Task 1

Prediction using Supervised ML
(Level – Beginner)

- Predict the percentage of an student based on the no. of study hours.
- This is a simple linear regression task as it involves just 2 variables.
- You can use R, Python, SAS Enterprise Miner or any other tool
- Data can be found at http://bit.ly/w-data
- What will be predicted score if a student studies for 9.25 hrs/ day?
- Sample Solution : https://bit.ly/2HxiGGJ
- Task submission:
  1. Host the code on GitHub Repository (public). Record the code and output in a video. Post the video on YouTube
  2. Share links of code (GitHub) and video (YouTube) as a post on **YOUR LinkedIn profile**, not TSF Network.
  3. Submit the LinkedIn link in Task Submission Form when shared.

Approach: A linear regression is a predictive modelling technique. It is used whenever there is a linear relation between the dependent and the independent variables. We will be using Python to tackle this task.

After importing the data using the Pandas library and removing the null values, we will assign the data into two variables (one encompassing the number of hours and the other indicating the scores).

We will then visualize this data using a scatterplot and start working on developing the actual regression.

For this, we first split the data into training and test data sets. We then train the linear regression on our training data set and notice that there is a positive correlation between the two variables.

We then predict the test result set and then use it to find the probable score if the student studies 9.25 hours a day.

Task 2

Approach: Exploratory Data Analysis is a method of evaluating or comprehending data in order to derive insights or key characteristics. EDA can be divided into two categories: graphical analysis and non-graphical analysis.

EDA is a critical component of any data science or machine learning process. You must explore the data, understand the relationships between variables, and the underlying structure of the data in order to build a reliable and valuable output based on it.

After loading our dataset using Python, we perform various visualizations to better position ourselves to perform data analysis and derive the desired insights.