"""

@author:Eswar

"""

"Import libraries"

import numpy as np

import pandas as pd

import os

import statsmodels.formula.api as stats

from statsmodels.formula.api import ols

import sklearn

from sklearn import linear\_model, datasets

from sklearn.metrics import mean\_squared\_error

import plotly.plotly as py

import plotly.graph\_objs as go

from plotly.offline import download\_plotlyjs,plot,iplot

"Read dataset file in CSV format"

power=pd.read\_csv("Folds5x2\_pp.csv")

print(power.head())

print(power.describe())

"Plot Heatmap"

corr\_matrix\_power=power.corr()

trace\_corr\_power=go.Heatmap(z=np.array(corr\_matrix\_power),x=corr\_matrix\_power.columns,y=corr\_matrix\_power.columns)

data\_power=[trace\_corr\_power]

plot(data\_power)

"Store independent variables in X and dependent variable in Y"

X= power.drop(['PE'],axis=1)

Y= power['PE']

"Linear Regression Model Implementation"

from sklearn.model\_selection import train\_test\_split

X\_train, X\_test, Y\_train, Y\_test = train\_test\_split(X,Y,test\_size=0.20)

from sklearn.linear\_model import LinearRegression

lm = LinearRegression()

lm.fit(X\_train,Y\_train)

Y\_pred= lm.predict(X\_test)

result\_lm=pd.DataFrame({

'Actual':Y\_test,

'Predict':Y\_pred

})

result\_lm['Diff']=Y\_test - Y\_pred

result\_lm.head()

print("Estimated Intercept is",lm.intercept\_)

print("Coefficients of the model:",lm.coef\_)

coef=zip(X.columns,lm.coef\_)

coef\_df = pd.DataFrame(list(zip(X.columns,lm.coef\_)),columns=['features','coefficients'])

print(coef\_df)

"Results"

from sklearn.metrics import mean\_absolute\_error

print('Mean Absolute Error:',mean\_absolute\_error(Y\_test,Y\_pred) )

print('Mean Squared Error:',mean\_squared\_error(Y\_test,Y\_pred) )

print('Root Mean Squared Error:', np.sqrt(mean\_squared\_error(Y\_test,Y\_pred)) )

print('Variance score:%.2f'%lm.score(X\_test,Y\_test))

import seaborn as sns

sns.regplot(x='Actual', y='Predict',data=result\_lm)