# Assignment 2

## Part A – Decision Tree:

The data set available for this assignment is based on the U.S. congress voting record from 1984. The data set consists of the votes (yes or no) on sixteen issues for each of the 435 members of congress. from the voting record. You will use this data to learn a decision tree that predicts the political party of the representative based on his /her vote.

1. Use the voting data to train a decision tree to predict political party (Democrat or Republican) based on the voting record. Use 25% of the members of congress for training and the rest for testing. Rerun this experiment five times and notice the impact of different random splits of the data into training and test sets. Report the sizes and accuracies of these trees in each experiment.

2. Measure the impact of training set size on the accuracy and the size of the learned tree. Consider training set sizes in the range (30-70%). Because of the high variance due to random splits repeat the experiment with five different random seeds for each training set size then report the mean, maximum and minimum accuracies at each training set size. Also measure the mean, max and min tree size.
   - Start with training data size 30%, 40% … Until you reach 70%.
   - The data set contained many missing values , i.e., votes in which a member of congress failed to participate. To solve those issue insert–for each absent vote–the voting decision of the majority.

# Part B - SVM:

The attached dataset **"heart.csv"** contain 303 records of patients have heart disease or not according to features in it. You are required to build **SVM** model to predict whether patient have heart disease or not **(target).**

**Cost Function:**

$$c(x, y, f(x)) = \begin{cases} 0, & \text{if } y * f(x) \geq 1 \\ 1 - y * f(x), & \text{else} \end{cases}$$

**Update Weights Equations:**

1- **If Point is correctly classified.**

$$w = w - \alpha \cdot (2\lambda w)$$

2- **If Point is not correctly classified.**

$$w = w + \alpha \cdot (y_i \cdot x_i - 2\lambda w)$$

a) Split dataset into training and testing sets.
b) Try different set of features and choose the best features.
c) Try different values of learning rate and see how this changes the accuracy of the model.
d) Implement accuracy function (correctly predicted values / test set size).

**Note: The final grade is proportional to the accuracy of your results.**

## Important Notes:

- You can only use “pandas”, “numpy” and “matplotlib” libraries. ***(Don’t use “sklearn”)***
- The maximum number of students in a team is 4 and minimum 3.
- No late submission is allowed.
- Cheating students will take negative grades and no excuses will be accepted.
- Deadline is on Sunday 27/12/2021 at 11:59 PM