



Decision  
Tree

---

Garry Ka Lok CHU

# Decision Tree

---

Decision tree is used to classify a target variable or to predict the value of a target variable based on the tree model constructed by the training dataset.



# Decision Tree

Can Josh Anderson score in Game 1 against Toronto Maple Leafs?



Consider the following training set:

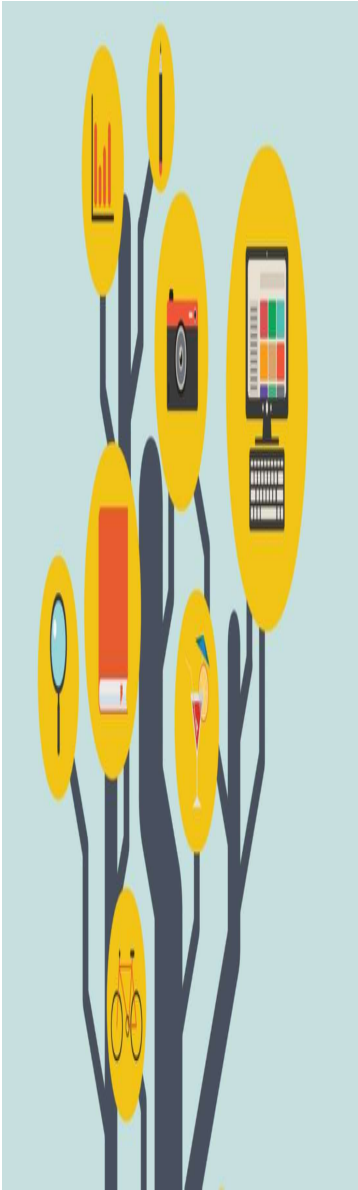
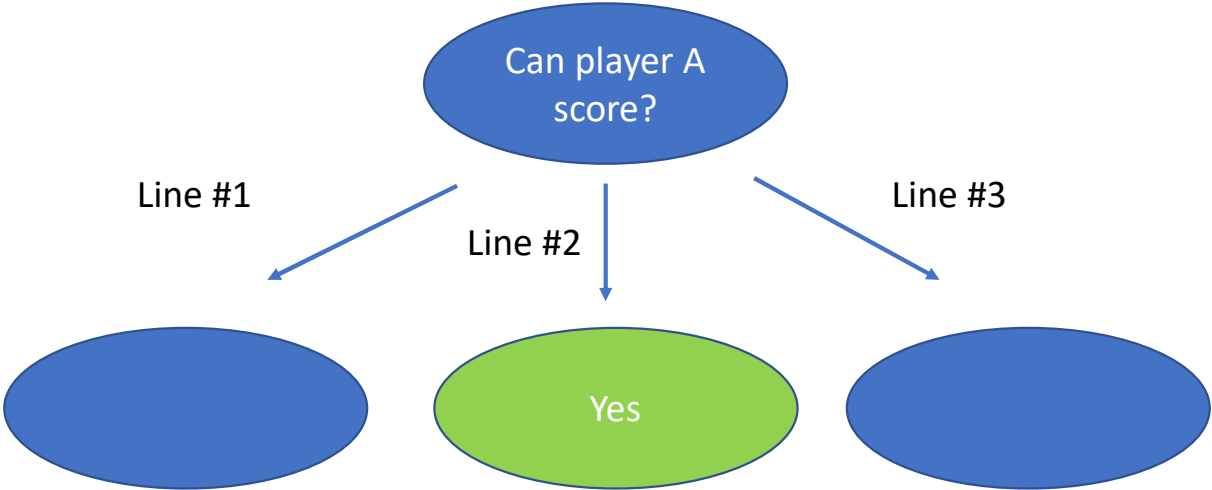
Game	Which line is player A playing?	Home/Away game	Opposing team's goalie GAA is below 2.00	Player A can score
1	1	Away	No	No
2	1	Away	Yes	No
3	2	Home	No	Yes
4	3	Home	No	Yes
5	3	Away	No	Yes
6	3	Away	Yes	No
7	2	Away	Yes	Yes
8	1	Away	No	No
9	1	Home	No	Yes
10	3	Away	No	Yes
11	1	Home	Yes	Yes
12	2	Home	Yes	Yes
13	2	Away	No	Yes
14	3	Home	Yes	No

# Decision Tree

Game	Which line is player A playing?	Home/Away game	Opposing team's goalie GAA is below 2.00	Player A can score
1	1	Away	No	No
2	1	Away	Yes	No
3	2	Home	No	Yes
4	3	Home	No	Yes
5	3	Away	No	Yes
6	3	Away	Yes	No
7	2	Away	Yes	Yes
8	1	Away	No	No
9	1	Home	No	Yes
10	3	Away	No	Yes
11	1	Home	Yes	Yes
12	2	Home	Yes	Yes
13	2	Away	No	Yes
14	3	Home	Yes	No

First, we consider any subset that is pure: one particular factor gives one outcome only. In this case, Player A playing in line #2 can score.

