

Import Library

```
import numpy as np
import pandas as pd
import matplotlib.pyplot as plt

from sklearn.model_selection import train_test_split
```

Load Dataset from local directory

```
from google.colab import files
uploaded = files.upload()
```

Choose Files Breast_cancer_data.csv

- **Breast_cancer_data.csv**(text/csv) - 19654 bytes, last modified: 4/19/2023 - 100% done
Saving Breast_cancer_data.csv to Breast_cancer_data.csv

Importing the dataset

```
dataset = pd.read_csv('Breast_cancer_data.csv')
print(dataset.shape)
print(dataset.head(5))
```

```
(569, 6)
   mean_radius  mean_texture  mean_perimeter  mean_area  mean_smoothness  \
0         17.99         10.38         122.80        1001.0         0.11840
1         20.57         17.77         132.90        1326.0         0.08474
2         19.69         21.25         130.00        1203.0         0.10960
3         11.42         20.38          77.58         386.1         0.14250
4         20.29         14.34         135.10        1297.0         0.10030

   diagnosis
0          0
1          0
2          0
3          0
4          0
```

Segregating Dataset

```
X = dataset.iloc[:, :-1].values
y = dataset.iloc[:, -1].values
```

Splitting Dataset into Train & Test

```
X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.2, random_state=0)
```

Training with XCBOOST

```
from xgboost import XGBClassifier
model = XGBClassifier()
model.fit(X_train, y_train)
```

```
XGBClassifier
XGBClassifier(base_score=None, booster=None, callbacks=None,
               colsample_bylevel=None, colsample_bynode=None,
               colsample_bytree=None, early_stopping_rounds=None,
               enable_categorical=False, eval_metric=None, feature_types=None,
               gamma=None, gpu_id=None, grow_policy=None, importance_type=None,
               interaction_constraints=None, learning_rate=None, max_bin=None,
               max_cat_threshold=None, max_cat_to_onehot=None,
               max_delta_step=None, max_depth=None, max_leaves=None,
               min_child_weight=None, missing=nan, monotone_constraints=None,
               n_estimators=100, n_jobs=None, num_parallel_tree=None,
               predictor=None, random_state=None, ...)
```

Confusion Matrix

```
from sklearn.metrics import confusion_matrix, accuracy_score
y_pred = model.predict(X_test)
cm = confusion_matrix(y_test, y_pred)
print(cm)
accuracy_score(y_test, y_pred)
```

```
[[42  5]
 [ 3 64]]
0.9298245614035088
```

K-Fold Cross Validation

```
from sklearn.model_selection import cross_val_score
accuracies = cross_val_score(estimator = classifier, X = X_train, y = y_train, cv=10)
print(accuracies.mean()*100)
```

```
-----
NameError                                Traceback (most recent call last)
<ipython-input-10-07e50b19e5af> in <cell line: 2>()
      1 from sklearn.model_selection import cross_val_score
----> 2 accuracies = cross_val_score(estimator = classifier, X = X_train, y = y_train)
      3 print(accuracies.mean()*100)
```

NameError: name 'classifier' is not defined

SEARCH STACK OVERFLOW

[Colab paid products](#) - [Cancel contracts here](#)

0s completed at 10:22 PM

