**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, Mathematics 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.
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. 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**

**Heart disease classification**

Content

The dataset contains several parameters which are considered important during the prediction of heart disease  
The parameters included are :

1. Age
2. Sex (0 – female and 1 – male)
3. cp: chest pain type  
   -- Value 1: typical angina  
   -- Value 2: atypical angina  
   -- Value 3: non-anginal pain  
   -- Value 4: asymptomatic
4. trestbps: resting blood pressure (in mm Hg on admission to the hospital)
5. 12 chol: serum cholesterol in mg/dl
6. fbs: (fasting blood sugar > 120 mg/dl) (1 = true; 0 = false)
7. restecg: resting electrocardiographic results  
   -- Value 0: normal  
   -- Value 1: having ST-T wave abnormality (T wave inversions and/or ST elevation or depression of > 0.05 mV)  
   -- Value 2: showing probable or definite left ventricular hypertrophy by Estes' criteria
8. thalach: maximum heart rate achieved
9. exang: exercise induced angina (1 = yes; 0 = no)
10. oldpeak = ST depression induced by exercise relative to rest
11. slope: the slope of the peak exercise ST segment  
    -- Value 1: upsloping  
    -- Value 2: flat  
    -- Value 3: downsloping
12. ca: number of major vessels (0-3) colored by flourosopy
13. thal: 3 = normal; 6 = fixed defect; 7 = reversable defect
14. num: diagnosis of heart disease (angiographic disease status)  
    -- Value 0: < 50% diameter narrowing  
    -- Value 1: > 50% diameter narrowing (**Predicted value)**

Create a ML models to predict the ‘target’ class and plot the confusion matrix(hint: sklearn.metrics -> plot\_confusin\_matrix <https://scikit-learn.org/stable/modules/generated/sklearn.metrics.plot_confusion_matrix.html>) for all the classifier models and also calculate the accuracy score for each. Build the following models:

1. Support Vector Machines (SVM)
2. K – nearest neighbor classifier (K-NN)
3. SVM with PCA
4. K-NN with PCA

# Plot the classification plot for all the four models and bar plot for Count of male and female having heart disease, Scatter plot between Age and Maximum heart rate.

Also write an inference on the models you prepared.