

REPORT ON

BREAST CANCER DETECTION

Breast cancer is a type of cancer which is diagnosed when breast cells mutate and become cancerous cells that multiply and form tumors. Breast cancer typically affects women and people assigned female at birth (AFAB) age 50 and older, but it can also affect men and people assigned male at birth (AMAB), as well as younger women.

Objective:

The main objective of this project is to develop a computer-aided detection (CAD) system that can assist healthcare professionals in accurately identifying and diagnosing breast cancer from mammogram images.

So this model right here provides the accuracy rate of detecting the Breast Cancer when the required datasets are provided.

This Project is coded in python programming where the program is programmed in 5 sectors they are:-

1. Loading the Dataset:-

The given set of data with features are loaded in the form of either Mammogram Images or CSV File

2. Cleaning the loaded data:-

The collected data undergo preprocessing techniques such as resizing, normalization, and noise removal to enhance their quality and suitability for analysis.

3. Conversion of Categorical Data:-

The Interpreters or the Model can read only the Numeric Data, Therefore all the Categorical datas are converted into Numeric Data and Features such as texture, shape, and intensity are extracted from the mammogram images using techniques like gray-level co-occurrence matrix (GLCM), Gabor filters, and wavelet transforms.

4. Fitting the training Data through the relevant Training Model:-

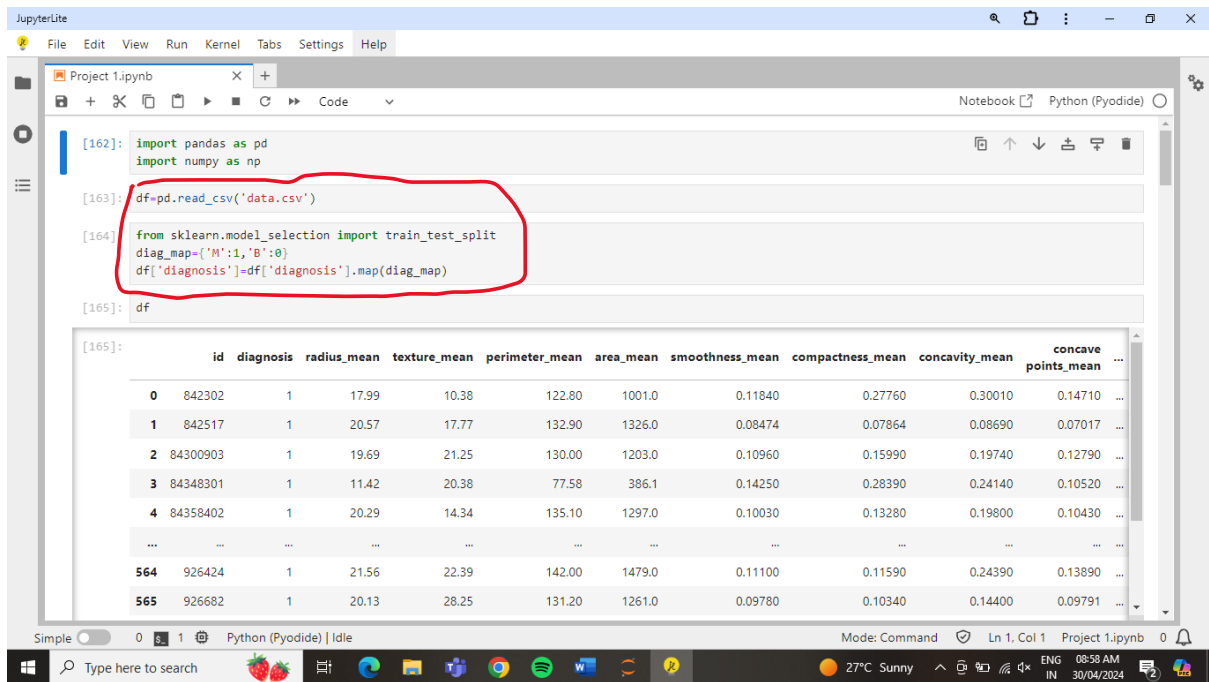
The trained Model is Fitted into the relevant training Models like KNN, NAÏVE BAYES, LOGISTIC REGRESSION and much more

5. Prediction:-

The performance of the developed models is evaluated using metrics like accuracy, sensitivity, specificity, and area under the receiver operating characteristic curve (AUC-ROC) through cross-validation techniques and Once the model achieves satisfactory performance, it is deployed as a user-friendly application or integrated into existing healthcare systems for real-time breast cancer detection.

