

# [Titel der Ausarbeitung]

[Untertitel der Ausarbeitung]

**Ausarbeitung**

im Studiengang Medieninformatik

Fachbereich Medien  
Hochschule Düsseldorf

Direnc Timur  
Matrikel-Nr.: 123456  
Datum: Juni 2022

Till Pilarczyk  
Matrikel-Nr.: 765335  
Datum: Juni 2022

Prüfer  
Prof. Dr.-Ing. Thomas Rakow

# Eidesstattliche Erklärung

Ich erkläre hiermit an Eides statt, dass ich die vorliegende Ausarbeitung selbständig und ohne unzulässige fremde Hilfe angefertigt habe. Die verwendeten Quellen sind vollständig zitiert. Diese Arbeit wurde weder in gleicher noch ähnlicher Form einem anderen Prüfungsamt vorgelegt oder veröffentlicht. Ich erkläre mich ausdrücklich damit einverstanden, dass diese Arbeit mittels eines Dienstes zur Erkennung von Plagiaten überprüft wird.

*Ort, Datum*

*Direnc Timur*

*Ort, Datum*

*Till Pilarczyk*

## **Kontaktinformationen**

Direnc Timur

direnc.timur@study.hs-duesseldorf.de

—

Till Pilarczyk

till.pilarczyk@study.hs-duesseldorf.de

# Zusammenfassung

[Titel der Ausarbeitung]

Direnc Timur

Dies ist die Zusammenfassung Ihrer Arbeit ...

# Abstract

[Titel der Ausarbeitung]

Direnc Timur

Diese möchten Sie natürlich auch auf Englisch bereitstellen . . .

# Inhaltsverzeichnis

<b>1</b>	<b>Einleitung</b>	<b>1</b>
<b>2</b>	<b>Aktuelle Forschung</b>	<b>2</b>
2.1	Whitepaper künstliche Intelligenz in der Hochschullehre . . . . .	2
2.1.1	Lehren und Lernen mit KI in der Hochschulausbildung . . . . .	2
2.2	Pass/Fail Prediction with Moodle . . . . .	4
2.3	Klassifizierungen . . . . .	4
2.4	Daten . . . . .	4
2.5	Algorithmen . . . . .	5
2.6	Model Building an Evaluation . . . . .	5
2.7	Ergebnisse . . . . .	5
2.7.1	Ergebnisse mit 6 Wochen Daten und 10 Wochen Daten . . . . .	5
2.8	Korrelation der Variablen . . . . .	5
2.9	Diskussion . . . . .	6
2.10	Ergebnis . . . . .	6
2.11	AIED . . . . .	6
2.12	Klassifizierungen . . . . .	10
2.13	Daten . . . . .	10
2.14	Algorithmen . . . . .	10
2.15	Model Building an Evaluation . . . . .	11
2.16	Ergebnisse . . . . .	11
2.16.1	Ergebnisse mit 6 Wochen Daten und 10 Wochen Daten . . . . .	11
2.17	Korrelation der Variablen . . . . .	11
2.18	Diskussion . . . . .	11
2.19	Ergebnis . . . . .	12
<b>3</b>	<b>Einleitung</b>	<b>14</b>
3.1	Exposé . . . . .	14
3.2	Hinweise . . . . .	15
3.3	Inhalt . . . . .	15
3.4	Organisatorisches . . . . .	15
3.5	Bewertungskriterien . . . . .	15
<b>4</b>	<b>Stile</b>	<b>16</b>
4.1	Text . . . . .	16
4.2	Abbildungen und Tabellen . . . . .	16
4.3	Zitieren . . . . .	17
4.4	Listen . . . . .	17

<b>5</b>	<b>Tools</b>	<b>18</b>
5.1	L <sup>A</sup> T <sub>E</sub> X . . . . .	18
<b>6</b>	<b>Infrastruktur</b>	<b>19</b>
6.1	GitLab-Server . . . . .	19
6.2	Versionsverwaltung mit Git . . . . .	19
<b>A</b>	<b>Tipps zu häufig gemachten Fehlern</b>	<b>21</b>
A.1	Abbildungen, Tabellen, Listings, etc. . . . .	21
A.2	Text . . . . .	21
A.3	Diverses . . . . .	21

# Abbildungsverzeichnis

4.1	Logo Hochschule Düsseldorf . . . . .	17
-----	--------------------------------------	----

# Tabellenverzeichnis

4.1	Beispieltabelle . . . . .	17
-----	---------------------------	----



# Abkürzungsverzeichnis

**BSI** Bundesamt für Sicherheit in der Informationstechnik

# Kapitel 1

## Einleitung

Aufgabenstellung beschreiben und in DAW einordnen

Historische Entwicklung. Gibt es erst seit 25 Jahren, ist also noch in den Kinderschuhen... es wird viel geforscht Diagramm finden warum das Thema Wichtig ist, wie viele

Hochschulen setzten KIs ein im verlaufe der Jahren ([https : //hochschulforumdigitalisierung.de/sites/default/files/2018-09/HE.pdf](https://hochschulforumdigitalisierung.de/sites/default/files/2018-09/HE.pdf) Seite9)

Als erstes .....

# Kapitel 2

## Aktuelle Forschung

In diesem Kapitel werden einige Paper vorgestellt, in den sich mit dem Thema KI an Hochschulen oder KIs an Hochschulen, die mit Zugriffsdaten arbeiten vorgestellt.

