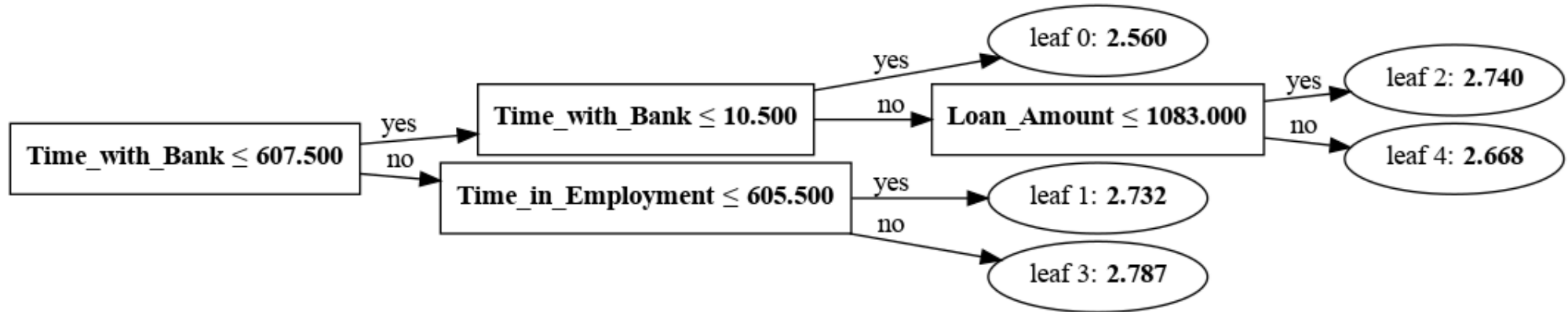


# Machine Learning: Gradient Boosting Decision Tree

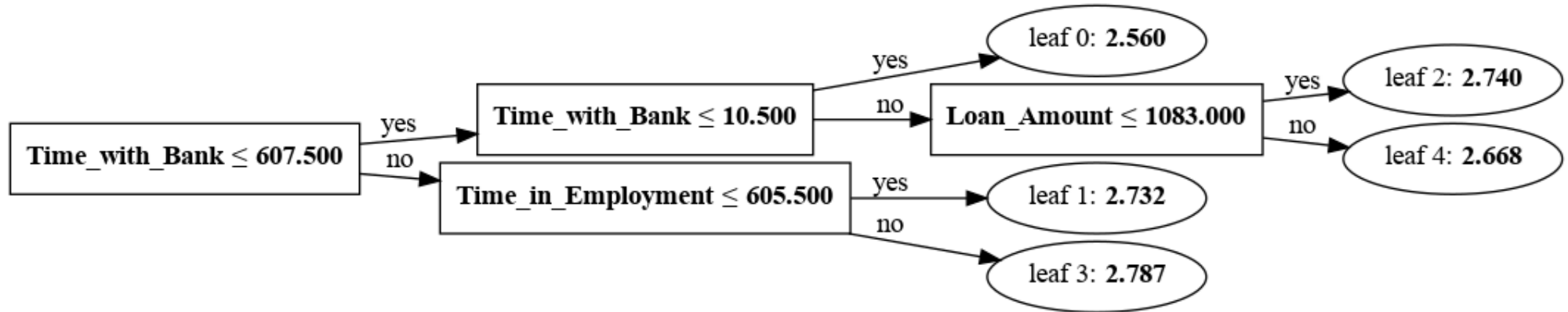
# 1. Decision Tree

•Decision Tree (Classification Tree) is a Machine Learning algorithm suitable for classification problems (eg. Predicting a „Good“ or „Bad“ outcome in Credit Risk Modelling)



