1 Implement a Decision Tree

The decision tree algorithm is extremely powerful, interpretable, and widely used. In this assignment, you will implement a decision tree algorithm from scratch and apply it to a data set and see how well it performs. The task is to learn to predict from one of two data sets (your choice!) 1) Predict party affiliation (Democrat/Republican) from 1984 Congressional Voting Records, or 2) Predict whether someone has diabetes.

https://www.kaggle.com/devvret/congressional-voting-records

(a) The algorithm is as follows:

- For the current node N
  - If all training examples classified correctly then
    * Stop
  - Otherwise
    * Compute the splitting criterion for each attribute
    * Split the current node according to the attribute which maximizes the splitting criterion
    * Recursively apply to each newly created child node N_i

(b) Choose and implement a splitting strategy You can implement either Information Gain or GINI as discussed in class.

(c) Report the error rate After growing the tree on the training set, apply it to your test set and report the error rate. I recommend implementing the entire homework in an object oriented fashion. Create a class, and create a fit function. Use the skills you remember from tree traversals in your algorithms class. There will likely be some recursive methods. You are not required to print the tree.

(d) Push Jupyter Notebook to github

2 Deliverables

Complete the tasks above, and send an email with a link to your repo to conradkennington@boisestate.edu by the due date.