



University of Bamberg  
Faculty of Information Systems and Applied Computer Science  
Chair of Mobile Systems

Master's Degree in Computing in the Humanities

Master's Thesis

# **Balancing Performance and Resource-Awareness: Optimizing the Model Selection and Ensemble Process in Machine Learning**

Author

**Niklas Diller**

Reviewers

**Prof. Dr. Daniela Nicklas**

**Michael Freitag**

Bamberg, 8. Juli 2024



---

## Zusammenfassung

Dieses Dokument repräsentiert Struktur und Designvorlage des neuen Abteilungsstils für studentische Arbeiten.



---

# Content

<b>1</b>	<b>Introduction</b>	<b>1</b>
<b>2</b>	<b>Anwendung des Style-Pakets</b>	<b>3</b>
2.1	Einbinden des Pakets . . . . .	3
2.2	Änderungen gegenüber book . . . . .	3
2.3	Einbinden von Abbildungen . . . . .	4
2.4	Erstellen und Anpassen des Indexes . . . . .	5
2.5	Internationalisierung . . . . .	5
2.6	Einbetten sämtlicher Schriften . . . . .	6
<b>3</b>	<b>Festlegungen der Abteilung</b>	<b>9</b>
3.1	Verzeichnisreihenfolge . . . . .	9
3.2	Zitate und Literaturreferenzen . . . . .	9
3.3	Die offizielle Titelseite . . . . .	10
	<b>Glossar</b>	<b>11</b>
	<b>Abbreviations</b>	<b>13</b>
	<b>Abbildungen</b>	<b>15</b>
	<b>Literature</b>	<b>17</b>
	<b>Index</b>	<b>19</b>



# 1 Introduction

A common and widely known challenge in training and using machine learning models is to find a suitable balance between overfitting and underfitting [vRV<sup>+</sup>10]. The goal for coming up with valuable predictors for a specific problem therefore is often to create models that neither simply reproduce results gathered from the training data, nor make too broad generalizations. Overfitting would mean that the model would not only learn the underlying pattern of data, but also random noise that comes with it. This results in a bad performing model for new and unseen data, as the model is constantly trying to match its' predictions with the training data too closely. Underfitting on the other hand happens if a model fails to completely cover the complexity of a specific problem. Because new data points are unfamiliar to the model, the resulting predictions will also be poor and not accurate [?]. It becomes a crucial task for machine learning engineers to balance the amount of the training data and the hyperparameters for machine learning models to find a suitable golden mean of creating models that correctly represent the underlying patterns but is at the same time capable of handling new data points.

A similar crucial balancing act in machine learning that seems to find much less attention, however, is the equipoise of the concepts of performance and resource awareness. Both principles are essential for a valuable machine learning model: Predictors need to produce accurate results in order to be of value and research has laid and increasingly focus on designing resource aware machine learning systems [3]. While the two concepts don't necessarily cancel each other out, there seems to be a trade-off between them when exploring and training models. Good performing models often make use of a big set of features and complex algorithms and in order to produce a stream of up-to-date predictions, they are often called in short time intervals. Machine learning models of this kind, however, tend to demand a larger number in computational resources, e.g. CPU time, memory consumption or network utilization. When only looking for the best performing model, i.e. the model that produces the most accurate results, the predictions might become too expensive in regards to memory or battery usage, or simply too computationally complex [4, p. 2]. Especially in the context of edge computing, this balance between good performance and good resource utilization becomes crucial. Edge computing is a computing paradigm that focuses on processing data at the same place where it is produced: at the edge of the network. In contrast to cloud computing, which works by sending all raw data to a central node to be processed there, edge computing operates on the data near the sensors that produce it. That way, less data needs to be sent on a anyway limited network bandwidth, accounting for faster response times, and less network pressure [5].

The main motivation for this work are two projects that the Chair of Mobile Systems of the University of Bamberg is carrying out. Both raise the demand of a model set retrieval system that lets the user decide, how significant resource awareness is supposed to be in relation to performance. These projects will be introduced in greater detail in the following chapter. For a model retrieval system to be working, several varying machine learning models need to exist first. It was therefore decided to implement a model training pipeline along with the retrieval system itself. As many questions regarding the conceptualization of different parts of the use cases are still open, this work also aims to elaborate various theoretical ideas and concepts regarding resource awareness, prediction horizons and model scores.

