnant</th>\n",

" <th>thyroid\_surgery</th>\n",

" <th>I131\_treatment</th>\n",

" <th>query\_hypothyroid</th>\n",

" <th>...</th>\n",

" <th>tumor</th>\n",

" <th>hypopituitary</th>\n",

" <th>psych</th>\n",

" <th>TSH</th>\n",

" <th>T3</th>\n",

" <th>TT4</th>\n",

" <th>T4U</th>\n",

" <th>FTI</th>\n",

" <th>TBG</th>\n",

" <th>target</th>\n",

" </tr>\n",

" </thead>\n",

" <tbody>\n",

" </tbody>\n",

"</table>\n",

"<p>0 rows × 23 columns</p>\n",

"</div>\n",

" <button class=\"colab-df-convert\" onclick=\"convertToInteractive('df-f93d3ba2-ae96-478c-86f7-685a60172ed7')\"\n",

" title=\"Convert this dataframe to an interactive table.\"\n",

" style=\"display:none;\">\n",

" \n",

" <svg xmlns=\"http://www.w3.org/2000/svg\" height=\"24px\"viewBox=\"0 0 24 24\"\n",

" width=\"24px\">\n",

" <path d=\"M0 0h24v24H0V0z\" fill=\"none\"/>\n",

" <path d=\"M18.56 5.44l.94 2.06.94-2.06 2.06-.94-2.06-.94-.94-2.06-.94 2.06-2.06.94zm-11 1L8.5 8.5l.94-2.06 2.06-.94-2.06-.94L8.5 2.5l-.94 2.06-2.06.94zm10 10l.94 2.06.94-2.06 2.06-.94-2.06-.94-.94-2.06-.94 2.06-2.06.94z\"/><path d=\"M17.41 7.96l-1.37-1.37c-.4-.4-.92-.59-1.43-.59-.52 0-1.04.2-1.43.59L10.3 9.45l-7.72 7.72c-.78.78-.78 2.05 0 2.83L4 21.41c.39.39.9.59 1.41.59.51 0 1.02-.2 1.41-.59l7.78-7.78 2.81-2.81c.8-.78.8-2.07 0-2.86zM5.41 20L4 18.59l7.72-7.72 1.47 1.35L5.41 20z\"/>\n",

" </svg>\n",

" </button>\n",

" \n",

" <style>\n",

" .colab-df-container {\n",

" display:flex;\n",

" flex-wrap:wrap;\n",

" gap: 12px;\n",

" }\n",

"\n",

" .colab-df-convert {\n",

" background-color: #E8F0FE;\n",

" border: none;\n",

" border-radius: 50%;\n",

" cursor: pointer;\n",

" display: none;\n",

" fill: #1967D2;\n",

" height: 32px;\n",

" padding: 0 0 0 0;\n",

" width: 32px;\n",

" }\n",

"\n",

" .colab-df-convert:hover {\n",

" background-color: #E2EBFA;\n",

" box-shadow: 0px 1px 2px rgba(60, 64, 67, 0.3), 0px 1px 3px 1px rgba(60, 64, 67, 0.15);\n",

" fill: #174EA6;\n",

" }\n",

"\n",

" [theme=dark] .colab-df-convert {\n",

" background-color: #3B4455;\n",

" fill: #D2E3FC;\n",

" }\n",

"\n",

" [theme=dark] .colab-df-convert:hover {\n",

" background-color: #434B5C;\n",

" box-shadow: 0px 1px 3px 1px rgba(0, 0, 0, 0.15);\n",

" filter: drop-shadow(0px 1px 2px rgba(0, 0, 0, 0.3));\n",

" fill: #FFFFFF;\n",

" }\n",

" </style>\n",

"\n",

" <script>\n",

" const buttonEl =\n",

" document.querySelector('#df-f93d3ba2-ae96-478c-86f7-685a60172ed7 button.colab-df-convert');\n",

" buttonEl.style.display =\n",

" google.colab.kernel.accessAllowed ? 'block' : 'none';\n",

"\n",

" async function convertToInteractive(key) {\n",

" const element = document.querySelector('#df-f93d3ba2-ae96-478c-86f7-685a60172ed7');\n",

" const dataTable =\n",

" await google.colab.kernel.invokeFunction('convertToInteractive',\n",

" [key], {});\n",

" if (!dataTable) return;\n",

"\n",

" const docLinkHtml = 'Like what you see? Visit the ' +\n",

" '<a target=\"\_blank\" href=https://colab.research.google.com/notebooks/data\_table.ipynb>data table notebook</a>'\n",

" + ' to learn more about interactive tables.';\n",

" element.innerHTML = '';\n",

" dataTable['output\_type'] = 'display\_data';\n",

" await google.colab.output.renderOutput(dataTable, element);\n",

" const docLink = document.createElement('div');\n",

" docLink.innerHTML = docLinkHtml;\n",

" element.appendChild(docLink);\n",

" }\n",

" </script>\n",

" </div>\n",

" </div>\n",

" "

]

},

"metadata": {},

"execution\_count": 25

}

],

"source": [

"data[data.age>100]"

]

},

{

"cell\_type": "code",

"execution\_count": null,

"metadata": {

"id": "eOoCYT1Nt4Uz"

},

"outputs": [],

"source": [

"data['age']=np.where((data.age>100), np.nan, data.age)"

]

},

{

"cell\_type": "code",

"execution\_count": null,

"metadata": {

"colab": {

"base\_uri": "https://localhost:8080/",

"height": 627

},

"id": "EbRE1uLquZu6",

"outputId": "edd9438d-0485-45cc-acd6-4598499bec58"

},

"outputs": [

{

"output\_type": "execute\_result",

"data": {

"text/plain": [

" age sex on\_thyroxine query\_on\_thyroxine on\_antithyroid\_meds sick \\\n",

"4 32.0 F f f f f \n",

"18 63.0 F t f f t \n",

"32 41.0 M f f f f \n",

"33 71.0 F t f f f \n",

"39 55.0 F t f f f \n",

"... ... .. ... ... ... ... \n",

"9153 64.0 M f f f f \n",

"9157 60.0 M f f t f \n",

"9158 64.0 M f f f f \n",

"9162 36.0 F f f f f \n",

"9169 69.0 M f f f f \n",

"\n",

" pregnant thyroid\_surgery I131\_treatment query\_hypothyroid ... tumor \\\n",

"4 f f f f ... f \n",

"18 f f f f ... f \n",

"32 f f f f ... f \n",

"33 f f f f ... f \n",

"39 f f f t ... f \n",

"... ... ... ... ... ... ... \n",

"9153 f f f f ... f \n",

"9157 f f f f ... f \n",

"9158 f f f t ... f \n",

"9162 f f f f ... f \n",

"9169 f f f f ... f \n",

"\n",

" hypopituitary psych TSH T3 TT4 T4U FTI TBG \\\n",

"4 f f NaN NaN NaN NaN NaN 36.0 \n",

"18 f f 68.000000 NaN 48.0 1.02 47.0 NaN \n",

"32 f f 0.050000 1.6 39.0 1.00 39.0 NaN \n",

"33 f f 0.050000 NaN 126.0 1.38 91.0 NaN \n",

"39 f f 9.599999 2.4 136.0 1.48 92.0 NaN \n",

"... ... ... ... ... ... ... ... ... \n",

"9153 f f 0.810000 NaN 31.0 0.55 56.0 NaN \n",

"9157 f f 0.180000 NaN 28.0 0.87 32.0 NaN \n",

"9158 f f NaN NaN 44.0 0.53 83.0 NaN \n",

"9162 f f NaN NaN 84.0 1.26 67.0 NaN \n",

"9169 f f NaN NaN 113.0 1.27 89.0 NaN \n",

"\n",

" target \n",

"4 miscellaneous \n",

"18 hypothyroid conditions \n",

"32 miscellaneous \n",

"33 binding protein \n",

"39 replacement therapy \n",

"... ... \n",

"9153 general health \n",

"9157 general health \n",

"9158 binding protein \n",

"9162 binding protein \n",

"9169 binding protein \n",

"\n",

"[2223 rows x 23 columns]"

],

"text/html": [

"\n",

" <div id=\"df-fe9d8c19-09b5-4db6-befd-9cd4a9c3e49d\">\n",

" <div class=\"colab-df-container\">\n",

" <div>\n",

"<style scoped>\n",

" .dataframe tbody tr th:only-of-type {\n",

" vertical-align: middle;\n",

" }\n",

"\n",

" .dataframe tbody tr th {\n",

" vertical-align: top;\n",

" }\n",

"\n",

" .dataframe thead th {\n",

" text-align: right;\n",

" }\n",

"</style>\n",

"<table border=\"1\" class=\"dataframe\">\n",

" <thead>\n",

" <tr style=\"text-align: right;\">\n",

" <th></th>\n",

" <th>age</th>\n",

" <th>sex</th>\n",

" <th>on\_thyroxine</th>\n",

" <th>query\_on\_thyroxine</th>\n",

" <th>on\_antithyroid\_meds</th>\n",

" <th>sick</th>\n",

" <th>pregnant</th>\n",

" <th>thyroid\_surgery</th>\n",

" <th>I131\_treatment</th>\n",

" <th>query\_hypothyroid</th>\n",

" <th>...</th>\n",

" <th>tumor</th>\n",

" <th>hypopituitary</th>\n",

" <th>psych</th>\n",

" <th>TSH</th>\n",

" <th>T3</th>\n",

" <th>TT4</th>\n",

" <th>T4U</th>\n",

" <th>FTI</th>\n",

" <th>TBG</th>\n",

" <th>target</th>\n",

" </tr>\n",

" </thead>\n",

" <tbody>\n",

" <tr>\n",

" <th>4</th>\n",

" <td>32.0</td>\n",

" <td>F</td>\n",

" <td>f</td>\n",

" <td>f</td>\n",

" <td>f</td>\n",

" <td>f</td>\n",

" <td>f</td>\n",

" <td>f</td>\n",

" <td>f</td>\n",

" <td>f</td>\n",

" <td>...</td>\n",

" <td>f</td>\n",

" <td>f</td>\n",

" <td>f</td>\n",

" <td>NaN</td>\n",

" <td>NaN</td>\n",

" <td>NaN</td>\n",

" <td>NaN</td>\n",

" <td>NaN</td>\n",

" <td>36.0</td>\n",

" <td>miscellaneous</td>\n",

" </tr>\n",

" <tr>\n",

" <th>18</th>\n",

" <td>63.0</td>\n",

" <td>F</td>\n",

" <td>t</td>\n",

" <td>f</td>\n",

" <td>f</td>\n",

" <td>t</td>\n",

" <td>f</td>\n",

" <td>f</td>\n",

" <td>f</td>\n",

" <td>f</td>\n",

" <td>...</td>\n",

" <td>f</td>\n",

" <td>f</td>\n",

" <td>f</td>\n",

" <td>68.000000</td>\n",

" <td>NaN</td>\n",

" <td>48.0</td>\n",

" <td>1.02</td>\n",

" <td>47.0</td>\n",

" <td>NaN</td>\n",

" <td>hypothyroid conditions</td>\n",

" </tr>\n",

" <tr>\n",

" <th>32</th>\n",

" <td>41.0</td>\n",

" <td>M</td>\n",

" <td>f</td>\n",

" <td>f</td>\n",

" <td>f</td>\n",

" <td>f</td>\n",

" <td>f</td>\n",

" <td>f</td>\n",

" <td>f</td>\n",

" <td>f</td>\n",

" <td>...</td>\n",

" <td>f</td>\n",

" <td>f</td>\n",

" <td>f</td>\n",

" <td>0.050000</td>\n",

" <td>1.6</td>\n",

" <td>39.0</td>\n",

" <td>1.00</td>\n",

" <td>39.0</td>\n",

" <td>NaN</td>\n",

" <td>miscellaneous</td>\n",

" </tr>\n",

" <tr>\n",

" <th>33</th>\n",

" <td>71.0</td>\n",

" <td>F</td>\n",

" <td>t</td>\n",

" <td>f</td>\n",

" <td>f</td>\n",

" <td>f</td>\n",

" <td>f</td>\n",

" <td>f</td>\n",

" <td>f</td>\n",

" <td>f</td>\n",

" <td>...</td>\n",

" <td>f</td>\n",

" <td>f</td>\n",

" <td>f</td>\n",

" <td>0.050000</td>\n",

" <td>NaN</td>\n",

" <td>126.0</td>\n",

" <td>1.38</td>\n",

" <td>91.0</td>\n",

" <td>NaN</td>\n",

" <td>binding protein</td>\n",

" </tr>\n",

" <tr>\n",

" <th>39</th>\n",

" <td>55.0</td>\n",

" <td>F</td>\n",

" <td>t</td>\n",

" <td>f</td>\n",

" <td>f</td>\n",

" <td>f</td>\n",

" <td>f</td>\n",

" <td>f</td>\n",

" <td>f</td>\n",

" <td>t</td>\n",

" <td>...</td>\n",

" <td>f</td>\n",

" <td>f</td>\n",

" <td>f</td>\n",

" <td>9.599999</td>\n",

" <td>2.4</td>\n",

" <td>136.0</td>\n",

" <td>1.48</td>\n",

" <td>92.0</td>\n",

" <td>NaN</td>\n",

" <td>replacement therapy</td>\n",

" </tr>\n",

" <tr>\n",

" <th>...</th>\n",

" <td>...</td>\n",

" <td>...</td>\n",

" <td>...</td>\n",

" <td>...</td>\n",

" <td>...</td>\n",

" <td>...</td>\n",

" <td>...</td>\n",

" <td>...</td>\n",

" <td>...</td>\n",

" <td>...</td>\n",

" <td>...</td>\n",

" <td>...</td>\n",

" <td>...</td>\n",

" <td>...</td>\n",

" <td>...</td>\n",

" <td>...</td>\n",

" <td>...</td>\n",

" <td>...</td>\n",

" <td>...</td>\n",

" <td>...</td>\n",

" <td>...</td>\n",

" </tr>\n",

" <tr>\n",

" <th>9153</th>\n",

" <td>64.0</td>\n",

" <td>M</td>\n",

" <td>f</td>\n",

" <td>f</td>\n",

" <td>f</td>\n",

" <td>f</td>\n",

" <td>f</td>\n",

" <td>f</td>\n",

" <td>f</td>\n",

" <td>f</td>\n",

" <td>...</td>\n",

" <td>f</td>\n",

" <td>f</td>\n",

" <td>f</td>\n",

" <td>0.810000</td>\n",

" <td>NaN</td>\n",

" <td>31.0</td>\n",

" <td>0.55</td>\n",

" <td>56.0</td>\n",

" <td>NaN</td>\n",

" <td>general health</td>\n",

" </tr>\n",

" <tr>\n",

" <th>9157</th>\n",

" <td>60.0</td>\n",

" <td>M</td>\n",

" <td>f</td>\n",

" <td>f</td>\n",

" <td>t</td>\n",

" <td>f</td>\n",

" <td>f</td>\n",

" <td>f</td>\n",

" <td>f</td>\n",

" <td>f</td>\n",

" <td>...</td>\n",

" <td>f</td>\n",

" <td>f</td>\n",

" <td>f</td>\n",

" <td>0.180000</td>\n",

" <td>NaN</td>\n",

" <td>28.0</td>\n",

" <td>0.87</td>\n",

" <td>32.0</td>\n",

" <td>NaN</td>\n",

" <td>general health</td>\n",

" </tr>\n",

" <tr>\n",

" <th>9158</th>\n",

" <td>64.0</td>\n",

" <td>M</td>\n",

" <td>f</td>\n",

" <td>f</td>\n",

" <td>f</td>\n",

" <td>f</td>\n",

" <td>f</td>\n",

" <td>f</td>\n",

" <td>f</td>\n",

" <td>t</td>\n",

" <td>...</td>\n",

" <td>f</td>\n",

" <td>f</td>\n",

" <td>f</td>\n",

" <td>NaN</td>\n",

" <td>NaN</td>\n",

" <td>44.0</td>\n",

" <td>0.53</td>\n",

" <td>83.0</td>\n",

" <td>NaN</td>\n",

" <td>binding protein</td>\n",

" </tr>\n",

" <tr>\n",

" <th>9162</th>\n",

" <td>36.0</td>\n",

" <td>F</td>\n",

" <td>f</td>\n",

" <td>f</td>\n",

" <td>f</td>\n",

" <td>f</td>\n",

" <td>f</td>\n",

" <td>f</td>\n",

" <td>f</td>\n",

" <td>f</td>\n",

" <td>...</td>\n",

" <td>f</td>\n",

" <td>f</td>\n",

" <td>f</td>\n",

" <td>NaN</td>\n",

" <td>NaN</td>\n",

" <td>84.0</td>\n",

" <td>1.26</td>\n",

" <td>67.0</td>\n",

" <td>NaN</td>\n",

" <td>binding protein</td>\n",

" </tr>\n",

" <tr>\n",

" <th>9169</th>\n",

" <td>69.0</td>\n",

" <td>M</td>\n",

" <td>f</td>\n",

" <td>f</td>\n",

" <td>f</td>\n",

" <td>f</td>\n",

" <td>f</td>\n",

" <td>f</td>\n",

" <td>f</td>\n",

" <td>f</td>\n",

" <td>...</td>\n",

" <td>f</td>\n",

" <td>f</td>\n",

" <td>f</td>\n",

" <td>NaN</td>\n",

" <td>NaN</td>\n",

" <td>113.0</td>\n",

" <td>1.27</td>\n",

" <td>89.0</td>\n",

" <td>NaN</td>\n",

" <td>binding protein</td>\n",

" </tr>\n",

" </tbody>\n",

"</table>\n",

"<p>2223 rows × 23 columns</p>\n",

"</div>\n",

" <button class=\"colab-df-convert\" onclick=\"convertToInteractive('df-fe9d8c19-09b5-4db6-befd-9cd4a9c3e49d')\"\n",

" title=\"Convert this dataframe to an interactive table.\"\n",

" style=\"display:none;\">\n",

" \n",

" <svg xmlns=\"http://www.w3.org/2000/svg\" height=\"24px\"viewBox=\"0 0 24 24\"\n",

" width=\"24px\">\n",

" <path d=\"M0 0h24v24H0V0z\" fill=\"none\"/>\n",

" <path d=\"M18.56 5.44l.94 2.06.94-2.06 2.06-.94-2.06-.94-.94-2.06-.94 2.06-2.06.94zm-11 1L8.5 8.5l.94-2.06 2.06-.94-2.06-.94L8.5 2.5l-.94 2.06-2.06.94zm10 10l.94 2.06.94-2.06 2.06-.94-2.06-.94-.94-2.06-.94 2.06-2.06.94z\"/><path d=\"M17.41 7.96l-1.37-1.37c-.4-.4-.92-.59-1.43-.59-.52 0-1.04.2-1.43.59L10.3 9.45l-7.72 7.72c-.78.78-.78 2.05 0 2.83L4 21.41c.39.39.9.59 1.41.59.51 0 1.02-.2 1.41-.59l7.78-7.78 2.81-2.81c.8-.78.8-2.07 0-2.86zM5.41 20L4 18.59l7.72-7.72 1.47 1.35L5.41 20z\"/>\n",

" </svg>\n",

" </button>\n",

" \n",

" <style>\n",

" .colab-df-container {\n",

" display:flex;\n",

" flex-wrap:wrap;\n",

" gap: 12px;\n",

" }\n",

"\n",

" .colab-df-convert {\n",

" background-color: #E8F0FE;\n",

" border: none;\n",

" border-radius: 50%;\n",

" cursor: pointer;\n",

" display: none;\n",

" fill: #1967D2;\n",

" height: 32px;\n",

" padding: 0 0 0 0;\n",

" width: 32px;\n",

" }\n",

"\n",

" .colab-df-convert:hover {\n",

" background-color: #E2EBFA;\n",

" box-shadow: 0px 1px 2px rgba(60, 64, 67, 0.3), 0px 1px 3px 1px rgba(60, 64, 67, 0.15);\n",

" fill: #174EA6;\n",

" }\n",

"\n",

" [theme=dark] .colab-df-convert {\n",

" background-color: #3B4455;\n",

" fill: #D2E3FC;\n",

" }\n",

"\n",

" [theme=dark] .colab-df-convert:hover {\n",

" background-color: #434B5C;\n",

" box-shadow: 0px 1px 3px 1px rgba(0, 0, 0, 0.15);\n",

" filter: drop-shadow(0px 1px 2px rgba(0, 0, 0, 0.3));\n",

" fill: #FFFFFF;\n",

" }\n",

" </style>\n",

"\n",

" <script>\n",

" const buttonEl =\n",

" document.querySelector('#df-fe9d8c19-09b5-4db6-befd-9cd4a9c3e49d button.colab-df-convert');\n",

" buttonEl.style.display =\n",

" google.colab.kernel.accessAllowed ? 'block' : 'none';\n",

"\n",

" async function convertToInteractive(key) {\n",

" const element = document.querySelector('#df-fe9d8c19-09b5-4db6-befd-9cd4a9c3e49d');\n",

" const dataTable =\n",

" await google.colab.kernel.invokeFunction('convertToInteractive',\n",

" [key], {});\n",

" if (!dataTable) return;\n",

"\n",

" const docLinkHtml = 'Like what you see? Visit the ' +\n",

" '<a target=\"\_blank\" href=https://colab.research.google.com/notebooks/data\_table.ipynb>data table notebook</a>'\n",

" + ' to learn more about interactive tables.';\n",

" element.innerHTML = '';\n",

" dataTable['output\_type'] = 'display\_data';\n",

" await google.colab.output.renderOutput(dataTable, element);\n",

" const docLink = document.createElement('div');\n",

" docLink.innerHTML = docLinkHtml;\n",

" element.appendChild(docLink);\n",

" }\n",

" </script>\n",

" </div>\n",

" </div>\n",

" "

]

},

"metadata": {},

"execution\_count": 27

}

],

"source": [

"data"

]

},

{

"cell\_type": "markdown",

"metadata": {

"id": "ngq7ttEXkqVN"

},

"source": [

"Splitting the data values as x and y"

]

},

{

"cell\_type": "code",

"execution\_count": null,

"metadata": {

"id": "bFlI340quhCy"

},

"outputs": [],

"source": [

"x=data.iloc[:,0:-1]\n",

"y= data.iloc[:,-1]"

]

},

{

"cell\_type": "code",

"execution\_count": null,

"metadata": {

"colab": {

"base\_uri": "https://localhost:8080/"

},

"id": "bKzmJvmku9BB",

"outputId": "a194228e-06c5-4f14-88b3-a2aceaa97758"

},

"outputs": [

{

"output\_type": "execute\_result",

"data": {

"text/plain": [

"age 0\n",

"sex 90\n",

"on\_thyroxine 0\n",

"query\_on\_thyroxine 0\n",

"on\_antithyroid\_meds 0\n",

"sick 0\n",

"pregnant 0\n",

"thyroid\_surgery 0\n",

"I131\_treatment 0\n",

"query\_hypothyroid 0\n",

"query\_hyperthyroid 0\n",

"lithium 0\n",

"goitre 0\n",

"tumor 0\n",

"hypopituitary 0\n",

"psych 0\n",

"TSH 150\n",

"T3 594\n",

"TT4 97\n",

"T4U 178\n",

"FTI 177\n",

"TBG 2125\n",

"target 0\n",

"dtype: int64"

]

},

"metadata": {},

"execution\_count": 29

}

],

"source": [

"data.isnull().sum()"

]

},

{

"cell\_type": "code",

"execution\_count": null,

"metadata": {

"colab": {

"base\_uri": "https://localhost:8080/"

},

"id": "-SEYKlkxvDZs",

"outputId": "e6721e01-d2e2-46c5-cb06-bf0a81aaadd8"

},

"outputs": [

{

"output\_type": "execute\_result",

"data": {

"text/plain": [

"array(['F', 'M', nan], dtype=object)"

]

},

"metadata": {},

"execution\_count": 30

}

],

"source": [

"x['sex'].unique()"

]

},

{

"cell\_type": "code",

"execution\_count": null,

"metadata": {

"id": "FVVtnhksvYzu"

},

"outputs": [],

"source": [

"x['sex'].replace(np.nan, 'F', inplace=True)"

]

},

{

"cell\_type": "code",

"execution\_count": null,

"metadata": {

"colab": {

"base\_uri": "https://localhost:8080/"

},

"id": "-57Uxi\_UvkIr",

"outputId": "f674cc50-cf5e-4c4f-9559-4683017d0d3b"

},

"outputs": [

{

"output\_type": "execute\_result",

"data": {

"text/plain": [

"F 1687\n",

"M 536\n",

"Name: sex, dtype: int64"

]

},

"metadata": {},

"execution\_count": 32

}

],

"source": [

"x['sex'].value\_counts()"

]

},

{

"cell\_type": "code",

"execution\_count": null,

"metadata": {

"colab": {

"base\_uri": "https://localhost:8080/"

},

"id": "mqjBEKEevqeK",

"outputId": "4ab916e5-5429-4d0f-e010-d3b6fa4fb7d5"

},

"outputs": [

{

"output\_type": "execute\_result",

"data": {

"text/plain": [

"age 0\n",

"sex 0\n",

"on\_thyroxine 0\n",

"query\_on\_thyroxine 0\n",

"on\_antithyroid\_meds 0\n",

"sick 0\n",

"pregnant 0\n",

"thyroid\_surgery 0\n",

"I131\_treatment 0\n",

"query\_hypothyroid 0\n",

"query\_hyperthyroid 0\n",

"lithium 0\n",

"goitre 0\n",

"tumor 0\n",

"hypopituitary 0\n",

"psych 0\n",

"TSH 150\n",

"T3 594\n",

"TT4 97\n",

"T4U 178\n",

"FTI 177\n",

"TBG 2125\n",

"dtype: int64"

]

},

"metadata": {},

"execution\_count": 33

}

],

"source": [

"x.isnull().sum()"

]

},

{

"cell\_type": "code",

"execution\_count": null,

"metadata": {

"colab": {

"base\_uri": "https://localhost:8080/"

},

"id": "zX9zLjK2vwyT",

"outputId": "fe454f9e-2039-46d5-f38e-71fcfab4ae6b"

},

"outputs": [

{

"output\_type": "stream",

"name": "stdout",

"text": [

"<class 'pandas.core.frame.DataFrame'>\n",

"Int64Index: 2223 entries, 4 to 9169\n",

"Data columns (total 23 columns):\n",

" # Column Non-Null Count Dtype \n",

"--- ------ -------------- ----- \n",

" 0 age 2223 non-null float64\n",

" 1 sex 2133 non-null object \n",

" 2 on\_thyroxine 2223 non-null object \n",

" 3 query\_on\_thyroxine 2223 non-null object \n",

" 4 on\_antithyroid\_meds 2223 non-null object \n",

" 5 sick 2223 non-null object \n",

" 6 pregnant 2223 non-null object \n",

" 7 thyroid\_surgery 2223 non-null object \n",

" 8 I131\_treatment 2223 non-null object \n",

" 9 query\_hypothyroid 2223 non-null object \n",

" 10 query\_hyperthyroid 2223 non-null object \n",

" 11 lithium 2223 non-null object \n",

" 12 goitre 2223 non-null object \n",

" 13 tumor 2223 non-null object \n",

" 14 hypopituitary 2223 non-null object \n",

" 15 psych 2223 non-null object \n",

" 16 TSH 2073 non-null float64\n",

" 17 T3 1629 non-null float64\n",

" 18 TT4 2126 non-null float64\n",

" 19 T4U 2045 non-null float64\n",

" 20 FTI 2046 non-null float64\n",

" 21 TBG 98 non-null float64\n",

" 22 target 2223 non-null object \n",

"dtypes: float64(7), object(16)\n",

"memory usage: 416.8+ KB\n"

]

}

],

"source": [

"data.info()"

]

},

{

"cell\_type": "code",

"execution\_count": null,

"metadata": {

"id": "Lt\_08b5vv92J"

},

"outputs": [],

"source": [

"x['age']=x['age'].astype('float')\n",

"x['TSH']=x['TSH'].astype('float')\n",

"x['T3']=x['T3'].astype('float')\n",

"x['TT4']=x['TT4'].astype('float')\n",

"x['T4U']=x['T4U'].astype('float')\n",

"x['FTI']=x['FTI'].astype('float')\n",

"x['TBG']=x['TBG'].astype('float')"

]

},

{

"cell\_type": "markdown",

"metadata": {

"id": "4yrhwGYLlEfW"

},

"source": [

"Converting Categorical to numerical values"

]

},

{

"cell\_type": "code",

"execution\_count": null,

"metadata": {

"colab": {

"base\_uri": "https://localhost:8080/"

},

"id": "2QUufog5wn9J",

"outputId": "3172d444-0172-4a5f-deb4-566d05848155"

},

"outputs": [

{

"output\_type": "stream",

"name": "stderr",

"text": [

"<ipython-input-36-d44f8bcef9c7>:3: DeprecationWarning: In a future version, `df.iloc[:, i] = newvals` will attempt to set the values inplace instead of always setting a new array. To retain the old behavior, use either `df[df.columns[i]] = newvals` or, if columns are non-unique, `df.isetitem(i, newvals)`\n",

" x.iloc[:, 1:16] = ordinal\_encoder.fit\_transform(x.iloc[:, 1:16])\n"

]

}

],

"source": [

"from sklearn.preprocessing import OrdinalEncoder, LabelEncoder\n",

"ordinal\_encoder = OrdinalEncoder(dtype = 'int64')\n",

"x.iloc[:, 1:16] = ordinal\_encoder.fit\_transform(x.iloc[:, 1:16])"

]

},

{

"cell\_type": "code",

"execution\_count": null,

"metadata": {

"colab": {

"base\_uri": "https://localhost:8080/",

"height": 300

},

"id": "8rg4MtIYxWGd",

"outputId": "84ea8fe6-b990-43e9-adcd-5fb47d7ef3df"

},

"outputs": [

{

"output\_type": "execute\_result",

"data": {

"text/plain": [

" age sex on\_thyroxine query\_on\_thyroxine on\_antithyroid\_meds sick \\\n",

"4 32.0 0 0 0 0 0 \n",

"18 63.0 0 1 0 0 1 \n",

"32 41.0 1 0 0 0 0 \n",

"33 71.0 0 1 0 0 0 \n",

"39 55.0 0 1 0 0 0 \n",

"\n",

" pregnant thyroid\_surgery I131\_treatment query\_hypothyroid ... goitre \\\n",

"4 0 0 0 0 ... 0 \n",

"18 0 0 0 0 ... 0 \n",

"32 0 0 0 0 ... 0 \n",

"33 0 0 0 0 ... 0 \n",

"39 0 0 0 1 ... 0 \n",

"\n",

" tumor hypopituitary psych TSH T3 TT4 T4U FTI TBG \n",

"4 0 0 0 NaN NaN NaN NaN NaN 36.0 \n",

"18 0 0 0 68.000000 NaN 48.0 1.02 47.0 NaN \n",

"32 0 0 0 0.050000 1.6 39.0 1.00 39.0 NaN \n",

"33 0 0 0 0.050000 NaN 126.0 1.38 91.0 NaN \n",

"39 0 0 0 9.599999 2.4 136.0 1.48 92.0 NaN \n",

"\n",

"[5 rows x 22 columns]"