# Decision Tree

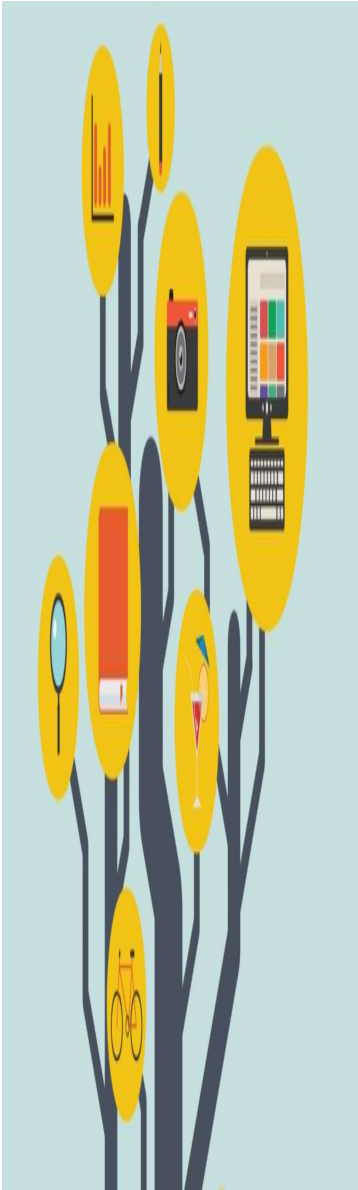
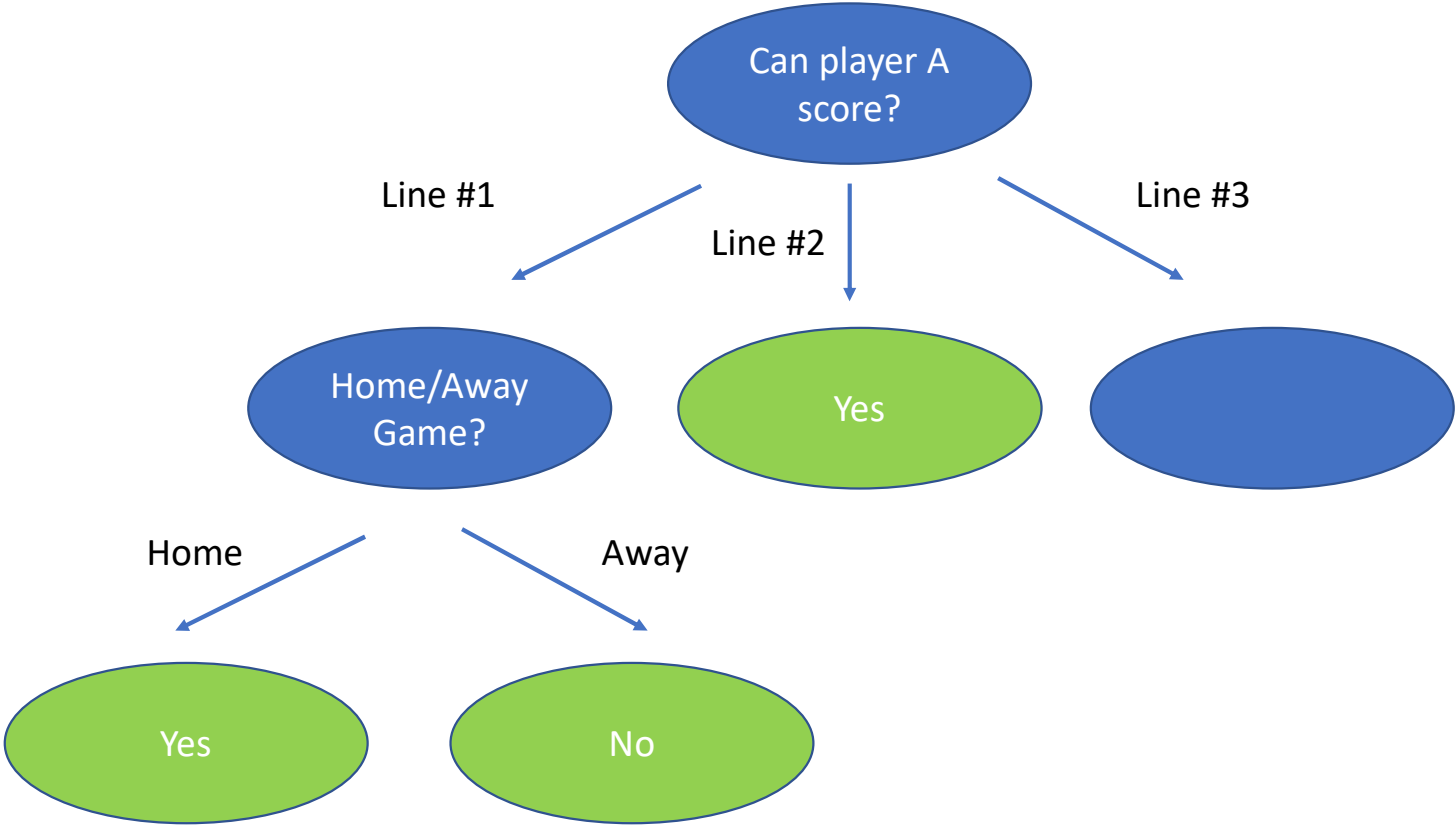


