

# Tutorial 5

DSA1101

Introduction to Data Science

October 5, 2018

## Exercise 1. *Decision Trees*

In this week's tutorial, we will revisit the customer churn example presented in week 5 of the course. Customer churn is the loss of clients or customers. Banks, telephone service companies, internet service providers, pay TV companies and insurance firms often use customer churn analysis and customer churn rates as one of their key business metrics.

- (a) The customer churn dataset is available as the CSV file `churn.csv` on the course website. Read in and note the dimension of the dataset.

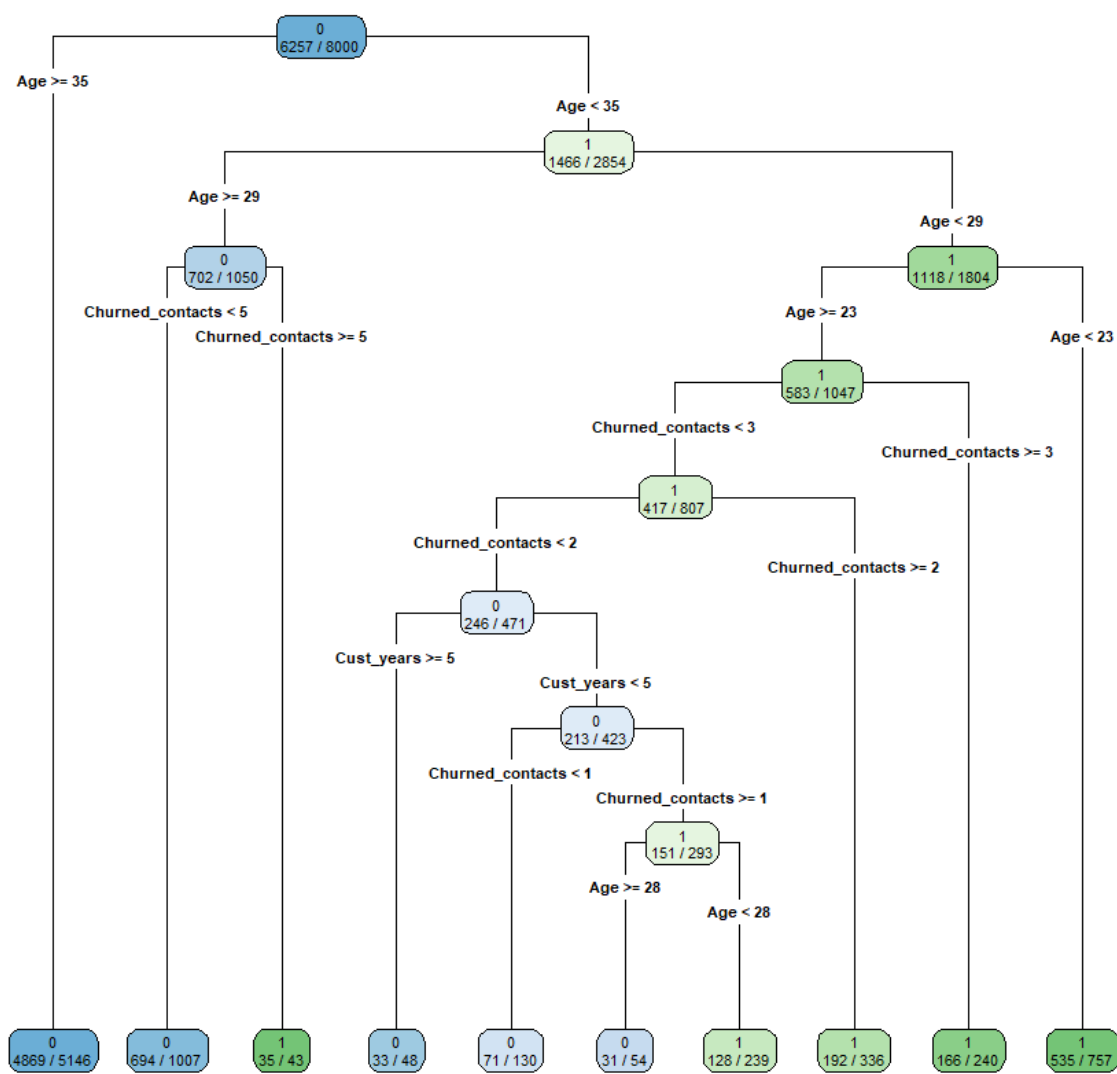
```
1 churn=read.csv("churn.csv")
2
3 head(churn)
4 dim(churn)
```

- (b) Build a decision tree for predicting customer churn, using the feature variables `Age`, `Married`, `Cust_years` and `Churned_contacts`.

```
1 library ("rpart")
2 library ("rpart.plot")
3
4 fit <- rpart ( Churned ~ Age + Married + Cust_years
5 + Churned_contacts,
6 method = "class",
7 data = churn,
8 control =rpart.control(cp=0.002),
9 parms = list(split = 'information'))
```

- (c) Visualize the decision tree built in part (b), using the `rpart.plot` function.

```
1 rpart.plot(fit , type=4, extra=2,
2 clip.right.labs =FALSE , varlen =0, faclen =0)
```



**Exercise 2.** Consider the decision tree in exercise 1 to predict binary variable **Churned**. Use the tree to predict customer churn for the following observations:

	Age	Married	Cust_years	Churned_contacts
2821	26	1	2	2
96	23	1	3	3
5085	56	1	5	2
758	36	1	5	2
487	45	0	2	1
987	28	0	2	2
6061	22	1	3	0
3745	22	0	3	2
4709	60	1	2	1
2769	32	0	3	1

By following the tree paths, the predicted values for Churned are 1, 1, 0, 0, 0, 1, 1, 1, 0, 0