This work will focus on answering the question on how to manage a meaningful balance between performance and resource awareness in the model selection process. In particular the creation of

ensembles or model sets will be discussed from this perspective while applying them to the use cases that this work will refer to. The chapters of this thesis are structured as followed:



## 2 Anwendung des Style-Pakets

Nachfolgend ein kurzer Leitfaden für die Anwendung des vorliegenden Style-Pakets für LaTeX-Dokumente für Arbeiten in der Abteilung Informationssysteme

- Autor: Markus Schmees
- Anpassungen für studentische Arbeiten C.v.O. Universität Oldenburg: Marco Grawunder und Richard Hackelbusch
- Anpassungen für studentische Arbeiten Otto Friedrich Universität Bamberg: Jan Benno Meyer zu Holte
- Version: 1.4
- Stand: 12.11.2014

### 2.1 Einbinden des Pakets

Die einfachste Möglichkeit besteht darin, das vorgegebene Dokument mit dem Titel

```
BEARBEITER.tex
```

zu verwenden und um eigene Inhalte zu ergänzen. BEARBEITER sollte durch die Art der Arbeit und den Namen des Autors ersetzt werden. Alternativ dazu kann der Autor im eigenen  $\text{\LaTeX}$ -Hauptdokument, das die Dokumentenklasse `book`, eine `11pt`-Schrift sowie die Blattgröße `a4paper` verwenden muss, zusätzlich am Ende der Paketdeklarationen das Abteilungsstylepaket durch folgenden Befehl einbinden

```
\usepackage{abtmobithesis/abtmobithesis}
```

Hierbei aber darauf achten, dass das Paket als **letztes Paket** eingebunden wird, da es diverse  $\text{\LaTeX}$ -Befehle überschreibt, die von anderen Paketen ebenfalls überschrieben werden können. Das resultierende Hauptdokument sollte dann die in der folgenden Abb. 2.1 dargestellte Struktur besitzen

### 2.2 Änderungen gegenüber `book`

Dem Erweiterungspaket zugrunde liegt der Dokumentenstil `book`. Dieser wurde mit Hilfe des vorliegenden Pakets an Besonderheiten der Mobi-Formatvorlage angepasst. Dabei wurden folgende Dinge verändert:

- Abschnittnummerierung auf Ebene 3 gesetzt (bis 1.1.1.1)
- Aufführung im Inhaltsverzeichnis bis Ebene 1 (bis 1.1)
- Ränder, Header, Textweite und Texthöhe angepasst

```

1  \documentclass[11pt,a4paper]{book} % Basisdokumentenklasse
2  \usepackage[latin1]{inputenc}      % F\"{u}r Umlaute
3  \usepackage[T1]{fontenc}           % und {\ss}
4  \usepackage[ngerman]{babel}        % Deutsche Standardbezeichner und Trennung
5  \usepackage{makeidx}               % Indexpaket
6  \usepackage{times}                 % Sch\"{o}nere Schriften
7  \usepackage{longtable}             % Lange Tabelle f\"{u}r Abk\"{u}rzungsverzeichnis
8  \usepackage{graphics}              % EPS-Grafiken einbinden
9  \usepackage{abtmobithesis/abtmobithesis} % Das Stylepaket am Ende einbinden!

10 % Das eigentliche Dokument
11 \begin{document}
12 ... Inhalt ...
13 \end{document}

```

*Abbildung 2.1: Einbinden des Style-Pakets*

- Linken Einzug um 0,37cm bei neuem Absatz eingefügt
- Abstand der Absätze auf 6pt eingestellt
- Seitenzahlen und Abschnitte im Header dargestellt
- Abstand der Überschriften zum Text eingestellt
- Farbe und Schrift der Überschriften festgelegt
- Verzeichnisdeklarationen für korrekte Headeranzeige überschrieben
- Abstand der Abbildungen zu den Unterschriften angepasst
- Schriftart und Style für Abbildungsunterschriften eingestellt
- „Offizielle“ Titelseite integriert
- Seitenzahltypen (römisch und arabisch) kombiniert
- Fußnoten linksbündig und mit Abstand zwischen Nummer und Text
- Sämtliche Listen linksbündig
- Geringere Einrückung der Unterstichpunkte im Index
- Umwandlung der Überschriften für Verzeichnisse in gleichwertige Überschriften ohne „-verzeichnis“