Some of the screenshots of the Model is displayed below:-

1. Loading the Dataset:-



The screenshot shows a Jupyter Notebook with the following code cells:

```
[162]: import pandas as pd
import numpy as np

[163]: df=pd.read_csv('data.csv')

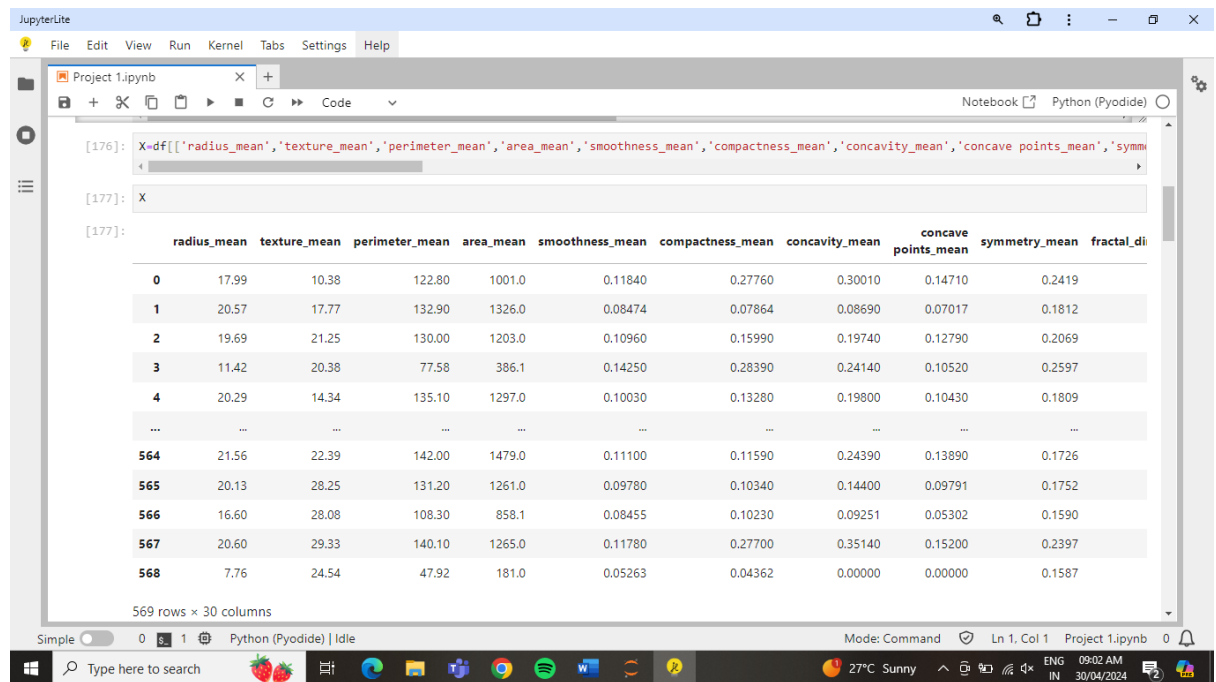
[164]: from sklearn.model_selection import train_test_split
diag_map={'M':1,'B':0}
df['diagnosis']=df['diagnosis'].map(diag_map)

[165]: df
```

The output of cell [165] is a preview of the DataFrame:

	id	diagnosis	radius_mean	texture_mean	perimeter_mean	area_mean	smoothness_mean	compactness_mean	concavity_mean	concave points_mean	...
0	842302	1	17.99	10.38	122.80	1001.0	0.11840	0.27760	0.30010	0.14710	...
1	842517	1	20.57	17.77	132.90	1326.0	0.08474	0.07864	0.08690	0.07017	...
2	84300903	1	19.69	21.25	130.00	1203.0	0.10960	0.15990	0.19740	0.12790	...
3	84348301	1	11.42	20.38	77.58	386.1	0.14250	0.28390	0.24140	0.10520	...
4	84358402	1	20.29	14.34	135.10	1297.0	0.10030	0.13280	0.19800	0.10430	...
...
564	926424	1	21.56	22.39	142.00	1479.0	0.11100	0.11590	0.24390	0.13890	...
565	926682	1	20.13	28.25	131.20	1261.0	0.09780	0.10340	0.14400	0.09791	...