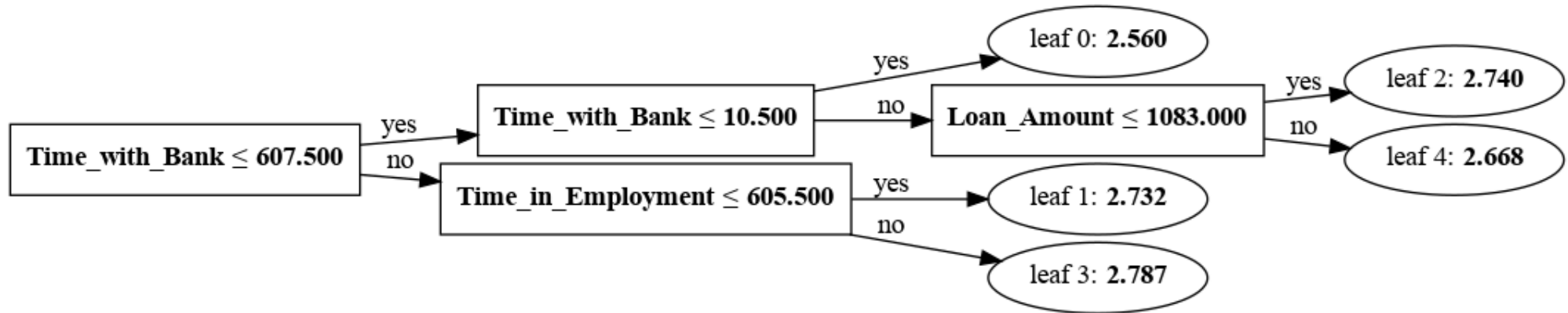
# 1. Decision Tree

•Decision Trees consist of: Root Node, Internal Nodes and Leaf Nodes



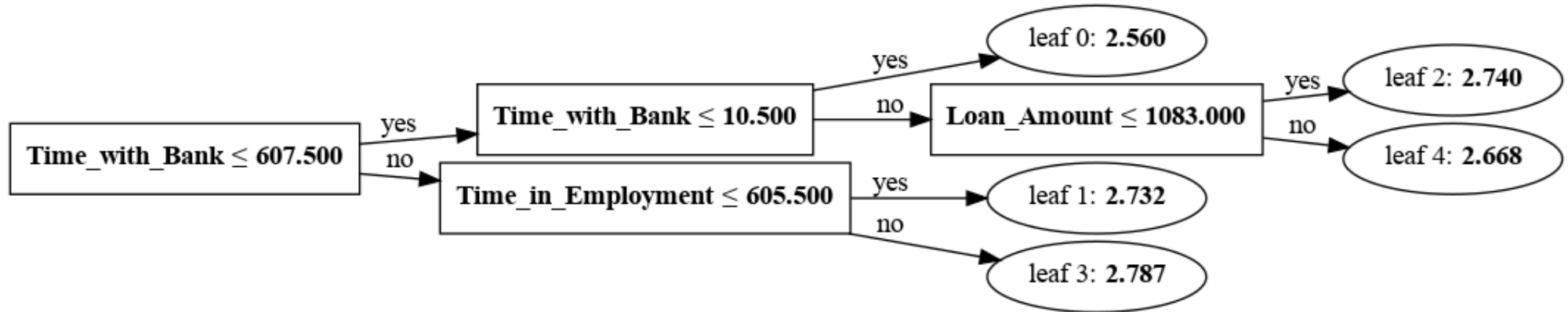
# 1. Decision Tree

• Predictive variables are used to segment the data. The leaf nodes represent a segment of the data that is used to calculate the probability of the outcome (eg „Good“ or „Bad“)



# 1. Decision Tree

•The values of the Leaf Nodes represent:  $\log(P(\text{Good}) / P(\text{Bad}))$



# 1. Decision Tree

How are predictive variables used to segment the data?

- Decision Trees segment the data by using predictive variables. Predictive variables with high predictive power are used early on. Predictive power can be measured by **Gini Impurity**.