],

"text/html": [

"\n",

" <div id=\"df-11bd6d41-511b-4034-b5fe-86e464f2e101\">\n",

" <div class=\"colab-df-container\">\n",

" <div>\n",

"<style scoped>\n",

" .dataframe tbody tr th:only-of-type {\n",

" vertical-align: middle;\n",

" }\n",

"\n",

" .dataframe tbody tr th {\n",

" vertical-align: top;\n",

" }\n",

"\n",

" .dataframe thead th {\n",

" text-align: right;\n",

" }\n",

"</style>\n",

"<table border=\"1\" class=\"dataframe\">\n",

" <thead>\n",

" <tr style=\"text-align: right;\">\n",

" <th></th>\n",

" <th>age</th>\n",

" <th>sex</th>\n",

" <th>on\_thyroxine</th>\n",

" <th>query\_on\_thyroxine</th>\n",

" <th>on\_antithyroid\_meds</th>\n",

" <th>sick</th>\n",

" <th>pregnant</th>\n",

" <th>thyroid\_surgery</th>\n",

" <th>I131\_treatment</th>\n",

" <th>query\_hypothyroid</th>\n",

" <th>...</th>\n",

" <th>goitre</th>\n",

" <th>tumor</th>\n",

" <th>hypopituitary</th>\n",

" <th>psych</th>\n",

" <th>TSH</th>\n",

" <th>T3</th>\n",

" <th>TT4</th>\n",

" <th>T4U</th>\n",

" <th>FTI</th>\n",

" <th>TBG</th>\n",

" </tr>\n",

" </thead>\n",

" <tbody>\n",

" <tr>\n",

" <th>4</th>\n",

" <td>32.0</td>\n",

" <td>0</td>\n",

" <td>0</td>\n",

" <td>0</td>\n",

" <td>0</td>\n",

" <td>0</td>\n",

" <td>0</td>\n",

" <td>0</td>\n",

" <td>0</td>\n",

" <td>0</td>\n",

" <td>...</td>\n",

" <td>0</td>\n",

" <td>0</td>\n",

" <td>0</td>\n",

" <td>0</td>\n",

" <td>NaN</td>\n",

" <td>NaN</td>\n",

" <td>NaN</td>\n",

" <td>NaN</td>\n",

" <td>NaN</td>\n",

" <td>36.0</td>\n",

" </tr>\n",

" <tr>\n",

" <th>18</th>\n",

" <td>63.0</td>\n",

" <td>0</td>\n",

" <td>1</td>\n",

" <td>0</td>\n",

" <td>0</td>\n",

" <td>1</td>\n",

" <td>0</td>\n",

" <td>0</td>\n",

" <td>0</td>\n",

" <td>0</td>\n",

" <td>...</td>\n",

" <td>0</td>\n",

" <td>0</td>\n",

" <td>0</td>\n",

" <td>0</td>\n",

" <td>68.000000</td>\n",

" <td>NaN</td>\n",

" <td>48.0</td>\n",

" <td>1.02</td>\n",

" <td>47.0</td>\n",

" <td>NaN</td>\n",

" </tr>\n",

" <tr>\n",

" <th>32</th>\n",

" <td>41.0</td>\n",

" <td>1</td>\n",

" <td>0</td>\n",

" <td>0</td>\n",

" <td>0</td>\n",

" <td>0</td>\n",

" <td>0</td>\n",

" <td>0</td>\n",

" <td>0</td>\n",

" <td>0</td>\n",

" <td>...</td>\n",

" <td>0</td>\n",

" <td>0</td>\n",

" <td>0</td>\n",

" <td>0</td>\n",

" <td>0.050000</td>\n",

" <td>1.6</td>\n",

" <td>39.0</td>\n",

" <td>1.00</td>\n",

" <td>39.0</td>\n",

" <td>NaN</td>\n",

" </tr>\n",

" <tr>\n",

" <th>33</th>\n",

" <td>71.0</td>\n",

" <td>0</td>\n",

" <td>1</td>\n",

" <td>0</td>\n",

" <td>0</td>\n",

" <td>0</td>\n",

" <td>0</td>\n",

" <td>0</td>\n",

" <td>0</td>\n",

" <td>0</td>\n",

" <td>...</td>\n",

" <td>0</td>\n",

" <td>0</td>\n",

" <td>0</td>\n",

" <td>0</td>\n",

" <td>0.050000</td>\n",

" <td>NaN</td>\n",

" <td>126.0</td>\n",

" <td>1.38</td>\n",

" <td>91.0</td>\n",

" <td>NaN</td>\n",

" </tr>\n",

" <tr>\n",

" <th>39</th>\n",

" <td>55.0</td>\n",

" <td>0</td>\n",

" <td>1</td>\n",

" <td>0</td>\n",

" <td>0</td>\n",

" <td>0</td>\n",

" <td>0</td>\n",

" <td>0</td>\n",

" <td>0</td>\n",

" <td>1</td>\n",

" <td>...</td>\n",

" <td>0</td>\n",

" <td>0</td>\n",

" <td>0</td>\n",

" <td>0</td>\n",

" <td>9.599999</td>\n",

" <td>2.4</td>\n",

" <td>136.0</td>\n",

" <td>1.48</td>\n",

" <td>92.0</td>\n",

" <td>NaN</td>\n",

" </tr>\n",

" </tbody>\n",

"</table>\n",

"<p>5 rows × 22 columns</p>\n",

"</div>\n",

" <button class=\"colab-df-convert\" onclick=\"convertToInteractive('df-11bd6d41-511b-4034-b5fe-86e464f2e101')\"\n",

" title=\"Convert this dataframe to an interactive table.\"\n",

" style=\"display:none;\">\n",

" \n",

" <svg xmlns=\"http://www.w3.org/2000/svg\" height=\"24px\"viewBox=\"0 0 24 24\"\n",

" width=\"24px\">\n",

" <path d=\"M0 0h24v24H0V0z\" fill=\"none\"/>\n",

" <path d=\"M18.56 5.44l.94 2.06.94-2.06 2.06-.94-2.06-.94-.94-2.06-.94 2.06-2.06.94zm-11 1L8.5 8.5l.94-2.06 2.06-.94-2.06-.94L8.5 2.5l-.94 2.06-2.06.94zm10 10l.94 2.06.94-2.06 2.06-.94-2.06-.94-.94-2.06-.94 2.06-2.06.94z\"/><path d=\"M17.41 7.96l-1.37-1.37c-.4-.4-.92-.59-1.43-.59-.52 0-1.04.2-1.43.59L10.3 9.45l-7.72 7.72c-.78.78-.78 2.05 0 2.83L4 21.41c.39.39.9.59 1.41.59.51 0 1.02-.2 1.41-.59l7.78-7.78 2.81-2.81c.8-.78.8-2.07 0-2.86zM5.41 20L4 18.59l7.72-7.72 1.47 1.35L5.41 20z\"/>\n",

" </svg>\n",

" </button>\n",

" \n",

" <style>\n",

" .colab-df-container {\n",

" display:flex;\n",

" flex-wrap:wrap;\n",

" gap: 12px;\n",

" }\n",

"\n",

" .colab-df-convert {\n",

" background-color: #E8F0FE;\n",

" border: none;\n",

" border-radius: 50%;\n",

" cursor: pointer;\n",

" display: none;\n",

" fill: #1967D2;\n",

" height: 32px;\n",

" padding: 0 0 0 0;\n",

" width: 32px;\n",

" }\n",

"\n",

" .colab-df-convert:hover {\n",

" background-color: #E2EBFA;\n",

" box-shadow: 0px 1px 2px rgba(60, 64, 67, 0.3), 0px 1px 3px 1px rgba(60, 64, 67, 0.15);\n",

" fill: #174EA6;\n",

" }\n",

"\n",

" [theme=dark] .colab-df-convert {\n",

" background-color: #3B4455;\n",

" fill: #D2E3FC;\n",

" }\n",

"\n",

" [theme=dark] .colab-df-convert:hover {\n",

" background-color: #434B5C;\n",

" box-shadow: 0px 1px 3px 1px rgba(0, 0, 0, 0.15);\n",

" filter: drop-shadow(0px 1px 2px rgba(0, 0, 0, 0.3));\n",

" fill: #FFFFFF;\n",

" }\n",

" </style>\n",

"\n",

" <script>\n",

" const buttonEl =\n",

" document.querySelector('#df-11bd6d41-511b-4034-b5fe-86e464f2e101 button.colab-df-convert');\n",

" buttonEl.style.display =\n",

" google.colab.kernel.accessAllowed ? 'block' : 'none';\n",

"\n",

" async function convertToInteractive(key) {\n",

" const element = document.querySelector('#df-11bd6d41-511b-4034-b5fe-86e464f2e101');\n",

" const dataTable =\n",

" await google.colab.kernel.invokeFunction('convertToInteractive',\n",

" [key], {});\n",

" if (!dataTable) return;\n",

"\n",

" const docLinkHtml = 'Like what you see? Visit the ' +\n",

" '<a target=\"\_blank\" href=https://colab.research.google.com/notebooks/data\_table.ipynb>data table notebook</a>'\n",

" + ' to learn more about interactive tables.';\n",

" element.innerHTML = '';\n",

" dataTable['output\_type'] = 'display\_data';\n",

" await google.colab.output.renderOutput(dataTable, element);\n",

" const docLink = document.createElement('div');\n",

" docLink.innerHTML = docLinkHtml;\n",

" element.appendChild(docLink);\n",

" }\n",

" </script>\n",

" </div>\n",

" </div>\n",

" "

]

},

"metadata": {},

"execution\_count": 37

}

],

"source": [

"x.head()"

]

},

{

"cell\_type": "code",

"execution\_count": null,

"metadata": {

"colab": {

"base\_uri": "https://localhost:8080/",

"height": 300

},

"id": "6wa8lB8Txa0S",

"outputId": "70354eab-0da0-4efb-b0d1-5ae4b44e6841"

},

"outputs": [

{

"output\_type": "execute\_result",

"data": {

"text/plain": [

" age sex on\_thyroxine query\_on\_thyroxine on\_antithyroid\_meds sick \\\n",

"4 32.0 0 0 0 0 0 \n",

"18 63.0 0 1 0 0 1 \n",

"32 41.0 1 0 0 0 0 \n",

"33 71.0 0 1 0 0 0 \n",

"39 55.0 0 1 0 0 0 \n",

"\n",

" pregnant thyroid\_surgery I131\_treatment query\_hypothyroid ... goitre \\\n",

"4 0 0 0 0 ... 0 \n",

"18 0 0 0 0 ... 0 \n",

"32 0 0 0 0 ... 0 \n",

"33 0 0 0 0 ... 0 \n",

"39 0 0 0 1 ... 0 \n",

"\n",

" tumor hypopituitary psych TSH T3 TT4 T4U FTI TBG \n",

"4 0 0 0 0 0 0 0 0 36.0 \n",

"18 0 0 0 68.0 0 48.0 1.02 47.0 0 \n",

"32 0 0 0 0.05 1.6 39.0 1.0 39.0 0 \n",

"33 0 0 0 0.05 0 126.0 1.38 91.0 0 \n",

"39 0 0 0 9.599999 2.4 136.0 1.48 92.0 0 \n",

"\n",

"[5 rows x 22 columns]"

],

"text/html": [

"\n",

" <div id=\"df-09b8fa28-09da-48e9-b271-69d79bf9e5ad\">\n",

" <div class=\"colab-df-container\">\n",

" <div>\n",

"<style scoped>\n",

" .dataframe tbody tr th:only-of-type {\n",

" vertical-align: middle;\n",

" }\n",

"\n",

" .dataframe tbody tr th {\n",

" vertical-align: top;\n",

" }\n",

"\n",

" .dataframe thead th {\n",

" text-align: right;\n",

" }\n",

"</style>\n",

"<table border=\"1\" class=\"dataframe\">\n",

" <thead>\n",

" <tr style=\"text-align: right;\">\n",

" <th></th>\n",

" <th>age</th>\n",

" <th>sex</th>\n",

" <th>on\_thyroxine</th>\n",

" <th>query\_on\_thyroxine</th>\n",

" <th>on\_antithyroid\_meds</th>\n",

" <th>sick</th>\n",

" <th>pregnant</th>\n",

" <th>thyroid\_surgery</th>\n",

" <th>I131\_treatment</th>\n",

" <th>query\_hypothyroid</th>\n",

" <th>...</th>\n",

" <th>goitre</th>\n",

" <th>tumor</th>\n",

" <th>hypopituitary</th>\n",

" <th>psych</th>\n",

" <th>TSH</th>\n",

" <th>T3</th>\n",

" <th>TT4</th>\n",

" <th>T4U</th>\n",

" <th>FTI</th>\n",

" <th>TBG</th>\n",

" </tr>\n",

" </thead>\n",

" <tbody>\n",

" <tr>\n",

" <th>4</th>\n",

" <td>32.0</td>\n",

" <td>0</td>\n",

" <td>0</td>\n",

" <td>0</td>\n",

" <td>0</td>\n",

" <td>0</td>\n",

" <td>0</td>\n",

" <td>0</td>\n",

" <td>0</td>\n",

" <td>0</td>\n",

" <td>...</td>\n",

" <td>0</td>\n",

" <td>0</td>\n",

" <td>0</td>\n",

" <td>0</td>\n",

" <td>0</td>\n",

" <td>0</td>\n",

" <td>0</td>\n",

" <td>0</td>\n",

" <td>0</td>\n",

" <td>36.0</td>\n",

" </tr>\n",

" <tr>\n",

" <th>18</th>\n",

" <td>63.0</td>\n",

" <td>0</td>\n",

" <td>1</td>\n",

" <td>0</td>\n",

" <td>0</td>\n",

" <td>1</td>\n",

" <td>0</td>\n",

" <td>0</td>\n",

" <td>0</td>\n",

" <td>0</td>\n",

" <td>...</td>\n",

" <td>0</td>\n",

" <td>0</td>\n",

" <td>0</td>\n",

" <td>0</td>\n",

" <td>68.0</td>\n",

" <td>0</td>\n",

" <td>48.0</td>\n",

" <td>1.02</td>\n",

" <td>47.0</td>\n",

" <td>0</td>\n",

" </tr>\n",

" <tr>\n",

" <th>32</th>\n",

" <td>41.0</td>\n",

" <td>1</td>\n",

" <td>0</td>\n",

" <td>0</td>\n",

" <td>0</td>\n",

" <td>0</td>\n",

" <td>0</td>\n",

" <td>0</td>\n",

" <td>0</td>\n",

" <td>0</td>\n",

" <td>...</td>\n",

" <td>0</td>\n",

" <td>0</td>\n",

" <td>0</td>\n",

" <td>0</td>\n",

" <td>0.05</td>\n",

" <td>1.6</td>\n",

" <td>39.0</td>\n",

" <td>1.0</td>\n",

" <td>39.0</td>\n",

" <td>0</td>\n",

" </tr>\n",

" <tr>\n",

" <th>33</th>\n",

" <td>71.0</td>\n",

" <td>0</td>\n",

" <td>1</td>\n",

" <td>0</td>\n",

" <td>0</td>\n",

" <td>0</td>\n",

" <td>0</td>\n",

" <td>0</td>\n",

" <td>0</td>\n",

" <td>0</td>\n",

" <td>...</td>\n",

" <td>0</td>\n",

" <td>0</td>\n",

" <td>0</td>\n",

" <td>0</td>\n",

" <td>0.05</td>\n",

" <td>0</td>\n",

" <td>126.0</td>\n",

" <td>1.38</td>\n",

" <td>91.0</td>\n",

" <td>0</td>\n",

" </tr>\n",

" <tr>\n",

" <th>39</th>\n",

" <td>55.0</td>\n",

" <td>0</td>\n",

" <td>1</td>\n",

" <td>0</td>\n",

" <td>0</td>\n",

" <td>0</td>\n",

" <td>0</td>\n",

" <td>0</td>\n",

" <td>0</td>\n",

" <td>1</td>\n",

" <td>...</td>\n",

" <td>0</td>\n",

" <td>0</td>\n",

" <td>0</td>\n",

" <td>0</td>\n",

" <td>9.599999</td>\n",

" <td>2.4</td>\n",

" <td>136.0</td>\n",

" <td>1.48</td>\n",

" <td>92.0</td>\n",

" <td>0</td>\n",

" </tr>\n",

" </tbody>\n",

"</table>\n",

"<p>5 rows × 22 columns</p>\n",

"</div>\n",

" <button class=\"colab-df-convert\" onclick=\"convertToInteractive('df-09b8fa28-09da-48e9-b271-69d79bf9e5ad')\"\n",

" title=\"Convert this dataframe to an interactive table.\"\n",

" style=\"display:none;\">\n",

" \n",

" <svg xmlns=\"http://www.w3.org/2000/svg\" height=\"24px\"viewBox=\"0 0 24 24\"\n",

" width=\"24px\">\n",

" <path d=\"M0 0h24v24H0V0z\" fill=\"none\"/>\n",

" <path d=\"M18.56 5.44l.94 2.06.94-2.06 2.06-.94-2.06-.94-.94-2.06-.94 2.06-2.06.94zm-11 1L8.5 8.5l.94-2.06 2.06-.94-2.06-.94L8.5 2.5l-.94 2.06-2.06.94zm10 10l.94 2.06.94-2.06 2.06-.94-2.06-.94-.94-2.06-.94 2.06-2.06.94z\"/><path d=\"M17.41 7.96l-1.37-1.37c-.4-.4-.92-.59-1.43-.59-.52 0-1.04.2-1.43.59L10.3 9.45l-7.72 7.72c-.78.78-.78 2.05 0 2.83L4 21.41c.39.39.9.59 1.41.59.51 0 1.02-.2 1.41-.59l7.78-7.78 2.81-2.81c.8-.78.8-2.07 0-2.86zM5.41 20L4 18.59l7.72-7.72 1.47 1.35L5.41 20z\"/>\n",

" </svg>\n",

" </button>\n",

" \n",

" <style>\n",

" .colab-df-container {\n",

" display:flex;\n",

" flex-wrap:wrap;\n",

" gap: 12px;\n",

" }\n",

"\n",

" .colab-df-convert {\n",

" background-color: #E8F0FE;\n",

" border: none;\n",

" border-radius: 50%;\n",

" cursor: pointer;\n",

" display: none;\n",

" fill: #1967D2;\n",

" height: 32px;\n",

" padding: 0 0 0 0;\n",

" width: 32px;\n",

" }\n",

"\n",

" .colab-df-convert:hover {\n",

" background-color: #E2EBFA;\n",

" box-shadow: 0px 1px 2px rgba(60, 64, 67, 0.3), 0px 1px 3px 1px rgba(60, 64, 67, 0.15);\n",

" fill: #174EA6;\n",

" }\n",

"\n",

" [theme=dark] .colab-df-convert {\n",

" background-color: #3B4455;\n",

" fill: #D2E3FC;\n",

" }\n",

"\n",

" [theme=dark] .colab-df-convert:hover {\n",

" background-color: #434B5C;\n",

" box-shadow: 0px 1px 3px 1px rgba(0, 0, 0, 0.15);\n",

" filter: drop-shadow(0px 1px 2px rgba(0, 0, 0, 0.3));\n",

" fill: #FFFFFF;\n",

" }\n",

" </style>\n",

"\n",

" <script>\n",

" const buttonEl =\n",

" document.querySelector('#df-09b8fa28-09da-48e9-b271-69d79bf9e5ad button.colab-df-convert');\n",

" buttonEl.style.display =\n",

" google.colab.kernel.accessAllowed ? 'block' : 'none';\n",

"\n",

" async function convertToInteractive(key) {\n",

" const element = document.querySelector('#df-09b8fa28-09da-48e9-b271-69d79bf9e5ad');\n",

" const dataTable =\n",

" await google.colab.kernel.invokeFunction('convertToInteractive',\n",

" [key], {});\n",

" if (!dataTable) return;\n",

"\n",

" const docLinkHtml = 'Like what you see? Visit the ' +\n",

" '<a target=\"\_blank\" href=https://colab.research.google.com/notebooks/data\_table.ipynb>data table notebook</a>'\n",

" + ' to learn more about interactive tables.';\n",

" element.innerHTML = '';\n",

" dataTable['output\_type'] = 'display\_data';\n",

" await google.colab.output.renderOutput(dataTable, element);\n",

" const docLink = document.createElement('div');\n",

" docLink.innerHTML = docLinkHtml;\n",

" element.appendChild(docLink);\n",

" }\n",

" </script>\n",

" </div>\n",

" </div>\n",

" "

]

},

"metadata": {},

"execution\_count": 38

}

],

"source": [

"x.replace(np.nan, '0', inplace=True)\n",

"x.head()"

]

},

{

"cell\_type": "code",

"execution\_count": null,

"metadata": {

"id": "ehGyPkt-x4Ui"

},

"outputs": [],

"source": [

"lable\_encoder = LabelEncoder()\n",

"y\_dt= lable\_encoder.fit\_transform(y)"

]

},

{

"cell\_type": "code",

"execution\_count": null,

"metadata": {

"colab": {

"base\_uri": "https://localhost:8080/",

"height": 424

},

"id": "L5ICM1NOyFik",

"outputId": "17fe8908-906a-4fdb-f8c6-b795d36f4516"

},

"outputs": [

{

"output\_type": "execute\_result",

"data": {

"text/plain": [

" target\n",

"0 5\n",

"1 4\n",

"2 5\n",

"3 1\n",

"4 6\n",

"... ...\n",

"2218 2\n",

"2219 2\n",

"2220 1\n",

"2221 1\n",

"2222 1\n",

"\n",

"[2223 rows x 1 columns]"

],

"text/html": [

"\n",

" <div id=\"df-afc4f68b-b7e6-4715-8c07-afd624c9eb4c\">\n",

" <div class=\"colab-df-container\">\n",

" <div>\n",

"<style scoped>\n",

" .dataframe tbody tr th:only-of-type {\n",

" vertical-align: middle;\n",

" }\n",

"\n",

" .dataframe tbody tr th {\n",

" vertical-align: top;\n",

" }\n",

"\n",

" .dataframe thead th {\n",

" text-align: right;\n",

" }\n",

"</style>\n",

"<table border=\"1\" class=\"dataframe\">\n",

" <thead>\n",

" <tr style=\"text-align: right;\">\n",

" <th></th>\n",

" <th>target</th>\n",

" </tr>\n",

" </thead>\n",

" <tbody>\n",

" <tr>\n",

" <th>0</th>\n",

" <td>5</td>\n",

" </tr>\n",

" <tr>\n",

" <th>1</th>\n",

" <td>4</td>\n",

" </tr>\n",

" <tr>\n",

" <th>2</th>\n",

" <td>5</td>\n",

" </tr>\n",

" <tr>\n",

" <th>3</th>\n",

" <td>1</td>\n",

" </tr>\n",

" <tr>\n",

" <th>4</th>\n",

" <td>6</td>\n",

" </tr>\n",

" <tr>\n",

" <th>...</th>\n",

" <td>...</td>\n",

" </tr>\n",

" <tr>\n",

" <th>2218</th>\n",

" <td>2</td>\n",

" </tr>\n",

" <tr>\n",

" <th>2219</th>\n",

" <td>2</td>\n",

" </tr>\n",

" <tr>\n",

" <th>2220</th>\n",

" <td>1</td>\n",

" </tr>\n",

" <tr>\n",

" <th>2221</th>\n",

" <td>1</td>\n",

" </tr>\n",

" <tr>\n",

" <th>2222</th>\n",

" <td>1</td>\n",

" </tr>\n",

" </tbody>\n",

"</table>\n",

"<p>2223 rows × 1 columns</p>\n",

"</div>\n",

" <button class=\"colab-df-convert\" onclick=\"convertToInteractive('df-afc4f68b-b7e6-4715-8c07-afd624c9eb4c')\"\n",

" title=\"Convert this dataframe to an interactive table.\"\n",

" style=\"display:none;\">\n",

" \n",

" <svg xmlns=\"http://www.w3.org/2000/svg\" height=\"24px\"viewBox=\"0 0 24 24\"\n",

" width=\"24px\">\n",

" <path d=\"M0 0h24v24H0V0z\" fill=\"none\"/>\n",

" <path d=\"M18.56 5.44l.94 2.06.94-2.06 2.06-.94-2.06-.94-.94-2.06-.94 2.06-2.06.94zm-11 1L8.5 8.5l.94-2.06 2.06-.94-2.06-.94L8.5 2.5l-.94 2.06-2.06.94zm10 10l.94 2.06.94-2.06 2.06-.94-2.06-.94-.94-2.06-.94 2.06-2.06.94z\"/><path d=\"M17.41 7.96l-1.37-1.37c-.4-.4-.92-.59-1.43-.59-.52 0-1.04.2-1.43.59L10.3 9.45l-7.72 7.72c-.78.78-.78 2.05 0 2.83L4 21.41c.39.39.9.59 1.41.59.51 0 1.02-.2 1.41-.59l7.78-7.78 2.81-2.81c.8-.78.8-2.07 0-2.86zM5.41 20L4 18.59l7.72-7.72 1.47 1.35L5.41 20z\"/>\n",

" </svg>\n",

" </button>\n",

" \n",

" <style>\n",

" .colab-df-container {\n",

" display:flex;\n",

" flex-wrap:wrap;\n",

" gap: 12px;\n",

" }\n",

"\n",

" .colab-df-convert {\n",

" background-color: #E8F0FE;\n",

" border: none;\n",

" border-radius: 50%;\n",

" cursor: pointer;\n",

" display: none;\n",

" fill: #1967D2;\n",

" height: 32px;\n",

" padding: 0 0 0 0;\n",

" width: 32px;\n",

" }\n",

"\n",

" .colab-df-convert:hover {\n",

" background-color: #E2EBFA;\n",

" box-shadow: 0px 1px 2px rgba(60, 64, 67, 0.3), 0px 1px 3px 1px rgba(60, 64, 67, 0.15);\n",

" fill: #174EA6;\n",

" }\n",

"\n",

" [theme=dark] .colab-df-convert {\n",

" background-color: #3B4455;\n",

" fill: #D2E3FC;\n",

" }\n",

"\n",

" [theme=dark] .colab-df-convert:hover {\n",

" background-color: #434B5C;\n",

" box-shadow: 0px 1px 3px 1px rgba(0, 0, 0, 0.15);\n",

" filter: drop-shadow(0px 1px 2px rgba(0, 0, 0, 0.3));\n",

" fill: #FFFFFF;\n",

" }\n",

" </style>\n",

"\n",

" <script>\n",

" const buttonEl =\n",

" document.querySelector('#df-afc4f68b-b7e6-4715-8c07-afd624c9eb4c button.colab-df-convert');\n",

" buttonEl.style.display =\n",

" google.colab.kernel.accessAllowed ? 'block' : 'none';\n",

"\n",

" async function convertToInteractive(key) {\n",

" const element = document.querySelector('#df-afc4f68b-b7e6-4715-8c07-afd624c9eb4c');\n",

" const dataTable =\n",

" await google.colab.kernel.invokeFunction('convertToInteractive',\n",

" [key], {});\n",

" if (!dataTable) return;\n",

"\n",

" const docLinkHtml = 'Like what you see? Visit the ' +\n",

" '<a target=\"\_blank\" href=https://colab.research.google.com/notebooks/data\_table.ipynb>data table notebook</a>'\n",

" + ' to learn more about interactive tables.';\n",

" element.innerHTML = '';\n",

" dataTable['output\_type'] = 'display\_data';\n",

" await google.colab.output.renderOutput(dataTable, element);\n",

" const docLink = document.createElement('div');\n",

" docLink.innerHTML = docLinkHtml;\n",

" element.appendChild(docLink);\n",

" }\n",

" </script>\n",

" </div>\n",

" </div>\n",

" "

]

},

"metadata": {},

"execution\_count": 40

}

],

"source": [

"y=pd.DataFrame(y\_dt, columns=['target'])\n",

"y"

]

},

{

"cell\_type": "code",

"execution\_count": null,

"metadata": {

"colab": {

"base\_uri": "https://localhost:8080/"

},

"id": "JheAMxgMyN2O",

"outputId": "4b07a51d-c884-45d2-d2b1-c9090bd28021"

},

"outputs": [

{

"output\_type": "execute\_result",

"data": {

"text/plain": [

"target\n",

"4 0.266757\n",

"2 0.196131\n",

"1 0.169141\n",

"6 0.151147\n",

"5 0.126406\n",

"3 0.081871\n",

"0 0.008547\n",

"dtype: float64"

]

},

"metadata": {},

"execution\_count": 41

}

],

"source": [

"y.value\_counts(normalize=True)"

]

},

{

"cell\_type": "markdown",

"source": [

"# Exploration Data Analysis"

],

"metadata": {

"id": "XAPcIzjNlWeE"

}

},

{

"cell\_type": "markdown",

"metadata": {

"id": "T0osKOZflV-w"

},

"source": [

"Checking the Correlation"

]

},

{

"cell\_type": "code",

"execution\_count": null,

"metadata": {

"colab": {

"base\_uri": "https://localhost:8080/",

"height": 885

},

"id": "up0wmkEeyawn",

"outputId": "f04541a5-823e-4888-e6c6-280a0d321b3d"

},

"outputs": [

{

"output\_type": "stream",

"name": "stderr",

"text": [

"<ipython-input-42-97e8ff478756>:3: FutureWarning: The default value of numeric\_only in DataFrame.corr is deprecated. In a future version, it will default to False. Select only valid columns or specify the value of numeric\_only to silence this warning.\n",

" corrmat = x.corr()\n"

]

},

{

"output\_type": "execute\_result",

"data": {

"text/plain": [

"<Axes: >"

]

},

"metadata": {},

"execution\_count": 42

},

{

"output\_type": "display\_data",

"data": {

"text/plain": [

"<Figure size 900x800 with 2 Axes>"

],

"image/png": "\n"

},

"metadata": {}

}

],

"source": [

"import seaborn as sns\n",

"import matplotlib.pyplot as plt\n",

"corrmat = x.corr()\n",

"f,ax = plt.subplots(figsize =(9,8))\n",

"sns.heatmap(corrmat, ax= ax, cmap =\"YlGnBu\", linewidths = 0.1)"

]

},

{

"cell\_type": "markdown",

"metadata": {

"id": "mcSW8\_HOliHh"

},

"source": [

"Splitting the train and test split"

]

},

{

"cell\_type": "code",

"execution\_count": null,

"metadata": {

"id": "GXAWTG7yywwV"

},

"outputs": [],

"source": [

"from sklearn.model\_selection import train\_test\_split\n",

"x\_train,x\_test,y\_train,y\_test=train\_test\_split(x,y,test\_size=0.20,random\_state=0)"

]

},

{

"cell\_type": "code",

"execution\_count": null,

"metadata": {

"colab": {

"base\_uri": "https://localhost:8080/"

},

"id": "vR3AgxSqzKyI",

"outputId": "05ed92d1-0c84-4d34-c1e6-c169aafc9e2c"

},

"outputs": [

{

"output\_type": "execute\_result",

"data": {

"text/plain": [

"target\n",

"4 472\n",

"2 360\n",

"1 298\n",

"6 269\n",

"5 226\n",

"3 138\n",

"0 15\n",

"dtype: int64"

]

},

"metadata": {},

"execution\_count": 44

}

],

"source": [

"y\_train.value\_counts()"

]

},

{

"cell\_type": "code",

"execution\_count": null,

"metadata": {

"id": "398e4Vouzgyj"

},

"outputs": [],

"source": [

"from imblearn.over\_sampling import SMOTE\n",

"os = SMOTE(random\_state=0,k\_neighbors=1)\n",

"x\_bal,y\_bal=os.fit\_resample(x\_train,y\_train)\n",

"x\_test\_bal,y\_test\_bal=os.fit\_resample(x\_test,y\_test)"