2. Cleaning the loaded data:-



The screenshot shows a Jupyter Notebook with the following code cells:

```
[176]: X=df[['radius_mean','texture_mean','perimeter_mean','area_mean','smoothness_mean','compactness_mean','concavity_mean','concave points_mean','symmetry_mean','fractal_dimension']]

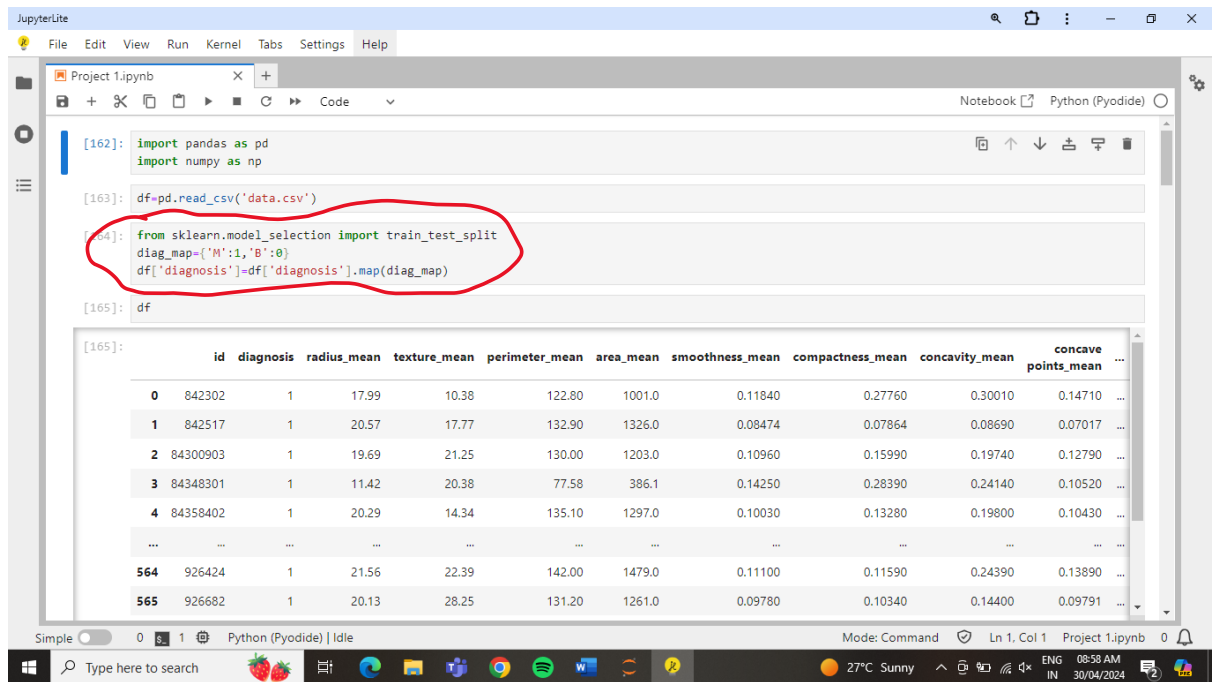
[177]: X
```

The output of cell [177] is a preview of the cleaned DataFrame:

	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	
1	20.57	17.77	132.90	1326.0	0.08474	0.07864	0.08690	0.07017	0.1812	
2	19.69	21.25	130.00	1203.0	0.10960	0.15990	0.19740	0.12790	0.2069	
3	11.42	20.38	77.58	386.1	0.14250	0.28390	0.24140	0.10520	0.2597	
4	20.29	14.34	135.10	1297.0	0.10030	0.13280	0.19800	0.10430	0.1809	
...
564	21.56	22.39	142.00	1479.0	0.11100	0.11590	0.24390	0.13890	0.1726	
565	20.13	28.25	131.20	1261.0	0.09780	0.10340	0.14400	0.09791	0.1752	
566	16.60	28.08	108.30	858.1	0.08455	0.10230	0.09251	0.05302	0.1590	
567	20.60	29.33	140.10	1265.0	0.11780	0.27700	0.35140	0.15200	0.2397	
568	7.76	24.54	47.92	181.0	0.05263	0.04362	0.00000	0.00000	0.1587	