# Decision Tree

Game	Which line is player A playing?	Home/Away game	Opposing team's goalie GAA is below 2.00	Player A can score
1	1	Away	No	No
8	1	Away	No	No
2	1	Away	Yes	No
9	1	Home	No	Yes
11	1	Home	Yes	Yes
13	2	Away	No	Yes
7	2	Away	Yes	Yes
3	2	Home	No	Yes
12	2	Home	Yes	Yes
5	3	Away	No	Yes
10	3	Away	No	Yes
4	3	Home	No	Yes
6	3	Away	Yes	No
14	3	Home	Yes	No

If possible, continue to find any pure subset. In this example, if Player A is playing in line #1, he can score when it is a home game.

# Decision Tree



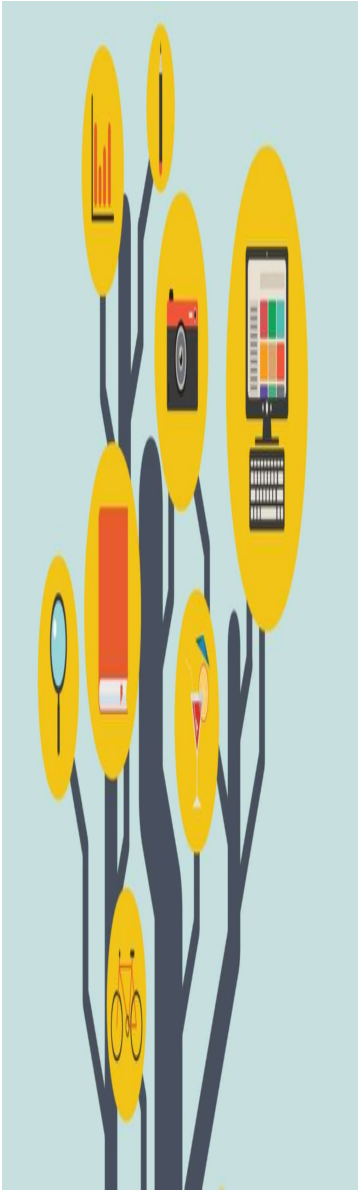
