KNN CLASSIFICATION IMPLEMENTATION

CODE

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

data = pd.read_csv("/content/drive/MyDrive/JISNIT/Courses/ML/LectureNotes/data/Titanic.csv")
x = data.drop('Survived', axis = 1)
y = data['Survived']
x.drop(['Name', 'Ticket', 'Cabin'], axis = 1, inplace = True)
print(data.shape)
print(data.isna().sum())
# Missing value Imputation
# data = data.dropna(axis = 0, how ='any')
# Checking for missing value
# print(data.isna().sum())
# print(data.shape)
# numeric value imputation with mean
x['Age'] = x['Age'].fillna(x['Age'].mean())
x['Embarked'] = x['Embarked'].fillna(x['Embarked'].mode()[0])
x = pd.get_dummies(x, columns = ['Sex', 'Embarked'], prefix = ['Sex', 'Embarked'], drop_first = True)
print(x.head())
from sklearn.model_selection import train_test_split
x_train, x_test, y_train, y_test = train_test_split(x, y, test_size = 0.2, random_state = 0)
from sklearn.preprocessing import StandardScaler
std_x = StandardScaler()
x_train = std_x.fit_transform(x_train)
x_test = std_x.transform(x_test)
from sklearn.neighbors import KNeighborsClassifier
classifier = KNeighborsClassifier(n_neighbors = 5)
classifier.fit(x_train, y_train)
y_pred = classifier.predict(x_test)
from sklearn.metrics import classification_report, confusion_matrix
print(confusion_matrix(y_test, y_pred))
print(classification_report(y_test, y_pred))
```

OUTPUT

```
(891, 12)
PassengerId      0
Survived         0
Pclass           0
Name             0
Sex              0
Age            177
SibSp            0
Parch            0
Ticket           0
Fare             0
Cabin          687
Embarked         2
dtype: int64
PassengerId    0
Survived       0
Pclass         0
Name           0
Sex            0
Age            0
SibSp          0
Parch          0
Ticket         0
Fare           0
Cabin          0
Embarked       0
dtype: int64
(183, 12)
   PassengerId  Pclass   Age  SibSp  Parch     Fare  Sex_male  Embarked_Q  \
0            1       3  22.0      1      0   7.2500         1           0
1            2       1  38.0      1      0  71.2833         0           0
2            3       3  26.0      0      0   7.9250         0           0
3            4       1  35.0      1      0  53.1000         0           0
4            5       3  35.0      0      0   8.0500         1           0

   Embarked_S
0           1
1           0
2           1
3           1
4           1
[[99 11]
 [19 50]]
              precision    recall  f1-score   support

           0       0.84      0.90      0.87       110
           1       0.82      0.72      0.77        69

    accuracy                           0.83       179
   macro avg       0.83      0.81      0.82       179
weighted avg       0.83      0.83      0.83       179
```