STAT 574 PROJECT INFORMATION

For the project you should do the following:

1. Find a data set on some interesting topic that lends itself to Machine Learning techniques that we studied in the course. Please be considerate of other students' feelings and don't choose a topic that is too medical or too controversial and might shock someone or spark a heated discussion.
2. Email me a brief description of your topic and a snippet of your data set for approval by 10 PM on Wednesday, April 10, 2024, sent to stat574s24@gmail.com with "Project" in the subject line.
3. Techniques to choose from are limited to the following list (plus their extensions and variations):

    1) Decision tree
    2) Random forest
    3) Gradient boosting
    4) K-nearest neighbor
    5) Support vector machine
    6) Naïve Bayes
    7) Artificial neural network
    8) Recurrent neural network
    9) Convolutional neural network
    10) Anomaly detection
    11) Change-point detection
    12) Text mining

4. Choose one technique (possibly with some variations) and use SAS, R, and Python (as many software programs as possible) to run the analysis.
5. Prepare an **8-minute PowerPoint presentation** to talk about your project.
6. Email me your slides for approval **at least one day before** the day of your presentation.
7. Give your PowerPoint presentation at one of the sessions on May 1 or May 8 (see the schedule of presentations for details).
8. When listening to presentations of others, you will be required to fill out a questionnaire (see the posted file). Your attendance will be graded based on how well you complete the questionnaires. Presenters' grades will be based on your answers to those questionnaires. Please prepare 14 questionnaires (they are two per page). You will be listening to 14 presentations and filling out 14 questionnaires.
9. Write a report for your project. The report will be due in pdf format and sent to stat574s24@gmail.com (with "Report" in the subject line) by 10 PM on **the day of your presentation**.
10. Your score for the project will be based on: slides (25 points), presentation (25 points), attendance (25 points), and report (25 points).