





BREAST CANCER CLASSIFICATION

RANDOM FOREST CLASSIFIER & TREE DECISION

by Rofiqo Azzahra





LIST OF CONTENTS

O1
ABOUT ME

03
DESCRIPTION

DATA PROCESS & VISUALIZATION

O2
TOOLS USED

G4 FLOWCHART

CONTACT

Introducing ABOUT ME

Hello, everyone! I am Rofiqo Azzahra, a Bachelor of Information Systems with a 3.66 GPA. I'm a Project/Document Administrator with over three years of experience in the telecommunications industry.

I am highly interested in Data Science because of its ability to analyze complex data and drive impactful decisions. I'm excited to learn more and apply my skills in this field.



Rofiqo Azzahra

https://www.linkedin.com/in/rofiqo-azzahra

TOOLS USED





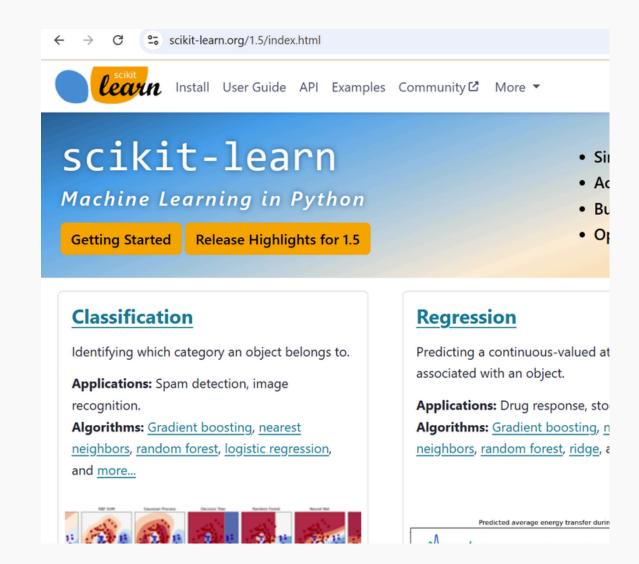






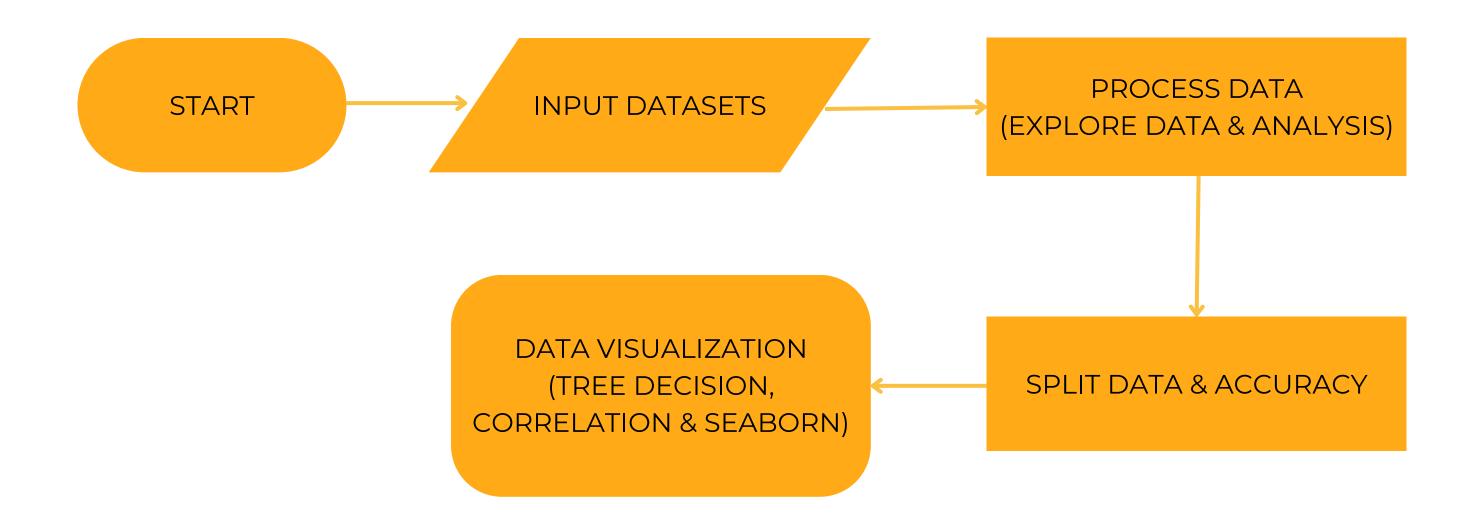
DESCRIPTION

- **Breast cancer** is a disease that occurs when cells in the breast grow out of control and form tumors. The tumors can spread to other parts of the body, which can be fatal.
- **Scikit-learn** is a free, open-source Python library that helps implement machine learning models. It's a popular choice for beginners because of its extensive documentation and ability to work well in production.
- Pandas (styled as pandas) is a <u>software library</u> written for the <u>Python</u> <u>programming language</u> for data manipulation and <u>analysis</u>.
- **Google Colab** is a hosted Jupyter Notebook service that requires no setup to use and provides free of charge access to computing resources, including GPUs and TPUs. Colab is especially well suited to machine learning, data science, and education.



https://scikit-learn.org/1.5/index.html

FLOWCHART



https://www.linkedin.com/in/rofiqo-azzahra

DATA PROCESS

import libraries and load datasets

```
#import data from sklearn
from sklearn import datasets
import pandas as pd
#load datasets and convert into a Dataframe
cancer = datasets.load_breast_cancer()
                    #input for machine learning
x = cancer.data
y = cancer.target
                    #ouput of machine learning
#convert feature and target data into Dataframe
df_x = pd.DataFrame(x, columns=cancer.feature_names)
df y = pd.Series(y, name='target')
#combine features and targets in one Dataframes
df = pd.concat([df_x, df_y], axis = 1)
df.head(10) #show 10 line of Dataframe
```

	mean radius	mean texture	mean perimeter	mean area	mean smoothness	mean compactness	mean concavity	mean concave points	mean symmetry	mean fractal dimension	
0	17.99	10.38	122.80	1001.0	0.11840	0.27760	0.30010	0.14710	0.2419	0.07871	
1	20.57	17.77	132.90	1326.0	0.08474	0.07864	0.08690	0.07017	0.1812	0.05667	
2	19.69	21.25	130.00	1203.0	0.10960	0.15990	0.19740	0.12790	0.2069	0.05999	
3	11.42	20.38	77.58	386.1	0.14250	0.28390	0.24140	0.10520	0.2597	0.09744	
4	20.29	14.34	135.10	1297.0	0.10030	0.13280	0.19800	0.10430	0.1809	0.05883	
5	12.45	15.70	82.57	477.1	0.12780	0.17000	0.15780	0.08089	0.2087	0.07613	
6	18.25	19.98	119.60	1040.0	0.09463	0.10900	0.11270	0.07400	0.1794	0.05742	
7	13.71	20.83	90.20	577.9	0.11890	0.16450	0.09366	0.05985	0.2196	0.07451	
8	13.00	21.82	87.50	519.8	0.12730	0.19320	0.18590	0.09353	0.2350	0.07389	
9	12.46	24.04	83.97	475.9	0.11860	0.23960	0.22730	0.08543	0.2030	0.08243	

overview

DATA PROCESS

#Show basic information of data
df.info()

```
(class 'pandas.core.frame.DataFrame'>
                                                                                                   float64
                                                                                     569 non-null
                                                          13 area error
langeIndex: 569 entries, 0 to 568
                                                                                                   float64
                                                          14 smoothness error
                                                                                     569 non-null
                                                                                     569 non-null
                                                                                                 float64
                                                          15 compactness error
Data columns (total 31 columns):
                                                                                                 float64
                                                          16 concavity error
                                                                                     569 non-null
    Column
                              Non-Null Count Dtype
                                                          17 concave points error
                                                                                     569 non-null
                                                                                                 float64
                                                                                                 float64
                                                          18 symmetry error
