```
import numpy as np
import pandas as pd
from sklearn.model_selection import train_test_split
from sklearn.linear_ import LogisticRegression
from sklearn.metrics import accuracy_score
credit_card_data  = pd.read_csv('https://www.kaggle.com/datasets')
credit_card_data,head() #for 1st 5 rows
credit_card_data.tail() # for ladt 5 rows
credit_card_data.info()#dataset info
#check missing value in each colomn
credit_card_data,isnull().sum()
credit_card_data['Class'].value_counts() # distribution of legal fraud
transcation
#separet data for analysis
legit = credit_card_data[credit_card_data.Class == 0]
fraud = credit_card_data[credit_card_data.Class == 1]
print(legit.shape)
print(fraud.shape)
legit.Amount.describe()
fraud.Amount.describe()
#compare values for both transacation
credit_card_data.groupby('Class').mean()
legit_sample = legit.sample(n=492) # random sample 492
new_dataset = pd.concat(legit_sample,fraud,axis=0)
new_dataset.head() #for 1st 5 rows
new_datset.tail() # for last 5 rows
new_dataset['Class'].value_counts()
new_dataset.groupby('Class').mean()
X = new_dataset.drop(columns='Class',axis=1)
Y = new_dataset['Class']
print(X)
print(YX_train,X_test,Y_train , Y_test = train_test_spli(X , Y ,
test_size=0.2, stratify=Y, random_state=2)
print(X.shape,X.train.shape,X_test.shape)
model = LogisticRegression()
#for training logistic regression model
model.fit(X_train, Y_train)
#find perfomance of model , accuracy
X_train_prediction = mpdel.predict(X_train)
training_data_accuracy = accuracy_score(X_train_prediction, Y_train)
print('Accuracy on Training data : ', training_data_accuracy)
X_test_prediction = model.predict(X_test)
test_data_accuracy = accuracy_score(X_test_prediction,Y_test)
print('Accuracy score on Test Data : ', test_data_accuracy)
```