**Project preparations**

1. You will have to prepare a project report with the following content:
   1. Table of content
   2. Table of figures/graphs
   3. Abstract
   4. Chapters (Introduction, Software-libraries used, Algorithm etc.)
   5. Conclusion
   6. Font style: Times New Roman
   7. Font size: Chapter heading: 16 (bold), Topics: 14 (bold), content: 14
   8. Margin: 1” on all sides
   9. The front page will contain the logo of your college and the company - Knowledge solutions India.
2. You will have to prepare a presentation of each individual candidate, explaining their parts in the project and will have to present it.
3. The report will be submitted to us and your respective college(if they ask) . In the companies copy the reports will be submitted in my name:

Gurvansh Singh

M.Tech

Knowledge Solutions India

1. **Time period for the project is 2 weeks, the code file should also be submitted. And one submission per group.**
2. **In the submission email mention your batch no. 8, your group no. and the names of your teammates**

**Project – 2**

**Insurance cost Prediction**

## **Content**

**Columns**

* age: age of primary beneficiary
* sex: insurance contractor gender, female, male
* bmi: Body mass index, providing an understanding of body, weights that are relatively high or low relative to height,  
  objective index of body weight (kg / m ^ 2) using the ratio of height to weight, ideally 18.5 to 24.9
* children: Number of children covered by health insurance / Number of dependents
* smoker: Smoking
* region: the beneficiary's residential area in the US, northeast, southeast, southwest, northwest.
* charges: Individual medical costs billed by health insurance(**target – y)**

## **Inspiration**

Can you accurately predict insurance costs?

Create a ML models to predict the ‘Insurance cost’ with minimum MSE and RMSE and maximum R-Square score. Build the following models:

1. Multiple linear regressor (MLR)
2. Random Forest Regressor (RFR)
3. MLR with PCA
4. RFR with PCA

Plot the actual and predicted values for all the four models and plot the most significant features (output of PCA) against the output y\_pred. Also write an inference on the models you prepared.