## 2.3 Einbinden von Abbildungen

Wie bei der normalen Verwendung von  $\text{\LaTeX}$  können auch hier über die entsprechenden Zusatzpakete Grafiken eingebunden werden. Immer darauf achten, dass diese als Vektorgrafiken vorliegen, damit sie problemlos skaliert und damit auch ohne Qualitätsverlust gedruckt werden können. Nachfolgend zeigt Abb. 2.2 am Beispiel des Uni-Logos, wie eine Abbildung eingebunden und angezeigt werden kann (vgl. dazu auch Abschnitt 2.6).



*Abbildung 2.2: Eine schöne Abbildung*

## 2.4 Erstellen und Anpassen des Indexes

Um das grundlegende Indexfile zu erzeugen, muss der Befehl `makeindex` innerhalb des Haupt- $\text{\LaTeX}$ -Dokuments aufgeführt sein. Dieser erzeugt aus sämtlichen `index`-Befehlen die Datei

```
dissertation.idx.
```

Um diese Datei ebenfalls passend zu formatieren, wurde ein entsprechendes Stylefile namens

```
abtmobithesis.ist
```

erstellt. Die Anwendung dieses Styles auf den Index erfolgt mit Hilfe des folgenden Befehls

```
makeindex -s abtmobithesis/abtmobithesis.ist -o content/index.tex \
dissertation.idx
```

Dieser Befehl sorgt dafür, dass die Indexdaten entsprechend formatiert und in die Datei `index.tex` in das Verzeichnis `content` abgelegt werden. Von dort kann der neue Index dann problemlos per `input` eingelesen werden. Also bitte nicht den traditionell erzeugten Index über den Befehl `printindex` einbinden! Der Auszug aus dem  $\text{\LaTeX}$ -Hauptdokument zur Erzeugung und Einbindung des Indexes sieht folgendermaßen aus

## 2.5 Internationalisierung

Durch Einbinden entsprechender Sprachpakete wird  $\text{\LaTeX}$  an Besonderheiten bestimmter Sprachräume, z.B. deren Trennung oder Verzeichnisbezeichnung angepasst. Durch die individuelle Gestaltung bestimmter Verzeichnisnamen wurden die zugehörigen Befehle aber überschrieben. Das bedeutet nun, dass der vorliegende Stil zwar problemlos für deutsche Arbeiten anwendbar ist, aber diese Verzeichnisbezeichnung für Dissertationen in fremder Sprache angepasst werden müssen. Diese betreffen also insbesondere die folgenden Paket- und Verzeichnisdeklarationen in den Zeilen 3, 11, 14, 18, 19, 23, 25, 29 und 30.

```

1  ...
2  % Indexdatei erzeugen
3  \makeindex
4
5  % Das eigentliche Dokument
6  \begin{document}
7
8  ...
9
10 \index{Irgendwas}           % Einstufiger Indexeintrag
11 \index{Irgendwas!Ganzanderes} % Zweistufiger Indexeintrag
12
13 ...
14
15 % Index als letztes Verzeichnis
16 \cleardoublepage           % Auf ungerader Seite beginnen
17 \addcontentsline{toc}{chapter}{Index} % Index im Inhaltsverzeichnis anzeigen
18 \input{content/index.tex}   % Formatierten Index einbinden
19 \end{document}

```

*Abbildung 2.3: Index erstellen, formatieren und integrieren*

## 2.6 Einbetten sämtlicher Schriften

Bitte unbedingt folgende Dokumenteneinstellung beibehalten

```
\documentclass[11pt,a4paper]{book}
```

Grundsätzlich bietet es sich an, die Arbeit mit `pdflatex` direkt als PDF zu erzeugen. Bei der Verwendung von `pdflatex` können als Bildformate PDF, PNG oder JPG (Achtung: Die beiden letzteren sind nicht im Vektorformat und damit nicht gut für den Ausdruck geeignet.) für Bilder verwendet werden.

Es sollten aber sämtliche verwendete Schriftarten in das resultierende PDF-Dokument eingebettet werden. `pdflatex` und `ps2pdf` sparen unter Umständen einige Schriften aus, sollten also nur mit Bedacht verwendet werden.