                                                                                     569 non-null
    mean radius
                                              float64
                              569 non-null
                                                          19 fractal dimension error 569 non-null
                                                                                                 float64
    mean texture
                              569 non-null
                                            float64
                                                                                                 float64
                                                             worst radius
                                                                                     569 non-null
                                            float64
                              569 non-null
    mean perimeter
                                                                                     569 non-null
                                                                                                 float64
                                                          21 worst texture
                              569 non-null
                                              float64
    mean area
                                                                                                 float64
                                                             worst perimeter
                                                                                     569 non-null
    mean smoothness
                              569 non-null
                                              float64
                                                                                                   float64
                                                                                     569 non-null
                                                             worst area
                                                                                                   float64
                                                             worst smoothness
                                                                                     569 non-null
                              569 non-null
                                              float64
    mean compactness
                                                                                                   float64
                                                             worst compactness
                                                                                     569 non-null
                              569 non-null
                                              float64
    mean concavity
                                                          26 worst concavity
                                                                                                 float64
                                                                                     569 non-null
    mean concave points
                              569 non-null
                                              float64
                                                          27 worst concave points
                                                                                                 float64
                                                                                     569 non-null
                              569 non-null
                                              float64
    mean symmetry
                                                          28 worst symmetry
                                                                                     569 non-null
                                                                                                 float64
    mean fractal dimension
                              569 non-null
                                              float64
                                                             worst fractal dimension 569 non-null
                                                                                                 float64
    radius error
                              569 non-null
                                              float64
                                                          30 target
                                                                                     569 non-null
                                                                                                   int64
    texture error
                              569 non-null
                                              float64
                                                         dtypes: float64(30), int64(1)
                              569 non-null
                                              float64
    perimeter error
                                                         memory usage: 137.9 KB
```

