**Machine Learning Project**

**Problem Statement:**

**Glass Classification -** Can you correctly identify glass type?

This is a Glass Identification Data Set from UCI. It contains 10 attributes including id. The response is glass type (discrete 7 values)

**ML Methodology:**

**Linear Regression model:**

Linear Regression is a machine learning algorithm based on supervised learning. It performs a regression task. Regression models a target prediction value based on independent variables. It is mostly used for finding out the relationship between variables and forecasting. Different regression models differ based on – the kind of relationship between dependent and independent variables, they are considering and the number of independent variables being used.

**Dataset Description:**

Attribute Information:

1. Id number: 1 to 214 (removed from CSV file)
2. RI: refractive index
3. Na: Sodium (unit measurement: weight percent in corresponding oxide, as are attributes 4-10)
4. Mg: Magnesium
5. Al: Aluminum
6. Si: Silicon
7. K: Potassium
8. Ca: Calcium
9. Ba: Barium
10. Fe: Iron
11. Type of glass: (class attribute)
12. Building windows float processed
13. Building windows non-float processed
14. Vehicle windows float processed
15. Vehicle windows non-float processed (none in this database)
16. Containers
17. Tableware
18. Headlamps

**Pre-Processing:**

Pre-processing refers to the transformations applied to our data before feeding it to the algorithm

**In [1]:** %matplotlib inline

import matplotlib.pyplot as plt

import numpy as np

import pandas as pd

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

from sklearn import datasets

from sklearn.metrics import mean\_squared\_error,accuracy\_score,classification\_report

from sklearn.preprocessing import StandardScaler

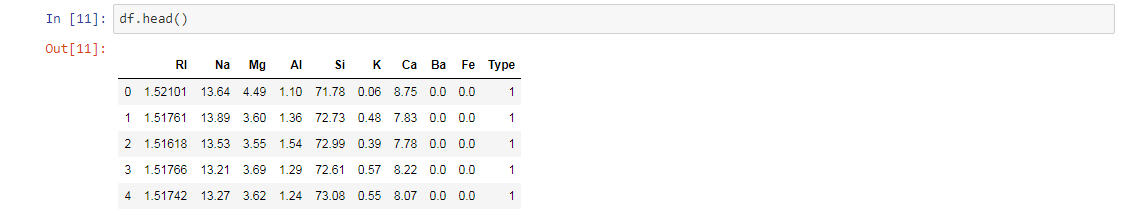
from sklearn.linear\_model import LinearRegression

**In [2]:** df=pd.read\_csv('d:\\glass.csv')

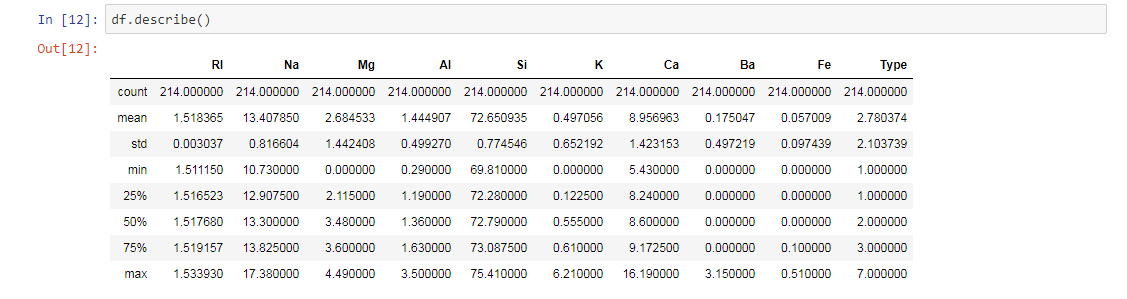
df.shape



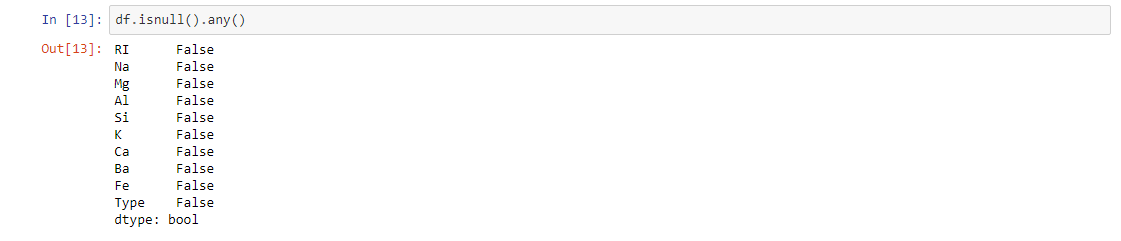
**In [3]:** df.head()