### 2.1 Whitepaper künstliche Intelligenz in der Hochschullehre

Kernthemen:

1. Lehren und Lernen mit KI in der Hochschulausbildung
2. Lehren und Lernen über KI in der Hochschulausbildung
3. KI und Ethik im Hochschulkontext
4. Zukunftsperspektiven für KI in der Hochschulbildung

#### 2.1.1 Lehren und Lernen mit KI in der Hochschulausbildung

Gründe, warum KI eingesetzt werden sollte:

- Verwendung von Daten zur Unterstützung von Studierenden bei der Weiterentwicklung des eigenen Lernverhaltens und Lehrenden bei der Verbesserung ihrer Didaktik.
- Förderung von kritischem und kreativem Denken
- Entdeckung neuer Erkenntnisse durch Analyse von Daten
- Vorzeitige Erkennung von Risiko-Studierenden

Auflistung von Problemen in Studium und Lehre und wie diese Probleme durch KI-Anwendungen adressiert werden können. Beispiele: Begrenzte Tutorzeit von Lehrenden für Studierende, Lösung virtuelle intelligente Assistenten; Aufwändige Bewertung von Texteinreichungen können durch sog. Automated Essay Scoring automatisiert werden.

## Learning Analytics

Learning Analytics als Prozess, um den Prozess des Lernens zu unterstützen. Einteilung von Learning Analytics in drei Varianten: Mikroebene, Mesoebene und Makroebene. Der Prozess kann unterschiedliche Zielgruppen unterstützen, wie Lernende, Lehrende oder auch gesamte Organisationen. Die Varianten beziehen sich dabei auf den Umfang und zeitlichen Kontext des Lernens. In der Mikroebene können Lernende bspw. im Rahmen einer Lernsitzung oder im Verlaufe eines Kurses unterstützt werden. In der Mesoebene können Lernende über ein gesamtes Semester begleitet und Daten kursübergreifend ausgewertet werden.

Learning Analytics könnte auch für die Erkennung von gefährdeten Lernenden eingesetzt werden. Dabei können auf Grundlage der bisherigen Leistungen Abbrüche oder Noten vorhergesagt werden. Mit

Paper listet Herausforderungen, die bei Learning Analytics beachtet werden müssen. Beispiele: Analysen betrachten meist nur einen Ausschnitt des Lernens; Systeme sollten sozialen Kontext berücksichtigen; Berücksichtigung persönlicher Eigenschaften.

## Arten von Künstlichen Intelligenzen in der Hochschulbildung

**Personalisierte adaptive Lernumgebungen (ALU)** Digitale Lernumgebungen für die autonome Lehre gut strukturierter Inhalte. Trennung in Domänenmodell (Fach), Lernermodell und Didaktikmodell. Das Lernermodell betrifft den Zustand des Lernenden und welches Wissen er besitzt. Das Didaktikmodell schlägt auf Basis des Lernenden Lernstrategien und Handlungsempfehlungen vor.

**Chatbots** Lernende können durch Chatbots moderiert und unterstützt werden, wobei sie gleichzeitig Lehrende entlasten.

**Empfehlungssysteme** Empfehlungssysteme sind bislang nur im Auswahlprozess eines Studiengangs im Einsatz.

**Edu-Robots** Meist Begleitungen oder Ersatz für Lehrende in Kursen.

**KI-Schreibbots** Generierung einzigartiger Texte im Hochschulkontext. Birgt die Herausforderung von erschwerter Überprüfbarkeit.

## Veränderung der Betreuung

Zunehmende Vermittlung von kritischem Denken, statt klassischem Wissen. Risiko durch nicht „Allwissenheit“ besteht. Beim Mentoring wichtig ist die Erkennung des Status des Lernenden und damit verknüpfte reaktive Handlungen. Aufteilung des Mentoring in die Phasen Vorbereitung, Lernprozess und Nachbereitung.

- Vorbereitung: Erhebung der Lerninhalte und Planung der Ausgestaltung des Lernprozesses.

- Lernprozess: Ausführung der geplanten individualisierten Aktivitäten zur An-eignung der Lerninhalte.
- Nachbereitung: Reflexion des Lernprozesses über Lernerfolg.

Intelligente Tutor- und Lernsysteme führen zur Verbesserung der Lernergebnisse. Fehlende Anwendung von KI, bislang nur Vorhersage von Erfolg und Misserfolg des Lernprozesses.

## 2.2 Pass/Fail Prediction with Moodle

Durch Zugriffsdaten, die von vielen Anwendungen erstellt und gespeichert werden, stehen viele Daten zu Verfügung, die genutzt werden können um das Lehren und das lernen in einem positiven zu verbessern.

LMS erklären ? oder bekannt ?

Das Paper [Quelle] beschäftigt sich mit den zwei Fragen, ob die Daten, die von Moodle Aktivitäten während der Dauer eines Kurses gesammelt werden ausreichen um Voraussagen zu können, wie gut ein Student in diesem Fach abschließt (Welche Note (nach dem Alphabet) und damit auch ob bestanden oder gefailt) Die zweite Frage, die versucht wird zu beantworten, beschäftigt sich damit herauszufinden, ob die selben Daten, jedoch nur für den Zeitraum von sechs bzw Zehn Wochen ausreichen um herauszufinden ob der Student das Fach bestehen wird oder nicht.

## Datensätze

Um die Fragen zu beantworten wurden Daten von 29 verschiedenen Kursen zu 9 verschiedenen Modulen zwischen den Jahren 2011 und 2018 ausgewählt. Alle Module kommen aus dem selben Fachbereich und dem selben Dozenten um eine homogene Maß an Daten anzusammeln. Insgesamt konnten so Kursdaten von 690 Studenten (410 waren Individuell, rest selber Student) gesammelt werden, wobei 83 davon, den Kurs frühzeitig verlassen haben.

Aufgrund des doch Speziellen Datensatzes können die Ergebnisse nur im Context des kleinen Studenten Kreises gesehen werden.