- $\text{Gini Impurity (Leaf)} = 1 - P^2(\text{Good}) - P^2(\text{Bad})$

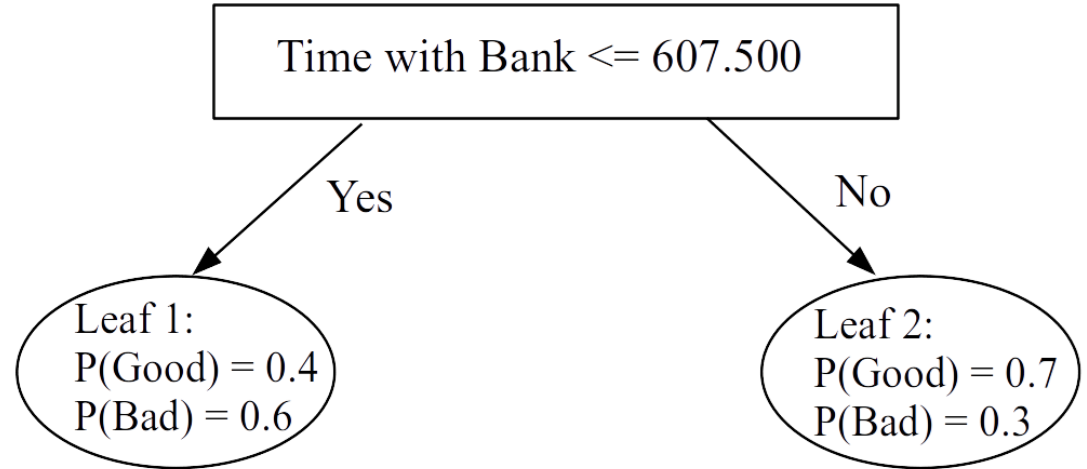
- $\text{Gini Impurity (Total)} = \text{sum (Weight * Gini Impurity (Leaf))} / \text{Number of Leaves}$

# 1. Decision Tree

Gini Impurity – Example:

•Gini Impurity (Leaf 1) = 0.48

•Gini Impurity (Leaf 2) = 0.42



•Assuming we have 100 elements in

each Leaf:

# 1. Decision Tree

How are the splits of predictive variables determined?

• Numerical variables are sorted in ascending order. The average of each pair is calculated and used as a test split. The Gini Impurity is calculated for each possible test split and the one with the lowest Gini impurity is used in the Decision Tree.

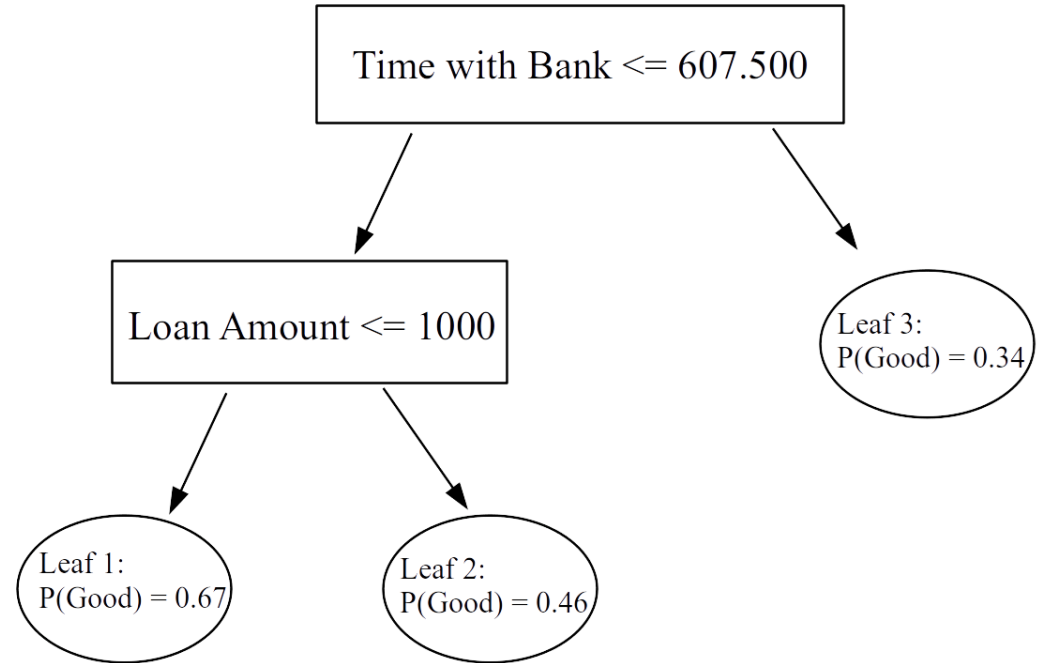
Gross Annual Income	Average Value	Gini Impurity
20.000	22.500	0.273
25.000		
30.000	27.500	0.302



# 1. Decision Tree

•When calculating further splits, the existing splits are considered as well.