Explore Data and Analysis (EDA)

EXPLORE DATA AND ANALYSIS (EDA)

#View a stastistical description of the data
df.describe()

	mean radius	mean texture	mean perimeter	mean area	mean smoothness	mean compactness	mean concavity
count	569.000000	569.000000	569.000000	569.000000	569.000000	569.000000	569.000000
mean	14.127292	19.289649	91.969033	654.889104	0.096360	0.104341	0.088799
std	3.524049	4.301036	24.298981	351.914129	0.014064	0.052813	0.079720
min	6.981000	9.710000	43.790000	143.500000	0.052630	0.019380	0.000000
25%	11.700000	16.170000	75.170000	420.300000	0.086370	0.064920	0.029560
50%	13.370000	18.840000	86.240000	551.100000	0.095870	0.092630	0.061540
75%	15.780000	21.800000	104.100000	782.700000	0.105300	0.130400	0.130700
max	28.110000	39.280000	188.500000	2501.000000	0.163400	0.345400	0.426800



Split Data

```
from sklearn.model_selection import train_test_split

#Split data
x_train, x_test, y_train, y_test = train_test_split(df_x, df_y, test_size = 0.2, random_state = 42)
```

Create & Train

```
from sklearn.ensemble import RandomForestClassifier

RandomForestClassifier (random_state=42)

#Create and train randomly

model = RandomForestClassifier(n_estimators=100, random_state=42)

model.fit(x_train, y_train)
```

DATA MODELLING

```
#Make predictions
y_pred = model.predict(x_test)
accuracy = accuracy_score(y_test, y_pred)

#Evaluate the model
print("Classification Report: ")
print(classification_report(y_test, y_pred))
print(f"Accuracy: {accuracy * 100:.2f}%")
Classification_report
```

Accuracy

Classification Report:											
	precision	recall	f1-score	support							
0	0.98	0.93	0.95	43							
1	0.96	0.99	0.97	71							
accuracy			0.96	114							
macro avg	0.97	0.96	0.96	114							
weighted avg	0.97	0.96	0.96	114							
Accuracy: 96.	49%										

