```
In [47]: import pandas as pd
    ...:
    ...: df=pd.read_csv('G:\Data Analysis\output.csv')
    ...: df=df.dropna()
    ...: X = df.iloc[:,[8,11,14,17,22,23,24,25,26,27,28,29,30,31,32,33,34,35,36,37]].values
    ...: y = df.iloc[:, 1].values

In [48]: from sklearn.model_selection import train_test_split
    ...: X_train, X_test, y_train, y_test = train_test_split(X, y, test_size = 0.3,
random_state = 0)

In [49]: from sklearn.linear_model import LinearRegression
    ...: regressor = LinearRegression()
    ...: regressor.fit(X_train, y_train)
Out[49]:
LinearRegression(copy_X=True, fit_intercept=True, n_jobs=None,
         normalize=False)

In [50]: import statsmodels.formula.api as sm
    ...:
    ...: model1=sm.OLS(y_train,X_train)
    ...: result=model1.fit()
    ...: print(result.summary())
                            OLS Regression Results
==============================================================================
Dep. Variable:                      y   R-squared:                       0.718
Model:                            OLS   Adj. R-squared:                  0.716
Method:                 Least Squares   F-statistic:                     288.2
Date:                Sat, 06 Jul 2019   Prob (F-statistic):              0.00
Time:                        19:08:44   Log-Likelihood:                -4783.3
No. Observations:                2279   AIC:                             9607.
Df Residuals:                    2259   BIC:                             9721.
Df Model:                          20
Covariance Type:            nonrobust
==============================================================================
                 coef    std err          t      P>|t|      [0.025      0.975]
------------------------------------------------------------------------------
x1             0.0043      0.006      0.759      0.448      -0.007       0.016
x2             0.0153      0.005      2.946      0.003       0.005       0.026
x3            -0.6731      0.127     -5.304      0.000      -0.922      -0.424
x4             0.0153      0.005      3.024      0.003       0.005       0.025
x5            -0.1794      0.049     -3.658      0.000      -0.276      -0.083
x6             0.1038      0.063      1.651      0.099      -0.019       0.227
x7             0.0761      0.051      1.482      0.138      -0.025       0.177
x8            -0.1075      0.045     -2.415      0.016      -0.195      -0.020
x9            -0.1070      0.054     -1.984      0.047      -0.213      -0.001
x10            0.0012      0.000      2.694      0.007       0.000       0.002
x11         3.384e-05      0.001      0.058      0.954      -0.001       0.001
x12            0.0018      0.000      4.370      0.000       0.001       0.003
x13            0.2305      0.102      2.258      0.024       0.030       0.431
x14            0.0007      0.000      2.683      0.007       0.000       0.001
x15           -0.0005      0.000     -1.460      0.145      -0.001       0.000
x16           -0.0017      0.000     -6.445      0.000      -0.002      -0.001
x17            0.0004      0.000      0.836      0.403      -0.001       0.001
x18           -0.0017      0.001     -3.310      0.001      -0.003      -0.001
x19           -0.0001      0.000     -0.277      0.782      -0.001       0.001
x20            0.0011      0.001      0.860      0.390      -0.001       0.003
```

```
==============================================================================
Omnibus:                     3072.586   Durbin-Watson:                   2.018
Prob(Omnibus):                  0.000   Jarque-Bera (JB):          1000006.122
Skew:                           7.374   Prob(JB):                         0.00
Kurtosis:                     104.555   Cond. No.                     2.99e+03
==============================================================================

Warnings:
[1] Standard Errors assume that the covariance matrix of the errors is correctly specified.
[2] The condition number is large, 2.99e+03. This might indicate that there are
strong multicollinearity or other numerical problems.
