5/4/21

*Aidan Duffy*

# ASSIGNMENT 1T: DECISION TREES

Please use this template in creating your response. Retain the gray text and MS Word headings, and supply your responses where indicated. Your materials, in black 12-point Times New Roman, should not exceed 5 pages excluding this gray text, the references and your figures. You may add appendices, which should be referred to in the body of the paper, and which will be read on an as-needed basis. Note the evaluation criteria below, and leave plenty of time for editing so that your paper responds to them and you obtain the most favorable grade. Responses considered “good” should go beyond the minimum of what’s requested.

Go [here](https://colab.research.google.com/drive/1DGZdokU6DOXCtDjYjGU_yev0_6KZOP2o?usp=sharing), and copy to your Google drive.

## 1. SUMMARY DESCRIPTION

Describe a decision tree application that you will implement by modifying the code you copied.

The decision tree application I decided to create was for retail banking institutions looking to expand the asset customer side, or customers who take out loans, of their business. The decision tree will help maximize the likelihood of folks targeted with their advertising in actually purchasing a loan. The dataset includes generic personal information like their age, how long they have been working, and where they live as well as relevant financial information such as do they have an investment or credit card account with this retail bank.

Note: Apologies for the partial completeness – I fell behind on work this week and got behind on lecture so was a bit confused, but I wanted to at least submit something.

## 2. DATA SOURCE

Explain the source of your data (e.g., by hand; Kaggle). Point to a URL so we can see it.

I decided to use a data set I found after some research on Kaggle. To view the dataset, click [**here**](https://www.kaggle.com/itsmesunil/bank-loan-modelling).

## 3. EXAMPLES

Create three different input sets to the anticipated application and show their outputs.

Your response goes below.

Input 1: High spending, low income. (CCAvg = 8, Income = 50)

Output 1: Has a loan

Input 2: Low spending, low income (CCAvg = 3, Income = 50)

Output 2: Has a loan

Input 3: High spending, high income. (CCAvg = 8, Income = 180).

Output 3: Does not have a loan

## 4. SOURCE

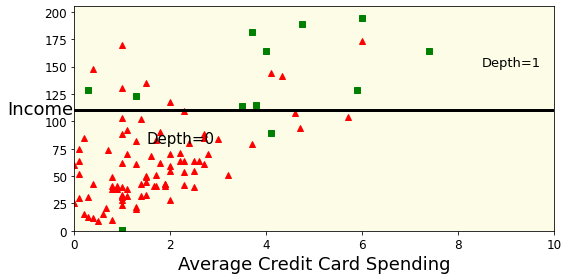
Please supply the URL of your (shared) Colab code.

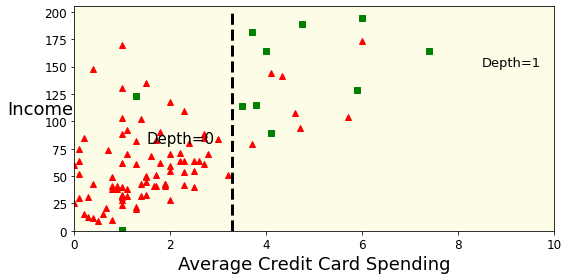
Please click [**here**](https://colab.research.google.com/drive/1NO2RPA2n1GCeDp7391tElyR_TCUjWLOg?usp=sharing) to access my colab code.

## 5. DATA ALTERATION FOR CHANGED RESULTS

Alter the data so that it changes in the outputs for the above inputs. Explain why. Limit: 2 normal paragraphs (remember that you can use appendices for reference material).

When the data was tweaked, I had to alter the categorizing because there were less folks with higher incomes now in the lower credit card spending areas which were not present before. I had to separate the groups this time not based on income but rather on credit card spending. I simply removed the lowest CCAvg, non-loan holder from the group. This was an individual who luckily had a very large income. Note the differences below:





## 6. INCONSISTENT DATA

Alter the data so that it contains inconsistencies, and show changes in the I/O of the examples in part 3. Explain. Limit: 2 normal paragraphs (remember that you can use appendices for reference material).

Your response goes here.

## 7. BENEFITS

In at most a page and a half (of 12-point text), explain the pros and cons of decision trees *applied to the application you have chosen*. You may build on the work of others but (1) show clearly that you understand this work and (2) observe all plagiarism rules scrupulously, including clear citations.

I think had I introduced additional features for the tree to consider, rather than just two, it would have aided in the simplification of this process overall. However, it is pretty clear that banks who wish to target folks in their marketing campaigns in hopes of expanding their loaning operations, the utilization of this type of program would ultimately be very beneficial. This can be seen in the two charts above. There would be a very low number of false positives/false negatives when using this program. Thus, the banks could look at their preexisting datasets, determine the factor or factors which best determine whether or not an individual may be interested in purchasing a personal loan, and target them in terms of offers and advertisements. Ultimately, it is obvious that this would be incredibly beneficial for institutions seeking to expand these departments. In addition, this could be used for other aspects of retail and investment banking, such as for offers on credit card, savings accounts, etc.

## EVALUATION