DATA VISUALIZATION

```
import matplotlib.pyplot as plt
import seaborn as sns

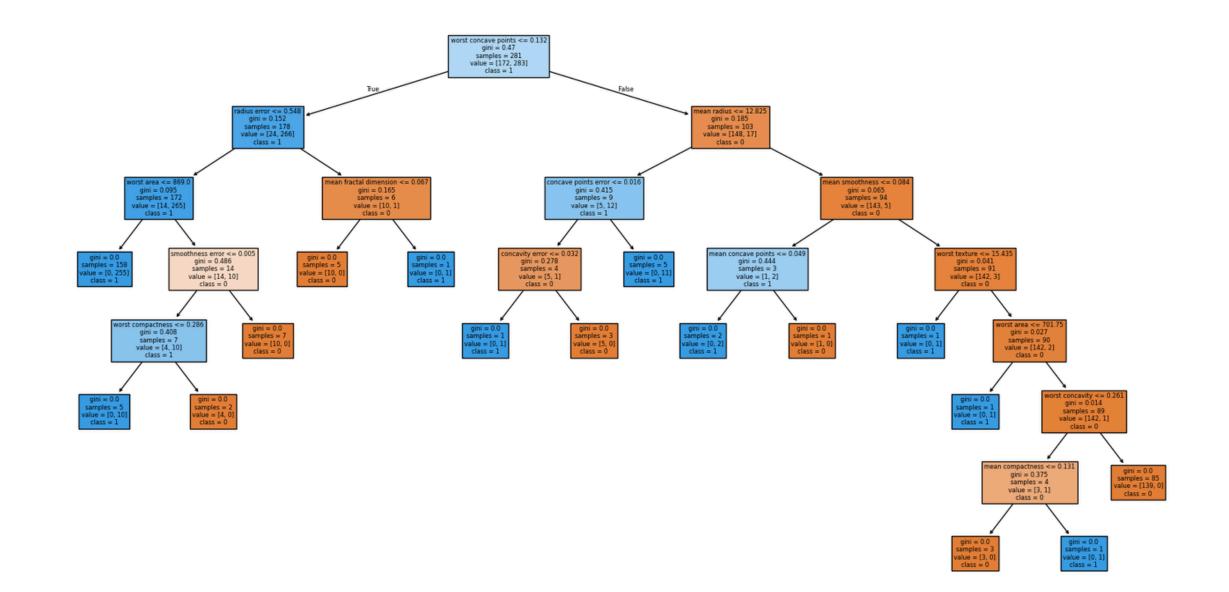
#
sns.countplot(x='target', data=df)
plt.title('Count of Target Classes')
plt.xlabel('Target Class')
plt.ylabel('Count')
plt.show()
```



DATA VISUALIZATION

```
import matplotlib.pyplot as plt
from sklearn import tree

plt.figure(figsize=(20, 10))
tree.plot_tree(model.estimators_[0], feature_names=df_x.columns, class_names=['0', '1'], filled=True)
plt.show()
```



Decision Tree

DATA VISUALIZATION

```
corr = df.corr()
plt.figure(figsize=(21, 16))
sns.heatmap(corr, annot=True, cmap='coolwarm')
plt.title('Correlation Heatmap')
plt.show()
```

