**CS 622, Introduction to Machine Learning Project Proposal**

**Student Full Name: Sayyed Ahmad Naghavi Nozad Student NSHE: 2001868645**

# Member names:

This team has only one member.

# Problem definition

In data analysis, outliers or anomalies (or even in some cases, novelties) are those rare items that deviate significantly from a well-known notion of normal behavior. Such examples may cause suspicions that these data samples might be deliberately created out of a discrepant mechanism or not be in harmony with the remainder of data. Outlier detection refers to methodologies that aim at discovering such rare and exceptional examples in data.

# List of competing methods

* **LOF (Local Outlier Factor):** This technique uses the distance values of each object to its nearest neighbors to compute local densities. However, LOF has a drawback: the scores obtained through this approach are not globally comparable between all objects in the same dataset or even in different datasets.
* **LoOP (Local Outlier Probability):** This is an enhanced version of LOF. LoOP gives each object a score in the interval [0,1], which is the probability of the object being an outlier and is widely interpretable among various situations.
* **iForest (Isolation Forest):** The main idea behind iForest was elicited from a well-known ensemble method called Random Forests, which is mostly employed in classification problems. In this isolation-based approach, some isolation trees based on some optimal parameters are built and then by traversing these trees, some anomaly degrees are attained based on which outliers are distinguished.

# Data description

We check the competing methods on the tailored datasets from the “Outlier Detection Data Sets (ODDS)” repository belonging to the Stony Brook University, New York.

# Experiment design

As we are following mostly unsupervised methods, we will follow the usual evaluation metrics in this area, including AUROC (Area Under Receiver Operating Curve) and AUPRC (Area Under Precision-Recall Curve). Other evaluation metrics might be considered too.