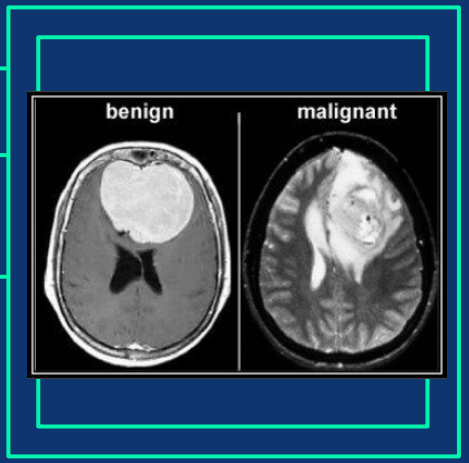
CLASSIFICATION OF MALIGNANT/BENIGN CANCERS

By: Pratishrut Kamal

TABLE OF CONTENTS



PROBLEM

Most hospitals require a classification for whether a scan of a possibly malignant tumor is actually malignant or whether it is benign, we want to create a model that can accurately (100% of the time) be able to

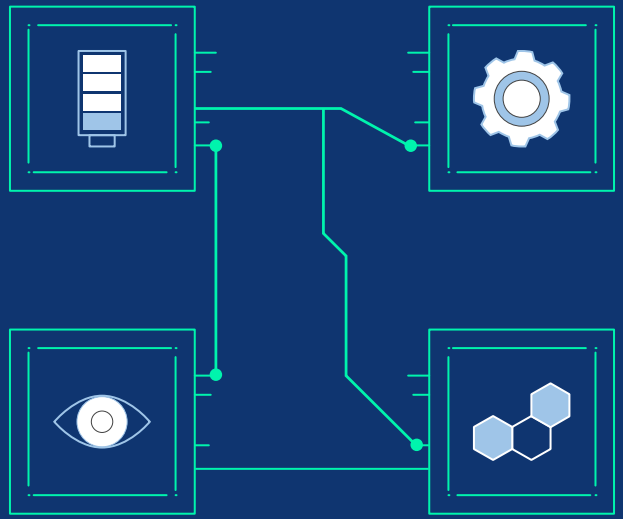
DATA ANALYSIS

CANCER.CSV

Contains numeric values based on CT Scans of Tumors

REMOVAL OF DATA POINTS

This decision was made with the intention of increasing the accuracy by decreasing the amount of variables and prioritizing the important and consistent data points

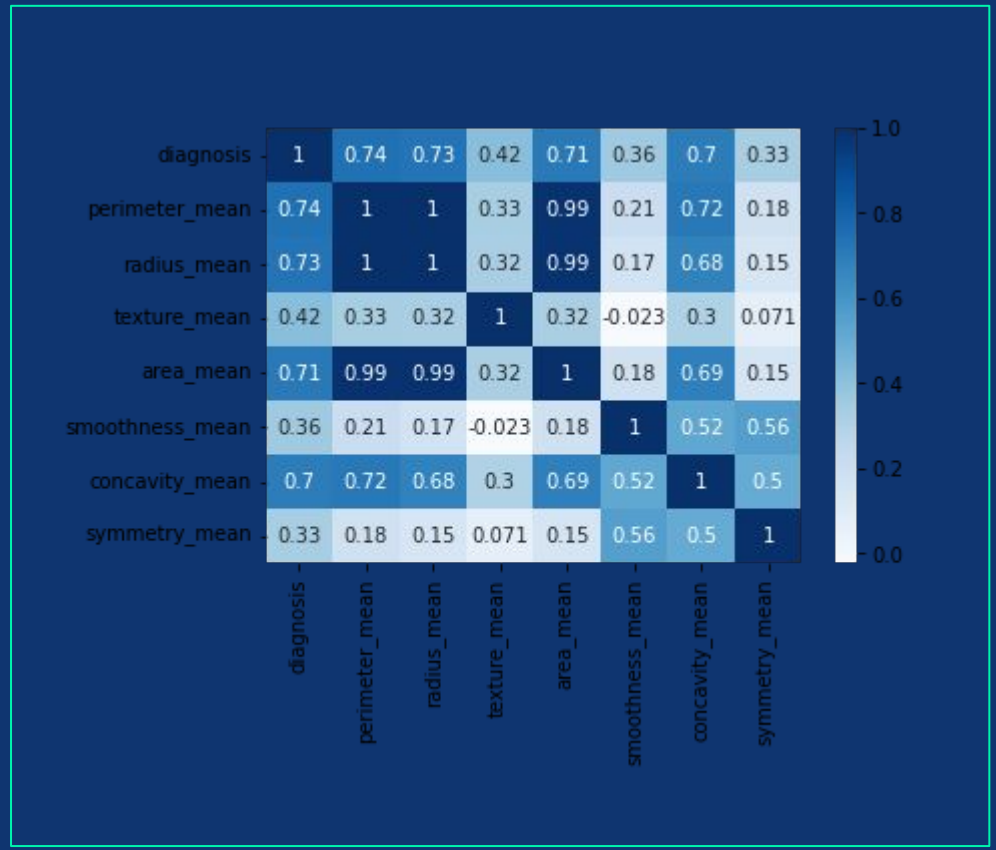


STANDARDSCALER

Standard Scaler is a module from sci-kit learn that will give every data point an equal weightage when making the predictions

DATA ANALYSIS

The Data that was provided for this project consisted of observed numerics relating to the CT Scan of a Tumor in determining whether it is Malignant or Benign, these metrics were: the average perimeter, radius, texture, area, smoothness, concavity, and symmetry of the cells in the tumor



MACHINE LEARNING MODEL

Two machine learning models were used for this project, Logistic Regression and Neural Networks, both yielding a 100% accuracy in detection which is optimal in a real medical setting

```
print(classification_report(y_test, pred_log))
```

	precision	recall	f1-score	support
0	1.00	1.00	1.00	71
1	1.00	1.00	1.00	43
accuracy			1.00	114
macro avg	1.00	1.00	1.00	114
weighted avg	1.00	1.00	1.00	114

```
from sklearn.neural_network import MLPClassifier
mlpc = MLPClassifier(hidden_layer_sizes=(11,11,11),max_iter=500)
mlpc.fit(X_train, y_train)
pred_mlpc=mlpc.predict(X_test)
print(classification_report(y_test, pred_mlpc))
```

	precision	recall	f1-score	support
0	1.00	1.00	1.00	71
1	1.00	1.00	1.00	43
accuracy			1.00	114
macro avg	1.00	1.00	1.00	114
weighted avg	1.00	1.00	1.00	114

QUESTIONS ANSWERED

- We have answered the correlation between certain metrics of CT Scans of tumors with the nature of the tumor

QUESTIONS LEFT UNANSWERED

- Are there any other classifications that we can make from just this data about tumors?
- Are there other machine learning models that can also perform the same task and the same accuracy and efficiency?

EXTRA DATA

- Extra data that could help this classification algorithm would be more related to predicting other data points in the data set that we were provided

THANK YOU

slidesgo