]

},

{

"cell\_type": "code",

"execution\_count": null,

"metadata": {

"colab": {

"base\_uri": "https://localhost:8080/",

"height": 488

},

"id": "nFx9RFjydgpN",

"outputId": "19f16bb0-2d50-4fc1-9bb4-01ba714896d6"

},

"outputs": [

{

"output\_type": "execute\_result",

"data": {

"text/plain": [

" age sex on\_thyroxine query\_on\_thyroxine on\_antithyroid\_meds \\\n",

"0 48.000000 0 0 0 0 \n",

"1 72.000000 0 0 0 0 \n",

"2 31.000000 0 0 0 0 \n",

"3 14.000000 0 0 0 0 \n",

"4 54.000000 0 0 0 0 \n",

"... ... ... ... ... ... \n",

"3299 72.621678 0 1 0 0 \n",

"3300 63.726045 0 1 0 0 \n",

"3301 61.000000 0 1 0 0 \n",

"3302 58.962148 0 1 0 0 \n",

"3303 28.149989 0 1 0 0 \n",

"\n",

" sick pregnant thyroid\_surgery I131\_treatment query\_hypothyroid ... \\\n",

"0 0 0 0 0 0 ... \n",

"1 0 0 0 0 0 ... \n",

"2 0 0 0 0 0 ... \n",

"3 0 0 0 0 0 ... \n",

"4 0 0 0 0 0 ... \n",

"... ... ... ... ... ... ... \n",

"3299 0 0 0 0 0 ... \n",

"3300 0 0 0 0 0 ... \n",

"3301 0 0 0 0 0 ... \n",

"3302 0 0 0 0 0 ... \n",

"3303 0 0 0 0 0 ... \n",

"\n",

" goitre tumor hypopituitary psych TSH T3 TT4 \\\n",

"0 0 0 0 0 0.0 0.0 207.0 \n",

"1 0 0 0 0 13.7 0.0 67.0 \n",

"2 0 0 0 0 1.8 3.7 109.0 \n",

"3 0 0 0 0 0.0 0.0 155.0 \n",

"4 0 0 0 1 3.3 0.7 76.0 \n",

"... ... ... ... ... ... ... ... \n",

"3299 0 0 0 0 0.068916 0.0 183.108392 \n",

"3300 0 0 0 0 0.225344 2.258907 202.958842 \n",

"3301 0 0 0 0 0.073103 0.362554 150.842368 \n",

"3302 0 0 0 0 0.081666 0.017412 125.068134 \n",

"3303 0 0 0 0 14.285991 0.153002 113.76501 \n",

"\n",

" T4U FTI TBG \n",

"0 1.66 125.0 0.0 \n",

"1 0.75 89.0 0.0 \n",

"2 1.62 67.0 0.0 \n",

"3 0.93 166.0 25.0 \n",

"4 0.72 105.0 0.0 \n",

"... ... ... ... \n",

"3299 0.924867 198.134965 0.0 \n",

"3300 0.934109 217.136977 0.0 \n",

"3301 0.851576 177.369471 0.0 \n",

"3302 0.780303 161.015141 0.0 \n",

"3303 1.1334 100.340004 0.0 \n",

"\n",

"[3304 rows x 22 columns]"

],

"text/html": [

"\n",

" <div id=\"df-7a33792e-0b53-437d-aec9-c739242a587b\">\n",

" <div class=\"colab-df-container\">\n",

" <div>\n",

"<style scoped>\n",

" .dataframe tbody tr th:only-of-type {\n",

" vertical-align: middle;\n",

" }\n",

"\n",

" .dataframe tbody tr th {\n",

" vertical-align: top;\n",

" }\n",

"\n",

" .dataframe thead th {\n",

" text-align: right;\n",

" }\n",

"</style>\n",

"<table border=\"1\" class=\"dataframe\">\n",

" <thead>\n",

" <tr style=\"text-align: right;\">\n",

" <th></th>\n",

" <th>age</th>\n",

" <th>sex</th>\n",

" <th>on\_thyroxine</th>\n",

" <th>query\_on\_thyroxine</th>\n",

" <th>on\_antithyroid\_meds</th>\n",

" <th>sick</th>\n",

" <th>pregnant</th>\n",

" <th>thyroid\_surgery</th>\n",

" <th>I131\_treatment</th>\n",

" <th>query\_hypothyroid</th>\n",

" <th>...</th>\n",

" <th>goitre</th>\n",

" <th>tumor</th>\n",

" <th>hypopituitary</th>\n",

" <th>psych</th>\n",

" <th>TSH</th>\n",

" <th>T3</th>\n",

" <th>TT4</th>\n",

" <th>T4U</th>\n",

" <th>FTI</th>\n",

" <th>TBG</th>\n",

" </tr>\n",

" </thead>\n",

" <tbody>\n",

" <tr>\n",

" <th>0</th>\n",

" <td>48.000000</td>\n",

" <td>0</td>\n",

" <td>0</td>\n",

" <td>0</td>\n",

" <td>0</td>\n",

" <td>0</td>\n",

" <td>0</td>\n",

" <td>0</td>\n",

" <td>0</td>\n",

" <td>0</td>\n",

" <td>...</td>\n",

" <td>0</td>\n",

" <td>0</td>\n",

" <td>0</td>\n",

" <td>0</td>\n",

" <td>0.0</td>\n",

" <td>0.0</td>\n",

" <td>207.0</td>\n",

" <td>1.66</td>\n",

" <td>125.0</td>\n",

" <td>0.0</td>\n",

" </tr>\n",

" <tr>\n",

" <th>1</th>\n",

" <td>72.000000</td>\n",

" <td>0</td>\n",

" <td>0</td>\n",

" <td>0</td>\n",

" <td>0</td>\n",

" <td>0</td>\n",

" <td>0</td>\n",

" <td>0</td>\n",

" <td>0</td>\n",

" <td>0</td>\n",

" <td>...</td>\n",

" <td>0</td>\n",

" <td>0</td>\n",

" <td>0</td>\n",

" <td>0</td>\n",

" <td>13.7</td>\n",

" <td>0.0</td>\n",

" <td>67.0</td>\n",

" <td>0.75</td>\n",

" <td>89.0</td>\n",

" <td>0.0</td>\n",

" </tr>\n",

" <tr>\n",

" <th>2</th>\n",

" <td>31.000000</td>\n",

" <td>0</td>\n",

" <td>0</td>\n",

" <td>0</td>\n",

" <td>0</td>\n",

" <td>0</td>\n",

" <td>0</td>\n",

" <td>0</td>\n",

" <td>0</td>\n",

" <td>0</td>\n",

" <td>...</td>\n",

" <td>0</td>\n",

" <td>0</td>\n",

" <td>0</td>\n",

" <td>0</td>\n",

" <td>1.8</td>\n",

" <td>3.7</td>\n",

" <td>109.0</td>\n",

" <td>1.62</td>\n",

" <td>67.0</td>\n",

" <td>0.0</td>\n",

" </tr>\n",

" <tr>\n",

" <th>3</th>\n",

" <td>14.000000</td>\n",

" <td>0</td>\n",

" <td>0</td>\n",

" <td>0</td>\n",

" <td>0</td>\n",

" <td>0</td>\n",

" <td>0</td>\n",

" <td>0</td>\n",

" <td>0</td>\n",

" <td>0</td>\n",

" <td>...</td>\n",

" <td>0</td>\n",

" <td>0</td>\n",

" <td>0</td>\n",

" <td>0</td>\n",

" <td>0.0</td>\n",

" <td>0.0</td>\n",

" <td>155.0</td>\n",

" <td>0.93</td>\n",

" <td>166.0</td>\n",

" <td>25.0</td>\n",

" </tr>\n",

" <tr>\n",

" <th>4</th>\n",

" <td>54.000000</td>\n",

" <td>0</td>\n",

" <td>0</td>\n",

" <td>0</td>\n",

" <td>0</td>\n",

" <td>0</td>\n",

" <td>0</td>\n",

" <td>0</td>\n",

" <td>0</td>\n",

" <td>0</td>\n",

" <td>...</td>\n",

" <td>0</td>\n",

" <td>0</td>\n",

" <td>0</td>\n",

" <td>1</td>\n",

" <td>3.3</td>\n",

" <td>0.7</td>\n",

" <td>76.0</td>\n",

" <td>0.72</td>\n",

" <td>105.0</td>\n",

" <td>0.0</td>\n",

" </tr>\n",

" <tr>\n",

" <th>...</th>\n",

" <td>...</td>\n",

" <td>...</td>\n",

" <td>...</td>\n",

" <td>...</td>\n",

" <td>...</td>\n",

" <td>...</td>\n",

" <td>...</td>\n",

" <td>...</td>\n",

" <td>...</td>\n",

" <td>...</td>\n",

" <td>...</td>\n",

" <td>...</td>\n",

" <td>...</td>\n",

" <td>...</td>\n",

" <td>...</td>\n",

" <td>...</td>\n",

" <td>...</td>\n",

" <td>...</td>\n",

" <td>...</td>\n",

" <td>...</td>\n",

" <td>...</td>\n",

" </tr>\n",

" <tr>\n",

" <th>3299</th>\n",

" <td>72.621678</td>\n",

" <td>0</td>\n",

" <td>1</td>\n",

" <td>0</td>\n",

" <td>0</td>\n",

" <td>0</td>\n",

" <td>0</td>\n",

" <td>0</td>\n",

" <td>0</td>\n",

" <td>0</td>\n",

" <td>...</td>\n",

" <td>0</td>\n",

" <td>0</td>\n",

" <td>0</td>\n",

" <td>0</td>\n",

" <td>0.068916</td>\n",

" <td>0.0</td>\n",

" <td>183.108392</td>\n",

" <td>0.924867</td>\n",

" <td>198.134965</td>\n",

" <td>0.0</td>\n",

" </tr>\n",

" <tr>\n",

" <th>3300</th>\n",

" <td>63.726045</td>\n",

" <td>0</td>\n",

" <td>1</td>\n",

" <td>0</td>\n",

" <td>0</td>\n",

" <td>0</td>\n",

" <td>0</td>\n",

" <td>0</td>\n",

" <td>0</td>\n",

" <td>0</td>\n",

" <td>...</td>\n",

" <td>0</td>\n",

" <td>0</td>\n",

" <td>0</td>\n",

" <td>0</td>\n",

" <td>0.225344</td>\n",

" <td>2.258907</td>\n",

" <td>202.958842</td>\n",

" <td>0.934109</td>\n",

" <td>217.136977</td>\n",

" <td>0.0</td>\n",

" </tr>\n",

" <tr>\n",

" <th>3301</th>\n",

" <td>61.000000</td>\n",

" <td>0</td>\n",

" <td>1</td>\n",

" <td>0</td>\n",

" <td>0</td>\n",

" <td>0</td>\n",

" <td>0</td>\n",

" <td>0</td>\n",

" <td>0</td>\n",

" <td>0</td>\n",

" <td>...</td>\n",

" <td>0</td>\n",

" <td>0</td>\n",

" <td>0</td>\n",

" <td>0</td>\n",

" <td>0.073103</td>\n",

" <td>0.362554</td>\n",

" <td>150.842368</td>\n",

" <td>0.851576</td>\n",

" <td>177.369471</td>\n",

" <td>0.0</td>\n",

" </tr>\n",

" <tr>\n",

" <th>3302</th>\n",

" <td>58.962148</td>\n",

" <td>0</td>\n",

" <td>1</td>\n",

" <td>0</td>\n",

" <td>0</td>\n",

" <td>0</td>\n",

" <td>0</td>\n",

" <td>0</td>\n",

" <td>0</td>\n",

" <td>0</td>\n",

" <td>...</td>\n",

" <td>0</td>\n",

" <td>0</td>\n",

" <td>0</td>\n",

" <td>0</td>\n",

" <td>0.081666</td>\n",

" <td>0.017412</td>\n",

" <td>125.068134</td>\n",

" <td>0.780303</td>\n",

" <td>161.015141</td>\n",

" <td>0.0</td>\n",

" </tr>\n",

" <tr>\n",

" <th>3303</th>\n",

" <td>28.149989</td>\n",

" <td>0</td>\n",

" <td>1</td>\n",

" <td>0</td>\n",

" <td>0</td>\n",

" <td>0</td>\n",

" <td>0</td>\n",

" <td>0</td>\n",

" <td>0</td>\n",

" <td>0</td>\n",

" <td>...</td>\n",

" <td>0</td>\n",

" <td>0</td>\n",

" <td>0</td>\n",

" <td>0</td>\n",

" <td>14.285991</td>\n",

" <td>0.153002</td>\n",

" <td>113.76501</td>\n",

" <td>1.1334</td>\n",

" <td>100.340004</td>\n",

" <td>0.0</td>\n",

" </tr>\n",

" </tbody>\n",

"</table>\n",

"<p>3304 rows × 22 columns</p>\n",

"</div>\n",

" <button class=\"colab-df-convert\" onclick=\"convertToInteractive('df-7a33792e-0b53-437d-aec9-c739242a587b')\"\n",

" title=\"Convert this dataframe to an interactive table.\"\n",

" style=\"display:none;\">\n",

" \n",

" <svg xmlns=\"http://www.w3.org/2000/svg\" height=\"24px\"viewBox=\"0 0 24 24\"\n",

" width=\"24px\">\n",

" <path d=\"M0 0h24v24H0V0z\" fill=\"none\"/>\n",

" <path d=\"M18.56 5.44l.94 2.06.94-2.06 2.06-.94-2.06-.94-.94-2.06-.94 2.06-2.06.94zm-11 1L8.5 8.5l.94-2.06 2.06-.94-2.06-.94L8.5 2.5l-.94 2.06-2.06.94zm10 10l.94 2.06.94-2.06 2.06-.94-2.06-.94-.94-2.06-.94 2.06-2.06.94z\"/><path d=\"M17.41 7.96l-1.37-1.37c-.4-.4-.92-.59-1.43-.59-.52 0-1.04.2-1.43.59L10.3 9.45l-7.72 7.72c-.78.78-.78 2.05 0 2.83L4 21.41c.39.39.9.59 1.41.59.51 0 1.02-.2 1.41-.59l7.78-7.78 2.81-2.81c.8-.78.8-2.07 0-2.86zM5.41 20L4 18.59l7.72-7.72 1.47 1.35L5.41 20z\"/>\n",

" </svg>\n",

" </button>\n",

" \n",

" <style>\n",

" .colab-df-container {\n",

" display:flex;\n",

" flex-wrap:wrap;\n",

" gap: 12px;\n",

" }\n",

"\n",

" .colab-df-convert {\n",

" background-color: #E8F0FE;\n",

" border: none;\n",

" border-radius: 50%;\n",

" cursor: pointer;\n",

" display: none;\n",

" fill: #1967D2;\n",

" height: 32px;\n",

" padding: 0 0 0 0;\n",

" width: 32px;\n",

" }\n",

"\n",

" .colab-df-convert:hover {\n",

" background-color: #E2EBFA;\n",

" box-shadow: 0px 1px 2px rgba(60, 64, 67, 0.3), 0px 1px 3px 1px rgba(60, 64, 67, 0.15);\n",

" fill: #174EA6;\n",

" }\n",

"\n",

" [theme=dark] .colab-df-convert {\n",

" background-color: #3B4455;\n",

" fill: #D2E3FC;\n",

" }\n",

"\n",

" [theme=dark] .colab-df-convert:hover {\n",

" background-color: #434B5C;\n",

" box-shadow: 0px 1px 3px 1px rgba(0, 0, 0, 0.15);\n",

" filter: drop-shadow(0px 1px 2px rgba(0, 0, 0, 0.3));\n",

" fill: #FFFFFF;\n",

" }\n",

" </style>\n",

"\n",

" <script>\n",

" const buttonEl =\n",

" document.querySelector('#df-7a33792e-0b53-437d-aec9-c739242a587b button.colab-df-convert');\n",

" buttonEl.style.display =\n",

" google.colab.kernel.accessAllowed ? 'block' : 'none';\n",

"\n",

" async function convertToInteractive(key) {\n",

" const element = document.querySelector('#df-7a33792e-0b53-437d-aec9-c739242a587b');\n",

" const dataTable =\n",

" await google.colab.kernel.invokeFunction('convertToInteractive',\n",

" [key], {});\n",

" if (!dataTable) return;\n",

"\n",

" const docLinkHtml = 'Like what you see? Visit the ' +\n",

" '<a target=\"\_blank\" href=https://colab.research.google.com/notebooks/data\_table.ipynb>data table notebook</a>'\n",

" + ' to learn more about interactive tables.';\n",

" element.innerHTML = '';\n",

" dataTable['output\_type'] = 'display\_data';\n",

" await google.colab.output.renderOutput(dataTable, element);\n",

" const docLink = document.createElement('div');\n",

" docLink.innerHTML = docLinkHtml;\n",

" element.appendChild(docLink);\n",

" }\n",

" </script>\n",

" </div>\n",

" </div>\n",

" "

]

},

"metadata": {},

"execution\_count": 46

}

],

"source": [

"x\_bal"

]

},

{

"cell\_type": "code",

"execution\_count": null,

"metadata": {

"id": "XYt-XkIf0jxz"

},

"outputs": [],

"source": [

"from sklearn.preprocessing import StandardScaler\n",

"sc = StandardScaler()\n",

"x\_bal = sc.fit\_transform(x\_bal)\n",

"x\_test\_bal= sc.transform(x\_test\_bal)"

]

},

{

"cell\_type": "code",

"execution\_count": null,

"metadata": {

"colab": {

"base\_uri": "https://localhost:8080/"

},

"id": "AicxxfIP26dM",

"outputId": "5aaa2459-99b3-449c-9857-da2dc3d6c119"

},

"outputs": [

{

"output\_type": "execute\_result",

"data": {

"text/plain": [

"array([[-0.19378269, -0.43366098, -0.4237673 , ..., 1.97247536,\n",

" 0.10098737, -0.18730348],\n",

" [ 1.06590801, -0.43366098, -0.4237673 , ..., -0.47069577,\n",

" -0.33015118, -0.18730348],\n",

" [-1.08606361, -0.43366098, -0.4237673 , ..., 1.86508322,\n",

" -0.59362474, -0.18730348],\n",

" ...,\n",

" [ 0.48854977, -0.43366098, 2.35978568, ..., -0.19798331,\n",

" 0.72816787, -0.18730348],\n",

" [ 0.38158878, -0.43366098, 2.35978568, ..., -0.38933866,\n",

" 0.53230725, -0.18730348],\n",

" [-1.23565244, -0.43366098, 2.35978568, ..., 0.55865798,\n",

" -0.19434248, -0.18730348]])"

]

},

"metadata": {},

"execution\_count": 48

}

],

"source": [

"x\_bal"

]

},

{

"cell\_type": "code",

"execution\_count": null,

"metadata": {

"colab": {

"base\_uri": "https://localhost:8080/"

},

"id": "JFsDMFfM2825",

"outputId": "7a3289f2-5cf7-4dd7-f9a1-1758910882b6"

},

"outputs": [

{

"output\_type": "execute\_result",

"data": {

"text/plain": [

"array([[-0.98108938, -0.43366098, -0.4237673 , ..., -0.28275953,\n",

" -0.06667762, -0.18730348],\n",

" [ 1.06590801, -0.43366098, -0.4237673 , ..., -0.49754381,\n",

" 4.41237289, -0.18730348],\n",

" [-0.40373114, 2.30594875, -0.4237673 , ..., 0.09311295,\n",

" -0.89302651, -0.18730348],\n",

" ...,\n",

" [ 0.12750926, -0.43366098, 2.35978568, ..., -0.24411648,\n",

" 1.28952497, -0.18730348],\n",

" [ 0.12047491, -0.43366098, 2.35978568, ..., -0.07171349,\n",

" 0.53743196, -0.18730348],\n",

" [ 0.47883698, -0.43366098, 2.35978568, ..., 0.16271716,\n",

" 0.7485795 , -0.18730348]])"

]

},

"metadata": {},

"execution\_count": 49

}

],

"source": [

"x\_test\_bal"

]

},

{

"cell\_type": "code",

"execution\_count": null,

"metadata": {

"colab": {

"base\_uri": "https://localhost:8080/"

},

"id": "JwqLZ3173Pey",

"outputId": "ec9c5df0-cb5c-41fb-d100-1fe7df0067d7"

},

"outputs": [

{

"output\_type": "execute\_result",

"data": {

"text/plain": [

"target\n",

"0 472\n",

"1 472\n",

"2 472\n",

"3 472\n",

"4 472\n",

"5 472\n",

"6 472\n",

"dtype: int64"

]

},

"metadata": {},

"execution\_count": 50

}

],

"source": [

"y\_bal.value\_counts()"

]

},

{

"cell\_type": "code",

"execution\_count": null,

"metadata": {

"id": "bU8msrQy3WO-"

},

"outputs": [],

"source": [

"columns=['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','TT4','T4U','FTI','TBG']"

]

},

{

"cell\_type": "code",

"execution\_count": null,

"metadata": {

"id": "MV0Ud\_l43lLE"

},

"outputs": [],

"source": [

"x\_test\_bal= pd.DataFrame(x\_test\_bal,columns=columns)"

]

},

{

"cell\_type": "code",

"execution\_count": null,

"metadata": {

"id": "WSXH5jvG3ywR"

},

"outputs": [],

"source": [

"x\_bal= pd.DataFrame(x\_bal,columns=columns)"

]

},

{

"cell\_type": "code",

"execution\_count": null,

"metadata": {

"colab": {

"base\_uri": "https://localhost:8080/",

"height": 488

},

"id": "yPKkJOVQ3-Ks",

"outputId": "7e61d4b7-5654-4902-d8ce-25ec1f156b8f"

},

"outputs": [

{

"output\_type": "execute\_result",

"data": {

"text/plain": [

" age sex on\_thyroxine query\_on\_thyroxine \\\n",

"0 -0.193783 -0.433661 -0.423767 -0.104957 \n",

"1 1.065908 -0.433661 -0.423767 -0.104957 \n",

"2 -1.086064 -0.433661 -0.423767 -0.104957 \n",

"3 -1.978345 -0.433661 -0.423767 -0.104957 \n",

"4 0.121140 -0.433661 -0.423767 -0.104957 \n",

"... ... ... ... ... \n",

"3299 1.098538 -0.433661 2.359786 -0.104957 \n",

"3300 0.631632 -0.433661 2.359786 -0.104957 \n",

"3301 0.488550 -0.433661 2.359786 -0.104957 \n",

"3302 0.381589 -0.433661 2.359786 -0.104957 \n",

"3303 -1.235652 -0.433661 2.359786 -0.104957 \n",

"\n",

" on\_antithyroid\_meds sick pregnant thyroid\_surgery \\\n",

"0 -0.074012 -0.140546 -0.13829 -0.281129 \n",

"1 -0.074012 -0.140546 -0.13829 -0.281129 \n",

"2 -0.074012 -0.140546 -0.13829 -0.281129 \n",

"3 -0.074012 -0.140546 -0.13829 -0.281129 \n",

"4 -0.074012 -0.140546 -0.13829 -0.281129 \n",

"... ... ... ... ... \n",

"3299 -0.074012 -0.140546 -0.13829 -0.281129 \n",

"3300 -0.074012 -0.140546 -0.13829 -0.281129 \n",

"3301 -0.074012 -0.140546 -0.13829 -0.281129 \n",

"3302 -0.074012 -0.140546 -0.13829 -0.281129 \n",

"3303 -0.074012 -0.140546 -0.13829 -0.281129 \n",

"\n",

" I131\_treatment query\_hypothyroid ... goitre tumor \\\n",

"0 -0.23 -0.234344 ... -0.052263 -0.131306 \n",

"1 -0.23 -0.234344 ... -0.052263 -0.131306 \n",

"2 -0.23 -0.234344 ... -0.052263 -0.131306 \n",

"3 -0.23 -0.234344 ... -0.052263 -0.131306 \n",

"4 -0.23 -0.234344 ... -0.052263 -0.131306 \n",

"... ... ... ... ... ... \n",

"3299 -0.23 -0.234344 ... -0.052263 -0.131306 \n",

"3300 -0.23 -0.234344 ... -0.052263 -0.131306 \n",

"3301 -0.23 -0.234344 ... -0.052263 -0.131306 \n",

"3302 -0.23 -0.234344 ... -0.052263 -0.131306 \n",

"3303 -0.23 -0.234344 ... -0.052263 -0.131306 \n",

"\n",

" hypopituitary psych TSH T3 TT4 T4U \\\n",

"0 -0.0174 -0.114831 -0.322733 -1.014991 1.340750 1.972475 \n",

"1 -0.0174 -0.114831 0.056471 -1.014991 -0.709600 -0.470696 \n",

"2 -0.0174 -0.114831 -0.272911 1.290950 -0.094495 1.865083 \n",

"3 -0.0174 -0.114831 -0.322733 -1.014991 0.579192 0.012569 \n",

"4 -0.0174 8.708456 -0.231392 -0.578732 -0.577792 -0.551240 \n",

"... ... ... ... ... ... ... \n",

"3299 -0.0174 -0.114831 -0.320825 -1.014991 0.990849 -0.001212 \n",

"3300 -0.0174 -0.114831 -0.316496 0.392821 1.281566 0.023602 \n",

"3301 -0.0174 -0.114831 -0.320710 -0.789038 0.518302 -0.197983 \n",

"3302 -0.0174 -0.114831 -0.320473 -1.004140 0.140829 -0.389339 \n",

"3303 -0.0174 -0.114831 0.072691 -0.919636 -0.024709 0.558658 \n",

"\n",

" FTI TBG \n",

"0 0.100987 -0.187303 \n",

"1 -0.330151 -0.187303 \n",

"2 -0.593625 -0.187303 \n",

"3 0.592006 1.862310 \n",

"4 -0.138534 -0.187303 \n",

"... ... ... \n",

"3299 0.976857 -0.187303 \n",

"3300 1.204426 -0.187303 \n",

"3301 0.728168 -0.187303 \n",

"3302 0.532307 -0.187303 \n",

"3303 -0.194342 -0.187303 \n",

"\n",

"[3304 rows x 22 columns]"

