

# **-Machine Learning: Boosting-Algorithmen-**

## Seminararbeit

Student:	David Erdös	67906
Universität:	Hochschule Karlsruhe – Technik und Wirtschaft	
Studiengang:	Informatik Bachelor	
Semester:	Wintersemester 2023	
Dozent:	Prof. Dr. Baier	
Bearbeitet am:	1. Dezember 2023	

# Inhaltsverzeichnis

<b>1</b>	<b>Einleitung</b>	<b>1</b>
1.1	Motivation und Zielsetzung . . . . .	1
1.2	Struktur der Arbeit . . . . .	1
1.3	Ziel und Umfang der Arbeit . . . . .	1
<b>2</b>	<b>Grundlagen des Machine Learning (3-4 Seiten)</b>	<b>2</b>
2.1	Modern Approaches in Machine Learning . . . . .	2
2.2	Role of Boosting Algorithms in ML . . . . .	2
2.3	Boosting Algorithms in Tabular Data Analysis . . . . .	2
<b>3</b>	<b>Boosting</b>	<b>3</b>
3.1	Was ist Boosting? . . . . .	3
3.2	Wie funktioniert Boosting? . . . . .	3
3.3	Rolle der Boosting Algorithmen in ML . . . . .	4
3.4	Boosting Algorithmen in Tabellendaten Analyse . . . . .	4
<b>4</b>	<b>AdaBoost (5-6 Seiten)</b>	<b>5</b>
4.1	Theoretische Grundlagen . . . . .	5
4.2	Algorithmus-Struktur und Funktionsweise . . . . .	5
4.3	Beispielanwendung mit Erläuterung . . . . .	5
<b>5</b>	<b>Gradient Boosting (5-6 Seiten)</b>	<b>6</b>
5.1	Theoretische Grundlagen . . . . .	6
5.2	Algorithmus-Struktur und Funktionsweise . . . . .	6
5.3	Beispielanwendung mit Erläuterung . . . . .	6
<b>6</b>	<b>Vergleich von AdaBoost und Gradient Boosting (4-5 Seiten)</b>	<b>7</b>
6.1	Gemeinsamkeiten und Unterschiede . . . . .	7
6.2	Performance-Analyse in Benchmarks . . . . .	7
6.3	Anwendungsbeispiele und Fallstudien . . . . .	7
<b>7</b>	<b>Aktuelle Trends und Entwicklungen (2-3 Seiten)</b>	<b>8</b>
7.1	Neueste Forschungsergebnisse . . . . .	8
7.2	Zukünftige Potenziale von Boosting-Algorithmen . . . . .	8
<b>8</b>	<b>Fazit und Ausblick (2-3 Seiten)</b>	<b>9</b>
8.1	Zusammenfassung der Erkenntnisse . . . . .	9
8.2	Reflexion über die Bedeutung für die Praxis . . . . .	9
8.3	Ausblick auf zukünftige Forschungsthemen . . . . .	9
	<b>Literaturverzeichnis</b>	<b>11</b>
	<b>Abbildungsverzeichnis</b>	<b>13</b>



# 1 Einleitung

In der dynamischen Welt des maschinellen Lernens (ML) haben sich Boosting-Algorithmen als revolutionär erwiesen, indem sie beeindruckende Leistungen in einer Vielzahl von Anwendungen erbringen [1, Kapitel 1.1.2]. Besonders bei tabellarischen Datensätzen erzielen sie in fast allen Bereichen die besten Ergebnisse.

In dieser Seminararbeit liegt der Schwerpunkt auf AdaBoost und GradientBoosting, die gemäß Géron [2, S. 192] zu den bekanntesten Vertretern der Boosting-Familie gehören. Die Arbeit wird die beiden Algorithmen detailliert untersuchen und verständlich darstellen.

