**Quiz 5**

*Content: Decision Trees*

*Functions in R: rpart(), rpart.plot().*

Consider again the dataset <bank-sample.csv> given in the lecture of Topic 5. One would want to build a model (a decision tree) which help to predict if a customer will subscribe the bank’s service, based on 8 factors (8 columns): *job, marital, education, default, housing, loan, contact* and *poutcome*.

To answer the questions below, please use the R code of Topic 5 to understand the data and to plot the tree as given in the photo below.

A screenshot of a computer

Description automatically generated

**Question 1** (correct answer in yellow highlight)

The node circled in red color (**no 1789/2000**) in the figure above means

1. The sample size is 2000 but only 1789 are available for data analysis.
2. The sample size is 2000 and 1789 of them belong to category “no” of the response variable “subscribe” (did not subscribe). The rest belongs to other category.
3. The sample size is 2000 but only 1789 of them has information about “poutcome”.
4. There are 1789 among 2000 observations in this tree.

**Question 2** (correct answer in yellow highlight)

The leaf node in the square (no 1763/1942) means

1. Among 1789 customers saying “no” to the subscription of the bank service, the probability that they belong to poutcome = failure, other or unknown is 1763/1942.
2. Among all the 2000 customers, there are 1942 of them has poutcome = failure, other or unknown. Among those 1942 customers, 1763 of them did not subscribe the bank service.

**Question 3** (correct answer in yellow highlight)

The node where the red arrow is pointing (yes 32/58) means

1. If a customer has poutcome = success then without information of any other factors, we can predict that the customer will subscribe the bank service with probability 32/58.
2. Among 2000 customers, there are total of 58 customers has poutcome = success. And among these 58 customers, there are 32 of the subscribed the bank service.
3. Among the 1789 customers that did not subscribe the bank service, there are 58 of them has poutcome = success.

**Question 4** (correct answer in yellow highlight)

The leaf node in the red triangle (yes 8/8) means

1. If a customer has education = primary or unknown, then s/he is predicted to subscribe the bank service with probability 100%.
2. There are 8 customers that has poutcome = success and education = primary or unknown but there is no information if they subscribed the bank service or not.
3. There are 58 customers has poutcome = success in the sample, and there are 8 among those 58 customers has education = primary or unknown that all subscribed the bank service.
4. If a customer has poutcome = success and education = primary or unknown, then we can predict that s/he will subscribe the bank service with probability 100%.

**Question 5** (correct answer in yellow highlight)

The leaf node above the purple arrow (no 26/45) means

1. Among 1789 customers that did not subscribe the bank service, there are 26 of them that have the job = admin, blue-collar, management, retired, services or technician.
2. If a customer that has poutcome = success; and education = secondary or tertiary; and job = admin, blue-collar, management, retired, services or technician; then we can predict that the probability s/he will not subscribe the bank service is 26/45.
3. Among 1789 customers that did not subscribe the bank service, the probability of them that have poutcome = success; and education = secondary or tertiary; and job = admin, blue-collar, management, retired, services or technician is 26/45.
4. Among 2000 customers, there are 58 of them that have poutcome = success; and among those 58, there are 50 of them that have education = secondary or tertiary; and there are 45 among those 50 that have job = admin, blue-collar, management, retired, services or technician; and there are 26 among those 45 that did not subscribe the bank service.