multiple-linear-regression-1

December 17, 2023

0.1 This is a project of predicting agricultural yield with respect to different factors such as soil quality, seed variety, amount of fertilizer and weather conditions

0.1.1 Importing neccessary libraries

[1]:

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

[27]:

```
test = pd.read_csv("agricultural_yield_test.csv")
test.head(10)
```

[27]:

```
   Soil_Quality  Seed_Variety  Fertilizer_Amount_kg_per_hectare  Sunny_Days  \
0     93.304721             0                        132.522218   96.670922
1     83.674653             1                         57.283997   99.007556
2     65.963033             1                        227.895479  104.844272
3     78.692834             1                        176.314126   90.136191
4     72.415684             1                        160.070418  101.221668
5     62.697082             1                        268.439392   99.970710
6     92.942845             1                        271.488755   97.328051
7     75.518716             1                         84.047465   94.753411
8     93.673816             1                         79.386039   97.558543
9     86.607111             1                         76.430841   98.945328

   Rainfall_mm  Irrigation_Schedule  Yield_kg_per_hectare
0   602.386237                    3            278.986563
1   466.518251                    8            836.434840
2   510.320495                    4            785.881787
3   354.350914                    5            807.884526
4   443.993788                   10           1064.542374
5   485.223056                    8           1003.229410
6   447.403574                    2            758.716916
7   690.571998                    3            541.177487
8   674.733545                    4            567.855540
9   562.697441                    6            765.078954
```

[28]:
```
train = pd.read_csv("agricultural_yield_train.csv")
train.head(10)
```

[28]:
```
   Soil_Quality  Seed_Variety  Fertilizer_Amount_kg_per_hectare  Sunny_Days  \
0     96.415657             1                        147.853040   94.593926
1     92.352626             0                        281.565396   90.504644
2     63.714785             1                        137.864940   97.329340
3     90.084256             1                        100.946659  113.404828
4     81.600341             1                        223.088908   83.048176
5     65.394343             1                        104.484889   95.922140
6     71.035412             1                         78.523089   80.591350
7     69.335886             1                        135.922769  119.827004
8     86.424758             1                        165.704149   98.069899
9     60.884139             0                         88.220811   91.645943

   Rainfall_mm  Irrigation_Schedule  Yield_kg_per_hectare
0   444.267569                    3            683.759119
1   517.585491                    7            678.714861
2   420.310945                    8            934.691975
3   547.817646                    7            905.842541
4   434.726333                    6            897.584665
5   462.036153                    4            634.978213
6   536.457354                    3            593.545375
7   384.350380                    2            750.353033
8   553.276585                    5            803.008654
9   624.455301                    8            461.788596
```

0.1.2 Seperating test data into dependent and independent variables

[29]:
```
xtest=␣
  ↪test[['Soil_Quality','Seed_Variety','Sunny_Days','Fertilizer_Amount_kg_per_hectare','Rainfa
ytest=test[['Yield_kg_per_hectare']]
```

0.1.3 Seperating train data into dependent and independent variables

[30]:
```
xtrain=␣
  ↪train[['Soil_Quality','Seed_Variety','Sunny_Days','Fertilizer_Amount_kg_per_hectare','Rainf
ytrain=train[['Yield_kg_per_hectare']]
```

0.1.4 Model fitting

[31]:
```
from sklearn import linear_model
```

[32]:
```
model=linear_model.LinearRegression()
```

[36]:
```
model.fit(xtrain,ytrain)
```

```
[36]: LinearRegression()
```

```
[37]: model.coef_
```

```
[37]: array([[  1.54365383, 300.46370108,   1.99230463,   0.80880424,
         -0.50553551,  49.98672072]])
```

```
[38]: model.intercept_
```

```
[38]: array([48.33483511])
```

```
[39]: model.score(xtest,ytest)
```

```
[39]: 0.9355847925651241
```

0.1.5 Prediction

```
[43]: model.predict(xtest)
```

```
[43]: array([[337.58000764],
              [885.60020937],
              [785.78893206],
              …,
              [790.49868167],
              [510.052485  ],
              [317.74372373]])
```

0.1.6 Here we have different data set for both testing and training. We have fitted the model in trainiing data to train the model and later used testing data to predict the yield using our trained data. Here we got model score 0.9355847925651241 which is good . So we used testing data to predict our agriculture yeild

```
[ ]:
```