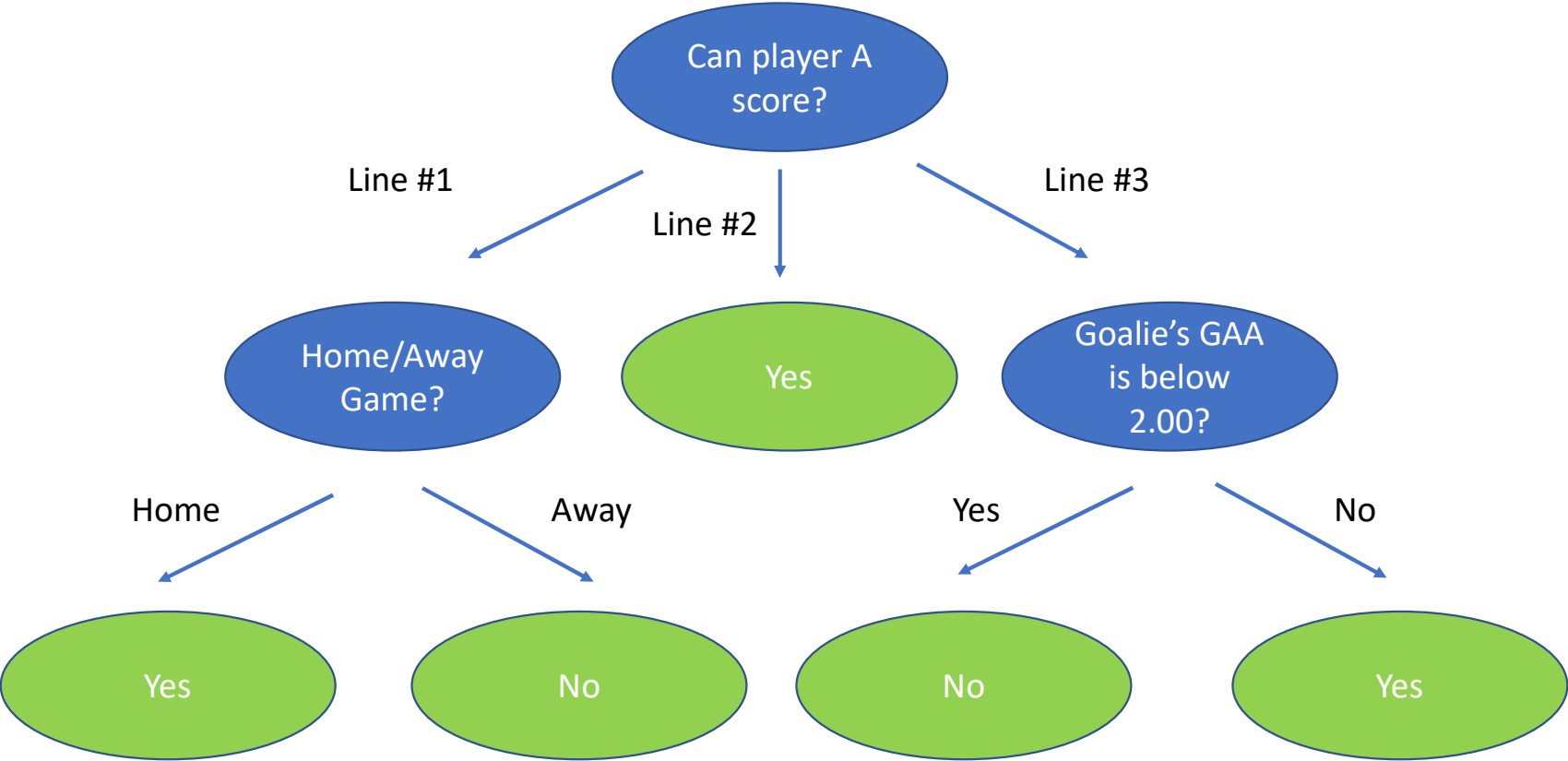


# Decision Tree

Game	Which line is player A playing?	Home/Away game	Opposing team's goalie GAA is below 2.00	Player A can score
1	1	Away	No	No
8	1	Away	No	No
2	1	Away	Yes	No
9	1	Home	No	Yes
11	1	Home	Yes	Yes
13	2	Away	No	Yes
7	2	Away	Yes	Yes
3	2	Home	No	Yes
12	2	Home	Yes	Yes
5	3	Away	No	Yes
10	3	Away	No	Yes
4	3	Home	No	Yes
6	3	Away	Yes	No
14	3	Home	Yes	No

Moreover, if Player A is playing in line #3, he can score when the opposing team's goalie GAA is not below 2.00.