## 2.3 Klassifizierungen

Um die Erste Frage zu beantworten wurde eine KI entwickelt. Diese KI kann insgesamt 5 verschiedene Vorhersagen treffen. "Distinction"(80%+) "Merit"65%- 79%) "Pass"(50% - 64%) "Fail"(1% bis 49%) "Early Exit"(0% oder keine Note)

Um die Zweite Frage zu beantworten wurden die ersten 3 Klassen zur Klasse "Pass" zusammengefasst und die letzten beiden zu "Fail"

## 2.4 Daten

Es wurden nur Zugriffsdaten von Studenten zum Trainieren und Validieren der KI genommen, die von Dozenten mussten entfernt werden Es wurden folgende Zugriffsdaten verwendet: LISTE im PAPER!!!

## 2.5 Algorithmen

Genauer Beschreiben ? Random Forest: Gradient Boosting: K nearest Neighbours: Linear Discriminant Analysis:

Dies sind verschiedene Algorithmen die benutzt werden können um Daten in Klassen zu Ordnen.

## 2.6 Model Building an Evaluation

Data-Split in 70-30

Kappa Statistic gib an wie gut die vom Modell vorausgesagten Klassen mit den richtigen Klassen übereinstimmen, zur Kontrolle wird ein zu einem Zufälligen Klassifiziere, der anhand der Häufigkeit der vorkommen von Klassen rät um welche Klasse es sich handelt genommen 0.02 = Fair 0.21-0.40. Moderat 0.41-0.60 Zustimmend 0.61-0.80 wesentliche Zustimmung 0.81-1.0 Perfekte Zustimmung

no-information Rate = Genauigkeit wenn immer die Klasse mit den meisten haben in dem Testset gerraten wird

## 2.7 Ergebnisse

Das Training und Testen der KI hat ergeben, dass alle Algorithmen eine sehr ähnliche Genauigkeit haben. Die Genauigkeit liegt bei dem Modell zu 95% in dem Bereich 53% bis 67%. Des Weiteren waren alle Algorithmen besser als die no-information Rate. Am besten konnte vorausgesagt werden, wenn ein Student besonders gut den Kurs besteht oder den Kurs frühzeitig beendet. Alles andere konnte nicht gut predicted werden.

Wenn anhand von den Zugriffsdaten für die gesamte Länge des Kurses, konnte, wenn nur festgestellt werden soll ob ein Student besteht oder nicht, mithilfe des Random Forest Algorithmus wurde eine Genauigkeit von 92% erreicht (148 von 152 Studenten die bestanden haben und 42 von 54 die nicht bestanden haben konnten korrekt vorhergesagt werden)

### 2.7.1 Ergebnisse mit 6 Wochen Daten und 10 Wochen Daten

Wenn die KI nur mit 6 Wochen alten Daten voraussagen treffen soll, hat diese nicht besser abgeschnitten als die no-information rate. Wurden jedoch 10 Wochen alte Ergebnisse genommen, konnte die deutlich bessere vorhersagen treffen als die no-information rate. Auch hier waren alle Algorithmen besser, am besten war jedoch der Random Forest Algorithmus mit 82.18%

## 2.8 Korrelation der Variablen

Pass/Fail all Data: Regular: und Assignment Submitted

Random Forest /Pass/Fail 10 Weeks Regular / Assignment Views Course Views  
pm.Early am.late

## 2.9 Diskussion

Frage 1: Wenn alle Daten benutzt werden ist es möglich die Notesignificant besser zu bestimmen als mit der no-information-rate am besten war dabei der Random Forest mit einer Genauigkeit von 60,5. Dies deutet darauf hin, dass die Moodleaktivität benutzt werden kann um die Noten vorherzusagen. Dabei war die Unterscheidung, ob ein Student sehr gut besteht oder abbricht am besten zu bestimmen, das kann daran liegen, dass dort das Nutzerverhalten extrem eindeutig ist. Wenn man nur bestimmen möchte ob die Studenten bestehen oder nicht können die Daten sehr gut genutzt werden. Dort hat man eine Genauigkeit von 92.2% Dieses Ergebnis stimmt auch mit vielen anderen Studien überein.

Das gute Ergebnis kann daran liegen, dass durch den längeren Zeitraum mehr Daten vorhanden waren. Im Gegensatz zu anderen Studien waren hier weniger als 83 Prozent der Daten passiv sonst sind das oft über 95 Prozent.

Frage 2: Für den kurzen Zeitraum von 6 Wochen hat keiner einen Algorithmus eine Verbesserung gebracht. Mit Daten von 10 Wochen hat die Performance etwas verbessert. Jedoch auch hier kann nur weniger als die Hälfte der failing student vorhergesagt werden. Dies deutet darauf hin, dass LMS Zugriffsdaten hilfreich für ein Frühwarnsystem sein können, jedoch werden noch mehr Daten benötigt z. B. eine noch genauere LMS Nutzungsdaten oder Assessment results, studentische Interaktion mit anderen. Dies wurde auch in anderen Studien schon gezeigt, dass diese Daten helfen können.

Trotz der hohen false-positiven kann es hilfreich sein so ein Frühwarnsystem einzurichten, da dadurch trotzdem einige Studenten erreicht, die dadurch Hilfe bekommen können und das die Schulen und den Studenten weniger kostet, als wenn diese einfach am Ende durchfallen.

Es wurde gezeigt, dass das reguläre Login darauf hinweist, dass ein Student eher besteht. Viele Forscher meinen dies zeigt darauf, dass die Studenten ein besseres Lern-/Zeit Management haben und dies ein wichtiger Faktor für den Erfolg ist.

Um die Modelle zu verbessern könnten feinere Zugriffsdaten gesammelt werden, sowie Ergebnisse von Aufgaben beachtet werden. Dozenten sollten Aufgaben früh in ihren Kurs einbauen und die Möglichkeiten von LMS nutzen um studentische Interaktivität zu fördern.

