

# Assignment-3

**What to submit?** - Submit a R Markdown file that contains R codes, output and responses as required below. The file submitted should be Word or PDF.

*Max points: 100*

In this analysis you will develop logistic regression model based on the data set provided to predict whether or not the specimens are genuine.

**Data Set Information:** Data (A6DATA.csv) were extracted from images that were taken from genuine and forged banknote-like specimens. For digitization, an industrial camera usually used for print inspection was used. The final images have 400x 400 pixels. Due to the object lens and distance to the investigated object gray-scale pictures with a resolution of about 660 dpi were gained. Wavelet Transform tool were used to extract features from images.

**Attribute Information:**

V1: variance of Wavelet Transformed image (continuous)  
V2: skewness of Wavelet Transformed image (continuous)  
V3: kurtosis of Wavelet Transformed image (continuous)  
V4: entropy of image (continuous)  
V5: class (0-forged, 1-genuine)

Read the A6DATA.csv data file into RStudio. Run `set.seed(681)` for partitioning of the dataset into training (60%) and testing (40%). Report on the number of forged and genuine banknote-like specimens in the training and testing data. To split data in to train and test, use following for consistency in results:

Develop a logistic regression model using the training data. Provide final logistic regression model (with only significant variables), equation for calculating probability that specimen is genuine, confusion matrix for both training & testing data, misclassification error for both training & testing data, and discuss performance of the model.