# Decision Tree



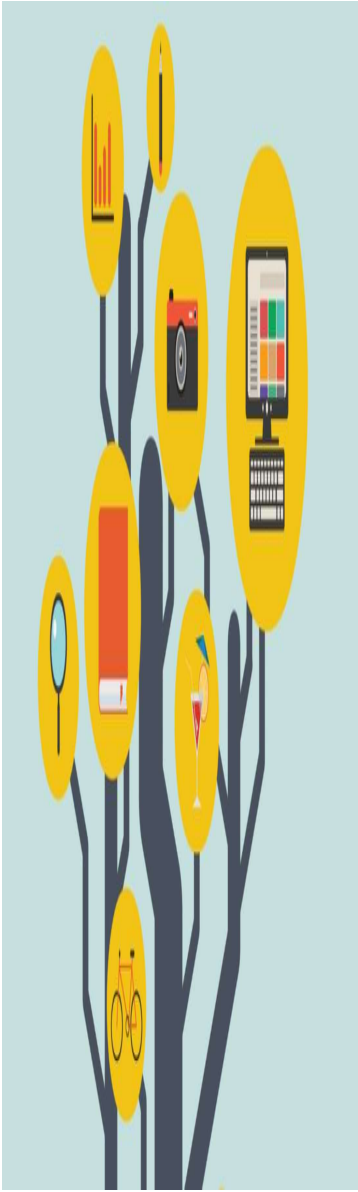
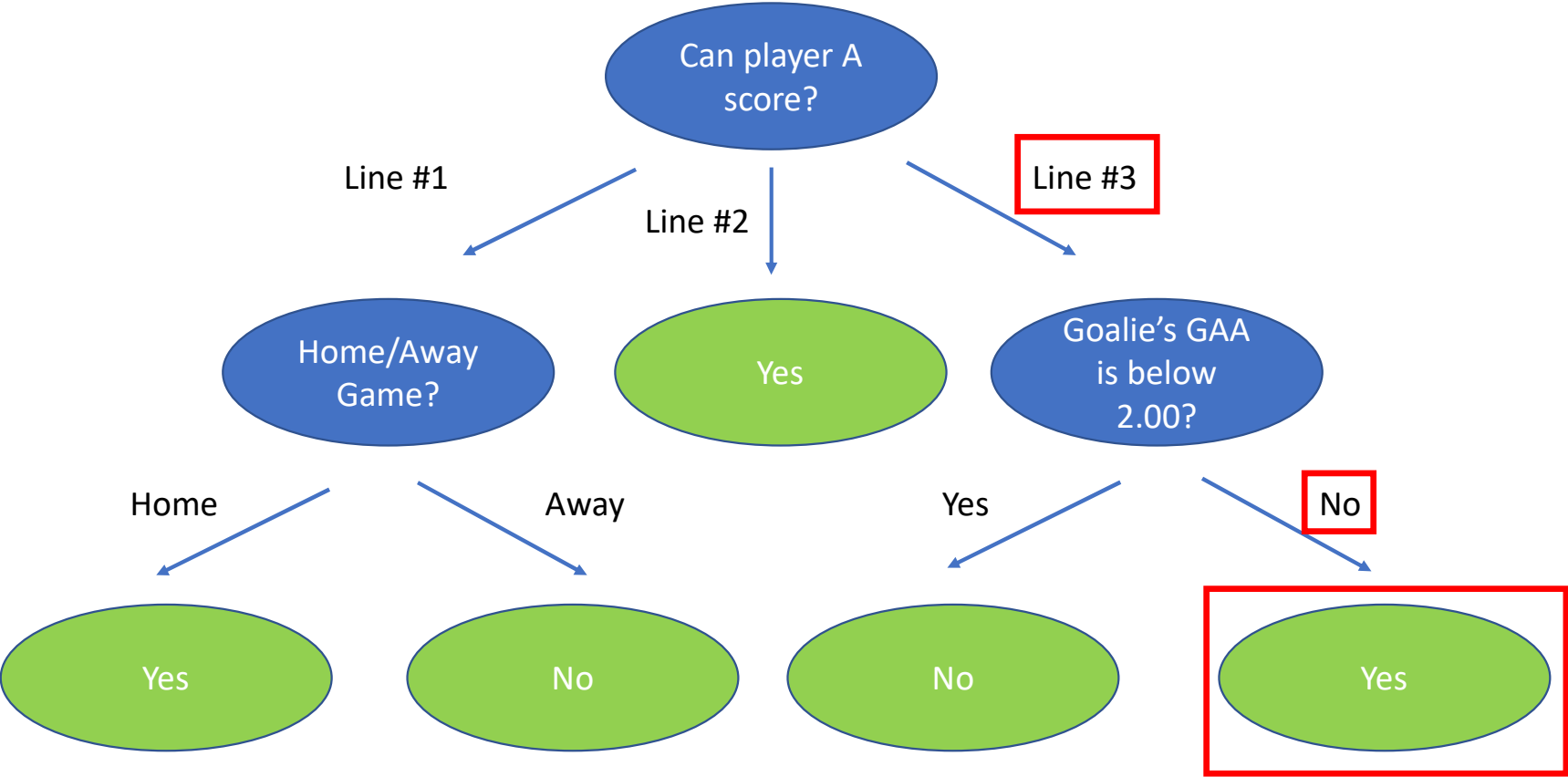
# Decision Tree

Can Josh Anderson score in Game 1 against Toronto Maple Leafs?