Eine Möglichkeit, dennoch sämtliche Schriftarten in das resultierende Dokument einzubetten, bietet eine Kombination aus `dvips` und `ghostscript` an. Zunächst das resultierende DVI-Dokument `dissertation.dvi` in ein passendes PS-Dokument umwandeln mit dem folgenden Befehl

```
dvips -P pdf -t A4 -z -Pdownload35 dissertation.dvi
```

Die resultierende Datei `dissertation.ps` darf daraufhin NICHT mit `ps2pdf` sondern muss – wie der folgende Befehl zeigt – mit `ghostscript` in ein PDF umgewandelt werden

```
gs -dSAFER -dNOPAUSE -dBATCH -sDEVICE=pdfwrite \
    -sPAPERSIZE=a4 -dPDFSETTINGS=/printer \
```

```

1  \documentclass[11pt,a4paper]{book}
2  ...
3  \usepackage[ngerman]{babel}
4  ...
5  \usepackage{abtmobithesis/abtmobithesis}
6  ...
7  % Das eigentliche Dokument
8  \begin{document}
9  ...
10 % Verzeichnisse am Ende, erst das Glossar
11 \addonchapter{Glossar} %
12 \input{content/glossar.tex}
13
14 % Dann die Abk\u{u}rzungen
15 \addonchapter{Abk\u{u}rzungen}
16 \input{content/abkuerzungen.tex}
17
18 % Weiter mit Abbildungen
19 \cleardoublepage
20 \addcontentsline{toc}{chapter}{Abbildungen}
21 \renewcommand{\listfigurename}{Abbildungen}
22 \listoffigures
23
24 % Schlie\u{ss}lich Literatur
25 \cleardoublepage
26 \addcontentsline{toc}{chapter}{Literatur}
27 \bibliographystyle{alphadin}
28 \renewcommand{\bibname}{Literatur}
29 \bibliography{content/bibliographie}
30
31 % Und ganz am Ende der Index
32 \cleardoublepage
33 \addcontentsline{toc}{chapter}{Index}
34 \input{content/index.tex}
35
36 % Und schlie\u{ss}lich noch die Versicherung f\u{u}r das Pr\u{u}fungsamt
37 \cleardoublepage
38 \versicherung{Bamberg}
39
40 \end{document}

```

*Abbildung 2.4: Anzupassende Pakete bzw. Verzeichnisbezeichner*

```
-dCompatibilityLevel=1.3 -dMaxSubsetPct=100 \  
  
-dSubsetFonts=true -dEmbedAllFonts=true \  
  
-sOutputFile=dissertation.pdf dissertation.ps
```

Das hieraus resultierende PDF-Dokument `dissertation.pdf` enthält schließlich sowohl das gewünschte Format als auch sämtliche Schriften. Bei dem Umweg über EPS ist allerdings zu beachten, dass Vektorgraphiken nicht als PDF sondern als EPS-Dateien eingebunden werden müssen.

## 3 Festlegungen der Abteilung

Über das reine Design hinaus sind nachfolgend Festlegungen zur Struktur einer Arbeit (insbesondere Studentischer Arbeiten) der Abteilung Informationssysteme aufgeführt.

### 3.1 Verzeichnisreihenfolge

Die Reihenfolge der Verzeichnisse soll dem vorliegenden Dokument folgen, die Verzeichnisse dazu wie folgt angeordnet sein. Im Inhaltsverzeichnis, das neuerdings nur noch „Inhalt“ heißt, werden nur die nachfolgend fett dargestellten Punkte aufgeführt.

1. Zusammenfassung
2. Abstract
3. Inhalt (vormals Inhaltsverzeichnis)
4. **Der eigentliche Inhalt** mit laufenden Kapitelnummern
5. **Glossar** und Folgende jeweils ohne Kapitelnummer
6. **Abkürzungen**
7. **Abbildungen**
8. **Literatur**
9. **Index**

Hierbei wurde auf die traditionelle Verwendung der Begriffe Inhaltsverzeichnis, Literaturverzeichnis und Abbildungsverzeichnis verzichtet und stattdessen nur noch Inhalt, Literatur und Abbildungen verwendet.

