

```

import numpy as np
import pandas as pd
from sklearn.model_selection import train_test_split
from sklearn.linear_model import LinearRegression
from sklearn.metrics import r2_score
import matplotlib.pyplot as plt
import seaborn as sns
import warnings
warnings.filterwarnings("ignore")
df=pd.read_csv('/content/Acidrain Dataset.csv')
print(df)
shape=df.shape
print(shape)
df.describe()
x=df.iloc[:,[0,3]]
y=df.iloc[:,4]
print(x)
print(y)
from sklearn.preprocessing import OneHotEncoder
ohe=OneHotEncoder(sparse=False)
x=ohe.fit_transform(df[['Normality']])
Y=ohe.fit_transform(df[['pH']])
print(x)
print(Y)
ohe.categories_
from sklearn.compose import make_column_transformer
col_trans=make_column_transformer(
    (OneHotEncoder(handle_unknown='ignore'), ['Normality']),
    remainder='passthrough')
x_train,x_test,y_train,y_test=train_test_split(x,y,test_size=0.25,random_state=0)
print("X_train:",x_train.shape)
print("X_test:",x_test.shape)
print("Y_train:",y_train.shape)
print("Y_test:",y_test.shape)
linreg=LinearRegression()
linreg.fit(x_train,y_train)
y_pred=linreg.predict(x_test)
y_pred
Accuracy=r2_score(y_test,y_pred)*100
print(" Accuracy of the model is %.2f" %Accuracy)
plt.scatter(y_test,y_pred);
plt.xlabel('Actual');
plt.ylabel('Predicted');
sns.regplot(x=y_test,y=y_pred,ci=None,color='red');
pred_df=pd.DataFrame({'Actual Value':y_test,'Predicted Value':y_pred,'Difference':y_test-y_pred})
print(pred_df)

```