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

dataset = pd.read_csv("Expected_EndOfDay_ubs (2).csv")
X=dataset.iloc[:,:-1].values
#print (X)
Y = dataset.iloc[:,3].values


from sklearn.preprocessing import LabelEncoder,OneHotEncoder
# labelencoder and categorical featurees for feature field "INSTRUMENT".
Instrument_label = LabelEncoder()
X[:,0] =Instrument_label.fit_transform(X[:,0])


Instrument_onehot = OneHotEncoder(categorical_features=[5])



# labelencoder and categorical featurees for feature field "ACCOUNT TYPE".
Acct_type_labEncoder=LabelEncoder()
X[:,2] = Acct_type_labEncoder.fit_transform(X[:,2])

Acct_type_onehot= OneHotEncoder(categorical_features=[2]) # acct type =( E or I)
X = Acct_type_onehot.fit_transform(X).toarray()

from sklearn.cross_validation import train_test_split
# random_state means for single "epocha" here same records we be train again and
#again,i.e 10 records
X_train,X_test,Y_train,Y_test = train_test_split(X,Y,test_size=0.0001,random_state =0)

# why i had choose liner regisssion, because i have histrioc data and to predict depende labeli.e [

# I need to fit the liner model,so i have called linearegression
from sklearn.linear_model import LinearRegression
# to understand i have writen my name has dilip

Dilip_LReg = LinearRegression()

Dilip_LReg.fit(X_train, Y_train)

y_pred = Dilip_LReg.predict(X_test)
print(y_pred)
print(Y_test)
#print(X_train,X_test,Y_train,Y_test)
colors = ("black","red")

plt.scatter(y_pred,Y_test,c = colors)

 plt.xlabel("y_pred")

 plt.ylabel("Y_test")

 plt.show()
```