],

"text/html": [

"\n",

" <div id=\"df-2a6bd918-9e66-4225-bb49-8bb815432426\">\n",

" <div class=\"colab-df-container\">\n",

" <div>\n",

"<style scoped>\n",

" .dataframe tbody tr th:only-of-type {\n",

" vertical-align: middle;\n",

" }\n",

"\n",

" .dataframe tbody tr th {\n",

" vertical-align: top;\n",

" }\n",

"\n",

" .dataframe thead th {\n",

" text-align: right;\n",

" }\n",

"</style>\n",

"<table border=\"1\" class=\"dataframe\">\n",

" <thead>\n",

" <tr style=\"text-align: right;\">\n",

" <th></th>\n",

" <th>age</th>\n",

" <th>sex</th>\n",

" <th>on\_thyroxine</th>\n",

" <th>query\_on\_thyroxine</th>\n",

" <th>on\_antithyroid\_meds</th>\n",

" <th>sick</th>\n",

" <th>pregnant</th>\n",

" <th>thyroid\_surgery</th>\n",

" <th>I131\_treatment</th>\n",

" <th>query\_hypothyroid</th>\n",

" <th>...</th>\n",

" <th>goitre</th>\n",

" <th>tumor</th>\n",

" <th>hypopituitary</th>\n",

" <th>psych</th>\n",

" <th>TSH</th>\n",

" <th>T3</th>\n",

" <th>TT4</th>\n",

" <th>T4U</th>\n",

" <th>FTI</th>\n",

" <th>TBG</th>\n",

" </tr>\n",

" </thead>\n",

" <tbody>\n",

" <tr>\n",

" <th>0</th>\n",

" <td>-0.193783</td>\n",

" <td>-0.433661</td>\n",

" <td>-0.423767</td>\n",

" <td>-0.104957</td>\n",

" <td>-0.074012</td>\n",

" <td>-0.140546</td>\n",

" <td>-0.13829</td>\n",

" <td>-0.281129</td>\n",

" <td>-0.23</td>\n",

" <td>-0.234344</td>\n",

" <td>...</td>\n",

" <td>-0.052263</td>\n",

" <td>-0.131306</td>\n",

" <td>-0.0174</td>\n",

" <td>-0.114831</td>\n",

" <td>-0.322733</td>\n",

" <td>-1.014991</td>\n",

" <td>1.340750</td>\n",

" <td>1.972475</td>\n",

" <td>0.100987</td>\n",

" <td>-0.187303</td>\n",

" </tr>\n",

" <tr>\n",

" <th>1</th>\n",

" <td>1.065908</td>\n",

" <td>-0.433661</td>\n",

" <td>-0.423767</td>\n",

" <td>-0.104957</td>\n",

" <td>-0.074012</td>\n",

" <td>-0.140546</td>\n",

" <td>-0.13829</td>\n",

" <td>-0.281129</td>\n",

" <td>-0.23</td>\n",

" <td>-0.234344</td>\n",

" <td>...</td>\n",

" <td>-0.052263</td>\n",

" <td>-0.131306</td>\n",

" <td>-0.0174</td>\n",

" <td>-0.114831</td>\n",

" <td>0.056471</td>\n",

" <td>-1.014991</td>\n",

" <td>-0.709600</td>\n",

" <td>-0.470696</td>\n",

" <td>-0.330151</td>\n",

" <td>-0.187303</td>\n",

" </tr>\n",

" <tr>\n",

" <th>2</th>\n",

" <td>-1.086064</td>\n",

" <td>-0.433661</td>\n",

" <td>-0.423767</td>\n",

" <td>-0.104957</td>\n",

" <td>-0.074012</td>\n",

" <td>-0.140546</td>\n",

" <td>-0.13829</td>\n",

" <td>-0.281129</td>\n",

" <td>-0.23</td>\n",

" <td>-0.234344</td>\n",

" <td>...</td>\n",

" <td>-0.052263</td>\n",

" <td>-0.131306</td>\n",

" <td>-0.0174</td>\n",

" <td>-0.114831</td>\n",

" <td>-0.272911</td>\n",

" <td>1.290950</td>\n",

" <td>-0.094495</td>\n",

" <td>1.865083</td>\n",

" <td>-0.593625</td>\n",

" <td>-0.187303</td>\n",

" </tr>\n",

" <tr>\n",

" <th>3</th>\n",

" <td>-1.978345</td>\n",

" <td>-0.433661</td>\n",

" <td>-0.423767</td>\n",

" <td>-0.104957</td>\n",

" <td>-0.074012</td>\n",

" <td>-0.140546</td>\n",

" <td>-0.13829</td>\n",

" <td>-0.281129</td>\n",

" <td>-0.23</td>\n",

" <td>-0.234344</td>\n",

" <td>...</td>\n",

" <td>-0.052263</td>\n",

" <td>-0.131306</td>\n",

" <td>-0.0174</td>\n",

" <td>-0.114831</td>\n",

" <td>-0.322733</td>\n",

" <td>-1.014991</td>\n",

" <td>0.579192</td>\n",

" <td>0.012569</td>\n",

" <td>0.592006</td>\n",

" <td>1.862310</td>\n",

" </tr>\n",

" <tr>\n",

" <th>4</th>\n",

" <td>0.121140</td>\n",

" <td>-0.433661</td>\n",

" <td>-0.423767</td>\n",

" <td>-0.104957</td>\n",

" <td>-0.074012</td>\n",

" <td>-0.140546</td>\n",

" <td>-0.13829</td>\n",

" <td>-0.281129</td>\n",

" <td>-0.23</td>\n",

" <td>-0.234344</td>\n",

" <td>...</td>\n",

" <td>-0.052263</td>\n",

" <td>-0.131306</td>\n",

" <td>-0.0174</td>\n",

" <td>8.708456</td>\n",

" <td>-0.231392</td>\n",

" <td>-0.578732</td>\n",

" <td>-0.577792</td>\n",

" <td>-0.551240</td>\n",

" <td>-0.138534</td>\n",

" <td>-0.187303</td>\n",

" </tr>\n",

" <tr>\n",

" <th>...</th>\n",

" <td>...</td>\n",

" <td>...</td>\n",

" <td>...</td>\n",

" <td>...</td>\n",

" <td>...</td>\n",

" <td>...</td>\n",

" <td>...</td>\n",

" <td>...</td>\n",

" <td>...</td>\n",

" <td>...</td>\n",

" <td>...</td>\n",

" <td>...</td>\n",

" <td>...</td>\n",

" <td>...</td>\n",

" <td>...</td>\n",

" <td>...</td>\n",

" <td>...</td>\n",

" <td>...</td>\n",

" <td>...</td>\n",

" <td>...</td>\n",

" <td>...</td>\n",

" </tr>\n",

" <tr>\n",

" <th>3299</th>\n",

" <td>1.098538</td>\n",

" <td>-0.433661</td>\n",

" <td>2.359786</td>\n",

" <td>-0.104957</td>\n",

" <td>-0.074012</td>\n",

" <td>-0.140546</td>\n",

" <td>-0.13829</td>\n",

" <td>-0.281129</td>\n",

" <td>-0.23</td>\n",

" <td>-0.234344</td>\n",

" <td>...</td>\n",

" <td>-0.052263</td>\n",

" <td>-0.131306</td>\n",

" <td>-0.0174</td>\n",

" <td>-0.114831</td>\n",

" <td>-0.320825</td>\n",

" <td>-1.014991</td>\n",

" <td>0.990849</td>\n",

" <td>-0.001212</td>\n",

" <td>0.976857</td>\n",

" <td>-0.187303</td>\n",

" </tr>\n",

" <tr>\n",

" <th>3300</th>\n",

" <td>0.631632</td>\n",

" <td>-0.433661</td>\n",

" <td>2.359786</td>\n",

" <td>-0.104957</td>\n",

" <td>-0.074012</td>\n",

" <td>-0.140546</td>\n",

" <td>-0.13829</td>\n",

" <td>-0.281129</td>\n",

" <td>-0.23</td>\n",

" <td>-0.234344</td>\n",

" <td>...</td>\n",

" <td>-0.052263</td>\n",

" <td>-0.131306</td>\n",

" <td>-0.0174</td>\n",

" <td>-0.114831</td>\n",

" <td>-0.316496</td>\n",

" <td>0.392821</td>\n",

" <td>1.281566</td>\n",

" <td>0.023602</td>\n",

" <td>1.204426</td>\n",

" <td>-0.187303</td>\n",

" </tr>\n",

" <tr>\n",

" <th>3301</th>\n",

" <td>0.488550</td>\n",

" <td>-0.433661</td>\n",

" <td>2.359786</td>\n",

" <td>-0.104957</td>\n",

" <td>-0.074012</td>\n",

" <td>-0.140546</td>\n",

" <td>-0.13829</td>\n",

" <td>-0.281129</td>\n",

" <td>-0.23</td>\n",

" <td>-0.234344</td>\n",

" <td>...</td>\n",

" <td>-0.052263</td>\n",

" <td>-0.131306</td>\n",

" <td>-0.0174</td>\n",

" <td>-0.114831</td>\n",

" <td>-0.320710</td>\n",

" <td>-0.789038</td>\n",

" <td>0.518302</td>\n",

" <td>-0.197983</td>\n",

" <td>0.728168</td>\n",

" <td>-0.187303</td>\n",

" </tr>\n",

" <tr>\n",

" <th>3302</th>\n",

" <td>0.381589</td>\n",

" <td>-0.433661</td>\n",

" <td>2.359786</td>\n",

" <td>-0.104957</td>\n",

" <td>-0.074012</td>\n",

" <td>-0.140546</td>\n",

" <td>-0.13829</td>\n",

" <td>-0.281129</td>\n",

" <td>-0.23</td>\n",

" <td>-0.234344</td>\n",

" <td>...</td>\n",

" <td>-0.052263</td>\n",

" <td>-0.131306</td>\n",

" <td>-0.0174</td>\n",

" <td>-0.114831</td>\n",

" <td>-0.320473</td>\n",

" <td>-1.004140</td>\n",

" <td>0.140829</td>\n",

" <td>-0.389339</td>\n",

" <td>0.532307</td>\n",

" <td>-0.187303</td>\n",

" </tr>\n",

" <tr>\n",

" <th>3303</th>\n",

" <td>-1.235652</td>\n",

" <td>-0.433661</td>\n",

" <td>2.359786</td>\n",

" <td>-0.104957</td>\n",

" <td>-0.074012</td>\n",

" <td>-0.140546</td>\n",

" <td>-0.13829</td>\n",

" <td>-0.281129</td>\n",

" <td>-0.23</td>\n",

" <td>-0.234344</td>\n",

" <td>...</td>\n",

" <td>-0.052263</td>\n",

" <td>-0.131306</td>\n",

" <td>-0.0174</td>\n",

" <td>-0.114831</td>\n",

" <td>0.072691</td>\n",

" <td>-0.919636</td>\n",

" <td>-0.024709</td>\n",

" <td>0.558658</td>\n",

" <td>-0.194342</td>\n",

" <td>-0.187303</td>\n",

" </tr>\n",

" </tbody>\n",

"</table>\n",

"<p>3304 rows × 22 columns</p>\n",

"</div>\n",

" <button class=\"colab-df-convert\" onclick=\"convertToInteractive('df-2a6bd918-9e66-4225-bb49-8bb815432426')\"\n",

" title=\"Convert this dataframe to an interactive table.\"\n",

" style=\"display:none;\">\n",

" \n",

" <svg xmlns=\"http://www.w3.org/2000/svg\" height=\"24px\"viewBox=\"0 0 24 24\"\n",

" width=\"24px\">\n",

" <path d=\"M0 0h24v24H0V0z\" fill=\"none\"/>\n",

" <path d=\"M18.56 5.44l.94 2.06.94-2.06 2.06-.94-2.06-.94-.94-2.06-.94 2.06-2.06.94zm-11 1L8.5 8.5l.94-2.06 2.06-.94-2.06-.94L8.5 2.5l-.94 2.06-2.06.94zm10 10l.94 2.06.94-2.06 2.06-.94-2.06-.94-.94-2.06-.94 2.06-2.06.94z\"/><path d=\"M17.41 7.96l-1.37-1.37c-.4-.4-.92-.59-1.43-.59-.52 0-1.04.2-1.43.59L10.3 9.45l-7.72 7.72c-.78.78-.78 2.05 0 2.83L4 21.41c.39.39.9.59 1.41.59.51 0 1.02-.2 1.41-.59l7.78-7.78 2.81-2.81c.8-.78.8-2.07 0-2.86zM5.41 20L4 18.59l7.72-7.72 1.47 1.35L5.41 20z\"/>\n",

" </svg>\n",

" </button>\n",

" \n",

" <style>\n",

" .colab-df-container {\n",

" display:flex;\n",

" flex-wrap:wrap;\n",

" gap: 12px;\n",

" }\n",

"\n",

" .colab-df-convert {\n",

" background-color: #E8F0FE;\n",

" border: none;\n",

" border-radius: 50%;\n",

" cursor: pointer;\n",

" display: none;\n",

" fill: #1967D2;\n",

" height: 32px;\n",

" padding: 0 0 0 0;\n",

" width: 32px;\n",

" }\n",

"\n",

" .colab-df-convert:hover {\n",

" background-color: #E2EBFA;\n",

" box-shadow: 0px 1px 2px rgba(60, 64, 67, 0.3), 0px 1px 3px 1px rgba(60, 64, 67, 0.15);\n",

" fill: #174EA6;\n",

" }\n",

"\n",

" [theme=dark] .colab-df-convert {\n",

" background-color: #3B4455;\n",

" fill: #D2E3FC;\n",

" }\n",

"\n",

" [theme=dark] .colab-df-convert:hover {\n",

" background-color: #434B5C;\n",

" box-shadow: 0px 1px 3px 1px rgba(0, 0, 0, 0.15);\n",

" filter: drop-shadow(0px 1px 2px rgba(0, 0, 0, 0.3));\n",

" fill: #FFFFFF;\n",

" }\n",

" </style>\n",

"\n",

" <script>\n",

" const buttonEl =\n",

" document.querySelector('#df-2a6bd918-9e66-4225-bb49-8bb815432426 button.colab-df-convert');\n",

" buttonEl.style.display =\n",

" google.colab.kernel.accessAllowed ? 'block' : 'none';\n",

"\n",

" async function convertToInteractive(key) {\n",

" const element = document.querySelector('#df-2a6bd918-9e66-4225-bb49-8bb815432426');\n",

" const dataTable =\n",

" await google.colab.kernel.invokeFunction('convertToInteractive',\n",

" [key], {});\n",

" if (!dataTable) return;\n",

"\n",

" const docLinkHtml = 'Like what you see? Visit the ' +\n",

" '<a target=\"\_blank\" href=https://colab.research.google.com/notebooks/data\_table.ipynb>data table notebook</a>'\n",

" + ' to learn more about interactive tables.';\n",

" element.innerHTML = '';\n",

" dataTable['output\_type'] = 'display\_data';\n",

" await google.colab.output.renderOutput(dataTable, element);\n",

" const docLink = document.createElement('div');\n",

" docLink.innerHTML = docLinkHtml;\n",

" element.appendChild(docLink);\n",

" }\n",

" </script>\n",

" </div>\n",

" </div>\n",

" "

]

},

"metadata": {},

"execution\_count": 54

}

],

"source": [

"x\_bal"

]

},

{

"cell\_type": "code",

"execution\_count": null,

"metadata": {

"colab": {

"base\_uri": "https://localhost:8080/",

"height": 424

},

"id": "XS9aRf\_Arl5W",

"outputId": "3e7451d0-b39d-4cf5-e16a-24d2df1379c9"

},

"outputs": [

{

"output\_type": "execute\_result",

"data": {

"text/plain": [

" target\n",

"0 1\n",

"1 4\n",

"2 1\n",

"3 3\n",

"4 2\n",

"... ...\n",

"3299 6\n",

"3300 6\n",

"3301 6\n",

"3302 6\n",

"3303 6\n",

"\n",

"[3304 rows x 1 columns]"

],

"text/html": [

"\n",

" <div id=\"df-fb2900fc-be7f-4c31-8a95-094fa3c254a3\">\n",

" <div class=\"colab-df-container\">\n",

" <div>\n",

"<style scoped>\n",

" .dataframe tbody tr th:only-of-type {\n",

" vertical-align: middle;\n",

" }\n",

"\n",

" .dataframe tbody tr th {\n",

" vertical-align: top;\n",

" }\n",

"\n",

" .dataframe thead th {\n",

" text-align: right;\n",

" }\n",

"</style>\n",

"<table border=\"1\" class=\"dataframe\">\n",

" <thead>\n",

" <tr style=\"text-align: right;\">\n",

" <th></th>\n",

" <th>target</th>\n",

" </tr>\n",

" </thead>\n",

" <tbody>\n",

" <tr>\n",

" <th>0</th>\n",

" <td>1</td>\n",

" </tr>\n",

" <tr>\n",

" <th>1</th>\n",

" <td>4</td>\n",

" </tr>\n",

" <tr>\n",

" <th>2</th>\n",

" <td>1</td>\n",

" </tr>\n",

" <tr>\n",

" <th>3</th>\n",

" <td>3</td>\n",

" </tr>\n",

" <tr>\n",

" <th>4</th>\n",

" <td>2</td>\n",

" </tr>\n",

" <tr>\n",

" <th>...</th>\n",

" <td>...</td>\n",

" </tr>\n",

" <tr>\n",

" <th>3299</th>\n",

" <td>6</td>\n",

" </tr>\n",

" <tr>\n",

" <th>3300</th>\n",

" <td>6</td>\n",

" </tr>\n",

" <tr>\n",

" <th>3301</th>\n",

" <td>6</td>\n",

" </tr>\n",

" <tr>\n",

" <th>3302</th>\n",

" <td>6</td>\n",

" </tr>\n",

" <tr>\n",

" <th>3303</th>\n",

" <td>6</td>\n",

" </tr>\n",

" </tbody>\n",

"</table>\n",

"<p>3304 rows × 1 columns</p>\n",

"</div>\n",

" <button class=\"colab-df-convert\" onclick=\"convertToInteractive('df-fb2900fc-be7f-4c31-8a95-094fa3c254a3')\"\n",

" title=\"Convert this dataframe to an interactive table.\"\n",

" style=\"display:none;\">\n",

" \n",

" <svg xmlns=\"http://www.w3.org/2000/svg\" height=\"24px\"viewBox=\"0 0 24 24\"\n",

" width=\"24px\">\n",

" <path d=\"M0 0h24v24H0V0z\" fill=\"none\"/>\n",

" <path d=\"M18.56 5.44l.94 2.06.94-2.06 2.06-.94-2.06-.94-.94-2.06-.94 2.06-2.06.94zm-11 1L8.5 8.5l.94-2.06 2.06-.94-2.06-.94L8.5 2.5l-.94 2.06-2.06.94zm10 10l.94 2.06.94-2.06 2.06-.94-2.06-.94-.94-2.06-.94 2.06-2.06.94z\"/><path d=\"M17.41 7.96l-1.37-1.37c-.4-.4-.92-.59-1.43-.59-.52 0-1.04.2-1.43.59L10.3 9.45l-7.72 7.72c-.78.78-.78 2.05 0 2.83L4 21.41c.39.39.9.59 1.41.59.51 0 1.02-.2 1.41-.59l7.78-7.78 2.81-2.81c.8-.78.8-2.07 0-2.86zM5.41 20L4 18.59l7.72-7.72 1.47 1.35L5.41 20z\"/>\n",

" </svg>\n",

" </button>\n",

" \n",

" <style>\n",

" .colab-df-container {\n",

" display:flex;\n",

" flex-wrap:wrap;\n",

" gap: 12px;\n",

" }\n",

"\n",

" .colab-df-convert {\n",

" background-color: #E8F0FE;\n",

" border: none;\n",

" border-radius: 50%;\n",

" cursor: pointer;\n",

" display: none;\n",

" fill: #1967D2;\n",

" height: 32px;\n",

" padding: 0 0 0 0;\n",

" width: 32px;\n",

" }\n",

"\n",

" .colab-df-convert:hover {\n",

" background-color: #E2EBFA;\n",

" box-shadow: 0px 1px 2px rgba(60, 64, 67, 0.3), 0px 1px 3px 1px rgba(60, 64, 67, 0.15);\n",

" fill: #174EA6;\n",

" }\n",

"\n",

" [theme=dark] .colab-df-convert {\n",

" background-color: #3B4455;\n",

" fill: #D2E3FC;\n",

" }\n",

"\n",

" [theme=dark] .colab-df-convert:hover {\n",

" background-color: #434B5C;\n",

" box-shadow: 0px 1px 3px 1px rgba(0, 0, 0, 0.15);\n",

" filter: drop-shadow(0px 1px 2px rgba(0, 0, 0, 0.3));\n",

" fill: #FFFFFF;\n",

" }\n",

" </style>\n",

"\n",

" <script>\n",

" const buttonEl =\n",

" document.querySelector('#df-fb2900fc-be7f-4c31-8a95-094fa3c254a3 button.colab-df-convert');\n",

" buttonEl.style.display =\n",

" google.colab.kernel.accessAllowed ? 'block' : 'none';\n",

"\n",

" async function convertToInteractive(key) {\n",

" const element = document.querySelector('#df-fb2900fc-be7f-4c31-8a95-094fa3c254a3');\n",

" const dataTable =\n",

" await google.colab.kernel.invokeFunction('convertToInteractive',\n",

" [key], {});\n",

" if (!dataTable) return;\n",

"\n",

" const docLinkHtml = 'Like what you see? Visit the ' +\n",

" '<a target=\"\_blank\" href=https://colab.research.google.com/notebooks/data\_table.ipynb>data table notebook</a>'\n",

" + ' to learn more about interactive tables.';\n",

" element.innerHTML = '';\n",

" dataTable['output\_type'] = 'display\_data';\n",

" await google.colab.output.renderOutput(dataTable, element);\n",

" const docLink = document.createElement('div');\n",

" docLink.innerHTML = docLinkHtml;\n",

" element.appendChild(docLink);\n",

" }\n",

" </script>\n",

" </div>\n",

" </div>\n",

" "

]

},

"metadata": {},

"execution\_count": 55

}

],

"source": [

"y\_bal"

]

},

{

"cell\_type": "code",

"execution\_count": null,

"metadata": {

"colab": {

"base\_uri": "https://localhost:8080/"

},

"id": "6lcelQLk3AYD",

"outputId": "1e4b3298-690f-487f-b5d4-5abaef4c5d03"

},

"outputs": [

{

"output\_type": "stream",

"name": "stderr",

"text": [

"<ipython-input-56-9bf3ba303cb1>:4: 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().\n",

" rfr.fit(x\_bal,y\_bal)\n"

]

},

{

"output\_type": "execute\_result",

"data": {

"text/plain": [

"((3304, 22), (3304, 1), (847, 22), (847, 1))"

]

},

"metadata": {},

"execution\_count": 56

}

],

"source": [

"from sklearn.ensemble import RandomForestClassifier\n",

"from sklearn.metrics import accuracy\_score, classification\_report\n",

"rfr = RandomForestClassifier()\n",

"rfr.fit(x\_bal,y\_bal)\n",

"y\_pred = rfr.predict(x\_test\_bal)\n",

"accuracy\_score(y\_test\_bal,y\_pred)\n",

"x\_bal.shape,y\_bal.shape,x\_test\_bal.shape,y\_test\_bal.shape "

]

},

{

"cell\_type": "code",

"execution\_count": null,

"metadata": {

"colab": {

"base\_uri": "https://localhost:8080/"

},

"id": "jGmB8QG76Mob",

"outputId": "b110d5f5-4bcc-4469-882c-5a8b23d0365e"

},

"outputs": [

{

"output\_type": "execute\_result",

"data": {

"text/plain": [

"0.9598583234946871"

]

},

"metadata": {},

"execution\_count": 57

}

],

"source": [

"test\_score=accuracy\_score(y\_test\_bal,y\_pred)\n",

"test\_score"

]

},

{

"cell\_type": "code",

"execution\_count": null,

"metadata": {

"colab": {

"base\_uri": "https://localhost:8080/"

},

"id": "ABSdPL26\_Knm",

"outputId": "2df84162-7a77-4a91-a5c2-37b7c3b7ab0f"

},

"outputs": [

{

"output\_type": "execute\_result",

"data": {

"text/plain": [

"1.0"

]

},

"metadata": {},

"execution\_count": 58

}

],

"source": [

"train\_score = accuracy\_score(y\_bal,rfr.predict(x\_bal))\n",

"train\_score"

]

},

{

"cell\_type": "markdown",

"metadata": {

"id": "5pkTiCAdmIlJ"

},

"source": [

"Performing Feature Importance"

]

},

{

"cell\_type": "code",

"execution\_count": null,

"metadata": {

"id": "FokMBgXu\_oYM"

},

"outputs": [],

"source": [

"from sklearn.inspection import permutation\_importance\n",

"results = permutation\_importance(rfr,x\_bal,y\_bal, scoring='accuracy')"

]

},

{

"cell\_type": "code",

"execution\_count": null,

"metadata": {

"colab": {

"base\_uri": "https://localhost:8080/",

"height": 1000

},

"id": "nTpcgGgKATm2",

"outputId": "0bc98d18-f1f5-4ca8-cceb-a91745918e0c"

},

"outputs": [

{

"output\_type": "stream",

"name": "stdout",

"text": [

"feature: age score: 0.0\n",

"feature: sex score: 0.0\n",

"feature: on\_thyroxine score: 0.0\n",

"feature: query\_on\_thyroxine score: 0.0\n",

"feature: on\_antithyroid\_meds score: 0.0\n",

"feature: sick score: 0.0001210653753026758\n",

"feature: pregnant score: 0.0002421307506053516\n",

"feature: thyroid\_surgery score: 0.0003026634382566895\n",

"feature: I131\_treatment score: 0.000605326876513268\n",

"feature: query\_hypothyroid score: 0.0007869249394672817\n",

"feature: query\_hyperthyroid score: 0.0010290556900726332\n",

"feature: lithium score: 0.0017554479418885993\n",

"feature: goitre score: 0.007142857142857162\n",

"feature: tumor score: 0.009745762711864425\n",

"feature: hypopituitary score: 0.030266343825665863\n",

"feature: psych score: 0.047760290556900745\n",

"feature: TSH score: 0.07675544794188864\n",

"feature: T3 score: 0.07717917675544791\n",

"feature: TT4 score: 0.11803874092009688\n",

"feature: T4U score: 0.15805084745762715\n",

"feature: FTI score: 0.17070217917675545\n",

"feature: TBG score: 0.27300242130750607\n"

]

},

{

"output\_type": "display\_data",

"data": {

"text/plain": [

"<Figure size 1000x1000 with 1 Axes>"

],

"image/png": "\n"

},

"metadata": {}

}

],

"source": [

"feature\_importance=['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','TT4','T4U','FTI','TBG']\n",

"importance = results.importances\_mean\n",

"importance = np.sort(importance)\n",

"for i,v in enumerate(importance):\n",

" i=feature\_importance[i]\n",

" print('feature: {:<20} score: {}'. format(i,v))\n",

"plt.figure(figsize=(10,10))\n",

"plt.bar(feature\_importance, height = importance)\n",

"plt.xticks(rotation=30, ha='right')\n",

"plt.show()"

]

},

{

"cell\_type": "code",

"execution\_count": null,

"metadata": {

"id": "R3O9g3aBKy7x"

},

"outputs": [],

"source": [

"#x\_bal.drop(['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','TT4','T4U','FTI','TBG'], axis=1,inplace=True)"

]

},

{

"cell\_type": "code",

"execution\_count": null,

"metadata": {

"id": "Rn8d8S3uKEpO"

},

"outputs": [],

"source": [

"#x\_test\_bal.drop(['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','TT4','T4U','FTI','TBG'],axis=1,inplace=True)"

]

},

{

"cell\_type": "code",

"execution\_count": null,

"metadata": {

"colab": {

"base\_uri": "https://localhost:8080/",

"height": 300

},

"id": "BRptpem9FF1h",

"outputId": "5b4fe15b-9c89-4c4b-af0d-f8cdcb6cd6ae"

},

"outputs": [

{

"output\_type": "execute\_result",

"data": {

"text/plain": [

" age sex on\_thyroxine query\_on\_thyroxine on\_antithyroid\_meds \\\n",

"0 -0.193783 -0.433661 -0.423767 -0.104957 -0.074012 \n",

"1 1.065908 -0.433661 -0.423767 -0.104957 -0.074012 \n",

"2 -1.086064 -0.433661 -0.423767 -0.104957 -0.074012 \n",

"3 -1.978345 -0.433661 -0.423767 -0.104957 -0.074012 \n",

"4 0.121140 -0.433661 -0.423767 -0.104957 -0.074012 \n",

"\n",

" sick pregnant thyroid\_surgery I131\_treatment query\_hypothyroid \\\n",

"0 -0.140546 -0.13829 -0.281129 -0.23 -0.234344 \n",

"1 -0.140546 -0.13829 -0.281129 -0.23 -0.234344 \n",

"2 -0.140546 -0.13829 -0.281129 -0.23 -0.234344 \n",

"3 -0.140546 -0.13829 -0.281129 -0.23 -0.234344 \n",

"4 -0.140546 -0.13829 -0.281129 -0.23 -0.234344 \n",

"\n",

" ... goitre tumor hypopituitary psych TSH T3 \\\n",

"0 ... -0.052263 -0.131306 -0.0174 -0.114831 -0.322733 -1.014991 \n",

"1 ... -0.052263 -0.131306 -0.0174 -0.114831 0.056471 -1.014991 \n",

"2 ... -0.052263 -0.131306 -0.0174 -0.114831 -0.272911 1.290950 \n",

"3 ... -0.052263 -0.131306 -0.0174 -0.114831 -0.322733 -1.014991 \n",

"4 ... -0.052263 -0.131306 -0.0174 8.708456 -0.231392 -0.578732 \n",

"\n",

" TT4 T4U FTI TBG \n",

"0 1.340750 1.972475 0.100987 -0.187303 \n",

"1 -0.709600 -0.470696 -0.330151 -0.187303 \n",

"2 -0.094495 1.865083 -0.593625 -0.187303 \n",

"3 0.579192 0.012569 0.592006 1.862310 \n",

"4 -0.577792 -0.551240 -0.138534 -0.187303 \n",

"\n",

"[5 rows x 22 columns]"

],

"text/html": [

"\n",

" <div id=\"df-c79d19ee-11d3-43b2-8647-0602431a932c\">\n",

" <div class=\"colab-df-container\">\n",

" <div>\n",

"<style scoped>\n",

" .dataframe tbody tr th:only-of-type {\n",

" vertical-align: middle;\n",

" }\n",

"\n",

" .dataframe tbody tr th {\n",

" vertical-align: top;\n",

" }\n",

"\n",

" .dataframe thead th {\n",

" text-align: right;\n",

" }\n",

"</style>\n",

"<table border=\"1\" class=\"dataframe\">\n",

" <thead>\n",

" <tr style=\"text-align: right;\">\n",

" <th></th>\n",

" <th>age</th>\n",

" <th>sex</th>\n",

" <th>on\_thyroxine</th>\n",

" <th>query\_on\_thyroxine</th>\n",

" <th>on\_antithyroid\_meds</th>\n",

" <th>sick</th>\n",

" <th>pregnant</th>\n",

" <th>thyroid\_surgery</th>\n",

" <th>I131\_treatment</th>\n",

" <th>query\_hypothyroid</th>\n",

" <th>...</th>\n",

" <th>goitre</th>\n",

" <th>tumor</th>\n",

" <th>hypopituitary</th>\n",

" <th>psych</th>\n",

" <th>TSH</th>\n",

" <th>T3</th>\n",

" <th>TT4</th>\n",

" <th>T4U</th>\n",

" <th>FTI</th>\n",

" <th>TBG</th>\n",

" </tr>\n",

" </thead>\n",

" <tbody>\n",

" <tr>\n",

" <th>0</th>\n",

" <td>-0.193783</td>\n",

" <td>-0.433661</td>\n",

" <td>-0.423767</td>\n",

" <td>-0.104957</td>\n",

" <td>-0.074012</td>\n",

" <td>-0.140546</td>\n",

" <td>-0.13829</td>\n",

" <td>-0.281129</td>\n",

" <td>-0.23</td>\n",

" <td>-0.234344</td>\n",

" <td>...</td>\n",

" <td>-0.052263</td>\n",

" <td>-0.131306</td>\n",

" <td>-0.0174</td>\n",

" <td>-0.114831</td>\n",

" <td>-0.322733</td>\n",

" <td>-1.014991</td>\n",

" <td>1.340750</td>\n",

" <td>1.972475</td>\n",

" <td>0.100987</td>\n",

" <td>-0.187303</td>\n",

" </tr>\n",

" <tr>\n",

" <th>1</th>\n",

" <td>1.065908</td>\n",

" <td>-0.433661</td>\n",

" <td>-0.423767</td>\n",

" <td>-0.104957</td>\n",

" <td>-0.074012</td>\n",

" <td>-0.140546</td>\n",

" <td>-0.13829</td>\n",

" <td>-0.281129</td>\n",

" <td>-0.23</td>\n",

" <td>-0.234344</td>\n",

" <td>...</td>\n",

" <td>-0.052263</td>\n",

" <td>-0.131306</td>\n",

" <td>-0.0174</td>\n",

" <td>-0.114831</td>\n",

" <td>0.056471</td>\n",

" <td>-1.014991</td>\n",

" <td>-0.709600</td>\n",

" <td>-0.470696</td>\n",

" <td>-0.330151</td>\n",

" <td>-0.187303</td>\n",

" </tr>\n",

" <tr>\n",

" <th>2</th>\n",

" <td>-1.086064</td>\n",

" <td>-0.433661</td>\n",

" <td>-0.423767</td>\n",

" <td>-0.104957</td>\n",

" <td>-0.074012</td>\n",

" <td>-0.140546</td>\n",

" <td>-0.13829</td>\n",

" <td>-0.281129</td>\n",

" <td>-0.23</td>\n",

" <td>-0.234344</td>\n",

" <td>...</td>\n",

" <td>-0.052263</td>\n",

" <td>-0.131306</td>\n",

" <td>-0.0174</td>\n",

" <td>-0.114831</td>\n",

" <td>-0.272911</td>\n",

" <td>1.290950</td>\n",

" <td>-0.094495</td>\n",

" <td>1.865083</td>\n",

" <td>-0.593625</td>\n",

" <td>-0.187303</td>\n",

" </tr>\n",

" <tr>\n",

" <th>3</th>\n",

" <td>-1.978345</td>\n",

" <td>-0.433661</td>\n",

" <td>-0.423767</td>\n",

" <td>-0.104957</td>\n",

" <td>-0.074012</td>\n",

" <td>-0.140546</td>\n",

" <td>-0.13829</td>\n",

" <td>-0.281129</td>\n",

" <td>-0.23</td>\n",

" <td>-0.234344</td>\n",

" <td>...</td>\n",

" <td>-0.052263</td>\n",

" <td>-0.131306</td>\n",

" <td>-0.0174</td>\n",

" <td>-0.114831</td>\n",

" <td>-0.322733</td>\n",

" <td>-1.014991</td>\n",

" <td>0.579192</td>\n",

" <td>0.012569</td>\n",

" <td>0.592006</td>\n",

" <td>1.862310</td>\n",

" </tr>\n",

" <tr>\n",

" <th>4</th>\n",

" <td>0.121140</td>\n",

" <td>-0.433661</td>\n",

" <td>-0.423767</td>\n",

" <td>-0.104957</td>\n",

" <td>-0.074012</td>\n",

" <td>-0.140546</td>\n",

" <td>-0.13829</td>\n",

" <td>-0.281129</td>\n",

" <td>-0.23</td>\n",

" <td>-0.234344</td>\n",

" <td>...</td>\n",

" <td>-0.052263</td>\n",

" <td>-0.131306</td>\n",

" <td>-0.0174</td>\n",

" <td>8.708456</td>\n",

" <td>-0.231392</td>\n",

" <td>-0.578732</td>\n",

" <td>-0.577792</td>\n",

" <td>-0.551240</td>\n",

" <td>-0.138534</td>\n",

" <td>-0.187303</td>\n",

" </tr>\n",

" </tbody>\n",

"</table>\n",

"<p>5 rows × 22 columns</p>\n",

"</div>\n",

" <button class=\"colab-df-convert\" onclick=\"convertToInteractive('df-c79d19ee-11d3-43b2-8647-0602431a932c')\"\n",

" title=\"Convert this dataframe to an interactive table.\"\n",

" style=\"display:none;\">\n",

" \n",

" <svg xmlns=\"http://www.w3.org/2000/svg\" height=\"24px\"viewBox=\"0 0 24 24\"\n",