Correlation

		Correlation Heatmap																													
mean radius -	1	0.32	1	0.99	0.17	0.51	0.68	0.82	0.15	-0.31	0.68	0.097	0.67	0.74	-0.22	0.21	0.19	0.38	-0.1	-0.043	0.97	0.3	0.97	0.94	0.12	0.41	0.53	0.74	0.16 0	.0071	-0.73
mean texture -	0.32	1	0.33	0.32	-0.023	0.24	0.3	0.29	0.071	-0.076	0.28	0.39	0.28	0.26	0.0066	0.19	0.14	0.16	0.0091	0.054	0.35	0.91	0.36	0.34	0.078	0.28	0.3	0.3	0.11	0.12	-0.42
mean perimeter -	1	0.33	1	0.99	0.21	0.56	0.72	0.85	0.18	-0.26	0.69	0.087	0.69	0.74	-0.2	0.25	0.23	0.41	-0.082	0.0055	0.97	0.3	0.97	0.94	0.15	0.46	0.56	0.77	0.19	0.051	-0.74
mean area -	0.99	0.32	0.99	1	0.18	0.5	0.69	0.82	0.15	-0.28	0.73	0.066	0.73	0.8	-0.17	0.21	0.21	0.37	-0.072	-0.02	0.96	0.29	0.96	0.96	0.12	0.39	0.51	0.72	0.14 0	.0037	-0.71
mean smoothness -	0.17	-0.023	0.21	0.18	1	0.66	0.52	0.55	0.56		0.3	0.068	0.3	0.25	0.33	0.32	0.25	0.38	0.2	0.28	0.21	0.036	0.24	0.21	0.81	0.47	0.43	0.5	0.39	0.5	-0.36
mean compactness -	0.51	0.24	0.56	0.5	0.66	1	0.88	0.83	0.6	0.57	0.5	0.046	0.55	0.46	0.14	0.74	0.57	0.64	0.23	0.51	0.54	0.25	0.59	0.51	0.57	0.87	0.82	0.82	0.51	0.69	-0.6
mean concavity -	0.68	0.3	0.72	0.69	0.52	0.88	1	0.92	0.5	0.34	0.63	0.076	0.66	0.62	0.099	0.67	0.69	0.68	0.18	0.45	0.69	0.3	0.73	0.68	0.45	0.75	0.88	0.86	0.41	0.51	-0.7
mean concave points -	0.82	0.29	0.85	0.82	0.55	0.83	0.92	1	0.46	0.17		0.021	0.71	0.69	0.028	0.49	0.44	0.62	0.095	0.26	0.83	0.29	0.86	0.81	0.45	0.67	0.75	0.91	0.38	0.37	-0.78
mean symmetry -	0.15	0.071	0.18	0.15	0.56	0.6	0.5	0.46	1	0.48	0.3	0.13	0.31	0.22	0.19	0.42	0.34	0.39	0.45	0.33	0.19	0.091	0.22	0.18	0.43	0.47	0.43	0.43	0.7	0.44	-0.33
mean fractal dimension -	-0.31	-0.076	-0.26	-0.28		0.57	0.34	0.17	0.48	1 0	.0001	10.16	0.04	-0.09	0.4	0.56	0.45	0.34	0.35	0.69	-0.25	0.051	-0.21	-0.23	0.5	0.46	0.35	0.18	0.33	0.77	0.013
radius error -	0.68	0.28	0.69	0.73	0.3	0.5	0.63	0.7	0.3 0	.0001	1 1	0.21	0.97	0.95	0.16	0.36	0.33	0.51	0.24	0.23	0.72	0.19	0.72	0.75	0.14	0.29	0.38	0.53	0.095	0.05	-0.57
texture error -	-0.097	0.39	-0.087-	-0.066	0.068	0.046	0.076	0.021	0.13	0.16	0.21	1	0.22	0.11	0.4	0.23	0.19	0.23	0.41	0.28	-0.11	0.41	-0.1 -	0.083	0.074	-0.092	0.069	-0.12	-0.13 -	0.0460	0.0083
perimeter error -	0.67	0.28	0.69	0.73	0.3	0.55	0.66	0.71	0.31	0.04	0.97	0.22	1	0.94	0.15	0.42	0.36	0.56	0.27	0.24		0.2	0.72	0.73	0.13	0.34	0.42	0.55	0.11	0.085	-0.56
area error -	0.74	0.26	0.74	0.8	0.25	0.46	0.62	0.69	0.22	-0.09	0.95	0.11	0.94	1	0.075	0.28	0.27	0.42	0.13	0.13	0.76	0.2	0.76	0.81	0.13	0.28	0.39	0.54	0.074	0.018	-0.55
smoothness error -	-0.22	0.0066	-0.2	-0.17	0.33	0.14	0.099	0.028	0.19	0.4	0.16	0.4	0.15	0.075	1	0.34	0.27	0.33	0.41	0.43	-0.23	0.075	-0.22	-0.18	0.31	-0.056-	0.058	-0.1	0.11	0.1	0.067
compactness error -	0.21	0.19	0.25	0.21	0.32	0.74	0.67	0.49	0.42	0.56	0.36	0.23	0.42	0.28	0.34	1	0.8	0.74	0.39	0.8	0.2	0.14	0.26	0.2	0.23	0.68	0.64	0.48	0.28		-0.29
concavity error -	0.19	0.14	0.23	0.21	0.25	0.57	0.69	0.44	0.34	0.45	0.33	0.19	0.36	0.27	0.27	0.8	1	0.77	0.31	0.73	0.19	0.1	0.23	0.19	0.17	0.48	0.66	0.44	0.2	0.44	-0.25