## 1.1 Motivation und Zielsetzung

Die Motivation für diese Arbeit liegt in der zunehmenden Bedeutung, die Boosting-Algorithmen in der modernen Datenanalyse zukommt. Insbesondere in Situationen, in denen tabellenartige Datensätze verarbeitet werden, haben sich diese Algorithmen als besonders effektiv erwiesen. Dies spiegelt sich auch in ihrer Beliebtheit in Data-Science-Wettbewerben und realen Anwendungsbeispielen wider. Das Ziel dieser Seminararbeit ist es, ein tiefes Verständnis für die Funktionsweise von AdaBoost und GradientBoosting zu entwickeln und deren Einsatzmöglichkeiten anhand konkreter Beispiele zu demonstrieren. Hierbei wird besonderer Wert darauf gelegt, die zugrundeliegenden Konzepte in einer für Bachelorstudenten der Informatik im 6. Semester verständlichen Form darzulegen.

## 1.2 Struktur der Arbeit

Die Seminararbeit gliedert sich in mehrere Abschnitte, die schrittweise aufeinander aufbauen. Zunächst wird ein Überblick über die Grundlagen des maschinellen Lernens gegeben, gefolgt von einer detaillierten Betrachtung der Boosting-Algorithmen, insbesondere AdaBoost und GradientBoosting. Anschließend erfolgt ein Vergleich dieser beiden Methoden, wobei deren Stärken und Schwächen in verschiedenen Anwendungsszenarien beleuchtet werden. Abschließend wird ein Ausblick auf zukünftige Entwicklungen und mögliche Forschungsrichtungen im Bereich des Boostings gegeben.

[3, text] [1, text] [2, text] [4, text] [5, text]

## 1.3 Ziel und Umfang der Arbeit

## **2 Grundlagen des Machine Learning (3-4 Seiten)**

### **2.1 Modern Approaches in Machine Learning**

Überblick über aktuelle Trends und Innovationen im Machine Learning Vorstellung fortgeschrittener Techniken und Methoden Diskussion über die Bedeutung von Deep Learning und künstlichen neuronalen Netzen

### **2.2 Role of Boosting Algorithms in ML**

Einführung in Boosting-Algorithmen und ihre Relevanz Spezifische Betrachtung von AdaBoost und Gradient Boosting Vergleich von Boosting-Algorithmen mit anderen fortgeschrittenen Methoden

### **2.3 Boosting Algorithms in Tabular Data Analysis**

Bedeutung von tabellenartigen Datensätzen in fortgeschrittenen ML-Anwendungen Analyse, wie AdaBoost und Gradient Boosting bei tabellenartigen Daten effektiv sind Fallstudien und Beispiele aus der Praxis, die den Einsatz dieser Algorithmen zeigen

## 3 Boosting

### 3.1 Was ist Boosting?

Boosting ist einer der bekanntesten und meistgenutzten Algorithmen im Bereich Machine Learning. **Definition:** ‘Der Begriff “Boosting” bezieht sich auf eine Familie von Algorithmen, die weak learners (schwache Lerner) in strong learners (starke Lerner) umwandeln.’[6]

Diese Definition ist leicht an einem Beispiel veranschaulichbar: Wie würde man erkennen, ob es jetzt gerade regnet? Folgende Kriterien wären nützlich:

- Ist der Boden nass?
- Sind Wolken am Himmel zu sehen?
- Gibt es eine hohe Luftfeuchtigkeit?
- Gibt es Personen, die einen Regenschirm an sich tragen?
- Liegt die Außentemperatur über 0 Grad Celsius?

Diese Regeln können mit hoher Zuverlässigkeit aussagen, ob es gerade regnet oder nicht. Individuell hingegen, ist an einer einzigen Regel nur sehr unzuverlässig die Antwort festzumachen.

Beispielsweise ist ein nasser Boden zwar eine Voraussetzung und ein guter erster Filter, allerdings könnte der Boden genauso gut durch einen Rasensprenger nass sein.

Die Temperatur ist hingegen ein relativ schlechtes Indiz für die Frage, ob es gerade regnet. Es unterscheidet aber den Fall Regen und Schnee und ist somit trotzdem essentiell für die Klassifikation.

