

## Class 8: Tree-based Models and Their Application to Targeted Marketing

Dr Wei Miao

UCL School of Management

October 25, 2023

## Section 1

# Decision Tree

# Introduction to Decision Tree

- A **decision tree** is a flowchart-like tree structure.
- Used in classification and regression.
- Consists of nodes representing decisions and leaves representing outcomes.

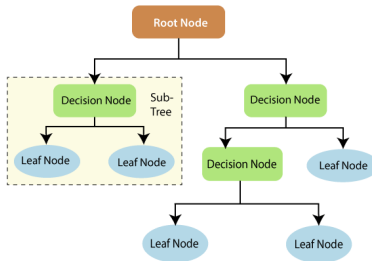


Figure 1: How a decision tree looks like. Source: [Medium](#).

## Example: Predict Customer Response to Marketing Offers

- Tesco made marketing offers to customers in the data, and the variable Response represents whether or not customers responded to our offer.
- **Business objective:** From data\_full, we want to train a decision tree model to predict the outcome variable Response based on Recency and totalspending (for simplicity).
- **Data collection and cleaning:**

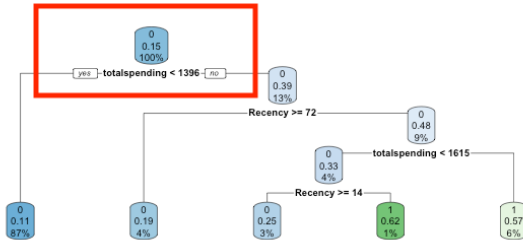
```
1 pacman::p_load(dplyr,modelsummary)
2 data_demo <- read.csv(file = "https://www.dropbox.com/s/a0v38lpydls2emy
3                       header = T)
4 data_purchase <- read.csv(file = "https://www.dropbox.com/s/de435r8zdx
5
6 data_full <- data_purchase %>%
7   left_join(data_demo, by = c('ID' = 'ID')) %>%
8   mutate(totalspending = MntWines + MntFruits +
9         MntMeatProducts + MntFishProducts +
10        MntSweetProducts + MntGoldProds)
```

# Implementation of Decision Tree in R

- Package `rpart` provides efficient implementation of decision trees in R
- Package `rpart.plot` provides nice visualizations of decision trees

```
1 # Load the necessary packages
2 pacman::p_load(rpart,rpart.plot)
3
4 # Below example shows how to train a decision tree
5 tree1 <- rpart(
6   formula = Response ~ Recency + totalspending,
7   data     = data_full,
8   method   = "class" # classification task; or 'anova' for regression
9 )
10
11 # visualize the tree
12 rpart.plot(tree1)
```

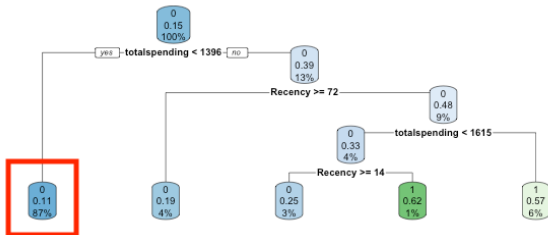
# How Decision Tree Works: Step 1



- ① Decision tree (DT) will try to split customers into 2 groups based on each unique value of each variable, and see which split can lead to customers being most differentiated in terms of Response.
- After this step, DT finds that total spending is the best variable and 1396 is the best cutoff.
  - DT therefore splits customers into 2 groups based on 1396.<sup>1</sup>

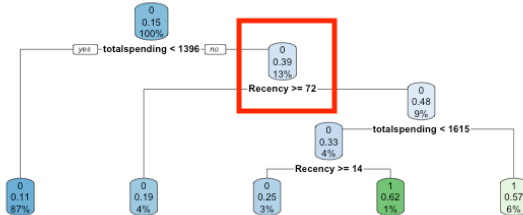
<sup>1</sup>In each node, the 3 numbers are (1) predicted outcome (2) predicted probability of outcome being 1, and (3) percentage of customers in the node

## How Decision Tree Works: Step 2



- ② For customers in the left branch ( $\text{totalspending} < 1396$ ), DT will continue to split based on each unique value of each variable, and see which split can result in the customers to be most different in terms of Response.
- However, DT couldn't find a cutoff that sufficiently differentiate customers, so DT stops in the left branch.

