```
# Title: KNN Algorithm


#Import Libraries
import numpy as np
import pandas as pd
from sklearn.neighbors import KNeighborsClassifier
import matplotlib.pyplot as plt
import seaborn as sns



df = pd.read_csv('/content/Data.csv')


print(df.shape)
```

```
(99, 13)
```

```
Start coding or generate with AI.


print(df.info())
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 99 entries, 0 to 98
Data columns (total 13 columns):
 #   Column                Non-Null Count  Dtype  
---  ------                --------------  -----  
 0   type                  99 non-null     object 
 1   fixed acidity         98 non-null     float64
 2   volatile acidity      98 non-null     float64
 3   citric acid           99 non-null     float64
 4   residual sugar        98 non-null     float64
 5   chlorides             98 non-null     float64
 6   free sulfur dioxide   99 non-null     float64
 7   total sulfur dioxide  99 non-null     float64
 8   density               99 non-null     float64
 9   pH                    98 non-null     float64
 10  sulphates             99 non-null     float64
 11  alcohol               99 non-null     float64
 12  quality               99 non-null     int64  
dtypes: float64(11), int64(1), object(1)
memory usage: 10.2+ KB
None
```

```
df["fixed acidity"].fillna("7.2",inplace=True)
df["volatile acidity"].fillna("0.23",inplace=True)
df["citric acid"].fillna("0.32",inplace=True)
df["residual sugar"].fillna("1.45",inplace=True)
df["chlorides"].fillna("0.044",inplace=True)
df["pH"].fillna("3.22",inplace=True)
df["sulphates"].fillna("0.52",inplace=True)
df.isnull().sum()
```

```
type                  0
fixed acidity         0
volatile acidity      0
citric acid           0
residual sugar        0
chlorides             0
free sulfur dioxide   0
total sulfur dioxide  0
density               0
pH                    0
sulphates             0
alcohol               0
quality               0
dtype: int64
```

```
print(df.head(5))
```

```
    type fixed acidity volatile acidity  ...  sulphates alcohol quality
0  white             7             0.27  ...       0.45     8.8       6
1  white           6.3              0.3  ...       0.49     9.5       6
2  white           8.1             0.28  ...       0.44    10.1       6
3  white           7.2             0.23  ...       0.40     9.9       6
4  white           7.2             0.23  ...       0.40     9.9       6
```

```
[5 rows x 13 columns]
```

```
#data Preprocessing
  #separate features fromm labels. (Vertical Split)
x = df.drop("quality",axis=1)
x=x.drop("type",axis=1)
y = df.quality
```

```
y.head(3)
```

```
0    6
1    6
2    6
Name: quality, dtype: int64
```

```
x.head(3)
```

	fixed acidity	volatile acidity	citric acid	residual sugar	chlorides	free sulfur dioxide	total sulfur dioxide	density	pH	sulphates	alcohol
0	7	0.27	0.36	20.7	0.045	45.0	170.0	1.0010	3	0.45	8.8
1	6.3	0.3	0.34	1.6	0.049	14.0	132.0	0.9940	3.3	0.49	9.5

```
from sklearn.model_selection import train_test_split
x_train,x_test,y_train,y_test = train_test_split(x,y,test_size=0.30,random_state=42)
```

```
print(x_train.shape)
print(x_test.shape)
print(y_train.shape)
print(y_test.shape)
```

```
(69, 11)
(30, 11)
(69,)
(30,)
```

```
model = KNeighborsClassifier(n_neighbors = 4)
```

```
model.fit(x_train,y_train)
```

```
KNeighborsClassifier(algorithm='auto', leaf_size=30, metric='minkowski',
                     metric_params=None, n_jobs=None, n_neighbors=4, p=2,
                     weights='uniform')
```

```
y_pred = model.predict(x_test)
```

```
from sklearn.metrics import accuracy_score
Accuracy = accuracy_score(y_test,y_pred)*100
Accuracy
```

```
33.33333333333333
```

```
from sklearn.metrics import confusion_matrix
cm=confusion_matrix(y_test,y_pred)
print(cm)
```

```
[[1 9 0 0]
 [4 9 1 1]
 [0 3 0 0]
 [0 2 0 0]]
```

```
import seaborn as sns
```

```
sns.heatmap(cm,annot=True)
```

<matplotlib.axes._subplots.AxesSubplot at 0x7f0bae2b83d0>