

Random Forests

Learning & Softcomputing

Alexander Belzer
und
Maximilian Wendt

Advisor:
Prof. Ulrich Hoffmann



- 1. Ensemble Learning
- 2. Decision Tree
- 3. How Random Forests Work
- 4. Advantages



Is one expert smarter than 1000 educated people?



1. Ensemble Learning

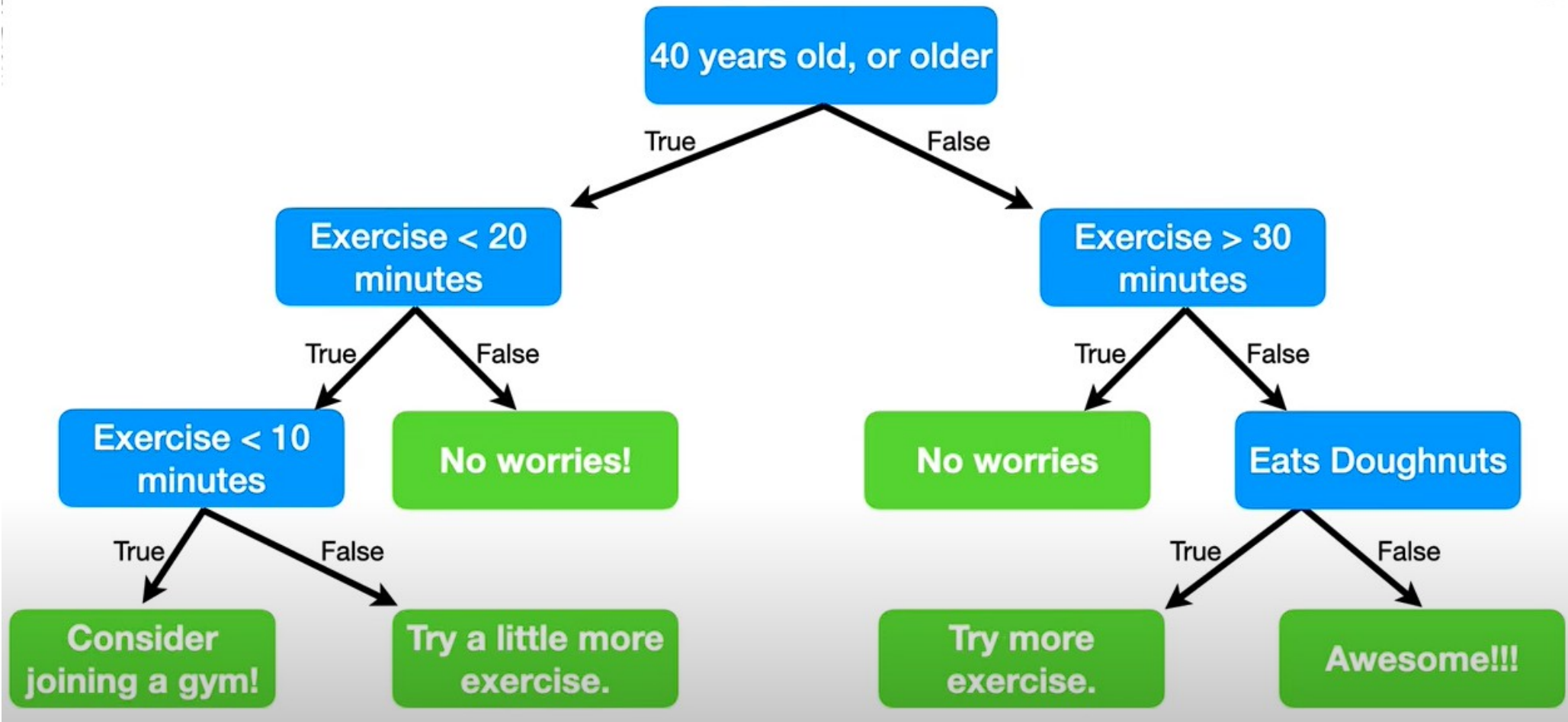
→ 2. Decision Tree

2.1 Building the Tree

3. How Random Forests Work

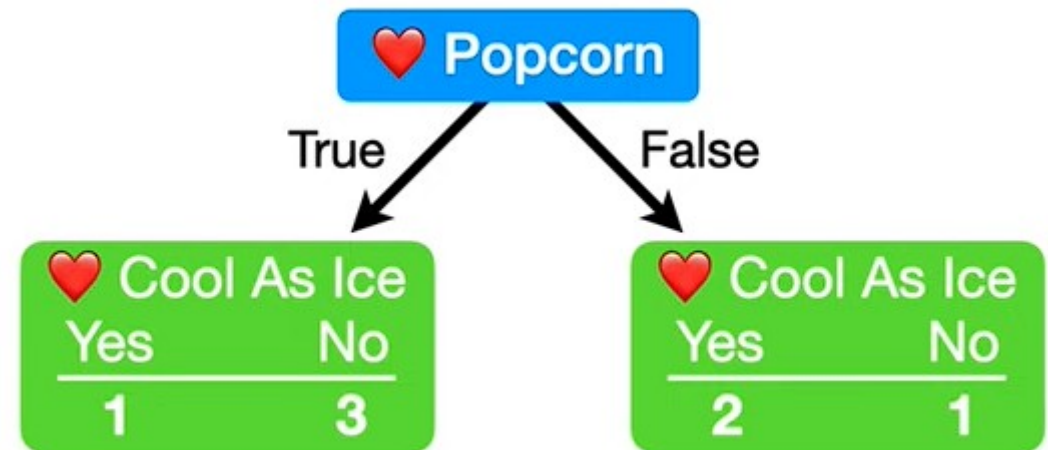
4. Advantages

2 Decision Tree



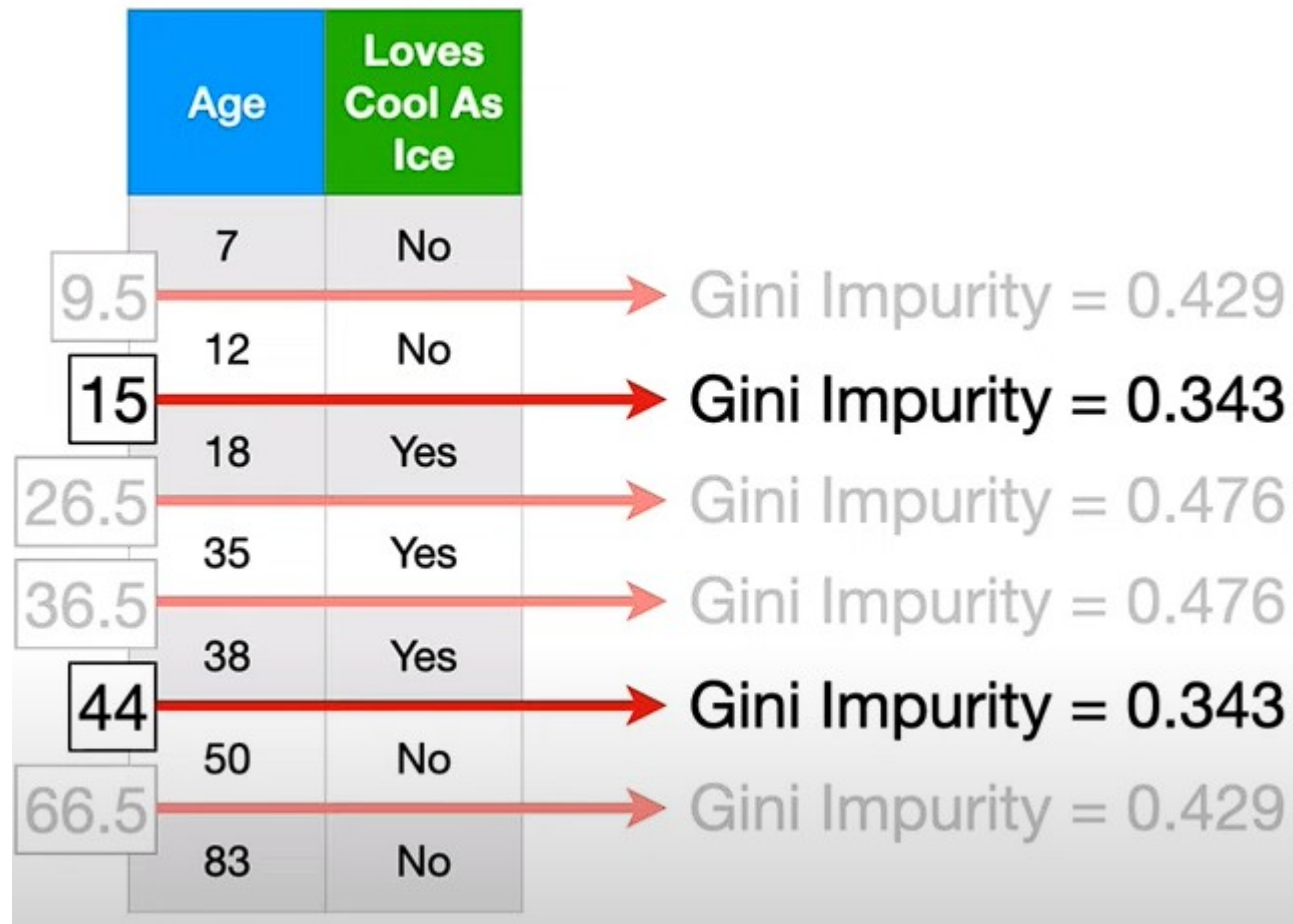
2.1 Building the Tree

Loves Popcorn	Loves Soda	Age	Loves Cool As Ice
Yes	Yes	7	No
Yes	No	12	No
No	Yes	18	Yes
No	Yes	35	Yes
Yes	Yes	38	Yes
Yes	No	50	No
No	No	83	No



$$\text{Gini Impurity} = 1 - P(\text{Yes})^2 - P(\text{No})^2$$

2.1 Building the Tree





1. Ensemble Learning

2. Decision Tree

→ 3. How Random Forests Work

3.1 Build the Forest

3.2 Estimate Accuracy

3.3 Adjust Accuracy

4. Advantages

3.1 Build the Forest