**In [4]:** df.describe()



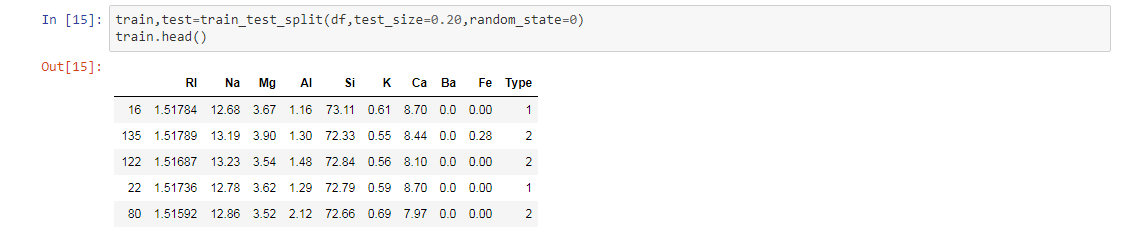
**In [5]:** df.isnull().any()



**Building, Training and Evaluation of all ML models:**

**In [6]:** train,test=train\_test\_split(df,test\_size=0.20,random\_state=0)

train.head()



**In [7]:** X\_train=train[['Na','Al','Si']]

y\_train=train.Type

**In [8]:** X\_test=test[['Na','Al','Si']]

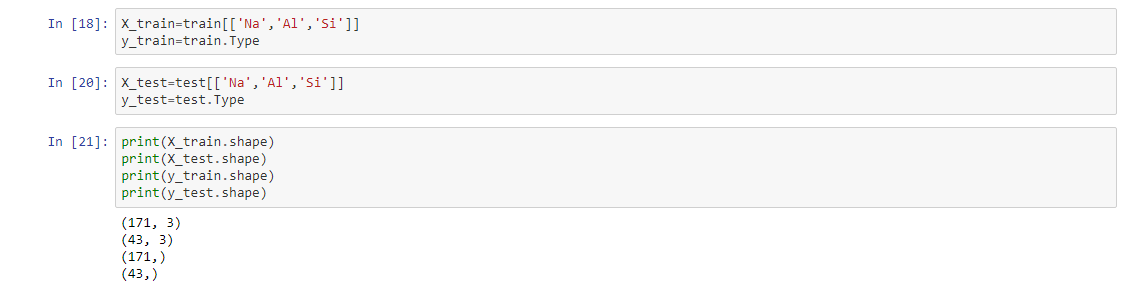
y\_test=test.Type

**In [9]:** print(X\_train.shape)

print(X\_test.shape)

print(y\_train.shape)

print(y\_test.shape)



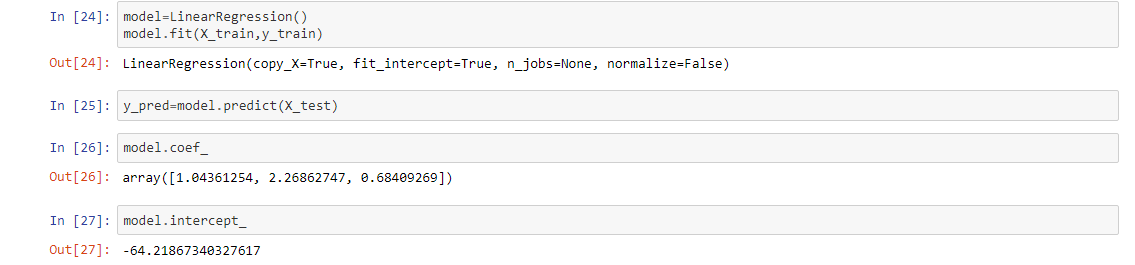
**In [10]:** model=LinearRegression()

model.fit(X\_train,y\_train)

