**ADMN5006 Assignment 1**

**Eyas Abu Elhouf, Katrina Ong, Melvin Cheriyan**

**Pranav Gujar, Bista Bijo, Praveen Kumar**

**Outline**

1. Business Use Case
   1. Detecting bankrupt companies
      1. Credit agencies
      2. Loan applications
      3. Investors
2. Applicable machine learning methods (Classification)
   1. Logistic Regression
   2. Naïve Bayes
   3. Random Forest Classifier
3. Methodology:
   1. Data Exploration:
      1. Check data types
      2. Explore Statistics and distribution
      3. Checking outliers
      4. Checking Missing Values
   2. Data Cleaning (On entire dataset)
      1. Remove Duplicates
      2. Remove records with more than 5 missing values
   3. Model Building
      1. Train-Test Split (75%, 25%)
      2. Data preparation after train-test split
         1. Impute Missing Values
            1. Fit the imputer to the training data
            2. Fill in missing values in both train and test data
         2. Treat Outliers
            1. Find upper and lower bounds of training data
            2. Substitute outliers in both training and test data using upper and lower bounds
         3. Resampling Imbalanced Training Data
            1. Used SMOTE oversampling method from Imbalanced Learn
      3. Train the model
   4. Validate the Model
      1. Performance metrics
   5. Hyperparameter Tuning
   6. Model Selection