**Exam 2 Instructions**

1. **Dataset**

For the coding exam 2, you will pick your dataset. The dataset should be related to the classification problem. To receive full credit for this part, your datasets should satisfy the following conditions:

* **Data set should be imbalanced data set with respect to the target variable. The minority class(es) should not have more than 30% observations.**
* At least 15 features (columns)
* At least 1000 instances (rows)
* At least two categorical/ordinal columns.
* Between 5 to 10 percent missing values across the dataset.

1. **Project Description:**

* Read data into Jupyter notebook, use pandas to import data into a data frame
* Preprocess data: Explore data, check for missing data and apply data scaling etc.

1. **Select metric for evaluation**

**Classification Task:**

1. **Basic Algorithms**
2. Naive Algorithms
   * 1. **Accuracy**: Predict the majority class (class 0).
     2. **G-Mean**: Predict a uniformly random class.
     3. **F-Measure**: Predict the minority class (class 1).
     4. **ROC AUC**: Predict a stratified random class.
     5. **PR ROC**: Predict a stratified random class.
     6. **Brier Score**: Predict majority class prior.
   1. Logistic Regression
   2. Decision Tree
   3. k-Nearest Neighbors
   4. Support Vector Machine
   5. Random Forest
   6. Extra Trees
   7. Gradient Boosting
   8. XgBoost
   9. Stackin**g Classifiers**
3. **Cost Sensitive Algorithms**
   1. Logistic Regression
   2. Decision Trees
   3. Support Vector Machines
   4. Random Forest
   5. XGBoost
   6. Extra Trees
   7. Bagging decision tree with under sampling
4. **Data Sampling Algorithms (**pick one under sampling/oversampling)
   1. Logistic Regression
   2. Decision Tree
   3. k-Nearest Neighbors
   4. Support Vector Machine
   5. Random Forest
   6. Easy Ensemble Classifier
   7. XgBoost
   8. Stacking Classifiers

* Choose the final model – Make prediction on test set using final model. Report the score on test set using final model.

**Deliverables:**

* Use "download as" in the "file" menu to convert your ipython file to a .pdf file
* Submit two files each for Regression and Classification task: .ipynb, and .pdf files to the eLearning