**In [11]:** y\_pred=model.predict(X\_test)

**In [12]:** model.coef\_

**In [13]:** model.intercept\_

****

**In [14]:** print(y\_test[0:5])

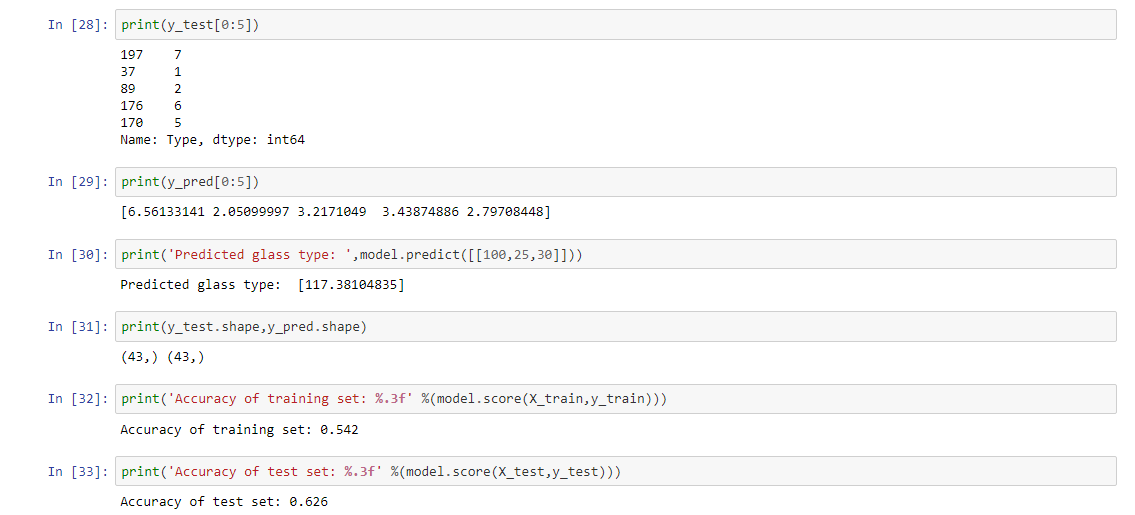
**In [15]:** print(y\_pred[0:5])

**In [16]:** print('Predicted glass type: ',model.predict([[100,25,30]]))

**In [17]:** print(y\_test.shape,y\_pred.shape)

**In [18]:** print('Accuracy of training set: %.3f' %(model.score(X\_train,y\_train)))

**In [19]:** print('Accuracy of test set: %.3f' %(model.score(X\_test,y\_test)))



**In [20]:** mse=mean\_squared\_error(y\_test,y\_pred)

print('Mean Squared Error: %.3f' %mse)



**In [21]:** y\_pred=y\_pred.round()

y\_test=y\_test.astype(float).round()

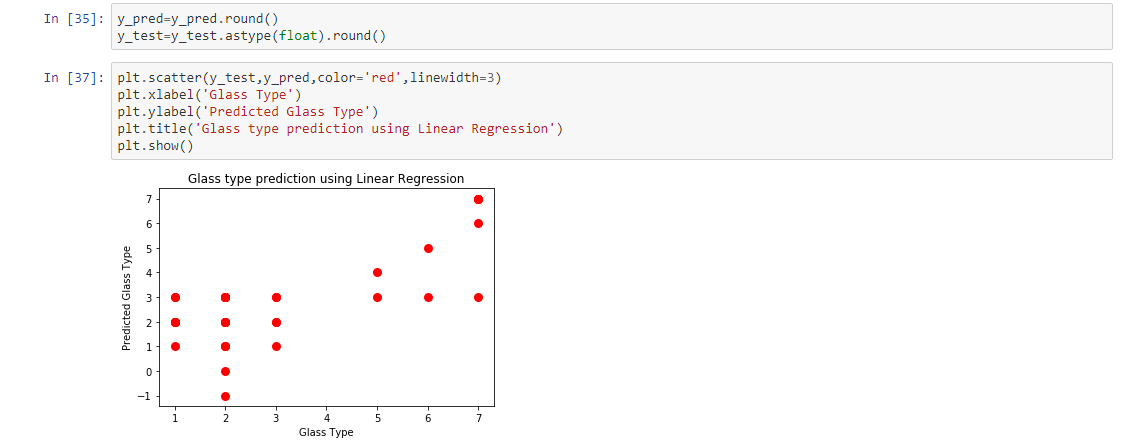
**In [22]:** plt.scatter(y\_test,y\_pred,color='red',linewidth=3)

plt.xlabel('Glass Type')

plt.ylabel('Predicted Glass Type')

plt.title('Glass type prediction using Linear Regression')

plt.show()



**Conclusion:**

We predict the glass type with the attributes Na - Sodium, Al - Aluminum, Si – Silicon by using Linear Regression model.