" width=\"24px\">\n",

" <path d=\"M0 0h24v24H0V0z\" fill=\"none\"/>\n",

" <path d=\"M18.56 5.44l.94 2.06.94-2.06 2.06-.94-2.06-.94-.94-2.06-.94 2.06-2.06.94zm-11 1L8.5 8.5l.94-2.06 2.06-.94-2.06-.94L8.5 2.5l-.94 2.06-2.06.94zm10 10l.94 2.06.94-2.06 2.06-.94-2.06-.94-.94-2.06-.94 2.06-2.06.94z\"/><path d=\"M17.41 7.96l-1.37-1.37c-.4-.4-.92-.59-1.43-.59-.52 0-1.04.2-1.43.59L10.3 9.45l-7.72 7.72c-.78.78-.78 2.05 0 2.83L4 21.41c.39.39.9.59 1.41.59.51 0 1.02-.2 1.41-.59l7.78-7.78 2.81-2.81c.8-.78.8-2.07 0-2.86zM5.41 20L4 18.59l7.72-7.72 1.47 1.35L5.41 20z\"/>\n",

" </svg>\n",

" </button>\n",

" \n",

" <style>\n",

" .colab-df-container {\n",

" display:flex;\n",

" flex-wrap:wrap;\n",

" gap: 12px;\n",

" }\n",

"\n",

" .colab-df-convert {\n",

" background-color: #E8F0FE;\n",

" border: none;\n",

" border-radius: 50%;\n",

" cursor: pointer;\n",

" display: none;\n",

" fill: #1967D2;\n",

" height: 32px;\n",

" padding: 0 0 0 0;\n",

" width: 32px;\n",

" }\n",

"\n",

" .colab-df-convert:hover {\n",

" background-color: #E2EBFA;\n",

" box-shadow: 0px 1px 2px rgba(60, 64, 67, 0.3), 0px 1px 3px 1px rgba(60, 64, 67, 0.15);\n",

" fill: #174EA6;\n",

" }\n",

"\n",

" [theme=dark] .colab-df-convert {\n",

" background-color: #3B4455;\n",

" fill: #D2E3FC;\n",

" }\n",

"\n",

" [theme=dark] .colab-df-convert:hover {\n",

" background-color: #434B5C;\n",

" box-shadow: 0px 1px 3px 1px rgba(0, 0, 0, 0.15);\n",

" filter: drop-shadow(0px 1px 2px rgba(0, 0, 0, 0.3));\n",

" fill: #FFFFFF;\n",

" }\n",

" </style>\n",

"\n",

" <script>\n",

" const buttonEl =\n",

" document.querySelector('#df-c79d19ee-11d3-43b2-8647-0602431a932c button.colab-df-convert');\n",

" buttonEl.style.display =\n",

" google.colab.kernel.accessAllowed ? 'block' : 'none';\n",

"\n",

" async function convertToInteractive(key) {\n",

" const element = document.querySelector('#df-c79d19ee-11d3-43b2-8647-0602431a932c');\n",

" const dataTable =\n",

" await google.colab.kernel.invokeFunction('convertToInteractive',\n",

" [key], {});\n",

" if (!dataTable) return;\n",

"\n",

" const docLinkHtml = 'Like what you see? Visit the ' +\n",

" '<a target=\"\_blank\" href=https://colab.research.google.com/notebooks/data\_table.ipynb>data table notebook</a>'\n",

" + ' to learn more about interactive tables.';\n",

" element.innerHTML = '';\n",

" dataTable['output\_type'] = 'display\_data';\n",

" await google.colab.output.renderOutput(dataTable, element);\n",

" const docLink = document.createElement('div');\n",

" docLink.innerHTML = docLinkHtml;\n",

" element.appendChild(docLink);\n",

" }\n",

" </script>\n",

" </div>\n",

" </div>\n",

" "

]

},

"metadata": {},

"execution\_count": 63

}

],

"source": [

"x\_bal.head()"

]

},

{

"cell\_type": "code",

"execution\_count": null,

"metadata": {

"colab": {

"base\_uri": "https://localhost:8080/",

"height": 300

},

"id": "BydesEELFKaR",

"outputId": "f5557ee7-e0d8-4674-ac4e-16308c8550ac"

},

"outputs": [

{

"output\_type": "execute\_result",

"data": {

"text/plain": [

" age sex on\_thyroxine query\_on\_thyroxine on\_antithyroid\_meds \\\n",

"0 -0.981089 -0.433661 -0.423767 -0.104957 -0.074012 \n",

"1 1.065908 -0.433661 -0.423767 -0.104957 -0.074012 \n",

"2 -0.403731 2.305949 -0.423767 -0.104957 -0.074012 \n",

"3 1.013421 -0.433661 -0.423767 -0.104957 -0.074012 \n",

"4 -0.981089 2.305949 -0.423767 -0.104957 -0.074012 \n",

"\n",

" sick pregnant thyroid\_surgery I131\_treatment query\_hypothyroid \\\n",

"0 -0.140546 -0.13829 3.557090 -0.23 -0.234344 \n",

"1 -0.140546 -0.13829 -0.281129 -0.23 -0.234344 \n",

"2 -0.140546 -0.13829 -0.281129 -0.23 -0.234344 \n",

"3 -0.140546 -0.13829 -0.281129 -0.23 -0.234344 \n",

"4 -0.140546 -0.13829 3.557090 -0.23 -0.234344 \n",

"\n",

" ... goitre tumor hypopituitary psych TSH T3 \\\n",

"0 ... -0.052263 -0.131306 -0.0174 -0.114831 -0.079157 0.231463 \n",

"1 ... -0.052263 -0.131306 -0.0174 -0.114831 -0.319965 4.344763 \n",

"2 ... 19.134030 -0.131306 -0.0174 -0.114831 -0.298099 -1.014991 \n",

"3 ... -0.052263 -0.131306 -0.0174 -0.114831 -0.306126 -0.391764 \n",

"4 ... -0.052263 -0.131306 -0.0174 -0.114831 1.144260 -0.391764 \n",

"\n",

" TT4 T4U FTI TBG \n",

"0 -0.358111 -0.282760 -0.066678 -0.187303 \n",

"1 3.566845 -0.497544 4.412373 -0.187303 \n",

"2 -1.105024 0.093113 -0.893027 -0.187303 \n",

"3 -0.402047 0.039417 -0.270271 -0.187303 \n",

"4 -1.222187 -0.390152 -0.905003 -0.187303 \n",

"\n",

"[5 rows x 22 columns]"

],

"text/html": [

"\n",

" <div id=\"df-009b5380-e901-4e79-8bbd-ac8623e89487\">\n",

" <div class=\"colab-df-container\">\n",

" <div>\n",

"<style scoped>\n",

" .dataframe tbody tr th:only-of-type {\n",

" vertical-align: middle;\n",

" }\n",

"\n",

" .dataframe tbody tr th {\n",

" vertical-align: top;\n",

" }\n",

"\n",

" .dataframe thead th {\n",

" text-align: right;\n",

" }\n",

"</style>\n",

"<table border=\"1\" class=\"dataframe\">\n",

" <thead>\n",

" <tr style=\"text-align: right;\">\n",

" <th></th>\n",

" <th>age</th>\n",

" <th>sex</th>\n",

" <th>on\_thyroxine</th>\n",

" <th>query\_on\_thyroxine</th>\n",

" <th>on\_antithyroid\_meds</th>\n",

" <th>sick</th>\n",

" <th>pregnant</th>\n",

" <th>thyroid\_surgery</th>\n",

" <th>I131\_treatment</th>\n",

" <th>query\_hypothyroid</th>\n",

" <th>...</th>\n",

" <th>goitre</th>\n",

" <th>tumor</th>\n",

" <th>hypopituitary</th>\n",

" <th>psych</th>\n",

" <th>TSH</th>\n",

" <th>T3</th>\n",

" <th>TT4</th>\n",

" <th>T4U</th>\n",

" <th>FTI</th>\n",

" <th>TBG</th>\n",

" </tr>\n",

" </thead>\n",

" <tbody>\n",

" <tr>\n",

" <th>0</th>\n",

" <td>-0.981089</td>\n",

" <td>-0.433661</td>\n",

" <td>-0.423767</td>\n",

" <td>-0.104957</td>\n",

" <td>-0.074012</td>\n",

" <td>-0.140546</td>\n",

" <td>-0.13829</td>\n",

" <td>3.557090</td>\n",

" <td>-0.23</td>\n",

" <td>-0.234344</td>\n",

" <td>...</td>\n",

" <td>-0.052263</td>\n",

" <td>-0.131306</td>\n",

" <td>-0.0174</td>\n",

" <td>-0.114831</td>\n",

" <td>-0.079157</td>\n",

" <td>0.231463</td>\n",

" <td>-0.358111</td>\n",

" <td>-0.282760</td>\n",

" <td>-0.066678</td>\n",

" <td>-0.187303</td>\n",

" </tr>\n",

" <tr>\n",

" <th>1</th>\n",

" <td>1.065908</td>\n",

" <td>-0.433661</td>\n",

" <td>-0.423767</td>\n",

" <td>-0.104957</td>\n",

" <td>-0.074012</td>\n",

" <td>-0.140546</td>\n",

" <td>-0.13829</td>\n",

" <td>-0.281129</td>\n",

" <td>-0.23</td>\n",

" <td>-0.234344</td>\n",

" <td>...</td>\n",

" <td>-0.052263</td>\n",

" <td>-0.131306</td>\n",

" <td>-0.0174</td>\n",

" <td>-0.114831</td>\n",

" <td>-0.319965</td>\n",

" <td>4.344763</td>\n",

" <td>3.566845</td>\n",

" <td>-0.497544</td>\n",

" <td>4.412373</td>\n",

" <td>-0.187303</td>\n",

" </tr>\n",

" <tr>\n",

" <th>2</th>\n",

" <td>-0.403731</td>\n",

" <td>2.305949</td>\n",

" <td>-0.423767</td>\n",

" <td>-0.104957</td>\n",

" <td>-0.074012</td>\n",

" <td>-0.140546</td>\n",

" <td>-0.13829</td>\n",

" <td>-0.281129</td>\n",

" <td>-0.23</td>\n",

" <td>-0.234344</td>\n",

" <td>...</td>\n",

" <td>19.134030</td>\n",

" <td>-0.131306</td>\n",

" <td>-0.0174</td>\n",

" <td>-0.114831</td>\n",

" <td>-0.298099</td>\n",

" <td>-1.014991</td>\n",

" <td>-1.105024</td>\n",

" <td>0.093113</td>\n",

" <td>-0.893027</td>\n",

" <td>-0.187303</td>\n",

" </tr>\n",

" <tr>\n",

" <th>3</th>\n",

" <td>1.013421</td>\n",

" <td>-0.433661</td>\n",

" <td>-0.423767</td>\n",

" <td>-0.104957</td>\n",

" <td>-0.074012</td>\n",

" <td>-0.140546</td>\n",

" <td>-0.13829</td>\n",

" <td>-0.281129</td>\n",

" <td>-0.23</td>\n",

" <td>-0.234344</td>\n",

" <td>...</td>\n",

" <td>-0.052263</td>\n",

" <td>-0.131306</td>\n",

" <td>-0.0174</td>\n",

" <td>-0.114831</td>\n",

" <td>-0.306126</td>\n",

" <td>-0.391764</td>\n",

" <td>-0.402047</td>\n",

" <td>0.039417</td>\n",

" <td>-0.270271</td>\n",

" <td>-0.187303</td>\n",

" </tr>\n",

" <tr>\n",

" <th>4</th>\n",

" <td>-0.981089</td>\n",

" <td>2.305949</td>\n",

" <td>-0.423767</td>\n",

" <td>-0.104957</td>\n",

" <td>-0.074012</td>\n",

" <td>-0.140546</td>\n",

" <td>-0.13829</td>\n",

" <td>3.557090</td>\n",

" <td>-0.23</td>\n",

" <td>-0.234344</td>\n",

" <td>...</td>\n",

" <td>-0.052263</td>\n",

" <td>-0.131306</td>\n",

" <td>-0.0174</td>\n",

" <td>-0.114831</td>\n",

" <td>1.144260</td>\n",

" <td>-0.391764</td>\n",

" <td>-1.222187</td>\n",

" <td>-0.390152</td>\n",

" <td>-0.905003</td>\n",

" <td>-0.187303</td>\n",

" </tr>\n",

" </tbody>\n",

"</table>\n",

"<p>5 rows × 22 columns</p>\n",

"</div>\n",

" <button class=\"colab-df-convert\" onclick=\"convertToInteractive('df-009b5380-e901-4e79-8bbd-ac8623e89487')\"\n",

" title=\"Convert this dataframe to an interactive table.\"\n",

" style=\"display:none;\">\n",

" \n",

" <svg xmlns=\"http://www.w3.org/2000/svg\" height=\"24px\"viewBox=\"0 0 24 24\"\n",

" width=\"24px\">\n",

" <path d=\"M0 0h24v24H0V0z\" fill=\"none\"/>\n",

" <path d=\"M18.56 5.44l.94 2.06.94-2.06 2.06-.94-2.06-.94-.94-2.06-.94 2.06-2.06.94zm-11 1L8.5 8.5l.94-2.06 2.06-.94-2.06-.94L8.5 2.5l-.94 2.06-2.06.94zm10 10l.94 2.06.94-2.06 2.06-.94-2.06-.94-.94-2.06-.94 2.06-2.06.94z\"/><path d=\"M17.41 7.96l-1.37-1.37c-.4-.4-.92-.59-1.43-.59-.52 0-1.04.2-1.43.59L10.3 9.45l-7.72 7.72c-.78.78-.78 2.05 0 2.83L4 21.41c.39.39.9.59 1.41.59.51 0 1.02-.2 1.41-.59l7.78-7.78 2.81-2.81c.8-.78.8-2.07 0-2.86zM5.41 20L4 18.59l7.72-7.72 1.47 1.35L5.41 20z\"/>\n",

" </svg>\n",

" </button>\n",

" \n",

" <style>\n",

" .colab-df-container {\n",

" display:flex;\n",

" flex-wrap:wrap;\n",

" gap: 12px;\n",

" }\n",

"\n",

" .colab-df-convert {\n",

" background-color: #E8F0FE;\n",

" border: none;\n",

" border-radius: 50%;\n",

" cursor: pointer;\n",

" display: none;\n",

" fill: #1967D2;\n",

" height: 32px;\n",

" padding: 0 0 0 0;\n",

" width: 32px;\n",

" }\n",

"\n",

" .colab-df-convert:hover {\n",

" background-color: #E2EBFA;\n",

" box-shadow: 0px 1px 2px rgba(60, 64, 67, 0.3), 0px 1px 3px 1px rgba(60, 64, 67, 0.15);\n",

" fill: #174EA6;\n",

" }\n",

"\n",

" [theme=dark] .colab-df-convert {\n",

" background-color: #3B4455;\n",

" fill: #D2E3FC;\n",

" }\n",

"\n",

" [theme=dark] .colab-df-convert:hover {\n",

" background-color: #434B5C;\n",

" box-shadow: 0px 1px 3px 1px rgba(0, 0, 0, 0.15);\n",

" filter: drop-shadow(0px 1px 2px rgba(0, 0, 0, 0.3));\n",

" fill: #FFFFFF;\n",

" }\n",

" </style>\n",

"\n",

" <script>\n",

" const buttonEl =\n",

" document.querySelector('#df-009b5380-e901-4e79-8bbd-ac8623e89487 button.colab-df-convert');\n",

" buttonEl.style.display =\n",

" google.colab.kernel.accessAllowed ? 'block' : 'none';\n",

"\n",

" async function convertToInteractive(key) {\n",

" const element = document.querySelector('#df-009b5380-e901-4e79-8bbd-ac8623e89487');\n",

" const dataTable =\n",

" await google.colab.kernel.invokeFunction('convertToInteractive',\n",

" [key], {});\n",

" if (!dataTable) return;\n",

"\n",

" const docLinkHtml = 'Like what you see? Visit the ' +\n",

" '<a target=\"\_blank\" href=https://colab.research.google.com/notebooks/data\_table.ipynb>data table notebook</a>'\n",

" + ' to learn more about interactive tables.';\n",

" element.innerHTML = '';\n",

" dataTable['output\_type'] = 'display\_data';\n",

" await google.colab.output.renderOutput(dataTable, element);\n",

" const docLink = document.createElement('div');\n",

" docLink.innerHTML = docLinkHtml;\n",

" element.appendChild(docLink);\n",

" }\n",

" </script>\n",

" </div>\n",

" </div>\n",

" "

]

},

"metadata": {},

"execution\_count": 64

}

],

"source": [

"x\_test\_bal.head()"

]

},

{

"cell\_type": "markdown",

"source": [

"# Model Building"

],

"metadata": {

"id": "4zKl1tlhlqzA"

}

},

{

"cell\_type": "markdown",

"metadata": {

"id": "7cFywyBamZtq"

},

"source": [

"RandomForest Model-1"

]

},

{

"cell\_type": "code",

"execution\_count": null,

"metadata": {

"colab": {

"base\_uri": "https://localhost:8080/"

},

"id": "lhkfNgIwM2be",

"outputId": "5f7b775e-7bf1-4c61-9e34-e8d042c72edc"

},

"outputs": [

{

"output\_type": "stream",

"name": "stderr",

"text": [

"<ipython-input-65-0f47a2fbaab0>:2: 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().\n",

" rfr1.fit(x\_bal,y\_bal)\n"

]

}

],

"source": [

"rfr1 = RandomForestClassifier()\n",

"rfr1.fit(x\_bal,y\_bal)\n",

"y\_pred=rfr1.predict(x\_test\_bal)"

]

},

{

"cell\_type": "code",

"execution\_count": null,

"metadata": {

"colab": {

"base\_uri": "https://localhost:8080/"

},

"id": "V1t0Qu50LJRN",

"outputId": "a862fdf9-6ac6-42e5-d440-94e6a8adafaf"

},

"outputs": [

{

"output\_type": "stream",

"name": "stdout",

"text": [

" precision recall f1-score support\n",

"\n",

" 0 1.00 1.00 1.00 121\n",

" 1 1.00 0.95 0.97 121\n",

" 2 0.94 0.98 0.96 121\n",

" 3 0.94 0.91 0.92 121\n",

" 4 0.99 0.99 0.99 121\n",

" 5 0.87 0.91 0.89 121\n",

" 6 0.99 0.98 0.99 121\n",

"\n",

" accuracy 0.96 847\n",

" macro avg 0.96 0.96 0.96 847\n",

"weighted avg 0.96 0.96 0.96 847\n",

"\n"

]

}

],

"source": [

"print(classification\_report(y\_test\_bal,y\_pred))"

]

},

{

"cell\_type": "code",

"execution\_count": null,

"metadata": {

"id": "ViixHG14zmY0"

},

"outputs": [],

"source": [

"train\_score = accuracy\_score(y\_bal,rfr1.predict(x\_bal))"

]

},

{

"cell\_type": "code",

"execution\_count": null,

"metadata": {

"colab": {

"base\_uri": "https://localhost:8080/"

},

"id": "P3ScT\_GEz2QR",

"outputId": "13eb9edf-04d2-45e3-9ef3-59600711d0eb"

},

"outputs": [

{

"output\_type": "execute\_result",

"data": {

"text/plain": [

"1.0"

]

},

"metadata": {},

"execution\_count": 68

}

],

"source": [

"train\_score"

]

},

{

"cell\_type": "markdown",

"metadata": {

"id": "mifMDyuamyqr"

},

"source": [

"XGBClassfier Model-2"

]

},

{

"cell\_type": "code",

"execution\_count": null,

"metadata": {

"colab": {

"base\_uri": "https://localhost:8080/",

"height": 248

},

"id": "iOV8DC2Iz631",

"outputId": "fdd4d8d8-f79f-44a5-e390-b0c1c427f008"

},

"outputs": [

{

"output\_type": "execute\_result",

"data": {

"text/plain": [

"XGBClassifier(base\_score=None, booster=None, callbacks=None,\n",

" colsample\_bylevel=None, colsample\_bynode=None,\n",

" colsample\_bytree=None, early\_stopping\_rounds=None,\n",

" enable\_categorical=False, eval\_metric=None, feature\_types=None,\n",

" gamma=None, gpu\_id=None, grow\_policy=None, importance\_type=None,\n",

" interaction\_constraints=None, learning\_rate=None, max\_bin=None,\n",

" max\_cat\_threshold=None, max\_cat\_to\_onehot=None,\n",

" max\_delta\_step=None, max\_depth=None, max\_leaves=None,\n",

" min\_child\_weight=None, missing=nan, monotone\_constraints=None,\n",

" n\_estimators=100, n\_jobs=None, num\_parallel\_tree=None,\n",

" objective='multi:softprob', predictor=None, ...)"

],

"text/html": [

"<style>#sk-container-id-1 {color: black;background-color: white;}#sk-container-id-1 pre{padding: 0;}#sk-container-id-1 div.sk-toggleable {background-color: white;}#sk-container-id-1 label.sk-toggleable\_\_label {cursor: pointer;display: block;width: 100%;margin-bottom: 0;padding: 0.3em;box-sizing: border-box;text-align: center;}#sk-container-id-1 label.sk-toggleable\_\_label-arrow:before {content: \"▸\";float: left;margin-right: 0.25em;color: #696969;}#sk-container-id-1 label.sk-toggleable\_\_label-arrow:hover:before {color: black;}#sk-container-id-1 div.sk-estimator:hover label.sk-toggleable\_\_label-arrow:before {color: black;}#sk-container-id-1 div.sk-toggleable\_\_content {max-height: 0;max-width: 0;overflow: hidden;text-align: left;background-color: #f0f8ff;}#sk-container-id-1 div.sk-toggleable\_\_content pre {margin: 0.2em;color: black;border-radius: 0.25em;background-color: #f0f8ff;}#sk-container-id-1 input.sk-toggleable\_\_control:checked~div.sk-toggleable\_\_content {max-height: 200px;max-width: 100%;overflow: auto;}#sk-container-id-1 input.sk-toggleable\_\_control:checked~label.sk-toggleable\_\_label-arrow:before {content: \"▾\";}#sk-container-id-1 div.sk-estimator input.sk-toggleable\_\_control:checked~label.sk-toggleable\_\_label {background-color: #d4ebff;}#sk-container-id-1 div.sk-label input.sk-toggleable\_\_control:checked~label.sk-toggleable\_\_label {background-color: #d4ebff;}#sk-container-id-1 input.sk-hidden--visually {border: 0;clip: rect(1px 1px 1px 1px);clip: rect(1px, 1px, 1px, 1px);height: 1px;margin: -1px;overflow: hidden;padding: 0;position: absolute;width: 1px;}#sk-container-id-1 div.sk-estimator {font-family: monospace;background-color: #f0f8ff;border: 1px dotted black;border-radius: 0.25em;box-sizing: border-box;margin-bottom: 0.5em;}#sk-container-id-1 div.sk-estimator:hover {background-color: #d4ebff;}#sk-container-id-1 div.sk-parallel-item::after {content: \"\";width: 100%;border-bottom: 1px solid gray;flex-grow: 1;}#sk-container-id-1 div.sk-label:hover label.sk-toggleable\_\_label {background-color: #d4ebff;}#sk-container-id-1 div.sk-serial::before {content: \"\";position: absolute;border-left: 1px solid gray;box-sizing: border-box;top: 0;bottom: 0;left: 50%;z-index: 0;}#sk-container-id-1 div.sk-serial {display: flex;flex-direction: column;align-items: center;background-color: white;padding-right: 0.2em;padding-left: 0.2em;position: relative;}#sk-container-id-1 div.sk-item {position: relative;z-index: 1;}#sk-container-id-1 div.sk-parallel {display: flex;align-items: stretch;justify-content: center;background-color: white;position: relative;}#sk-container-id-1 div.sk-item::before, #sk-container-id-1 div.sk-parallel-item::before {content: \"\";position: absolute;border-left: 1px solid gray;box-sizing: border-box;top: 0;bottom: 0;left: 50%;z-index: -1;}#sk-container-id-1 div.sk-parallel-item {display: flex;flex-direction: column;z-index: 1;position: relative;background-color: white;}#sk-container-id-1 div.sk-parallel-item:first-child::after {align-self: flex-end;width: 50%;}#sk-container-id-1 div.sk-parallel-item:last-child::after {align-self: flex-start;width: 50%;}#sk-container-id-1 div.sk-parallel-item:only-child::after {width: 0;}#sk-container-id-1 div.sk-dashed-wrapped {border: 1px dashed gray;margin: 0 0.4em 0.5em 0.4em;box-sizing: border-box;padding-bottom: 0.4em;background-color: white;}#sk-container-id-1 div.sk-label label {font-family: monospace;font-weight: bold;display: inline-block;line-height: 1.2em;}#sk-container-id-1 div.sk-label-container {text-align: center;}#sk-container-id-1 div.sk-container {/\* jupyter's `normalize.less` sets `[hidden] { display: none; }` but bootstrap.min.css set `[hidden] { display: none !important; }` so we also need the `!important` here to be able to override the default hidden behavior on the sphinx rendered scikit-learn.org. See: https://github.com/scikit-learn/scikit-learn/issues/21755 \*/display: inline-block !important;position: relative;}#sk-container-id-1 div.sk-text-repr-fallback {display: none;}</style><div id=\"sk-container-id-1\" class=\"sk-top-container\"><div class=\"sk-text-repr-fallback\"><pre>XGBClassifier(base\_score=None, booster=None, callbacks=None,\n",

" colsample\_bylevel=None, colsample\_bynode=None,\n",

" colsample\_bytree=None, early\_stopping\_rounds=None,\n",

" enable\_categorical=False, eval\_metric=None, feature\_types=None,\n",

" gamma=None, gpu\_id=None, grow\_policy=None, importance\_type=None,\n",

" interaction\_constraints=None, learning\_rate=None, max\_bin=None,\n",

" max\_cat\_threshold=None, max\_cat\_to\_onehot=None,\n",

" max\_delta\_step=None, max\_depth=None, max\_leaves=None,\n",

" min\_child\_weight=None, missing=nan, monotone\_constraints=None,\n",

" n\_estimators=100, n\_jobs=None, num\_parallel\_tree=None,\n",

" objective=&#x27;multi:softprob&#x27;, predictor=None, ...)</pre><b>In a Jupyter environment, please rerun this cell to show the HTML representation or trust the notebook. <br />On GitHub, the HTML representation is unable to render, please try loading this page with nbviewer.org.</b></div><div class=\"sk-container\" hidden><div class=\"sk-item\"><div class=\"sk-estimator sk-toggleable\"><input class=\"sk-toggleable\_\_control sk-hidden--visually\" id=\"sk-estimator-id-1\" type=\"checkbox\" checked><label for=\"sk-estimator-id-1\" class=\"sk-toggleable\_\_label sk-toggleable\_\_label-arrow\">XGBClassifier</label><div class=\"sk-toggleable\_\_content\"><pre>XGBClassifier(base\_score=None, booster=None, callbacks=None,\n",

" colsample\_bylevel=None, colsample\_bynode=None,\n",

" colsample\_bytree=None, early\_stopping\_rounds=None,\n",

" enable\_categorical=False, eval\_metric=None, feature\_types=None,\n",

" gamma=None, gpu\_id=None, grow\_policy=None, importance\_type=None,\n",

" interaction\_constraints=None, learning\_rate=None, max\_bin=None,\n",

" max\_cat\_threshold=None, max\_cat\_to\_onehot=None,\n",

" max\_delta\_step=None, max\_depth=None, max\_leaves=None,\n",

" min\_child\_weight=None, missing=nan, monotone\_constraints=None,\n",

" n\_estimators=100, n\_jobs=None, num\_parallel\_tree=None,\n",

" objective=&#x27;multi:softprob&#x27;, predictor=None, ...)</pre></div></div></div></div></div>"

]

},

"metadata": {},

"execution\_count": 69

}

],

"source": [

"from xgboost import XGBClassifier\n",

"xgb = XGBClassifier()\n",

"xgb.fit(x\_bal,y\_bal)"

]

},

{

"cell\_type": "code",

"execution\_count": null,

"metadata": {

"id": "\_V8Hjhzb0TSI"

},

"outputs": [],

"source": [

"y\_pred-xgb.predict(x\_test\_bal)"

]

},

{

"cell\_type": "code",

"execution\_count": null,

"metadata": {

"colab": {

"base\_uri": "https://localhost:8080/"

},

"id": "MZRSBlPd0m\_G",

"outputId": "8e3c8002-52d4-4944-e7bc-11683177afe1"

},

"outputs": [

{

"output\_type": "stream",

"name": "stdout",

"text": [

" precision recall f1-score support\n",

"\n",

" 0 1.00 1.00 1.00 121\n",

" 1 1.00 0.95 0.97 121\n",

" 2 0.94 0.98 0.96 121\n",

" 3 0.94 0.91 0.92 121\n",

" 4 0.99 0.99 0.99 121\n",

" 5 0.87 0.91 0.89 121\n",

" 6 0.99 0.98 0.99 121\n",

"\n",

" accuracy 0.96 847\n",

" macro avg 0.96 0.96 0.96 847\n",

"weighted avg 0.96 0.96 0.96 847\n",

"\n"

]

}

],

"source": [

"print(classification\_report(y\_test\_bal,y\_pred))"

]

},

{

"cell\_type": "code",

"execution\_count": null,

"metadata": {

"colab": {

"base\_uri": "https://localhost:8080/"

},

"id": "xbdHc4Ed0wyA",

"outputId": "ccd9faa9-175e-4963-df9b-988e9af509d6"

},

"outputs": [

{

"output\_type": "execute\_result",

"data": {

"text/plain": [

"1.0"

]

},

"metadata": {},

"execution\_count": 72

}

],

"source": [

"train\_score = accuracy\_score(y\_bal, xgb.predict(x\_bal))\n",

"train\_score"

]

},

{

"cell\_type": "markdown",

"metadata": {

"id": "SNfJYkVOm9oO"

},

"source": [

"SVC Model-3"

]

},

{

"cell\_type": "code",

"execution\_count": null,

"metadata": {

"id": "E5D6skJu1BDG"

},

"outputs": [],

"source": [

"from sklearn.svm import SVC\n",

"from sklearn.metrics import accuracy\_score, classification\_report\n",

"\n",

"sv= SVC()"

]

},

{

"cell\_type": "code",

"execution\_count": null,

"metadata": {

"colab": {

"base\_uri": "https://localhost:8080/",

"height": 129

},

"id": "hC1DX9OJ1aor",

"outputId": "33316847-124f-45cc-babe-435ef82d9840"

},

"outputs": [

{

"output\_type": "stream",

"name": "stderr",

"text": [

"/usr/local/lib/python3.9/dist-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().\n",

" y = column\_or\_1d(y, warn=True)\n"

]

},

{

"output\_type": "execute\_result",

"data": {

"text/plain": [

"SVC()"