## 2.10 Ergebnis

Moodle Data kann benutzt werden um vorherzusagen, wie gut ein Student besteht und ob dieser besteht. Um dies jedoch praktisch zu benutzen, wäre ein Frühwarnsystem gut, Moodle Daten reichen dazu nicht ausreichen aus. Die getesteten Modelle sollten daher mit mehr Daten, genaueren Zugriffsdaten und Aufgaben ergebnisse miteinbezogen werden. Des Weiteren waren alle Zugriffsdaten nur von Kursen die von einem Dozenten geleitet wurden. Daher ist das Ergebnis nicht gut Generalisierbar

## 2.11 AIED

Artificial Intelligence in Education (AIED)

Paper beschäftigt sich damit den State of the Art anzuschauen (vergangenen 25 Jahre )um rückschlüsse auf wie sich das thema in den Nächsten Jahren entwickelt wird (auch 25). Die Aktuellen trends werden dabei Interpretier in einer Utopie und Dystopie Zum Schluss sollen noch sieben wichtige Risiken Herausforderungen und Chancen ermittelt werden für AIED um erfolgreich zu sein.

Die Autoren wissen jedoch auch das es nicht möglich sein wird etwas vorherzusagen, jedoch soll das Paper in zukunft dazu dienen um zu wissen was man sich früher bei dem Thema gedacht hat.

#### **Current Trends and Developments**

Nehmen AIED gibt es viele Faktoren ausßerhalb von AIED die AIED beeinflussen. GERman Computer Sciende Association hat fünf große Ziele definiert für die Computer wissenschaft. In dieser liste stehen auch herausforderungen die relevant für AIED sind. Zum Beispiel der Schutz von Lerndaten der Nutezrverhalten.

Viel Einfluss wird auch die Big Data und Industrie 4.0 haben. Da es unumstreitbar ist, das AIED viele Daten brauch um besser zu funktionieren

#### **Educational trends and Developments Relevant to AIED**

Die Dedactic ist ein nicht schnell änders Feld.

TODO:

Unterschiedliche Kulturen lernen Anders, daher ist es schwierig eine Software zu haben die in vielen Ländern eingesetzt werden kann. Außerdem ist es dadruch schwiie-rig eine allgegenwörtige, universelle Interaktionsmethode mit der Software/Hardware herzustellen, welche nötig ist damit Lehrende diese in ihere Lehre einbinden können. Ein Weitres Problem ist es, dass Lerherende, die von Ihnen verwendete Technologie verstehen und deren Vorteile kennen müssen, dies passiert jedoch in den seltesten Fällen. Des Weitren ist die Schulinfrastruktur nicht ausreichend genug vorhanden um verschiedene technologien einsetzen pädagogisch sinnvoll einsetzen zu können.

#### **Two Szenarios**

Das Paper nennt zwei extreme Szenarien die in 25 Jahren eintreten könnten und die uns vorauge führen sollen, in welche Rcithungen AIED gehen kann

#### **Utopie**

Zwei personen unterhalten sich darüber wie hilfreich deren verwendete Technologie ihnen bei ihrem lernen geholfen hat, wie sie dadurch inhalte besser verstanden haben, Sich die anwendungen an deren Lernart angepasst hat und das sie sie motivieren mit spielen und sie Spaß beim lernen haben, sowie auch am Wochendende Hilfe von Digi-talen Tutoren haben können und Aufgaben kontrollieren lassen und das die Lehrere keinen einsicht in die Daten haben und deren Daten nur verwendet wird um ihnen zu helfen und um anonymisiert weitergeleitet werden um die tools zu verbessern. Und das deren Daten benutzt werden um die Tools zu verbessern.



#### **Distopie**

Zwei Personen sitzen in der Bibliothek und bevorzugen Bücher anstelle von Tools. Das Problem ist hier, dass die Tools nicht für deren Land ausgerichtet sind und deren Lehrer diese nicht benutzen. Des Weiteren vermuten die Personen, dass die Lehrer auch besorgt um deren Job ist. Denn in den letzten 10 Jahren haben 50 % der Lehrer von Mathematik und Wissenschaftsfächern ihren Job verloren. Die Tools, die benutzt werden, besonders in Ethikfächern oder ähnliches, sollte mit Vorsicht genutzt werden, da diese die Antworten der Nutzer speichern und weiterverkauft. Es gibt Personen, die daher keinen Job gefunden haben. Daher werden auch viele Tools und Foren von wenigen Personen benutzt und man bekommt kaum Hilfe.

#### **Discussion und Conclusion**

Beide Szenarien folgen den aktuellen Technologie Trends und Herausforderungen, mit denen sich AIED Technologien auseinandersetzen müssen.

**Challenges: Intercultural and Global Dimensions.** Auch wenn in Zukunft es theoretisch möglich ist, jedem Schüler/Student AIED Systeme zukommen zu lassen. Bleibt die Herausforderung, dass es unterschiedliche Kulturen gibt und jede dieser Kulturen auf eine andere Art und Weise das Wissen vermittelt. Dies liegt zu einem an der Sprache, verschiedene Lehrpläne, die pädagogische Kultur. Dies trifft vor allem auf, wenn typische Schüler-Lehrer Interaktionen auftreten, so wie es bei vielen AIED der Fall ist /sein soll.

Daher ist es notwendig, die kulturellen Verschiedenheiten genauer zu erforschen und AIED Systeme genau zu designen und zu implementieren, damit mit den gesammelten Daten nichts Falsches angestellt werden kann.

**Challenges: Practical Impact.** Im Moment gibt es schon Kooperationen und benutzte AIED Systeme, dies muss in Zukunft verstärkt werden. Besonders in Schulen. Dafür wird jedoch Personal in den Schulen benötigt, falls es Probleme mit dem System gibt, oder Personen Hilfe beim Umgang mit diesem brauchen, sowie eine dauerhafte Erreichbarkeit und Benutzbarkeit.

