

Consider the `drug200.csv` data file (posted under the In-Class 6 assignment link). This file contains information basic demographic and health information on 200 patients. The goal is to predict the drug type using patient's demographic and health information. **In Python**, answer the following:

1. (3 points) Using the pandas library, read the csv data file and create a data-frame called `drug`.
2. (3 points) Create the frequency table of the target variable `Drug`.
3. (4 points) Using the `where` function from `numpy`, create a new variable called `Drug_numb` as follows:
 - if `Drug = drugA`, then `Drug_numb = 1`
 - if `Drug = drugB`, then `Drug_numb = 2`
 - if `Drug = drugC`, then `Drug_numb = 3`
 - if `Drug = drugX`, then `Drug_numb = 4`
 - if `Drug = DrugY`, then `Drug_numb = 5`
4. (10 points) Change `Sex`, `BP` and `Cholesterol` from labels to dummy variables.
5. (5 points) Using `Age`, `Sex` (dummy variable), `BP` (dummy variables), `Cholesterol` (dummy variable), and `Na_to_K` as the input variables, and `Drug_numb` as the target variable, split the data into two data-frames (taking into account the proportion of 1s, 2s, 3s, 4s, and 5s) `train` (80%) and `test` (20%).
6. (8 points) Using `train` data-frame and the one-vs-all multi-class classification strategy with the random forest model (with 500 trees and the maximum depth of each tree equal to 3), build a multi-class classification model. Then, use this model to make predictions on the `test` data-frame. Compute the classification report of this model.
7. (8 points) Using `train` data-frame and the one-vs-all multi-class classification strategy with the AdaBoost model (with 500 trees, the maximum depth of each tree equal to 3, and learning rate equal to 0.01), build a multi-class classification model. Then, use this model to make predictions on the `test` data-frame. Compute the classification report of this model.
8. (3 points) Using the results from part 6 and 7, what model would use to predict `Drug`? Be specific.