

Proseminar  
*Advanced topics in  
machine learning*

*Bagging, Boosting, and Ensemble  
Learning*

Author

*Florian Zager*

Matrikelnummer: *2489018*

Date:

Karlsruhe, December 26, 2023

## **Abstract — Zusammenfassung**

Mandatory. Short summary of the report.

# Contents

<b>1</b>	<b>Introduction</b>	<b>1</b>
<b>2</b>	<b>Ensemble Learning</b>	<b>1</b>
2.1	Bagging . . . . .	2
2.2	Random Forest . . . . .	2
2.3	Out-of-bag . . . . .	2
2.4	Boosting . . . . .	2
2.5	Gradient Boosting . . . . .	2
2.6	Extreme Gradient Boosting . . . . .	2
<b>3</b>	<b>Examples</b>	<b>2</b>
3.1	Example 1 . . . . .	2
3.2	Example 2 . . . . .	2
<b>4</b>	<b>Summary and conclusion</b>	<b>2</b>

# 1 Introduction

Mandatory. Questions like: What is the topic of this work, what's the broader context (topic of the proseminar), why is it relevant?

- History of ensemble learning + papers of people "inventing" it
- Goal of the report
- learning more about bagging + boosting
- get to know most popular types of both methods
- learn how to practically use them
- when to use which technique
- 

# 2 Ensemble Learning

Ensemble learning is an advanced machine learning approach that combines the strengths of multiple smaller learning algorithms to improve predictive performance. The concept behind ensemble learning is analogous to the "wisdom of crowds". Which describes, that a crowd, on average, makes collectively better decisions, than any single member of it.

Just as a diverse group of people can provide a more accurate collective decision than an individual, in ensemble learning, a combination of learning algorithms often predicts more accurately than an individual learning algorithm. This approach is based on the principle that a diverse set of learning algorithms can capture different patterns or trends in the data, leading to more robust and accurate predictions.

To be more precise, ensemble methods use multiple smaller learning algorithms, which specialize in small aspects of the problem. However, by combining these algorithms, the ensemble often achieves better predictive performance than the used algorithm could achieve alone because they complement each others strengths and weaknesses.

So the goal of ensemble learning is to achieve a better predictive performance. Nevertheless, it comes at the cost of increased computational resources for training as well as prediction and storage.

Overall, there are many different ensemble methods, such as bagging and boosting, which we will go into more detail in this report. However, there are many more like stacking and blending.

- Whats ensemble learning?
- wisdom of crowds

## **2.1 Bagging**

- Whats the idea behind it?
- How to train bagging? (+ graphic)
- How does the prediction work? (+ graphic)
- when to use it
- advantages and challenges

## **2.2 Random Forest**

- difference to bagging

## **2.3 Out-of-bag**

- explain

## **2.4 Boosting**

- Whats the idea behind it?
- How to train boosting? (+ graphic)
- How does the prediction work? (+ graphic)
- when to use it
- advantages and challenges of using it

## **2.5 Gradient Boosting**

- difference to boosting

## **2.6 Extreme Gradient Boosting**

- difference to gradient boosting

## **3 Examples**

### **3.1 Example 1**

### **3.2 Example 2**

## **4 Summary and conclusion**

Mandatory. Short summary of the most important aspects of the report. If possible: What are open challenges?

- Bagging vs. Boosting - whats the difference?

## References