TEST 3

Data Manipulation and Evaluation of SVM Classifier

For this exam you will be responsible for performing exploratory data analysis, creating, and testing an SVM classifier to predict the biodegradability of molecules given 41 different attributes. You will need to provide evidence of your work starting with the manipulation and the exploration of the data. You should end with a thorough evaluation of the model. You should explain what you attempted and the reasoning for why you attempted certain strategies.

You are free to do this assignment with any resources you can obtain. That means you can use the internet, your book, class code, and class notes. The only thing you may not use is other groups as a resource for code or algorithmic solutions. You must do your own work and own research. You may share sources (I found this site to be helpful, I found a solution using chapter 4 in our book, etc.). You may not copy and paste code; you should understand the code you write.

You can use IDLE, Jupiter Notebook or Anaconda to do this assignment. Just upload your file(s) to canvas when you are completed. You may provide images of EDA, visualizations of models and examples for how you cleaned your dataset. Your group should turn in evidence of each portion of this exam in the form of source code, images, and documentation.

You should implement and attempt to do feature selection, feature extraction, regularization of the model, normalization of the data (feature scaling), evaluation with the use of different metrics, a validation set to test hyperparameters of the SVM model and a brief explanation of why you think your model would be generalizable.

You and your partner should turn in one exam with both names. This is a chance for you to show me what you know! The more things you incorporate from class the better you will do on this exam.