Im Anwendungsfall hat jeder weak learner eine Vorhersage. Da weak learners schon dem Namen entsprechend simpel gehalten sind, ist die Vorhersage meist ein boolescher Wert. Durch, im simpelsten Fall, mehrheitliche Abstimmung der weak learners kann ein strong learner geschaffen werden.[6] Oder als schlussfolgernde Definition formuliert:

**Boosting bezeichnet den Prozess die Vorhersagen mehrerer weak learners zu einem strong learner zu verschmelzen.**

### 3.2 Wie funktioniert Boosting?

Die Anschlussfrage die sich stellt ist natürlich, wie funktioniert Boosting im Konkretefall.

- Definition und Grundkonzept von Boosting im Machine Learning
- Historische Entwicklung und theoretischer Hintergrund
- Unterschiede zu anderen Ensemble-Methoden wie Bagging
- Typische Einsatzgebiete und Anwendungen von Boosting

### **3.3 Rolle der Boosting Algorithmen in ML**

- Überblick über verschiedene Boosting-Algorithmen
- Bedeutung und Einfluss von Boosting-Algorithmen in modernen ML-Ansätzen
- Vergleich der Leistung von Boosting-Algorithmen mit anderen ML-Techniken
- Einsatzgebiete von Boosting-Algorithmen in komplexen Problemstellungen

### **3.4 Boosting Algorithmen in Tabellendaten Analyse**

- Relevanz von tabellenartigen Datensätzen in der Datenanalyse
- Effektivität von Boosting-Algorithmen bei der Analyse tabellarischer Daten
- Beispiele und Fallstudien zur Anwendung von Boosting in Tabellendaten
- Herausforderungen und Lösungsansätze beim Einsatz von Boosting in dieser Domäne

Ensemble-Learning[7]