**Challenges: Privacy.** Wenn private Firmen AIED Systeme bereitstellen, spielt die Datenprivatsphäre eine besondere Rolle. Es gibt bereits Gesetze und Regelungen in und zwischen Ländern zur Regelung der Datenprivatsphäre, diese werden in Zukunft noch eine wichtige Rolle spielen. Durch die strikten Regeln wird es auch schwerer, Gewinn mit kommerziellen AIED Systemen zu machen, denn wenn man kein Gewinn mit den Daten machen kann. Werbung kann man schlecht schalten, das stört den Lernfluss. Zu hohe Lizenzkosten können sich einige Schulen und Lehrinstitutionen nicht leisten. Wenn mit den LernDaten Gewinne erzielt werden sollen, müssen transparente und klar definierte und kontrollierte Regeln in Kraft treten.

**Challenges: Interaction Methods.** Es muss genau erforscht werden, wie Menschen mit den Computersystemen umgehen und versuchen festzustellen, wie dies in Zukunft aussehen wird, da sich dies stetig ändert und es nicht bringt, AIED Systeme für heute zu entwickeln, sondern für morgen.

**Challenges: Collaboration at scale:** Derzeit gibt es schon Online Kurse, die weltweit genutzt werden, AIED Systeme sollten daher auch die Möglichkeit haben, weltweit eingesetzt zu werden.

**Challenges: Effectiveness in multiple Domains: ...?? TODO:**

Challenges: Role of AIED in educational technology: hmmm ??

Durch Zugriffsdaten, die von vielen Anwendungen erstellt und gespeichert werden, stehen viele Daten zu Verfügung, die genutzt werden können um das Lehren und das Lernen in einem positiven zu verbessern.

LMS erklären ? oder bekannt ?

Das Paper [Quelle] beschäftigt sich mit den zwei Fragen, ob die Daten, die von Moodle Aktivitäten während der Dauer eines Kurses gesammelt werden ausreichen um Voraussagen zu können, wie gut ein Student in diesem Fach abschließt (Welche Note (nach dem Alphabet) und damit auch ob bestanden oder gefailt) Die zweite Frage, die versucht wird zu beantworten, beschäftigt sich damit herauszufinden, ob die selben Daten, jedoch nur für den Zeitraum von sechs bzw Zehn Wochen ausreichen um herauszufinden ob der Student das Fach bestehen wird oder nicht.

## Datensätze

Um die Fragen zu beantworten wurden Daten von 29 verschiedenen Kursen zu 9 verschiedenen Modulen zwischen den Jahren 2011 und 2018 ausgewählt. Alle Module kommen aus dem selben Fachbereich und dem selben Dozenten um eine homogene Maß an Daten anzusammeln. Insgesamt konnten so Kursdaten von 690 Studenten (410 waren Individuell, rest selber Student) gesammelt werden, wobei 83 davon, den Kurs frühzeitig verlassen haben.

Aufgrund des doch speziellen Datensatzes können die Ergebnisse nur im Context des kleinen Studenten Kreises gesehen werden.

## 2.12 Klassifizierungen

Um die Erste Frage zu beantworten wurde eine KI entwickelt. Diese KI kann insgesamt 5 verschiedene Vorhersagen treffen. "Distinction" (80%+) "Merit" 65%- 79%) "Pass" (50% - 64%) "Fail" (1% bis 49%) "Early Exit" (0% oder keine Note)

Um die Zweite Frage zu beantworten wurden die ersten 3 Klassen zur Klasse "Pass" zusammengefasst und die letzten beiden zu "Fail"

## 2.13 Daten

Es wurden nur Zugriffsdaten von Studenten zum Trainieren und Validieren der KI genommen, die von Dozenten mussten entfernt werden Es wurden folgende Zugriffsdaten verwendet: LISTE im PAPER!!!

## 2.14 Algorithmen

Genauer Beschreiben ? Random Forest: Gradient Boosting: K nearest Neighbours: Linear Discriminant Analysis:

Dies sind verschiedene Algorithmen die benutzt werden können um Daten in Klassen zu Ordnen.

## 2.15 Model Building an Evaluation

Data-Split in 70-30

Kappa Statistic gibt an wie gut die vom Modell vorausgesagten Klassen mit den richtigen Klassen übereinstimmen, zur Kontrolle wird ein zufälliger Klassifizierer, der anhand der Häufigkeit der Vorkommen von Klassen ratet um welche Klasse es sich handelt genommen. 0.02 = Fair 0.21-0.40. Moderat 0.41-0.60. Zustimmend 0.61-0.80. Wesentliche Zustimmung 0.81-1.0. Perfekte Zustimmung.

no-information Rate = Genauigkeit wenn immer die Klasse mit den meisten Labels in dem Testset geraten wird

## 2.16 Ergebnisse

Das Training und Testen der KI hat ergeben, dass alle Algorithmen eine sehr ähnliche Genauigkeit haben. Die Genauigkeit liegt bei dem Modell zu 95% in dem Bereich 53% bis 67%. Des Weiteren waren alle Algorithmen besser als die no-information Rate. Am besten konnte vorausgesagt werden, wenn ein Student besonders gut den Kurs besteht oder den Kurs frühzeitig beendet. Alles andere konnte nicht gut predicted werden.

Wenn anhand von den Zugriffsdaten für die gesamte Länge des Kurses, konnte, wenn nur festgestellt werden soll ob ein Student besteht oder nicht, mithilfe des Random Forest Algorithmus wurde eine Genauigkeit von 92% erreicht (148 von 152 Studenten die bestanden haben und 42 von 54 die nicht bestanden haben konnten korrekt vorhergesagt werden).

