Project

BUSA 523

due: March 9th, 11:59 PM

- **Dataset information:**
  https://archive.ics.uci.edu/ml/datasets/Haberman%27s+Survival

1. Perform Exploratory data analysis. **(20)**
   - What can you conclude from methods of central tendency and deviation of the variables?
   - Are there any correlated variables?
   - Interpret the skewness of the variables.

2. Develop Supervised learning models. **(50)**
   - Logistic Regression
   - Random Forest
   - Support Vector Machine
   - Neural Network

   Compare the 4 methods based on the performance metrics and identify the best model for this problem.
   - Use 10 - fold cross validation.
   - Use GridSearchCV for searching through hyper parameters.

3. Evaluate if the algorithms are overfitting? **(10)**

4. Develop a function which can implement machine learning. **(20)**

   *def **machine_learning** (data, algorithm):*
   .....
   .....
   plot ROC curve
   return** **performance metrics

   Function call: machine_learning (data, neural_network)