### 3.2 Zitate und Literaturreferenzen

Im eigentlichen Text kann man verschiedene Zitate verwenden. Das machen auch z.B. [?], [?], [?] oder [?]. Den jeweiligen Zitaten soll dazu der Bibliographiestil `alphadin` (mit kleinem „a“) zugrunde liegen. Dieser repräsentiert einen genormten Zitierstil und wird durch folgende Zeile im  $\text{\LaTeX}$ -Dokument eingestellt

```
\bibliographystyle{alphadin}
```

Wer trotz des verkürzten Zitierstils in ausführlicher Weise auf die jeweiligen Autoren hinweisen möchte, hat die Möglichkeit, dies z.B. folgendermaßen zu machen: „wie auch von Appelrath, Boles, Kleinfeld und andere in [?] beschrieben. ...“

Bei den Literaturangaben wird künftig auf die ISBN- bzw. ISSN-Nummern verzichtet, da bei Dissertationen ohnehin davon auszugehen ist, dass begutachtete bzw. verlegte Literatur verwendet wird.

### 3.3 Die offizielle Titelseite

Der `\maketitle`-Befehl wurde überschrieben. Für die neue Titelseite sind daher folgende Einstellungen zu machen

```
1 % Voreinstellungen für offizielle Titelseite
2 \title{Hier den Titel eingeben} % Titel der Dissertation
3 \author{Dipl.-Inform. Vorname Nachname} % Name des Autors
4 \arbeitstyp{Diplomarbeit} % oder Bachelorarbeit, Masterarbeit ...
5 \erstgutachter{Prof. Dr. Erster Gutachter} % Name des Erstgutachters
6 \zweitgutachter{Prof. Dr.-Ing. Zweiter Gutachter} % Name des Zweitgutachters
7
8 % Einbinden/Anzeigen der Titelseite innerhalb von \begin/end{document}
9 \maketitle
```

*Abbildung 3.1: Einstellungen für offizielle Titelseite*



## Glossar

Nachfolgend sind noch einmal wesentliche Begriffe dieser Arbeit zusammengefasst und erläutert. Eine ausführliche Erklärung findet sich jeweils in den einführenden Abschnitten sowie der jeweils darin angegebenen Literatur. Das im Folgenden im Rahmen der Erläuterung verwendete Symbol  $\sim$  bezieht sich jeweils auf den im Einzelnen vorgestellten Begriff, das Symbol  $\uparrow$  verweist auf einen ebenfalls innerhalb dieses Glossars erklärten Begriff.

**Auktion** Eine  $\sim$  ist das im  $\uparrow$ E-Commerce am Häufigsten eingesetzte Verfahren zur dynamischen  $\uparrow$ Preisfindung. Interessenten können dabei durch Abgabe von Geboten Preis, Dauer und Gewinner beeinflussen. Bei einer offenen  $\sim$  sind Bieter, Höhe der Gebote und der aktuelle Preis für alle Teilnehmer sichtbar, bei der geschlossenen (sealed)  $\sim$  erfolgt nur eine interne Benachrichtigung. Die bekanntesten Typen sind die traditionelle  $\uparrow$ Versteigerung sowie die  $\uparrow$ holländische,  $\uparrow$ umgekehrte und  $\uparrow$ verdeckte  $\sim$ .

**Behaviorismus** Der  $\sim$  ist eine  $\uparrow$ Lerntheorie, die davon ausgeht, dass Wissen als Struktur unabhängig vom  $\uparrow$ Lernenden existiert und dass sein Verhalten operant konditioniert ist, d.h. dass es als Konsequenz aus anderen Verhaltensweisen resultiert. Erfolgt eine positive Reaktion, behält der  $\uparrow$ Lernende neu erlerntes Verhalten bei, negative Reaktionen führen zu einer Verminderung dieses Verhaltens. Der  $\uparrow$ Lehrende bestimmt dabei das zu erlernende Wissen und ist für die Steuerung des  $\uparrow$ Lernprozesses zuständig.