### 2.16.1 Ergebnisse mit 6 Wochen Daten und 10 Wochen Daten

Wenn die KI nur mit 6 Wochen alten Daten voraussagen treffen soll, hat diese nicht besser abgeschnitten als die no-information rate. Wurden jedoch 10 Wochen alte Ergebnisse genommen, konnte die deutlich bessere vorherzusagen treffen als die no-information rate. Auch hier waren alle Algorithmen besser, am besten war jedoch der Random Forest Algorithmus mit 82.18%.

## 2.17 Korrelation der Variablen

Pass/Fail all Data: Regular: und Assignment Submitted

Random Forest /Pass/Fail 10 Weeks Regular / Assignment Views Course Views  
pm.Early am.late

## 2.18 Diskussion

Frage 1: Wenn alle Daten benutzt werden ist es möglich die Notensignifikant besser zu bestimmen als mit der no-information-rate. Am besten war dabei der Random Forest mit einer Genauigkeit von 60,5. Dies deutet darauf hin, dass die Moodle Aktivität benutzt werden kann um die Noten vorherzusagen. Dabei war die Unterscheidung, ob

ein Student sehr gut besteht oder abbricht am besten zu bestimmen, das kann daran liegen, das dort das Nutzerverhalten extrem eindeutig ist. Wenn man nur bestimmen möchte ob die Studenten bestehen oder nicht können die Daten sehr gut genutzt werden. Dort hat man eine Genauigkeit von 92.2% Dieses Ergebnis stimmt auch mit vielen anderen Studien überein.

Das Gute Ergebnis kann daran liegen, das durch den längeren Zeitraum mehr Datenvorhanden waren. Im Gegensatz zu anderen Studien waren hier weniger als 83 Prozent der Daten passiv sonst sind das oft über 95 Prozent.

Frage 2: Für den kurzen Zeitraum von 6 Wochen hat keiner einen Algorithmus eine Verbesserung gebracht. Mit Daten von 10 Wochen hat die Performance etwas verbessert. Jedoch auch hier kann nur weniger als die Hälfte der Failing Student vorhergesagt werden. Dies deutet darauf hin, das LMS Zugriffsdaten hilfreich für ein Frühwarnsystem sein können, Jedoch werden noch mehr Daten benötigt z. B. eine noch genauere LMS Nutzungsdaten oder Assessment results, Studentische Interaktion mit anderen. Dies wurde auch in anderen Studien schon gezeigt, das diese Daten helfen können.

Trotz der hohen false-positiven kann es hilfreich sein so ein Frühwarnsystem einzurichten, da dadurch trotzdem einige Studenten erreicht, die dadurch Hilfe bekommen können und das die Schulen und den Studenten weniger kostet, als wenn diese einfach am Ende durchfallen.

Es wurde gezeigt, das das Reguläre Login darauf hinweist, das ein Student eher besteht. Viele Forscher meinen dies zeigt darauf, das die Studenten ein besseres Lern/Zeit Management haben und dies ein wichtiger Faktor für den Erfolg ist.

Um die Modelle zu verbessern könnten feinere Zugriffsdaten gesammelt werden, sowie Ergebnisse von Aufgaben beachtet werden. Dozenten sollten Aufgaben früh in ihren Kurs einbauen und die Möglichkeiten von LMS nutzen um Studentische Interaktivität zu fördern.

## 2.19 Ergebnis

Moodle Data kann benutzt werden um vorherzusagen, wie gut ein Student besteht und ob dieser besteht. Um dies jedoch praktisch zu benutzen, wäre ein Frühwarnsystem gut, Moodle Daten reichen dazu nicht ausreichen aus. Die getesteten Modelle sollten daher mit mehr Daten, genaueren Zugriffsdaten und Aufgaben ergebnisse miteinbezogen werden. Des Weiteren waren alle Zugriffsdaten nur von Kursen die von einem Dozenten geleitet wurden. Daher ist das Ergebnis nicht gut Generalisierbar

*[https://diid.hhu.de/wp-content/uploads/2019/10/DIID-Precis\\_Kieslich-et-al\\_Fin.pdf](https://diid.hhu.de/wp-content/uploads/2019/10/DIID-Precis_Kieslich-et-al_Fin.pdf)*

Fürs Fazit wie viele Leute KI in welchen Bereichen akzeptieren - Keiner will Dropout-detection

# Kapitel 3

## Einleitung

„The user’s going to pick dancing  
pigs over security every time.“

BRUCE SCHNEIER  
(\*1963)

Das Verfassen einer eigenständigen Ausarbeitung ...

### 3.1 Exposé

Bevor eine Ausarbeitung begonnen werden kann, MÜSSEN sich Kandidat/-in und Prüfer auf eine Aufgabenstellung einigen. Als Grundlage dienen hier eigene Ideen des Kandidaten / der Kandidatin und Vorschläge des Prüfers. Diese Aufgabenstellung MUSS vom Kandidaten/von der Kandidatin in ein Exposé überführt werden, welches auf Basis der vorliegenden L<sup>A</sup>T<sub>E</sub>X-Vorlage erstellt und folgende Form / Struktur haben MUSS:

**Kontext und Motivation** Eine inhaltliche Einleitung in das Themengebiet der Arbeit mit Referenzierung der wichtigsten Literatur sowie eine Motivation, z. B. durch Aufarbeitung von Literatur getrieben.

**Ziele** Aufzählung und Kurzbeschreibung konkreter Ziele (im Sinne einer Spezifikation)

**Vorgehen** Eine Beschreibung wie die einzelnen Ziele erreicht werden sollen (im Sinne einer Implementierung) und wie die Zielerreichung validiert werden soll.

**Projektplan** Ein leichtgewichtiger und realistischer Projektplan basierend auf Meilensteinen und untergeordneten Aufgaben, welcher nach Finalisierung des Exposé durch den Kandidaten/die Kandidatin in Gitlab (siehe Abschnitt 6.1) eingepflegt werden MUSS.

Das Exposé MUSS explizit (Email, direkte mündliche Absprache, etc.) vom Prüfer akzeptiert werden.