],

"text/html": [

"<style>#sk-container-id-2 {color: black;background-color: white;}#sk-container-id-2 pre{padding: 0;}#sk-container-id-2 div.sk-toggleable {background-color: white;}#sk-container-id-2 label.sk-toggleable\_\_label {cursor: pointer;display: block;width: 100%;margin-bottom: 0;padding: 0.3em;box-sizing: border-box;text-align: center;}#sk-container-id-2 label.sk-toggleable\_\_label-arrow:before {content: \"▸\";float: left;margin-right: 0.25em;color: #696969;}#sk-container-id-2 label.sk-toggleable\_\_label-arrow:hover:before {color: black;}#sk-container-id-2 div.sk-estimator:hover label.sk-toggleable\_\_label-arrow:before {color: black;}#sk-container-id-2 div.sk-toggleable\_\_content {max-height: 0;max-width: 0;overflow: hidden;text-align: left;background-color: #f0f8ff;}#sk-container-id-2 div.sk-toggleable\_\_content pre {margin: 0.2em;color: black;border-radius: 0.25em;background-color: #f0f8ff;}#sk-container-id-2 input.sk-toggleable\_\_control:checked~div.sk-toggleable\_\_content {max-height: 200px;max-width: 100%;overflow: auto;}#sk-container-id-2 input.sk-toggleable\_\_control:checked~label.sk-toggleable\_\_label-arrow:before {content: \"▾\";}#sk-container-id-2 div.sk-estimator input.sk-toggleable\_\_control:checked~label.sk-toggleable\_\_label {background-color: #d4ebff;}#sk-container-id-2 div.sk-label input.sk-toggleable\_\_control:checked~label.sk-toggleable\_\_label {background-color: #d4ebff;}#sk-container-id-2 input.sk-hidden--visually {border: 0;clip: rect(1px 1px 1px 1px);clip: rect(1px, 1px, 1px, 1px);height: 1px;margin: -1px;overflow: hidden;padding: 0;position: absolute;width: 1px;}#sk-container-id-2 div.sk-estimator {font-family: monospace;background-color: #f0f8ff;border: 1px dotted black;border-radius: 0.25em;box-sizing: border-box;margin-bottom: 0.5em;}#sk-container-id-2 div.sk-estimator:hover {background-color: #d4ebff;}#sk-container-id-2 div.sk-parallel-item::after {content: \"\";width: 100%;border-bottom: 1px solid gray;flex-grow: 1;}#sk-container-id-2 div.sk-label:hover label.sk-toggleable\_\_label {background-color: #d4ebff;}#sk-container-id-2 div.sk-serial::before {content: \"\";position: absolute;border-left: 1px solid gray;box-sizing: border-box;top: 0;bottom: 0;left: 50%;z-index: 0;}#sk-container-id-2 div.sk-serial {display: flex;flex-direction: column;align-items: center;background-color: white;padding-right: 0.2em;padding-left: 0.2em;position: relative;}#sk-container-id-2 div.sk-item {position: relative;z-index: 1;}#sk-container-id-2 div.sk-parallel {display: flex;align-items: stretch;justify-content: center;background-color: white;position: relative;}#sk-container-id-2 div.sk-item::before, #sk-container-id-2 div.sk-parallel-item::before {content: \"\";position: absolute;border-left: 1px solid gray;box-sizing: border-box;top: 0;bottom: 0;left: 50%;z-index: -1;}#sk-container-id-2 div.sk-parallel-item {display: flex;flex-direction: column;z-index: 1;position: relative;background-color: white;}#sk-container-id-2 div.sk-parallel-item:first-child::after {align-self: flex-end;width: 50%;}#sk-container-id-2 div.sk-parallel-item:last-child::after {align-self: flex-start;width: 50%;}#sk-container-id-2 div.sk-parallel-item:only-child::after {width: 0;}#sk-container-id-2 div.sk-dashed-wrapped {border: 1px dashed gray;margin: 0 0.4em 0.5em 0.4em;box-sizing: border-box;padding-bottom: 0.4em;background-color: white;}#sk-container-id-2 div.sk-label label {font-family: monospace;font-weight: bold;display: inline-block;line-height: 1.2em;}#sk-container-id-2 div.sk-label-container {text-align: center;}#sk-container-id-2 div.sk-container {/\* jupyter's `normalize.less` sets `[hidden] { display: none; }` but bootstrap.min.css set `[hidden] { display: none !important; }` so we also need the `!important` here to be able to override the default hidden behavior on the sphinx rendered scikit-learn.org. See: https://github.com/scikit-learn/scikit-learn/issues/21755 \*/display: inline-block !important;position: relative;}#sk-container-id-2 div.sk-text-repr-fallback {display: none;}</style><div id=\"sk-container-id-2\" class=\"sk-top-container\"><div class=\"sk-text-repr-fallback\"><pre>SVC()</pre><b>In a Jupyter environment, please rerun this cell to show the HTML representation or trust the notebook. <br />On GitHub, the HTML representation is unable to render, please try loading this page with nbviewer.org.</b></div><div class=\"sk-container\" hidden><div class=\"sk-item\"><div class=\"sk-estimator sk-toggleable\"><input class=\"sk-toggleable\_\_control sk-hidden--visually\" id=\"sk-estimator-id-2\" type=\"checkbox\" checked><label for=\"sk-estimator-id-2\" class=\"sk-toggleable\_\_label sk-toggleable\_\_label-arrow\">SVC</label><div class=\"sk-toggleable\_\_content\"><pre>SVC()</pre></div></div></div></div></div>"

]

},

"metadata": {},

"execution\_count": 74

}

],

"source": [

"sv.fit(x\_bal,y\_bal)"

]

},

{

"cell\_type": "code",

"execution\_count": null,

"metadata": {

"id": "u5PQxSCB1iFm"

},

"outputs": [],

"source": [

"y\_pred = sv.predict(x\_test\_bal)"

]

},

{

"cell\_type": "code",

"execution\_count": null,

"metadata": {

"colab": {

"base\_uri": "https://localhost:8080/"

},

"id": "9bIXSmUq1srb",

"outputId": "4ad72dbe-15f5-4098-c449-30bb83f0bec3"

},

"outputs": [

{

"output\_type": "stream",

"name": "stdout",

"text": [

" precision recall f1-score support\n",

"\n",

" 0 0.89 1.00 0.94 121\n",

" 1 0.86 0.74 0.79 121\n",

" 2 0.75 0.88 0.81 121\n",

" 3 0.78 0.83 0.80 121\n",

" 4 0.84 0.60 0.70 121\n",

" 5 0.75 0.75 0.75 121\n",

" 6 0.91 0.97 0.94 121\n",

"\n",

" accuracy 0.82 847\n",

" macro avg 0.82 0.82 0.82 847\n",

"weighted avg 0.82 0.82 0.82 847\n",

"\n"

]

}

],

"source": [

"print(classification\_report(y\_test\_bal,y\_pred))"

]

},

{

"cell\_type": "code",

"execution\_count": null,

"metadata": {

"colab": {

"base\_uri": "https://localhost:8080/"

},

"id": "azOQQWow2I01",

"outputId": "5042a7d7-8c6b-4f86-ea61-a2b280bcb50c"

},

"outputs": [

{

"output\_type": "execute\_result",

"data": {

"text/plain": [

"0.8450363196125908"

]

},

"metadata": {},

"execution\_count": 77

}

],

"source": [

"train\_score=accuracy\_score(y\_bal,sv.predict(x\_bal))\n",

"train\_score"

]

},

{

"cell\_type": "markdown",

"metadata": {

"id": "ve51KCgWnHo0"

},

"source": [

"Grid\_Search for RandomForest"

]

},

{

"cell\_type": "code",

"execution\_count": null,

"metadata": {

"id": "NFvw9eVh2dLB"

},

"outputs": [],

"source": [

"params={\n",

" 'n\_estimators': [100,200,500],\n",

" 'criterion': ['gini','entropy'],\n",

" 'max\_depth' : [x for x in range(1,20)]\n",

"}"

]

},

{

"cell\_type": "code",

"execution\_count": null,

"metadata": {

"id": "Yj5F-4EF3B5K"

},

"outputs": [],

"source": [

"from sklearn.model\_selection import GridSearchCV\n",

"grid\_search = GridSearchCV(rfr1, params, scoring='accuracy',cv=5,n\_jobs=-1)"

]

},

{

"cell\_type": "code",

"execution\_count": null,

"metadata": {

"colab": {

"base\_uri": "https://localhost:8080/",

"height": 172

},

"id": "hN0aXIo33nkM",

"outputId": "57ae119b-3218-4fb6-b763-4e2b34ce1202"

},

"outputs": [

{

"output\_type": "stream",

"name": "stderr",

"text": [

"/usr/local/lib/python3.9/dist-packages/sklearn/model\_selection/\_search.py:909: 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().\n",

" self.best\_estimator\_.fit(X, y, \*\*fit\_params)\n"

]

},

{

"output\_type": "execute\_result",

"data": {

"text/plain": [

"GridSearchCV(cv=5, estimator=RandomForestClassifier(), n\_jobs=-1,\n",

" param\_grid={'criterion': ['gini', 'entropy'],\n",

" 'max\_depth': [1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12,\n",

" 13, 14, 15, 16, 17, 18, 19],\n",

" 'n\_estimators': [100, 200, 500]},\n",

" scoring='accuracy')"

],

"text/html": [

"<style>#sk-container-id-3 {color: black;background-color: white;}#sk-container-id-3 pre{padding: 0;}#sk-container-id-3 div.sk-toggleable {background-color: white;}#sk-container-id-3 label.sk-toggleable\_\_label {cursor: pointer;display: block;width: 100%;margin-bottom: 0;padding: 0.3em;box-sizing: border-box;text-align: center;}#sk-container-id-3 label.sk-toggleable\_\_label-arrow:before {content: \"▸\";float: left;margin-right: 0.25em;color: #696969;}#sk-container-id-3 label.sk-toggleable\_\_label-arrow:hover:before {color: black;}#sk-container-id-3 div.sk-estimator:hover label.sk-toggleable\_\_label-arrow:before {color: black;}#sk-container-id-3 div.sk-toggleable\_\_content {max-height: 0;max-width: 0;overflow: hidden;text-align: left;background-color: #f0f8ff;}#sk-container-id-3 div.sk-toggleable\_\_content pre {margin: 0.2em;color: black;border-radius: 0.25em;background-color: #f0f8ff;}#sk-container-id-3 input.sk-toggleable\_\_control:checked~div.sk-toggleable\_\_content {max-height: 200px;max-width: 100%;overflow: auto;}#sk-container-id-3 input.sk-toggleable\_\_control:checked~label.sk-toggleable\_\_label-arrow:before {content: \"▾\";}#sk-container-id-3 div.sk-estimator input.sk-toggleable\_\_control:checked~label.sk-toggleable\_\_label {background-color: #d4ebff;}#sk-container-id-3 div.sk-label input.sk-toggleable\_\_control:checked~label.sk-toggleable\_\_label {background-color: #d4ebff;}#sk-container-id-3 input.sk-hidden--visually {border: 0;clip: rect(1px 1px 1px 1px);clip: rect(1px, 1px, 1px, 1px);height: 1px;margin: -1px;overflow: hidden;padding: 0;position: absolute;width: 1px;}#sk-container-id-3 div.sk-estimator {font-family: monospace;background-color: #f0f8ff;border: 1px dotted black;border-radius: 0.25em;box-sizing: border-box;margin-bottom: 0.5em;}#sk-container-id-3 div.sk-estimator:hover {background-color: #d4ebff;}#sk-container-id-3 div.sk-parallel-item::after {content: \"\";width: 100%;border-bottom: 1px solid gray;flex-grow: 1;}#sk-container-id-3 div.sk-label:hover label.sk-toggleable\_\_label {background-color: #d4ebff;}#sk-container-id-3 div.sk-serial::before {content: \"\";position: absolute;border-left: 1px solid gray;box-sizing: border-box;top: 0;bottom: 0;left: 50%;z-index: 0;}#sk-container-id-3 div.sk-serial {display: flex;flex-direction: column;align-items: center;background-color: white;padding-right: 0.2em;padding-left: 0.2em;position: relative;}#sk-container-id-3 div.sk-item {position: relative;z-index: 1;}#sk-container-id-3 div.sk-parallel {display: flex;align-items: stretch;justify-content: center;background-color: white;position: relative;}#sk-container-id-3 div.sk-item::before, #sk-container-id-3 div.sk-parallel-item::before {content: \"\";position: absolute;border-left: 1px solid gray;box-sizing: border-box;top: 0;bottom: 0;left: 50%;z-index: -1;}#sk-container-id-3 div.sk-parallel-item {display: flex;flex-direction: column;z-index: 1;position: relative;background-color: white;}#sk-container-id-3 div.sk-parallel-item:first-child::after {align-self: flex-end;width: 50%;}#sk-container-id-3 div.sk-parallel-item:last-child::after {align-self: flex-start;width: 50%;}#sk-container-id-3 div.sk-parallel-item:only-child::after {width: 0;}#sk-container-id-3 div.sk-dashed-wrapped {border: 1px dashed gray;margin: 0 0.4em 0.5em 0.4em;box-sizing: border-box;padding-bottom: 0.4em;background-color: white;}#sk-container-id-3 div.sk-label label {font-family: monospace;font-weight: bold;display: inline-block;line-height: 1.2em;}#sk-container-id-3 div.sk-label-container {text-align: center;}#sk-container-id-3 div.sk-container {/\* jupyter's `normalize.less` sets `[hidden] { display: none; }` but bootstrap.min.css set `[hidden] { display: none !important; }` so we also need the `!important` here to be able to override the default hidden behavior on the sphinx rendered scikit-learn.org. See: https://github.com/scikit-learn/scikit-learn/issues/21755 \*/display: inline-block !important;position: relative;}#sk-container-id-3 div.sk-text-repr-fallback {display: none;}</style><div id=\"sk-container-id-3\" class=\"sk-top-container\"><div class=\"sk-text-repr-fallback\"><pre>GridSearchCV(cv=5, estimator=RandomForestClassifier(), n\_jobs=-1,\n",

" param\_grid={&#x27;criterion&#x27;: [&#x27;gini&#x27;, &#x27;entropy&#x27;],\n",

" &#x27;max\_depth&#x27;: [1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12,\n",

" 13, 14, 15, 16, 17, 18, 19],\n",

" &#x27;n\_estimators&#x27;: [100, 200, 500]},\n",

" scoring=&#x27;accuracy&#x27;)</pre><b>In a Jupyter environment, please rerun this cell to show the HTML representation or trust the notebook. <br />On GitHub, the HTML representation is unable to render, please try loading this page with nbviewer.org.</b></div><div class=\"sk-container\" hidden><div class=\"sk-item sk-dashed-wrapped\"><div class=\"sk-label-container\"><div class=\"sk-label sk-toggleable\"><input class=\"sk-toggleable\_\_control sk-hidden--visually\" id=\"sk-estimator-id-3\" type=\"checkbox\" ><label for=\"sk-estimator-id-3\" class=\"sk-toggleable\_\_label sk-toggleable\_\_label-arrow\">GridSearchCV</label><div class=\"sk-toggleable\_\_content\"><pre>GridSearchCV(cv=5, estimator=RandomForestClassifier(), n\_jobs=-1,\n",

" param\_grid={&#x27;criterion&#x27;: [&#x27;gini&#x27;, &#x27;entropy&#x27;],\n",

" &#x27;max\_depth&#x27;: [1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12,\n",

" 13, 14, 15, 16, 17, 18, 19],\n",

" &#x27;n\_estimators&#x27;: [100, 200, 500]},\n",

" scoring=&#x27;accuracy&#x27;)</pre></div></div></div><div class=\"sk-parallel\"><div class=\"sk-parallel-item\"><div class=\"sk-item\"><div class=\"sk-label-container\"><div class=\"sk-label sk-toggleable\"><input class=\"sk-toggleable\_\_control sk-hidden--visually\" id=\"sk-estimator-id-4\" type=\"checkbox\" ><label for=\"sk-estimator-id-4\" class=\"sk-toggleable\_\_label sk-toggleable\_\_label-arrow\">estimator: RandomForestClassifier</label><div class=\"sk-toggleable\_\_content\"><pre>RandomForestClassifier()</pre></div></div></div><div class=\"sk-serial\"><div class=\"sk-item\"><div class=\"sk-estimator sk-toggleable\"><input class=\"sk-toggleable\_\_control sk-hidden--visually\" id=\"sk-estimator-id-5\" type=\"checkbox\" ><label for=\"sk-estimator-id-5\" class=\"sk-toggleable\_\_label sk-toggleable\_\_label-arrow\">RandomForestClassifier</label><div class=\"sk-toggleable\_\_content\"><pre>RandomForestClassifier()</pre></div></div></div></div></div></div></div></div></div></div>"

]

},

"metadata": {},

"execution\_count": 80

}

],

"source": [

"grid\_search.fit(x\_bal,y\_bal)"

]

},

{

"cell\_type": "code",

"execution\_count": null,

"metadata": {

"colab": {

"base\_uri": "https://localhost:8080/"

},

"id": "xq9NU-qV37eJ",

"outputId": "03bf8058-a3e9-40c3-e266-4861ff2de54f"

},

"outputs": [

{

"output\_type": "execute\_result",

"data": {

"text/plain": [

"{'criterion': 'entropy', 'max\_depth': 19, 'n\_estimators': 500}"

]

},

"metadata": {},

"execution\_count": 81

}

],

"source": [

"grid\_search.best\_params\_"

]

},

{

"cell\_type": "code",

"execution\_count": null,

"metadata": {

"id": "UlnLv-0-5BYr"

},

"outputs": [],

"source": [

"rfr\_gs=RandomForestClassifier(criterion= 'entropy', max\_depth=16, n\_estimators= 200)"

]

},

{

"cell\_type": "code",

"execution\_count": null,

"metadata": {

"colab": {

"base\_uri": "https://localhost:8080/",

"height": 129

},

"id": "RcN2pPmG5lgj",

"outputId": "4f63086a-c18e-4674-e5c5-108a53ac2522"

},

"outputs": [

{

"output\_type": "stream",

"name": "stderr",

"text": [

"<ipython-input-83-e295d669d733>:1: 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().\n",

" rfr\_gs.fit(x\_bal,y\_bal)\n"

]

},

{

"output\_type": "execute\_result",

"data": {

"text/plain": [

"RandomForestClassifier(criterion='entropy', max\_depth=16, n\_estimators=200)"

],

"text/html": [

"<style>#sk-container-id-4 {color: black;background-color: white;}#sk-container-id-4 pre{padding: 0;}#sk-container-id-4 div.sk-toggleable {background-color: white;}#sk-container-id-4 label.sk-toggleable\_\_label {cursor: pointer;display: block;width: 100%;margin-bottom: 0;padding: 0.3em;box-sizing: border-box;text-align: center;}#sk-container-id-4 label.sk-toggleable\_\_label-arrow:before {content: \"▸\";float: left;margin-right: 0.25em;color: #696969;}#sk-container-id-4 label.sk-toggleable\_\_label-arrow:hover:before {color: black;}#sk-container-id-4 div.sk-estimator:hover label.sk-toggleable\_\_label-arrow:before {color: black;}#sk-container-id-4 div.sk-toggleable\_\_content {max-height: 0;max-width: 0;overflow: hidden;text-align: left;background-color: #f0f8ff;}#sk-container-id-4 div.sk-toggleable\_\_content pre {margin: 0.2em;color: black;border-radius: 0.25em;background-color: #f0f8ff;}#sk-container-id-4 input.sk-toggleable\_\_control:checked~div.sk-toggleable\_\_content {max-height: 200px;max-width: 100%;overflow: auto;}#sk-container-id-4 input.sk-toggleable\_\_control:checked~label.sk-toggleable\_\_label-arrow:before {content: \"▾\";}#sk-container-id-4 div.sk-estimator input.sk-toggleable\_\_control:checked~label.sk-toggleable\_\_label {background-color: #d4ebff;}#sk-container-id-4 div.sk-label input.sk-toggleable\_\_control:checked~label.sk-toggleable\_\_label {background-color: #d4ebff;}#sk-container-id-4 input.sk-hidden--visually {border: 0;clip: rect(1px 1px 1px 1px);clip: rect(1px, 1px, 1px, 1px);height: 1px;margin: -1px;overflow: hidden;padding: 0;position: absolute;width: 1px;}#sk-container-id-4 div.sk-estimator {font-family: monospace;background-color: #f0f8ff;border: 1px dotted black;border-radius: 0.25em;box-sizing: border-box;margin-bottom: 0.5em;}#sk-container-id-4 div.sk-estimator:hover {background-color: #d4ebff;}#sk-container-id-4 div.sk-parallel-item::after {content: \"\";width: 100%;border-bottom: 1px solid gray;flex-grow: 1;}#sk-container-id-4 div.sk-label:hover label.sk-toggleable\_\_label {background-color: #d4ebff;}#sk-container-id-4 div.sk-serial::before {content: \"\";position: absolute;border-left: 1px solid gray;box-sizing: border-box;top: 0;bottom: 0;left: 50%;z-index: 0;}#sk-container-id-4 div.sk-serial {display: flex;flex-direction: column;align-items: center;background-color: white;padding-right: 0.2em;padding-left: 0.2em;position: relative;}#sk-container-id-4 div.sk-item {position: relative;z-index: 1;}#sk-container-id-4 div.sk-parallel {display: flex;align-items: stretch;justify-content: center;background-color: white;position: relative;}#sk-container-id-4 div.sk-item::before, #sk-container-id-4 div.sk-parallel-item::before {content: \"\";position: absolute;border-left: 1px solid gray;box-sizing: border-box;top: 0;bottom: 0;left: 50%;z-index: -1;}#sk-container-id-4 div.sk-parallel-item {display: flex;flex-direction: column;z-index: 1;position: relative;background-color: white;}#sk-container-id-4 div.sk-parallel-item:first-child::after {align-self: flex-end;width: 50%;}#sk-container-id-4 div.sk-parallel-item:last-child::after {align-self: flex-start;width: 50%;}#sk-container-id-4 div.sk-parallel-item:only-child::after {width: 0;}#sk-container-id-4 div.sk-dashed-wrapped {border: 1px dashed gray;margin: 0 0.4em 0.5em 0.4em;box-sizing: border-box;padding-bottom: 0.4em;background-color: white;}#sk-container-id-4 div.sk-label label {font-family: monospace;font-weight: bold;display: inline-block;line-height: 1.2em;}#sk-container-id-4 div.sk-label-container {text-align: center;}#sk-container-id-4 div.sk-container {/\* jupyter's `normalize.less` sets `[hidden] { display: none; }` but bootstrap.min.css set `[hidden] { display: none !important; }` so we also need the `!important` here to be able to override the default hidden behavior on the sphinx rendered scikit-learn.org. See: https://github.com/scikit-learn/scikit-learn/issues/21755 \*/display: inline-block !important;position: relative;}#sk-container-id-4 div.sk-text-repr-fallback {display: none;}</style><div id=\"sk-container-id-4\" class=\"sk-top-container\"><div class=\"sk-text-repr-fallback\"><pre>RandomForestClassifier(criterion=&#x27;entropy&#x27;, max\_depth=16, n\_estimators=200)</pre><b>In a Jupyter environment, please rerun this cell to show the HTML representation or trust the notebook. <br />On GitHub, the HTML representation is unable to render, please try loading this page with nbviewer.org.</b></div><div class=\"sk-container\" hidden><div class=\"sk-item\"><div class=\"sk-estimator sk-toggleable\"><input class=\"sk-toggleable\_\_control sk-hidden--visually\" id=\"sk-estimator-id-6\" type=\"checkbox\" checked><label for=\"sk-estimator-id-6\" class=\"sk-toggleable\_\_label sk-toggleable\_\_label-arrow\">RandomForestClassifier</label><div class=\"sk-toggleable\_\_content\"><pre>RandomForestClassifier(criterion=&#x27;entropy&#x27;, max\_depth=16, n\_estimators=200)</pre></div></div></div></div></div>"

]

},

"metadata": {},

"execution\_count": 83

}

],

"source": [

"rfr\_gs.fit(x\_bal,y\_bal)"

]

},

{

"cell\_type": "code",

"execution\_count": null,

"metadata": {

"id": "r-v0AMZg5rkd"

},

"outputs": [],

"source": [

"y\_pred=rfr\_gs.predict(x\_test\_bal)"

]

},

{

"cell\_type": "code",

"execution\_count": null,

"metadata": {

"colab": {

"base\_uri": "https://localhost:8080/"

},

"id": "otKosoTL6QXK",

"outputId": "062a6c71-ffe6-4fee-f529-31136463ee06"

},

"outputs": [

{

"output\_type": "stream",

"name": "stdout",

"text": [

" precision recall f1-score support\n",

"\n",

" 0 1.00 0.99 1.00 121\n",

" 1 0.98 0.98 0.98 121\n",

" 2 0.96 0.98 0.97 121\n",

" 3 0.94 0.92 0.93 121\n",

" 4 0.99 0.99 0.99 121\n",

" 5 0.88 0.90 0.89 121\n",

" 6 0.99 0.98 0.99 121\n",

"\n",

" accuracy 0.96 847\n",

" macro avg 0.96 0.96 0.96 847\n",

"weighted avg 0.96 0.96 0.96 847\n",

"\n"

]

}

],

"source": [

"print(classification\_report(y\_test\_bal,y\_pred))"

]

},

{

"cell\_type": "code",

"execution\_count": null,

"metadata": {

"colab": {

"base\_uri": "https://localhost:8080/"

},

"id": "I0QZUd-56lPC",

"outputId": "a67fa86c-e941-4f34-9333-4c6c2a2fec66"

},

"outputs": [

{

"output\_type": "execute\_result",

"data": {

"text/plain": [

"1.0"

]

},

"metadata": {},

"execution\_count": 86

}

],

"source": [

"train\_score= accuracy\_score(y\_bal,rfr\_gs.predict(x\_bal))\n",

"train\_score\n"

]

},

{

"cell\_type": "markdown",

"metadata": {

"id": "685HGF9kOktX"

},

"source": [

"Grid\_Search for XGB"

]

},

{

"cell\_type": "code",

"execution\_count": null,

"metadata": {

"id": "QSzhX4ySOtD-"

},

"outputs": [],

"source": [

"params={\n",

" 'n\_estimators': [100, 200, 500],\n",

" 'learning\_rate': [0.01,0.05,0.1],\n",

" 'booster':['gbtree', 'gblinear'],\n",

" 'ganma': [0,0.5, 1],\n",

"\n",

"\n",

" \n",

"}"

]

},

{

"cell\_type": "code",

"execution\_count": null,

"metadata": {

"id": "9Igl9mvhPvuL"

},

"outputs": [],

"source": [

"from sklearn.model\_selection import GridSearchCV\n",

"grid\_xgb = GridSearchCV(xgb, params, scoring='accuracy',cv=3,n\_jobs=-1)"

]

},

{

"cell\_type": "code",

"execution\_count": null,

"metadata": {

"colab": {

"base\_uri": "https://localhost:8080/",

"height": 170

},

"id": "FWVTqfKCQNaV",

"outputId": "aa5e5f67-c007-47cf-b971-3945fae28d11"

},

"outputs": [

{

"output\_type": "stream",

"name": "stdout",

"text": [

"[12:15:42] WARNING: ../src/learner.cc:767: \n",

"Parameters: { \"ganma\" } are not used.\n",

"\n"

]

},

{

"output\_type": "execute\_result",

"data": {

"text/plain": [

"GridSearchCV(cv=3,\n",

" estimator=XGBClassifier(base\_score=None, booster=None,\n",

" callbacks=None, colsample\_bylevel=None,\n",

" colsample\_bynode=None,\n",

" colsample\_bytree=None,\n",

" early\_stopping\_rounds=None,\n",

" enable\_categorical=False, eval\_metric=None,\n",

" feature\_types=None, gamma=None,\n",

" gpu\_id=None, grow\_policy=None,\n",

" importance\_type=None,\n",

" interaction\_constraints=None,\n",

" learning\_rate=None,...\n",

" max\_delta\_step=None, max\_depth=None,\n",

" max\_leaves=None, min\_child\_weight=None,\n",

" missing=nan, monotone\_constraints=None,\n",

" n\_estimators=100, n\_jobs=None,\n",

" num\_parallel\_tree=None,\n",

" objective='multi:softprob', predictor=None, ...),\n",

" n\_jobs=-1,\n",

" param\_grid={'booster': ['gbtree', 'gblinear'],\n",

" 'ganma': [0, 0.5, 1],\n",

" 'learning\_rate': [0.01, 0.05, 0.1],\n",

" 'n\_estimators': [100, 200, 500]},\n",

" scoring='accuracy')"

],

"text/html": [

"<style>#sk-container-id-5 {color: black;background-color: white;}#sk-container-id-5 pre{padding: 0;}#sk-container-id-5 div.sk-toggleable {background-color: white;}#sk-container-id-5 label.sk-toggleable\_\_label {cursor: pointer;display: block;width: 100%;margin-bottom: 0;padding: 0.3em;box-sizing: border-box;text-align: center;}#sk-container-id-5 label.sk-toggleable\_\_label-arrow:before {content: \"▸\";float: left;margin-right: 0.25em;color: #696969;}#sk-container-id-5 label.sk-toggleable\_\_label-arrow:hover:before {color: black;}#sk-container-id-5 div.sk-estimator:hover label.sk-toggleable\_\_label-arrow:before {color: black;}#sk-container-id-5 div.sk-toggleable\_\_content {max-height: 0;max-width: 0;overflow: hidden;text-align: left;background-color: #f0f8ff;}#sk-container-id-5 div.sk-toggleable\_\_content pre {margin: 0.2em;color: black;border-radius: 0.25em;background-color: #f0f8ff;}#sk-container-id-5 input.sk-toggleable\_\_control:checked~div.sk-toggleable\_\_content {max-height: 200px;max-width: 100%;overflow: auto;}#sk-container-id-5 input.sk-toggleable\_\_control:checked~label.sk-toggleable\_\_label-arrow:before {content: \"▾\";}#sk-container-id-5 div.sk-estimator input.sk-toggleable\_\_control:checked~label.sk-toggleable\_\_label {background-color: #d4ebff;}#sk-container-id-5 div.sk-label input.sk-toggleable\_\_control:checked~label.sk-toggleable\_\_label {background-color: #d4ebff;}#sk-container-id-5 input.sk-hidden--visually {border: 0;clip: rect(1px 1px 1px 1px);clip: rect(1px, 1px, 1px, 1px);height: 1px;margin: -1px;overflow: hidden;padding: 0;position: absolute;width: 1px;}#sk-container-id-5 div.sk-estimator {font-family: monospace;background-color: #f0f8ff;border: 1px dotted black;border-radius: 0.25em;box-sizing: border-box;margin-bottom: 0.5em;}#sk-container-id-5 div.sk-estimator:hover {background-color: #d4ebff;}#sk-container-id-5 div.sk-parallel-item::after {content: \"\";width: 100%;border-bottom: 1px solid gray;flex-grow: 1;}#sk-container-id-5 div.sk-label:hover label.sk-toggleable\_\_label {background-color: #d4ebff;}#sk-container-id-5 div.sk-serial::before {content: \"\";position: absolute;border-left: 1px solid gray;box-sizing: border-box;top: 0;bottom: 0;left: 50%;z-index: 0;}#sk-container-id-5 div.sk-serial {display: flex;flex-direction: column;align-items: center;background-color: white;padding-right: 0.2em;padding-left: 0.2em;position: relative;}#sk-container-id-5 div.sk-item {position: relative;z-index: 1;}#sk-container-id-5 div.sk-parallel {display: flex;align-items: stretch;justify-content: center;background-color: white;position: relative;}#sk-container-id-5 div.sk-item::before, #sk-container-id-5 div.sk-parallel-item::before {content: \"\";position: absolute;border-left: 1px solid gray;box-sizing: border-box;top: 0;bottom: 0;left: 50%;z-index: -1;}#sk-container-id-5 div.sk-parallel-item {display: flex;flex-direction: column;z-index: 1;position: relative;background-color: white;}#sk-container-id-5 div.sk-parallel-item:first-child::after {align-self: flex-end;width: 50%;}#sk-container-id-5 div.sk-parallel-item:last-child::after {align-self: flex-start;width: 50%;}#sk-container-id-5 div.sk-parallel-item:only-child::after {width: 0;}#sk-container-id-5 div.sk-dashed-wrapped {border: 1px dashed gray;margin: 0 0.4em 0.5em 0.4em;box-sizing: border-box;padding-bottom: 0.4em;background-color: white;}#sk-container-id-5 div.sk-label label {font-family: monospace;font-weight: bold;display: inline-block;line-height: 1.2em;}#sk-container-id-5 div.sk-label-container {text-align: center;}#sk-container-id-5 div.sk-container {/\* jupyter's `normalize.less` sets `[hidden] { display: none; }` but bootstrap.min.css set `[hidden] { display: none !important; }` so we also need the `!important` here to be able to override the default hidden behavior on the sphinx rendered scikit-learn.org. See: https://github.com/scikit-learn/scikit-learn/issues/21755 \*/display: inline-block !important;position: relative;}#sk-container-id-5 div.sk-text-repr-fallback {display: none;}</style><div id=\"sk-container-id-5\" class=\"sk-top-container\"><div class=\"sk-text-repr-fallback\"><pre>GridSearchCV(cv=3,\n",