## Abbreviations

ADDIE	Analysis, Design, Development, Implementation, Evaluation
ADEPT	Advanced Decision Environment for Process Tasks
ADL	Advanced Distributed Learning Initiative
AGB	Allgemeine Geschäftsbedingungen
AGOF	Arbeitsgemeinschaft Online Forschung
AICC	Aviation Industry Computer Based Training Committee
API	Application Programming Interface
ARIS	Architektur integrierter Informationssysteme
ASC	Accredited Standards Committee
ASP	Application Service Providing
ASTD	American Society for Training and Development
AXIS	Apache eXtensible Interaction System
B2B	Business-to-Business
B2C	Business-to-Consumer
BDSG	Bundesdatenschutzgesetz
BGB	Bürgerliches Gesetzbuch
BITKOM	Bundesverband Informationswirtschaft, Telekommunikation und neue Medien
BMBF	Bundesministerium für Bildung und Forschung
BME	Bundesverband Materialwirtschaft, Einkauf und Logistik
BPEL4WS	Business Process Execution Language for Web Services
BSCW	Basic Support for Cooperative Work
BSI	Bundesamt für Sicherheit in der Informationstechnik
CAL	Computer Aided/Assisted Learning
CBT	Computer Based Training
CC	Creative Commons
CD	Compact Disc
CELab	Labor für Content Engineering
CMI	Computer Managed Instruction
CMS	Content Management System
CORBA	Common Object Request Broker Architecture
CPU	Central Processing Unit
CSCL	Computer Supported Collaborative Learning
CSCW	Computer Supported Cooperative Work
CSS	Customer Support Services
CRM	Customer Relationship Management
CUL	Computerunterstütztes Lernen
DBMS	Datenbankmanagementsystem
DCOM	Distributed Component Object Model
DFN	Deutsches Forschungsnetz
DIN	Deutsches Institut für Normung
DREL	Digital Rights Expression Language
DRM	Digital Rights Management
DVD	Digital Video Disc
E2B	Education-to-Business

E2C	Education-to-Consumer
E2E	Education-to-Education
EAI	Enterprise Application Integration
EAN	European Article Numbering
EBPP	Electronic Bill Presentment and Payment
ebXML	Electronic Business Extensible Markup Language
ECA	Event-Condition-Action
EC	Electronic Cash
ECC	E-Learning Courseware Certification
EDI	Electronic Data Interchange
EDIFACT	EDI for Administration, Commerce and Transport
EFQM	European Foundation for Quality Management
EGBGB	Einführungsgesetz zum Bürgerlichen Gesetzbuch
EITO	European Information Technology Observatory
ELAN	E-Learning Academic Network Niedersachsen

## Abbildungen

2.1	Einbinden des Style-Pakets . . . . .	4
2.2	Eine schöne Abbildung . . . . .	5
2.3	Index erstellen, formatieren und integrieren . . . . .	6
2.4	Anzupassende Pakete bzw. Verzeichnisbezeichner . . . . .	7
3.1	Einstellungen für offizielle Titelseite . . . . .	10



## Literature

- [vRV<sup>+</sup>10] W. M. P. van der Aalst, V. Rubin, H. M. W. Verbeek, B. F. van Dongen, E. Kindler, and C. W. Günther. Process mining: A two-step approach to balance between underfitting and overfitting. *Software & Systems Modeling*, 9(1):87–111, January 2010.





---

## Index

### A

Abbildungen .....	2
Änderungen .....	2
Anwendung .....	1

### E

Einbinden .....	1
Abbildungen .....	2

### F

Festlegungen .....	7
--------------------	---

### I

Indexerstellung .....	3
Internationalisierung .....	4

### L

Literatur .....	7
Literaturreferenzen .....	7

### P

Paket .....	1
-------------	---

### S

Schriften .....	5
-----------------	---

### T

Titelseite .....	8
------------------	---

### V

Verzeichnisreihenfolge .....	7
------------------------------	---

### Z

Zitate .....	7
--------------	---



## Versicherung

Hiermit versichere ich, dass ich diese Arbeit selbständig verfasst und keine anderen als die angegebenen Quellen und Hilfsmittel benutzt habe. Außerdem versichere ich, dass ich die allgemeinen Prinzipien wissenschaftlicher Arbeit und Veröffentlichung, wie sie in den Leitlinien guter wissenschaftlicher Praxis der Otto Friedrich Universität Bamberg festgelegt sind, befolgt habe.

Bamberg, den 8. Juli 2024

---

Niklas Diller