569 rows x 30 columns

3. Conversion of Categorical Data:-



```
[162]: import pandas as pd
import numpy as np

[163]: df=pd.read_csv('data.csv')

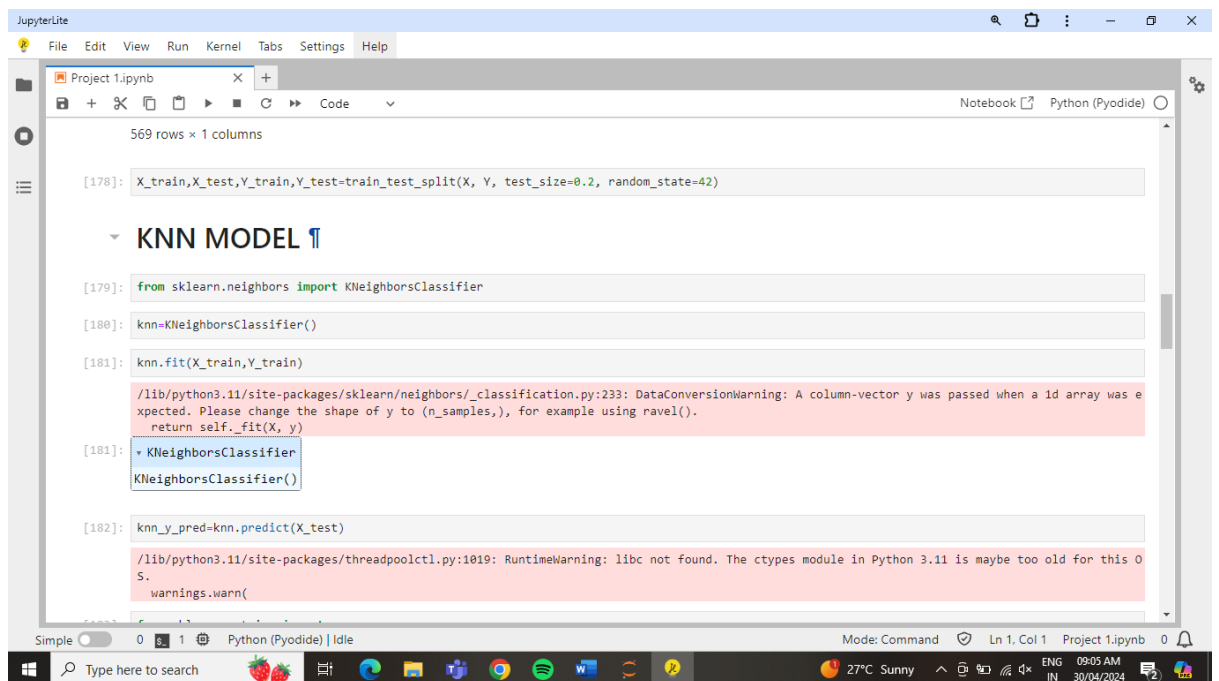
[164]: from sklearn.model_selection import train_test_split
diag_map={'M':1,'B':0}
df['diagnosis']=df['diagnosis'].map(diag_map)

[165]: df
```

	id	diagnosis	radius_mean	texture_mean	perimeter_mean	area_mean	smoothness_mean	compactness_mean	concavity_mean	concave points_mean	...
0	842302	1	17.99	10.38	122.80	1001.0	0.11840	0.27760	0.30010	0.14710	...
1	842517	1	20.57	17.77	132.90	1326.0	0.08474	0.07864	0.08690	0.07017	...
2	84300903	1	19.69	21.25	130.00	1203.0	0.10960	0.15990	0.19740	0.12790	...
3	84348301	1	11.42	20.38	77.58	386.1	0.14250	0.28390	0.24140	0.10520	...
4	84358402	1	20.29	14.34	135.10	1297.0	0.10030	0.13280	0.19800	0.10430	...
...
564	926424	1	21.56	22.39	142.00	1479.0	0.11100	0.11590	0.24390	0.13890	...
565	926682	1	20.13	28.25	131.20	1261.0	0.09780	0.10340	0.14400	0.09791	...