•The Gini Impurity of the Loan Amount split is calculated only on the population that has Time with Bank  $\leq 607.500$



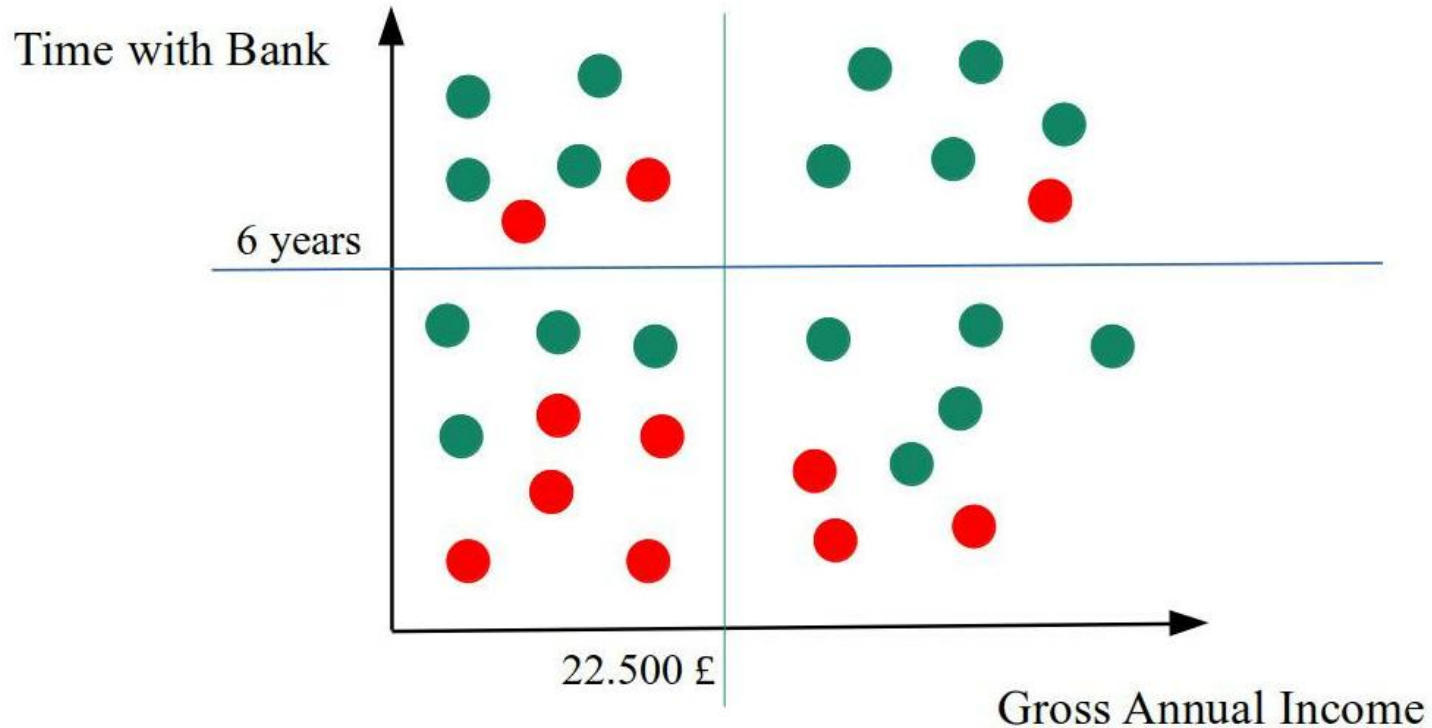
# 1. Decision Tree

How much further should Decision Trees split the data?

