**The Sparks Foundation**

**Akanksha Pawar**

**Task 1 (Prediction using Supervised ML)**

Predict the percentage of a student based on the no. of study hours.

This is a simple linear regression task as it involves just 2 variables, that is, scores and no. of study hours.

**Data**

|  |  |
| --- | --- |
| **Hours** | **Scores** |
| 2.5 | 21 |
| 5.1 | 47 |
| 3.2 | 27 |
| 8.5 | 75 |
| 3.5 | 30 |
| 1.5 | 20 |
| 9.2 | 88 |
| 5.5 | 60 |
| 8.3 | 81 |
| 2.7 | 25 |
| 7.7 | 85 |
| 5.9 | 62 |
| 4.5 | 41 |
| 3.3 | 42 |
| 1.1 | 17 |
| 8.9 | 95 |
| 2.5 | 30 |
| 1.9 | 24 |
| 6.1 | 67 |
| 7.4 | 69 |
| 2.7 | 30 |
| 4.8 | 54 |
| 3.8 | 35 |
| 6.9 | 76 |
| 7.8 | 86 |

Percentage of a student is dependent upon the no. of hours studied. Therefore, scores is a dependent variable and hours is an independent variable.

**Y = a + b\*X**  where, Y = scores and X = hours of study

The regression results are as follows:

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| SUMMARY OUTPUT | |  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |  |
| *Regression Statistics* | |  |  |  |  |  |  |  |  |
| Multiple R | 0.976191 |  |  |  |  |  |  |  |  |
| R Square | 0.952948 |  |  |  |  |  |  |  |  |
| Adjusted R Square | 0.950902 |  |  |  |  |  |  |  |  |
| Standard Error | 5.603059 |  |  |  |  |  |  |  |  |
| Observations | 25 |  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |  |
| ANOVA |  |  |  |  |  |  |  |  |  |
|  | *df* | *SS* | *MS* | *F* | *Significance F* |  |  |  |  |
| Regression | 1 | 14624.17 | 14624.17 | 465.8229 | 9.13E-17 |  |  |  |  |
| Residual | 23 | 722.0683 | 31.39427 |  |  |  |  |  |  |
| Total | 24 | 15346.24 |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |  |
|  | *Coefficients* | *Standard Error* | *t Stat* | *P-value* | *Lower 95%* | *Upper 95%* | *Lower 95.0%* | *Upper 95.0%* |  |
| Intercept | 2.483673 | 2.531663 | 0.981044 | 0.336779 | -2.75347 | 7.720817 | -2.75347 | 7.720817 |  |
| X Variable 1 | 9.775803 | 0.452941 | 21.58293 | 9.13E-17 | 8.838823 | 10.71278 | 8.838823 | 10.71278 |  |

Therefore, the regression equation is

**Y = 2.483673 + 9.775803X**

Question: What will be predicted score if a student studies for 9.25 hrs/ day?

Answer: X = 9.25, then Y =2.483673 + 9.775803\*9.25

= 2.483673 + 90.4261778

= 92.9098508

**If the student studies for 9.25hrs/day the predicted score will be 92.91(approx.)**

**Scatter plot**