" estimator=XGBClassifier(base\_score=None, booster=None,\n",

" callbacks=None, colsample\_bylevel=None,\n",

" colsample\_bynode=None,\n",

" colsample\_bytree=None,\n",

" early\_stopping\_rounds=None,\n",

" enable\_categorical=False, eval\_metric=None,\n",

" feature\_types=None, gamma=None,\n",

" gpu\_id=None, grow\_policy=None,\n",

" importance\_type=None,\n",

" interaction\_constraints=None,\n",

" learning\_rate=None,...\n",

" max\_delta\_step=None, max\_depth=None,\n",

" max\_leaves=None, min\_child\_weight=None,\n",

" missing=nan, monotone\_constraints=None,\n",

" n\_estimators=100, n\_jobs=None,\n",

" num\_parallel\_tree=None,\n",

" objective=&#x27;multi:softprob&#x27;, predictor=None, ...),\n",

" n\_jobs=-1,\n",

" param\_grid={&#x27;booster&#x27;: [&#x27;gbtree&#x27;, &#x27;gblinear&#x27;],\n",

" &#x27;ganma&#x27;: [0, 0.5, 1],\n",

" &#x27;learning\_rate&#x27;: [0.01, 0.05, 0.1],\n",

" &#x27;n\_estimators&#x27;: [100, 200, 500]},\n",

" scoring=&#x27;accuracy&#x27;)</pre><b>In a Jupyter environment, please rerun this cell to show the HTML representation or trust the notebook. <br />On GitHub, the HTML representation is unable to render, please try loading this page with nbviewer.org.</b></div><div class=\"sk-container\" hidden><div class=\"sk-item sk-dashed-wrapped\"><div class=\"sk-label-container\"><div class=\"sk-label sk-toggleable\"><input class=\"sk-toggleable\_\_control sk-hidden--visually\" id=\"sk-estimator-id-7\" type=\"checkbox\" ><label for=\"sk-estimator-id-7\" class=\"sk-toggleable\_\_label sk-toggleable\_\_label-arrow\">GridSearchCV</label><div class=\"sk-toggleable\_\_content\"><pre>GridSearchCV(cv=3,\n",

" estimator=XGBClassifier(base\_score=None, booster=None,\n",

" callbacks=None, colsample\_bylevel=None,\n",

" colsample\_bynode=None,\n",

" colsample\_bytree=None,\n",

" early\_stopping\_rounds=None,\n",

" enable\_categorical=False, eval\_metric=None,\n",

" feature\_types=None, gamma=None,\n",

" gpu\_id=None, grow\_policy=None,\n",

" importance\_type=None,\n",

" interaction\_constraints=None,\n",

" learning\_rate=None,...\n",

" max\_delta\_step=None, max\_depth=None,\n",

" max\_leaves=None, min\_child\_weight=None,\n",

" missing=nan, monotone\_constraints=None,\n",

" n\_estimators=100, n\_jobs=None,\n",

" num\_parallel\_tree=None,\n",

" objective=&#x27;multi:softprob&#x27;, predictor=None, ...),\n",

" n\_jobs=-1,\n",

" param\_grid={&#x27;booster&#x27;: [&#x27;gbtree&#x27;, &#x27;gblinear&#x27;],\n",

" &#x27;ganma&#x27;: [0, 0.5, 1],\n",

" &#x27;learning\_rate&#x27;: [0.01, 0.05, 0.1],\n",

" &#x27;n\_estimators&#x27;: [100, 200, 500]},\n",

" scoring=&#x27;accuracy&#x27;)</pre></div></div></div><div class=\"sk-parallel\"><div class=\"sk-parallel-item\"><div class=\"sk-item\"><div class=\"sk-label-container\"><div class=\"sk-label sk-toggleable\"><input class=\"sk-toggleable\_\_control sk-hidden--visually\" id=\"sk-estimator-id-8\" type=\"checkbox\" ><label for=\"sk-estimator-id-8\" class=\"sk-toggleable\_\_label sk-toggleable\_\_label-arrow\">estimator: XGBClassifier</label><div class=\"sk-toggleable\_\_content\"><pre>XGBClassifier(base\_score=None, booster=None, callbacks=None,\n",

" colsample\_bylevel=None, colsample\_bynode=None,\n",

" colsample\_bytree=None, early\_stopping\_rounds=None,\n",

" enable\_categorical=False, eval\_metric=None, feature\_types=None,\n",

" gamma=None, gpu\_id=None, grow\_policy=None, importance\_type=None,\n",

" interaction\_constraints=None, learning\_rate=None, max\_bin=None,\n",

" max\_cat\_threshold=None, max\_cat\_to\_onehot=None,\n",

" max\_delta\_step=None, max\_depth=None, max\_leaves=None,\n",

" min\_child\_weight=None, missing=nan, monotone\_constraints=None,\n",

" n\_estimators=100, n\_jobs=None, num\_parallel\_tree=None,\n",

" objective=&#x27;multi:softprob&#x27;, predictor=None, ...)</pre></div></div></div><div class=\"sk-serial\"><div class=\"sk-item\"><div class=\"sk-estimator sk-toggleable\"><input class=\"sk-toggleable\_\_control sk-hidden--visually\" id=\"sk-estimator-id-9\" type=\"checkbox\" ><label for=\"sk-estimator-id-9\" class=\"sk-toggleable\_\_label sk-toggleable\_\_label-arrow\">XGBClassifier</label><div class=\"sk-toggleable\_\_content\"><pre>XGBClassifier(base\_score=None, booster=None, callbacks=None,\n",

" colsample\_bylevel=None, colsample\_bynode=None,\n",

" colsample\_bytree=None, early\_stopping\_rounds=None,\n",

" enable\_categorical=False, eval\_metric=None, feature\_types=None,\n",

" gamma=None, gpu\_id=None, grow\_policy=None, importance\_type=None,\n",

" interaction\_constraints=None, learning\_rate=None, max\_bin=None,\n",

" max\_cat\_threshold=None, max\_cat\_to\_onehot=None,\n",

" max\_delta\_step=None, max\_depth=None, max\_leaves=None,\n",

" min\_child\_weight=None, missing=nan, monotone\_constraints=None,\n",

" n\_estimators=100, n\_jobs=None, num\_parallel\_tree=None,\n",

" objective=&#x27;multi:softprob&#x27;, predictor=None, ...)</pre></div></div></div></div></div></div></div></div></div></div>"

]

},

"metadata": {},

"execution\_count": 89

}

],

"source": [

"grid\_xgb.fit(x\_bal,y\_bal)"

]

},

{

"cell\_type": "code",

"execution\_count": null,

"metadata": {

"colab": {

"base\_uri": "https://localhost:8080/"

},

"id": "JoJreYK4Qify",

"outputId": "53a3a4c5-70fc-4740-f8a7-2ea659d5c04a"

},

"outputs": [

{

"output\_type": "execute\_result",

"data": {

"text/plain": [

"{'booster': 'gbtree', 'ganma': 0, 'learning\_rate': 0.1, 'n\_estimators': 200}"

]

},

"metadata": {},

"execution\_count": 90

}

],

"source": [

"grid\_xgb.best\_params\_"

]

},

{

"cell\_type": "code",

"execution\_count": null,

"metadata": {

"id": "CdF602TPUar1"

},

"outputs": [],

"source": [

"xgb1=XGBClassifier(booster='gbtree', gamna= 0, learning\_rate= 0.1, n\_estimators= 500)"

]

},

{

"cell\_type": "code",

"execution\_count": null,

"metadata": {

"colab": {

"base\_uri": "https://localhost:8080/",

"height": 300

},

"id": "OYLLd972X8YZ",

"outputId": "6217270b-b0fb-419c-acb9-4ce0a8244e00"

},

"outputs": [

{

"output\_type": "stream",

"name": "stdout",

"text": [

"[12:15:51] WARNING: ../src/learner.cc:767: \n",

"Parameters: { \"gamna\" } are not used.\n",

"\n"

]

},

{

"output\_type": "execute\_result",

"data": {

"text/plain": [

"XGBClassifier(base\_score=None, booster='gbtree', callbacks=None,\n",

" colsample\_bylevel=None, colsample\_bynode=None,\n",

" colsample\_bytree=None, early\_stopping\_rounds=None,\n",

" enable\_categorical=False, eval\_metric=None, feature\_types=None,\n",

" gamma=None, gamna=0, gpu\_id=None, grow\_policy=None,\n",

" importance\_type=None, interaction\_constraints=None,\n",

" learning\_rate=0.1, max\_bin=None, max\_cat\_threshold=None,\n",

" max\_cat\_to\_onehot=None, max\_delta\_step=None, max\_depth=None,\n",

" max\_leaves=None, min\_child\_weight=None, missing=nan,\n",

" monotone\_constraints=None, n\_estimators=500, n\_jobs=None,\n",

" num\_parallel\_tree=None, objective='multi:softprob', ...)"

],

"text/html": [

"<style>#sk-container-id-6 {color: black;background-color: white;}#sk-container-id-6 pre{padding: 0;}#sk-container-id-6 div.sk-toggleable {background-color: white;}#sk-container-id-6 label.sk-toggleable\_\_label {cursor: pointer;display: block;width: 100%;margin-bottom: 0;padding: 0.3em;box-sizing: border-box;text-align: center;}#sk-container-id-6 label.sk-toggleable\_\_label-arrow:before {content: \"▸\";float: left;margin-right: 0.25em;color: #696969;}#sk-container-id-6 label.sk-toggleable\_\_label-arrow:hover:before {color: black;}#sk-container-id-6 div.sk-estimator:hover label.sk-toggleable\_\_label-arrow:before {color: black;}#sk-container-id-6 div.sk-toggleable\_\_content {max-height: 0;max-width: 0;overflow: hidden;text-align: left;background-color: #f0f8ff;}#sk-container-id-6 div.sk-toggleable\_\_content pre {margin: 0.2em;color: black;border-radius: 0.25em;background-color: #f0f8ff;}#sk-container-id-6 input.sk-toggleable\_\_control:checked~div.sk-toggleable\_\_content {max-height: 200px;max-width: 100%;overflow: auto;}#sk-container-id-6 input.sk-toggleable\_\_control:checked~label.sk-toggleable\_\_label-arrow:before {content: \"▾\";}#sk-container-id-6 div.sk-estimator input.sk-toggleable\_\_control:checked~label.sk-toggleable\_\_label {background-color: #d4ebff;}#sk-container-id-6 div.sk-label input.sk-toggleable\_\_control:checked~label.sk-toggleable\_\_label {background-color: #d4ebff;}#sk-container-id-6 input.sk-hidden--visually {border: 0;clip: rect(1px 1px 1px 1px);clip: rect(1px, 1px, 1px, 1px);height: 1px;margin: -1px;overflow: hidden;padding: 0;position: absolute;width: 1px;}#sk-container-id-6 div.sk-estimator {font-family: monospace;background-color: #f0f8ff;border: 1px dotted black;border-radius: 0.25em;box-sizing: border-box;margin-bottom: 0.5em;}#sk-container-id-6 div.sk-estimator:hover {background-color: #d4ebff;}#sk-container-id-6 div.sk-parallel-item::after {content: \"\";width: 100%;border-bottom: 1px solid gray;flex-grow: 1;}#sk-container-id-6 div.sk-label:hover label.sk-toggleable\_\_label {background-color: #d4ebff;}#sk-container-id-6 div.sk-serial::before {content: \"\";position: absolute;border-left: 1px solid gray;box-sizing: border-box;top: 0;bottom: 0;left: 50%;z-index: 0;}#sk-container-id-6 div.sk-serial {display: flex;flex-direction: column;align-items: center;background-color: white;padding-right: 0.2em;padding-left: 0.2em;position: relative;}#sk-container-id-6 div.sk-item {position: relative;z-index: 1;}#sk-container-id-6 div.sk-parallel {display: flex;align-items: stretch;justify-content: center;background-color: white;position: relative;}#sk-container-id-6 div.sk-item::before, #sk-container-id-6 div.sk-parallel-item::before {content: \"\";position: absolute;border-left: 1px solid gray;box-sizing: border-box;top: 0;bottom: 0;left: 50%;z-index: -1;}#sk-container-id-6 div.sk-parallel-item {display: flex;flex-direction: column;z-index: 1;position: relative;background-color: white;}#sk-container-id-6 div.sk-parallel-item:first-child::after {align-self: flex-end;width: 50%;}#sk-container-id-6 div.sk-parallel-item:last-child::after {align-self: flex-start;width: 50%;}#sk-container-id-6 div.sk-parallel-item:only-child::after {width: 0;}#sk-container-id-6 div.sk-dashed-wrapped {border: 1px dashed gray;margin: 0 0.4em 0.5em 0.4em;box-sizing: border-box;padding-bottom: 0.4em;background-color: white;}#sk-container-id-6 div.sk-label label {font-family: monospace;font-weight: bold;display: inline-block;line-height: 1.2em;}#sk-container-id-6 div.sk-label-container {text-align: center;}#sk-container-id-6 div.sk-container {/\* jupyter's `normalize.less` sets `[hidden] { display: none; }` but bootstrap.min.css set `[hidden] { display: none !important; }` so we also need the `!important` here to be able to override the default hidden behavior on the sphinx rendered scikit-learn.org. See: https://github.com/scikit-learn/scikit-learn/issues/21755 \*/display: inline-block !important;position: relative;}#sk-container-id-6 div.sk-text-repr-fallback {display: none;}</style><div id=\"sk-container-id-6\" class=\"sk-top-container\"><div class=\"sk-text-repr-fallback\"><pre>XGBClassifier(base\_score=None, booster=&#x27;gbtree&#x27;, callbacks=None,\n",

" colsample\_bylevel=None, colsample\_bynode=None,\n",

" colsample\_bytree=None, early\_stopping\_rounds=None,\n",

" enable\_categorical=False, eval\_metric=None, feature\_types=None,\n",

" gamma=None, gamna=0, gpu\_id=None, grow\_policy=None,\n",

" importance\_type=None, interaction\_constraints=None,\n",

" learning\_rate=0.1, max\_bin=None, max\_cat\_threshold=None,\n",

" max\_cat\_to\_onehot=None, max\_delta\_step=None, max\_depth=None,\n",

" max\_leaves=None, min\_child\_weight=None, missing=nan,\n",

" monotone\_constraints=None, n\_estimators=500, n\_jobs=None,\n",

" num\_parallel\_tree=None, objective=&#x27;multi:softprob&#x27;, ...)</pre><b>In a Jupyter environment, please rerun this cell to show the HTML representation or trust the notebook. <br />On GitHub, the HTML representation is unable to render, please try loading this page with nbviewer.org.</b></div><div class=\"sk-container\" hidden><div class=\"sk-item\"><div class=\"sk-estimator sk-toggleable\"><input class=\"sk-toggleable\_\_control sk-hidden--visually\" id=\"sk-estimator-id-10\" type=\"checkbox\" checked><label for=\"sk-estimator-id-10\" class=\"sk-toggleable\_\_label sk-toggleable\_\_label-arrow\">XGBClassifier</label><div class=\"sk-toggleable\_\_content\"><pre>XGBClassifier(base\_score=None, booster=&#x27;gbtree&#x27;, callbacks=None,\n",

" colsample\_bylevel=None, colsample\_bynode=None,\n",

" colsample\_bytree=None, early\_stopping\_rounds=None,\n",

" enable\_categorical=False, eval\_metric=None, feature\_types=None,\n",

" gamma=None, gamna=0, gpu\_id=None, grow\_policy=None,\n",

" importance\_type=None, interaction\_constraints=None,\n",

" learning\_rate=0.1, max\_bin=None, max\_cat\_threshold=None,\n",

" max\_cat\_to\_onehot=None, max\_delta\_step=None, max\_depth=None,\n",

" max\_leaves=None, min\_child\_weight=None, missing=nan,\n",

" monotone\_constraints=None, n\_estimators=500, n\_jobs=None,\n",

" num\_parallel\_tree=None, objective=&#x27;multi:softprob&#x27;, ...)</pre></div></div></div></div></div>"

]

},

"metadata": {},

"execution\_count": 92

}

],

"source": [

"xgb1.fit(x\_bal,y\_bal)"

]

},

{

"cell\_type": "code",

"execution\_count": null,

"metadata": {

"id": "aQP1zJlqYNRf"

},

"outputs": [],

"source": [

"y\_pred= xgb1.predict(x\_test\_bal)"

]

},

{

"cell\_type": "code",

"execution\_count": null,

"metadata": {

"colab": {

"base\_uri": "https://localhost:8080/"

},

"id": "S4hk72DGYZJC",

"outputId": "f063a8c8-ef44-4e5b-ce33-be88fb7abbbb"

},

"outputs": [

{

"output\_type": "stream",

"name": "stdout",

"text": [

" precision recall f1-score support\n",

"\n",

" 0 0.98 1.00 0.99 121\n",

" 1 0.98 0.98 0.98 121\n",

" 2 0.95 0.98 0.97 121\n",

" 3 0.93 0.88 0.91 121\n",

" 4 0.99 0.98 0.98 121\n",

" 5 0.86 0.89 0.88 121\n",

" 6 0.99 0.98 0.98 121\n",

"\n",

" accuracy 0.96 847\n",

" macro avg 0.96 0.96 0.96 847\n",

"weighted avg 0.96 0.96 0.96 847\n",

"\n"

]

}

],

"source": [

"print(classification\_report(y\_test\_bal,y\_pred))"

]

},

{

"cell\_type": "code",

"execution\_count": null,

"metadata": {

"colab": {

"base\_uri": "https://localhost:8080/"

},

"id": "Mu2ycZQpYmVY",

"outputId": "22d9e074-ec88-4268-811a-08187dd39696"

},

"outputs": [

{

"output\_type": "execute\_result",

"data": {

"text/plain": [

"1.0"

]

},

"metadata": {},

"execution\_count": 95

}

],

"source": [

"train\_score= accuracy\_score(y\_bal,xgb1.predict(x\_bal))\n",

"train\_score"

]

},

{

"cell\_type": "markdown",

"source": [

"# Performance Testing & Hyperparameter Tuning"

],

"metadata": {

"id": "m3ZOCCu1n6NX"

}

},

{

"cell\_type": "markdown",

"metadata": {

"id": "nYOJSDZ5Y7Mv"

},

"source": [

" Grid\_Search for SVC"

]

},

{

"cell\_type": "code",

"execution\_count": null,

"metadata": {

"id": "3Q3PqZbIZL13"

},

"outputs": [],

"source": [

"params = {\n",

" 'C' : [0.01, 1, 10,100, 1000],\n",

" 'kernel': ['rbf', 'linear'],\n",

" 'gamma': [1, 0.1, 0.01, 0.001, 0.0001],\n",

"}"

]

},

{

"cell\_type": "code",

"execution\_count": null,

"metadata": {

"id": "z05pevf7oLmo"

},

"outputs": [],

"source": [

"from sklearn.svm import SVC\n",

"from sklearn.metrics import accuracy\_score, classification\_report\n",

"\n",

"sv= SVC()"

]

},

{

"cell\_type": "code",

"execution\_count": null,

"metadata": {

"id": "ow4fP03fZytw"

},

"outputs": [],

"source": [

"from sklearn.model\_selection import GridSearchCV\n",

"grid\_svc = GridSearchCV(sv,params,scoring='accuracy',cv=5,n\_jobs=-1)"

]

},

{

"cell\_type": "code",

"execution\_count": null,

"metadata": {

"colab": {

"base\_uri": "https://localhost:8080/",

"height": 488

},

"id": "UvU0IjgrrV8F",

"outputId": "76405db7-e69f-4c36-f6e2-945ddd7efd66"

},

"outputs": [

{

"output\_type": "execute\_result",

"data": {

"text/plain": [

" age sex on\_thyroxine query\_on\_thyroxine \\\n",

"0 -0.193783 -0.433661 -0.423767 -0.104957 \n",

"1 1.065908 -0.433661 -0.423767 -0.104957 \n",

"2 -1.086064 -0.433661 -0.423767 -0.104957 \n",

"3 -1.978345 -0.433661 -0.423767 -0.104957 \n",

"4 0.121140 -0.433661 -0.423767 -0.104957 \n",

"... ... ... ... ... \n",

"3299 1.098538 -0.433661 2.359786 -0.104957 \n",

"3300 0.631632 -0.433661 2.359786 -0.104957 \n",

"3301 0.488550 -0.433661 2.359786 -0.104957 \n",

"3302 0.381589 -0.433661 2.359786 -0.104957 \n",

"3303 -1.235652 -0.433661 2.359786 -0.104957 \n",

"\n",

" on\_antithyroid\_meds sick pregnant thyroid\_surgery \\\n",

"0 -0.074012 -0.140546 -0.13829 -0.281129 \n",

"1 -0.074012 -0.140546 -0.13829 -0.281129 \n",

"2 -0.074012 -0.140546 -0.13829 -0.281129 \n",

"3 -0.074012 -0.140546 -0.13829 -0.281129 \n",

"4 -0.074012 -0.140546 -0.13829 -0.281129 \n",

"... ... ... ... ... \n",

"3299 -0.074012 -0.140546 -0.13829 -0.281129 \n",

"3300 -0.074012 -0.140546 -0.13829 -0.281129 \n",

"3301 -0.074012 -0.140546 -0.13829 -0.281129 \n",

"3302 -0.074012 -0.140546 -0.13829 -0.281129 \n",

"3303 -0.074012 -0.140546 -0.13829 -0.281129 \n",

"\n",

" I131\_treatment query\_hypothyroid ... goitre tumor \\\n",

"0 -0.23 -0.234344 ... -0.052263 -0.131306 \n",

"1 -0.23 -0.234344 ... -0.052263 -0.131306 \n",

"2 -0.23 -0.234344 ... -0.052263 -0.131306 \n",

"3 -0.23 -0.234344 ... -0.052263 -0.131306 \n",

"4 -0.23 -0.234344 ... -0.052263 -0.131306 \n",

"... ... ... ... ... ... \n",

"3299 -0.23 -0.234344 ... -0.052263 -0.131306 \n",

"3300 -0.23 -0.234344 ... -0.052263 -0.131306 \n",

"3301 -0.23 -0.234344 ... -0.052263 -0.131306 \n",

"3302 -0.23 -0.234344 ... -0.052263 -0.131306 \n",

"3303 -0.23 -0.234344 ... -0.052263 -0.131306 \n",

"\n",

" hypopituitary psych TSH T3 TT4 T4U \\\n",

"0 -0.0174 -0.114831 -0.322733 -1.014991 1.340750 1.972475 \n",

"1 -0.0174 -0.114831 0.056471 -1.014991 -0.709600 -0.470696 \n",

"2 -0.0174 -0.114831 -0.272911 1.290950 -0.094495 1.865083 \n",

"3 -0.0174 -0.114831 -0.322733 -1.014991 0.579192 0.012569 \n",

"4 -0.0174 8.708456 -0.231392 -0.578732 -0.577792 -0.551240 \n",

"... ... ... ... ... ... ... \n",

"3299 -0.0174 -0.114831 -0.320825 -1.014991 0.990849 -0.001212 \n",

"3300 -0.0174 -0.114831 -0.316496 0.392821 1.281566 0.023602 \n",

"3301 -0.0174 -0.114831 -0.320710 -0.789038 0.518302 -0.197983 \n",

"3302 -0.0174 -0.114831 -0.320473 -1.004140 0.140829 -0.389339 \n",

"3303 -0.0174 -0.114831 0.072691 -0.919636 -0.024709 0.558658 \n",

"\n",

" FTI TBG \n",

"0 0.100987 -0.187303 \n",

"1 -0.330151 -0.187303 \n",

"2 -0.593625 -0.187303 \n",

"3 0.592006 1.862310 \n",

"4 -0.138534 -0.187303 \n",

"... ... ... \n",

"3299 0.976857 -0.187303 \n",

"3300 1.204426 -0.187303 \n",

"3301 0.728168 -0.187303 \n",

"3302 0.532307 -0.187303 \n",

"3303 -0.194342 -0.187303 \n",

"\n",

"[3304 rows x 22 columns]"