## How Decision Tree Works: Step 3 ...



- ③ For customers in the right branch ( $\text{totalspending} \geq 1396$ ), DT will continue to split based on each unique value of each variable, and see which split can result in the customers to be most different in terms of Response.
  - After this step, DT finds *Recency* is the best variable and 72 is the best cutoff. DT further splits customers into 2 groups.
- ④ This process continues until DT determines that there is no need to further split customers.



# Advantages of Decision Trees

- They are very interpretable.
- Making predictions is fast.
- It's easy to understand what variables are important in making the prediction. The internal nodes (splits) are those variables that most largely reduce the SSE (criteria for split).

## Section 2

# Random Forest

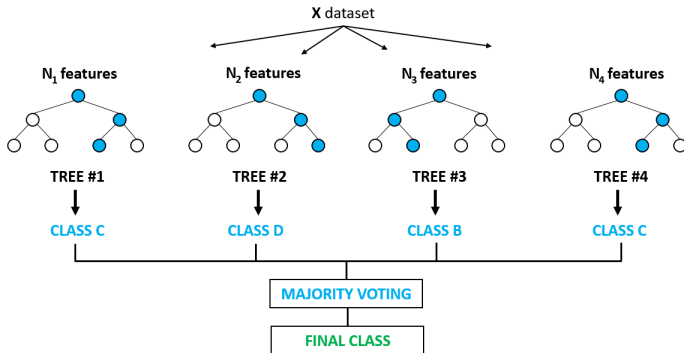
## Disadvantages of Decision Trees

- Single regression trees tend to have high variance (overfitting), resulting in unstable predictions.
- Due to the high variance, single regression trees tend to have poor predictive accuracy.

# Random Forest

- To overcome the overfitting tendency of a single decision tree, random forest has been developed by (Breiman 2001).
  - Instead of using all customers, each tree is grown to a random subsample of customers
  - Instead of using all features for splitting, each tree is grown to a random subset of features

# Visualization of Random Forest



For a new customer,

- Each tree gives a prediction of the outcome
- Random forest takes the average of all trees' predictions as the final prediction

# Implementation of Random Forest in R

- Package ranger provides implementation of random forest in R.
- ranger() is the function in the package to train a random forest; refer to its help function for more details.
- The following code shows how to train a random forest consisting of 500 decision trees, where the outcome variable is mpg, and the predictors are 5 car attribute variables.

```
1  pacman::p_load(ranger)
2  randomforest1 <- ranger(
3      formula    = Response ~ totalspending + Recency,
4      data       = data_full, # dataset to train the model
5      num.trees  = 500, # 500 decision trees
6      seed       = 888, # make sure of replication
7      probability = TRUE
8  )
```

# Make Predictions from Random Forest

- After we train the predictive model, we can use `predict()` function to make predictions
  - The 1st argument is the trained model object
  - The 2nd argument is the dataset to make predictions on

```
1 # Make predictions on the mtcars
2 prediction_rf <- predict(randomforest1,
3                           data = data_full)
4
5 # Because prediction_rf is a list object
6 # Need to use $ to extract the predicted value as a numeric vector
7 prediction_rf$predictions
```

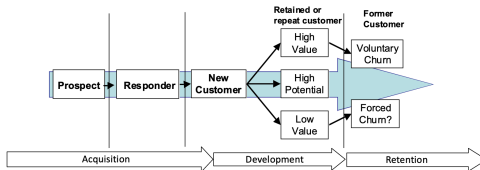
## Section 3

# Improve Marketing Efficiency Using Supervised Learning



# Customer Life Cycle

- Acquisition (Tesco Case Study)
  - Use predictive analytics to target responsive customers to reduce marketing costs
- Development
  - Use predictive analytics to recommend products to customers (personalized recommendation system); for each customer, promote the item with the highest purchase probability
- Retention
  - Use predictive analytics to find valuable customers who are likely to churn and conduct targeted churn management



# Workflow



- 1 Define a business objective: target responsive customers in acquisition stage to reduce customer acquisition costs
- 2 Collect data
- 3 Clean and prepare data
- 4 Analyze data using predictive analytics
- 5 Conduct break-even analyses to show the profitability of the proposed marketing campaign

## After-Class Reading

- (optional) Varian, Hal R. "Big data: New tricks for econometrics." Journal of Economic Perspectives 28, no. 2 (2014): 3-28
- (next week) Predictive Analytics for Tesco
- (recommended) [Decision tree in R](#)
- (recommended) [Random forest in R](#)