

IS 507 Assignment 8

Due by Sunday, December 5, 2021 11:59PM

Problem 1: The Excel spreadsheet Employee.csv contains one sheet named Employee, which is data attempting to explain employee retention. These are data from a sample of 4,654 employees and consists of 9 variables for each employee. These are:

- 1) Education Level (Categorical)
- 2) Joining Year (Year Employee joined the company)
- 3) City (Office Location)
- 4) Payment Tier (1-Highest Level; 2-Mid Level; 3-Lowest Level)
- 5) Age (Current Age of Employee)
- 6) Gender (Gender of Employee-Male/Female)
- 7) Ever Benched (Left off of a project for 1 or more months)
- 8) ExperienceInCurrentDomain (Level of Experience)
- 9) LeaveOrNot (1-Leaves Company in 2 years; 0-Does not Leave Company in 2 years)-Dep. Var.

Note: To use the categorical variables you will first need to convert them into numeric data. There are several different ways of completing this task, make sure you know your reference category.

Develop a **Logistic Regression model** to classify the LeaveOrNot event from the other variables.

- a) Create a logistic regression model and explain the significant odds ratios in terms of LeaveOrNot.
- b) Create a confusion matrix and explain how well the model is classifying the Leaves Company in 2 years events.
- c) Create an ROC curve and calculate the c-statistic (auc). What does this mean about the model?
- d) What are the differences between the information in part a and part b?
- e) How does this model differ from the linear discriminate analysis you ran in Assignment 7?