Assignment 2

Hrishikesh Dubey RA1911028010069

```
In [106]:  %matplotlib inline

           import math
           import numpy as np
           import pandas as pd
           import matplotlib.pyplot as plt

           df = pd.read_csv("seed_data.csv")
           df.head(10)
```

Out[106]:

	A	P	C	LK	WK	A_Coef	LKG	target
0	15.26	14.84	0.8710	5.763	3.312	2.221	5.220	0
1	14.88	14.57	0.8811	5.554	3.333	1.018	4.956	0
2	14.29	14.09	0.9050	5.291	3.337	2.699	4.825	0
3	13.84	13.94	0.8955	5.324	3.379	2.259	4.805	0
4	16.14	14.99	0.9034	5.658	3.562	1.355	5.175	0
5	14.38	14.21	0.8951	5.386	3.312	2.462	4.956	0
6	14.69	14.49	0.8799	5.563	3.259	3.586	5.219	0
7	14.11	14.10	0.8911	5.420	3.302	2.700	5.000	0
8	16.63	15.46	0.8747	6.053	3.465	2.040	5.877	0
9	16.44	15.25	0.8880	5.884	3.505	1.969	5.533	0

```
In [113]:  X = df.iloc[:, 1]
           y = df.iloc[:, 7]
```

```
In [123]:  df.plot.scatter('P','target', figsize=(15,4))
```