Das Exposé dient im Anschluss als Grundlage für die Einleitung der Ausarbeitung. Dabei sollen explizit die Abschnitte Kontext und Motivation sowie Ziele in der

Einleitung in möglicherweise überarbeiteter Fassung übernommen werden. Zudem MUSS eine Einleitung eine Erläuterung des weiteren Aufbaus der Arbeit beinhalten.

## 3.2 Hinweise

Bitte denken Sie daran, dass Sie die eidesstattliche Erklärung vor Abgabe unterschreiben.

## 3.3 Inhalt

Die Arbeit MUSS – neben dem Hauptteil – nachfolgende Inhalte berücksichtigen:

- Titelseite
- Eidesstattliche Erklärung
- Zusammenfassung und Abstract (Englisch)
- Inhaltsverzeichnis, Abbildungsverzeichnis, Tabellenverzeichnis, Abkürzungsverzeichnis und Literaturverzeichnis
- Einleitung (siehe Abschnitt 3.1)
- Aufarbeitung verwandter und relevanter Literatur unter Angabe der konkreten Vorgehensweise bei der Literaturrecherche
- Kritische Betrachtung der eigenen Arbeit
- Fazit bestehend aus einer reflektierten Zusammenfassung und einem Ausblick

Weiterhin MÜSSEN, falls anwendbar, vom Prüfer vorgegebene Richtlinien für Coding-Style und Code-Dokumentation sowie Gestaltung eingehalten werden.

## 3.4 Organisatorisches

- Es gilt die jeweils aktuelle Prüfungsordnung (§15 in BMI PO vom 04.08.2010 bzw. §15 in MMI PO vom 16.06.2011). Lesen Sie aufmerksam die für Sie geltende Prüfungsordnung und richten Sie sich nach den dort definierten Vorgaben (es sei denn Sie haben mit dem Prüfer eine Abweichung abgesprochen).
- Abzugeben gebunden als Ausdruck (beidseitig bedruckt) und elektronisch als PDF

## 3.5 Bewertungskriterien

Die Bewertung einer Arbeit erfolgt unter anderem auf Grundlage von **Schwierigkeitsgrad**, **wissenschaftlicher Arbeitstechnik**, **ingenieurmäßiger Vorgehensweise**, **Stil** und **Form**.



# Kapitel 4

## Stile

„The wise know their weakness  
too well to assume infallibility;  
and he who knows most, knows  
best how little he knows.“

Thomas Jefferson  
(1743–1826)

Nachfolgend sind einige Beispiele zum Styling von Inhalten aufgeführt. Eine gute Einführung in das Arbeiten mit  $\text{\LaTeX}$  bietet die Ausarbeitung<sup>1</sup> von Jürgens und Feuerstack der FernUniversität in Hagen.

### 4.1 Text

Dies ist ein Beispiel für *kursiven* und **fetten** Text.

Abkürzungen werden in der Datei `acronyms.tex` definiert und können dann vereinfacht genutzt werden. Alle tatsächlich eingesetzten Abkürzungen werden automatisch im Abkürzungsverzeichnis aufgeführt. Eine Abkürzung wird bei der ersten Verwendung zusätzlich ausgeschreiben dargestellt. Ein Beispiel: Das Bundesamt für Sicherheit in der Informationstechnik (BSI) stellte fest ... und weiterhin beobachtet das BSI ...

### 4.2 Abbildungen und Tabellen

Abbildung 4.1 zeigt eine einfache Abbildung.

Tabelle 4.1 zeigt eine einfache Tabelle.

---

<sup>1</sup>[https://www.fernuni-hagen.de/imperia/md/content/zmi\\_2010/a026\\_latex\\_einf.pdf](https://www.fernuni-hagen.de/imperia/md/content/zmi_2010/a026_latex_einf.pdf), aufgerufen am 16.06.2017



Abbildung 4.1: Logo Hochschule Düsseldorf

Eins	1
Zwei	2

Tabelle 4.1: Beispieltabelle

## 4.3 Zitieren

Die benötigte Literatur wird in der Datei `literatur.bib` gepflegt. Das Literaturverzeichnis wird automatisch generiert. Dies ist ein Zitat von (Eckert, 2014) ... laut Eckert, 2014, Seite 42 ist dieses Vorgehen empfehlenswert.

Der Zitierstil MUSS nach APA (American Psychological Association)<sup>2</sup> Style erfolgen.

## 4.4 Listen

Unsortierte Liste:

- Eins
- Zwei
- Drei

Nummerierte Liste:

1. Element
2. Element
3. Element

---

<sup>2</sup><http://www.apastyle.org/>, abgerufen am 16.06.2017

# Kapitel 5

## Tools

„Man is still the most  
extraordinary computer of all.“

John F. Kennedy  
(1917–1963)

Nachfolgende Hinweise und Empfehlungen zum Einsatz von Tools vereinfachen den Umgang mit  $\text{\LaTeX}$ .

### 5.1 $\text{\LaTeX}$

Das Verfassen von Dokumenten mit  $\text{\LaTeX}$  kann durch unterschiedlichste Tools unterstützt werden. Da  $\text{\LaTeX}$  grundsätzlich textbasiert arbeitet können jegliche Inhalte auch in einem einfachen Texteditor erstellt und angepasst werden.

Mittels unterschiedlichster Editoren kann die Erstellung und Pflege von Dokumenten mit  $\text{\LaTeX}$  vereinfacht werden. Unter Linux bietet der Editor „Kile“<sup>1</sup> eine Vielzahl nützlicher Funktionen. Für Apple OS X und Microsoft Windows ist „Texmaker“<sup>2</sup> empfehlenswert. Unbedingt empfohlen wird die Verwendung einer Rechtschreibprüfung, die typischerweise in den Editoren integriert sind.

