**DS-5220 Course project- Survey of supervised classification techniques for income prediction using “Adult dataset”.**

**Team-25**

**Group members:**

Joyce Marathi- [marathi.v@northeastern.edu](mailto:marathi.v@northeastern.edu)

Rohit Barve- [barve.r@northeastern.edu](mailto:barve.r@northeastern.edu)

Lasya Manthripragada – [manthripragada.l@northeastern.edu](mailto:manthripragada.l@northeastern.edu)

**Introduction**:

This project’s goal is to find the best model to predict if the income of a person is greater than 50k or lesser. The data that the model is based on is ‘adult dataset’ from UCI machine learning repository [1]. The data pre-processing includes handling missing values, data imbalances etc.,

**Machine learning models**:

We will initially use various machine learning models to train the data

1. SVM
2. Random forest
3. Naïve bayes
4. Decision tree

Based on the results, we can select the best models to perform feature selection and note if it makes any difference once we train the resultant data.

**References:**

1. Dataset: https://archive.ics.uci.edu/ml/datasets/Adult