- Away Game on May 20
- Leafs goalie Campbell: GAA 2.15
- Anderson: line 3



# Decision Tree



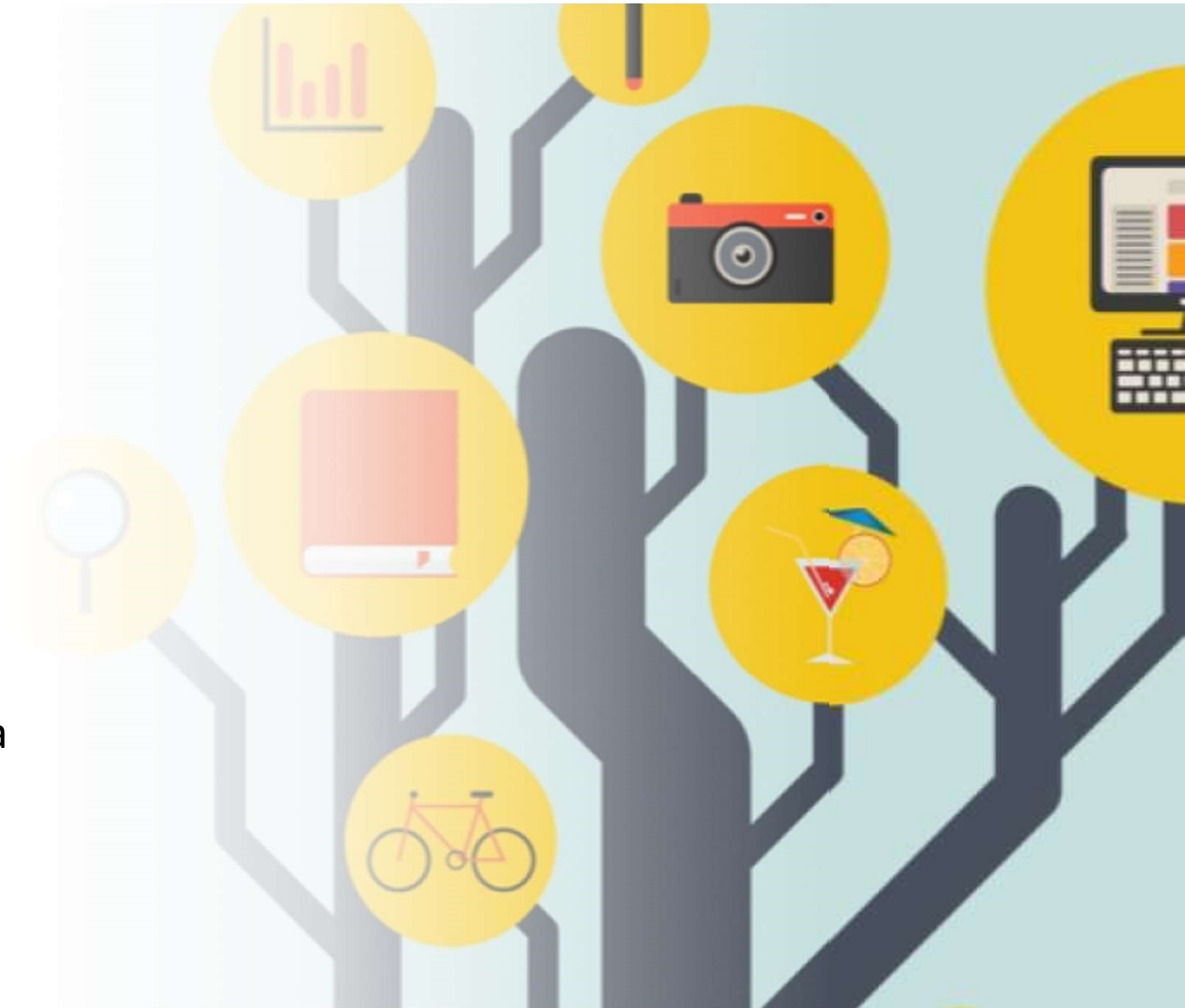
---

# Decision Tree

---

## Further Questions:

- Can we separate the subsets if they are impure?
- Is there any method to measure the impurity of a set?
- May we use decision tree with numerical data?



---

# Decision Tree

---

## Further Topics:

- Terminology in Decision Trees
- Random Forests