4. Fitting the training Data through the relevant Training Model:-

Modelling is done through KNN prediction as it is efficient in predicting the trained and tested data



```
[178]: X_train,X_test,Y_train,Y_test=train_test_split(X, Y, test_size=0.2, random_state=42)

KNN MODEL ¶

[179]: from sklearn.neighbors import KNeighborsClassifier

[180]: knn=KNeighborsClassifier()

[181]: knn.fit(X_train,Y_train)

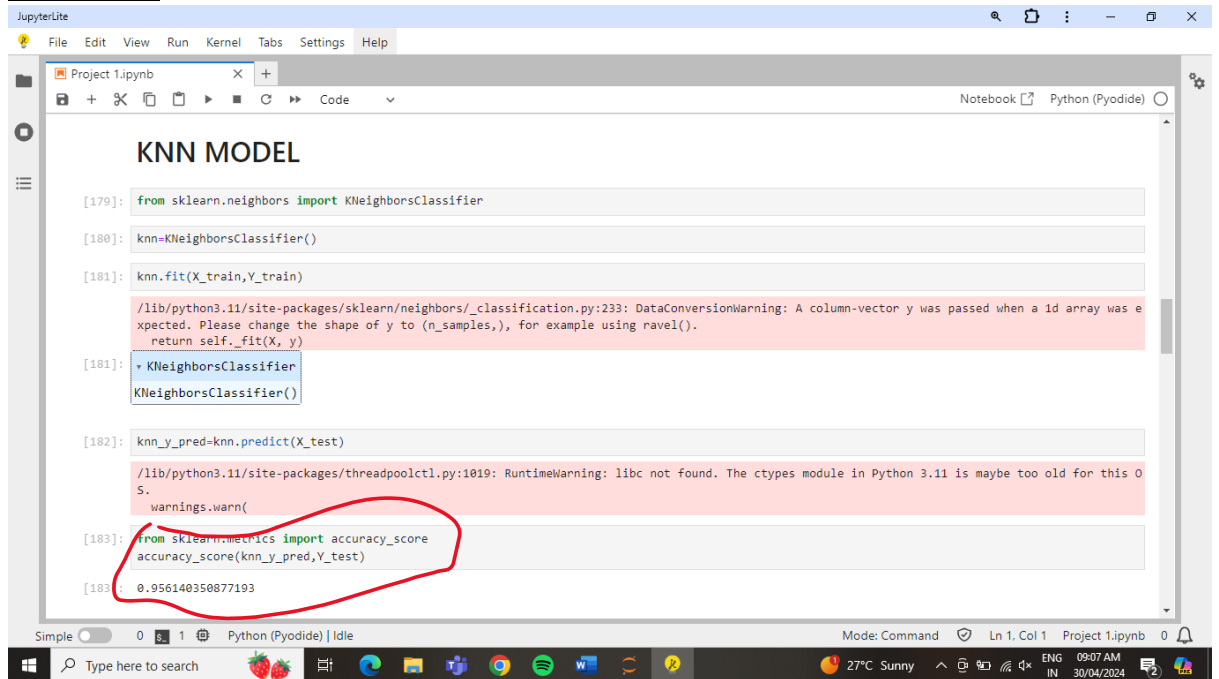
/lib/python3.11/site-packages/sklearn/neighbors/_classification.py:233: DataConversionWarning: A column-vector y was passed when a 1d array was expected. Please change the shape of y to (n_samples,), for example using ravel().
return self._fit(X, y)

[181]: KNeighborsClassifier
KNeighborsClassifier()

[182]: knn_y_pred=knn.predict(X_test)

/lib/python3.11/site-packages/threadpoolctl.py:1019: RuntimeWarning: libc not found. The ctypes module in Python 3.11 is maybe too old for this OS.
warnings.warn(
```

5. Prediction:-



```
[179]: from sklearn.neighbors import KNeighborsClassifier
[180]: knn=KNeighborsClassifier()
[181]: knn.fit(X_train,Y_train)
/lib/python3.11/site-packages/sklearn/neighbors/_classification.py:233: 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().
return self._fit(X, y)
[181]: KNeighborsClassifier
KNeighborsClassifier()
[182]: knn_y_pred=knn.predict(X_test)
/lib/python3.11/site-packages/threadpoolctl.py:1019: RuntimeWarning: libc not found. The ctypes module in Python 3.11 is maybe too old for this OS.
warnings.warn(
[183]: from sklearn.metrics import accuracy_score
accuracy_score(knn_y_pred,Y_test)
[183]: 0.956140350877193
```

As per the modelling KNN prediction this model gives 95% accuracy.

Conclusion:-

This project successfully demonstrates the feasibility and effectiveness of using Machine Learning techniques for automated breast cancer detection. The developed model can potentially aid healthcare professionals in making more accurate and timely diagnoses, leading to better patient outcomes.