```

```
In [51]: c=0
    ...: y_pred = regressor.predict(X_test)
    ...: for i in range(len(y_pred)):
    ...:     y_pred[i]=(y_pred[i]<=y_test[i]+1.5  and y_pred[i]>=y_test[i]-1.5)
    ...:     if(y_pred[i]):
    ...:         c+=1
    ...:
    ...: acc=float(c/len(y_test))

In [52]: acc
Out[52]: 0.7656090071647902

In [53]: def backwardElimination(x, sl):
    ...:     numVars = len(x[0])
    ...:     for i in range(0, numVars):
    ...:         regressor_OLS = sm.OLS(y, x).fit()
    ...:         maxVar = max(regressor_OLS.pvalues).astype(float)
    ...:         if maxVar > sl:
    ...:             for j in range(0, numVars - i):
    ...:                 if (regressor_OLS.pvalues[j].astype(float) == maxVar):
    ...:                     x = np.delete(x, j, 1)
    ...:     regressor_OLS.summary()
    ...:     return x
    ...:
    ...:
    ...: SL = 0.05
    ...: import numpy as np
    ...: X=np.append(arr=np.ones((3256,1)).astype(int),values =X ,axis=1)
    ...: X_opt = X[:,[0,1,2,3,4,5,6,7,8,9,10,11,12,13,14,15,16,17,18,19,20]]
    ...: X_Modeled = backwardElimination(X_opt, SL)

In [54]: from sklearn.model_selection import train_test_split
    ...: X_train, X_test, y_train, y_test = train_test_split(X_Modeled, y, test_size = 0.3,
random_state = 0)
    ...:
    ...: from sklearn.linear_model import LinearRegression
    ...: regressor = LinearRegression()
    ...: regressor.fit(X_train, y_train)
Out[54]:
LinearRegression(copy_X=True, fit_intercept=True, n_jobs=None,
         normalize=False)

In [55]: model1=sm.OLS(y_train,X_train)
    ...: result=model1.fit()
    ...: print(result.summary())
                            OLS Regression Results
==============================================================================
```

```
Dep. Variable:                      y   R-squared:                       0.717
Model:                            OLS   Adj. R-squared:                  0.716
Method:                 Least Squares   F-statistic:                     442.7
Date:                Sat, 06 Jul 2019   Prob (F-statistic):              0.00
Time:                        19:09:15   Log-Likelihood:                -4787.0
No. Observations:                2279   AIC:                             9600.
Df Residuals:                    2266   BIC:                             9675.
Df Model:                          13
Covariance Type:            nonrobust
==============================================================================
                 coef    std err          t      P>|t|      [0.025      0.975]
------------------------------------------------------------------------------
x1             0.0183      0.004      4.358      0.000       0.010       0.027
x2            -0.6494      0.123     -5.265      0.000      -0.891      -0.408
x3             0.0162      0.005      3.364      0.001       0.007       0.026
x4            -0.1856      0.048     -3.881      0.000      -0.279      -0.092
x5             0.0700      0.053      1.333      0.183      -0.033       0.173
x6            -0.1175      0.044     -2.667      0.008      -0.204      -0.031
x7             0.0014      0.000      3.238      0.001       0.001       0.002
x8             0.0019      0.000      4.839      0.000       0.001       0.003
x9             0.2500      0.099      2.524      0.012       0.056       0.444
x10            0.0006      0.000      2.576      0.010       0.000       0.001
x11           -0.0006      0.000     -2.055      0.040      -0.001   -2.85e-05
x12           -0.0017      0.000     -6.624      0.000      -0.002      -0.001
x13           -0.0017      0.000     -3.391      0.001      -0.003      -0.001
==============================================================================
Omnibus:                     3070.587   Durbin-Watson:                   2.022
Prob(Omnibus):                  0.000   Jarque-Bera (JB):           992958.038
Skew:                           7.368   Prob(JB):                        0.00
Kurtosis:                     104.191   Cond. No.                     2.48e+03
==============================================================================

Warnings:
[1] Standard Errors assume that the covariance matrix of the errors is correctly specified.
[2] The condition number is large, 2.48e+03. This might indicate that there are
strong multicollinearity or other numerical problems.
```

```
In [56]: c=0
    ...: y_pred = regressor.predict(X_test)
    ...: for i in range(len(y_pred)):
    ...:     y_pred[i]=(y_pred[i]<=y_test[i]+2 and y_pred[i]>=y_test[i]-2)
    ...:     if(y_pred[i]):
    ...:         c+=1
    ...:
    ...: accAfterBackElimination=float(c/len(y_test))

In [57]: accAfterBackElimination
Out[57]: 0.9068577277379734

In [58]:
```