

## **CS5227DM Datamining: Assignment on Classification.**

In this assignment you will prepare classifiers to participate in the Kaggle competition “Titanic: Machine Learning from Disaster” available at

<https://www.kaggle.com/c/titanic>

Description of the prediction task and the data are available on the website.

You are required to build three different classifiers for this prediction task and compare their results. The three classifiers should include a Decision tree-based classifier (E.g., J48, ID3), a Rule-based classifier (E.g., FOIL [1], CPAR [2]) and a Bayesian classifier (E.g., Naive Bayes, Bayesian Nets). You may use implementations from Machine learning libraries such as WEKA.

Do the following for each of the three classifiers,

1. Briefly describe the classification algorithm.
2. Use the data in “train.csv” to train and evaluate the your classifier. Use 10-fold cross validation and report precision, recall and F1 measure.
3. Retrain you classifier using the entire training set in “train.csv” and use the trained classifier to make prediction for the unseen examples in “test.csv”. Submit your predictions to Kaggle and provide a screen-shot of your result.

### **References:**

1. J. R. Quinlan and R. M. Cameron-Jones. FOIL: A midterm report. In Proc. 1993 European Conf. Machine Learning, pp. 3–20, Vienna, Austria, 1993.
2. X.Yin and J.Han, "CPAR: Classification Based on Predictive Association Rules", Proceedings of SIAM International conference on Data Mining, 2003, pp.331-335.