**CENG 4515 – CLASSIFICATION-CLUSTERING HOMEWORK**

Download the “mole2\_dataset.csv”. The data is about the moles. There are some features which will help you identify whether the mole is velus, melanoma, seborrheic keratosis or basal cell carcinoma . (Last column is your response variable).

**1.** Split your data into train (80%) and test (20%) datasets using caret’s createDataPartition function.

**2.** Use the skim\_to\_wide function in skimr package, and provide descriptive stats for each column. Comment!

**3.** Predict and impute the missing values with k-Nearest Neighbors using preProcess function in caret. Comment!

**4.** After you impute missing values, use variable transformations. Convert all the numeric variables to range between 0 and 1, by setting method=range in preProcess function.

**5.** Use caret’s featurePlot() function to visually examine how the predictors influence the predictor variable. Comment!

**6.**

**a.** Use train() function to build the machine learning model. Choose knn algorithm.

**b.** Make predictions for test data using the predict() function.

**c.** Construct the confusion matrix to compare the predictions (data) vs the actuals (reference).

**7.**

**a.** Use train() function to build the machine learning model. Choose random forest algorithm.

**b.** Make predictions for test data using the predict() function.

**c.** Construct the confusion matrix to compare the predictions (data) vs the actuals (reference).

**8.**

**a.** Use train() function to build the machine learning model. Choose naïve bayes classification algorithm.

**b.** Make predictions for test data using the predict() function.

**c.** Construct the confusion matrix to compare the predictions (data) vs the actuals (reference).

**9.** Compare and make more and more comments about the final results you find in steps “6-8”.

**Submit as html file you create in rmarkdown ☺**