Index	Blood Pressure	Smokes cigarettes	Obese	Cancer in the Family	Has Cancer
1	121/82	Yes	Yes	No	Yes
2	126/84	Yes	No	No	No
3	135/89	Yes	No	Yes	Yes
4	124/82	Yes	Yes	Yes	Yes
5	122/83	No	Yes	No	No
6	145/92	No	No	No	No

3.1 Build The Forest



Original Data

Index	Blood Pressure	Smokes cigarettes	Obese	Cancer in the Family	Has Cancer
1	121/82	Yes	Yes	No	Yes
2	126/84	Yes	No	No	No
3	135/89	Yes	No	Yes	Yes
4	124/82	Yes	Yes	Yes	Yes
5	122/83	No	Yes	No	No
6	145/92	No	No	No	No

Bootstrapped Sample

Index	Blood Pressure	Smokes cigarettes	Obese	Cancer in the Family	Has Cancer
1	121/82	Yes	Yes	No	Yes
3	135/89	Yes	No	Yes	Yes
3	135/89	Yes	No	Yes	Yes
5	122/83	No	Yes	No	No
1	122/82	Yes	Yes	No	Yes
2	126/88	Yes	No	No	No

3.1 Build The Forest



Index		Smoke s cigarett es		Cancer in the Family	Has Cancer
1		Yes		No	Yes
3		Yes		Yes	Yes
3		Yes		Yes	Yes
5		No		No	No
1		Yes		No	Yes
2		Yes		No	No

3.1 Build The Forest