],

"text/html": [

"\n",

" <div id=\"df-57e0d662-073b-4344-9d8b-4f7362445a83\">\n",

" <div class=\"colab-df-container\">\n",

" <div>\n",

"<style scoped>\n",

" .dataframe tbody tr th:only-of-type {\n",

" vertical-align: middle;\n",

" }\n",

"\n",

" .dataframe tbody tr th {\n",

" vertical-align: top;\n",

" }\n",

"\n",

" .dataframe thead th {\n",

" text-align: right;\n",

" }\n",

"</style>\n",

"<table border=\"1\" class=\"dataframe\">\n",

" <thead>\n",

" <tr style=\"text-align: right;\">\n",

" <th></th>\n",

" <th>age</th>\n",

" <th>sex</th>\n",

" <th>on\_thyroxine</th>\n",

" <th>query\_on\_thyroxine</th>\n",

" <th>on\_antithyroid\_meds</th>\n",

" <th>sick</th>\n",

" <th>pregnant</th>\n",

" <th>thyroid\_surgery</th>\n",

" <th>I131\_treatment</th>\n",

" <th>query\_hypothyroid</th>\n",

" <th>...</th>\n",

" <th>goitre</th>\n",

" <th>tumor</th>\n",

" <th>hypopituitary</th>\n",

" <th>psych</th>\n",

" <th>TSH</th>\n",

" <th>T3</th>\n",

" <th>TT4</th>\n",

" <th>T4U</th>\n",

" <th>FTI</th>\n",

" <th>TBG</th>\n",

" </tr>\n",

" </thead>\n",

" <tbody>\n",

" <tr>\n",

" <th>0</th>\n",

" <td>-0.193783</td>\n",

" <td>-0.433661</td>\n",

" <td>-0.423767</td>\n",

" <td>-0.104957</td>\n",

" <td>-0.074012</td>\n",

" <td>-0.140546</td>\n",

" <td>-0.13829</td>\n",

" <td>-0.281129</td>\n",

" <td>-0.23</td>\n",

" <td>-0.234344</td>\n",

" <td>...</td>\n",

" <td>-0.052263</td>\n",

" <td>-0.131306</td>\n",

" <td>-0.0174</td>\n",

" <td>-0.114831</td>\n",

" <td>-0.322733</td>\n",

" <td>-1.014991</td>\n",

" <td>1.340750</td>\n",

" <td>1.972475</td>\n",

" <td>0.100987</td>\n",

" <td>-0.187303</td>\n",

" </tr>\n",

" <tr>\n",

" <th>1</th>\n",

" <td>1.065908</td>\n",

" <td>-0.433661</td>\n",

" <td>-0.423767</td>\n",

" <td>-0.104957</td>\n",

" <td>-0.074012</td>\n",

" <td>-0.140546</td>\n",

" <td>-0.13829</td>\n",

" <td>-0.281129</td>\n",

" <td>-0.23</td>\n",

" <td>-0.234344</td>\n",

" <td>...</td>\n",

" <td>-0.052263</td>\n",

" <td>-0.131306</td>\n",

" <td>-0.0174</td>\n",

" <td>-0.114831</td>\n",

" <td>0.056471</td>\n",

" <td>-1.014991</td>\n",

" <td>-0.709600</td>\n",

" <td>-0.470696</td>\n",

" <td>-0.330151</td>\n",

" <td>-0.187303</td>\n",

" </tr>\n",

" <tr>\n",

" <th>2</th>\n",

" <td>-1.086064</td>\n",

" <td>-0.433661</td>\n",

" <td>-0.423767</td>\n",

" <td>-0.104957</td>\n",

" <td>-0.074012</td>\n",

" <td>-0.140546</td>\n",

" <td>-0.13829</td>\n",

" <td>-0.281129</td>\n",

" <td>-0.23</td>\n",

" <td>-0.234344</td>\n",

" <td>...</td>\n",

" <td>-0.052263</td>\n",

" <td>-0.131306</td>\n",

" <td>-0.0174</td>\n",

" <td>-0.114831</td>\n",

" <td>-0.272911</td>\n",

" <td>1.290950</td>\n",

" <td>-0.094495</td>\n",

" <td>1.865083</td>\n",

" <td>-0.593625</td>\n",

" <td>-0.187303</td>\n",

" </tr>\n",

" <tr>\n",

" <th>3</th>\n",

" <td>-1.978345</td>\n",

" <td>-0.433661</td>\n",

" <td>-0.423767</td>\n",

" <td>-0.104957</td>\n",

" <td>-0.074012</td>\n",

" <td>-0.140546</td>\n",

" <td>-0.13829</td>\n",

" <td>-0.281129</td>\n",

" <td>-0.23</td>\n",

" <td>-0.234344</td>\n",

" <td>...</td>\n",

" <td>-0.052263</td>\n",

" <td>-0.131306</td>\n",

" <td>-0.0174</td>\n",

" <td>-0.114831</td>\n",

" <td>-0.322733</td>\n",

" <td>-1.014991</td>\n",

" <td>0.579192</td>\n",

" <td>0.012569</td>\n",

" <td>0.592006</td>\n",

" <td>1.862310</td>\n",

" </tr>\n",

" <tr>\n",

" <th>4</th>\n",

" <td>0.121140</td>\n",

" <td>-0.433661</td>\n",

" <td>-0.423767</td>\n",

" <td>-0.104957</td>\n",

" <td>-0.074012</td>\n",

" <td>-0.140546</td>\n",

" <td>-0.13829</td>\n",

" <td>-0.281129</td>\n",

" <td>-0.23</td>\n",

" <td>-0.234344</td>\n",

" <td>...</td>\n",

" <td>-0.052263</td>\n",

" <td>-0.131306</td>\n",

" <td>-0.0174</td>\n",

" <td>8.708456</td>\n",

" <td>-0.231392</td>\n",

" <td>-0.578732</td>\n",

" <td>-0.577792</td>\n",

" <td>-0.551240</td>\n",

" <td>-0.138534</td>\n",

" <td>-0.187303</td>\n",

" </tr>\n",

" <tr>\n",

" <th>...</th>\n",

" <td>...</td>\n",

" <td>...</td>\n",

" <td>...</td>\n",

" <td>...</td>\n",

" <td>...</td>\n",

" <td>...</td>\n",

" <td>...</td>\n",

" <td>...</td>\n",

" <td>...</td>\n",

" <td>...</td>\n",

" <td>...</td>\n",

" <td>...</td>\n",

" <td>...</td>\n",

" <td>...</td>\n",

" <td>...</td>\n",

" <td>...</td>\n",

" <td>...</td>\n",

" <td>...</td>\n",

" <td>...</td>\n",

" <td>...</td>\n",

" <td>...</td>\n",

" </tr>\n",

" <tr>\n",

" <th>3299</th>\n",

" <td>1.098538</td>\n",

" <td>-0.433661</td>\n",

" <td>2.359786</td>\n",

" <td>-0.104957</td>\n",

" <td>-0.074012</td>\n",

" <td>-0.140546</td>\n",

" <td>-0.13829</td>\n",

" <td>-0.281129</td>\n",

" <td>-0.23</td>\n",

" <td>-0.234344</td>\n",

" <td>...</td>\n",

" <td>-0.052263</td>\n",

" <td>-0.131306</td>\n",

" <td>-0.0174</td>\n",

" <td>-0.114831</td>\n",

" <td>-0.320825</td>\n",

" <td>-1.014991</td>\n",

" <td>0.990849</td>\n",

" <td>-0.001212</td>\n",

" <td>0.976857</td>\n",

" <td>-0.187303</td>\n",

" </tr>\n",

" <tr>\n",

" <th>3300</th>\n",

" <td>0.631632</td>\n",

" <td>-0.433661</td>\n",

" <td>2.359786</td>\n",

" <td>-0.104957</td>\n",

" <td>-0.074012</td>\n",

" <td>-0.140546</td>\n",

" <td>-0.13829</td>\n",

" <td>-0.281129</td>\n",

" <td>-0.23</td>\n",

" <td>-0.234344</td>\n",

" <td>...</td>\n",

" <td>-0.052263</td>\n",

" <td>-0.131306</td>\n",

" <td>-0.0174</td>\n",

" <td>-0.114831</td>\n",

" <td>-0.316496</td>\n",

" <td>0.392821</td>\n",

" <td>1.281566</td>\n",

" <td>0.023602</td>\n",

" <td>1.204426</td>\n",

" <td>-0.187303</td>\n",

" </tr>\n",

" <tr>\n",

" <th>3301</th>\n",

" <td>0.488550</td>\n",

" <td>-0.433661</td>\n",

" <td>2.359786</td>\n",

" <td>-0.104957</td>\n",

" <td>-0.074012</td>\n",

" <td>-0.140546</td>\n",

" <td>-0.13829</td>\n",

" <td>-0.281129</td>\n",

" <td>-0.23</td>\n",

" <td>-0.234344</td>\n",

" <td>...</td>\n",

" <td>-0.052263</td>\n",

" <td>-0.131306</td>\n",

" <td>-0.0174</td>\n",

" <td>-0.114831</td>\n",

" <td>-0.320710</td>\n",

" <td>-0.789038</td>\n",

" <td>0.518302</td>\n",

" <td>-0.197983</td>\n",

" <td>0.728168</td>\n",

" <td>-0.187303</td>\n",

" </tr>\n",

" <tr>\n",

" <th>3302</th>\n",

" <td>0.381589</td>\n",

" <td>-0.433661</td>\n",

" <td>2.359786</td>\n",

" <td>-0.104957</td>\n",

" <td>-0.074012</td>\n",

" <td>-0.140546</td>\n",

" <td>-0.13829</td>\n",

" <td>-0.281129</td>\n",

" <td>-0.23</td>\n",

" <td>-0.234344</td>\n",

" <td>...</td>\n",

" <td>-0.052263</td>\n",

" <td>-0.131306</td>\n",

" <td>-0.0174</td>\n",

" <td>-0.114831</td>\n",

" <td>-0.320473</td>\n",

" <td>-1.004140</td>\n",

" <td>0.140829</td>\n",

" <td>-0.389339</td>\n",

" <td>0.532307</td>\n",

" <td>-0.187303</td>\n",

" </tr>\n",

" <tr>\n",

" <th>3303</th>\n",

" <td>-1.235652</td>\n",

" <td>-0.433661</td>\n",

" <td>2.359786</td>\n",

" <td>-0.104957</td>\n",

" <td>-0.074012</td>\n",

" <td>-0.140546</td>\n",

" <td>-0.13829</td>\n",

" <td>-0.281129</td>\n",

" <td>-0.23</td>\n",

" <td>-0.234344</td>\n",

" <td>...</td>\n",

" <td>-0.052263</td>\n",

" <td>-0.131306</td>\n",

" <td>-0.0174</td>\n",

" <td>-0.114831</td>\n",

" <td>0.072691</td>\n",

" <td>-0.919636</td>\n",

" <td>-0.024709</td>\n",

" <td>0.558658</td>\n",

" <td>-0.194342</td>\n",

" <td>-0.187303</td>\n",

" </tr>\n",

" </tbody>\n",

"</table>\n",

"<p>3304 rows × 22 columns</p>\n",

"</div>\n",

" <button class=\"colab-df-convert\" onclick=\"convertToInteractive('df-57e0d662-073b-4344-9d8b-4f7362445a83')\"\n",

" title=\"Convert this dataframe to an interactive table.\"\n",

" style=\"display:none;\">\n",

" \n",

" <svg xmlns=\"http://www.w3.org/2000/svg\" height=\"24px\"viewBox=\"0 0 24 24\"\n",

" width=\"24px\">\n",

" <path d=\"M0 0h24v24H0V0z\" fill=\"none\"/>\n",

" <path d=\"M18.56 5.44l.94 2.06.94-2.06 2.06-.94-2.06-.94-.94-2.06-.94 2.06-2.06.94zm-11 1L8.5 8.5l.94-2.06 2.06-.94-2.06-.94L8.5 2.5l-.94 2.06-2.06.94zm10 10l.94 2.06.94-2.06 2.06-.94-2.06-.94-.94-2.06-.94 2.06-2.06.94z\"/><path d=\"M17.41 7.96l-1.37-1.37c-.4-.4-.92-.59-1.43-.59-.52 0-1.04.2-1.43.59L10.3 9.45l-7.72 7.72c-.78.78-.78 2.05 0 2.83L4 21.41c.39.39.9.59 1.41.59.51 0 1.02-.2 1.41-.59l7.78-7.78 2.81-2.81c.8-.78.8-2.07 0-2.86zM5.41 20L4 18.59l7.72-7.72 1.47 1.35L5.41 20z\"/>\n",

" </svg>\n",

" </button>\n",

" \n",

" <style>\n",

" .colab-df-container {\n",

" display:flex;\n",

" flex-wrap:wrap;\n",

" gap: 12px;\n",

" }\n",

"\n",

" .colab-df-convert {\n",

" background-color: #E8F0FE;\n",

" border: none;\n",

" border-radius: 50%;\n",

" cursor: pointer;\n",

" display: none;\n",

" fill: #1967D2;\n",

" height: 32px;\n",

" padding: 0 0 0 0;\n",

" width: 32px;\n",

" }\n",

"\n",

" .colab-df-convert:hover {\n",

" background-color: #E2EBFA;\n",

" box-shadow: 0px 1px 2px rgba(60, 64, 67, 0.3), 0px 1px 3px 1px rgba(60, 64, 67, 0.15);\n",

" fill: #174EA6;\n",

" }\n",

"\n",

" [theme=dark] .colab-df-convert {\n",

" background-color: #3B4455;\n",

" fill: #D2E3FC;\n",

" }\n",

"\n",

" [theme=dark] .colab-df-convert:hover {\n",

" background-color: #434B5C;\n",

" box-shadow: 0px 1px 3px 1px rgba(0, 0, 0, 0.15);\n",

" filter: drop-shadow(0px 1px 2px rgba(0, 0, 0, 0.3));\n",

" fill: #FFFFFF;\n",

" }\n",

" </style>\n",

"\n",

" <script>\n",

" const buttonEl =\n",

" document.querySelector('#df-57e0d662-073b-4344-9d8b-4f7362445a83 button.colab-df-convert');\n",

" buttonEl.style.display =\n",

" google.colab.kernel.accessAllowed ? 'block' : 'none';\n",

"\n",

" async function convertToInteractive(key) {\n",

" const element = document.querySelector('#df-57e0d662-073b-4344-9d8b-4f7362445a83');\n",

" const dataTable =\n",

" await google.colab.kernel.invokeFunction('convertToInteractive',\n",

" [key], {});\n",

" if (!dataTable) return;\n",

"\n",

" const docLinkHtml = 'Like what you see? Visit the ' +\n",

" '<a target=\"\_blank\" href=https://colab.research.google.com/notebooks/data\_table.ipynb>data table notebook</a>'\n",

" + ' to learn more about interactive tables.';\n",

" element.innerHTML = '';\n",

" dataTable['output\_type'] = 'display\_data';\n",

" await google.colab.output.renderOutput(dataTable, element);\n",

" const docLink = document.createElement('div');\n",

" docLink.innerHTML = docLinkHtml;\n",

" element.appendChild(docLink);\n",

" }\n",

" </script>\n",

" </div>\n",

" </div>\n",

" "

]

},

"metadata": {},

"execution\_count": 99

}

],

"source": [

"x\_bal"

]

},

{

"cell\_type": "code",

"execution\_count": null,

"metadata": {

"colab": {

"base\_uri": "https://localhost:8080/",

"height": 172

},

"id": "m0KWgdOwaw0s",

"outputId": "e6c04421-7c26-476b-de33-4800dca42b07"

},

"outputs": [

{

"output\_type": "stream",

"name": "stderr",

"text": [

"/usr/local/lib/python3.9/dist-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().\n",

" y = column\_or\_1d(y, warn=True)\n"

]

},

{

"output\_type": "execute\_result",

"data": {

"text/plain": [

"GridSearchCV(cv=5, estimator=SVC(), n\_jobs=-1,\n",

" param\_grid={'C': [0.01, 1, 10, 100, 1000],\n",

" 'gamma': [1, 0.1, 0.01, 0.001, 0.0001],\n",

" 'kernel': ['rbf', 'linear']},\n",

" scoring='accuracy')"

],

"text/html": [

"<style>#sk-container-id-7 {color: black;background-color: white;}#sk-container-id-7 pre{padding: 0;}#sk-container-id-7 div.sk-toggleable {background-color: white;}#sk-container-id-7 label.sk-toggleable\_\_label {cursor: pointer;display: block;width: 100%;margin-bottom: 0;padding: 0.3em;box-sizing: border-box;text-align: center;}#sk-container-id-7 label.sk-toggleable\_\_label-arrow:before {content: \"▸\";float: left;margin-right: 0.25em;color: #696969;}#sk-container-id-7 label.sk-toggleable\_\_label-arrow:hover:before {color: black;}#sk-container-id-7 div.sk-estimator:hover label.sk-toggleable\_\_label-arrow:before {color: black;}#sk-container-id-7 div.sk-toggleable\_\_content {max-height: 0;max-width: 0;overflow: hidden;text-align: left;background-color: #f0f8ff;}#sk-container-id-7 div.sk-toggleable\_\_content pre {margin: 0.2em;color: black;border-radius: 0.25em;background-color: #f0f8ff;}#sk-container-id-7 input.sk-toggleable\_\_control:checked~div.sk-toggleable\_\_content {max-height: 200px;max-width: 100%;overflow: auto;}#sk-container-id-7 input.sk-toggleable\_\_control:checked~label.sk-toggleable\_\_label-arrow:before {content: \"▾\";}#sk-container-id-7 div.sk-estimator input.sk-toggleable\_\_control:checked~label.sk-toggleable\_\_label {background-color: #d4ebff;}#sk-container-id-7 div.sk-label input.sk-toggleable\_\_control:checked~label.sk-toggleable\_\_label {background-color: #d4ebff;}#sk-container-id-7 input.sk-hidden--visually {border: 0;clip: rect(1px 1px 1px 1px);clip: rect(1px, 1px, 1px, 1px);height: 1px;margin: -1px;overflow: hidden;padding: 0;position: absolute;width: 1px;}#sk-container-id-7 div.sk-estimator {font-family: monospace;background-color: #f0f8ff;border: 1px dotted black;border-radius: 0.25em;box-sizing: border-box;margin-bottom: 0.5em;}#sk-container-id-7 div.sk-estimator:hover {background-color: #d4ebff;}#sk-container-id-7 div.sk-parallel-item::after {content: \"\";width: 100%;border-bottom: 1px solid gray;flex-grow: 1;}#sk-container-id-7 div.sk-label:hover label.sk-toggleable\_\_label {background-color: #d4ebff;}#sk-container-id-7 div.sk-serial::before {content: \"\";position: absolute;border-left: 1px solid gray;box-sizing: border-box;top: 0;bottom: 0;left: 50%;z-index: 0;}#sk-container-id-7 div.sk-serial {display: flex;flex-direction: column;align-items: center;background-color: white;padding-right: 0.2em;padding-left: 0.2em;position: relative;}#sk-container-id-7 div.sk-item {position: relative;z-index: 1;}#sk-container-id-7 div.sk-parallel {display: flex;align-items: stretch;justify-content: center;background-color: white;position: relative;}#sk-container-id-7 div.sk-item::before, #sk-container-id-7 div.sk-parallel-item::before {content: \"\";position: absolute;border-left: 1px solid gray;box-sizing: border-box;top: 0;bottom: 0;left: 50%;z-index: -1;}#sk-container-id-7 div.sk-parallel-item {display: flex;flex-direction: column;z-index: 1;position: relative;background-color: white;}#sk-container-id-7 div.sk-parallel-item:first-child::after {align-self: flex-end;width: 50%;}#sk-container-id-7 div.sk-parallel-item:last-child::after {align-self: flex-start;width: 50%;}#sk-container-id-7 div.sk-parallel-item:only-child::after {width: 0;}#sk-container-id-7 div.sk-dashed-wrapped {border: 1px dashed gray;margin: 0 0.4em 0.5em 0.4em;box-sizing: border-box;padding-bottom: 0.4em;background-color: white;}#sk-container-id-7 div.sk-label label {font-family: monospace;font-weight: bold;display: inline-block;line-height: 1.2em;}#sk-container-id-7 div.sk-label-container {text-align: center;}#sk-container-id-7 div.sk-container {/\* jupyter's `normalize.less` sets `[hidden] { display: none; }` but bootstrap.min.css set `[hidden] { display: none !important; }` so we also need the `!important` here to be able to override the default hidden behavior on the sphinx rendered scikit-learn.org. See: https://github.com/scikit-learn/scikit-learn/issues/21755 \*/display: inline-block !important;position: relative;}#sk-container-id-7 div.sk-text-repr-fallback {display: none;}</style><div id=\"sk-container-id-7\" class=\"sk-top-container\"><div class=\"sk-text-repr-fallback\"><pre>GridSearchCV(cv=5, estimator=SVC(), n\_jobs=-1,\n",

" param\_grid={&#x27;C&#x27;: [0.01, 1, 10, 100, 1000],\n",

" &#x27;gamma&#x27;: [1, 0.1, 0.01, 0.001, 0.0001],\n",

" &#x27;kernel&#x27;: [&#x27;rbf&#x27;, &#x27;linear&#x27;]},\n",

" scoring=&#x27;accuracy&#x27;)</pre><b>In a Jupyter environment, please rerun this cell to show the HTML representation or trust the notebook. <br />On GitHub, the HTML representation is unable to render, please try loading this page with nbviewer.org.</b></div><div class=\"sk-container\" hidden><div class=\"sk-item sk-dashed-wrapped\"><div class=\"sk-label-container\"><div class=\"sk-label sk-toggleable\"><input class=\"sk-toggleable\_\_control sk-hidden--visually\" id=\"sk-estimator-id-11\" type=\"checkbox\" ><label for=\"sk-estimator-id-11\" class=\"sk-toggleable\_\_label sk-toggleable\_\_label-arrow\">GridSearchCV</label><div class=\"sk-toggleable\_\_content\"><pre>GridSearchCV(cv=5, estimator=SVC(), n\_jobs=-1,\n",

" param\_grid={&#x27;C&#x27;: [0.01, 1, 10, 100, 1000],\n",

" &#x27;gamma&#x27;: [1, 0.1, 0.01, 0.001, 0.0001],\n",

" &#x27;kernel&#x27;: [&#x27;rbf&#x27;, &#x27;linear&#x27;]},\n",

" scoring=&#x27;accuracy&#x27;)</pre></div></div></div><div class=\"sk-parallel\"><div class=\"sk-parallel-item\"><div class=\"sk-item\"><div class=\"sk-label-container\"><div class=\"sk-label sk-toggleable\"><input class=\"sk-toggleable\_\_control sk-hidden--visually\" id=\"sk-estimator-id-12\" type=\"checkbox\" ><label for=\"sk-estimator-id-12\" class=\"sk-toggleable\_\_label sk-toggleable\_\_label-arrow\">estimator: SVC</label><div class=\"sk-toggleable\_\_content\"><pre>SVC()</pre></div></div></div><div class=\"sk-serial\"><div class=\"sk-item\"><div class=\"sk-estimator sk-toggleable\"><input class=\"sk-toggleable\_\_control sk-hidden--visually\" id=\"sk-estimator-id-13\" type=\"checkbox\" ><label for=\"sk-estimator-id-13\" class=\"sk-toggleable\_\_label sk-toggleable\_\_label-arrow\">SVC</label><div class=\"sk-toggleable\_\_content\"><pre>SVC()</pre></div></div></div></div></div></div></div></div></div></div>"

]

},

"metadata": {},

"execution\_count": 100

}

],

"source": [

"grid\_svc.fit(x\_bal,y\_bal)"

]

},

{

"cell\_type": "code",

"execution\_count": null,

"metadata": {

"colab": {

"base\_uri": "https://localhost:8080/"

},

"id": "Al7khj6DdpAF",

"outputId": "21bbe9ef-0555-41cc-f4f2-ca79f0842c04"

},

"outputs": [

{

"output\_type": "execute\_result",

"data": {

"text/plain": [

"{'C': 1000, 'gamma': 0.1, 'kernel': 'rbf'}"

]

},

"metadata": {},

"execution\_count": 101

}

],

"source": [

"grid\_svc.best\_params\_"

]

},

{

"cell\_type": "code",

"execution\_count": null,

"metadata": {

"id": "AHwZrz-cmsZ9"

},

"outputs": [],

"source": [

"sv1=SVC(C =1000,gamma =1, kernel= 'rbf')"

]

},

{

"cell\_type": "code",

"execution\_count": null,

"metadata": {

"colab": {

"base\_uri": "https://localhost:8080/",

"height": 129

},

"id": "GkYzErIonD20",

"outputId": "8314a0e2-e223-46e3-de9f-f28b45b798d8"

},

"outputs": [

{

"output\_type": "stream",

"name": "stderr",

"text": [

"/usr/local/lib/python3.9/dist-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().\n",

" y = column\_or\_1d(y, warn=True)\n"

]

},

{

"output\_type": "execute\_result",

"data": {

"text/plain": [

"SVC(C=1000, gamma=1)"

],

"text/html": [

"<style>#sk-container-id-8 {color: black;background-color: white;}#sk-container-id-8 pre{padding: 0;}#sk-container-id-8 div.sk-toggleable {background-color: white;}#sk-container-id-8 label.sk-toggleable\_\_label {cursor: pointer;display: block;width: 100%;margin-bottom: 0;padding: 0.3em;box-sizing: border-box;text-align: center;}#sk-container-id-8 label.sk-toggleable\_\_label-arrow:before {content: \"▸\";float: left;margin-right: 0.25em;color: #696969;}#sk-container-id-8 label.sk-toggleable\_\_label-arrow:hover:before {color: black;}#sk-container-id-8 div.sk-estimator:hover label.sk-toggleable\_\_label-arrow:before {color: black;}#sk-container-id-8 div.sk-toggleable\_\_content {max-height: 0;max-width: 0;overflow: hidden;text-align: left;background-color: #f0f8ff;}#sk-container-id-8 div.sk-toggleable\_\_content pre {margin: 0.2em;color: black;border-radius: 0.25em;background-color: #f0f8ff;}#sk-container-id-8 input.sk-toggleable\_\_control:checked~div.sk-toggleable\_\_content {max-height: 200px;max-width: 100%;overflow: auto;}#sk-container-id-8 input.sk-toggleable\_\_control:checked~label.sk-toggleable\_\_label-arrow:before {content: \"▾\";}#sk-container-id-8 div.sk-estimator input.sk-toggleable\_\_control:checked~label.sk-toggleable\_\_label {background-color: #d4ebff;}#sk-container-id-8 div.sk-label input.sk-toggleable\_\_control:checked~label.sk-toggleable\_\_label {background-color: #d4ebff;}#sk-container-id-8 input.sk-hidden--visually {border: 0;clip: rect(1px 1px 1px 1px);clip: rect(1px, 1px, 1px, 1px);height: 1px;margin: -1px;overflow: hidden;padding: 0;position: absolute;width: 1px;}#sk-container-id-8 div.sk-estimator {font-family: monospace;background-color: #f0f8ff;border: 1px dotted black;border-radius: 0.25em;box-sizing: border-box;margin-bottom: 0.5em;}#sk-container-id-8 div.sk-estimator:hover {background-color: #d4ebff;}#sk-container-id-8 div.sk-parallel-item::after {content: \"\";width: 100%;border-bottom: 1px solid gray;flex-grow: 1;}#sk-container-id-8 div.sk-label:hover label.sk-toggleable\_\_label {background-color: #d4ebff;}#sk-container-id-8 div.sk-serial::before {content: \"\";position: absolute;border-left: 1px solid gray;box-sizing: border-box;top: 0;bottom: 0;left: 50%;z-index: 0;}#sk-container-id-8 div.sk-serial {display: flex;flex-direction: column;align-items: center;background-color: white;padding-right: 0.2em;padding-left: 0.2em;position: relative;}#sk-container-id-8 div.sk-item {position: relative;z-index: 1;}#sk-container-id-8 div.sk-parallel {display: flex;align-items: stretch;justify-content: center;background-color: white;position: relative;}#sk-container-id-8 div.sk-item::before, #sk-container-id-8 div.sk-parallel-item::before {content: \"\";position: absolute;border-left: 1px solid gray;box-sizing: border-box;top: 0;bottom: 0;left: 50%;z-index: -1;}#sk-container-id-8 div.sk-parallel-item {display: flex;flex-direction: column;z-index: 1;position: relative;background-color: white;}#sk-container-id-8 div.sk-parallel-item:first-child::after {align-self: flex-end;width: 50%;}#sk-container-id-8 div.sk-parallel-item:last-child::after {align-self: flex-start;width: 50%;}#sk-container-id-8 div.sk-parallel-item:only-child::after {width: 0;}#sk-container-id-8 div.sk-dashed-wrapped {border: 1px dashed gray;margin: 0 0.4em 0.5em 0.4em;box-sizing: border-box;padding-bottom: 0.4em;background-color: white;}#sk-container-id-8 div.sk-label label {font-family: monospace;font-weight: bold;display: inline-block;line-height: 1.2em;}#sk-container-id-8 div.sk-label-container {text-align: center;}#sk-container-id-8 div.sk-container {/\* jupyter's `normalize.less` sets `[hidden] { display: none; }` but bootstrap.min.css set `[hidden] { display: none !important; }` so we also need the `!important` here to be able to override the default hidden behavior on the sphinx rendered scikit-learn.org. See: https://github.com/scikit-learn/scikit-learn/issues/21755 \*/display: inline-block !important;position: relative;}#sk-container-id-8 div.sk-text-repr-fallback {display: none;}</style><div id=\"sk-container-id-8\" class=\"sk-top-container\"><div class=\"sk-text-repr-fallback\"><pre>SVC(C=1000, gamma=1)</pre><b>In a Jupyter environment, please rerun this cell to show the HTML representation or trust the notebook. <br />On GitHub, the HTML representation is unable to render, please try loading this page with nbviewer.org.</b></div><div class=\"sk-container\" hidden><div class=\"sk-item\"><div class=\"sk-estimator sk-toggleable\"><input class=\"sk-toggleable\_\_control sk-hidden--visually\" id=\"sk-estimator-id-14\" type=\"checkbox\" checked><label for=\"sk-estimator-id-14\" class=\"sk-toggleable\_\_label sk-toggleable\_\_label-arrow\">SVC</label><div class=\"sk-toggleable\_\_content\"><pre>SVC(C=1000, gamma=1)</pre></div></div></div></div></div>"

]

},

"metadata": {},

"execution\_count": 103

}

],

"source": [

"sv1.fit(x\_bal,y\_bal)"

]

},

{

"cell\_type": "code",

"execution\_count": null,

"metadata": {

"id": "MbuLe4lknVVM"

},

"outputs": [],

"source": [

"y\_pred= sv1.predict(x\_test\_bal)"

]

},

{

"cell\_type": "code",

"execution\_count": null,

"metadata": {

"colab": {

"base\_uri": "https://localhost:8080/"

},

"id": "lVzQrN0An0Jk",

"outputId": "f859f461-8620-416e-8a47-3c6e0c89ec7d"

},

"outputs": [

{

"output\_type": "stream",

"name": "stdout",

"text": [

" precision recall f1-score support\n",

"\n",

" 0 0.96 0.60 0.73 121\n",

" 1 0.85 0.78 0.81 121\n",

" 2 0.85 0.88 0.87 121\n",

" 3 0.76 0.69 0.72 121\n",

" 4 0.47 0.87 0.61 121\n",

" 5 0.85 0.65 0.74 121\n",

" 6 0.92 0.82 0.86 121\n",

"\n",

" accuracy 0.76 847\n",

" macro avg 0.81 0.76 0.76 847\n",

"weighted avg 0.81 0.76 0.76 847\n",

"\n"

]

}

],

"source": [

"print(classification\_report(y\_test\_bal,y\_pred))"

]

},

{

"cell\_type": "code",

"execution\_count": null,

"metadata": {

"colab": {

"base\_uri": "https://localhost:8080/"

},

"id": "iXqhn\_rmoAua",

"outputId": "849624ea-0821-4abe-a921-db4def14e70d"

},

"outputs": [

{

"output\_type": "execute\_result",

"data": {

"text/plain": [

"0.9996973365617433"

]

},

"metadata": {},

"execution\_count": 106

}

],

"source": [

"train\_score= accuracy\_score(y\_bal,sv1.predict(x\_bal))\n",

"train\_score"

]

},

{

"cell\_type": "markdown",

"source": [

"Saving the model as thyroid1\_model.pkl"

],

"metadata": {

"id": "XYTY73c5DTkc"

}

},

{

"cell\_type": "code",

"execution\_count": null,

"metadata": {

"id": "Tz7v0YmzoRaX"

},

"outputs": [],

"source": [

"import pickle\n",

"pickle.dump(sv1,open('thyroid\_1\_model.pkl','wb'))"

]

},

{

"cell\_type": "code",

"execution\_count": null,

"metadata": {

"id": "ObdBgVx2pInn"

},

"outputs": [],

"source": [

"features = np.array([[0,0,0,0,0.000000,0.0,0.0,1.00,0.0,40.0]])\n",

"#print(lable\_encoder.inverse\_transform(xgb1.predict(features)))"

]

},

{

"cell\_type": "code",

"execution\_count": null,

"metadata": {

"colab": {

"base\_uri": "https://localhost:8080/"

},

"id": "Fn0YoNykp2pf",

"outputId": "2a39c85f-37cc-466e-ba6f-758a671773ac"

},

"outputs": [

{

"output\_type": "execute\_result",

"data": {

"text/plain": [

"numpy.ndarray"

]

},

"metadata": {},

"execution\_count": 139

}

],

"source": [

"type(features)"

]

},

{

"cell\_type": "code",

"execution\_count": null,

"metadata": {

"id": "xowBJvpzqCn4"

},

"outputs": [],

"source": [

"pickle.dump(lable\_encoder,open('lable\_encoder.pk1','wb'))"

]

},

{

"cell\_type": "code",

"execution\_count": null,

"metadata": {

"colab": {

"base\_uri": "https://localhost:8080/"

},

"id": "IPbhPPlGqiub",

"outputId": "9fcb0aae-ac85-4b97-f5ae-93cf8de8ce17"

},

"outputs": [

{

"output\_type": "execute\_result",

"data": {

"text/plain": [

"array(['miscellaneous', 'hypothyroid conditions', 'binding protein',\n",

" 'replacement therapy', 'general health', 'hyperthyroid conditions',\n",

" 'antithyroid treatment'], dtype=object)"

]

},

"metadata": {},

"execution\_count": 141

}

],

"source": [

"data['target'].unique()"

]

},

{

"cell\_type": "code",

"execution\_count": null,

"metadata": {

"colab": {

"base\_uri": "https://localhost:8080/"

},

"id": "UvzXhj\_LqtO2",

"outputId": "5640e4dd-344f-4460-a4a8-abf9aa58e499"

},

"outputs": [

{

"output\_type": "execute\_result",

"data": {

"text/plain": [

"array([5, 4, 1, 6, 2, 3, 0])"

]

},

"metadata": {},

"execution\_count": 142

}

],

"source": [

"y['target'].unique()"

]

}

],

"metadata": {

"colab": {

"provenance": []

},

"kernelspec": {

"display\_name": "Python 3",

"name": "python3"

},

"language\_info": {

"name": "python"

}

},

"nbformat": 4,

"nbformat\_minor": 0

}

## app.py

# -\*- coding: utf-8 -\*-  
*"""  
Created on Mon Jun 27 11:56:38 2022  
  
@author: SmartBridge-PC  
"""*from flask import Flask, render\_template, request  
import numpy as np  
import pickle  
import pandas as pd  
  
model = pickle.load(open(r"C:\Users\ELCOT\pycharmprojects\thyroid disease\thyroid\_1\_model (1).pkl", 'rb'))  
  
  
app = Flask(\_\_name\_\_)  
  
@app.route("/")  
def about():  
 return render\_template('submit.html')  
  
@app.route("/predict", methods=['POST', 'GET'])  
def predict():  
 x = [[float(x) for x in request.form.values()]]  
  
 print(x)  
 col = ['goitre','tumor','hypo','psych','TSH','T3','TT4','T4U','FTI','TBG']  
 x = pd.DataFrame(x, columns=col)  
  
  
 #print(x.shape)  
  
 print(x)  
 pred = model.predict(x)  
 #pred = le.inverse\_transform(pred)  
 print(pred[0])  
 return render\_template('submit.html', prediction\_text=str(pred[0]))  
if \_\_name\_\_ == '\_\_main\_\_':  
 app.run(debug=False)

Submit.html:

<!DOCUMENT html>  
<htmnl lang="en">  
 <head>  
 <meta charset="UTF-8">  
 <title>thyroid disease</title>  
 </head>  
 <body bgcolor="blue">  
 <h1>THYROID DISEASE CLASSIFICATION USING MACHINE LEARNING</h1>  
 <from action = "{{url\_for('predict')}}" method = "post">  
 GOITRE<input type="number" name="goitre" id="goitre">  
 TUMOR<input type="number" name="tumor" id="tumor">  
 PSYCH<input type="number" name="psych" id="psych">  
 TSH<input type="number" name="TSH" id="TSH">  
 T3<input type="number" name="T3" id="T3">  
 TT4<input type="number" name="TT4" id="TT4">  
 T4U<input type="number" name="T4U" id="T4U">  
 FTI<input type="number" name="FTI" id="FTI">  
 TBG<input type="number" name="TBG" id="TBG">  
<button type="submit">Predict</button>  
 </from>  
 <h1>{{ y }}</h1>  
 </body>  
</htmnl>