Out[123]: <matplotlib.axes._subplots.AxesSubplot at 0x22ed52d8898>
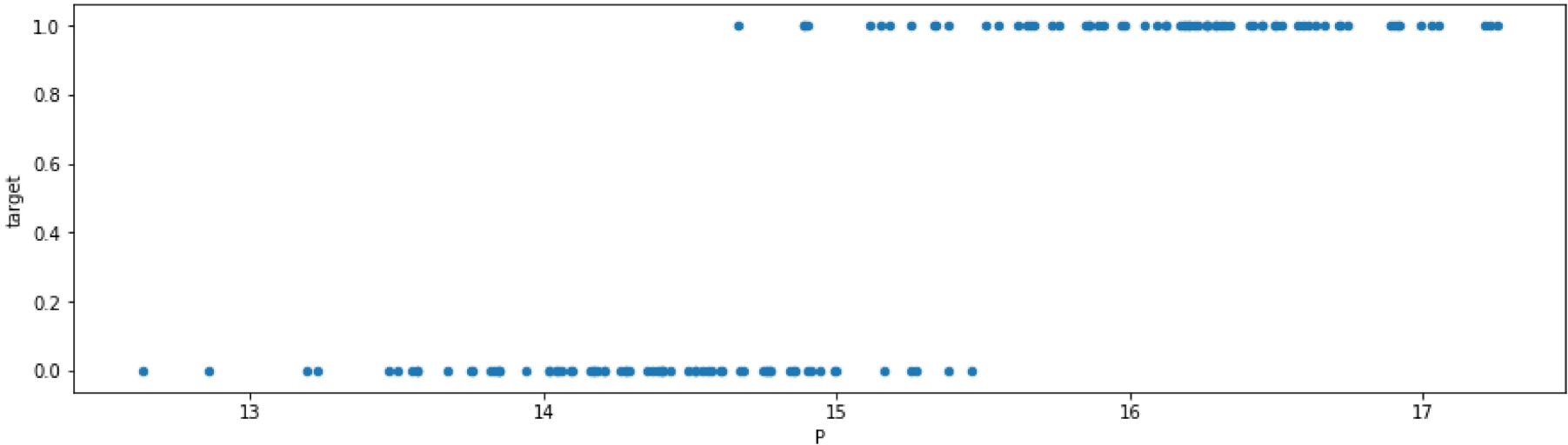
```python
In [114]: from sklearn.model_selection import train_test_split
          X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=1/3.0, rand
          om_state=0)


          X_train = X_train.values.reshape(-1, 1)
```

```python
In [115]: from sklearn.linear_model import LinearRegression

          regressor = LinearRegression()
          regressor.fit(X_train, y_train)

          y_pred = regressor.predict(X_test.values.reshape(-1, 1))

          y_pred
```

```
Out[115]: array([ 0.04578613, -0.17500894,  0.06902772,  0.18523565,  0.91734563,
                 -0.09753698,  0.92896643,  0.47575549,  0.56484823,  0.59583702,
                  1.08391034, -0.03555942, -0.03168582,  0.77014892,  0.7933905 ,
                  1.02580637,  0.9677024 ,  0.85149447,  0.0961429 ,  0.88248325,
                  1.18849748,  0.37891554,  0.99869119,  0.32468517,  0.06515412,
                  0.38278914, -0.0626746 ,  0.67718257,  0.1116373 ,  0.35954755,
                  0.16586766,  0.36342115,  0.89023045,  0.23171883,  0.75078093,
                  1.09165753, -0.4112984 ,  0.87086246,  1.21173906,  0.99481759,
                  0.23946602, -0.03943302, -0.28346968,  0.85149447,  1.28921102,
                  0.90572484,  0.66943537])
```

```
In [116]: y_test
```

```
Out[116]: 45      0
          59      0
          7       0
          50      0
          92      1
          27      0
          131     1
          137     1
          122     1
          8       0
          111     1
          16      0
          63      0
          76      1
          123     1
          97      1
          104     1
          110     1
          33      0
          91      1
          90      1
          22      0
          102     1
          24      0
          2       0
          51      0
          26      0
          71      1
          18      0
          10      0
          56      0
          43      0
          109     1
          48      0
          107     1
          83      1
          60      0
          106     1
          89      1
          78      1
          44      0
          30      0
          62      0
          121     1
          114     1
          73      1
          95      1
          Name: target, dtype: int64
```

```
In [122]: fig = plt.figure(figsize=(24, 4))
          plt.scatter(X_train, y_train)
          plt.plot(np.arange(12, 18, 1), regressor.predict(np.arange(12, 18, 1).reshape(-
          1, 1)), color='red')
          plt.xlabel('Perimeter of Seed')
          plt.ylabel('type')
          plt.show()
```
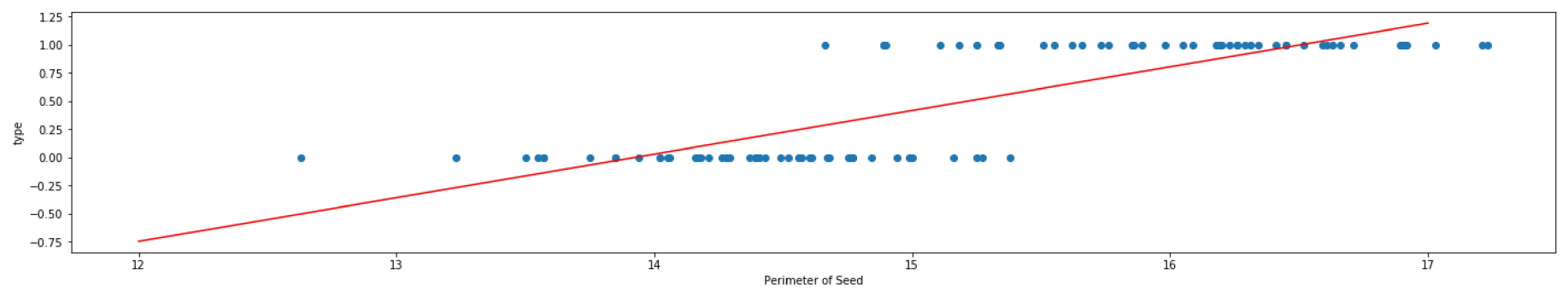


```
In [121]: from sklearn.model_selection import cross_val_score
          from sklearn.model_selection import KFold

          k = 5
          kf = KFold(n_splits=k, random_state=None)

          result = cross_val_score(regressor , X_train, y_train, cv = kf)

          print("Avg accuracy: {}".format(result.mean()))
```

Avg accuracy: 0.6298637222614076