---

<sup>1</sup><http://kile.sourceforge.net/>, aufgerufen am 16.06.2017

<sup>2</sup><http://www.xmlmath.net/texmaker/>, aufgerufen am 16.06.2017

# Kapitel 6

## Infrastruktur

„The secret of all victory lies in  
the organization of the  
non-obvious.“

Marcus Aurelius  
(121–180)

Die Erstellung einer Arbeit sollte in einer bereitgestellten Infrastruktur erfolgen, die insbesondere bei der Planung und Verwaltung einer Arbeit unterstützt.

### 6.1 GitLab-Server

Unterschiedliche Dienste, die im Kontext einer Arbeit von Nutzen sind, werden über einen GitLab-Server bereitgestellt. Jeder Kandidat erhält einen persönlichen Zugang und ein eigenes Repository. In diesem Repository werden Dokumente und eigene Inhalte der Arbeit zentral verwaltet und somit dem Betreuer zur Kontrolle übergeben. Der GitLab-Server stellt hierzu in erster Linie ein Repository bereit. Das Repository bzw. eine Versionsverwaltung im Allgemeinen hilft vor allem bei der Verwaltung von textbasierten Dateien, so z. B. Quellcode oder Dokumente in  $\text{\LaTeX}$ .

Zur Planung einer Arbeit und Kontrolle des Fortschritts erfolgt das Projektmanagement digital innerhalb von GitLab. Hierzu werden Milestones und Issues angelegt und während des Projektes gepflegt bzw. Fortschritte kontrolliert. Eine möglichst präzise Projektplanung hilft bei der Vermeidung von etwaigen zeitlichen Engpässen im Laufe der Erstellung einer Arbeit.

### 6.2 Versionsverwaltung mit Git

Die Verwaltung der Arbeit, die mittels  $\text{\LaTeX}$  verfasst wird, und aller zugehörigen Dateien bzw. Dokumente kann auf einfache und sehr transparente Weise mittels einer Versionsverwaltung erfolgen. Als Versionsverwaltung wird Git<sup>1</sup> eingesetzt. Git steht für alle gängigen Betriebssysteme bereit.

---

<sup>1</sup><https://git-scm.com/>, aufgerufen am 16.06.2017

Jegliche Änderungen und Ergänzungen werden von Git erkannt und aufgezeichnet. Erfolgte Änderungen sollten mittels sog. „Commits“ eingepflegt und beschrieben werden. Die Versionsverwaltung erfolgt in erster Linie auf dem lokalen System. Erfolgte Änderungen bzw. Fortschritte sollten – nicht nur als Backup – regelmäßig über den bereitgestellten GitLab-Server dem Betreuer zur Verfügung gestellt werden.

Die Arbeit mit Git kann sowohl auf der Kommandozeile als auch in Applikation mit UI erfolgen. Die Applikation „SourceTree“<sup>2</sup> ermöglicht beispielsweise die komfortable Verwaltung von Git-Repositories.

Grundsätzliche Tipps zum Umgang mit Git liefern die offizielle Dokumentation<sup>3</sup> und das „Git Cheat Sheet“<sup>4</sup>.

Gerade im Zusammenhang mit L<sup>A</sup>T<sub>E</sub>X entstehen viele temporäre Dateien, die nicht in der Versionsverwaltung landen sollten. Dazu sollte eine `gitignore` Konfiguration<sup>5</sup> erstellt werden.

---

<sup>2</sup><https://www.sourcetreeapp.com/>, aufgerufen am 16.06.2017

<sup>3</sup><https://git-scm.com/doc>, aufgerufen am 16.06.2017

<sup>4</sup><https://www.git-tower.com/blog/git-cheat-sheet/>, aufgerufen am 16.06.2017

<sup>5</sup><https://www.gitignore.io/>, aufgerufen am 04.07.2017

# Anhang A

## Tipps zu häufig gemachten Fehlern

### A.1 Abbildungen, Tabellen, Listings, etc.

1. Die Schriftgröße von Text in Abbildungen muss sich nach der Schriftgröße des regulären Textes richten.
2. Alle Abbildungen, Tabellen, Listings, etc. sind mit einer Beschriftung und Nummerierung zu versehen. Im Text muss mit Hilfe der Nummerierung auf die jeweilige Abbildung, Tabelle bzw. das Listing, etc. verwiesen und eine Erläuterung der Abbildung, Tabelle bzw. des Listings verfasst werden.

### A.2 Text

1. Es muss konsistent aus „Wir“ oder „Man“ Perspektive geschrieben werden.
2. Abkürzungen werden einmalig wie in Abschnitt 4.1 beschrieben eingeführt und verwendet.
3. Fachbegriffe müssen eingeführt und definiert werden. Der Fachbegriff kann z.B. einmal *kursiv* gedruckt und danach normal geschrieben werden. Für die Definition und Erklärung sollte einschlägige Literatur verwendet werden.
4. Es muss eine Rechtschreib- und Grammatikprüfung verwendet werden.
5. Es sollte Korrektur durch Dritte durchgeführt werden.
6. Es muss Groß-/Kleinschreibung im Literaturverzeichnis beachtet werden.
7. Es müssen Deutsche Anführungsstriche verwendet werden: „...“

### A.3 Diverses

1. Wenn es sich bei der Arbeit um einen Angriff dreht, dann muss (am Besten am Beginn der Arbeit) die Hackerethik zusammenfassend beschrieben und dabei konkret auf den Angriff bezogen werden.
2. Internetquellen sollen nicht in das Literaturverzeichnis, sondern über eine Fußnote unter Angabe der URL und dem letzten Abrufdatum dokumentiert werden.

# Literatur

Eckert, C. (2014). *IT-Sicherheit: Konzepte - Verfahren - Protokolle* (9. Aufl.). De Gruyter Oldenbourg. <https://google.com>