**Lab Assign-05- Logistic Regression Model**

* Read the dataset **"FlightDelays.csv"** that is provided to you.
* Build a suitable logistic regression model using R and predict the status of the flights in your test data set.
* Draw the ROCR and Lift curves of your model and comment on the effectiveness of your model.
* Develop some metrics to determine the accuracy of your classification model (Error % using confusion matrix)

**Notes**

1. The data set consists of 2201 airplane ﬂights in January 2004 from the Washington DC area into the NYC area. The characteristic of interest (the response) is whether or not a ﬂight has been delayed by more than 15min (coded as 0 for no delay, and 1 for delay).
2. **Explanatory Variable Description**
   1. Arrival airports (Kennedy, Newark, and LaGuardia);
   2. Departure airports (Reagan, Dulles, and Baltimore);
   3. Carriers
   4. Categorical variable for 16 different hours of departure (6 am to 10 pm)
   5. Weather conditions (0=good/1=bad);
   6. day of week (1 for Sunday and Monday; and 0 for all other days).
3. Here the objective is to identify ﬂights that are likely to be delayed. The binary classiﬁcation problem, which amounts to deciding whether a new case with given features is either a “success” (in this case, delayed) or a “failure” (in this case, on-time) is very common in data mining, and this is what makes the logistic regression approach so useful.
4. The R function **glm** is used to fit logistic (regression) models. The syntax for this command is given below:

glm is used to fit generalized linear models, specified by giving a symbolic description of the linear predictor and a description of the error distribution.

**glm(formula, family = gaussian, data, weights, subset,**

**na.action, start = NULL, etastart, mustart, offset,**

**control = list(...), model = TRUE, method = "glm.fit",**

**x = FALSE, y = TRUE, singular.ok = TRUE, contrasts = NULL, ...)**

Type >help(“glm”) in Rstudio to get detailed help on **glm** command.