Index	Blood Pressure	Smokes cigarettes	Obese	Cancer in the Family	Has Cancer
1	121/82	Yes	Yes	No	Yes
3	135/89	Yes	No	Yes	Yes
3	135/89	Yes	No	Yes	Yes
5	122/83	No	Yes	No	No
1	122/82	Yes	Yes	No	Yes
2	126/48	Yes	No	No	No

3.1 Build The Forest



Index	Blood Pressure	Smoke cigarettes	Obese	Cancer in the Family	Has Cancer
1	121/82	Yes	Yes	No	Yes
3	135/89	Yes	No	Yes	Yes
3	135/89	Yes	No	Yes	Yes
5	122/83	No	Yes	No	No
1	122/82	Yes	Yes	No	Yes
2	126/48	Yes	No	No	No

3.1 Build The Forest



Index	Blood Pressure	Smokes cigarettes	Obese	Cancer in the Family	Has Cancer
1	121/82	Yes	Yes	No	Yes
3	135/89	Yes	No	Yes	Yes
3	135/89	Yes	No	Yes	Yes
5	122/83	No	Yes	No	No
1	122/82	Yes	Yes	No	Yes
2	126/48	Yes	No	No	No



Use the Out-Of-Bag (OOB) Data!



Correlation and Strength are adjusted by the amount of variables randomly chosen for the Decision Tree's.



1. Ensemble Learning
2. Decision Tree
3. How Random Forests Work
- 4. Advantages



Speed and Accuracy