concave points error -	0.38	0.16	0.41	0.37	0.38	0.64	0.68	0.62	0.39	0.34	0.51	0.23	0.56	0.42	0.33	0.74	0.77	1	0.31	0.61	0.36	0.087	0.39	0.34	0.22	0.45	0.55	0.6	0.14	0.31	-0.41
symmetry error -	-0.1	0.0091	0.082	-0.072	0.2	0.23	0.18	0.095	0.45	0.35	0.24	0.41	0.27	0.13	0.41	0.39	0.31	0.31	1	0.37	-0.13	-0.077	-0.1	-0.11	0.013	0.06	0.037	-0.03	0.39	0.0780	0.0065
fractal dimension error -	-0.043	0.054	0.0055	5-0.02	0.28	0.51	0.45	0.26	0.33	0.69	0.23	0.28	0.24	0.13	0.43	0.8	0.73		0.37	1	-0.037	0.003	20.001-	0.023	0.17	0.39	0.38	0.22	0.11	0.59	0.078
worst radius -	0.97	0.35	0.97	0.96	0.21	0.54	0.69	0.83	0.19	-0.25	0.72	-0.11	0.7	0.76	-0.23	0.2	0.19	0.36	-0.13	-0.037	1	0.36	0.99	0.98	0.22	0.48	0.57	0.79	0.24	0.093	-0.78
worst texture -	0.3	0.91	0.3	0.29	0.036	0.25	0.3	0.29	0.091	-0.051	0.19	0.41	0.2	0.2	-0.075	0.14	0.1	0.087	-0.077	0.0032	20.36	1	0.37	0.35	0.23	0.36	0.37	0.36	0.23	0.22	-0.46
worst perimeter -	0.97	0.36	0.97	0.96	0.24		0.73	0.86	0.22	-0.21	0.72	-0.1	0.72	0.76	-0.22	0.26	0.23	0.39	-0.1	-0.001	0.99	0.37	1	0.98	0.24	0.53	0.62	0.82	0.27	0.14	-0.78
worst area -	0.94	0.34	0.94	0.96	0.21	0.51	0.68	0.81	0.18	-0.23	0.75	0.083	0.73	0.81	-0.18	0.2	0.19	0.34	-0.11	-0.023	0.98	0.35	0.98	1	0.21	0.44	0.54	0.75	0.21	0.08	-0.73
worst smoothness -	0.12	0.078	0.15	0.12	0.81	0.57	0.45	0.45	0.43	0.5	0.14	0.074	0.13	0.13	0.31	0.23	0.17	0.22	-0.013	0.17	0.22	0.23	0.24	0.21	1	0.57	0.52	0.55	0.49	0.62	-0.42
worst compactness -	0.41	0.28	0.46	0.39	0.47	0.87	0.75	0.67	0.47	0.46	0.29	0.092	0.34	0.28	-0.056	0.68	0.48	0.45	0.06	0.39	0.48	0.36	0.53	0.44	0.57	1	0.89	0.8	0.61	0.81	-0.59
worst concavity -	0.53	0.3	0.56	0.51	0.43	0.82	0.88	0.75	0.43	0.35	0.38	0.069	0.42	0.39	-0.058	0.64	0.66	0.55	0.037	0.38	0.57	0.37	0.62	0.54	0.52	0.89	1	0.86	0.53	0.69	-0.66
worst concave points -	0.74	0.3	0.77	0.72	0.5	0.82	0.86	0.91	0.43	0.18	0.53	-0.12	0.55	0.54	-0.1	0.48	0.44	0.6	-0.03	0.22	0.79	0.36	0.82	0.75	0.55	0.8	0.86	1	0.5	0.51	-0.79
worst symmetry -	0.16	0.11	0.19	0.14	0.39	0.51	0.41	0.38		0.33	0.095	-0.13	0.11	0.074	-0.11	0.28	0.2	0.14	0.39	0.11	0.24	0.23	0.27	0.21	0.49	0.61	0.53	0.5	1	0.54	-0.42
worst fractal dimension	0.0071	0.12	0.0510	0.0037	0.5	0.69	0.51	0.37	0.44	0.77	0.05	0.046	0.085	0.018	0.1		0.44	0.31	0.078	0.59	0.093	0.22	0.14	0.08	0.62	0.81	0.69	0.51	0.54	1	-0.32
target -	-0.73	-0.42	-0.74	-0.71	-0.36	-0.6	-0.7	-0.78	-0.33	0.013	-0.57	0.0083	-0.56	-0.55	0.067	-0.29	-0.25	-0.41	0.0065	0.078	-0.78	-0.46	-0.78	-0.73	-0.42	-0.59	-0.66	-0.79	0.42	-0.32	1
	mean radius	mean texture	mean perimeter	mean area	mean smoothness	mean compactness	mean concavity	mean concave points	mean symmetry	ean fractal dimension	radius error	texture error	perimeter error	area error	smoothness error	compactness error	concavity error	concave points error	symmetry error	actal dimension error	worst radius	worst texture	worst perimeter	worst area	worst smoothness	worst compactness	worst concavity	worst concave points	worst symmetry	orst fractal dimension	target

0000

THANK YOU

0000