- The level of segmentation of the population is determined by the hyperparameters of the model:
- Maximum Depth – how many splits can a Decision Tree make before coming to a prediction
- Minimum Child Sample – how many elements at least are required in each leaf
- Number of Leaves – how many segments can the Decision Tree split the data into

# 1. Decision Tree

How is data segmented – a different viewpoint:



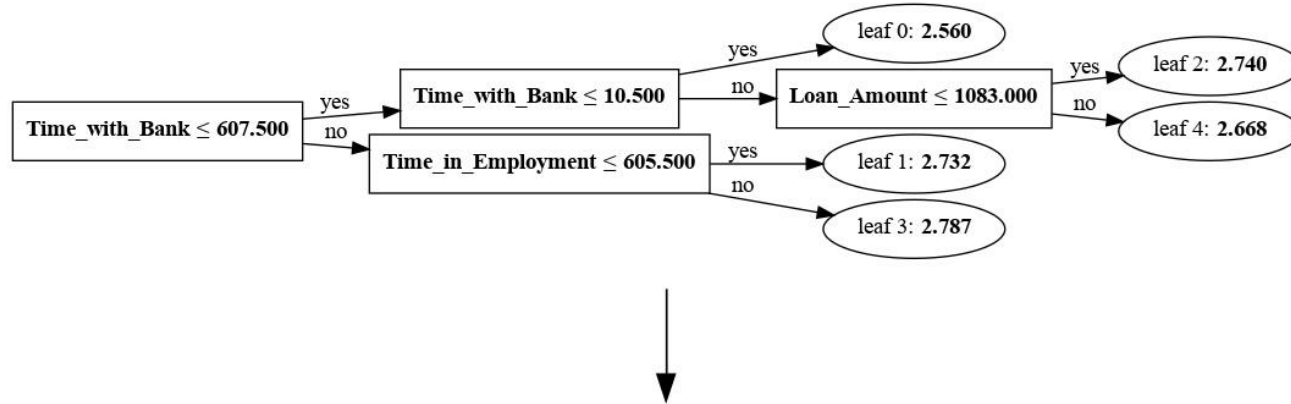
## 2. Gradient Boosting

The Gradient Boosting Decision Tree algorithm uses multiple Decision Trees for the prediction of the outcome.

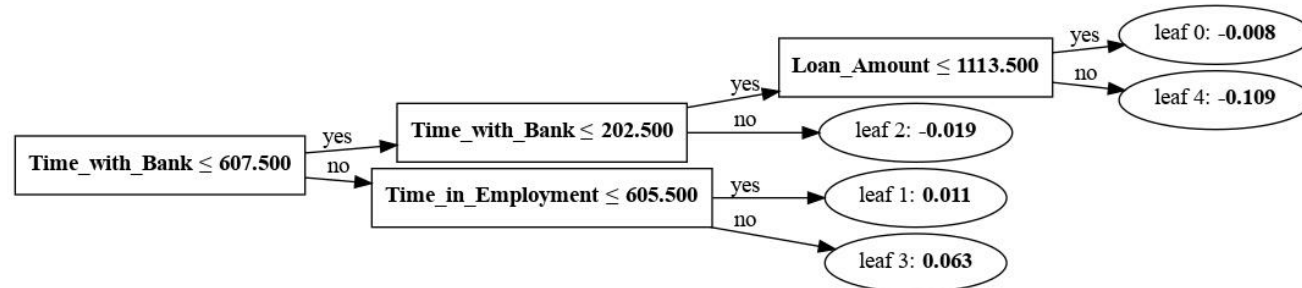
- The classification is performed by multiple Decision Trees in sequential order.
- Each next Decision Tree takes the outcome of the previous Decision Tree into consideration and aims to reduce the overall error rate.
- The total number of Decision Trees used in Gradient Boosting is also a hyperparameter of the Gradient Boosting model.

## 2. Gradient Boosting

Decision Tree 1



Decision Tree 2



### 3. References

1. J. Starmer (2021): Decision and Classification Trees

Link: [https://www.youtube.com/watch?v=\\_L39rN6gz7Y](https://www.youtube.com/watch?v=_L39rN6gz7Y)

2. J. Starmer (2019): Gradient Boost Part 3 (of 4): Classification

Link: <https://www.youtube.com/watch?v=jxuNLH5dXCs>

3. J. Starmer (2019): Gradient Boost Part 4 (of 4): Classification Details

Link: <https://www.youtube.com/watch?v=StWY5QWMXCw>