**Introduction:**

As part of the internship program with The Sparks Foundation, a beginner task was given to make a prediction using Supervised ML. The dataset used was obtained from <http://bit.ly/w-data>.

The dataset contains two variables, which makes it very easy to use the linear regression model. The linear regression model was chosen because it allows for estimating the relationship between the two variables, in this case, hours and Scores. Therefore it allows estimating how a dependent variable changes as independent variables changes.

**Objectives**

The objective of the study was to predict the percentage of a student based on the no. of study hours.

Methodology/Workflow

Import necessary libraries such as

* **pandas** to store and manipulate the dataset
* matplotlib for visualization
* sklearn for machine learning models

Loading datasets from the URL link

Perform Data pre-processing

Perform exploratory data Analysis (EDA)

Perform Machine-learning models

**Conclusion**

The prediction shows that the student who studied for 9.25 hours/day will get a score of 92.9%.