# Practical Data Science
# (Model Evaluation & Selection)

**Solve the following problems.**

**Problem 1: Model Evaluation with different resampling schemes**

Given the following data with Age as predictor variable and Survived as target variable:

| Id | Age | Survived |
|---|---|---|
| 1 | 25 | 1 |
| 2 | 23 | 0 |
| 3 | 30 | 1 |
| 4 | 35 | 1 |
| 5 | 32 | 0 |
| 6 | 28 | 1 |
| 7 | 13 | 0 |
| 8 | 12 | 0 |

Do the following:

a) Create a data frame with above sample data and do required type conversions

b) Build tree model with gender feature using caret train method(without trainControl argument)

c) Explore the object returned by train method
   a. **Model summary:** Find the summary of model to understand the default evaluation strategy used and accuracy of model
   b. **Final model:** Find out the actual model built by accessing finalModel field of returned object and also check how many observations used in building final model
   c. **Bias and variance:** Find out the bias and variance of finalModel across resampling iterations by accessing resample field of the returned object
   d. **Train and validation data:** Find the details of train and validation used across resampling iterations via control$index and control$indexout fields of the returned object
   e. **Confusion matrices:** Find out more deeper details about accuracy of models across resampling iterations via resampledCM field of returned object

d) Explore all the details expected by part-c above for each of the following options by using train control object in train method:
   i. Repeated holdout(LGOCV) with 3 iterations and 75% train data.
   ii. 4-fold cross validation(cv with number =4)

iii. 4-fold cross validation with 3 repeats(repeatedcv with number=4 and repeats=3)
iv. Leave one out cross validation(LOOCV)
v. Bootstrapping with 3 iterations(boot with number = 3)

**Problem 2: Model Selection**

Given the following data with Age, Gender, Location as predictor variables and Survived as target variable:

| Id | Age | Gender | Location | Survived |
|---|---|---|---|---|
| 1 | 25 | F | S | 1 |
| 2 | 23 | M | Q | 0 |
| 3 | 30 | F | Q | 1 |
| 4 | 35 | M | C | 1 |
| 5 | 32 | F | S | 0 |
| 6 | 28 | F | S | 1 |
| 7 | 13 | M | Q | 0 |
| 8 | 12 | F | C | 0 |

Do the following:

a) Create a data frame with above sample data and do required type conversions

b) Build tree model with
   a. Features: gender
   b. Evalution strategy: 10-fold cross validation

c) Build tree model with
   a. Features: gender and age
   b. Evalution strategy: 10-fold cross validation

d) Build tree model with
   a. Features: gender, age and location
   b. Evalution strategy: 10-fold cross validation

e) Find out the bias and variance of the above 3 models. Suggest the final model for deployment.

f) Use the deployed model to predict the outcomes for following test data:

| Id | Age | Gender | Location |
|---|---|---|---|
| 9 | 26 | F | S |
| 10 | 36 | F | Q |