## 4 AdaBoost (5-6 Seiten)

### 4.1 Theoretische Grundlagen

### 4.2 Algorithmus-Struktur und Funktionsweise

### 4.3 Beispielanwendung mit Erläuterung



## 5 Gradient Boosting (5-6 Seiten)

### 5.1 Theoretische Grundlagen

### 5.2 Algorithmus-Struktur und Funktionsweise

### 5.3 Beispielanwendung mit Erläuterung

## **6 Vergleich von AdaBoost und Gradient Boosting (4-5 Seiten)**

### **6.1 Gemeinsamkeiten und Unterschiede**

### **6.2 Performance-Analyse in Benchmarks**

### **6.3 Anwendungsbeispiele und Fallstudien**

## **7 Aktuelle Trends und Entwicklungen (2-3 Seiten)**

### **7.1 Neueste Forschungsergebnisse**

### **7.2 Zukünftige Potenziale von Boosting-Algorithmen**

## 8 Fazit und Ausblick (2-3 Seiten)

8.1 Zusammenfassung der Erkenntnisse

8.2 Reflexion über die Bedeutung für die Praxis

8.3 Ausblick auf zukünftige Forschungsthemen

## Glossar

**AdaBoost** Ein Machine Learning-Algorithmus, der auf dem Prinzip des Boosting basiert.. 1, 2

**Gradient Boosting** Eine Methode des maschinellen Lernens, die für Regression und Klassifikation verwendet wird.. 1, 2

## Literaturverzeichnis

- [1] R. E. Schapire und Y. Freund, *Boosting : foundations and algorithms* (Adaptive Computation and Machine Learning Series). Cambridge, Mass. [u.a.]: MIT Press, 2012, Includes bibliographical references and index, ISBN: 9780262017183. Adresse: <http://www.gbv.de/dms/ilmenau/toc/672276232.PDF%20;%20https://zbmath.org/?q=an:1278.68021>.
- [2] A. [ Géron, *Praxiseinstieg Machine Learning mit Scikit-Learn und TensorFlow : Konzepte, Tools und Techniken für intelligente Systeme*, 1. Auflage, K. [ Rother, Hrsg. Heidelberg: O'Reilly, 2018, Äuthorized German translation of the English edition of 'Hands-on machine learning with Scikit-Learn and TensorFlow: concepts, tools, and techniques to build intelligent systems', ISBN 978-1-491-96228-9, (c) 2017 Impressum; ÜS-Bestseller zu Deep Learning Umschlag"Dieses Buch erscheint in Kooperation mit O'Reilly Media, Inc. unter dem Imprint Ö'Reilly". Rückseite der Titelseite. - "Deutschsprachige O'Reilly-Bücher werden vom dpunkt.verlag in Heidelberg publiziert, vermarktet und vertrieben.- Webseite [www.oreilly.de](http://www.oreilly.de); 201712; aa, ISBN: 9783960090618. Adresse: [http://www.gbv.de/dms/ilmenau/toc/898831717.PDF%20;%20http://vub.de/cover/data/isbn%3A9783960090618/medium/true/de/vub/cover.jpg%20;%20http://deposit.d-nb.de/cgi-bin/dokserv?id=6ba05f7dfc5242d6bbc6f653e199e877&prov=M&dok\\_var=1&dok\\_ext=htm%20;%20https://www.oreilly.de/buecher/13111/9783960090618-praxiseinstieg-machine-learning-mit-scikit-learn-und-tensorflow.html](http://www.gbv.de/dms/ilmenau/toc/898831717.PDF%20;%20http://vub.de/cover/data/isbn%3A9783960090618/medium/true/de/vub/cover.jpg%20;%20http://deposit.d-nb.de/cgi-bin/dokserv?id=6ba05f7dfc5242d6bbc6f653e199e877&prov=M&dok_var=1&dok_ext=htm%20;%20https://www.oreilly.de/buecher/13111/9783960090618-praxiseinstieg-machine-learning-mit-scikit-learn-und-tensorflow.html).
- [3] J. [ Frochte, *Maschinelles Lernen : Grundlagen und Algorithmen in Python* (Hanser eLibrary), 3., überarbeitete und erweiterte Auflage. München: Hanser, 2020, ISBN: 9783446463554. Adresse: <https://dx.doi.org/10.3139/9783446463554%20;%20https://doi.org/10.3139/9783446463554%20;%20https://www.hanser-elibrary.com/doi/book/10.3139/9783446463554%20;%20http://dx.doi.org/10.3139/9783446463554>.
- [4] G. [ James, *An introduction to statistical learning : with applications in Python* (Springer texts in statistics), D. [ Witten, T. [ Hastie, R. [ Tibshirani und J. E. [ Taylor, Hrsg. Cham, Switzerland: Springer, 2023, ISBN: 9783031387463.
- [5] T. [ Hastie, *The elements of statistical learning : data mining, inference, and prediction* (Springer series in statistics), Second edition, R. [ Tibshirani und J. H. [ Friedman, Hrsg. New York, NY: Springer, 2009, Hier auch später erschienene, unveränderte Nachdrucke; Literaturverzeichnis: Seite 699-727, ISBN: 9780387848570. Adresse: <http://www.gbv.de/dms/ilmenau/toc/572093853.PDF%20;%20https://zbmath.org/?q=an:1273.62005%20;%20https://swbplus.bsz-bw.de/bsz287727726kap.htm%20;%20https://swbplus.bsz-bw.de/bsz287727726vor.htm%20;%20https://swbplus.bsz-bw.de/bsz287727726inh.htm%20;%20https://swbplus.bsz-bw.de/bsz287727726cov.jpg>.
- [6] S. Ray. „Quick Introduction to Boosting Algorithms in Machine Learning.“ (2022), Adresse: <https://www.analyticsvidhya.com/blog/2015/11/quick-introduction-boosting-algorithms-machine-learning/> (besucht am 16.11.2023).

- [7] IBM. „Was ist Boosting.“ (2016), Adresse: <https://www.ibm.com/de-de/topics/boosting> (besucht am 16.11.2023).

## Abbildungsverzeichnis



## Tabellenverzeichnis