**NAME – DASHMESH SINGH**

**COLLEGE – DR B R AMBEDKAR NATIONAL INSTITUTE OF TECHNOLGY, JALANDHAR**

**DATA SCIENCE AND BUISNESS ANALYTICS**

**PREDICTION USING SUPERVISED MACHINE LEARNING**

**TASK – 1**

We need to predict the percentage of the student based on the number of study hours per day.

I will use Microsoft Excel to solve this supervised learning problem by using REGRESSION Model in Excel data analysis.

Let’s first analyze whether the given problem is linear or any other model problem.

The given data set is: <http://bit.ly/w-data>

Let’s visualize the data by plotting the scatter plot for given data as it is a simple problem in two variables.

It can be easily seen that it is a linear regression problem as both variables ‘HOURS’ and ‘PERCENTAGE’ has a linear relation with each other as in the scatter plot.

Now we need find the equation if this linear relation which we find with the help of the Microsoft Excel data analysis tool.

The result of the plotting of data on the excel is shown in the picture:

Graphical user interface, chart, scatter chart

Description automatically generated

The relation between both variables can be expressed in the form of a straight line:

y = mx + c

Percentage = m\*Hours + c

Where, m is the slope of the line and

c is the intercept of line on y-axis and these can be found out by using the analysis table.

The result of the data analysis is shown in the below picture:

Graphical user interface, application, table, Excel

Description automatically generated

The values of slope and intercept can be noted from the above arrow mark.

Now we can easily predict the percentage of the student by filling the numbers of hours in the cell of excel sheet below.

Graphical user interface, chart, scatter chart

Description automatically generated

**Q:** In the question we were asked to predict the percentage of the student who study for 9.25 hours/day?

**Sol:** We can find out the prediction of the marks of the student by putting the value of the hours = 9.25 in the excel sheet1. The result is shown in the figure below.

Chart, scatter chart

Description automatically generated

The prediction is found out to be equal to the **92.90985%**

**QUALITY OF THE REGRESSION**

The quality of the regression can be noted by the value of **R- square** along with lower 95% and upper 95% values.

The value of R- square is found out to be 0.952948197 which shows that the regression quality is quite good.

Hence, we have used the supervised learning to predict the score of the student based on the study hours per day.



The Excel sheet of the above solution is attached above.